Etude et utilisation des normes vectorielles en analyse numérique linéaire (1ere thèse) – Inéquations variationnelles et problèmes aux limites (2eme thèse)

François Robert

N° d'ordre

THESES

présentées à

LA FACULTE DES SCIENCES DE L'UNIVERSITE DE GRENOBLE

pour obtenir

LE GRADE DE DOCTEUR ES SCIENCES MATHEMATIQUES

par

François ROBERT

Première thèse :

Etude et utilisation de normes vectorielles en analyse numérique linéaire

Deuxième thèse :

Inéquations variationnelles

Thèses soutenues le 12 novembre 1968 devant la Commission d'Examen :

Monsieur	KUNTZMANN	Président
Monsieur	LIONS	Examinateur
Monsieur	GASTINEL	Examinateur
Monsieur	LAURENT	Examinateur

D'un charlatan

Un charlatan disait en plein marché
Qu'il montrerait le diable à tout le monde;
Si n'y eut nul, tant fut il empêché
Qui ne courût pour voir l'esprit immonde.
Lors une bourse assez large et profonde
Il leur déploie, et leur dit "Gens de bien,
Ouvrez vos yeux! Voyez! Y a t'il rien?
"Non", dit quelqu'un des plus près regardants.
"Et c'est, dit il, le diable, voyez vous bien?
Ouvrir sa bourse et ne rien voir dedans".

Mellin de Saint Gelais
1487-1558

... Est temps désormais
que crie à toutes gens mercis

(Villon)

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

Que Messieurs les Professeurs KUNTZMANN et GASTINEL veuillent bien trouver ici un témoignage de ma reconnaissance, pour l'existence même de cet Institut, pour les recherches qu'ils y développent, pour ce qu'ils donnent à leurs chercheurs en leur offrant un cadre de travail et une base de départ très favorables.

Monsieur KUNTZMANN a bien voulu accepter de présider le Jury : je lui exprime ma très respectueuse gratitude.

Monsieur GASTINEL m'a proposé le sujet de cette étude, qu'il a constamment soutenue de son intérêt, de sa réflexion et de ses conseils. Travailler sous sa direction, c'est aussi prendre part à la vie, à l'amitié, au cheminement de l'équipe qu'il anime par des discussions, suggestions, tentatives quotidiennes. Ma très déferente gratitude saura lui être fidèle.

Je suis très sensible à l'honneur que me fait Monsieur le Professeur LIONS en acceptant de venir à GRENOBLE faire partie du Jury . Il a bien voulu examiner ce travail et me proposer le second sujet. Je lui exprime ici ma reconnaissance.

Je voudrais aussi remercier Monsieur LAURENT, ainsi que Monsieur BERTRANDIAS, pour l'attention et les suggestions apportées à une partie de ce travail.

En outre, je témoigne de l'amitié qui me relie aux membres du Laboratoire, spécialement à ceux de l'équipe d'analyse numérique.

J'adresse enfin mes vifs remerciements à toutes les personnes ayant contribué à la réalisation matérielle de cette thèse, particulièrement à Madame DUSART, du Laboratoire de Calcul de Lille, pour l'attention et le soin apportés à la frappe des stencils.

INTRODUCTION

L'idée semble intéressante, qui consiste à munir un espace vectoriel d'une "norme" à valeurs dans le cône positif d'un espace vectoriel ordonné et non plus nécessairement dans $\mathbb{R}_+$; d'où l'expression, lourde mais parlante, de "norme vectorielle", que nous avons adoptée.

Cette idée n'est pas nouvelle, et le chapitre 12 de l'ouvrage de KANTOROVITCH, VULIK, PINSKER "Analyse fonctionnelle dans les espaces ordonnés (en russe; non traduit; Moscou 1950) est consacré à cette question, dans un contexte général d'analyse fonctionnelle. Sur un plan plus appliqué d'analyse numérique, plusieurs auteurs, tels FIEDLER et PTĀK, OSTROWSKI, SCHROEDER, FEINGOLD et VARGA (voir la bibliographie) utilisent, plus ou moins explicitement, certaines normes vectorielles sur $\mathbb{R}^n$. A un degré intermédiaire de généralité se place un cours de F.L. BAUER à l'Université de Stanford : "Theory of norms" (Août 1967).

Les deux pôles du présent travail peuvent être ainsi situés :

* Tout d'abord, une étude "à priori" de la notion de norme vectorielle (chapitres 1 à 4) de façon à préparer "proprement" le matériel pour la seconde partie. En particulier, nous introduisons la notion, essentielle, de norme vectorielle régulière.

** En second lieu, l'application effective de l'instrument ainsi élaboré à différentes questions d'analyse numérique linéaire (chapitres 5 à 7 et Annexe). Cette exploitation s'est voulue systématique, dans le sens d'une recherche, pour tel ou tel résultat, des hypothèses "les plus faibles" le conservant.

On trouvera, en tête de chaque chapitre, un résumé assez détaillé de son contenu. Profitons en pour n'indiquer ici que les lignes directrices et les résultats principaux de ce travail.

* Première partie

1) Une observation féconde est la suivante : même sur $\mathbb{R}^n$, et contrairement à ce qui se passe dans le cas des normes habituelles, il existe des normes vectorielles ne permettant pas de générer une norme vectorielle correspondante sur l'espace des matrices (n,n) (considérées comme opérateurs linéaires sur $\mathbb{R}^n$) : on peut exhiber de très simples contre exemples.

Ceci nous a amenés à dégager la notion de norme vectorielle régulière en caractérisant, par une propriété de somme directe de certains sous-espaces, la classe des normes vectorielles sur $\mathbb{R}^n$ ou $\mathbb{C}^n$ qui permettent de construire une norme vectorielle de matrices. Ainsi caractérisées, les normes vectorielles régulières sont donc, de fait, les seules utiles pour les applications. Effectivement on vérifie à postériori que les normes vectorielles utilisées par les auteurs cités plus haut sont toujours régulières.

2) Cette notion dégagée pour les espaces de dimension finie, se posait la question de savoir jusqu'à quel degré de généralité on pouvait étendre la notion de norme vectorielle sans perdre celle de norme vectorielle régulière. Sans prétendre répondre de façon exhaustive, nous montrons que l'on peut pratiquer cette extension en imposant à l'espace (de Riesz) des valeurs de la norme vectorielle d'être <u>archimédien</u>, <u>unitaire</u>, et en fait un peu plus, ce qui nous permet- c'est décisif,- d'utiliser la représentation de Yosida de tels espaces. La conclusion suivante éclaire quelque peu le problème :

Soit X l'espace vectoriel que nous munissons d'une norme vectorielle p. N'assujettir p à prendre ses valeurs que dans **un** $\mathbb{R}^I$ (où I est un ensemble quelconque), ordonné "naturellement", est <u>équivalent</u>

à définir sur $\mathfrak{X}$ une topologie localement convexe,ce qui est, pour nous, trop général. En effet, pour conserver la notion de norme vectorielle régulière, nous avons été amenés à restreindre l'espace des valeurs à être un $\mathfrak{F}(I)$, auquel est isomorphe, ainsi que nous le prouvons, tout espace dit "de Yosida régulier". Ce qui nous a incité à limiter ici la notion de norme vectorielle à celle de "norme à valeur dans un espace de Yosida régulier".

Ces deux questions qui font l'objet des 4 premiers chapitres, sont naturellement présentées dans l'ordre inverse :

La notion d'espace de Yosida régulier est précisée au chapitre 1; au chapitre 2 est définie, au plan de généralité situé ci-dessus, la notion de norme vectorielle, avec mise en évidence de la classe des normes vectorielles régulières; cependant qu'au chapitre 3 sont reformulés et complétés dans le détail l'ensemble de ces résultats dans un contexte d'espaces de dimension finie. Le chapitre 4 est consacré aux normes vectorielles de matrices et les résultats établis seront utilisés constamment dans la seconde partie; en particulier celui-ci : si M est la norme vectorielle de matrice générée par une norme vectorielle régulière de vecteurs, M(A) est toujours de rayon spectral supérieur ou égal à celui de A.

** Deuxième partie. Applications

Nous nous occupons principalement d'utiliser des normes vectorielles dans l'étude de conditions suffisantes de convergence pour des itérations linéaires. Le fait, pour une matrice d'itération, d'être <u>contractante relativement à une norme vectorielle donnée</u> est une situation caractéristique d'un point de vue numérique, et qui entraîne, outre la convergence de l'itération, la possibilité de contrôler effectivement, en <u>norme vectorielle</u>, l'approximation obtenue sur la solution recherchée. Ces questions font l'objet du chapitre 5.

Au chapitre 6, la <u>caractérisation</u> des matrices admettant, relativement à une norme vectorielle donnée, une <u>M-minorante</u> sert de définition à la classe des <u>blocs-H-matrices</u> relativement à cette norme vectorielle; classe ainsi nommée car elle redonne, dans le cas de la norme vectorielle type $p(x) = |x|$ (dite de Schroeder), la classe des H-matrices définie par A. OSTROWSKI. On pense ainsi avoir mis, dans leur "contexte naturel" différents résultats partiels que nous complétons, établis antérieurement par OSTROWSKI, FIEDLER et PTAK, FEINGOLD et VARGA.

Les blocs G-matrices et les matrices à bloc-diagonale dominante forment alors des sous-classes successives de celle des blocs-H-matrices. Cette dernière classe semble être, de fait, "la bonne classe" de matrices pour l'emploi d'une norme vectorielle donnée. On peut la caractériser de plusieurs manières; et tout d'abord par le fait qu'une certaine estimation en norme vectorielle de l'approximation obtenue sur la solution d'un système linéaire est bornée (donc utilisable) si et seulement si la matrice du système est une bloc-H-matrice.

Une première tentative d'extension de ces résultats a donné lieu à la notion de B-matrices et à la recherche de B-minorantes d'une matrice donnée. Une seconde tentative, utilisant les résultats du chapitre 7, est donnée en Annexe.

De fait, l'utilisation de normes vectorielles régulières constitue un moyen d'étude particulièrement commode des méthodes itératives classiques <u>par blocs</u> (Jacobi, Gauss-Seidel, sur-relaxation); cela fait l'objet du chapitre 7 où se trouve élaboré un résultat aussi général que possible (Th. 5) montrant qu'une certaine condition, numériquement exploitable, entraîne pour les matrices d'itération par blocs correspondantes, le fait d'être contractantes. Nous donnons diverses formulations particulières de ce théorème, dont l'une, qui s'énonce ainsi :

"Si A est une bloc-H-matrice, les matrices par bloc correspondantes de Jacobi et Gauss-Seidel sont convergentes",

est l'exacte transcription au cas des blocs d'un résultat connu antérieurement dans le cas des H-matrices et des méthodes par point. La remarque qui sert de point de départ à la démonstration du Théorème 5 et que certaines majorantes des matrices d'itération (par blocs) de Jacobi et Gauss-Seidel sont elles-mêmes <u>des matrices de Jacobi et Gauss-Seidel par point</u>. On peut alors leur appliquer (parce qu'elles sont non négatives) le théorème de Stein Rosenberg. On peut donner, de façon analogue, un résultat de convergence relatif à la sur-relaxation.

En conclusion, on peut retenir que la notion de norme vectorielle constitue un outil efficace et commode de l'analyse numérique, spécialement dans l'évaluation de bornes d'erreur. D'autres chapitres de l'analyse numérique sont certainement justiciables d'un traitement utilisant des normes vectorielles. Néammoins l'outil mis au point est particulièrement bien adapté au maniement des opérateurs linéaires décomposés en blocs.

*

* *

CHAPITRE 1 : ESPACES de YOSIDA REGULIERS

Le but de ce chapitre est de définir et d'étudier une classe d'espaces vectoriels ordonnés que nous nommons "espaces de Yosida réguliers".

Après quelques rappels de terminologie sur les espaces vectoriels ordonnés (§1-1) le paragraphe (1-2) regroupe les théorèmes fondamentaux sur la question : et en particulier le théorème de représentation de Yosida qui construit un isomorphisme de tout espace de Riesz archimédien et unitaire Y (espace dit "de Yosida") dans l'espace $\mathcal{B}(\mathcal{M})$ des fonctions réelles bornées définies sur l'ensemble $\mathcal{M}$ des idéaux maximaux de Y.

Au paragraphe (1-3) on définit, parmi les espaces de Yosida complètement réticulés, la classe des espaces de Yosida réguliers. On peut prouver (Proposition 1-3-10) que pour qu'un espace de Yosida complètement réticulé soit identifiable avec $\mathcal{B}(\mathcal{M})$ (i.e : pour que l'isomorphisme de Yosida soit surjectif) il faut et il suffit qu'il soit régulier.

La classe des espace de Yosida réguliers sera utilisée au chapitre 2 pour préciser la notion de norme vectorielle.

1- 1 - NOTATIONS et RAPPELS

Soit Y un espace de Riesz ([8],[57],[67]) : c'est un espace vectoriel sur $\mathbb{R}$,

* muni d'une relation d'ordre, que l'on notera $\leqslant$, compatible avec la structure d'espace vectoriel :

$$x \leqslant y \Rightarrow \forall \lambda > 0 \quad \lambda x \leqslant \lambda y$$
$$\Rightarrow \forall z \in y \quad x + z \leqslant y + z$$

* réticulé : pour tout couple (x,y) d'éléments de Y,

Sup(x,y) existe (le plus petit des majorants de $\{x,y\}$)

Inf(x,y) existe (le plus grand des minorants de $\{x,y\}$)

On posera, avec des notations usuelles :

$$x^+ = \mathrm{Sup}(x,0)$$
$$x^- = \mathrm{Sup}(-x,0) = (-x)^+$$
$$|x| = x^+ + x^- \qquad \text{d'où}$$
$$x = x^+ - x^-$$
$$|x| = \mathrm{Sup}(x,-x)$$

Remarque : L'application $x \longrightarrow p(x) = |x|$ de Y dans lui-même vérifie les axiomes :

i) $p(x) = 0 \Rightarrow x = 0$

ii) $p(\lambda x) = |\lambda|\, p(x) \quad \forall \lambda \in \mathbb{R},\ \forall x \in Y$

iii) $p(x+y) \leqslant p(x) + p(y) \quad \forall x, y \in Y$

C'est l' exemple-type d'une norme vectorielle.

Y est dit <u>complètement réticulé</u> si toute partie majorée a une borne supérieure. Alors toute partie minorée a une borne inférieure.

1- 1- 1 - Espace de Riesz unitaire [67]

L'espace de Riesz est dit <u>unitaire</u> s'il existe e dans Y ($e \geqslant 0$) tel que : pour tout x dans Y, il existe un nombre réel positif α_x tel que :

$$-\alpha_x e \leqslant x \leqslant \alpha_x e$$

e est dit unité. Si e est unité, μe $(0 < \mu \in \mathbb{R})$ est également unité

Exemples - Voir[8] . On rappelle que

* $\mathcal{B}([01])$ espace des fonctions numériques bornées définies sur $[01]$ est un espace de Riesz complètement réticulé et unitaire pour la relation d'ordre :

  $f \leq g$ si $f(\xi) \leq g(\xi)$ pour tout $\xi \in [01]$

* Par contre, $\mathcal{C}([01])$, sous-espace du précédent formé par les fonctions continues sur $[01]$ est, pour la relation d'ordre induite, un espace de Riesz unitaire mais qui <u>n'est pas complètement réticulé</u> ([8], p.21)

* l'espace, noté ℓ^{∞}, des suites réelles bornées est un espace de Riesz complètement réticulé et unitaire pour la relation d'ordre habituelle :

  $x = (\xi_i)_{i \in \mathbb{N}}$; $y = (\eta_i)_{i \in \mathbb{N}}$

  $x \leq y$ si, pour tout i dans $\mathbb{N}$, $\xi_i \leq \eta_i$

1- 1- 2 - Eléments étrangers [8]

Deux éléments x et y d'un espace de Riesz Y sont dits <u>étrangers</u> si

$$\mathrm{Inf}\,(|x|, |y|) = 0$$

On notera $A^{\perp}$ l'ensemble des éléments de Y étrangers à tous les éléments d'une partie donnée A de Y.

1- 1- 3 - Eléments nilpotents (dans un espace de Riesz Y unitaire) [67]

x appartenant à l'espace de Riesz unitaire Y est dit <u>nilpotent</u> si :

$$\forall n \in \mathbb{N} \qquad n|x| \leq e$$

L'ensemble $\mathcal{R}$ des éléments nilpotents de Y est appelé le <u>radical</u> de Y (c'est un sous-espace vectoriel).

1- 1- 4 - Espace de Riesz archimédien [67]

L'espace de Riesz Y est dit <u>Archimédien</u> si :

$$[\forall n \quad nx \leq y] \Rightarrow x \leq 0$$

Exemples : les espaces $\mathcal{F}$ ([01], $\mathcal{C}$ ([01]), ℓ^∞ introduits plus haut sont archimédiens. De même ^{n}, ordonné "composante à composante" est un espace de Riesz, complètement réticulé, unitaire et archimédien.

Contre exemple : il est bien connu, par contre que $\mathbb{R}^n$, ordonné lexicographiquement, est un espace de Riesz qui n'est pas archimédien pour $n \geqslant 2$

Pour $\mathbb{R}^2$

$$(x_1, x_2) \leqslant (y_1, y_2) \underset{df}{\Longleftrightarrow} [x_1 < y_1] \cup ([x_1 = y_1] \cap [x_2 \leqslant y_2])$$

Soit $x = (0, 1)$; $y = (1,0)$. On a, pour tout $n \in \mathbb{N}$

$$(0\ n) = nx \leqslant y = (1\ 0) \text{ et cependant } x \text{ n'est pas } \leqslant 0$$

Remarque * pour un espace de Riesz unitaire :
Archimédien $\Longleftrightarrow \mathcal{R} = \{0\}$ (radical réduit à zéro).

1- 1- 5 - Sous-espace vectoriel épais (ou : idéal) d'un espace de Riesz Y.

On appelle ainsi tout sous-espace vectoriel J de Y tel que :

$$y \in J \qquad |x| \leqslant |y| \quad \Rightarrow \quad x \in J$$

En particulier $|x|$ appartient à J dès que x appartient à J.

Exemples : * Le radical d'un espace de Riesz unitaire est un idéal.

* Dans $\mathcal{B}$(01) ou $\mathcal{C}$(01) tout idéal est formé de l'ensemble des fonctions s'annulant sur une partie donnée de [01].
* Dans $\mathbb{R}^k$, ordonné "composante à composante", l'ensemble des vecteurs ayant une ou plusieurs composantes d'indice fixé nulles forme un idéal.
* Dans l'espace ℓ^∞, le sous-espace J des suites convergeant vers zéro forme un idéal. Soit Λ_1 une sous-suite de $\mathbb{N}$; alors le sous-espace J_{Λ_1} des suites dont la restriction à Λ_1 converge vers zéro est un idéal contenant J. Prenant Λ_2 sous-suite de Λ_1, Λ_{r+1} sous-suite de Λ_r on forme donc une suite d'idéaux croissante strictement.

$$J \subset J_{\Lambda 1} \subset J_{\Lambda 2} \quad \cdots\cdot \qquad \subset J_{\Lambda r} \subset J_{\Lambda r+1} \quad \cdots\cdots$$

Idéaux triviaux de Y

Ce sont $\{0\}$ et Y lui-même.

Dans un espace de Riesz unitaire Y, le seul idéal contenant l'unité est l'idéal trivial Y.

1- 1- 6 - Espace de Riesz simple

Si les seuls idéaux sont les idéaux triviaux. Tout espace de Riesz simple Y a, s'il est unitaire, son radical réduit à zéro.

Un idéal non trivial est dit maximal s'il n'est contenu dans aucun autre idéal que lui-même et l'espace tout entier.

Exemples : * $\mathbb{R}$, muni de l'ordre habituel, est un espace de Riesz simple.

* Dans $\mathcal{B}([01])$ ou $\mathcal{E}([01])$ les idéaux maximaux sont les sous-vectoriels formés des fonctions s'annulant en un point fixé de $[01]$. Il y a bijection entre $[01]$ et l'ensemble $\mathcal{M}$ des idéaux maximaux.

* Dans ℓ^{∞} l'idéal des suites qui convergent vers zéro n'est pas maximal : il est strictement contenu, par exemple, dans l'idéal des suites dont la restriction aux indices pairs converge vers zéro.

1- 1- 7 - Quotient d'un espace de Riesz par un idéal [67]

Si J est un idéal d'un espace de Riesz Y quelconque, la relation $x - y \in J$ est alors une relation d'équivalence. L'espace quotient Y/J est un nouvel espace de Riesz, unitaire si Y l'est : l'unité $\overline{e}$ de Y/J est la classe de l'unité e de Y.

Y/J est simple si J est maximal.

Exemple : le quotient de $\mathcal{B}([01])$ par l'idéal des fonctions s'annulant en n points fixés de [01] est identifiable à $\mathbb{R}^n$.

1- 1- 8 - Bande dans un espace de Riesz complètement réticulé [8]

C'est tout idéal J ayant la propriété suivante :

pour toute partie A de J, majorée dans Y, sup (A) (qui existe dans Y complètement réticulé) appartient à J. Autrement dit : idéal contenant les bornes supérieures de ses parties majorées; il contient alors les bornes inférieures de ses parties minorées.

Remarque : Toute bande est un idéal; le contre exemple suivant met en évidence, dans un espace de Riesz complètement réticulé, un idéal qui n'est pas une bande:

ℓ^∞ est, pour l'ordination naturelle, un espace de Riesz complètement réticulé, archimédien et unitaire. Le sous-espace J des suites de ℓ^∞ qui convergent vers zéro est un idéal. Dans J, la partie A des suites nulles à partir d'un certain rang, et valant 1 avant, est majorée par la suite (111.....) qui est d'ailleurs la borne supérieure de A et qui n'appartient pas à J : J n'est pas une bande.

Par contre, dans $\mathcal{B}$ ([01]) tout idéal maximal est une bande. On caractérisera ultérieurement les idéaux maximaux qui sont des bandes.

Et l'on appelle [8] <u>bande engendrée par une partie A</u> d'un espace de Riesz complètement réticulé, la plus petite bande contenant A, que l'on note B(A).

1- 1- 9 - *Remarque :* Si J est un idéal maximal d'un espace de Riesz Y complètement réticulé, et si J n'est pas une bande, la bande B(J) engendrée par J est nécessairement l'espace Y tout entier. En effet si l'on avait :

$$J \subset B(J) \subset Y \quad \text{(inclusions strictes)}$$

J ne serait pas maximal puisque strictement contenu dans B(J).

1- 2 - RAPPEL des THEOREMES UTILISES

1- 2- 1 - Théorème de F. Riesz [8]

Soit A une partie d'un espace de Riesz Y complètement réticulé. L'ensemble $A^\perp$ des éléments de Y étrangers à tous les éléments de A est une bande. La bande $(A^\perp)^\perp$ coïncide avec la bande engendrée par A et Y est somme directe ordonnée de $A^\perp$ et $(A^\perp)^\perp$

Conséquence : Pour qu'un idéal maximal J d'un espace de Riesz Y complètement réticulé soit une bande, il faut et il suffit que : $J^{\perp} \neq \{0\}$

α) Si J est une bande : la bande $(J^{\perp})^{\perp}$ engendrée par J coïncide donc avec J. Y est alors somme directe de J et $J^{\perp}$. J étant maximal, ne coïncide pas avec Y. On conclut que $J^{\perp}$ n'est pas réduit à l'origine.

β) Si $J^{\perp} \neq \{0\}$, montrons que J est une bande. Sinon, puisque J est maximal, la bande qu'il engendre est tout l'espace (cf 1- 1- 9). D'où Y, somme directe de B(J) et de $J^{\perp}$. Puisque B(J) = Y, ceci entraîne que $J^{\perp}$ est réduit à l'origine, conclusion contraire à l'hypothèse.

1- 2- 2 - Théorème [67]

*Dans un espace de Riesz unitaire, tout idéal est contenu dans un idéal maximal (*utilise le lemme de Zorm*)*

Ce théorème permet d'exhiber un contre exemple d'idéal maximal qui n'est pas une bande :

Dans l'espace ℓ^{∞}, complètement réticulé et unitaire pour la relation d'ordre "naturelle", des suites réelles bornées, l'idéal J des suites convergeant vers zéro, est, d'après ce qui précède contenu dans un idéal maximal J_o. D'après (1- 1- 5) J_o, idéal non trivial, ne peut contenir l'unité de ℓ^{∞}, i.e (par exemple) la suite (1,1.....). Or cette unité est la borne supérieure des suites de J_o nulles à partir d'un certain rang et valant 1 avant; J_o par conséquent n'est pas une bande.

1- 2- 3 - Théorie de Yosida pour les espaces de Riesz unitaires [67] page 372.

Elle montre que tout espace Riesz Y, unitaire et archimédien, est isomorphe * à un sous-espace vectoriel de l'espace $\mathcal{B}(\mathcal{M})$ des fonctions réelles, bornées, définies sur l'ensemble $\mathcal{M}$ des idéaux maximaux de Y. Ce très utile théorème de représentation nécessite la démonstration des deux lemmes suivants:

---

* en tant qu'espace de Riesz

Lemme 1 : Tout espace de Riesz Y, unitaire et simple, est de la forme

$$Y = \{\alpha e \quad -\infty < \alpha < +\infty\}$$

où e est l'unité de Y. (Y est isomorphe à $\mathbb{R}$)

Lemme 2 : Dans un espace de Riesz unitaire, le radical coïncide avec l'intersection de tous les idéaux maximaux.

Soit donc Y un espace de Riesz unitaire de radical $\mathcal{R}$ et soit l'ensemble des idéaux maximaux de Y, d'où

$$\mathcal{R} = \bigcap_{J \in \mathcal{M}} J$$

Soit $J \in \mathcal{M}$; Y/J est simple puisque J est maximal, et unitaire puisque Y l'est. Donc (Lemme 1) Y/J est de la forme

$$Y/J = \{\alpha \bar{e} \quad \alpha \in \mathbb{R}\}$$

où $\bar{e}$ désigne la classe de l'unité e de Y.

Désignons par x(J) le nombre réel correspondant à $x \in Y$ par l'homomorphisme canonique

$$Y \longrightarrow Y/J$$

Alors l'application $J \in \mathcal{M} \longrightarrow x(J) \in \mathbb{R}$ est un homomorphisme de $\mathcal{M}$ dans $\mathbb{R}$, pour x fixé dans Y. L'ensemble de ces homomorphismes définit, lorsque x décrit Y, un homomorphisme de Y dans l'espace $\mathcal{B}(\mathcal{M})$ des fonctions réelles bornées définies sur $\mathcal{M}$. Cet homomorphisme sera isomorphisme ssi:

$$[x(J) = y(J) \quad \forall J \in \mathcal{M}] \Rightarrow [x = y]$$

soit $$[[x-y](J) = 0 \quad \forall J \in \mathcal{M}] \Rightarrow [x = y]$$

or $[x-y](J) = 0$ s'écrit aussi : "x-y appartient à J". Pour qu'il y ait isomorphisme il est donc nécessaire et suffisant que

$$\bigcap_{J \in \mathcal{M}} J = \{0\}$$

C'est-à-dire (Lemme 2) que le radical $\mathcal{R}$ soit réduit à zéro, ou encore (1- 1- 4) que Y soit archimédien. D'où le

Théorème de Yosida : *Soit Y un espace de Riesz unitaire et archimédien. La correspondance* $(x \in Y) \longrightarrow (J \in \mathcal{M} \longrightarrow x(J) \in \mathbb{R})$ *définit un isomorphisme* *d'espace de Riesz de Y dans l'espace* $\mathcal{B}(\mathcal{M})$ *des fonctions réelles bornées définies sur* $\mathcal{M}$ *et l'on a :*

a) $x(J) = 0 \iff x \in J$

b) $[x+y](J) = x(J) + y(J)$

c) $[\lambda x](J) = \lambda x(J)$

d) $x \leq y \iff \forall J\ x(J) \leq y(J)$

e) $e(J) = 1 \quad \forall J \in \mathcal{M}$

f) $|x(J)| = [|x|](J) \quad \forall J \in \mathcal{M}$

g) $x = 0 \iff \forall J \in \mathcal{M}\ x(J) = 0$

h) $[\mathrm{Sup}(x,y)](J) = \mathrm{Max}(x(J),y(J))$

k) $[\mathrm{Inf}(x,y)](J) = \mathrm{Min}\,(x(J), y(J))$

l) si $J_1 \neq J_2$ il existe $x \in Y : x(J_1) \neq x(J_2)$

Remarque : Si Y, espace de Riesz unitaire, n'est pas archimédien, $Y/\mathcal{R}$ quotient de Y par son radical est, lui, unitaire et de radical réduit à zéro, i.e. archimédien. On peut alors évidemment appliquer à Y/R la représentation de Yosida.

Exemples : * $\mathbb{R}^k$, muni de la relation d'ordre "composante à composante" :

$x = (x_1, \ldots\ x_k);\quad y = (y_1, \ldots y_k)$

$x \leq y \underset{df}{\iff} x_i \leq y_i (i = 1,2\ldots k)$

Tout idéal maximal est l'un des k sous-espaces défini par la nullité de l'une des k composantes. La représentation de Yosida est triviale :

$x \in \mathbb{R}^k \longrightarrow x(J_i) = x_i \qquad (i = 1,2\ldots k)$

* pour $\mathcal{B}([01])$ ou $\mathcal{C}([01])$ la représentation de Yosida est également triviale. Tout idéal J_τ maximal est formé des fonctions de $\mathcal{B}([01])$ (resp. $\mathcal{C}([01])$ s'annulant au point $\tau \in [01]$. La représentation de Yosida est alors la suivante :

$f(J_\tau) =$ "$f(\tau)$, valeur de f en τ".

1 - 3 - ESPACES de YOSIDA REGULIERS

1- 3- 1 - *Définition :* Nous appellerons <u>espace de Yosida</u> tout espace de

Riesz unitaire et archimédien (c'est-à-dire vérifiant les hypothèses du théorème de représentation de Yosida)

1- 3- 2 - Topologies sur un espace de Yosida

Proposition : Tout espace de Yosida Y peut être muni d'une topologie τ d'espace vectoriel localement convexe, par la famille $\{\tau_J\}_{J \in \mathcal{M}}$ des semi-normes ainsi définies:

$$x \in Y,\ J \in \mathcal{M} \longrightarrow \tau_J(x) = |x(J)| = [|x|](J)$$

Il est facile de vérifier que les τ_J sont des semi-normes; la topologie τ est séparée (point g) du théorème de Yosida) et compatible avec l'ordre, puisque :

$$|x| \leq |y| \Longrightarrow \forall\ J \in \mathcal{M} \quad \tau_J(x) \leq \tau_J(y)$$

On rappelle qu'une base de voisinages de l'origine, pour la topologie τ, est définie par l'ensemble des voisinages de la forme :

$$V = \begin{Bmatrix} J_1 \ldots\ldots J_r \\ \\ \varepsilon_1 \ldots\ldots\ \ r \end{Bmatrix} \underset{df}{=} \left\{ x \in Y \quad \tau_{J_i}(x) \leq \varepsilon_i \qquad \begin{matrix} \varepsilon_i > 0 \\ i = 1,2\ldots r \end{matrix} \right\}$$

Remarque : Du fait que Y est unitaire résulte que :

$\forall\ x \in Y \quad \exists\ k \in R_+ \quad \forall\ J \in \mathcal{M} \quad \tau_J(x) \leq k$. D'où la

Proposition : Tout espace de Yosida Y peut être muni de la norme ν ainsi définie :

$$x \in Y \longrightarrow \nu(x) = \sup_{J \in \mathcal{M}} \{\tau_J(x)\}$$

dont la topologie (de convergence uniforme) est plus fine que la topologie (de convergence ponctuelle) τ. ν respecte l'ordre puisque :

$$|x| \leq |y| \Rightarrow \nu(x) \leq \nu(y)$$

1- 3- 3 - *Proposition :* Pour tout idéal maximal J d'un espace de Yosida, on a :

$$J^{\perp} = \bigcap_{\substack{K \in \mathcal{M} \\ K \neq J}} K$$

α) $J^{\perp} \subset \bigcap_{\substack{K \in \mathcal{M} \\ K \neq J}} K$ soit $x \in J^{\perp}$ et $K \in \mathcal{M} - \{J\}$. Il existe y dans J tel que $y(K) \neq 0$ car sinon on aurait $J \subset K$, choix impossible puisque J et K sont maximaux et différents. x et y sont étrangers, soit

$\mathrm{Inf}(|x|, |y|) = 0$ d'où

$\mathrm{Min}(|x(K)|, |y(K)|) = 0$ ce qui entraîne, puisque

y(K) n'est pas nul, que x(K) = 0 : x appartient à K, d'où le résultat ;

β) $J^{\perp} \supset \bigcap\limits_{\substack{K \in \mathcal{M} \\ K \neq J}} K$ soit $x \in \bigcap\limits_{\substack{K \in \mathcal{M} \\ K \neq J}} K$. On a donc x(K) = 0

pour tout $K \in \mathcal{M} - \{J\}$. Soit $y \in J$. Il vient :

$\mathrm{Min}(|y(K)|, |x(K)|) = \mathrm{Min}(|y(K)|, 0) = 0$ pour tout $K \in \mathcal{M} - \{J\}$

$\mathrm{Min}(|y(J)|, |x(J)|) = \mathrm{Min}(0, |x(J)|) = 0$. x est étranger à tous les éléments de J d'où le résultat : $x \in J^{\perp}$

Remarque : Dans un espace de Yosida Y, tout idéal maximal J est fermé pour la topologie τ, et $J^{\perp}$ de même

En effet, soit $x \in \overline{J}^{\tau}$. Dans le voisinage de x :

$$V_x = \{y \in Y \quad \tau_J(x - y) \leq \varepsilon\}$$

nous savons exister y appartenant à J, vérifiant par conséquent

$$\tau_J(x-y) \leq \varepsilon \text{ soit}$$
$$|x(J) - y(J)| \leq \varepsilon$$

Or, par hypothèse, y(J) = 0 ($y \in J$). On a montré que, pour tout $\varepsilon > 0$, $|x(J)| \leq \varepsilon$. D'où x(J) = 0 ce qui montre que x appartient à J : J est fermé et $J^{\perp}$, intersection de fermés (lemme précédent) est également fermé.

1- 3- 4 - *Définition :* Nous appellerons espace de Yosida <u>régulier</u> tout espace de Yosida complètement réticulé, pour lequel tout idéal maximal est une bande.

D'après (1- 2- 1) les espaces de Yosida réguliers sont caractérisés, parmi les espaces de Yosida complètement réticulés, par la condition

$$\forall J \in \mathcal{M} \qquad J^{\perp} \neq \{0\}$$

Exemples : $\mathcal{B}([01])$ est un espace de Yosida régulier

$\mathcal{E}([01])$ est un espace de Yosida qui n'est pas régulier, puisque non complètement réticulé.

Enfin ℓ^∞, espace de Yosida complètement réticulé, n'est pas régulier : en effet, on a vu (1- 2- 2) que l'idéal maximal J_o contenant l'idéal J (formé des suites qui convergent vers zéro) n'est pas une bande.

1- 3- 5 - Proposition : *Dans un espace de Yosida régulier Y, $J^\perp$ est un sous-espace de dimension 1 pour tout $J \in \mathcal{M}$ et τ_J est une norme sur $J^\perp$*

On sait que, pour tout $J \in \mathcal{M}$, $J^\perp \neq \{0\}$. Soit donc x non nul appartenant à $J^\perp$; on a :

$x(K) = 0 \qquad \forall K \in \mathcal{M} - \{J\} \qquad (1- 3- 3)$

$x(J) \neq 0 \qquad$ puisque $x \neq 0$

Posons $\qquad e_J = \dfrac{x}{x(J)} \in J^\perp$

D'où $\qquad \begin{cases} e_J(K) = 0 & \forall K \in \mathcal{M} - \{J\} \\ e_J(J) = 1 \end{cases}$

Il est clair que tout $x \in J^\perp$ est proportionnel à e_J. De plus :

$$\{x \in J^\perp,\ \tau_J(x) = 0\} = \{0\}; \quad \tau_J \text{ est une norme sur } J^\perp$$

1 - 3- 6 - Corollaire : *Soit J un idéal maximal d'un espace de Yosida régulier Y. Alors tout $x \in Y$ s'écrit*

$$x = x_1 + \lambda e_J \qquad \text{où } x_1 \in J,\ \lambda \in \mathbb{R}$$

Cette décomposition est unique et si $y = y_1 + \mu e_J$, l'inégalité $x \leq y$ équivaut à

$$x_1 \leq y_1 \qquad \text{(dans J)}$$
$$\lambda \leq \mu \qquad (\text{dans } \mathbb{R})$$

Ce corollaire résulte du théorème de Riesz (1- 2- 1) et de la proposition précédente.

1 - 3- 7 - Proposition : *Soit Y un espace de Yosida régulier. La famille $\mathcal{B} = \{e_J\}_{J \in \mathcal{M}}$ a les propriétés suivantes :*

α) Toute partie finie de $\mathcal{B}$ est libre.

β) L'ensemble des combinaisons linéaires finies d'éléments de $\mathcal{B}$ est dense dans Y pour la topologie τ. ($\mathcal{B}$ est totale)

α) Soit A une partie finie de $\mathcal{B}$: Il existe une partie finie S de $\mathcal{M}$, telle que

$$A = \{e_J\}_{J \in S}$$

Soient $\alpha_J (J \in S)$ des réels tels que

$$\sum_{J \in S} \alpha_J \, e_J = 0$$

D'où pour tout $K \in S$

$$\sum_{J \in S} \alpha_J \, e_J(K) = 0$$

Le premier membre se réduit à $\alpha_K \, e_K(K) = \alpha_K$ d'où $\alpha_K = 0$ pour tout K dans S.

β) Soit $x \in Y$ et V_x un voisinage de x; on peut prendre V_x dans la base usuelle des voisinages de x : il existe une partie finie S de $\mathcal{M}$ et des réels $\varepsilon_J > 0$ $(J \in S)$ tels que

$$V_x = \{y \in Y \quad \tau_J(x-y) \leqslant \varepsilon_J\}_{J \in S}$$

soit
$$x^* = \sum_{K \in S} x(K) \, e_K$$

Alors $\tau_J(x - x^*) = |[x - x^*](J)| = |x(J) - x^*(J)|$, quantité nulle pour tous les $J \in S$, puisqu'alors $x(J) = x^*(J)$; par conséquent $x^* \in V_x$

Le même calcul montre que pour toute partie finie T de $\mathcal{M}$ contenant S, l'élément

$$x^{**} = \sum_{K \in T} x(K) \, e_K$$

appartient encore à V_x.

D'où la

1- 3- 8 - Proposition : *Dans un espace de Yosida régulier Y, la famille $\{x(J) \, e_J\}_{J \in \mathcal{M}}$, définie pour tout $x \in Y$, est sommable (pour la topologie τ) et de somme égale* *à x. On notera :*

$$x = \sum_{J \in \mathcal{H}} x(J)\, e_J$$

1 - 3- 9 - Notations

Soit Y un espace de Yosida régulier

Soit S une partie finie de $\mathcal{H}$, ayant k éléments.

On notera T(S) le sous-espace de dimension k engendré par les e_J $(J \in S)$:

$$T(S) = \underset{J \in S}{+} J^{\perp}$$

et

$$T = \bigcup_{S \in \mathcal{PF}(\mathcal{H})} T(S)$$

T est donc l'ensemble des combinaisons linéaires finies d'éléments de $\mathcal{B}$. D'après ce qui précède, T est dense dans Y pour la topologie τ.

1- 3- 10 - Proposition : *Soit Y un espace de Yosida complètement réticulé. Pour que Y soit régulier, il faut et il suffit que l'isomorphisme de Yosida (de Y dans $\mathcal{B}(\mathcal{H})$) soit surjectif.*

α) La condition est suffisante.

Soit $J \in \mathcal{H}$. Par hypothèse, il existe $x \in Y$, tel que

$$x(J) = 1$$
$$x(K) = 0 \quad \forall\, K \in \mathcal{H} - \{J\}$$

Ce qui montre que x $(\neq 0)$ appartient à $J^{\perp}$ (1- 3- 3). Donc $\forall\, J \in \mathcal{H}$, $J \neq \{0\}$. D'après (1- 3- 4) il en résulte que Y est régulier.

β) La condition est nécessaire : Supposons donc Y régulier, et soit $x^* \in \mathcal{B}(\mathcal{H})$; c'est l'application :

$J \in \mathcal{H} \longrightarrow x^*(J) \in \mathbb{R}$, et il existe k, réel positif, tel que :

$$\forall\, J \in \mathcal{H}, \ |x^*(J)| \leq k$$

Considérons, dans Y, l'ensemble A des sommes finies de la forme :

$$\sum_{J \in S} x^*(J)\, e_J \qquad S \in \mathcal{PF}(\mathcal{H})$$

Il vient

$$\forall S \in \mathcal{PF}(\mathcal{H}) \quad \sum_{J \in S} x^*(J)\, e_J \leq k e$$

A est donc une partie majorée de Y, complètement réticulé : A admet donc une borne supérieure x telle que :

$$\forall J \in \mathcal{H} \qquad x^*(J) \leq x(J)$$

Montrons que l'on a, en fait

$$\forall J \in \mathcal{H} \quad x^*(J) = x(J) \qquad\qquad (1)$$

En effet, s'il existait $J \in \mathcal{H}$ tel que

$$x^*(J) < x(J)$$

On construirait

$$y = x - [x(J) - x^*(J)]e_J \leq x$$
$$y \neq x$$

Alors

$$y(J) = x^*(J) < x(J)$$
$$y(K) = x(K) \qquad \forall K \in \mathcal{H} - \{J\}$$

y serait donc un majorant de A, inférieur à sa borne supérieure, chose absurde. On a donc prouvé que (1) est vraie. L'isomorphisme de Yosida est surjectif.

Exemples α) La représentation de Yosida, pour l'espace de Yosida régulier $\mathcal{B}([01])$ se réduit à l'application identique, évidemment surjective, de $\mathcal{B}([01])$ sur lui-même

β) Pour $\mathcal{C}([01])$, espace de Yosida non régulier, la représentation de Yosida, qui est encore l'application identique, plonge $\mathcal{C}([01])$

dans $\mathcal{B}([01])$: elle n'est évidemment pas surjective.

1- 4 - HOMOMORPHISMES CONTINUS d'un ESPACE de YOSIDA REGULIER dans $\mathbb{R}$.

1- 4- 1 - Définition : On appelle homomorphisme continu d'un espace de Yosida régulier Y dans $\mathbb{R}$, toute fonctionnelle linéaire ℓ définie sur Y, positive :

$$x \leqslant y \implies \ell(x) \leqslant \ell(y)$$

ce qui s'écrit aussi (ℓ, linéaire) :

$$z \geqslant 0 \implies \ell(z) \geqslant 0$$

et continue pour la topologie τ sur Y

$\mathcal{H}$ désignera l'ensemble de tels homomorphismes.

1- 4- 2 - Proposition : *Soit $\ell \in \mathcal{H}$. Il existe une partie finie S_ℓ de $\mathcal{M}$ telle que :*

$$\ell(e_J) = 0 \qquad si\ J \notin S_e$$

$$\ell(e_J) = \gamma_J > 0 \qquad si\ J \in S_e$$

et l'on a, pour tout x dans Y :

$$\ell(x) = \sum_{J \in S_\ell} \ell(e_J)x(J)$$

Soit $\varepsilon > 0$ dans $\mathbb{R}$. ℓ étant continu, il correspond à ε une partie finie S_ε de $\mathcal{M}$ et un voisinage de zéro V_ε :

$$V_\varepsilon = \{z \in Y : \tau_J(z) \leqslant \varepsilon_J \quad \forall J \in S_\varepsilon\}$$

tels que

$$z \in V_\varepsilon \implies |\ell(z)| \leqslant \varepsilon$$

Soit $x \in \bigcap_{J \in S_\varepsilon} J$ et $\mu > 0$ dans $\mathbb{R}$, d'où

$$y = \frac{\varepsilon x}{\mu}$$

y appartient aussi à $\bigcap_{J \in S_\varepsilon} J$ et par conséquent à V_ε. D'où $|\ell(y)| \leqslant \varepsilon$ soit

$$|\ell(x)| \leqslant \mu$$

pour tout $\mu > 0$; donc $\ell(x) = 0$. On a donc prouvé que le noyau $\ell^{-1}(0)$ de ℓ contient $\bigcap_{J \in S_\varepsilon} J$. Alors, pour tout J n'appartenant pas à S_ε, on a :

$$J^\perp = \bigcap_{\substack{K \in \mathcal{M} \\ K \neq J}} K \subset \bigcap_{K \in S_\varepsilon} K \subset \ell^{-1}(0) \qquad \text{et} \quad \ell(e_J) = 0$$

D'où S_ℓ, la partie finie minimale de $\mathcal{M}$ telle que :

$$\ell(e_J) = 0 \qquad \forall J \notin S_\ell$$

On pose $\gamma_J = \ell(e_J) > 0 \qquad \forall J \in S_\ell$, et l'on a pour tout $x \in Y$:

$$x = \sum_{J \in \mathcal{M}} x(J) e_J$$

D'où (ℓ continu)

$$\ell(x) = \sum_{J \in \mathcal{M}} x(J)\, \ell(e_J)$$

quantité qui se réduit donc à :

$$\ell(x) = \sum_{J \in S_\ell} x(J)\, \ell(e_J) = \sum_{J \in S_\ell} x(J)\, \gamma_J$$

1- 4- 3 - Définition : Soient $\ell_1, \ell_2 \in \mathcal{H}$. On dira que

$$\ell_1 \leq \ell_2 \qquad (20)$$

si $\forall x \geq 0 \quad \ell_1(x) \leq \ell_2(x)$

C'est une relation d'ordre partiel sur $\mathcal{H}$.

1- 4- 4 - Proposition : *Une condition nécessaire et suffisante pour que* $\ell_1 \leq \ell_2$ *est que :*

$$\forall J \in \mathcal{M} \qquad \ell_1(e_J) \leq \ell_2(e_J) \qquad (1)$$

La condition est évidemment nécessaire puisque $e_J \geq 0$. Elle est suffisante : supposons (1) vérifiée , et soit $x \geq 0$

$$\ell_2(x) = \sum_{J \in S_{\ell_2}} x(J)\, \ell_2(e_J) \qquad \text{avec } \ell_2(e_K) = 0 \text{ pour } K \notin S_{\ell_2}$$

$$\ell_1(x) = \sum_{J \in S_{\ell_1}} x(J)\, \ell_1(e_J) \qquad \text{avec } \ell_1(e_K) = 0 \text{ pour } K \notin S_{\ell_1}$$

Il résulte alors de (1) que $S_{\ell_1} \subset S_{\ell_2}$

$(\ell_2(e_J) = 0 \implies \ell_1(e_J) = 0)$

D'où le résultat

$$\ell_1(x) \leq \ell_2(x)$$

puisque tous les x(J) sont non négatifs.

*

* *

CHAPITRE 2 - NORMES VECTORIELLES et ESPACES VECTORIELLEMENT NORMES

Dans ce chapitre est définie (§2-2) la notion de norme vectorielle comme norme "à valeur dans un espace Yosida régulier". Les espaces vectoriellement normés constituent un type particulier d'espaces localement convexes. On précise (§2-3) la classe des normes vectorielles régulières et l'on peut prouver (§2-4) que si la norme vectorielle est régulière, toute semi-norme continue admet un plus petit majorant.

A titre d'application, on montre au paragraphe (2-5) comment cette propriété permet d'obtenir des formules d'erreur optimales en un certain sens.

2 - 1 - Introduction

$\mathbb{K}$ désigne $\mathbb{R}$ ou $\mathbb{C}$. Soit X un espace vectoriel sur $\mathbb{K}$ .Munissons le d'une topologie d'espace vectoriel localement convexe : on sait que cela revient à se donner sur X une famille $(\phi_i)_{i\in I}$ de semi-normes .

La relation "pour tout i, $\alpha_i \leq \beta_i$" (dans $\mathbb{R}$) est une relation d'ordre partiel entre les éléments $\alpha = (\alpha_i)_{i\in I}$ et $\beta = (\beta_i)_{i\in I}$ de $\mathbb{R}^I$. On notera $\alpha \leq \beta$

Muni de cet ordre, $\mathbb{R}^I$ est un espace de Riesz, complètement réticulé, archimédien (mais non unitaire). Pour $\alpha = (\alpha_i)_{i\in I}$, on notera $|\alpha| = (|\alpha_i|)_{i\in I}$

Soit alors p l'application de X dans $\mathbb{R}^I$

$$x \in X \longrightarrow p(x) = (\phi_i(x))_{i\in I}$$

Cette application vérifie les axiomes suivants:

$$(A) \begin{cases} \text{i) } p(\lambda x) = |\lambda| p(x) & \forall \lambda \in \mathbb{K}, \forall x \in X \\ \text{ii) } p(x+y) \leq p(x) + p(y) & \forall x, y \in X. \end{cases}$$

et, si la topologie Φ est séparée,

iii) $p(x) = 0 \Rightarrow x = 0$

Réciproquement, soit X un espace vectoriel sur $\mathbb{K}$ et I un ensemble quelconque. Soit p une application de X dans $\mathbb{R}^I$ satisfaisant aux axiomes (A)

D'où, pour tout $x \in X$ et tout $i \in I$, le réel positif ou nul (*)

$$\phi_i(x) = \big(p(x)\big)_i$$

Il en résulte que l'on peut munir X d'une topologie Φ d'espace localement convexe par la famille $(\phi_i)_{i \in I}$ de semi-normes sur X. De plus, si p satisfait à iii), Φ est séparée.

Ainsi il est équivalent de se donner, sur un espace vectoriel X, une topologie d'e.v.l.c. ou de se donner une application de X dans un $\mathbb{R}^I$, satisfaisant aux axiomes (A), la topologie étant séparée ssi l'axiome iii) est vérifié.

Nous allons, dans ce qui suit, particulariser beaucoup l'espace d'arrivée de l'application p.

2 - 2 - NORMES VECTORIELLES.

Avec les notations du Chapitre 1, soit Y un espace de Yosida régulier, $\mathcal{M}$ l'ensemble de ses idéaux maximaux. τ est, sur Y, la topologie de convergence ponctuelle définie par la famille $(\tau_J)_{J \in \mathcal{M}}$ des semi-normes :

$$u \in Y \longrightarrow \tau_J(u) = |u(J)|$$

et l'on sait que l'on peut identifier Y avec $\mathcal{B}(\mathcal{M})$, espace des fonctions numériques bornées définies sur $\mathcal{M}$, muni de la topologie de convergence ponctuelle.

(*) On sait que i) et ii) impliquent que

$$|p(x) - p(y)| \leq p(x-y) \qquad \forall x, y \in X$$

d'où, en particulier : $0 \leq p(x) \qquad \forall x \in X$

2 -2- 1 - Définition. Soit X un espace vectoriel sur $\mathbb{K}$. ($\mathbb{R}$ ou $\mathbb{C}$). On appellera <u>semi-norme vectorielle</u> toute application p, satisfaisant aux axiomes A, de X dans un espace de Yosida régulier Y.

On dira que p est une <u>norme vectorielle</u> si, de plus, l'axiome iii) est vérifié.

On particularise donc beaucoup par rapport au paragraphe précédent, puisque

$$\mathrm{Im}(p) \subset Y = \mathcal{B}(\mathcal{M}) \subset \mathbb{R}^{\mathcal{M}}$$

2 - 2- 2 - Topologies. D'après ce qui précède, X se trouve muni d'une topologie Φ localement convexe par la famille $(\phi_J)_{J\in\mathcal{M}}$ des semi-normes ainsi définies :

$$x \in X \longrightarrow \phi_J(x) = [p(x)]\ (J) = \tau_J(p(x))$$

Nous noterons $V_J = \{x \in X : \phi_J(x) = 0\}$

le noyau de ϕ_J. X a alors la propriété suivante :

$$\forall x \in X,\ \exists\ \nu\in R_+\ \ \forall\ J\in\mathcal{M}\qquad \phi_J(x) \leq \nu$$

puisque $Y = \mathcal{B}(\mathcal{M})$. D'où sur X, la semi-norme ϕ

$$x \in X \longrightarrow \phi(x) = \sup_{J\in\mathcal{M}} \{\phi_J(x)\} = \tau(p(x))$$

dont la topologie (de convergence uniforme) est plus fine que la topologie de convergence ponctuelle Φ.

2- 2- 3 - Taille finie. Lorsque $Y = \mathbb{R}^k$ (ordonné composante à composante), on dira que la semi-norme vectorielle est de taille k.

Il n'y a alors, pour définir Φ, qu'un nombre fini de semi-normes(k), Φ est équivalente à la topologie de la semi-norme ϕ.

Nous noterons $p_1 \dots\ p_k$ les dites semi-normes. $p_i(x)$ est la $i^{\text{ème}}$ composante de p(x) dans $\mathbb{R}^k$ (i = 1,2 ....k)

Par des raisonnements élémentaires on prouve la

2- 2- 4 - Proposition : *Pour qu'une semi-norme vectorielle soit une norme vectorielle, il faut et il suffit que l'une des propositions suivantes (équivalentes à l'axiome iii) soit vérifiée:*

a) $\bigcap_{J\in\mathcal{M}} V_J = \{0\}$ (V_J, noyau de ϕ_J)

b) Φ est séparée

c) la semi-norme ϕ est une norme.

d) Pour toute norme ψ monotone sur Y

l'application $h : x \in X \longrightarrow h(x) = \psi(p(x))$ est une norme sur X.

Rappelons qu'une norme ϕ sur un espace de Riesz est dite absolue [5] si elle respecte l'ordre :

$$|u| \leqslant |v| \implies \phi(u) \leqslant \phi(v)$$

Elle est dite monotone [5] (notion plus faible) si

$$0 \leqslant u \leqslant v \implies \phi(u) \leqslant \phi(v)$$

Nous emploierons les mêmes définitions pour les semi-normes vectorielles définies sur des espaces de Riesz., et nous n'utiliserons plus désormais que des normes vectorielles.

2 - 2- 5 - Définition. Un espace vectoriel X muni d'une norme vectorielle p sera dit vectoriellement normé.

Il est clair, d'après ce qui précède, que les espaces vectoriellement normés sont caractérisés, parmi les espaces vectoriels topologiques localement convexes, par la propriété suivante :

la famille $(\phi_J)_{J \in \mathcal{H}}$ des semi-normes définissant la topologie est telle que :

$$\forall x \in X \qquad \exists \nu \in \mathbb{R}_+ \qquad \forall J \in \mathcal{H} \qquad \phi_J(x) \leqslant \nu$$

2- 2- 6 - Composition - *Soit p uen norme vectorielle de X dans Y, et q une norme vectorielle monotone de Y dans Z. Alors l'application*

$$x \in X \longrightarrow r(x) = q(p(x)) \in Z$$

est une norme vectorielle de X dans Z (démonstration immédiate).

Remarque : *Si* $Y = \mathbb{R}^k$ *et* $Z = \mathbb{R}^\ell$ p *et* r *définissent deux topologies équivalentes sur X.*

- Exemples

1) Evidemment, les normes (vectorielles de taille 1)

2) Sur $\mathbb{R}^n$, l'exemple-type de norme vectorielle (de taille n) est le

suivant :

$$x = (\S_1 \ldots \S_n) \longrightarrow p(x) = |x| = (|\S_1| \ldots |\S_n|)$$

De même, sur tout espace de Yosida régulier Y la norme vectorielle (absolue)

$$u \in Y \longrightarrow p(u) = |u| = \mathrm{Sup}(u^+, u^-)$$

3) Sur $\mathbb{R}^4$, l'application

$$x = (\S_1 \ldots \S_4) \longrightarrow p(x) = (|\S_1| + |\S_2|, \ |\S_3| + |\S_4|)$$

est une norme vectorielle de taille 2. Par contre, l'application

$$x = (\S_1 \ldots \S_4) \longrightarrow q(x) = (|\S_1 + \S_2|, |\S_3 + \S_4|)$$

n'est qu'une semi-norme vectorielle.

4) Sur l'espace $\mathcal{M}_{mn}$ des matrices réelles de type (m,n) l'application

$$A = (a_{ij})_{\substack{1 \leq i \leq m \\ 1 \leq j \leq n}} \longrightarrow p(A) \in \mathbb{R}^m \text{ définie par}$$

$$p_i(A) = \underset{1 \leq j \leq n}{\mathrm{Max}} \{|a_{ij}|\} \qquad (i = 1,2 \ldots m)$$

est une norme vectorielle absolue de taille m

5) sur $C^2([01])$, espace des fonctions numériques ayant une dérivée seconde continue sur [01], la norme vectorielle de taille 3

$$u \in C^2([01]) \longrightarrow p(u) = \left(\underset{t \in [01]}{\mathrm{Sup}} |u''(t) + u(t)|, |u(o)|, |u'(o)| \right)$$

s'introduit parallèlement à l'équation différentielle

$$\begin{cases} u'' + u = f & [u \in C^2([01]) \qquad f \in C^\circ([01])] \\ u(o) = a, \quad u'(o) = b & [a, b \in \mathbb{R}] \end{cases}$$

6) Soit $\mathcal{B}(\Omega)$ l'espace des fonctions numériques bornées définies sur le carré [01] x [01]

soit $f \in \mathcal{B}(\Omega)$ d'où $g_f \in \mathcal{B}([01])$ définie par

$$y \in [01] \longrightarrow g_f(y) = \underset{x \in [01]}{\mathrm{Sup}} \{|f(x,y)|\}$$

L'application

$$f \in \mathcal{B}(\Omega) \longrightarrow p(f) = g_f \in \mathcal{B}([01])$$

est évidemment une norme vectorielle (absolue)

7) Sur ℓ^2, espace des suites de carré sommable, soit p l'application:

$$x = (\S_1 \ldots \S_k, \S_{k+1} \ldots) \longrightarrow p(x) = \left(|\S_1|, \ldots |\S_k|, \left(\sum_{k+1}^{\infty} |\S_j|^2\right)^{1/2} \right)$$

c'est une norme vectorielle de taille k+1.

De même sur ℓ^∞, espace des suites bornées :

$$x = (\S_1 \ldots \S_k \S_{k+1} \ldots) \longrightarrow p(x) = \left(|\S_1|, \ldots, |\S_k|, \sup_{k+1}^{\infty} |\S_j| \right)$$

ces deux derniers exemples sont empruntés à [36]

2- 2- 7 - Sous-espaces V_J, W_J, $W(S)$, W, Z_J

Soit p une norme vectorielle sur X. Nous allons définir un certain nombre de sous-espaces. On posera, pour tout $J \in \mathcal{M}$:

$V_J = p^{-1}(J)$ (noyau de la semi-norme ϕ_J)

$$W_J = p^{-1}(J^\perp) = p^{-1}\Big(\bigcap_{\substack{K \in \mathcal{M} \\ K \neq J}} K\Big) = \bigcap_{\substack{K \in \mathcal{M} \\ K \neq J}} p^{-1}(K) = \bigcap_{\substack{K \in \mathcal{M} \\ K \neq J}} V_K$$

Si p est de taille finie k, on notera ces sous-espaces V_i et W_i ($i = 1, 2 \ldots k$)

D'où les propriétés élémentaires :

a) V_J et W_J, *images réciproques d'idéaux par une norme vectorielle sont des sous-espaces vectoriels de X.*

b) V_J et W_J, *images réciproques de fermés de $Y(\tau)$ par p, application continue, sont des fermés de X (Φ)*

c) $V_J \cap W_J = \bigcap_{J \in \mathcal{M}} V_J = \{0\}$

Il en résulte que, sur W_j, ϕ_J est une norme. D'après la construction de W_J, il est clair que la topologie définie sur W_J par ϕ_J est la topologie induite par Φ

d) $W_J \cap W_K = \{0\}$ si $J \neq K$

e) la relation

(1) $\quad 0 = x_1 + \ldots + x_r \quad$ où $\quad x_i \in W_{J_i} \quad (i=1,2 \ldots r)$

entraîne $\quad 0 = x_1 = x_2 = \ldots = x_r$

En effet (1) s'écrit :

$$x_i = -\sum_{j \neq i} x_j$$

Or tous les x_j $(j\neq i)$ appartiennent à V_{J_i} d'où :

$$x_i \in W_{J_i} \cap V_{J_i} = \{0\}$$

f) D'où pour toute partie finie S de $\mathcal{M}$, la somme directe algébrique :

$$W(S) = \dotplus_{J\in S} W_J$$

et
$$W = \bigcup_{S\in \mathcal{F}(\mathcal{M})} W(S)$$

Ainsi W est l'ensemble des sommes finies d'éléments pris dans les W_J $(J\in\mathcal{M})$

Enfin on posera, pour tout $J\in\mathcal{M}$:

$$Z_J = W_J + V_J \qquad \text{(somme directe algébrique)}$$

g) On a, pour tout J dans $\mathcal{M}$:

$$Z_J \supset W$$

En effet soit $x\in W$. Il existe une partie finie S de $\mathcal{M}$ telle que x s'écrit, de façon unique

$$x = \sum_{K\in S} x_K \qquad \text{où } x_K \in W_K$$

α) si $J\notin S$ $\quad x \in V_J \subset Z_J$

β) si $J\in S$ $\quad x = x_J + \sum_{\substack{K\in S\\K\neq J}} x_K = x_J + y_J$

$x_J \in W_J$; $\quad y_J \in V_J$. Donc $x\in Z_J$.

h) Soit $x\in W$:

$$x = \sum_{K\in S} x_K \qquad (S\in\mathcal{F}(\mathcal{M});\ x_K\in W_K)$$

Alors $\quad$ α) si $J\notin S$ $\quad \phi_J(x) = 0$

β) si $J\in S$ $\quad \phi_J(x) = \phi_J(x_J)$

α) est évident. Quant à β) :

$$\phi_J(x_J) - \sum_{\substack{K\in S\\K\neq J}} \phi_J(x_K) \leq \phi_J\Big(\sum_{K\in S} x_K\Big) \leq \phi_J(x_J) + \sum_{\substack{K\in S\\K\neq J}} \phi_J(x_K)$$

or $\quad \phi_J(x_K) = 0$ si $J \neq K$, puisque $x_J \in W_J \subset V_K$

On rappelle qu'au chapitre 1, on a défini (1- 3- 9) de façon analogue à f) les sous-espaces suivants de l'espace de Yosida régulier Y :

$$T(S) = \underset{J \in S}{+} J^{\perp} \qquad \forall\, S \in \mathcal{PF}(\mathcal{M})$$

$$T = \bigcup_{S \in \mathcal{PF}(\mathcal{M})} T(S) \qquad \text{On a alors :}$$

k) $p(W) \subset T_+$

En effet soit $x = \sum_{K \in S} x_K \in W$ d'où

$[p(x)](J) = \phi_J(x) = 0$ si $J \notin S$, ce qui prouve que

$$p(x) \in \big(T(S)\big)_+ \subset T_+$$

2 -2- 8 - Proposition : *Pour que*

$$p(W) = T_+$$

il faut et il suffit que

$$\forall\, J \in \mathcal{M}\,, \qquad W_J \neq \{0\} \qquad (2)$$

c'est-à-dire que, pour tout J *dans* $\mathcal{M}$, *l'image réciproque par* p *de* $e_J \in J^{\perp}$ *soit non vide.*

Montrons que cette condition, évidemment nécessaire, est suffisante.

Soit donc, pour tout J dans $\mathcal{M}$, x_J tel que

$$p(x_J) = e_J \qquad \text{d'où} \quad x_J \in W_J$$

Soit $u \in T_+$. Il existe S, partie finie de $\mathcal{M}$ telle que

$$u = \sum_{J \in S} u(J) e_J \qquad (0 \leq u(J) \in \mathbb{R})$$

Montrons que l'élément

$$x = \sum_{J \in S} u(J) x_J,$$

qui appartient à W(S), donc à W, vérifie

$$p(x) = u \qquad (3)$$

En effet : $\phi_J(x) = u(J)\phi_J(x_J) = u(J)$ si $J \in S$

$$\phi_K(x) = 0 \quad \text{si } K \notin S \quad \text{d'où} \qquad (3)$$

Nous dirons, par abus de langage, que p est surjective lorsque (2) est vérifié. Ceci parce que, lorsque p est de taille finie k, tout $\mathbb{R}^k{}_+$ est atteint par p(x), $x \in W$.

2 - 3 - NORMES VECTORIELLES REGULIERES

Les normes vectorielles régulières sont en fait les normes vectorielles utilisables, à cause d'une propriété, étudiée en 2-4, exploitée constamment par la suite.

2-3- 1 - Définition - La norme vectorielle p est dite régulière si

$$\overline{W} = X$$

ou $\overline{W}$ désigne la fermeture de W pour la topologie Φ.

Exemple - La norme vectorielle de taille 3 sur $C^2([0,\pi])$

$$u \in C^2([0,\pi]) \to p(u) = \left(\sup_{0 \leq t \leq \pi} |u''(t) + u(t)|, |u(0)|, |u'(0)| \right)$$

W_1 est le sous-espace défini par $u(0) = u'(0) = 0$

W_2 est défini par $u'' + u = 0$, $u'(0) = 0$; W_2 est donc l'ensemble des fonctions de la forme $u(t) = A\cos t$ $(A \in \mathbb{R})$

W_3 défini par $u'' + u = 0$, $u(0) = 0$, est l'ensemble des fonctions de la forme $u(t) = B \sin t$. $(B \in \mathbb{R})$

Or pour tout $u \in C^2([0,\pi])$ on peut écrire :

$$u(t) = \underbrace{[u(t) - u(0)\cos t - u'(0)\sin t]}_{= u_1(t)} + \underbrace{[u(0)\cos t]}_{= u_2(t)} + \underbrace{[u'(0)\sin t]}_{= u_3(t)}$$

où $u_1 \in W_1$; $u_2 \in W_2$; $u_3 \in W_3$, ce qui montre que $W = \overline{W} = C^2([0,\pi])$; p est régulière.

2- 3- 2 - Propriété caractéristique : *Les propositions suivantes sont équivalentes :*

a) $\overline{W} = X$ (p est régulière)

b) $\forall J \in \mathcal{H}$ $\overline{Z}_J = X$

a) => b) On a vu que, dans tous les cas :

$$Z_J \supset W \qquad \text{d'où}$$

$$\overline{Z}_J \supset \overline{W} \qquad \text{d'où le résultat}$$

b) => a) Soit $x \in X$ et V_x un voisinage de x pour Φ

Montrons que, moyennant b), $W \cap V_x \neq \emptyset$

Se donner V_x revient à se donner une partie finie S de $\mathcal{M}$ et des réels positifs ε_J $(J \in S)$ avec

$$V_x = \{y \in X \qquad \forall J \in S \qquad \phi_J(x-y) \leq \varepsilon_J\}$$

D'après b) il existe, pour tout $J \in \mathcal{M}$, $y_J \in Z_J$ tel que

$$\phi_J(x-y_J) \leq \varepsilon_J \qquad (1)$$

avec $\qquad y_J = x_J + z_J \qquad$ ou $x_J \in W_J$, $z_J \in V_J$

(1) s'écrit

$$\phi_J(x-x_J) \leq \varepsilon_J$$

Posons $\qquad x^* = \sum\limits_{J \in S} x_J \in W(S) \subset W$

Alors $\qquad \phi_J(x^* - x) = \phi_J(x - x_J) \leq \varepsilon_J$ pour tout $J \in S$

par conséquent $\qquad x^* \in V_x \cap W$. Ainsi $V_x \cap W \neq \emptyset$. cqfd.

2- 3- 3 - Corollaire. *Toute norme vectorielle régulière est surjective.*

Soit p régulière. Supposons exister $J \in \mathcal{M}$ tel que

$$W_J = \{0\} \qquad \text{(p non surjective)}$$

On aurait donc

$$Z_J = W_J + V_J = V_J \qquad \text{or } V_J \text{ est fermé}$$

d'où

$$\overline{Z}_J = V_J \text{ et, p étant régulière}$$

$$\overline{Z}_J = X$$

La semi-norme ϕ_J serait donc partout nulle, cas trivial que nous éliminons.

2 - 4 - MAJORATION OPTIMALE d'une SEMI-NORME

Nous donnons, dans ce paragraphe, le résultat qui est à la base de toutes les applications que nous développerons. Commençons par la

2- 4- 1 - Définition

Soit ψ une semi-norme définie sur un espace vectoriel X muni d'une norme vectorielle p, à valeurs dans un espace de Yosida régulier Y. $\mathcal{H}$ désigne l'ensemble des homomorphismes continus (pour la topologie τ) de Y dans $\mathbb{R}$

Nous appellerons <u>majorant</u> de la semi-norme ψ, tout $\ell \in \mathcal{H}$ tel que

$$\forall x \in X, \quad \psi(x) \leq \ell(p(x)) \qquad (1)$$

Et nous nous proposons l'étude de tels majorants.

2- 4- 2 - Proposition : *La semi-norme ψ admet un majorant si et seulement si elle est continue pour la topologie Φ*

1 - Soit ψ admettant un majorant ℓ. D'après le chapitre 1, on sait qu'il existe une partie finie S de $\mathcal{M}$ et des réels positifs $\gamma_J (J \in S)$ tels que

$$\forall u \in Y \qquad \ell(u) = \sum_{J \in S} u(J)\gamma_J$$

On a alors, d'après (1)

$$\forall x \in X, \quad \psi(x) \leq \sum_{J \in S} [p(x)]\ (J)\ \gamma_J = \sum_{J \in S} \phi_J(x)\gamma_J \qquad (2)$$

Montrons que ψ est continue. A $\varepsilon > 0$ on associe dans X le voisinage de zéro suivant :

$$V_\varepsilon = \{x \in X \quad \forall J \in S \quad \phi_J(x) \leq \varepsilon/_{r\gamma_J}\}$$

avec r = card(S)

si $x - y \in V_\varepsilon$ on a, d'après (2) :

$$|\psi(x) - \psi(y)| \leq \psi(x-y) \leq \sum_{J \in S} \phi_J(x-y)\gamma_J \leq r\varepsilon / r = \varepsilon$$

2 - Soit ψ une semi-norme sur X, continue pour la topologie Φ. Montrons qu'elle admet un majorant.

$\varepsilon > 0$ étant donné dans , on sait exister S, partie finie de et des constantes positives $\varepsilon_J (J \quad S)$ telles que

$\varepsilon > 0$ étant donné dans $\mathbb{R}$, on sait exister S, partie finie de $\mathcal{H}$ et des constantes positives $\varepsilon_J (J \in S)$ telles que

$$[\forall J \in S \quad \phi_J(x) \leqslant \varepsilon_J] \Rightarrow \psi(x) \leqslant \varepsilon \tag{3}$$

Soit $z \in X$. Posons, si $\underset{J \in S}{\mathrm{Max}} (\phi_J(z)) \neq 0$:

$$x = \frac{\underset{J \in S}{\mathrm{Min}} (\varepsilon_J)}{\underset{J \in S}{\mathrm{Max}} (\phi_J(z))} z.$$

Alors $\phi_J(x) \leqslant \varepsilon_J$ pour tout $J \in S$ et par conséquent

$$\psi(x) \leqslant \varepsilon \quad \text{soit}$$

$$\psi(z) \leqslant \frac{\varepsilon}{\underset{J \in S}{\mathrm{Min}}(\varepsilon_J)} \underset{J \in S}{\mathrm{Max}} (\phi_J(z)) \leqslant \frac{\varepsilon}{\underset{J \in S}{\mathrm{Min}}(\varepsilon_J)} \sum_{J \in S} \phi_J(z)$$

Si $\underset{J \in S}{\mathrm{Max}} (\phi_J(z)) = 0$, (3) montre que $\psi(z) = 0$. On a donc, pour tout z dans X :

$$\psi(z) \leqslant \ell(p(z))$$

ψ est donc majoré par $\ell \in \mathcal{H}$ ainsi défini :

$$\begin{cases} \ell(e_K) = 0 & \text{si } K \notin S \\ \ell(e_K) = \dfrac{\varepsilon}{\underset{J \in S}{\mathrm{Min}}(\varepsilon_J)} & \text{si } K \in S \end{cases}$$

2 - 4 - 3 - Proposition : *Si ψ est une semi-norme continue elle s'annule sur tous les W_J sauf un nombre fini d'entre eux.*

Sur chacun de ces derniers, ψ est continue pour la norme ϕ_J.

Avec les notations introduites, on vérifie facilement, grâce à (3), que le noyau V_ψ de ψ contient le sous-espace

$$\bigcap_{J \in S} V_J$$

Or ce sous-espace contient lui-même tous les W_K pour $K \notin S$. Et puisque la topologie induite par Φ sur le fermé W_J est celle de la norme ϕ_J, on conclut que, pour tout $J \in S$:

$$0 \leqslant \underset{\substack{x \in W_J \\ x \neq 0}}{\mathrm{Sup}} \left\{ \frac{\psi(x)}{\phi_J(x)} \right\} < \infty$$

2- 4- 4 - Définition : Soit ψ une semi-norme continue sur X. Nous lui associons la partie (finie) S_ψ de $\mathcal{M}$ formée des seuls idéaux J telle que ψ n'est pas identiquement nulle sur W_J. Nous définissons alors ℓ_ψ, élément de $\mathcal{H}$ par

$$\begin{cases} \ell_\psi(e_K) = 0 & \text{si } K \notin S_\psi \\ \ell_\psi(e_J) = \sup\limits_{x \in W_J} \left\{ \dfrac{\psi(x)}{\phi_J(x)} \right\} \underset{df}{=} \gamma_J > 0 & \text{si } J \in S_\psi \end{cases}$$

2- 4- 5 - Position du problème

Nous nous posons la question de savoir à quelle condition sur p toute semi-norme continue admet un plus petit majorant (pour la relation d'ordre sur $\mathcal{H}$ définie en 1- 4- 3)

Les résultats suivants préparent une réponse :

2- 4- 6 : Lemme. *Soit ψ une semi-norme continue sur X. Tout majorant ℓ de ψ est alors tel que :*

$$\ell_\psi \leqslant \ell \qquad (4)$$

En effet, soit $J \in S_\psi$ et x non nul dans W_J, d'où

$p(x) = \phi_J(x)\, e_J$ Alors

$$\psi(x) \leqslant \ell(p(x)) = \phi_J(x)\, \ell(e_J)$$

On a donc :

$$\begin{cases} \forall J \in S_\psi & \ell(e_J) \geqslant \sup\limits_{\substack{x \in W_J \\ x \neq 0}} \left\{ \dfrac{\psi(x)}{\phi_J(x)} \right\} = \gamma_J = \ell_\psi(e_J) \\ \forall K \notin S_\psi & \ell(e_K) \geqslant 0 = \ell_\psi(e_K) \end{cases}$$

Ce qui, d'après (1- 4- 4), est équivalent à (4).

2- 4- 7 - Lemme. *Soit ψ une semi-norme continue sur X. On a :*

$$\forall x \in W \qquad \psi(x) \leqslant \ell_\psi(p(x)) \qquad (5)$$

En effet, tout $x \in W$ s'écrit

$$x = \sum_{J \in S} x_J \qquad \left(S \in \mathcal{P}\mathcal{F}(\mathcal{M}),\quad x_J \in W_J\right)$$

d'où

$$\psi(x) = \psi(\sum_{J\in S} x_J) \leq \sum_{J\in S} \psi(x_J) = \sum_{J\in S \cap S_\psi} \psi(x_J) \underset{df}{=} \alpha$$

puisque $\psi(x_J) = 0$ si $J \notin S_\psi$

or $\alpha \leq \sum_{J\in S \cap S\psi} \gamma_J \phi_J(x_J) = \sum_{J\in S \cap S\psi} \gamma_J \phi_J(x) \underset{df}{=} \beta$; or $\phi_J(x) = 0$ si $J \notin S$, d'où

$$\beta = \sum_{J\in S\psi} \gamma_J \phi_J(x) = \ell_\psi(p(x)) \qquad \text{d'où} \qquad (5)$$

Nous pouvons alors répondre au problème posé par le

2- 4- 8 - Théorème 1

Soit X un espace vectoriel sur $\mathbb{R}$ *ou* $\mathbb{C}$, *muni de la norme vectorielle p à valeurs dans un espace de Yosida régulier Y. X et Y sont respectivement munis des topologies* Φ *et* τ *de convergence ponctuelle.*

La régularité de p est alors condition suffisante pour que toute semi-norme ψ, *continue sur X, admette un plus petit majorant* (*) ; *ce plus petit majorant est* ℓ_ψ, *et l'ensemble des majorants de* ψ *est l'ensemble des* ℓ *tels que :*

$$\ell_\psi \leq \ell$$

En effet, les applications ψ et ℓ_ψ et p sont continues. Il résulte donc de (2- 4- 7) que l'on a

$$\forall x \in \overline{W} \qquad \psi(x) \leq \ell_\psi(p(x))$$

ce qui montre que si p est régulière ($\overline{W} = X$) ℓ_ψ est un majorant de ψ

Alors, il résulte du Lemme (2- 4- 6) que ℓ_ψ est le plus petit majorant de ψ puisque, d'après ce lemme, tout majorant ℓ de ψ vérifie

$$\ell_\psi \leq \ell$$

Et réciproquement il est trivial de voir alors que tout ℓ vérifiant l'inégalité ci-dessus est aussi majorant de ψ (puisque ℓ_ψ en est un)

(*) On verra ultérieurement (chapitre 3) que si X est de dimension finie la régularité de p est également condition nécessaire d'existence, pour tout ψ, d'un plus petit majorant.

2- 5 - APPLICATION : FORMULE d'ERREUR POUR UNE EQUATION DIFFERENTIELLE (simple exemple).

Soit l'équation différentielle suivante, à résoudre dans $\mathcal{E}^2([0,\pi])$

$$\begin{cases} u'' + u = f \\ u(0) = a; \quad u'(0) = b \end{cases}$$

où f est une fonction donnée, continue sur $[0,\pi]$.

On considère la norme vectorielle

$$v \in \mathcal{E}^2([0,\pi]) \longrightarrow p(v) = \left| \begin{array}{l} \sup\limits_{t \in [0,\pi]} \{ |v''(t) + v(t)| \} \underset{df}{=} \phi_\infty(v'' + v) \\ |v(0)| \\ |v'(0)| \end{array} \right.$$

On a vu (2- 3- 1) qu'elle est régulière et que :

$$W_1 = \{ v \ / \ v(0) = v'(0) = 0 \}$$

$$W_2 = \{ v \ / \ v(t) = A \cos t \}$$

$$W_3 = \{ v \ / \ v(t) = B \sin t \}$$

Soit, sur $\mathcal{E}^2([0,\pi])$, la semi-norme ψ:

$$v \in \mathcal{E}^2([0,\pi]) \longrightarrow \psi(v) = |v(\pi)|$$

on a

$$v(\pi) = -v(0) + \int_0^\pi [v''(t) + v(t)] \sin t \, dt$$

d'où

$$\forall \ v \in \mathcal{E}^2([0,\pi]) \qquad \psi(v) = |v(\pi)| \leqslant |v(0)| + \phi_\infty(v''+v) \int_0^\pi |\sin t| dt$$

ce qui montre que ψ est continue pour la topologie d'espace vectoriellement normé définie par p sur $\mathcal{E}^2([0 \ ])$

Le théorème 1 nous prouve alors l'existence de constantes minimales $\alpha \ \beta \ \gamma$ telles que

$$\forall \ v \in \mathcal{E}^2([0,\pi]) \qquad \psi(v) \leqslant [\alpha \ \beta \ \gamma] \ p(v)$$

et l'on sait que l'on a

$$\alpha = \sup_{v\in W_1}\left\{\frac{\psi(v)}{\phi_\infty(v''+v)}\right\} = \sup_{v\in W_1}\frac{\left|\int_0^\pi [v''(t)+v(t)]\,\mathrm{sint}\,dt\right|}{\sup_{0\leq t\leq\pi}\{|v''(t)+v(t)|\}}$$

D'où

$$\alpha \leq \int_0^\pi |\mathrm{sint}|\,dt = 2$$

Or pour $v_o(t) = 1 - \mathrm{cost}$,

$v_o'' + v_o = 1$

$v_o(0) = v_o'(0) = 0$

v_o appartient à W_1 et l'on a $\dfrac{\psi(v_o)}{\phi_\infty(v''_o + v_o)} = \dfrac{2}{1} = 2$

d'où $\alpha = 2$

$$\beta = \sup_{v\in W_2}\left\{\frac{\psi(v)}{|v(0)|}\right\} = \sup_{v\in W_2}\left\{\frac{|v(0)|}{|v(0)|}\right\} = 1$$

$$\gamma = \sup_{v\in W_3}\left\{\frac{\psi(v)}{|v'(0)|}\right\} = \sup_{v\in W_3}\left\{\frac{0}{|v'(0)|}\right\} = 0$$

d'où la formule optimale

$$\forall v\ \mathcal{C}^2([0\pi]) \qquad |v(\pi)| \leq 2\,\phi_\infty(v''+v) + |v(0)| \quad \text{(A)}$$

Nous allons faire le même calcul pour la semi-norme

$$v\in\mathcal{C}^2([0\pi]) \longrightarrow \phi(v) = |v'(\pi)|$$

On a

$$v'(\pi) = -v'(0) + \int_0^\pi [v''(t)+v(t)]\,\mathrm{cost}\,dt$$

d'où

$$\forall v\in\mathcal{C}^2([0\pi]) \qquad \phi(v) = |v'(\pi)| \leq |v'(0)| + \phi_\infty(v''+v)\int_0^\pi |\mathrm{cost}|\,dt$$

ce qui montre que ϕ est continue elle aussi pour la topologie d'espace vectoriellement normé défini par p sur $\mathcal{C}^2([0\pi])$

D'où (théorème 1) l'existence de constantes minimales ρ, μ, ε telles que :

$$\forall\, v \in \mathscr{C}^2([0\pi]) \qquad \phi(v) \leq [\rho\ \mu\ \varepsilon]\ p(v)$$

et l'on sait que l'on a :

$$\rho = \operatorname*{Sup}_{v\in W_1} \left\{ \frac{\phi(v)}{\phi_\infty(v''+v)} \right\} = \operatorname*{Sup}_{v\in W_1} \left\{ \frac{\left| \int_0^\pi [v''(t)+v(t)\ \cos t\ dt \right|}{\operatorname*{Sup}_{0\leq t\leq\pi} \left\{ |v''(t)+v(t)| \right\}} \right\}$$

d'où

$$\rho \leq \int_0^\pi |\cos t|\ dt = 2$$

Or soit v_1 la fonction suivante

$$v_1(t) = 1 - \cos t \quad \text{entre } 0 \text{ et } \pi/2$$

$$= -\cos t + 2 \sin t - 1 \quad \text{entre } \pi/2 \text{ et } \pi$$

on a

$$v_1''(t) + v_1(t) = 1 \text{ entre } 0 \text{ et } \pi/2$$

$$v_1''(t) + v_1(t) = -1 \quad \text{entre } \pi/2 \text{ et } \pi$$

v_1 est une fonction a dérivée première continue, a dérivée seconde discontinue en $\pi/2$. Or l'on a

$$\frac{\phi(v_1)}{\phi_\infty(v_1''+v_1)} = \frac{\left| \int_0^{\pi/2} \cos t\ dt - \int_{\pi/2}^{\pi} \cos t\ dt \right|}{1} = 2$$

v_1 peut être approchée uniformément par une suite de fonctions de W_1 ce qui prouve que

$$\rho = 2$$

$$\mu = \operatorname*{Sup}_{v\in W_2} \left\{ \frac{\phi(v)}{|v(0)|} \right\} = 0$$

$$\varepsilon = \operatorname*{Sup}_{v\in W_3} \left\{ \frac{\phi(v)}{|v'(0)|} \right\} = \operatorname*{Sup}_{v\ W_3} \left\{ \frac{|v'(0)|}{|v'(0)|} \right\}$$

d'où la formule optimale :

$$\forall\, v \in \mathscr{C}^2([0\pi]) \qquad |v'(\pi)| \leq 2\ \phi_\infty(v''+v) + |v'(0)| \qquad \text{(B)}$$

Soit alors $u_\nu(t)$ une solution numérique approchée de la solution exacte $u(t)$ de l'équation différentielle donnée. Posons $v(t) = u(t) - u_\nu(t)$ (fonction erreur). Les formules optimales (A) et (B) permettent d'écrire

$$|u(\pi) - u_\nu(\pi)| \leq 2\,\phi_\infty(u''+u-u''_\nu-u_\nu)+|u(0)-u_\nu(0)|$$

$$|u'(\pi) - u'_\nu(\pi)| \leq 2\phi_\infty(u''+u-u''_\nu-u_\nu) + |u'(0) - u'_\nu(0)|$$

Or, par définition $u'' + u = f$; $u(0) = a$; $u'(0) = b$; d'où

$$\boxed{|u(\pi) - u_\nu(\pi)| \leq 2\phi_\infty(u''_\nu + u_\nu - f) + |u_\nu(0) - a| \qquad (\mathrm{I})}$$

$$\boxed{|u'(\pi) - u'_\nu(\pi)| \leq 2\phi_\infty(u''_\nu + u_\nu - f) + |u'_\nu(0) - b| \qquad (\mathrm{II})}$$

Ces formules d'erreur sont <u>optimales</u> dans le sens que le jeu des constantes [(2, 1, 0) et (2, 0, 1)] qui figure dans chacune d'elle est le plus petit possible. On remarque que $u''_\nu + u_\nu - f$ est le résidu relatif à la solution approchée u_ν.

Exemple : Supposons que u_ν vérifie : $\begin{cases} u''_\nu + u_\nu = f \\ u_\nu(0) = a \end{cases}$

d'où, en appliquant (I)

$$u(\pi) = u_\nu(\pi)$$

la solution approchée prend, en π, la même valeur que la solution exacte.

Cela peut être vérifié directement, en posant

$$v = u - u_\nu$$

d'où $\begin{cases} v'' + v = 0 \\ v(0) = 0 \end{cases}$

ce qui montre que v, élément de W_3, est de la forme

$$v(t) = A \sin t$$

fonction qui s'annule bien en π.

CHAPITRE 3 : NORMES VECTORIELLES SUR $\mathbb{R}^n$ ou $\mathbb{C}^n$

Nous allons voir (§3- 1) ce que deviennent les résultats du chapitre précédent lorsque X est $\mathbb{R}^n$ ou $\mathbb{C}^n$, et Y $\mathbb{R}^k$ muni de la relation d'ordre "composante à composante".

Le problème principal est alors le suivant : $\mathbb{C}^n$ et $\mathbb{C}^m$ étant respectivement munis des normes vectorielles p et q, peut-on construire, associée à p et q, une norme vectorielle sur les matrices de type (m,n), considérées comme opérateurs linéaires? Nous prouverons (§3-2) qu'on peut le faire <u>si et seulement si</u> p est <u>régulière</u>. De plus, l'existence d'une norme vectorielle duale est caractéristique des normes vectorielles régulières (§3-3).

De fait, la régularité est une propriété essentielle pour une norme vectorielle, à cause des résultats précédents. Ceci justifie le fait que seules les normes vectorielles régulières sont utilisées en analyse numérique, comme nous le verrons dans les chapitres ultérieurs.

3 - 1 - NORMES VECTORIELLES sur $\mathbb{R}^n$ ou $\mathbb{C}^n$

3- 1- 1 - Notations.

$\mathbb{K}$ désigne $\mathbb{R}$ ou $\mathbb{C}$. Soit p une norme vectorielle de taille k sur $\mathbb{K}^n$: c'est une application de $\mathbb{K}^n$ dans $\mathbb{R}^k$ vérifiant les axiomes :

i) $p(\lambda x) = |\lambda|\, p(x) \quad \forall \lambda \in \mathbb{K} \;, \forall x \in \mathbb{K}^n$

ii) $p(x+y) \leq p(x)+p(y) \quad \forall x,y \in \mathbb{K}^n$

iii) $p(x) = 0 \Rightarrow x = 0$

Nous noterons $p_i(x)$ la ième composante de p(x) dans $\mathbb{R}^k$ (i =1,2...k). D'où les sous-espaces

$$V_i = \{x \in \mathbb{K}^n \quad p_i(x) = 0\} \quad \text{(noyau de la semi norme } p_i)$$

$$W_i = \bigcap_{\substack{j=1\\ j\neq i}}^{k} V_j \qquad \text{et la somme directe}$$

$$W = \sum_{i=1}^{k} W_i$$

Il est clair que se donner une norme vectorielle de taille k sur $\mathbb{K}^n$*, c'est se donner k semi-normes dont les noyaux* V_i *sont d'intersection réduite à zéro :* (séparation, i.e. axiome iii) :

$$\bigcap_{i=1}^{k} V_i = \{0\}$$

condition d'ailleurs équivalente à la suivante : *les boules unités* B_i *des semi-normes* p_i *sont d'intersection*

$$\bigcap_{i=1}^{k} B_i$$

bornée dans $\mathbb{K}^n$.

3- 1- 2 - Normes vectorielles surjectives.

Pour que $p(\mathbb{K}^n) = \mathbb{R}_+^k$ il faut et il suffit que (cf 2-2-8)

$$W_i \neq \{0\} \quad (i = 1, 2 \,....\, k)$$

Corollaire : *Une norme vectorielle de taille supérieure à la dimension de l'espace sur lequel elle est définie ne peut être surjective.*

En effet

$$\dim(W) = \sum_{i=1}^{k} \dim (W_i) \leq n$$

si $k > n$ cela entraîne l'existence d'un indice i entre 1 et k tel que

$$\dim (W_i) = 0 \qquad \text{soit } W_i = \{0\}$$

et d'après ce qui précède p n'est pas surjective.

Remarque : Si en supprimant une ou plusieurs semi-normes à la norme vectorielle p, on obtient encore une norme vectorielle (axiome iii) de séparation encore vérifié), p n'est pas surjective. En effet, pour toute semi-norme p_i supprimée, on a, relativement à p : $W_i = \{0\}$. Cas particulier de ceci : l'une des semi-normes est une norme sur $\mathbb{K}^n$. En supprimant toutes les autres, on obtient donc encore une norme (vectorielle de taille 1)

Exemple : Sur $\mathbb{R}^4$ la norme vectorielle

$$x = (\xi_1 \ldots \xi_4) \longrightarrow p(x) = (|\xi_1|, |\xi_1| + |\xi_2|, |\xi_3| + |\xi_4|)$$

n'est pas surjective ($W_1 = \{0\}$)

Mais

$$x = (\xi_1 \ldots \xi_4) \longrightarrow q(x) = (|\xi_1|, |\xi_2| + |\xi_3| , |\xi_3| + |\xi_4|)$$

est surjective.

3- 1- 3 - Normes vectorielles régulières (cf 2-3)

p est dite régulière si

$$W = \mathbb{K}^n$$

(dans un espace de dimension finie tout sous-espace est fermé). p peut être ainsi construite :

Décomposons $\mathbb{K}^n$ en la somme directe de k sous-espaces $E_1 \ldots E_k$

dont aucun n'est réduit à zéro. Nous noterons x_i la projection de $x \in \mathbb{K}^n$ dans E_i. Choisissons, sur chaque E_i, une norme Φ_i. Alors l'application

$$x \in \mathbb{K}^n \longrightarrow p_i(x) \underset{df}{=} \Phi_i(x_i) \tag{3}$$

est une semi-norme sur $\mathbb{K}^n$, et il est simple de vérifier que l'application

$$x \in \mathbb{K}^n \longrightarrow p(x) = \left(p_1(x), \ldots\ldots\ldots p_k(x)\right) \in \mathbb{R}_+^k \tag{4}$$

est une norme vectorielle de taille k sur $\mathbb{K}^n$, telle que

$$W_i = E_i \qquad (i = 1,2 \ldots\ldots k) \tag{5}$$

Nous dirons que p est <u>construite sur les W_i</u>

Elle est donc régulière. Inversement, les résultats du chapitre précédent montrent que toute norme vectorielle régulière sur $\mathbb{K}^n$ est ainsi construite.

Exemples : Sur $\mathbb{R}^4$, la norme vectorielle suivante est régulière

$$x = (\xi_1 \ldots\ldots \xi_4) \longrightarrow p(x) = (\mathrm{Max}\{|\xi_1|, |\xi_2|\}, |\xi_3| + |\xi_4|)$$

$$W_1 = \mathbb{R}_1 + \mathbb{R}_2 \qquad W_2 = \mathbb{R}_3 + \mathbb{R}_4 \quad ; \; W = \mathbb{R}^4$$

Par contre la norme vectorielle suivante est surjective mais non régulière :

$$x = (\xi_1 \ldots\ldots \xi_4) \longrightarrow q(x) = (|\xi_1|, |\xi_2| + |\xi_3|, |\xi_3| + |\xi_4|)$$

$$W_1 = \mathbb{R}_1 \quad ; \quad W_2 = \mathbb{R}_2 \quad ; \quad W_3 = \mathbb{R}_4 \qquad W \neq \mathbb{R}^4$$

La proposition 2- 3- 2 s'écrit alors

3- 1- 4 - Caractérisation

Soit p une norme vectorielle de taille k sur $\mathbb{K}^n$. Les propositions suivantes sont équivalentes :

a) $W = \mathbb{K}^n$ (p est régulière)

b) $V_i + W_i = \mathbb{K}^n$ $(i = 1, 2 \ldots\ldots k)$

Toute norme vectorielle régulière est surjective.

Remarque : Par conséquent, toute norme vectorielle régulière est nécessairement de taille $k \leq n$ (cf 3- 1- 2), dimension de l'espace sur lequel elle est définie.
Soit $k = n$, et supposons p surjective. Alors les W_i, de dimension ≥ 1 (puisque $\neq 0$) sont en nombre n d'où $\dim(W) \geq n$. On conclut, par conséquent que tous les W_i sont de dimension 1 et que $W = \mathbb{K}^n$.
Donc pour une norme vectorielle p de taille n sur $\mathbb{K}^n$ surjective <==> régulière <==> $\dim(W_i) = 1 \quad (i = 1...n)$
Il existe donc une matrice A, de type (n,n), non singulière, telle que

$$\forall x \in \mathbb{K}^n \qquad p(x) = |Ax| \qquad (6)$$

Enfin nous allons montrer que, pour des normes vectorielles sur des espaces de dimension finie, la propriété de majoration optimale des semi-normes est <u>caractéristique</u> des normes vectorielles régulières.

3- 1- 5 - Définition. Soit p une norme vectorielle de taille k sur $\mathbb{K}^n$, et ψ une semi-norme sur $\mathbb{K}^n$. On appellera <u>majorant</u> de ψ (relativement à p) tout $v \in \mathbb{R}^k$ tel que

$$\forall x \in \mathbb{K}^n \qquad \psi(x) \leq v^t p(x)$$

On désignera par Maj (ψ) l'ensemble des majorants de ψ. Maj (ψ) est une partie convexe, non bornée de $\mathbb{R}^k$

Exemple : Soit, sur $\mathbb{R}^4$, la norme vectorielle p

$$x = (\xi_1 \ldots\ldots \xi_4) \longrightarrow p(x) = (|\xi_1| + |\xi_2| + |\xi_3| \ , \ |\xi_3| + |\xi_4|)$$

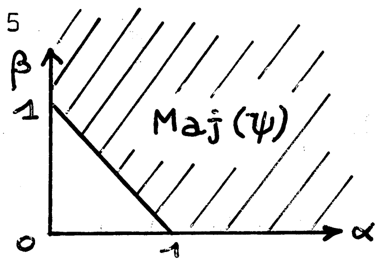
et la semi-norme

$$x = (\xi_1 \ldots \xi_4) \longrightarrow \psi(x) = |\xi_3|$$

Maj (ψ) est l'ensemble des couples (α,β) tels que

$$\forall (\xi_1 \ldots\ldots \xi_4) \qquad |\xi_3| \leq \alpha\,[|\xi_1| + |\xi_2| + |\xi_3|] + \beta\,[|\xi_3| + |\xi_4|]$$

soit $\begin{cases} \alpha + \beta \geqslant 1 \\ \alpha \geqslant 0 \\ \beta \geqslant 0 \end{cases}$



Cette partie de $\mathbb{R}^2{}_+$ <u>n'a pas de plus petit élément</u> pour la relation d'ordre composante à composante.

On a alors le

3- 1- 6 - Théorème 1-bis

Soit p une norme vectorielle de taille k sur $\mathbb{K}^n = \mathbb{R}^n$ *ou* $\mathbb{C}^n$*. Pour que l'ensemble des majorants de* ψ *admette, quelle que soit la semi-norme* ψ *sur* $\mathbb{K}^n$*, un plus petit élément* $\mathbf{v}_\psi$ *(pour la relation d'ordre composante à composante sur* $\mathbb{R}^k$*) il faut et il suffit que p soit régulière. La ième composante de* $\mathbf{v}_\psi$ *est alors donnée par*

$$0 \leqslant \gamma_i = \operatorname*{Sup}_{\substack{x \in W_i \\ x \neq 0}} \left\{ \frac{\psi(x)}{p_i(x)} \right\} \qquad (1)$$

et l'ensemble des majorants de ψ *est constitué des* $\mathbf{v}$ *de* $\mathbb{R}^k$ *tels que*

$$\mathbf{v}_\psi \leq \mathbf{v}$$

La condition est évidemment suffisante, d'après le théorème 1 (2-4-8). De plus, la valeur des composantes des $\mathbf{v}_\psi$ est bien, dans ce cas, donnée par (1)

Montrons que la condition, dans ce contexte d'espace de dimension finie, est, de plus, <u>nécessaire</u>.

Soit ψ une semi-norme sur $\mathbb{K}^n$, de noyau $Z_i = V_i + W_i$, pour i fixé entre 1 et k.

α) $Z_i \supset V_i$ D'où l'existence, <u>puisque $\mathbb{K}^n$ est de dimension finie</u>, de la constante non négative :

$$\lambda = \operatorname*{Sup}_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \left\{ \frac{\psi(x)}{p_i(x)} \right\} < \infty$$

ce qui montre que $\mathbf{v}_1 = (0 \,\ldots\ldots\, \lambda \,..\, 0) = \lambda e_i$ est un majorant de ψ. De même β) $Z_i \supset W_i$ or W_i est le noyau de la semi-norme $\sum_{j \neq i} p_j$.

D'où l'existence de la constante μ non négative :

$$\mu = \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \left\{ \frac{\psi(x)}{\sum_{j \neq i} p_j(x)} \right\}$$

ce qui montre que $v_2 = (\mu, \mu, \ldots, 0, \ldots \mu) = \sum_{j \neq i} \mu e_j$ est aussi un majorant de ψ

Par hypothèse Maj (ψ) admet un plus petit élément. Contenant v_1 et v_2 il contient donc $\mathrm{Inf}(v_1, v_2)$. Or

$$\mathrm{Inf}\ (v_1, v_2) = 0 \qquad \text{d'où}$$

$$\forall\, x \in \mathbb{K}^n \qquad \psi(x) = 0 \qquad \text{d'où} \qquad V_i + W_i = \mathbb{K}^n$$

On a donc prouvé que

$$V_i + W_i = \mathbb{K}^n \qquad (i = 1, 2 \ldots k)$$

c'est-à-dire (3-1-4) que p est <u>régulière</u>

Remarque : Au lieu d'imposer l'existence d'un plus petit majorant pour toute semi-norme, il suffit de l'imposer pour toutes les semi-normes de la forme

$$\psi_y(x) \underset{df}{=} |y^H x| \qquad (y \in \mathbb{K}^n \ ; \ y \neq 0)$$

On a alors le

3- 1- 7 - Corollaire *Soit p une norme vectorielle de taille k sur $\mathbb{K}^n$. Une condition nécessaire et suffisante pour que Maj (ψ_y) admette, pour tout y non nul dans $\mathbb{K}^n$, un plus petit élément, est que p soit régulière.*

D'après ce qui précède, la condition est évidemment suffisante. On démontre sa nécessité de façon très analogue à ce qui a été fait ci-dessus:

Soit p non régulière. Il existe (3-1-4) un indice i entre 1 et k tel que :

$$Z_i = W_i + W_i \subset \mathbb{K}^n \qquad \text{(inclusion stricte)}$$

Donc $\dim(Z_i) \leq n-1$. D'où l'existence, dans $\mathbb{K}^n$, d'un $y \neq 0$ tel que le noyau de ψ_y (de dimension n-1) contienne Z_i

De façon analogue à ce qui a été fait ci-dessus, on prouve que si ψ_y admettait un plus petit majorant, ψ_y serait partout nulle, conclusion absurde.

Ce corollaire permettra de montrer qu'une norme vectorielle sur $\mathbb{K}^n$ admet une duale <u>ssi</u> elle est régulière.

3 - 2 - NORMES VECTORIELLES de MATRICES.

Dans cette section, on construit une norme vectorielle de matrice considérée comme application linéaire d'un $\mathbb{K}^n$ dans un $\mathbb{K}^m$ respectivement munis de normes vectorielles p et q. Enonçons le problème de la façon suivante :

soit p une norme vectorielle de taille k sur $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$) et q une norme vectorielle de taille ℓ sur $\mathbb{K}^m$.

Comment construire, sur l'espace $\mathcal{M}_{mn}$ des matrices (m,n) à éléments dans $\mathbb{K}$, une norme vectorielle d'opérateur, de façon analogue à ce qui s'écrit, lorsque p et q sont des normes ordinaires :

$$A \in \mathcal{M}_{mn} \longrightarrow S_{qp}(A) = \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \left\{ \frac{q(Ax)}{p(x)} \right\}$$

Notations : nous noterons ici $p_j(x)$ $(j=1,\dots,k)$ et $q_i(y)$ $(i=1,\dots,\ell)$ les composantes de p(x) et q(y). Nous noterons

$$V_j^p \quad W_j^p \quad W^p \text{ (resp. } W_i^q \; W_i^q \quad W^q)$$

les sous-espaces habituels relatifs à p (resp. q)

3- 2- 1 - Définition

Nous appellerons <u>majorante</u> de $A \in \mathcal{M}_{mn}$ (relativement à p et q) toute matrice B de type (ℓ,k) telle que :

$$\forall\, x \in \mathbb{K}^n, \quad q(Ax) \leq B\, p(x) \qquad (1)$$

Désignons par Maj (A) l'ensemble des majorantes de A. Soit v_i^t la ième ligne de $B \in$ Maj(A) et ψ_i la semi-norme sur $\mathbb{K}^n$:

$$x \in \mathbb{K}^n \longrightarrow \psi_i(x) = q_i(Ax) \qquad (i = 1,2 \ldots.\ell)$$

(1) s'écrit alors

$$\forall\, x \in \mathbb{K}^n \quad \psi_i(x) \leq v_i^t\, p(x) \qquad (i = 1, 2 \ldots\ell)$$

Ainsi, dire que $B \in$ Maj(A), c'est dire que chacune de ses lignes, transposée, appartient à Maj $(\psi_i)(i=1,2,\ldots,\ell)$. Donc Maj (A) n'est pas vide. C'est une partie convexe, non bornée, de l'ensemble des matrices de type (ℓ,k). Maj(A) se trouve partiellement ordonné par la relation d'ordre "élément à élément" entre matrices réelles de même type. S'il existe, notons par M(A) le plus petit élément de Maj(A).

La démonstration du résultat suivant est élémentaire.

3- 2- 2 - Proposition : *Si* M(A) *existe pour tout* $A \in \mathcal{M}_{mn}$, *l'application*

$$A \longrightarrow M(A)$$

est une norme vectorielle de taille $\ell.k$ *sur l'espace* $\mathcal{M}_{mn}$ *de dimension* m n.

Dire que M(A) existe, c'est dire que chacune de ses lignes est le plus petit majorant de ψ_i (i = 1,2 .... k)

Et dire que M(A) existe quels que soient m, ℓ, q et A, c'est dire que p vérifie l'hypothèse du théorème 1-bis (3-1-6) c'est-à-dire, d'après ce théorème, que p est régulière.

D'où la :

3- 2- 3 - Proposition :

Soit p *une norme vectorielle de taille* k *sur* $\mathbb{K}^n$.

α) *pour que* Maj(A) *admette un plus petit élément* M(A) *quelle que soit la norme vectorielle* q, *de taille* ℓ *(*ℓ *quelconque) définie sur* $\mathbb{K}^m$

(m quelconque) et quelle que soit $A \in \mathcal{M}_{mn}$, *il faut et il suffit que* p *soit régulière*

β) *si* p *est régulière,* $\mathcal{M}_{aj}(A)$ *est l'ensemble des matrices* B *de type* (ℓ, k) *telles que*

$$M(A) \leqslant B$$

où $M(A) = [\mathfrak{m}_{ij}(A)]_{\substack{1 \leq i \leq \ell \\ 1 \leq j \leq k}}$ *est définie par :*

$$\mathfrak{m}_{ij}(A) = \sup_{\substack{x \in W_j \\ x \neq 0}} \left\{ \frac{q_i(Ax)}{p_j(x)} \right\}$$

γ) *l'application* $A \longrightarrow M(A)$ *est alors une norme vectorielle de taille* $\ell.k$ *sur* $\mathcal{M}_{mn}$

Voyons maintenant le cas, important pour les applications où p et q sont régulières.

3- 2- 4 - Corollaire : *Si* p *et* q *sont régulières,* M *est une norme vectorielle elle-même régulière*

En effet $\mathbb{K}^n = \overset{k}{\underset{j=1}{+}} W_j^p$; $\mathbb{K}^m = \overset{\ell}{\underset{i=1}{+}} W_i^q$

Soit $x \in \mathbb{K}^n$, $A \in \mathcal{M}_{mn}$ et $y = A\,x$

Désignons par P_j l'opérateur de projection de $\mathbb{K}^n$ dans W_j^p

Q_i " " " " $\mathbb{K}^m$ " W_i^q

$\sum_{j=1}^{k} P_j$ est la transformation identique dans $\mathbb{K}^n$

$\sum_{i=1}^{\ell} Q_i$ est la transformation identique dans $\mathbb{K}^m$

Posons $A_{ij} = Q_i\, A\, P_j$ de sorte que

$$A = \sum_{\substack{1 \leq j \leq k \\ 1 \leq i \leq \ell}} A_{ij}$$

Soit G_{ij} le sous-espace de $\mathcal{M}_{mn}$ décrit par A_{ij} lorsque A parcourt

$\mathcal{M}_{mn}$. On a alors

$$\mathcal{M}_{mn} = \underset{\substack{1\leq i\leq \ell \\ 1\leq j\leq k}}{+} G_{ij} \qquad (2)$$

et la relation

$$y = Ax \qquad \text{s'écrit}$$

$$Y_i = \sum_{j=1}^{k} A_{ij}\, x_j \qquad i = 1,\ldots\ldots \ell$$

$$x_j = P_j\, x \in W_j^p$$

$$Y_i = Q_i\, y \in W_i^q$$

Alors, pour $x = x_j \in W_j^p$, $Y_i = A_{ij} x_j$ et $q_i(Ax) = q_i(A_{ij}x_j)$ d'où

$$m_{ij}(A) = \sup_{\substack{x_j \in W_j^p \\ x_j \neq 0}} \left\{ \frac{q_i(A_{ij}x_j)}{p_j(x_j)} \right\} \overset{\text{df}}{=} S_{q_i p_j}(A_{ij})$$

Par conséquent, si l'on pose, relativement à la norme vectorielle M :

$$V_{ij}^M = \{A \in \mathcal{M}_{mn} \quad m_{ij}(A) = 0\} \qquad (1\leq i,j\leq k)$$

et

$$W_{ij}^M = \bigcap_{\substack{r\neq i \\ s\neq j}} V_{rs}^M \qquad (1\leq i,j\leq k)$$

On a

$$W_{ij}^M = G_{ij} \qquad (3)$$

$$1 \leq i\leq \ell$$
$$1 \leq j\leq k$$

D'où le résultat : M est régulière puisque, d'après (2) et (3) $\mathcal{M}_{mn}$ est la somme directe des W_{ij}^M

3 - 3 - APPLICATION : NORME VECTORIELLE DUALE D'UNE NORME VECTORIELLE REGULIERE

Soit p une norme vectorielle de taille k sur $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$). D'après le corollaire 3-1-7 la régularité de p est une condition équivalente à la suivante :

Pour tout $y \in \mathbb{K}^n - \{0\}$, il existe un plus petit élément $v_y \in \mathbb{R}^k$, dans l'ensemble des v tels que

$$\forall\, x \in \mathbb{K}^n \qquad |y^H x| \leq v^t p(x)$$

Supposons donc p régulière, d'où l'application p^* :

$$y \in \mathbb{K}^n - \{0\} \longrightarrow p^*(y) = v_y \in \mathbb{R}^k_+$$

Et l'on posera : $\quad p^*(0) = 0$

Sur $\mathbb{R}$, le module est une norme, évidemment régulière. Il résulte donc du corollaire précédent (3-2-4) que p^* est une norme vectorielle (de taille k) régulière sur $\mathbb{K}^n$. p^* sera appelé la duale de p. Désignons par p^*_i la ième semi-norme de p^*. On a :

$$y \in \mathbb{K}^n \longrightarrow p^*_i(y) = \sup_{\substack{x \in W_i^p \\ x \neq 0}} \left\{ \frac{|y^H x|}{p_i(x)} \right\}$$

Remarque : La restriction de p^*_i à W_i^p est la norme duale sur W_i^p, de la norme p_i.

3- 3- 1 - Proposition: $W_i^{p^*}$ *est le complément orthogonal de* V_i^p

$$(i = 1 \;\; 2 \ldots . k)$$

α) Soit $y \in W_i^{p^*}$; alors pour tout $x \in \mathbb{K}^n$

$$|y^H x| \leq [p^*(y)]^t p(x) = \qquad p_i^*(y)\; p_i(x) \qquad \text{et}$$

pour $x \in V_i^p \qquad p_i(x) = 0 \qquad$ d'où

$$y^H x = 0$$

Ainsi

$$W_i^{p^*} \subset (V_i^p)^\perp \qquad \textbf{(1)}$$

β) Réciproquement soit $y \in (V_i^p)^\perp$. Alors, pour tout $x \in V_i^p$,

$$y^H x = 0 \qquad \text{D'où } (j \neq i)$$

$$p_j^*(y) = \sup_{\substack{x \in W_j^p \\ x \neq 0}} \left\{ \frac{|y^H x|}{p_j(x)} \right\}$$

Or $W_j^p \subset V_i^p$ pour tous les $j \neq i$. D'où

$$P_j^*(y) = 0 \text{ pour tous les } j \neq i \text{, soit :}$$

$y \in W_i^{p^*}$ d'où l'inclusion inverse de (1) :

$$W_i^{p^*} \supset (V_i^p)^\perp$$

Remarque : a) $W_i^{p^*}$ et W_i^p, tous deux complémentaires de V_i^p, ont donc même dimension

b) D'où dim (W^{p^*}) = dim (W^p) = n : on retrouve que p^* est régulière.

3- 3- 2 - Proposition : W_i^p *est le complémentaire orthogonal de* $V_i^{p^*}$ (i = 1,2 .... k)

α) soit $x \in V_i^{p^*}$ Alors

$$p_i^*(x) = \sup_{\substack{y \in W_i^p \\ y \neq 0}} \left\{ \frac{|x^H y|}{p_i(y)} \right\} = 0$$

ce qui montre que :

$$W_i^p \subset \left(V_i^{p^*}\right)^\perp$$

β) or W_i^p a même dimension que $W_i^{p^*}$, complémentaire de $V_i^{p^*}$.

Donc W_i^p <u>est</u> le complémentaire orthogonal de $V_i^{p^*}$

Puisque p^* est régulière, elle admet elle même une duale, notée p^{**}. Il résulte de (3-3-1) et (3-3-2) que l'on a :

$$W_i^{p^{**}} = W_i^p \; ; \quad V_i^{p^{**}} = V_i^p \qquad (i = 1,2\ldots\ldots k)$$

or $x \in \mathbb{K}^n$ s'écrit, de façon unique

$$x = \mu + \nu \quad \text{où} \quad \mu \in W_i^{p^*} \quad \text{et} \quad \nu \in V_i^p$$

puisque $W_i^{p^*}$ et V_i^p sont complémentaires orthogonaux. D'où

$$p_i(x) = p_i(\mu)$$

$$p_i^{**}(x) = p_i^{**}(\mu)$$

Or, sur $W_i^{p^*}$, p_i^{**} coïncide avec p_i, duale de la norme p_i^*. On a donc prouvé la

3- 3- 3 - Proposition : *Une norme vectorielle régulière de taille* k *sur* $\mathbb{K}^n$ *coïncide avec sa biduale*

Exemple : Sur $\mathbb{R}^6$ soit p la norme vectorielle régulière de taille 3

$$x = (\xi_1 \ldots \xi_6) \longrightarrow p(x) = \left(|\xi_1| + |\xi_2|, (|\xi_3|^\alpha + |\xi_4|^\alpha)^{1/\alpha}, \text{Max}\{|\xi_5|, |\xi_6|\}\right)$$

avec $\alpha > 1$

Notant $y = (\eta_1 \ldots \eta_6)$ on a

$$p_1^*(y) = \sup_{\xi_1 \xi_2} \left\{ \frac{|\xi_1\eta_1 + \xi_2\eta_2|}{|\xi_1| + |\xi_2|} \right\} = \text{Max}\{|\eta_1|, |\eta_2|\}$$

$$p_2^*(y) = \sup_{\xi_3 \xi_4} \left\{ \frac{|\xi_3\eta_3 + \xi_4\eta_4|}{(|\xi_3|^\alpha + |\xi_4|^\alpha)^{1/\alpha}} \right\} = (|\eta_3|^\beta + |\eta_4|^\beta)^{1/\beta}$$

avec $1/_\alpha + 1/_\beta = 1$

$$p_3^*(y) = \sup_{\xi_5 \xi_6} \left\{ \frac{|\xi_5\eta_5 + \xi_6\eta_6|}{\text{Max}(|\xi_5|, |\xi_6|)} \right\} = |\eta_5| + |\eta_6|$$

D'où $$p^*(y) = \left(\text{Max}\{|\eta_1|, |\eta_2|\}, (|\eta_3|^\beta + |\eta_4|^\beta)^{1/\beta}, |\eta_5| + |\eta_6|\right)$$

Et il est clair que $p^{**} = p$.

CHAPITRE 4 : NORMES VECTORIELLES de MATRICES CARREES

Nous commençons l'étude des normes vectorielles sur l'algèbre des matrices carrées par le résultat suivant, établi en 4 - 1 : si p est une norme vectorielle surjective sur $\mathbb{K}^n$ (donc, en particulier, si elle est régulière) les majorantes de toute matrice A donnée sont non-négatives et de rayon spectral supérieur ou égal à celui de A. En (4 - 2) on étend aux normes vectorielles un résultat de N.Gastinel montrant que les normes de type $S_{\phi\phi}$ sont minimales parmi les normes sous-multiplicatives de matrices. L'extension se fait sans difficulté moyennant la proposition préliminaire (4- 2- 2)

En (4 - 3) on démontre une extension d'un théorème de Householder-Ostrowski concernant l'approximation d'un rayon spectral par une norme. La démonstration est un exemple de manipulation effective de normes vectorielles.

En (4 - 4) sont données diverses réponses au problème de l'approximation à ε près du rayon spectral d'une matrice A donnée par celui d'une majorante de A.

Enfin en (4 -5), on étudie en vue des applications aux itérations linéaires le comportement, lorsque m tend vers l'infini, de la norme vectorielle de la puissance $m^{\text{ième}}$ d'une matrice donnée.

4 - 1 - LE CAS "CARRE"

Soit p une norme vectorielle (non nécessairement régulière) de taille k sur $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$). $\mathcal{M}_{nn}$ désigne l'algèbre des matrices carrées de type (n,n) à élément dans $\mathbb{K}$. Soit $A \in \mathcal{M}_{nn}$.

4- 1- 1 - Définition

Nous appellerons <u>majorante de A (relativement à p)</u> toute matrice réelle B de type (kk) telle que

$$\forall\, x \in \mathbb{K}^n \qquad p(Ax) \leqslant B\, p(x) \qquad (1)$$

Avec les notations § (3-2) on a donc $m = n$ et $p = q$

4- 1- 2 - Proposition : *Si p est surjective, toute majorante d'une matrice A donnée* ***est*** *non négative*

En effet, supposons exister une majorante B de A, admettant un élément négatif, noté b_{ij}. On tire de (1), pour $x = x_j \in W_j$

$$p_i(Ax_j) \leqslant b_{ij}\, p_j(x_j)$$

En effet (2- 2- 7) $p_r(x) = p_r(x_r) = \begin{cases} 0 \text{ si } r \neq j \\ p_j(x_j) \text{ pour } r = j \end{cases}$

si x_j est différent de zéro, $p_j(x_j)$ est positif puisque p_j est une norme sur W_j, $b_{ij}\, p_j(x_j)$ est négatif et par conséquent $p_i(Ax_j)$ aussi ce qui est absurde. On conclut que

$$W_j = \{0\}$$

Ce qui montre que p n'est pas surjective (3- 1-)

Contre exemple : Pour la norme vectorielle suivante sur $\mathbb{R}^4$

$$x = \left|\begin{array}{l} \S_1 \\ \cdot \\ \cdot \\ \cdot \\ \S_4 \end{array}\right. \longrightarrow \qquad p(x) = \left|\begin{array}{c} |\S_1| \\ |\S_1| + |\S_2| \\ |\S_3| + |\S_4| \end{array}\right.$$

qui n'est pas surjective ($W_1 = \{0\}$), la matrice unité (6,6) admet une majorante dont certains éléments sont négatifs :

$$p(Ix) = \begin{vmatrix} |\xi_1| \\ |\xi_1| + |\xi_2| \\ |\xi_3| + |\xi_4| \end{vmatrix} \leqslant \begin{vmatrix} -0{,}5 & 2 & 0 \\ -3 & 5 & 1 \\ 0 & 0 & 1 \end{vmatrix} \begin{vmatrix} |\xi_1| \\ |\xi_1| + |\xi_2| \\ |\xi_3| + |\xi_4| \end{vmatrix}$$

4- 1- 3 - Proposition

Soit p *une norme vectorielle surjective de taille* k *sur* $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$). *Alors, pour tout* $A \in \mathcal{M}_{nn}$ *et toute majorante* B *de* A *(relativement à* p*) le rayon spectral de* B *n'est pas inférieur à celui de* A *:*

$$\rho\,(A) \leqslant \rho\,(B)$$

. Démonstration.

D'après la proposition précédente B est non négative. Supposons la de plus irréductible. D'après le théorème de Perron-Frobenius, [25] [64], sur les matrices non négatives, **B a** une valeur propre $\rho > 0$, égale à son rayon spectral $\rho(B)$ et il lui correspond un vecteur propre ayant toutes ses composantes positives :

$$Bu = \rho u \quad ; \quad \rho > 0 \quad ; \quad u = (u_1 \ldots u_k) \quad ; \quad u_i > 0 \quad (i = 1,2 \ldots k) \ ;$$

Soit alors λ une valeur propre de A et x_λ un vecteur propre relatif à λ. x n'est défini qu'à une homothétie près et il est toujours possible de le choisir tel que les relations suivantes soient vérifiées :

$$\begin{cases} p_1\,(x_\lambda) \leqslant u_1 \\ \vdots \\ p_j\,(x_\lambda) = u_j \\ \vdots \\ p_k\,(x_\lambda) \leqslant u_k \end{cases}$$

Ceci, parce que tous les u_i sont positifs : il existe au moins un indice j tel que l'égalité $p_j(x_\lambda) = u_j$ ait lieu (*)

Or B est une majorante de A relativement à p :

$$\forall\, x \in \mathbb{K}^n \qquad p(Ax) \leq B\, p(x)$$

D'où, pour $x = x_\lambda$

$$p(Ax_\lambda) = p(\lambda x_\lambda) = |\lambda| p(x_\lambda) \leq B\, p(x_\lambda)$$

(*) *Remarque :* Si $\mathbb{K} = \mathbb{R}$ p est une norme vectorielle sur $\mathbb{R}^n$ et A une matrice réelle. Mais x_λ peut appartenir à $\mathbb{C}^n$ et $p(x_\lambda)$ n'est pas défini. On conviendra alors de prolonger p à $\mathbb{C}^n$ de la façon suivante :

$$z = x + iy \longrightarrow \tilde{p}(z) \underset{df}{=} p(x) + p(y)$$

$$(z \in \mathbb{C}^n,\ x,\ y \in \mathbb{R}^n)$$

ce qui donne une norme vectorielle sur $\mathbb{C}^n$. D'autre part, si B est une majorante de A relativement à p :

$$\forall\, x \in \mathbb{R}^n \qquad p(Ax) \leq Bp(x)$$

on a, pour z pris dans $\mathbb{C}^n$

$$\tilde{p}(Az) = \tilde{p}(Ax + i\, Ay) \underset{df}{=} p(Ax) + p(Ay) \text{ puisque A est réelle}$$

$$\leq B\, [p(x) + p(y)] = B\, \tilde{p}(z)$$

ce qui montre que B est aussi une majorante de A relativement à $\tilde{p}$. On se trouve donc dans le cas $\mathbb{K} = \mathbb{C}$ et A réelle.

Comme, réciproquement, il est clair que toute majorante de A réelle relativement à $\tilde{p}$ est aussi une majorante de A relativement à p, il en résulte que, pour A réelle, l'ensemble des majorantes de A est le même relativement à p ou $\tilde{p}$.

Or $B \geq 0$ et $p(x_\lambda) \leq u$ d'où

$$|\lambda| p(x_\lambda) \leq B\, u = \rho u \qquad \text{d'où}$$

$$|\lambda| p_j(x_\lambda) \leq \rho u_j \qquad \text{soit}$$

$$|\lambda| \leq \rho \quad \text{puisque } p_j(x_\lambda) = u_j > 0$$

Ceci est vrai pour toute valeur propre λ de A. Le résultat est donc acquis si B est irréductible.

Si B est réductible, soit E la matrice de type (k,k) "pleine de 1". Alors, pour tout ε réel, positif, la matrice

$$B + \varepsilon E$$ (qui est aussi une majorante de A) est irréductible.

D'où, pour tout $\varepsilon > 0$

$$\rho(A) \leqslant \rho\,(B + \varepsilon E)$$

ce qui, par des arguments de continuité, entraîne que

$$\rho(A) \leqslant \rho\,(B)$$

e t le résultat acquis dans le cas général.

Remarques :

1) Si p est régulière, Maj(A) admet un plus petit élément M(A); donc pour toute majorante B de A :

$$0 \leqslant M(A) < B$$

d'où $$\rho(M(A) \leqslant \rho(B)$$

[En effet, sur l'ensemble des matrices non négatives, le rayon spectral est une fonction non décroissante de l'ensemble des éléments [64]]. D'où

$$\rho(A) \leqslant \rho\, M(A) = \min_{B \in \mathrm{Maj}(A)} \{\rho(B)\}$$

2) si p est une norme ordinaire, on sait que la plus petite majorante est le nombre

$$0 \leqslant \mathrm{Spp}(A) \underset{df}{=} \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \left\{ \frac{p(Ax)}{p(x)} \right\}$$

Le résultat précédent apparaît donc comme une extension de l'inégalité bien connue :

$$\forall\ A \qquad \rho(A) \leqslant \mathrm{Spp}(A)$$

Si p est régulière, la norme vectorielle M a une propriété importante : elle est sous-multiplicative. En effet

$$\forall x \in \mathbb{K}^n \qquad p(ABx) \leqslant M(A)p(Bx) \leqslant M(A)M(B)p(x)$$

d'où $$M(AB) \leqslant M(A)M(B)$$

puisque M(AB) est la plus petite des majorantes de AB.

Avant d'étudier ce type de normes vectorielles, nous résumons les résultats relatifs au cas "carré" pour une norme vectorielle régulière :

4- 1- 4 - Proposition :

$\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. Soit la décomposition en somme directe de $\mathbb{K}^n$:

$$\mathbb{K}^n = \bigoplus_{i=1}^{k} W_i$$

et p une norme vectorielle régulière (de taille k) construite sur cette décomposition (3- 1- 3). P_i $(i = 1, 2 \ldots k)$ désignant l'opérateur de projection de $\mathbb{K}^n$ dans W_i, toute matrice $A(n,n)$ appliquant $\mathbb{K}^n$ dans lui-même est décomposée en "blocs" $A_{ij} = P_i A P_j$ $\quad (1 \leq i,j \leq k)$:

1) L'ensemble des majorantes de A relativement à p est l'ensemble des matrices B de type (kk) telles que

$$M(A) \leq B$$

*où $M(A) = \left(m_{ij}(A)\right)$ est la **matrice** définie par :*

$$m_{ij}(A) = \sup_{\substack{x_j \in W_j \\ x_j \neq 0}} \left\{ \frac{p_i(A_{ij} x_j)}{p_j(x_j)} \right\} \underset{df}{=} S_{ij}(A_{ij})$$

$$1 \leq i,j \leq k$$

Ainsi M(A) est la plus petite des majorantes de A.

2) L'application $A \in \mathcal{M}_{nn} \longrightarrow M(A)$

est une norme vectorielle régulière de type (kk), sous-multiplicative sur l'algèbre $\mathcal{M}_{nn}$ des matrices de types (n,n)

(On dira que p <u>génère</u> M ou que M <u>est issue</u> de p)

3) Le rayon spectral de M(A) n'est pas inférieur à celui de A(*)

Exemples :

a) Sur $\mathbb{K}^n$ soit p la norme vectorielle type :

$$x = (\xi_1 \ldots \xi_n) \in \mathbb{K}^n \longrightarrow p(x) \; (|\xi_1| \ldots |\xi_n|) \in \mathbb{R}_+^n$$

---

* On retrouve les points 1) et 2) démontrés dans [19]; le point 3) dans [40] et [36] p. 478.

elle est régulière et l'on a :

$$\forall A \in \mathcal{M}_{nn} \qquad M(A) = |A|$$

où $|A|$ désigne la matrice réelle ayant pour éléments les modules des éléments correspondants de A. Toute majorante $\mathbf{B}$ de A est donnée par

$$|A| \leqslant \mathbf{B}$$

On retrouve donc un résultat bien connu [64] : si A et $\mathbf{B}$ sont deux matrices (n,n) telles que $|A| \leqslant \mathbf{B}$ alors on a

$$\rho(A) \leqslant \rho(\mathbf{B})$$

En particulier, pour tout A dans $\mathcal{M}_{nn}$,

$$\rho(A) \leqslant \rho(|A|)$$

b) pour des exemples numériques de majoration de rayons spectraux, se reporter au chapitre 5.

4- 1- 5 - *Remarques*

1) Soit p une norme vectorielle régulière de taille k sur $\mathbb{K}^n$, générant M sur $\mathcal{M}_{nn}$

Soit D une diagonale (kk) positive. Alors l'application

$$x \in \mathbb{K}^n \longrightarrow q(x) = D\,p(x)$$

est encore une norme vectorielle régulière de taille k sur $\mathbb{K}^n$ (*), générant sur $\mathcal{M}_{nn}$ la norme vectorielle N. On vérifie facilement que

$$\forall A \in \mathcal{M}_{nn} \qquad N(A) = DM(A)\,D^{-1}$$

d'où $\rho\big(N(A)\big) = \rho\,\big(M(A)\big)$

2) Soit p une norme vectorielle de taille k, régulière sur $\mathbb{K}^n$, générant M sur $\mathcal{M}_{nn}$. Soit H une matrice (n,n) non singulière. Alors l'application

$$x \in \mathbb{K}^n \longrightarrow q(x) = p(Hx)$$

est une norme vectorielle de taille k, régulière sur $\mathbb{K}^n$, générant sur $\mathcal{M}_{nn}$ la norme vectorielle N. Il est facile de prouver que :

$$\forall A \in \mathcal{M}_{nn} \qquad N(A) = M(HAH^{-1})$$

3) Soit p une norme vectorielle régulière de taille k sur $\mathbb{K}^n$

(*) telle que $V_i^q = V_i^p$, $W_i^q = W_i^p$ $(i = 1,2 \ldots k)$

générant sur $\mathcal{M}_{nn}$ la norme vectorielle M de taille k^2 (cf 4-1-3).

Montrons que l'on a, pour tout a,b pris dans $\mathbb{K}^n$

$$M(ab^H) = p(a)[p^*(b)]^t$$

En effet désignons par m_i^t la ième ligne de $M(ab^H)$. C'est le plus petit vecteur de $\mathbb{R}^k$ tel que :

$$\forall\, x \in \mathbb{K}^n \qquad p_i(ab^H x) \leq m_i^t\, p(x)$$

soit

$$\forall\, x \in \mathbb{K}^n \qquad |b^H x|\; p_i(a) \leq m_i^t\, p(x)$$

On conclut donc (cf(3-3)) que

$$m_i^t = p_i(a)[p_i^*(b)]^t$$

$$i = 1,2 \,\ldots\, k$$

d'où le résultat.

4 - 2 - NORMES VECTORIELLES SOUS-MULTIPLICATIVES MINIMALES

Soit ϕ une norme sur $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ où $\mathbb{C}$) et soit A une matrice de type (n,n). D'où la norme sous-multiplicative :

$$S_{\phi\phi}(A) = \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \left\{ \frac{\phi(Ax)}{\phi(x)} \right\}$$

On sait [26] que les normes $S_{\phi\phi}$ sont <u>minimales</u> dans l'ensemble des normes sous-multiplicatives sur l'algèbre des matrices carrées, en vertu des résultats suivants([26] p. 23-24)

Théorème A = *Soit* N *une norme sous-multiplicative sur* $\mathcal{M}_{nn}$ *; il existe alors une norme* ϕ *sur* $\mathbb{K}^n$ *telle que*

$$\forall\, A \in \mathcal{M}_{nn} \qquad S_{\phi\phi}(A) \leq N(A)$$

(On peut prendre $\phi(x) = N(xa^H)$; $a \in \mathbb{K}^n - \{0\}$)

Théorème B = *Soit* ϕ *et* ψ *deux normes sur* $\mathbb{K}^n$

$$\text{Si} \qquad \forall\, A \in \mathcal{M}_{nn} \qquad S_{\phi\phi}(A) \leq S_{\psi\psi}(A)$$

alors il existe une constante $\lambda > 0$ *dans* $\mathbb{R}$ *, telle que*

$$\forall\, x \in \mathbb{K}^n \qquad \phi(x) = \lambda\psi(x) \text{ et l'on a}$$

$$\forall A \in \mathcal{M}_{nn} \quad S_{\phi\phi}(A) = S_{\psi\psi}(A)$$

Nous allons montrer que ces résultats sont encore vrais dans le cas de normes vectorielles. Il nous faudra, préalablement, établir une proposition (4- 2- 2), triviale dans le cas des normes (vectorielles de taille 1) mais nécessaire dans le cas général pour permettre l'extension des théorèmes A et B

4- 2- 1 - Définition

On appelle norme vectorielle de taille k^2, sous-multiplicative, sur $\mathcal{M}_{nn}$, toute application S de l'espace $\mathcal{M}_{nn}$ (des matrices (n,n) opérant sur $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$)) dans l'espace des matrices réelles (k,k), telle que

$$\begin{aligned} \text{i)}\ & S(A) = 0 \Rightarrow A = 0 \\ \text{ii)}\ & S(\lambda A) = |\lambda|\ S(A) \\ \text{iii)}\ & S(A+B) \leqslant S(A) + S(B) \qquad (*) \\ \text{iiii)}\ & S(AB) \leqslant S(A)S(B) \end{aligned}$$

On notera $s_{ij}(A)$ l'élément en position (i,j) de S(A), et :

$$V_{ij}^S = \{A \in \mathcal{M}_{nn} \quad s_{ij}(A) = 0\}$$

$$W_{ij}^S = \bigcap_{\substack{1 \leqslant r \leqslant k \\ 1 \leqslant s \leqslant k \\ r \neq i,\, s \neq j}} V_{rs} = \{A \in \mathcal{M}_{nn}\ \forall (rs) \neq (i,j) \quad s_{rs}(A) = 0\}$$

$$(1 \leqslant i,j \leqslant k)$$

La norme vectorielle S est alors dite régulière si la somme directe de tous les W_{ij}^S ($\leqslant$i,j$\leqslant$k) coïncide avec $\mathcal{M}_{nn}$. Démontrons alors la :

4- 2- 2 - Proposition : *Toute norme vectorielle régulière* S, *de taille* k^2, *sous-multiplicative sur* $\mathcal{M}_{nn}$ *est ainsi construite :*
Il existe une décomposition de $\mathbb{K}^n$ *en* k *sous espaces :*

---

* On sait que ii) et iii) entraînent que $\forall$ A, B $|S(A)-S(B)| \leqslant S(A+B)$; d'où, en particulier : $S(A) \geqslant 0 \quad \forall A \in \mathcal{M}_{nn}$

$$\mathbb{K}^n = W_1 + \ldots\ldots\ldots + W_k$$

telle que

$$W^S_{ij} = \{P_i A P_j \quad A \in \mathcal{M}_{nn}\} \quad (1 \leq i,j \leq k)$$

où P_i *Désigne l'opérateur de projection de* $\mathbb{K}^n$ *dans* W_i

Démonstration

$W^S_{ij} \otimes W^S_{rs}$ désigne l'ensemble des matrices de la forme AB avec $A \in W^S_{ij}$ et $B \in W^S_{rs}$. Alors, la sous-multiplicativité entraîne que

$$\alpha)\ W^S_{ij} \otimes W^S_{rs} = \{0\} \quad \text{si } r \neq j$$

$$\beta)\ W^S_{ij} \otimes W^S_{js} \subset W^S_{is}$$

En effet

α) si $A \in W^S_{ij}$ et $B \in W^S_{rs}$

$$0 \leq S(AB) \leq S(A)S(B) = 0$$

soit $\quad S(AB) = 0 \quad$ d'où $AB = 0$

β) si $A \in W^S_{ij}$ et $B \in W^S_{js}$

$$S(AB) \leq S(A)S(B)$$

Or $S(A)S(B)$ n'a que des éléments nuls, sauf peut être celui en position (i,s)

De même, par conséquent, pour $S(AB)$ ce qui entraîne que $A\,B \in W^S_{is}$.

Nous supposons, de plus, S <u>régulière</u> : c'est-à-dire que $\mathcal{M}_{nn}$ est la somme directe des W^S_{ij} $(1 \leq i,j \leq k)$

Tout $A \in \mathcal{M}_{nn}$ s'écrit donc, de façon unique

$$A = \sum_{i,j=1}^{k} A_{ij}$$

où $A_{ij} \in W^S_{ij}$

Posons alors, pour $i = 1, 2 \ldots\ldots k$

$$V_i = \{x \in \mathbb{K}^n \quad \forall j \in \{1,2 \dots k\} \quad \forall A \in W^S_{ji} \quad Ax = 0\}$$

et $$W_i = \bigcap_{\substack{j=1 \\ j \neq i}}^{k} V_j$$

Les V_i et W_i sont des sous-espaces vectoriels de $\mathbb{K}^n$ et l'on a

$$\bigcap_{j=1}^{k} V_j = \{0\}$$

En effet soit $A = \sum_{i,j} A_{ij}$ et $x \in \bigcap_{j=1}^{k} V_j$; d'où

$$Ax = \sum_{ij} A_{ij}x; \quad \text{mais puisque } x \in V_j$$

$$A_{ij}x = 0 \qquad (j = 1, 2 \dots k)$$

Ainsi $Ax = 0$ quelle que soit $A \in \mathcal{M}_{nn}$, d'où $x = 0$

On en déduit les propriétés habituelles des sous-espaces V_i et W_i :

α) $V_i \cap W_i = \{0\}$; $\quad W_i \cap W_j = \{0\} \qquad$ si $i \neq j$

β) $0 = \sum_{i=1}^{k} x_i$, où $x_i \in W_i$ entraîne $x_1 = \dots\dots = x_k = 0$

d'où la somme directe

$$W = \dot{+}_{i=1}^{k} W_i$$

et les propriétés suivantes :

<u>a) les éléments de W^S_{ij} appliquent $\mathbb{K}^n$ dans W_i</u>

Soit $A \in W^S_{ij}$, $x \in \mathbb{K}^n$ et $y = Ax$. Alors, pour $B \in W^S_{rv}$ $(v \neq i)$ on a

$$BA = 0 \qquad \text{d'où} \quad BAx = 0 \qquad \text{soit } By = 0$$

y appartient donc à tous les V_v $(v \neq i)$ donc à W_i

<u>b) Les éléments de W^S_{ij} appliquent W_s $(s \neq j)$ sur $\{0\}$</u>

Soit $x \in W_s$ $(s \neq j)$ donc $x \in V_j$; de la définition de V_j résulte que

$$\forall i \in \{1,2 \dots n\}, \ \forall A \in W^S_{ij}, \qquad Ax = 0$$

c) Aucun des W_i n'est réduit à l'origine.

Supposons exister i tel que $W_i = \{0\}$. Soit $A_{ij} \in W_{ij}^S$. Alors pour tout x dans $\mathbb{K}^n$, on a :

$$A_{ij}\, x \in W_i \qquad \text{d'après a)}$$

soit $\forall\, x \in \mathbb{K}^n$ $\quad A_{ij}x = 0 \quad ,\qquad$ soit

$$A_{ij} = 0$$

On aurait donc $W_{ij}^S = \{0\}$, ce qui est impossible, puisque S, qui est régulière, est surjective (3-1-4 ; 3-1-2)

d) W est appliqué dans lui-même par tout $A \in \mathcal{M}_{nn}$

Soit $x \in W$. D'où

$$x = \sum_{s=1}^{k} x_s \qquad \text{avec } x_s \in W_s$$

Soit alors $\quad A = \sum_{i,j} A_{ij} \in \mathcal{M}_{nn} \qquad (A_{ij} \in W_{ij}^S)$

d'après b) la quantité $A_{ij} \sum_{s=1}^{k} x_s$ se réduit à $A_{ij}x_j$ et, d'après a) elle appartient à W_i. Posons

$$y_i = \sum_{j=1}^{k} A_{ij}x_j \in W_i \qquad \text{d'où}$$

$$y = \sum_{i=1}^{k} y_i$$

ce qui prouve que $y = A\, x \in W$.

Or $\{0\}$ et $\mathbb{K}^n$ lui-même sont les deux seuls sous-espaces de $\mathbb{K}^n$ invariants par toute matrice A de $\mathcal{M}_{nn}$. W ne peut être réduit à $\{0\}$ puisqu' aucun des W_i ne l'est (c). On conclut

$$W = \mathbb{K}^n \qquad (1)$$

Enfin $\quad W_{ij}^S = \mathcal{L}(W_j, W_i) \qquad (1 \leq i,j \leq k)$

En effet

α) $W_{ij}^s \subset \mathcal{L}(W_j W_i)$ puisque, d'après a) tout $A \in W_{ij}^S$ applique $\mathbb{K}^n$, donc à fortiori W_j, dans W_i

β) Or d'après (1) $\mathbb{K}^n$ est somme directe des W_i; par conséquent (4-1-2) $\mathcal{M}_{nn}$ est la somme directe des $\mathcal{L}(W_j W_i)$. Il est aussi, par hypothèse, somme directe des W_{ij}^S. Par conséquent l'inclusion α) est en fait une égalité.

Posant dim $(W_i) = \lambda_i$, on a dim $W_{ij}^S = \lambda_i \lambda_j$

4- 2- 3 - *Remarque :* Si l'on compare ce dernier résultat à la proposition 4- 1- 4, on peut se demander s'il existe sur $\mathbb{K}^n$, une norme vectorielle p, de taille k, construite sur les W_i :

$$W_i^p = W_i \qquad i = 1,2 \;.... \; k$$

et générant S sur $\mathcal{M}_{nn}$ (i.e. S(A) est, pour tout $A \in \mathcal{M}_{nn}$, la plus petite des majorantes de A relativement à p). La réponse est négative, comme le prouve l'exemple suivant : soit la partition suivante de toute matrice A appartenant à $\mathcal{M}_{nn}$:

$$A = \begin{vmatrix} A_{11} \cdots\cdots & A_{12} \\ \cdot & \\ \cdot & \\ A_{k1} \cdots\cdots & A_{kk} \end{vmatrix}$$

Posons $s_{ij}(A) = \sqrt{\lambda_i \lambda_j}\; \phi_\infty(A_{ij})$

$1 \leq i,j \leq k$

où $\phi_\infty(A_{ij})$ désigne le module maximum des éléments de A_{ij}, bloc de type (λ_i, λ_j), en position (i,j).

La relation matricielle

$$C = AB$$

s'écrit, selon ces blocs

$$C_{ij} = \sum_{r=1}^{k} A_{ir} B_{rj} \qquad 1 \leq i,j \leq k$$

L'application $A \in \mathcal{M}_{nn} \longrightarrow S(A) = \left(s_{ij}(A)\right)_{1 \leq i,j \leq k}$ est une norme vectorielle, régulière par construction, de taille k^2. Vérifions qu'elle est sous-multiplicative :

$$s_{ij}(C_{ij}) = \sqrt{\lambda_i\lambda_j}\ \phi_\infty\left(\sum_{r=1}^{k} A_{ir}B_{rj}\right) \leq \sqrt{\lambda_i\lambda_j}\sum_{r=1}^{k} \phi_\infty(A_{ir}B_{rj})$$

or $\quad \phi_\infty(A_{ir}B_{rj}) \leq \lambda_r\ \phi_\infty(A_{ir})\phi_\infty(B_{rj}) \quad$ d'où

$$s_{ij}(C_{ij}) \leq \sum_{r=1}^{k} [\sqrt{\lambda_i\lambda_r}\ \phi_\infty(A_{ir})][\sqrt{\lambda_r\lambda_j}\ \phi_\infty(B_{rj})] \quad \text{soit}$$

$$s_{ij}(C_{ij}) \leq \sum_{r=1}^{k} s_{ir}(A_{ir})s_{rj}(B_{rj}), \quad \text{soit}$$

matriciellement

$$S(AB) \leq S(A)\ S(B) : S \text{ est bien sous-multiplicative.}$$

Or si S était issue d'une norme vectorielle sur $\mathbb{K}^n$, on aurait (4-1-2)

$$S(I_n) = I_k$$

où I_n désigne l'unité (n,n) et I_k l'unité (k,k).

Or on a

$$S(I_n) = \text{diag}\ (\lambda_1 \ldots\ldots\ldots \lambda_k).$$

matrice diagonale, mais différente de l'unité, dès que l'un des λ_i est supérieur à 1; S, bien que régulière et sous-multiplicative, n'est donc pas issue d'une norme vectorielle sur $\mathbb{K}^n$.

En conservant les notations utilisées, nous sommes maintenant à même d'étendre aux normes vectorielles les théorèmes A et B rappelés en 4 - 2.

4- 2- 4 - Proposition :

Soit S *une norme vectorielle régulière, sous-multiplicative, de taille* k^2 *sur* $\mathcal{M}_{nn}$. *Il existe alors une norme vectorielle* p, *régulière de taille* k *sur* $\mathbb{K}^n$, *générant sur* $\mathcal{M}_{nn}$ *la norme vectorielle* P *telle que*

$$\forall A \in \mathcal{M}_{nn} \qquad P(A) \leq S(A) \qquad (2)$$

d'où $\qquad \rho(A) \leq \rho(P(A)) \leq \rho(S(A))$

En effet, il suffit de choisir $a \in \mathbb{K}^n - \{0\}$, $b \in \mathbb{R}^k_+$, $b > 0$ (toutes composantes positives) et de poser

$$p(x) = S(xa^H)b$$

Avec les notations de la proposition 4- 2- 2, on vérifie que p est une norme vectorielle de taille k, construite sur les W_i ($W_i^p = W_i$) donc régulière.

D'où

$$p(Ax) = S(Axa^H)b \leqslant S(A)S(xa^H)b = S(A)\ p(x)$$

Ainsi, S(A) est une majorante de A relativement à p d'où (2) puisque $\mathbf{P}(A)$ est la plus petite de telles majorantes.

4- 2- 5 - Proposition. *Soient* p *et* q *deux normes vectorielles régulières, de taille* k *sur* $\mathbb{K}^n$*, telles que*

$$W_i^p = W_i^q \quad (i = 1,2 \ldots k) \qquad (3)^*$$

On note P *(* resp. Q) *la norme vectorielle* (sur $\mathcal{M}_{nn}$) *issue de* p (resp.q).

Si pour tout $A \in \mathcal{M}_{nn}$,

$$P(A) \leqslant Q(A) \qquad (4)$$

alors il existe une constante positive λ *telle que* $p = \lambda q$ *et l'on a, en fait :*

$$\forall A \in \mathcal{M}_{nn} \qquad P(A) = Q(A)$$

En effet : p et q étant régulières et équivalentes, P et Q le sont aussi. Posons

$$W_i = W_i^p = W_i^q \qquad (i = 1,2 \ldots k)$$

---

* p et q étant régulieres, (3) est condition nécessaire et suffisante d'existence de $\alpha>0$ et $\beta \geq \alpha$ tels que

$$\forall\ x \in \mathbb{K}^n \qquad \alpha q(x) \leqslant p(x) \leqslant \beta q(x)$$

ceci car $\mathbb{K}^n$ est de dimension finie.

p et q sont dites <u>équivalentes</u>.

$$W_{ij} = W^P_{ij} = W^Q_{ij} \qquad (i,j = 1,\ 2 \ldots k)$$

Tout $A \in \mathcal{M}_{nn}$ s'écrit, de façon unique

$$A = \sum_{i,j} A_{ij} \qquad \text{où } A_{ij} \in W_{ij}$$

Et de (4) on tire, en particulier,

$$\forall A_{ii} \in W_{ii} \qquad S_{p_ip_i}(A_{ii}) \leqslant S_{q_iq_i}(A_{ii})$$

Il existe donc (th. B) une constante λ_i telle que

$$\forall x \in W_i \qquad p_i(x) = \lambda_i q_i(x). \quad \text{Alors}$$

$$\forall A_{ii} \in W_{ii} \qquad S_{p_ip_i}(A_{ii}) = S_{q_iq_i}(A_{ii}) \qquad (i = 1,2 \ldots k)$$

Mais alors

$$S_{p_ip_j}(A_{ij}) = \sup_{x_j \in W_j} \left\{ \frac{p_i(A_{ij}x_j)}{p_j(x_j)} \right\} = \frac{\lambda_i}{\lambda_j} S_{q_iq_j}(A_{ij})$$

d'où, à cause de (4),

$$\frac{\lambda_i}{\lambda_j} \leqslant 1$$

mais alors $\quad S_{p_jp_i}(A_{ji}) \leqslant S_{q_jq_i}(A_{ji}) \quad$ d'où

$$\frac{\lambda_i}{\lambda_j} \geqslant 1$$

Ce qui prouve que tous les λ_i sont égaux. Soit λ leur valeur commune, d'où le résultat.

4-2-6 - Corollaire : *Soit* q *une norme vectorielle régulière sur* $\mathbb{K}^n$, *générant* Q *sur* $\mathcal{M}_{nn}$. *Si* S *est, sur* $\mathcal{M}_{nn}$ *une norme vectorielle régulière et sous-multiplicative telle que*

$$\forall A \in \mathcal{M}_{nn} \qquad S(A) \leqslant Q(A)$$

Alors (4-2-4) il existe p, *norme vectorielle régulière sur* $\mathbb{K}^n$, *générant* P *sur* $\mathcal{M}_{nn}$ *telle que*

$$\forall A \in \mathcal{M}_{nn} \qquad P(A) \leqslant S(A) \qquad \textit{d'où}$$

$$\forall A \in \mathcal{M}_{nn} \qquad P(A) \leq S(A) \leq Q(A)$$

Alors (4-2-5) il existe $\lambda > 0$ *tel que* $p = \lambda q$ *et l'on a, en fait*

$$\forall A \in \mathcal{M}_{nn} \qquad P(A) = S(A) = Q(A)$$

4- 2- 7 - Conclusion : Les propositions (4-2-4), (4-2-5), (4-2-6) permettent de conclure :

Les normes vectorielles sur $\mathcal{M}_{nn}$ *issues de normes vectorielles régulières sur* $\mathbb{K}^n$ *(* $\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$ *) sont éléments minimaux de l'ensemble des normes vectorielles régulières et sous-multiplicatives sur* $\mathcal{M}_{nn}$.

4- 2- 8 - *Remarque :* Soit p, norme vectorielle régulière sur $\mathbb{K}^n$, générant P sur $\mathcal{M}_{nn}$.

Soit A fixé dans $\mathcal{M}_{nn}$. Alors P(A) est la plus petite des matrices N telles que

$$\forall X \in \mathcal{M}_{nn} \qquad P(AX) \leq N\, P(X) \qquad (5)$$

En effet, la sous-multiplicativité de P

$P(AX) \leq P(A)P(X)$ montre que P(A) vérifie (5).

Soit N vérifiant (5). Pour $X = I_n$, il vient

$$P(A) \leq N\, P(I_n) = N$$

d'où le résultat.

4 - 3 - GENERALISATION d'un THEOREME de HOUSEHOLDER-OSTROWSKI

Soit ϕ une norme sur $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$), et soit $\mathcal{M}_{nn}$ l'espace des matrices (n,n) à éléments dans $\mathbb{K}$. On sait que l'on a :

$$\forall A \in \mathcal{M}_{nn}, \ \rho(A) \leq S_{\phi\phi}(A) \underset{df}{=} \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \left\{ \frac{\phi(Ax)}{\phi(x)} \right\}$$

Householder [33] et Ostrowski [45] ont prouvé, de façons différentes, le

Théorème : *Pour tout* $A \in \mathcal{M}_{nn}$ *et tout* $\varepsilon > 0$, *il existe une norme* ϕ *sur* $\mathbb{K}^n$ *telle que*

$$S_{\phi\phi}(A) \leq \rho(A) + \varepsilon$$

Ce résultat s'étend de la façon suivante

4- 3- 1 - Théorème 2 :

Soient $A_{12} \ldots A_{k1}$ k *matrices (à éléments dans* $\mathbb{K}$*) dont le produit* $C_1 = A_{12}A_{23} \ldots A_{k1}$ *existe et est carré. Soit* ρ *le rayon spectral de* C_1

Alors, pour tout $\varepsilon>0$ *et tout jeu* $\gamma_1 \ldots \gamma_k$ *de constantes positives telles que*

$$\prod_{i=1}^{k} \gamma_i = \rho + \varepsilon$$

il existe k *normes* $\phi_1 \ldots \phi_k$ *sur des sous-espaces convenables telles que :*

$$\begin{cases} S_{\phi_1\phi_2}(A_{12}) \leq \gamma_1 \\ S_{\phi_2\phi_3}(A_{23}) \leq \gamma_2 \\ \vdots \\ S_{\phi_k\phi_1}(A_{k1}) \leq \gamma_k \end{cases}$$

(pour k = 1, on retrouve évidemment le théorème précédent)

Démonstration :

1- Nous allons d'abord construire k normes $\psi_1 \ldots \psi_k$ telles que

$$\begin{cases} S_{\psi_1\psi_2}(A_{12}) \leq (\rho + \varepsilon)^{1/k} \\ \vdots \\ S_{\psi_k\psi_1}(A_{k1}) \leq (\rho + \varepsilon)^{1/k} \end{cases}$$

et l'on construira ensuite les normes ϕ_i par homotétie sur les ψ_i.

Désignons par $(\lambda_i\lambda_j)$ le type de la matrice A_{ij}. Posons $n = \sum_{i=1}^{k} \lambda i$, d'où la matrice A(n,n) :

$$A = \begin{vmatrix} 0 & A_{12} & 0 & \cdots & 0 \\ 0 & 0 & A_{23} & & \vdots \\ & & 0 & \ddots & A_{k-1k} \\ A_{k1} & \cdots & \cdots & \cdots & 0 \end{vmatrix}_{(n,n)}$$

d'où

$$A^k = \begin{vmatrix} C_1 & & & \\ & \ddots & & \\ & & \ddots & \\ & & & C_k \end{vmatrix}_{(n,n)}$$

avec

$$\begin{cases} C_1 = A_{12}\ A_{23} \cdots\cdots A_{k1} \\ C_2 = \quad A_{23}\ A_{24} \cdots A_{k1} \cdot A_{12} \\ \vdots \\ C_k = \quad A_{k1}\ A_{12} \cdots\cdots A_{k-1\ k} \end{cases}$$

Soit W_i le sous-espace (de dimension λ_i) de $\mathbb{K}^n$ sur lequel opère C_i ($i = 1,2 \dots k$). Tous les C_i ont mêmes valeurs propres, donc même rayon spectral ρ. $\varepsilon > 0$ étant donné, il existe donc d'après le théorème rappelé ci-dessus, k normes $\tau_1 \dots \tau_k$, respectivement définies sur $W_1 \dots W_k$, telles que :

$$S_{\tau_i \tau_i}(C_i) \leq \rho + \varepsilon \quad (i = 1, 2 \dots k)$$

Pour x appartenant à $W_1 + \dots + W_k = \mathbb{K}^n$, posons :

$$\tau(x) = \begin{vmatrix} \tau_1(x_1) \\ \vdots \\ \tau_k(x_k) \end{vmatrix}$$

où x_i désigne la projection de x sur W_i. τ est une norme vectorielle régulière de taille k sur $\mathbb{K}^n$, générant M_τ sur $\mathcal{M}_{nn}$:

$$\tau(A^k x) \leqslant M_\tau(A^k)\tau(x) \leqslant (\rho+\varepsilon)\tau(x) \quad . \qquad (1)$$

avec

$$M_\tau(A^k) = \begin{vmatrix} S_{\tau_1\tau_1}(C_1) & & \\ & \ddots & \\ & & S_{\tau_k\tau_k}(C_k) \end{vmatrix}_{(k,k)}$$

qui est la plus petite majorante de A^k relativement à τ.

Soit alors P la matrice (k,k) de permutation suivante :

$$P = \begin{vmatrix} 0 & \cdots\cdots & 1 \\ 1 & \ddots & \\ & \ddots\ddots & \\ & 1 & 0 \end{vmatrix}_{(k\ k)}$$

d'où $\quad P^k = I \qquad$ et $\quad P^t = P^{-1}$

Posons

$$\psi(x) = (\rho+\varepsilon)^{\frac{k-1}{k}}\tau(x) + (\rho+\varepsilon)^{\frac{k-2}{k}} P\tau(Ax) + (\rho+\varepsilon)^{\frac{k-3}{k}} P^2\tau(A^2x) + \ldots P^{k-1}\tau(A^{k-1}x).$$

On peut montrer alors que l'application ψ est une norme vectorielle de taille k sur $\mathbb{K}^n$, régulière car construite sur les W_i. Autrement dit :

$$x = x_1 + \ldots\ldots x_k \longrightarrow \psi(x) = \begin{vmatrix} \psi_1(x_1) \\ \vdots \\ \psi_k(x_k) \end{vmatrix}$$

$(x_i \in W_i)$

où ψ_i est une norme sur W_i

Au lieu de le démontrer complètement - ce qui nécessiterait un formalisme assez lourd - contentons-nous de le vérifier pour k = 3. Il vient

$$\psi(x) = \begin{vmatrix} \psi_1(x_1) = (\rho+\varepsilon)^{2/3}\tau_1(x_1) + (\rho+\varepsilon)^{1/3}\tau_3(A_{31}x_1) + \tau_2(A_{23}A_{31}x_1) \\ \psi_2(x_2) = (\rho+\varepsilon)^{2/3}\tau_2(x_2) + (\rho+\varepsilon)^{1/3}\tau_1(A_{12}x_2) + \tau_3(A_{31}A_{12}x_2) \\ \psi_3(x_3) = (\rho+\varepsilon)^{2/3}\tau_3(x_3) + (\rho+\varepsilon)^{1/3}\tau_2(A_{23}x_3) + \tau_1(A_{12}A_{23}x_3) \end{vmatrix}$$

De façon générale, il est clair que la ième composante de $\psi(x)$ est la somme d'une norme de x_i (dont la valeur est $(\rho+\varepsilon)^{\frac{k-1}{k}}\tau_i(x_i)$) et de k-1 semi-normes de x_i, ce qui redonne bien une norme de x_i. On a, alors $\psi(Ax) = (\rho+\varepsilon)^{\frac{k-1}{k}}\tau(Ax) + (\rho+\varepsilon)^{\frac{k-2}{k}}P\tau(A^2x) + \ldots\ldots P^{k-1}\tau(A^kx)$

D'après (1) le dernier terme de cette somme est majoré par

$$(\rho+\varepsilon)\, P^{k-1}\tau(x)$$

Soit

$$\psi(Ax) \leqslant (\rho+\varepsilon)^{\frac{k-1}{k}}\tau(Ax) + (\rho+\varepsilon)^{\frac{k-2}{k}}P\tau(A^2x) + \ldots\ldots (\rho+\varepsilon)\, P^{k-1}\tau(x) \qquad (2)$$

Faisons passer le dernier terme de cette somme en première place; remarquons que $P^{k-1} = P^t$, et mettons $(\rho+\varepsilon)^{1/k}\, P^t$ en facteur.

(2) devient :

$$\psi(Ax) \leqslant (\rho+\varepsilon)^{\frac{1}{k}} P^t[(\rho+\varepsilon)^{\frac{k-1}{k}}\tau(x) + (\rho+\varepsilon)^{\frac{k-2}{k}} P\tau(Ax) + \ldots P^{k-1}\tau(A^{k-1}x)]$$

soit

$$\psi(Ax) \leqslant (\rho+\varepsilon)^{\frac{1}{k}} P^t\psi(x)$$

Ainsi, la matrice $(\rho+\varepsilon)^{\frac{1}{k}} P^t$ est une majorante de A relativement à ψ.

D'où :

$$M_\psi(A) \leqslant (\rho+\varepsilon)^{\frac{1}{k}} P^t \qquad \text{soit}$$

$$M_\psi(A) = \begin{vmatrix} 0 & S_{\psi_1\psi_2}(A_{12}) & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & S_{\psi_{k-1}\psi_k}(A_{k-1\,k}) \\ S_{\psi_k\psi_1}(A_{k1}) & & \cdots & 0 \end{vmatrix} \leqslant (\rho+\varepsilon)^{\frac{1}{k}} \begin{vmatrix} 0 & 1 & \cdots & 0 \\ \vdots & \ddots & \ddots & \vdots \\ \vdots & & \ddots & 1 \\ 1 & \cdots & \cdots & 0 \end{vmatrix}$$

D'où le résultat annoncé :

$$S_{\psi_1\psi_2}(A_{12}) \leqslant (\rho+\varepsilon)^{\frac{1}{k}} ; S_{\psi_2\psi_3}(A_{23}) \leqslant (\rho+\varepsilon)^{\frac{1}{k}} ; \ldots\ldots S_{\psi_k\psi_1}(A_{k1}) \leqslant (\rho+\varepsilon)^{\frac{1}{k}}$$

2 Soient alors $\gamma_1 \ldots\ldots \gamma_k$ des constantes positives de produit égal à $\rho+\varepsilon$. D'où par définition

$$\left\{\begin{array}{l} \lambda_1 = (\rho+\varepsilon)^{\frac{1}{k}} \\ \lambda_2 = \dfrac{(\rho+\varepsilon)^{\frac{2}{k}}}{\gamma_1} \\ \vdots \\ \lambda_i = \dfrac{(\rho+\varepsilon)^{\frac{i}{k}}}{\gamma_1 \ldots\ldots \gamma_{i-1}} \\ \vdots \\ \lambda_k = \dfrac{(\rho+\varepsilon)}{\gamma_1 \ldots \gamma_{k-1}} = \gamma_k \end{array}\right.$$

Posons $\phi_i = \lambda_i \psi_i \quad (i = 1,2 \ldots\ldots k)$

On a alors, pour $i \in \mathbb{N}$ (modulo k) :

$$S_{\phi_i\phi_{i+1}}(A_{i,i+1}) = \frac{\lambda_i}{\lambda_{i+1}} S_{\psi_i\psi_{i+1}}(A_{i,i+1}) \leqslant \frac{\lambda_i}{\lambda_{i+1}} (\rho+\varepsilon)^{\frac{1}{k}}$$

Or :
$$\frac{\lambda_i}{\lambda_{i+1}} = \frac{\gamma_i}{(\rho+\varepsilon)^{\frac{1}{k}}}$$

D'où le résultat final :

$$\left| \begin{array}{l} S_{\phi_1\phi_2}(A_{12}) \leqslant \gamma_1 \\ \vdots \\ S_{\phi_k\phi_1}(A_{k1}) \leqslant \gamma_k \end{array}\right.$$

4 - 4 - UN PROBLEME RECIPROQUE

Soit p une norme vectorielle régulière sur $\mathbb{K}^n$, générant sur $\mathcal{M}_{nn}$ la norme vectorielle M (cf 4-1-4). On sait que l'on a :

$$\forall A \in \mathcal{M}_{nn} \qquad \rho(A) \leq \rho\big(M(A)\big)$$

Etant donné $\varepsilon > 0$ *et une matrice* A, *existe-t-il une norme vectorielle* p *régulière sur* $\mathbb{K}^n$, *générant* M *sur* $\mathcal{M}_{nn}$, *et telle que :*

$$\rho\big(M(A)\big) \leq \rho(A) + \varepsilon$$

1 En vertu du théorème de Householder-Ostrowski, rappelé en 4-3, la réponse est positive : il suffit de prendre pour p la norme (vectorielle de taille 1) ϕ que ce théorème exhibe et qui est telle que :

$$\rho\ M(A) = M(A) = S_{\phi\phi}(A) \leq \rho(A) + \varepsilon$$

2 On peut aussi donner une autre réponse : soit J la forme de Jordan de A

$$J = \begin{vmatrix} J_1 & & \\ & \ddots & \\ & & J_k \end{vmatrix}$$

où

$$J_i = \begin{vmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{vmatrix} \quad (\tau_i, \tau_i); \quad \sum_{i=1}^{k} \tau_i = n$$

et

$$\rho(A) = \rho(J) = \operatorname*{Max}_i |\lambda_i|$$

k est le nombre de blocs diagonaux de J. Nous allons construire une norme vectorielle M, de taille k^2, répondant à la question.

Toujours d'après le théorème de Householder-Ostrowski cité plus haut, il existe sur $\mathbb{K}^{\tau_i}$ une norme ϕ_i telle que

$$(1) \qquad S_{\phi_i\phi_i}(J_i) \underset{df}{=} \sup_{x\in\mathbb{K}^{\lambda_i}} \left\{ \frac{\phi_i\,(J_i x)}{\phi_i(x)} \right\} \leqslant \rho(J_i) + \varepsilon = |\lambda_i| + \varepsilon$$

Soit W_i le sous-espace (de dimension τ_i) sur lequel opère $\mathbf{J}_i$, et x_i la projection de x pris dans $\mathbb{K}^n$ sur W_i. D'où la norme vectorielle q régulière de taille k sur $\mathbb{K}^n$:

$$x = \begin{vmatrix} x_1 \\ \vdots \\ x_k \end{vmatrix} \longrightarrow q(x) = \begin{vmatrix} \phi_1(x_1) \\ \vdots \\ \phi_k(x_k) \end{vmatrix}$$

q génère sur $\mathcal{M}_{nn}$ la norme vectorielle Q telle que

$$Q(J) = \text{diag}\left(S_{\phi_i\phi_i}(J_i)\right)$$

Alors, à cause de (1)

$$\rho\left(Q(J)\right) = \max_i \left(S_{\phi_i\phi_i}(J_i)\right) \leqslant \max_i(|\lambda_i|) + \varepsilon = \rho(A) + \varepsilon$$

J et A sont semblables : il existe U inversible telle que

$$J = U\,A\,U^{-1}$$

Définissons sur $\mathbb{K}^n$ la norme vectorielle p suivante

$$p(x) = q(Ux)$$

Alors (4-1-5) p est régulière et génère sur $\mathcal{M}_{nn}$ la norme vectorielle M telle que

$$\forall B \qquad M(B) = Q(U\,B\,U^{-1})$$

d'où

$$M(A) = Q(U\,A\,U^{-1}) = Q(J)$$

$$\text{et} \quad \rho\left(M(A)\right) = \rho\left(Q(J)\right) \leqslant \rho(A) + \varepsilon$$

3 Enfin, on peut répondre d'une troisième façon dans le cas où A est K-cyclique (cf[64] p. 39). En effet, à une similitude près,(*)

(*) que l'on traite comme précédemment

elle peut être prise sous la forme

$$A = \begin{vmatrix} 0 & A_{12} & & & \\ \vdots & 0 & A_{23} & & \\ \vdots & & \ddots & \ddots & \\ \vdots & & & 0 & A_{k-1k} \\ A_{k1} & \cdots & \cdots & \cdots & 0 \end{vmatrix}$$

où $A_{i\,i+1}$ est de type $(\lambda_i, \lambda_{i+1})$ $\quad (\sum_{i=1}^{k} \lambda_i = n)$

Soit C_1 la matrice

$$C_1 = A_{12}\, A_{23} \ldots\ldots A_{k-1\ k}\, A_{k1}$$

Il est facile de vérifier que l'on a, à cause de la forme particulière de A (cf 4-3)

$$\rho(C_1) = [\rho(A)]^k = \rho(A^k)$$

Par conséquent, d'après le théorème 2 (4-3-1), il existe, pour $\eta > 0$ donné, k normes $\psi_1, \ldots\ldots \psi_k$ définies sur des sous-espaces convenables telles que :

$$\begin{cases} S_{\psi_1\psi_2}(A_{12}) \leqslant \left(\rho(C_1) + \eta\right)^{1/k} \\ \vdots \\ S_{\psi_k\psi_1}(A_{k1}) \leqslant \left(\rho(C_1) + \eta\right)^{1/k} \end{cases}$$

Soit alors p la norme vectorielle régulière

$$x \in \mathbb{K}^n \longrightarrow p(x) = \begin{vmatrix} \psi_1(x_1) \\ \vdots \\ \psi_k(x_k) \end{vmatrix}$$

où x_i désigne la projection de x sur le sous-espace W_i où est définie ψ_i.

On a

$$p(Ax) \leqslant M(A)\, p(x)$$

où M(A) s'écrit

$$M(A) = \begin{vmatrix} 0 & S_{\psi_1\psi_2}(A_{12}) & & & \\ \vdots & 0 & S_{\psi_2\psi_3}(A_{23}) & & \\ \vdots & & \ddots & \ddots & \\ \vdots & & & 0 & S_{\psi_{k-1}\psi_k}(A_{k-1\ k}) \\ S_{\psi_k\psi_1}(A_{k1}) & - & - & - & 0 \end{vmatrix}$$

d'où

$$\rho(A) \leqslant \rho\Big(M(A)\Big) = \left[\prod_{i=1}^{k} S_{\psi_i\psi_{i+1}}(A_{i\ i+1})\right]^{1/k} \leqslant \Big(\rho(C_1) + \eta\Big)^{\frac{k}{k.k}}$$

or $\rho(C_1) = \Big(\rho(A)\Big)^k$. D'où

$$\rho\Big(M(A)\Big) \leqslant \Big[[\rho(A)]^k + \eta\Big]^{1/k}$$

$\varepsilon > 0$ étant donné, il est toujours possible de choisir η tel que

$$\Big[[\rho(A)]^k + \eta\Big]^{1/k} \leqslant \rho(A) + \varepsilon$$

Et l'on a bien construit une norme vectorielle p régulière sur $\mathbb{K}^n$ générant M sur $\mathcal{M}_{nn}$ telle que

$$\rho\Big(M(A)\Big) \leqslant \rho(A) + \varepsilon$$

Remarques : 1 les trois réponses apportées restent théoriques, la construction effective des normes vectorielles mises en évidence étant numériquement très difficile.

2 Un problème beaucoup plus ardu que le précédent, et auquel nous ne savons absolument pas répondre, même partiellement, est le suivant :

Etant donné $\varepsilon > 0$ et une matrice $A(n,n)$, existe-t-il une norme vectorielle p sur $\mathbb{K}^n$, de taille k donnée, régulière et construite sur des sous-espaces W_i donnés (*) telle que

$$\rho\ M(A) \leqslant \rho(A) + \varepsilon$$

(*) tels que $\overset{k}{\underset{i=1}{+}} W_i = \mathbb{K}^n$ (régularité de p)

(M étant la norme vectorielle de matrice générée par p).

4 -5 - NORMES VECTORIELLES de PUISSANCES de MATRICES

Soit T une matrice (n,n) à éléments dans $\mathbb{R}$ ou $\mathbb{C}$. Le comportement de T^m lorsque m tend vers l'infini [1] est élucidé dans Ostrowski [42] et Varga [64]. Rappelons rapidement leurs résultats :

Soit J la forme de Jordan de la matrice T :

$$T = U^{-1} J U \qquad \text{d'où}$$
$$T^m = U^{-1} J^m U$$

Désignons par $J_1 \ldots J_k$ les blocs diagonaux de J (les seuls non nuls) ; J_i (i = 1,2 .... k) est la matrice suivante, de type $(\tau_i \tau_i)$

$$J_i = \begin{vmatrix} \lambda_i & 1 & & \\ & \ddots & \ddots & \\ & & \ddots & 1 \\ & & & \lambda_i \end{vmatrix}$$

Sans restriction de généralité on peut supposer que $\lambda_1, \lambda_2 \ldots \lambda_\ell$ sont les valeurs propres de J de module égal à $\rho(J)$:

$$\rho(T) = \rho(J) = |\lambda_1| = |\lambda_2| = \ldots\ldots |\lambda_\ell| > |\lambda_{\ell+1}| \geq \ldots \geq |\lambda_k|$$

On peut également supposer que, parmi les ℓ premiers blocs diagonaux de J, $J_1 \ldots J_s$ sont ceux de taille (q) maximum

$$q = \tau_1 = \tau_2 = \ldots = \tau_s > \tau_{s+1} \geq \ldots \geq \tau_\ell$$

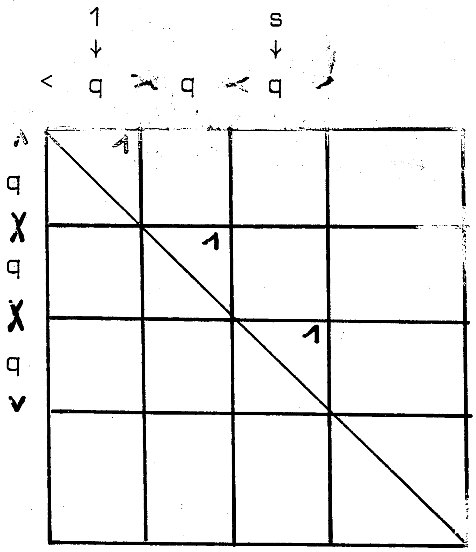
On trouve alors dans [42] et [64] le résultat suivant

$$J^m = C_m^{q-1} \; \rho(J)^{m-q+1} \; [K + S'_m]$$

où S'_m est une matrice qui tend vers zéro lorsque m tend vers l'infini, et K est la matrice:

---

(1) qui règle la convergence d'une itération linéaire : $x_{r+1} = T x_r + h$

$K =$ 

Autrement dit

$$K = \sum_{i=1}^{s} E_{1+(i-1)q,\ iq}$$

où $E_{\mu\nu}$ est la matrice élémentaire comportant un 1 en position (μ,ν) et des zéros ailleurs (K est de rang s)

T^m, alors, s'écrit :

$$T^m = U^{-1} J^m U = C_m^{q-1} \ [\rho(T)]^{m-q+1} [U^{-1} \ K \ U + S_m]$$

$$\text{avec } S_m = U^{-1} \ S'_m \ U$$

or $\qquad E_{\mu\nu} = e_\mu e_\nu^t \qquad$ (e_μ, e_ν, vecteurs de base)

Il en résulte que U^{-1} K U s'écrit

$$U^{-1} \ K \ U = \sum_{i=1}^{s} \underbrace{U^{-1} \ e_{1+(i-1)q}}_{a_i} \quad \underbrace{e_{iq}^t \ U}_{b_i^t} = \sum_{i=1}^{s} a_i b_i^t \ \underset{df}{=} \ \mathbf{H}$$

Les a_i sont linéairement indépendants, les b_i de même. Les a_i sont d'ailleurs vecteurs propres de T puisque $e_{1+(i-1)q}$ est vecteur propre ~~de~~ J $(i = 1,2..s)$

D'où enfin

$$\boxed{T^m = C_m^{q-1} \ [\rho(T)]^{m-q+1} \ [\sum_{i=1}^{s} a_i b_i^t + S_m] \qquad (1) \qquad S_m \to 0 \text{ quand } m \to \infty}$$

Ayant établi (1), Ostrowski [42] en déduit le :

4- 5- 1 - Théorème (*) *Soit* N *une norme sur l'espace vectoriel* $\mathcal{M}_{nn}$ *des matrices* (n.n) (**). *Alors pour tout* T *pris dans* $\mathcal{M}_{nn}$, *on a :*

$$\lim_{m\to\infty} [N(T^m)]^{1/m} = \rho(T)$$

En effet

$$[N(T^m)]^{1/m} = [C_m^{q-1}]^{1/m}\ [\rho(T)]^{\frac{m-q+1}{m}}\ [N(H + S_m)]^{1/m}$$

et lorsque m tend vers l'infini, le premier et le troisième facteur tendent vers l'unité, cependant que le second

$$[\rho(T)]^{\frac{m-q+1}{m}} \quad \text{tend vers } \rho(T)$$

le résultat précédent s'étend sans difficulté aux normes vectorielles :

4- 5- 2 - Proposition : *Soit* N *une norme vectorielle de taille* k^2 *sur l'espace vectoriel* $\mathcal{M}_{nn}$
$T \longrightarrow N(T)$, matrice de type (k,k)

Alors, pour tout T pris dans $\mathcal{M}_{nn}$,

$$\lim_{m\to\infty} [\rho(N(T^m))]^{1/m} = \rho(T)$$

En effet, on tire de (1)

$$[\rho(N(T^m))]^{1/m} = [C_m^{q-1}]^{1/m}[\rho(T)]^{\frac{m-q+1}{m}}\ [\rho(N(H + S_m))]^{1/m}$$

d'où le résultat lorsque m tend vers l'infini.

Remarque : Le résultat est valable pour toute norme vectorielle N de taille k^2. Si N est régulière et sous-multiplicative, on peut démontrer le résultat directement. En effet (4-2-4) on a alors :

(*) qui n'est qu'un cas particulier d'un théorème valable pour les opérateurs sur un espace de Banach. Voir [67] p. 211

(**) N, norme sur <u>l'espace vectoriel</u> $\mathcal{M}_{nn}$, n'est pas <u>à priori</u> supposée sous-multiplicative.

$\rho(T) \leqslant \rho\big(N(T)\big)$ d'où

$$\rho(T) = [\rho(T^m)]^{1/m} \leqslant [\rho[N(T^m)]]^{1/m} \leqslant [S_{\infty\infty}\big(N(T^m)\big)]^{1/m} \underset{df}{=} [\psi(T^m)]^{1/m}$$

où $S_{\infty\infty}$ est la norme de matrice issue de la norme de max sur $\mathbb{R}^k$. Puisque $S_{\infty\infty}$ est absolue, ψ est une norme sur $\mathcal{M}_{nn}$. D'où le résultat par application du théorème d'Ostrowski, rappelé ci-dessus, à la norme ψ.

Exemple : $N(T) = |T|$
d'où(*)

$$\lim_{m\to\infty} \rho(|T^m|)]^{1/m} = \rho(T)$$

(*) On trouve ce résultat démontré directement dans [13]

CHAPITRE 5 : APPLICATIONS. MATRICES CONTRACTANTES RELATIVEMENT à une NORME VECTORIELLE et CONVERGENCE d'ITERATIONS LINEAIRES.

On trouve la notion très générale d'opérateur contractant relativement à une norme vectorielle et le théorème de point fixe correspondant présentés dans l'ouvrage (en russe) de Kantorovitch, Vulik et Pinsker [36]. Ces auteurs donnent aussi les principales applications de ce théorème de point fixe à différents chapitres de l'analyse.

Nous nous attachons ici à reformuler et exploiter ces résultats d'un point de vue d'analyse numérique, dans le cadre des matrices carrées.

Une telle matrice est dite contractante (relativement à une norme vectorielle) si elle admet une majorante convergente. L'étude des matrices contractantes est faite au paragraphe 2. Les techniques numériques, utilisant des normes vectorielles, établies au paragraphe 1, permettent de prouver que telle matrice donnée est contractante.

Le fait que la matrice d'itération soit contractante est une condition suffisante de convergence d'une itération linéaire (5-2-3). On peut alors, dans cette situation caractéristique, donner des estimations en norme vectorielle de l'approximation obtenue sur la solution.

J.Schroeder [58] semble être le premier à avoir calculé effectivement de telles estimations. Il utilise en fait la norme vectorielle $p(x) = |x|$ et, pour obtenir les estimations recherchées, propose une itération d'erreur parallèle à l'itération principale. Nous montrons au paragraphe trois que l'algorithme de Schroeder peut être étendu à toute norme vectorielle régulière p (ce qui présente certains avantages numériques) et utilisé pour toute matrice d'itération contractante relativement à p. La meilleure version de cet algorithme redonne les estimations théoriques établies au paragraphe 2. Nous donnons une procédure Algol correspondante.

Enfin le paragraphe 4 concerne l'étude de la façon dont, dans une itération linéaire convergente, le vecteur d'erreur tend vers zéro. Là encore, l'utilisation de normes vectorielles permet de grouper les composantes par blocs et d'obtenir des résultats plus globaux.

5 - 1 - TECHNIQUES de MAJORATION d'un RAYON SPECTRAL

Soit $A \in \mathcal{M}_{nn}$ une matrice de type (n,n) à éléments dans $\mathbb{K}$ ($\mathbb{R}$ ou $\mathbb{C}$) dont nous désirons majorer le rayon spectral $\rho(A)$. La majoration

$$\rho(A) \leqslant S_{\phi\phi}(A) = \sup_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \left\{ \frac{\phi(Ax)}{\phi(x)} \right\}$$

où ϕ est une norme sur $\mathbb{K}^n$ est, dans la pratique, utilisée à peu de frais pour la norme du max :

$$\rho(A) \leqslant S_{\infty\infty}(A) = \max_{i=1,2\ldots n} \left\{ \sum_{j=1}^{n} |a_{ij}| \right\}$$

ou pour sa duale :

$$\rho(A) \leqslant S_{11}(A) = \max_{j=1,2\ldots n} \left\{ \sum_{i=1}^{n} |a_{ij}| \right\}$$

On utilise éventuellement la norme euclidienne qui donne une majoration d'un accès numérique plus difficile et sans véritable intérêt puisque s'exprimant elle-même en fonction d'un rayon spectral :

$$\rho(A) \leqslant S_{22}(A) = \sqrt{\rho(A^T A)}$$

[Si A est symétrique :

$$S_{22}(A) = \sqrt{\rho(A^2)} = \rho(A)$$]

Nous disposons maintenant des techniques (éventuellement plus fines) suivantes :

5-1-1 - Utilisation d'une norme vectorielle sur $\mathcal{M}_{nn}$ issue d'une norme vectorielle régulière sur $\mathbb{K}^n$.

Nous utilisons le résultat de (4-1-4) sous la forme suivante :

Décomposons A en blocs A_{ij} de type $(\lambda_i \lambda_j)$ avec $n = \sum_{i=1}^{k} \lambda_i$

$$A = \begin{vmatrix} A_{11} & \cdots\cdots & A_{1k} \\ & & \\ A_{k1} & \cdots\cdots & A_{kk} \end{vmatrix} \quad (1)$$

Choisissons sur chaque $\mathbb{K}^{\lambda_i}$ une norme ϕ_i d'où la matrice

$$M(A) = \begin{vmatrix} S_{\phi_1\phi_1}(A_{11}) & \cdots\cdots & S_{\phi_1\phi_k}(A_{1k}) \\ \vdots & & \\ S_{\phi_k\phi_1}(A_{k1}) & \cdots\cdots & S_{\phi_k\phi_k}(A_{kk}) \end{vmatrix}$$

où $$S_{\phi_i\phi_j}(A_{ij}) \overset{=}{df} \sup_{\substack{x_j \in \mathbb{K}^{\lambda_j} \\ x_j \neq 0}} \left\{ \frac{\phi_i(A_{ij}\, x_j)}{\phi_j(x_j)} \right\}$$

On sait (4-1-4) que le rayon spectral de $M(A)$ majore celui de A. Or $M(A)$ est une matrice dont le type $(k\ k)$ peut être bien inférieur au type (n,n) de A (Exemple courant : $n = 5o$, $k = 5$). Pratiquement les normes ϕ_i utilisées seront toutes soit la norme de max, soit sa duale; d'où

$$M_\infty(A) \underset{df}{=} \begin{vmatrix} S_{\infty\infty}(A_{11}) & \cdots\cdots\cdots & S_{\infty\infty}(A_{1k}) \\ \vdots & & \vdots \\ S_{\infty\infty}(A_{k1}) & \cdots\cdots\cdots & S_{\infty\infty}(A_{kk}) \end{vmatrix}$$

(on notera p_∞ la norme vectorielle régulière de taille k sur $\mathbb{K}^n$ qui génère M_∞)

et

$$M_1(A) \overset{=}{df} \begin{vmatrix} S_{11}(A_{11}) & \cdots\cdots\cdots & S_{11}(A_{1k}) \\ \vdots & & \vdots \\ S_{11}(A_{k1}) & \cdots\cdots\cdots & S_{11}(A_{kk}) \end{vmatrix}$$

qui sont aisément calculables, et dont on sait que

$$\rho(A) \leqslant \rho\big(M_\infty(A)\big) \quad ; \qquad \rho(A) \leqslant \rho\big(M_1(A)\big)$$

On peut alors, selon les cas, majorer à leur tour $\rho\big(M_\infty(A)\big)$ ou $\rho\big(M_1(A)\big)$ ou calculer éventuellement ces quantités directement, $M_1(A)$ et $M_\infty(A)$ étant de taille réduite par rapport à A.

On utilise beaucoup plus rarement, de la même manière, la norme vectorielle construite à partir de la norme euclidienne sur chaque $\mathbb{K}^{\lambda_i}$:

$$M_2(A) \underset{df}{=} \begin{vmatrix} S_{22}(A_{11}) & \cdots\cdots\cdots & S_{22}(A_{1k}) \\ \vdots & & \\ S_{22}(A_{k1}) & \cdots\cdots\cdots & S_{22}(A_{kk}) \end{vmatrix}$$

Procédures ALGOL

La procédure MATNØRMVEC, écrite ci-dessous, permet de calculer de telles normes vectorielles de matrices :

```
procédure MATNØRMVEC (A, M, N, K, LAMBDA);
    tableau A,M; entier tableau LAMBDA; entier N, K;
    début entier I, J, I1, J1, I2, J2; réel S, NØR;
    réelle procédure NØRME (W, P, Q);
    :
    :
    fin procédure NØRME;
    pour I:= 1 pas 1 a K faire
    pour J:= 1 pas 1 a K faire
    début tableau C[1:LAMBDA[I], 1: LAMBDA[J]];
    I1 := J1 := 0;
    pour I2 := 1 pas 1 a I - 1 faire
    I1 := I1 + LAMBDA [I2];
    pour J2 := 1 pas 1 a J - 1 faire
    J1 := J1 + LAMBDA [J2];
    pour J2 := 1 pas 1 a LAMBDA [I] faire
    pour J2 := 1 pas 1 a LAMBDA [J] faire
    C[I2,J2]:= A[I1 + I2, J1 + J2];
    M[I,J]:= NØRME(C,LAMBDA [I] , LAMBDA [J])
    fin
    fin procédure MATNØRMVEC;
```

Commentaires : A est une matrice réelle, carrée, donnée, de type (N,N). L'entier tableau LAMBDA [1:K] est donné et sert à définir le partionnement en blocs de la matrice A : LAMBDA [I] est en fait l'entier λ_i défini plus haut. Ainsi la somme des composantes (entières, positives) du tableau LAMBDA doit-elle être égale à N.

Alors l'effet de la procédure ci-dessus est le suivant : Selon que NØRME calcule $S_{\infty\infty}(W)$, $S_{11}(W)$ ou $S_{22}(W)$, MATNØRMVEC charge la matrice M de type (KK) avec (respectivement) $M_\infty(A)$, $M_1(A)$ ou $M_2(A)$, matrice relative au partionnement en blocs défini par LAMBDA.

Par exemple la

```
réelle procédure NØRME (W,P,Q); tableau W; entier P,Q;
    début NØR := 0.0;
          pour I2 := 1 pas 1 a P faire
          début S := 0.0;
                pour J2 := 1 pas 1 a Q faire
                S := S + ABS (W[I2,J2]);
                si NØR < S alors NØR := S
          fin;
          NØRME := NØR
    fin procédure NØRME;
```

calcule $S_{\infty\infty}(W)$ où W est une matrice réelle de type (P,Q). A noter que I2, J2, S et NØR sont des paramètres non locaux à NØRME déclarés dans MATNØRMVEC.

De même, la

```
réelle procédure NØRME (W, P, Q); tableau W; entier P,Q;
    début NØR := 0.0;
          pour J2 := 1 pas 1 a Q faire
          début S := 0.0;
                pour I2 := 1 pas 1 a P faire
                S := S + ABS(W[I2,J2]);
                si NØR < S alors NØR := S
          fin;
          NØRME := NØR
    fin procédure NØRME;
```

calcule $S_{11}(W)$ où W est de type (P,Q)

La réelle procédure NORME calculant $S_{22}(W)$ (par exemple par une méthode de la puissance sur $W^T W$) est beaucoup plus longue à écrire.

5-1- 2 - Utilisation d'une norme vectorielle régulière et sous-multiplicative.

On sait (4-2-4) que pour toute norme vectorielle S régulière et sous-multiplicative de matrice on a encore

$$\forall A \in \mathcal{M}_{nn} \quad \rho(A) \leq \rho\big(S(A)\big)$$

Dans la pratique, on se donne la décomposition en blocs (1) de A et l'on utilise la norme vectorielle S définie par

$$S(A) = \big(s_{ij}(A)\big)_{1 \leq i,j \leq k}$$

avec

$$s_{ij}(A) = \sqrt{\lambda_i \lambda_j}\ \phi_\infty(A_{ij})$$

où $\phi_\infty(A_{ij})$ désigne le module maximum des éléments du bloc A_{ij}, de type $(\lambda_i \lambda_j)$ (cf 4-2-3)

Remarques : - si k = 1 (un seul bloc, A) cela revient à utiliser la norme sous-multiplicative sur $\mathcal{M}_{nn}$:

$$A \longrightarrow s(A) = n\ \{\max_{1 \leq i,j \leq n} |a_{ij}|\} \in \mathbb{R}_+$$

- les normes utilisées étant absolues c'est en fait $\rho(|A|)$ que l'on majore. Or l'on sait que l'on a :

$$\forall A \in \mathcal{M}_{nn} \qquad \rho(A) \leq \rho(|A|)$$

5- 1- 3 - Exemples numériques

1) Soit A la matrice suivante, de rayon spectral égal à 10 (Exemple tiré de [12] II p. 226)

$$A = \begin{array}{|c|c|c|c|l}
5 & 4 & 1 & 1 & \\ \hline
4 & 5 & 1 & 1 & \alpha \\ \hline
1 & 1 & 4 & 2 & \beta \\ \hline
1 & 1 & 2 & 4 & \gamma \\
\multicolumn{4}{c}{\quad \alpha \quad \beta \quad \gamma} &
\end{array}$$

Les majorations classiques donnent :

$$10 = \rho(A) \lneq S_{\infty\infty}(A) = S_{11}(A) = 11$$

$$10 = \rho(A) \lneq s(A) = 4 \times 5 = 20$$

Avec les partionnements indiqués, il vient :

<u>Partionnement α</u>

$$M_\infty(A) = \begin{vmatrix} 5 & 6 \\ 4 & 7 \end{vmatrix} \qquad M_1(A) = \begin{vmatrix} 5 & 4 \\ 6 & 7 \end{vmatrix} \qquad S(A) = \begin{vmatrix} 5 & 4\sqrt{3} \\ 4\sqrt{3} & 15 \end{vmatrix}$$

de rayons spectraux respectifs

$$11 \qquad\qquad 11 \qquad\qquad 18 < 10 + \sqrt{73} < 19$$

<u>Partionnement β</u>

$$M_\infty(A) = \begin{vmatrix} 9 & 2 \\ 2 & 6 \end{vmatrix} \qquad M_1(A) = \begin{vmatrix} 9 & 2 \\ 2 & 6 \end{vmatrix} \qquad S(A) = \begin{vmatrix} 10 & 2 \\ 2 & 8 \end{vmatrix}$$

de rayons spectraux respectifs

$$10 \qquad\qquad 10 \qquad\qquad 11{,}2 < 9 + \sqrt{5} < 11{,}3$$

<u>Partionnement γ</u>

$$M_\infty(A) = \begin{vmatrix} 10 & 2 \\ 4 & 4 \end{vmatrix} \qquad M_1(A) = \begin{vmatrix} 10 & 4 \\ 2 & 4 \end{vmatrix} \qquad S(A) = \begin{vmatrix} 15 & 2\sqrt{3} \\ 2\sqrt{3} & 4 \end{vmatrix}$$

de rayonnements spectraux respectifs

$$11{,}1 < 7 + \sqrt{17} < 11{,}2 \qquad\qquad 7 + \sqrt{17} \qquad\qquad \frac{19 + \sqrt{99}}{2} < 14{,}5$$

On voit que les majorations obtenues peuvent être meilleures, sans que cela soit systématique, que les majorations classiques.

2) Soit la matrice (8,8) de rayon spectral égal à 4543, 696.....
([12] II p. 275)

$$A = \left|\begin{array}{ccccccc|c} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 2 & 3 & 4 & 5 & 6 & 7 & 8 \\ 1 & 3 & 6 & 10 & 15 & 21 & 28 & 36 \\ 1 & 4 & 10 & 20 & 35 & 56 & 84 & 120 \\ 1 & 5 & 15 & 35 & 70 & 126 & 210 & 330 \\ 1 & 6 & 21 & 56 & 126 & 252 & 462 & 792 \\ 1 & 7 & 28 & 84 & 210 & 462 & 924 & 1716 \\ \hline 1 & 8 & 36 & 120 & 330 & 792 & 1716 & 3432 \end{array}\right|$$

D'où les majorations classiques

$4543{,}696 = \rho(A) \leqslant S_{\infty\infty}(A) = S_{11}(A) = 6435$ (40% environ)

$4543{,}696 = \rho(A) \leqslant s(A) = 8 \times 3432 = 27456$ (500% environ)

Avec le partitionnement indiqué, il vient :

$$M_\infty(A) = \begin{vmatrix} 1716 & 1716 \\ 3003 & 3432 \end{vmatrix} \qquad M_1(A) = [M_\infty(A)]^t \qquad S(A) = \begin{vmatrix} 6468 & 1716.\sqrt{7} \\ 1716\sqrt{7} & 3432 \end{vmatrix}$$

de rayons spectraux respectifs égaux à

5000,.... 5000,.... 8544,....

d'où des majorations d'environ :

10% 10% 100%

3) On trouve dans le livre (en russe) de Kantorovitch [36] l'exemple suivant :

$$A = \left|\begin{array}{cc|cc} 1 & 1 & 1 & 1 \\ 1 & -1 & 1 & -1 \\ \hline 1 & 1 & -3 & 1 \\ 1 & -1 & 1 & 1 \end{array}\right| = \left|\begin{array}{c|c} A_{11} & A_{12} \\ \hline A_{21} & A_{22} \end{array}\right|$$

$$\rho(A) \leqslant \rho\big(M_2(A)\big)$$

avec

$$M_2(A) = \begin{vmatrix} \rho(A_{11}) & \rho(A_{12}) \\ \rho(A_{21}) & \rho(A_{22}) \end{vmatrix} = \begin{vmatrix} \sqrt{2} & \sqrt{2} \\ \sqrt{2} & 1+\sqrt{5} \end{vmatrix}$$

ce qui donne

$$3{,}78801 = \rho(A) \leqslant \rho\big(M_2(A)\big) = 4{,}0073$$

5 - 2 - MATRICES CONTRACTANTES (RELATIVEMENT à une NORME VECTORIELLE) et CONVERGENCE d'ITERATIONS LINEAIRES

Soit p une norme vectorielle régulière de taille k sur $\mathbb{K}^n$, et A une matrice (n,n) à éléments dans $\mathbb{K}$. Nous sommes amenés à poser la

5- 2- 1 - Définition

A est dite <u>contractante</u> (relativement à la norme vectorielle p) si elle admet, relativement à p, une majorante <u>convergente</u> (i.e de rayon spectral inférieur à l'unité).

Remarques : 1 Dire que A est contractante relativement à p c'est dire que

$$\rho\big(M(A)\big) < 1$$

En effet si la majorante M(A) est convergente, A est, par définition, contractante relativement à p. Réciproquement si A admet une majorante B convergente, on a, puisque M(A) est la plus petite des majorantes de A

$$M(A) \leq B$$

D'où la même inégalité sur les rayons spectraux de ces matrices non négatives :

$$\rho\big(M(A)\big) \leq \rho(B) < 1$$

2 Si la norme vectorielle p est une norme habituelle, (k = 1), la définition ci-dessus redonne bien la définition classique [38]. A est dite contractante relativement à la norme p s'il existe une constante α $(0<\alpha<1)$ telle que

$$\forall\, x \in \mathbb{K}^n \qquad p(Ax) \leq \alpha p(x)$$

ce qui s'écrit aussi

$$S_{pp}(A) \underset{df}{=} \sup_{\substack{x\in\mathbb{K}^n \\ x\neq 0}} \left\{ \frac{p(Ax)}{p(x)} \right\} < 1$$

3 Soit S une norme vectorielle régulière et sous-multiplicative sur $\mathcal{M}_{nn}$. Soit A une matrice (n,n) telle que

$$\rho\big(S(A)\big) < 1$$

Alors (4-2-4) il existe une norme vectorielle p, régulière sur $\mathbb{K}^n$, générant M sur $\mathcal{M}_{nn}$ telle que

$$M(A) \leq S(A)$$

d'où

$$\rho\big(M(A)\big) \leq \rho\big(S(A)\big) < 1$$

Ainsi, A est contractante relativement à p.

On a alors le

5-2- 2 - Lemme *Si* T *est contractante relativement à* p *, alors* $(I - T)^{-1}$ *existe et l'on a, pour toute majorante* B *convergente de* T :

$$M[(I - T)^{-1}] \leq [I - M(T)]^{-1} \leq (I - B)^{-1}$$

En effet

$\rho(T) \leq \rho\big(M(T)\big) \leq \rho(B) < 1$: ce qui montre que $(I - T)^{-1}$ existe.
On a alors, puisque B est convergente et non-négative :

$$(I - B)^{-1} = I + B + \ldots\ldots + B^r + \ldots \geq 0$$

Or

$$\forall x \in \mathbb{K}^n \quad [I - B]\, p(x) \leq p(x) - p(Tx) \leq p[(I - T)x]$$

D'où l'on tire, puisque $[I-B]^{-1}$ est non négative et $(I - T)$ inversible :

$$\forall\, y \in \mathbb{K}^n \qquad p\big((I - T)^{-1}y\big) \leq [I - B]^{-1}\, p(y)$$

$(I - B)^{-1}$ est donc une majorante de $(I - T)^{-1}$ relativement à p. Or (4-1-4) $M[(I - T)^{1}]$ est la plus petite des majorantes de $(I - T)^{-1}$ d'où

$$M[(I - T)^{-1}] \leq (I - B)^{-1}$$

en particulier

$$M[(I - T)^{-1}] \leq [I - M(T)]^{-1}$$

Et la dernière égalité à démontrer est conséquence évidente de

$$M(T) \leq B$$

Etablissons maintenant le

5- 2- 3 Théorème 3

Si T *est contractante relativement à* p, *elle est convergente : l'itération linéaire :*

$$x_{r+1} = Tx_r + h \qquad (h \in \mathbb{K}^n)$$

converge vers la solution unique $x = (I - T)^{-1}h$ *du système linéaire* $(I - T)x = h$. *De plus,* $r_\circ$ *étant un rang fixé à priori, on a, pour tout rang* $r \geq r_\circ$ *et pour toute majorante* B*(de* T*) convergente, les majorations d'erreur suivantes en norme vectorielle :*

$$p(\varepsilon_r) \underset{df}{=} p(x-x_r) \leq [M(T)]^{r-r_\circ}[I-M(T)]^{-1}p(x_{r_\circ+1}-x_{r_\circ}) \leq B^{r-r_\circ}[I-B]^{-1}p(x_{r_\circ+1}-x_{r_\circ})$$

Le premier point est trivial : par hypothèse T admet une majorante B convergente. D'où

$$\rho(T) \leq \rho(M(T)) \leq \rho(B)$$

T est convergente, $(I - T)^{-1}$ existe : l'itération envisagée converge vers la solution unique du système linéaire $(I - T)x = h$.

Posant $\qquad \varepsilon_r = x - x_r ; \qquad \rho_r = x_{r+1} - x_r$, il vient :

$$\varepsilon_{r+1} = T\varepsilon_r ; \quad \rho_{r+1} = T\rho_r ; \quad \rho_r = (I - T)\varepsilon_r$$

D'où

$$p(\varepsilon_r) = p(T\varepsilon_{r-1}) \leq Bp(\varepsilon_{r-1}) \leq B^2 p(\varepsilon_{r-2}) \leq \ldots\ldots \leq B^{r-r_\circ}p(\varepsilon_{r_\circ})$$

or $\qquad p(\varepsilon_{r_\circ}) = p[(I - T)^{-1}\rho_{r_\circ}] \leq M[(I - T)^{-1}]\, p(\rho_{r_\circ})$

Mais d'après le lemme précédent

$$M[(I - T)^{-1}] \leq [I - B]^{-1} \qquad \text{d'où enfin}$$

$$p(\varepsilon_r) \leq B^{r-r_\circ}[I - B]^{-1}p(x_{r_\circ+1} - x_{r_\circ})$$

en particulier,

$$p(\varepsilon_r) \leq M(T)^{r-r_\circ}[I - M(T)]^{-1}p(x_{r_\circ+1} - x_{r_\circ})$$

et la dernière inégalité à démontrer est conséquence évidente de

$$M(T) \leq B.$$

Remarques : 1 Ce théorème est l'exacte transcription en norme vectorielle d'un résultat classique lorsque p est une norme ordinaire (cf [9] [38]) : B est alors un nombre réel supposé tel que

$$S_{pp}(T) \underset{df}{=} \sup_{x\neq 0}\left\{\frac{p(Tx)}{p(x)}\right\} \leq B < 1$$

et l'estimation s'écrit

$$p(\varepsilon_r) \leq \frac{\{S_{pp}(T)\}^{r-r_\circ}}{1 - S_{pp}(T)}\, p(x_{r_\circ+1} - x_{r_\circ}) \leq \frac{B^{r-r_\circ}}{1-B}\, p(x_{r_\circ+1} - x_{r_\circ})$$

2 Les majorations d'erreur obtenues n'ont pas qu'un intérêt théorique : elles sont utilisables effectivement pour contrôler l'approximation obtenue sur la solution d'une itération linéaire (5-3; 7-4). A ce propos, l'estimation donnée par M(T) étant meilleure que celle que donne toute autre majorante de T, il pourrait sembler préférable d'utiliser systématiquement M(T). Mais, dans bien des cas pratiques (cf par exemple 7-4) M(T) est inaccessible numériquement alors que l'on sait facilement calculer une autre majorante convergente. C'est la raison pour laquelle nous avons donné ici ces doubles majorations, l'une avec M(T), l'autre avec B.

5- 2- 4 - Exemples

1) Soit la matrice (7,7) :

$$T = \frac{1}{10}\left|\begin{array}{cc|ccc|cc} 0 & 0 & -3 & -1 & 2 & -3 & 2 \\ 0 & 0 & 1,5 & 2 & -2 & 2,5 & -2 \\ \hline 1 & 1 & 0 & 0 & 0 & -4 & 0 \\ 0,5 & -1 & 0 & 0 & 0 & 2 & -1 \\ 1,5 & 0,5 & 0 & 0 & 0 & 4 & 0 \\ \hline 3 & -1 & 1 & -1,5 & 0,5 & 0 & 0 \\ -1 & 2 & 0,5 & 0,5 & -1 & 0 & 0 \end{array}\right|$$

(matrice d'une itération de Jacobi par blocs)

Avec les notations introduites en (5-1) il vient

$$S_{\infty\infty}(T) = 1,1 > 1$$

$$S_{11}(T) = 1,55 > 1$$

$$s(T) = 7 \times 0,4 = 2,8 > 1$$

Aucune de ces majorations ne permet de conclure à la convergence de T. Cependant en utilisant le partionnement indiqué, il vient

$$M_{\infty}(T) = \begin{vmatrix} 0 & 0.6 & 0.5 \\ 0.2 & 0 & 0.4 \\ 0.4 & 0.3 & 0 \end{vmatrix}$$

et $\rho\left(M_{\infty}(T)\right) \leq S_{11}\left(M_{\infty}(T)\right) = 0,9 < 1$

ce qui prouve que T est contractante (donc convergente).

Remarque : Dans ce cas S(T) est égale à

$$S(T) = \frac{1}{10} \begin{vmatrix} 0 & 3\sqrt{6} & 6 \\ 1.5\sqrt{6} & 0 & 4\sqrt{6} \\ 6 & 1.5\sqrt{6} & 0 \end{vmatrix}$$

et ni $S_{\infty\infty}\big(S(T)\big)$ ni $S_{11}\big(S(T)\big)$ ne permettent de conclure

2) Pour la matrice T' "voisine" de T :

$$T' = \frac{1}{10} \left|\begin{array}{cc|ccc|cc} 0 & 0 & -2 & -2 & 2 & -2.5 & 2.5 \\ 0 & 0 & 2 & 2 & -2 & 2.5 & -2.5 \\ \hline 1 & 1 & 0 & 0 & 0 & -3 & 3 \\ -1 & 1 & 0 & 0 & 0 & 3 & -3 \\ 1 & -1 & 0 & 0 & 0 & 3 & 3 \\ \hline 2 & -2 & 1 & -1 & 1 & 0 & 0 \\ 2 & 2 & 1 & 1 & -1 & 0 & 0 \end{array}\right|$$

On a

$$S_{\infty\infty}(T') = 1.1 > 1$$

$$S_{11}(T') = 1.4 > 1$$

$$s(T') = 7 \times 0.3 = 2.1 > 1$$

$$M_\infty(T') = \begin{vmatrix} 0 & 0.6 & 0.5 \\ 0.2 & 0 & 0.6 \\ 0.4 & 0.3 & 0 \end{vmatrix} \qquad M_1(T') = \begin{vmatrix} 0 & 0.4 & 0.5 \\ 0.3 & 0 & 0.9 \\ 0.4 & 0.2 & 0 \end{vmatrix}$$

Et aucune des techniques précédentes ne permet de conclure. Or

$$S(T') = \begin{vmatrix} 0 & 0.2\sqrt{6} & 0.5 \\ 0.1\sqrt{6} & 0 & 0.3\sqrt{6} \\ 0.4 & 0.1\sqrt{6} & 0 \end{vmatrix}$$

et

$$\rho\big(S(T')\big) \leqslant S_{\infty\infty}\big(S(T')\big) = \frac{1}{10}\,\mathrm{Max}\,\{2\sqrt{6} + 5,\ 4\sqrt{6},\ 4 + \sqrt{6}\}$$

$$= 0.2\sqrt{6} + 0.5 < 1$$

Ce qui prouve (cf 5-1-2 et 5-2-1, remarque 3) que T' est contractante relativement à une certaine norme vectorielle p, d'ailleurs assez laborieuse à construire (on peut prendre $p(x) = S(xa^{H})b$ $(a \in \mathbb{K}^n - \{0\}, b \in \mathbb{R}^k_+$; cf 4-2-4)

En tous cas on a prouvé ainsi la convergence de T'.

5 - 3 - L'ALGORITHME de SCHROEDER pour une MATRICE d'ITERATION CONTRACTANTE

J. Schroeder semble être le premier à avoir effectivement utilisé, pour une itération linéaire, des estimations <u>en norme vectorielle</u> de l'approximation obtenue sur la solution [58]. Il utilise la norme vectorielle type $p(x) = |x|$. Aussi proposons-nous ici de faire le lien entre l'algorithme de Schroeder et les estimations d'erreur données au théorème 3 (5-2-3)

5- 3- 1 - L'algorithme de Schroeder

Il donne, à chaque pas d'une itération linéaire, une majoration, composante à composante, de l'approximation obtenue sur la solution : Soit l'itération linéaire dans $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$)

$$x_{r+1} = Tx_r + h \qquad (1)$$

T matrice (n,n) à éléments dans $\mathbb{K}$

h vecteur fixé dans $\mathbb{K}^n$

La condition nécessaire et suffisante de convergence est évidemment $\rho(T) < 1$. Sous cette hypothèse, la suite des x_r converge vers x tel que

$$x = Tx + h$$

On rappelle que l'on pose :

$\varepsilon_r = x - x_r$ (vecteur, <u>inconnu</u>, d'erreur)

$\rho_r = x_{r+1} - x_r$ (vecteur, <u>connu</u>, résidu)

et l'on a

$$\varepsilon_{r+1} = T\varepsilon_r$$

$$\rho_{r+1} = T\rho_r$$

$$\rho_r = (I-T)\varepsilon_r$$

Donnons-nous une matrice B, non négative, de type (n,n) telle que :

$$B \geqslant |T|$$

où $|T|$ désigne la matrice (n,n) ayant pour éléments les modules des éléments correspondants de la matrice T.

On suppose $\rho(B)$ inférieur à l'unité, ce qui entraîne d'ailleurs la convergence de (1)

$$\rho(T) \leqslant \rho(|T|) \leqslant \rho(B) < 1$$

Parallèlement à l'itération (1), dite <u>itération principale</u>, on envisage, dans $\mathbb{R}^n_+$ l'itération suivante, que l'on appelle <u>itération préliminaire</u> :

$$\left.\begin{array}{l} \omega_o = 0 \\ \omega_{r+1} = B\omega_r + |x_{r+1} - x_r| = B\omega_r + |\rho_r| \geqslant 0 \end{array}\right\} \quad (2)^*$$

Sous l'hypothèse que B est irréductible et non-cyclique, J. Schroeder [58] montre qu'il existe un rang r_o tel que

$$\omega_{r_o+1} \leqslant \omega_{r_o} \qquad (3)$$

Il pose alors

$$z_{r_o} = \omega_{r_o}$$

$$z_{r+1} = Bz_r \qquad (r = r_o,\ r_o+1 \ldots) \qquad (4)$$

et l'on a la <u>majoration d'erreur</u> (cf [58]) :

$$\boxed{|\varepsilon_r| \leqslant z_r = B^{r-r_o} z_{r_o} \quad (r = r_o,\ r_o+1, \ldots)} \qquad (5)$$

ce qui était annoncé. L'itération (4) sera appelée <u>itération d'erreur</u>.

5- 3- 2 - Commentaires

a) Considérons, sur $\mathbb{K}^n$, la norme vectorielle type définie par $p(x) = |x|$. Elle génère sur $\mathcal{M}_{nn}$ la norme vectorielle $M(U) = |U|$ (cf 4-1-4, exemple a)

---

* là encore $|\rho_r|$ désigne le vecteur ayant pour composantes les modules des composantes correspondantes de ρ_r

Il est clair qu'avec la terminologie introduite en 5-2, l'hypothèse d'existence d'une matrice B telle que

$$|T| \leqslant B \qquad \rho(B) < 1$$

s'énonce ainsi "T est contractante relativement à la norme vectorielle $p(x) = |x|$".

b) l'itération préliminaire (2) est en fait destinée à produire un rang r_o et un vecteur z_{r_o} de $\mathbb{R}^n_+$ tel que

$$|\varepsilon_{r_o}| \leqslant z_{r_o} \tag{6}$$

En effet, (3) s'écrit

$$Bz_{r_o} + |\rho_{r_o}| \leqslant z_{r_o} \qquad \text{soit}$$

$$(I - B)\, z_{r_o} \geqslant |\rho_{r_o}| \tag{7}$$

Or d'après le lemme de (5-2-2) appliqué à la norme vectorielle $M(U)=|U|$ utilisée ici,

$$(I - B)^{-1} \geqslant |(I - T)^{-1}|$$

on tire donc de (7) .

$$z_{r_o} \geqslant (I - B)^{-1}|\rho_{r_o}| \geqslant |(I - T)^{-1}||\rho_{r_o}| \geqslant |(I - T)^{-1}\rho_{r_o}| = |\varepsilon_{r_o}|$$

c) Disposant, au rang r_o, de la majoration d'erreur (6), l'itération d'erreur (4) permet d'obtenir une majoration d'erreur aux rangs suivants. Cela se montre très simplement par récurrence. En effet, supposons que

$$|\varepsilon_r| \leqslant z_r$$

On a alors

$$|\varepsilon_{r+1}| = |T\varepsilon_r| \leqslant B|\varepsilon_r| \leqslant Bz_r = z_{r+1}$$

5- 3- 3 - Extension de l'algorithme

Soit p une norme vectorielle régulière de taille k sur $\mathbb{K}^n$, générant sur $\mathcal{M}_{nn}$ la norme vectorielle M, régulière, sous-multiplicative, de taille k^2. L'algorithme de Schroeder s'étend de façon très naturelle :

Soit l'itération (1), dite itération principale. Supposons T contractante relativement à p : elle admet une majorante B convergente.

Alors (Théorème 3) (1) est convergente; supposons connaître un rang r_o et un vecteur z_{r_o} de $\mathbb{R}^k_+$ tel que

$$p(\varepsilon_{r_o}) \leqslant z_{r_o} \qquad (8)$$

Alors l'itération d'erreur

$$z_{r+1} = Bz_r \qquad (r = r_o,\ r_o+1, ..)(9)$$

permet d'écrire

$$\boxed{p(\varepsilon_r) \leqslant z_r = B^{r-r_o}\, z_{r_o} \qquad (r = r_o,\ r_o+1...)(10)}$$

En effet, cette majoration de la norme vectorielle $p(\varepsilon_r)$ du vecteur inconnu d'erreur ε_r est vraie au rang r_o (8). Elle se démontre alors par récurrence pour les rangs ultérieurs : si l'on a

$$p(\varepsilon_r) \leqslant z_r$$

il vient

$$p(\varepsilon_{r+1}) = p(T\varepsilon_r) \leqslant Bp(\varepsilon_r) \leqslant Bz_r = z_{r+1}$$

Il est clair que le choix de $p(x) = |x|$, d'où $M(T) = |T|$ redonne la méthode de Schroeder standard. Pour obtenir le r_o et le z_{r_o} vérifiant (8) on peut envisager l'itération préliminaire (dans $\mathbb{R}^k_+$)

$$\omega_o = 0$$
$$\omega_{r+1} = B\omega_r + p(\rho_r) \qquad (11)$$

S'il existe (*) un rang r_o tel que

$$\omega_{r_o+1} \leqslant \omega_{r_o} \qquad (12)$$

on pose

$$z_{r_o} = \omega_{r_o}$$

et (8) est alors vérifiée. Cela se démontre comme précédemment, l'inégalité (12) s'écrivant

$$(I - B)\omega_{r_o} \geqslant p(\rho_{r_o})$$

(*) nous n'avons pas cherché à formuler des conditions suffisantes d'existence d'un tel r_o : en effet on verra au prochain paragraphe une manière de procéder plus avantageuse que l'itération préliminaire. Néanmoins, dans la plupart des expériences numériques effectuées un tel r_o a été atteint (cf 7-4)

D'où

$$\omega_{r_o} \geqslant (I - B)^{-1} \, p(\rho_{r_o}) \qquad (13)$$

D'où enfin, par l'utilisation du Lemme 5-2-2

$$\omega_{r_o} \geqslant M[(I-T)^{-1}] \, p(\rho_{r_o}) \geqslant p[(I-T)^{-1}\rho_{r_o}] = p(\varepsilon_{r_o})$$

5- 3- 4 - Commentaires

Le choix de la norme vectorielle type, dite de Schroeder ($p(x) = |x|$) conduit à une itération d'erreur de même taille que l'itération principale, ce qui revient à doubler le volume des calculs. On obtient alors, à chaque pas de l'itération, une majoration de l'erreur sur chaque composante du vecteur itéré, ce qui, dans des cas d'utilisation effective (gros systèmes linéaires, provenant de la discrétisation de problèmes différentiels, et conduisant à des itérations lentement convergentes) donne un volume considérable de résultats partiels très semblables.

Par contre, le choix d'une norme vectorielle de taille k bien inférieure à n, taille de l'itération principale (exemple courant : n = 2500, k =50) permet de doser l'importance relative du calcul d'erreur. On obtient alors des majorations d'erreur globales, par blocs de composantes du vecteur itéré. On peut, de plus, se permettre alors de fixer <u>à priori</u> le rang r_o à partir duquel sont obtenues les majorations d'erreur et de calculer un z_{r_o} correspondant (I.E. vérifiant (8)) par résolution directe d'un système (k,k). Il suffit en effet de résoudre :

$$(I - B)z_{r_o} = p(\rho_{r_o}) \qquad (14)$$

pour obtenir

$$z_{r_o} = (I - B)^{-1}p(\rho_{r_o}) \geqslant M[(I - T)^{-1}]p(\rho_{r_o}) \geqslant p(\varepsilon_{r_o})$$

Remarque : r_o étant supposé obtenu par une itération préliminaire (11), soit ω_{r_o} la majoration correspondante de $p(\varepsilon_{r_o})$ et soit $z_{r_o} = (I - B)^{-1}p(\rho_{r_o})$. (13) s'écrit

$$z_{r_o} \leqslant \omega_{r_o}$$

D'où

$$B^{r-r_o} z_{r_o} \leqslant B^{r-r_o} \omega_{r_o}.$$

Autrement dit :

Règle : *Il est toujours préférable de fixer à priori* r_o *et de calculer* z_{r_o} *par (14). On obtient alors la majoration*

$$p(\varepsilon_r) \leqslant B^{r-r_o} (I - B)^{-1} p(x_{r_o+1} - x_{r_o})$$

qui est celle donnée au Théorème 3 (5-2-3)

5 - 3 - 5 - Mise en oeuvre pratique

A cause de l'inégalité

$$\rho(T) \leqslant \rho\big(M(T)\big) \leqslant \rho(B)$$

le rayon spectral de B est généralement supérieur à celui de T : l'itération converge plus lentement que l'itération principale, d'où une certaine "dérive" des majorations : avec l'accroissement du rang de l'itération, les majorations obtenues deviennent relativement de plus en plus lâches (voir les exemples numériques du paragraphe 7-4). Il est numériquement préférable de redémarrer régulièrement une nouvelle itération d'erreur par résolution d'un système linéaire (14); cela permet de suivre beaucoup plus finement la décroissance de l'erreur qu'en poursuivant la même itération d'erreur. Le fait de pouvoir obtenir des majorations d'erreur à tout rang fixé à priori de l'itération principale, rend particulièrement simple cette technique de réenclanchement périodique d'une nouvelle itération d'erreur.

C'est d'ailleurs ainsi que l'on opère dans la procédure ALGOL donnée ci-après.

5- 3- 6 - Procédure ALGOL

```
PROCEDURE SCHCONTRITER(T,X,H,N,K,LAMBDA,EPS,
   NT,TILT,BOND,RES,NOMBTOURMAX,SURPASS) ;
TABLEAU T,X,H,RES ; ENTIER TABLEAU LAMBDA ; REEL EPS ;
ENTIER N,K,NT,TILT,BOND,NOMBTOURMAX ; ETIQUETTE SURPASS ;

DEBUT TABLEAU B,D[1:K,1:K] ; TABLEAU Y,Z[1:N] ;
   ENTIER I,J,I1,I2,J1,J2 ;
   REEL STOC ; BOOLEEN BOUL ;

   PROCEDURE MATVEC(T,X,N) ; TABLEAU T,X ;
   ENTIER N ;
   DEBUT POUR I1:=1 PAS 1 JUSQUA N FAIRE Y[I1]:=X[I1] ;
      POUR I1:=1 PAS 1 JUSQUA N FAIRE
      DEBUT STOC:=0.0 ;
         POUR I2:=1 PAS 1 JUSQUA N FAIRE
         STOC:=STOC+T[I1,I2]*Y[I2] ;
         X[I1]:=STOC
      FIN
   FIN PROCEDURE MATVEC ;

   PROCEDURE SOMVEC(X,Y,N) ; TABLEAU X,Y ; ENTIER N ;
   POUR I1:=1 PAS 1 JUSQUA N FAIRE X[I1]:=X[I1]+Y[I1] ;

PROCEDURE MATNORMVEC(A,M,N,K,LAMBDA) ;
TABLEAU A,M ; ENTIER TABLEAU LAMBDA ; ENTIER N,K ;
   DEBUT ENTIER I,J,I1,J1,I2,J2 ; REEL S,NOR ;

      REEL PROCEDURE NORME(W,P,Q) ;
      TABLEAU W ; ENTIER P,Q ;
      DEBUT NOR:=0.0 ;
         POUR I2:=1 PAS 1 JUSQUA P FAIRE
         DEBUT S:=0.0 ;
            POUR J2:=1 PAS 1 JUSQUA Q FAIRE
            S:=S+ABS(W[I2,J2]) ;
            SI NOR < S ALORS NOR:=S
         FIN ;
         NORME:=NOR
      FIN PROCEDURE NORME ;

      POUR I:=1 PAS 1 JUSQUA K FAIRE
      POUR J:=1 PAS 1 JUSQUA K FAIRE
      DEBUT TABLEAU C[1:LAMBDA[I],1:LAMBDA[J]] ;
      I1:=J1:=0 ;
      POUR I2:=1 PAS 1 JUSQUA I-1 FAIRE
      I1:=I1+LAMBDA[I2] ;
      POUR J2:=1 PAS 1 JUSQUA J-1 FAIRE
      J1:=J1+LAMBDA[J2] ;
      POUR I2:=1 PAS 1 JUSQUA LAMBDA[I] FAIRE
      POUR J2:=1 PAS 1 JUSQUA LAMBDA[J] FAIRE
      C[I2,J2]:=A[I1+I2,J1+J2] ;
      M[I,J]:=NORME(C,LAMBDA[I],LAMBDA[J])
   FIN
FIN PROCEDURE MATNORMVEC ;
```

```
PROCEDURE GRESOLPIV(A,B,X,N,IMPOSSIBLE) ; VALEUR N ;
   TABLEAU A,B,X ; ENTIER N ; ETIQUETTE IMPOSSIBLE ;
COMMENTAIRE RESOLUTION DU SYSTEME LINEAIRE AX=B, D'ORDRE N, PAR LA
   METODE DE GAUSS. DANS LA PHASE TRIANGULARISATION ON UTILISE LA
   REGLE DUPIVOT MAXIMUM.
   A ET B SONT CONSERVES. SIA EST SINGULIERE, SORTIE PAR IMPOSSIBLE ;
   CHARGEMENT: DEBUT TABLEAU A1[1:N,1:N],B1[1:N] ;
      ENTIER I,J,K ; REEL R,TX ;
               POUR I:=1 PAS 1 JUSQUA N FAIRE
            DEBUT
               B1[I]:=B[I] ;
               POUR J:=1 PAS 1 JUSQUA N FAIRE
               A1[I,J]:=A[I,J] ;
            FIN ;
   TRIANGULARISATION: POUR K:=1 PAS 1 JUSQUA N-1 FAIRE
      DEBUT ENTIER M ; REEL A,B ;
               M:=K ; A:=A1[K,K] ;
   PIVOTMAX: POUR I:=K+1 PAS 1 JUSQUA N FAIRE
         DEBUT
               B:=A1[I,K] ;
               SI ABS(A) < ABS(B) ALORS
            DEBUT
               A:=B ; M:=I ;
            FIN ;
         FIN ;
               SI ABS(A1[M,K])=0 ALORS ALLERA  IMPOSSIBLE ;
   ECH DE LIGNES: POUR J:=K PAS 1 JUSQUA N FAIRE
            DEBUT
               R:=A1[K,J] ; A1[K,J]:=A1[M,J] ; A1[M,J]:=R ;
            FIN ;
               R:=B1[K] ; B1[K]:=B1[M] ; B1[M]:=R ;
   CALCUL: POUR I:=K+1 PAS 1 JUSQUA N FAIRE
            DEBUT
               SI ABS(A1[I,K])=0 ALORS ALLERA  END ;
               R:=A1[I,K]/A1[K,K] ;
               POUR J:=K+1 PAS 1 JUSQUA N FAIRE
               A1[I,J]:=A1[I,J]-R*A1[K,J] ;
               B1[I]:=B1[I]-R*B1[K] ;
      END: FIN ;
      FIN TRIANGULARISATION ;
   RESOLSYSTRI: SI ABS(A1[N,N])=0 ALORS ALLERA  IMPOSSIBLE ;
               POUR I:=N PAS -1 JUSQUA 1 FAIRE
            DEBUT TX:=0 ;
               POUR J:=N PAS -1 JUSQUA I+1 FAIRE
               TX:=TX-X[J]*A1[I,J] ;
               X[I]:=(B1[I]+TX)/A1[I,I] ;
            FIN RESOLSYSTRI ;
      FIN GRESOLPIV ;

   PROCEDURE NORMVEC(X,N,K,LAMBDA,P) ;
   TABLEAU X,P ; ENTIER TABLEAU LAMBDA ; ENTIER N,K ;
   DEBUT J1:=0 ;
      POUR I1:=1 PAS 1 JUSQUA K FAIRE
      DEBUT STOC:=0.0 ;
         POUR I2:=1 PAS 1 JUSQUA LAMBDA[I1] FAIRE
```

```
        DEBUT J1:=J1+1 ;
        SI STOC < ABS(X[J1]) ALORS STOC:=ABS(X[J1])
        FIN ;
        P[I1]:=STOC
     FIN
  FIN PROCEDURE NORMVEC ;

  POUR NT:=1 PAS 1 JUSQUA TILT FAIRE
  DEBUT MATVEC(T,X,N) ;
     SOMVEC(X,H,N)
  FIN ;
  MATNORMVEC(T,B,N,K,LAMBDA) ;
  POUR I:=1 PAS 1 JUSQUA K FAIRE
  DEBUT D[I,I]:=1.0-B[I,I] ;
     POUR J:=1 PAS 1 JUSQUA I-1,I+1 PAS 1 JUSQUA K FAIRE
     D[I,J]:=-B[I,J]
  FIN ;
   NT:= TILT ;
  RETI: NT:=NT+1 ;
  POUR I:=1 PAS 1 JUSQUA N FAIRE Z[I]:=-X[I] ;
  MATVEC(T,X,N) ;
  SOMVEC(X,H,N) ;
  SOMVEC(Z,X,N) ;
  NORMVEC(Z,N,K,LAMBDA,RES) ;
  GRESOLPIV(D,RES,RES,K,SCHTROUMPF) ;
  SCHTROUMPF: MATVEC(B,RES,K) ;
ITER: NT:=NT+1 ;
  SI NT > NOMBTOURMAX ALORS ALLERA  SURPASS ;
  MATVEC(T,X,N) ;
  SOMVEC(X,H,N) ;
  MATVEC(B,RES,K) ;
  BOUL:= VRAI ; I:=0 ;
  POUR I:=I+1 TANTQUE  BOUL ET (I ≤ K) FAIRE
      BOUL:= BOUL ET (RES[I] < EPS)  ;
  SI BOUL ALORS ALLERA  EXIT ;
  SI (NT-TILT) ./. BOND*BOND ≠ NT-TILT ALORS ALLERA
  ITER SINON ALLERA  RETI ;
  EXIT: FIN  PROCEDURE SCHCONTRITER ;
```

<u>Commentaires</u> tableaux et variables à charger avant l'activation de la procédure :

les entiers N, K, NØMBTØURMAX, BØND, TILT

les tableaux T[1:N, 1:N] X[1:N] H[1:N]

l'entier tableau LAMBDA [1:K]

le réel EPS

après l'exécution le vecteur solution est porté par le tableau X[1:N]
les estimations de l'approximation obtenue sont en RES[1:K]

<u>Activation</u> : On suppose chargés la matrice T de l'itération principale, le vecteur de départ (porté par X) et le vecteur de second membre (porté par H). L'itération principale répond alors à la (pseudo) instruction Algol :

$$X := TX + H$$

K est la taille de la norme vectorielle p_∞ (qui génère M_∞ cf(5-11)) utilisée. L'entier tableau LAMBDA définit le partionnement utilisé : ses composantes portent les dimensions des sous-espaces W_i sur lesquels p est construite. Dans l'exemple suivant

$$x = \left|\begin{array}{c} \S_1 \\ \S_2 \\ \hline \S_3 \\ \S_4 \\ \S_5 \\ \hline \S_6 \\ \S_7 \end{array}\right. \longrightarrow p_\infty(x) = \left|\begin{array}{l} \mathrm{Max}\{|\S_1|, |\S_2|\} \\ \mathrm{Max}\{|\S_3|, |\S_4|, |\S_5|\} \\ \mathrm{Max}\{|\S_6|, |\S_7|\} \end{array}\right.$$

on a donc N = 7 ; K = 3 et LAMBDA est le tableau [2,3,2]

L'activation de la procédure SCHCØNTRITER a les effets suivants :

1) exécution de TILT-1 pas de l'itération principale.
2) au rang TILT de l'itération principale est calculée la première estimation d'erreur

$$z_{TILT} = [I - M_\infty(T)]^{-1}\ p_\infty(x_{TILT+1} - x_{TILT})$$

par résolution, grâce à GRESOLPIV, d'un système linéaire.

3) l'itération principale et l'itération d'erreur se déroulent parallèlement jusqu'au rang TILT + BØND - 1

4) au rang TILT + BOND de l'itération principale, s'enclanche une nouvelle itération d'erreur par résolution directe de

$$z_{\text{TILT+BØND}} = [I - M_\infty(T)]^{-1}\ p_\infty(x_{\text{TILT+BØND+1}} - x_{\text{TILT+BØND}})$$

5) le processus se poursuit : déroulelement parallèle de l'itération principale et de l'itération d'erreur avec réenclanchement régulier d'une nouvelle itération d'erreur tous les BØND pas de l'itération principale. On peut sortir de la procédure de deux façons différentes :

α) au rang NØMBTØURMAX de l'itération principale, les composantes de l'estimation d'erreur obtenues ne sont pas toutes inférieures à EPS. On sort alors de la procédure par déroutement sur l'étiquette SURPASS.

β) à un rang inférieur ou égal à NØMBTØURMAX, les composantes de l'estimation obtenue sont toutes inférieures à EPS. C'est la sortie normale de la procédure.

Dans les deux cas, NT porte le numéro de la dernière itération effectuée et X porte les N composantes du dernier vecteur itéré obtenu.

Exemple   Soit T la matrice (7,7)

$$T = \frac{1}{10}\left|\begin{array}{cc|ccc|cc}
0 & 0 & -3 & -1 & 2 & -3 & 2 \\
0 & 0 & 1.5 & 2 & -2 & 2.5 & -2 \\
\hline
1 & 1 & 0 & 0 & 0 & -4 & 0 \\
0.5 & -1 & 0 & 0 & 0 & 2 & -1 \\
1.5 & 0.5 & 0 & 0 & 0 & 4 & 0 \\
\hline
3 & -1 & 1 & -1.5 & 0.5 & 0 & 0 \\
-1 & 2 & 0.5 & 0.5 & -1 & 0 & 0
\end{array}\right|$$

Il a été prouvé en 5-2-4 qu'avec le partionnement indiqué, $\rho\left(M_\infty(T)\right) < 1$: T est contractante relativement à p . D'où l'itération linéaire :

$$x_{r+1} = Tx_r + h$$

EPS = 0.0001

Itération n°	vecteur itéré	Erreur	Norme vectorielle de l'erreur	Majoration obtenue
	1.3 0.8	0.3 -0.2	0.3	6.2442393
1	1.2 0.94 0.4	0.2 -0.06 -0.6	0.6	4.3271888
	0.8 0.9	-0.2 -0.1	0.2	4.6958523
	0.9874689 1.0094609	-0.0125311 0.0094609	0.0125311	0.0965634
6	0.9919601 1.0055149 1.0079033	-0.0080398 0.0055149 0.0079033	0.0080398	0.0706371
	1.0028827 0.9996058	0.0028827 -0.0003941	0.0028827	0.0762366
	1.0004715 0.9996650	0.0004715 -0.0003349	0.0004715	0.0037770
11	1.0002398 0.9997992 0.9997298	0.0002398 -0.0002007 -0.0002701	0.0002701	0.0026771
	0.9997902 1.0000517	-0.0002097 0.0000517	0.0002097	0.0026956
	0.9999835 1.0000110	-0.0000165 0.0000110	0.0000165	0.0001374
16	0.9999937 1.0000069 1.0000085	-0.0000062 0.0000069 0.0000085	0.0000085	0.0000937
	1.0000111 0.9999966	0.0000111 -0.0000033	0.0000111	0.0000879
	0.9999988 1.0000022	-0.0000011 0.0000022	0.0000022	0.0000791
17	0.9999949 1.0000006 1.0000025	-0.0000050 0.0000006 0.0000025	0.0000050	0.0000532
	0.9999926 1.0000030	-0.0000073 0.0000030	0.0000073	0.0000588

où $h = (1.3, 0.8, 1.2, 0.95, 0.4, 0.8, 0.9)$
de solution $x = (1,1,1,1,1,1,1)$

La norme vectorielle utilisée est celle donnée à titre d'exemple dans les commentaires ci-dessus. On a donc N = 7, K = 3 et LAMBDA est le tableau (2,3,2)

le vecteur de départ de l'itération principal a été pris nul

TILT a été pris égal à 1

BØND a été pris égal à 5

NØMBTØURMAX a été pris égal à 50

EPS a été pris égal à 0.0001

On a obtenu les résultats suivants (pour cet exemple, les résultats ont été sortis tous les TILT + BØND rangs de l'itération principale)

A l'itération n°17, toutes les composantes de l'estimation d'erreur sont inférieures à EPS = 0.0001 ce qui n'est pas le cas à l'itération n° 16

5-4 - MODE de CONVERGENCE d'une ITERATION LINEAIRE

Soit l'itération linéaire suivante dans $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$) que nous supposons convergente $(\rho(T) < 1)$:

$$x_{r+1} = Tx_r + h \qquad (1)$$

soit $x = Tx + h$ la limite de cette itération;

On rappelle que l'on pose

$$\varepsilon_r = x - x_r, \qquad \text{d'où}$$

$$\varepsilon_r = T\varepsilon_{r-1} = T^2\varepsilon_{r-2} = \ldots\ldots T^r\varepsilon_o$$

où ε_o est le vecteur fixe $x - x_o$. Ainsi la façon dont le vecteur d'erreur tend vers zéro est donnée par le comportement de T^r lorsque r tend vers l'infini.

En (4-5) nous avons rappelé la formule établie par Ostrowski [42]

$$T^r = C_r^{q-1}[\rho(T)]^{r-q+1}\Big[\sum_{i=1}^{s} a_i b_i^t + S_r\Big]$$

Nous reprenons les mêmes notations qu'en 4-4; en particulier S_r est une matrice qui tend vers zéro lorsque r tend vers l'infini. Il vient alors

$$\varepsilon_r = T^r\varepsilon_o = C_r^{q-1}[\rho(T)]^{r-q+1}\Big[\sum_{i=1}^{s} a_i \underbrace{b_i^t\varepsilon_o}_{\beta_i} + \underbrace{S_r \cdot \varepsilon_o}_{\gamma_r}\Big]$$

Soit

$$\boxed{\varepsilon_r = C_r^{q-1}[\rho(T)]^{r-q+1}\Big[\sum_{i=1}^{s} \beta_i a_i + \gamma_r\Big]}$$

$$\gamma_r \longrightarrow 0 \text{ quand } r \longrightarrow \infty$$

Puisque $\rho(T) < 1$, la quantité $C_r^{q-1}[\rho(T)]^{r-q+1}$ tend vers zéro lorsque r tend vers l'infini. Il en résulte que ε_r <u>s'annule en s'alignant sur le vecteur</u> $\varepsilon = \sum_{i=1}^{s} \beta_i a_i$, vecteur fixe, combinaison linéaire des vecteurs propres a_i de T; ε dépend donc de ε_o, donc du vecteur de départ x_o par les $\beta_i = b_i^t\varepsilon_o$

Il est donc numériquement inaccessible puisque ε_0 est inconnu, sauf dans le cas où s = 1 : c'est le cas "numérique" où il n'y a qu'un seul bloc de rayon spectral $\rho(T)$ et de taille maximum q(cf 4-5).

On a alors

$$T^r = C_r^{q-1}[\rho(T)]^{r-q+1} [ab^t + S_r]$$

d'où

$$\varepsilon_r = T^r\varepsilon_0 = C_r^{q-1}[\rho(T)]^{r-q+1}[(b^t\varepsilon_0)a + \gamma_r]$$

$$\gamma_r \longrightarrow 0$$
$$r \longrightarrow \infty$$

Dans ce cas ε_r s'annule en s'alignant (*) sur le vecteur propre a de T (relatif à la valeur propre de plus grand module $\rho(T)$). Il est donc indépendant de ε_0, donc de x_0, et peut être obtenu numériquement en direction par une "méthode de la puissance". En effet soit y_0 un vecteur quelconque, non nul, de $\mathbb{K}^n$; le rapport

$$\frac{T^r y_0}{\|T^r y_0\|} = \frac{ab^t y_0 + S_r y_0}{\|ab^t y_0 + S_r y_0\|}$$

tend, puisque S_r tend vers zéro, vers le vecteur

$$\frac{(b^t y_0)\ a}{|(b^t y_0)|\|a\|} = \frac{\varepsilon\ a}{\|a\|} \qquad \text{où } \varepsilon = \pm 1$$

On peut donc obtenir ainsi les cosinus directeurs de a. Mais utilisons plutôt une norme vectorielle p sur $\mathbb{K}^n$. On a alors

$$p(\varepsilon_r) = C_r^{q-1}[\rho(T)]^{r-q+1}\ p[(b^t\varepsilon_0)a + \gamma_r)]$$

Ainsi $p(\varepsilon_r)$s'annule en s'alignant sur p(a) (si du moins $b^t\varepsilon_0 \neq 0$) c'est-à-dire sur

$$\lim_{r\to\infty}\left\{\frac{p(T^r y_0)}{\|T^r y_0\|}\right\}$$

(*) si du moins $b^t\varepsilon_0 \neq 0$ ce qu'on peut raisonnablement supposer réalisé.

Remarque : si p est régulière, elle génère la norme vectorielle M sur $\mathcal{M}_{nn}$; on a alors

$$M(T^r) = C_r^{q-1}[\rho(T)]^{r-q+1} \quad M(ab^t + S_r)$$

ainsi $M(T^r)$ s'annule en devenant proportionnelle à

$M(ab^t) = p(a)\ [p^*(b)]^t$ (cf 4-1-5) ; $M(ab^t)$ a donc toutes ses colonnes proportionnelles à p(a).

Exemples numériques

1

$$T = \left|\begin{array}{cc|ccc} 0.9 & & & & \\ & 0.7 & & & \\ \hline & & 0.5 & & \\ & & & 0.3 & \\ & & & & 0.1 \end{array}\right|$$

avec le partitionnement indiqué, on a utilisé la norme vectorielle M_∞ (cf 5-1-1).

à l'itération 175, on a obtenu, en machine

$$\frac{M_\infty(T^{175})}{\| T^{175} y_0 \|} = \begin{vmatrix} 1 & 0 \\ 0 & 0 \end{vmatrix}$$

Or (cf 4-5), a est le vecteur propre $\left|\begin{array}{c} 1 \\ 0 \\ \hline 0 \\ 0 \\ 0 \end{array}\right|$ de T relatif à la plus grande valeur propre 0.9. De plus b = a d'où $p(a)[p^*(b)]^t = \begin{pmatrix} 1 \\ 0 \end{pmatrix}(1\ \ 0) = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}$

2 Soit l'exemple très simple suivant

$$J = \frac{1}{10} \left|\begin{array}{cc|cc} 1 & 1 & & \\ & 1 & & \\ \hline & & 9 & 1 \\ & & & 9 \end{array}\right| \qquad U = \begin{vmatrix} 4 & 0 & -1 & -1 \\ 0 & 4 & -1 & -1 \\ -1 & -1 & 4 & 0 \\ -1 & -1 & 0 & 4 \end{vmatrix} \quad \text{d'où } U^{-1} = \frac{1}{24} \begin{vmatrix} 7 & 1 & 2 & 2 \\ 1 & 7 & 2 & 2 \\ 2 & 2 & 7 & 1 \\ 2 & 2 & 1 & 7 \end{vmatrix}$$

et $$T = U^{-1}JU = \frac{1}{240} \left(\begin{array}{cc|cc} -10 & -6 & 57 & 65 \\ -34 & -6 & 63 & 71 \\ \hline -71 & -63 & 246 & 58 \\ -65 & -57 & 30 & 250 \end{array}\right)$$

a est le vecteur propre de T : $a = U^{-1}e_4 = \frac{1}{24}\begin{vmatrix} 2 \\ 2 \\ 1 \\ 7 \end{vmatrix}$

d'où, avec le partitionnement indiqué et la norme vectorielle M_∞

$$p(a) = \frac{1}{24}\begin{vmatrix} 2 \\ 7 \end{vmatrix}$$

la matrice limite doit donc avoir toutes ses colonnes proportionnelles au vecteur $\begin{vmatrix} 2 \\ 7 \end{vmatrix}$. Or on a obtenu, avec deux vecteurs de départ y_0 différents, les deux matrices "limites" suivantes :

$$\begin{vmatrix} 0.205 & 0.411 \\ 0.714 & 1.43 \end{vmatrix} \quad \text{et} \quad \begin{vmatrix} 0.004 & 0.008 \\ 0.014 & 0.028 \end{vmatrix}$$

Soit alors l'itération

$$x_{r+1} = Tx_r + h$$

où T est la matrice précédente et $h = \begin{vmatrix} -378 \\ -414 \\ -198 \\ -162 \end{vmatrix}$

dont la solution est $x = \begin{vmatrix} -240 \\ -240 \\ 240 \\ 240 \end{vmatrix}$

A l'itération numéro 100, on a obtenu

$$\varepsilon_{100} = \begin{vmatrix} 0.0429 \\ 0.0430 \\ 0.1406 \\ 0.0313 \end{vmatrix} \qquad \text{d'où } p(\varepsilon_{100}) = \begin{vmatrix} 0.0430 \\ 0.1406 \end{vmatrix}$$

$p(\varepsilon_{100})$ est bien très sensiblement proportionnel au vecteur $\begin{vmatrix} 2 \\ 7 \end{vmatrix}$

3 Soit H_n la matrice de Hilbert de type (n.n) [de terme général

$h_{ij} = \frac{1}{i+j-1}$]. Divisons la par $S_{\infty\infty}(H_n)$ de façon à obtenir une matrice T_n de rayon spectral inférieur à l'unité :

$$T_n = \frac{1}{S_{\infty\infty}(H_n)} \qquad H_n = \frac{1}{\sum_{i=1}^{n} \frac{1}{i}} \qquad \begin{vmatrix} 1 & \frac{1}{2} & \cdots\cdots & \frac{1}{n} \\ \frac{1}{2} & \frac{1}{3} & \cdots\cdots & \frac{1}{n+1} \\ & & & \\ \frac{1}{n} & \cdots\cdots & & \frac{1}{2n-1} \end{vmatrix}$$

Pour n = 10, 20, 30, on a obtenu les résultats suivants

a) n = 10. Norme vectorielle M_∞ correspondant à des blocs (2x2).

p(a) est proportionnel à

$$\begin{vmatrix} 1.3861 \\ 0.6281 \\ 0.4265 \\ 0.3268 \\ 0.2661 \end{vmatrix}$$

b) n = 20. Norme vectorielle M_∞ correspondant à des blocs (2x2)

p(a) est proprotionnel à

$$\begin{vmatrix} 1.5295 \\ 0.7368 \\ 0.5179 \\ 0.4065 \\ 0.3371 \\ 0.2891 \\ 0.2536 \\ 0.2262 \\ 0.2043 \\ 0.1865 \end{vmatrix}$$

c) n = 30. Norme vectorielle M_∞ correspondant à des blocs (3x3) p(a) est proportionnel à

$$\begin{vmatrix} 1.6384 \\ 0.8122 \\ 0.5807 \\ 0.4614 \\ 0.3864 \\ 0.3340 \\ 0.2950 \\ 0.2647 \\ 0.2403 \\ 0.2203 \end{vmatrix}$$

Tous ces exemples montrent que dans une itération linéaire ayant T_n pour matrice d'itération, les premières composantes du vecteur itéré sont celles qui convergent le moins vite vers leur valeur limite. L'utilisation que nous avons faite de normes vectorielles permet de <u>mesurer</u> ce phénomène numérique.

CHAPITRE 6 : MINORANTES, M-MINORANTES et BLOC-H-MATRICES

Alors que les chapitres précédents traitaient de majorantes, nous nous intéresserons ici à l'étude des minorantes d'une matrice A donnée : soit p une norme vectorielle régulière de taille k sur $\mathbb{K}^n$; on appellera minorante (relativement à p) d'une matrice A de type (n,n) toute matrice réelle N de type (k,k) telle que

$$\forall\, x \in \mathbb{K}^n \qquad Np(x) \leq p(Ax)$$

Après avoir rappelé au paragraphe 6-1, la notion de M-matrice, nous caractérisons, en 6-2, la classe des matrices admettant une M-minorante (i.e une minorante qui soit une M-matrice) par la propriété suivante : la matrice de Jacobi par blocs associée est contractante. Lorsque la norme vectorielle utilisée est la norme vectorielle type $p(x) = |x|$ cette classe de matrices coïncide avec celle des H-matrices définie par A. Ostrowski ([43][44]). D'où la désignation de bloc-H-matrice que nous avons adoptée dans le cas général pour désigner une matrice admettant une M-minorante.

Au paragraphe 6-3 nous montrons qu'un résultat de V.V. Gudkov permet de définir une classe de matrices intermédiaire, du point de vue de l'inclusion, entre celle des H-matrices et celle des matrices à diagonale dominante : c'est la classe des G-matrices.

Dans le cas des blocs, on a des définitions analogues et les mêmes inclusions pour la classe des bloc-H-matrices, celle des bloc-G-matrices, et la classe des matrices à bloc diagonale dominante (§ 6-4)

Le paragraphe (6-5) constitue une première tentative d'extension de ces résultats (*) par la définition de B-matrices et de B-minorantes.

(*) nous en formulerons une autre en Annexe.

Malheureusement nous manquons d'une caractérisation, dans le cas général, des matrices qui admettent une B-minorante. Nous donnons néammoins cette caractérisation dans le cas où tous les sous-espaces sur lesquels est construite la norme vectorielle sont de même dimension.

L'application des résultats précédents à l'estimation (en norme vectorielle) de l'approximation obtenue sur la solution d'un système linéaire fait l'objet du paragraphe (6-6) et conduit à une nouvelle caractérisation des blocs H-matrices.

La classe des blocs-H-matrices, qui sera mise en évidence d'une autre manière encore au chapitre suivant, est donc la "bonne" classe de matrices pour l'utilisation d'une norme vectorielle donnée.

6 - 1 - RAPPELS et REMARQUES sur les M-MATRICES

Z_o désignera la classe des matrices réelles de type (k,k) à éléments hors diagonaux négatifs ou nuls. Z désigne la partie de Z_o formée des matrices de Z_o à diagonale strictement positive.

Rappelons le théorème suivant, qui sert de définition à la classe des M-matrices ([18] p. 387)

6 - 1 - 1 - Théorème

Soit $N \in Z_o$. *Les propositions suivantes sont équivalentes :*

1°/ *Il existe* $u \geq 0$ *tel que* $Nu > 0$

2°/ *Il existe* $u > 0$ *tel que* $Nu > 0$

3°/ *Il existe une diagonale positive* Δ *telle que* $N\Delta e > 0$
(e : *vecteur dont toutes les composantes sont égales à* 1)

4°/ *Il existe une diagonale positive* Δ *telle que* $W = N\Delta$ *est une matrice à diagonale positive dominante.*

5°/ *Pour toute matrice diagonale* R *telle que* $R \geq N$, *l'inverse* R^{-1} *existe et* $\rho\left(R^{-1}(D-N)\right) < 1$, *où* D *est la diagonale de* N.

6°/ *si* $B \in Z_o$ *et* $B \geq N$ *alors* B^{-1} *existe*

7°/ *Toute valeur propre réelle de* N *est positive*

8°/ *Tous les mineurs principaux de* N *sont positifs*

9°/ *Il existe une séquence strictement croissante de sous-ensembles de* $K = \{1,2\ldots k\}$ *soit*

$\emptyset \neq M_1 \subset M_2 \ldots\ldots \subset M_k = K$ *telle que les mineurs principaux* $\text{dét}\left(N(M_i)\right)$ *sont* > 0

10°/ *Il existe une matrice de permutation* P *telle que* PNP^T *peut être écrite sous la forme d'un produit* RS *où* R, *triangulaire inférieure appartient à* Z *et* S, *triangulaire supérieure, appartient également à* Z.

11°/ L'inverse N^{-1} *existe et* $N^{-1} \geq 0$

12°/ La partie réelle de toute valeur propre de N *est positive*

13°/ Pour tout vecteur $x \geq 0$, *il existe un indice* i *tel que* $x_i y_i > 0$ *pour* $y = Nx$

6-1-2 - Définition

On appelle M-matrice toute matrice de Z_0 vérifiant l'une des conditions équivalentes ci-dessus énoncées.

Il est facile de voir que toute M-matrice appartient à Z.

Soit N une matrice de Z, D sa diagonale (positive). On posera :

$$U = I - D^{-1}N \qquad \text{de sorte que :}$$
$$N = D(I-U) \qquad (U \geq 0)$$

6-1-3 - Proposition *Pour que* N *soit une* M-*matrice, il faut et il suffit que* $\rho(U)$ *soit inférieur à l'unité*

La condition est évidemment nécessaire; à partir du point 5) du théorème précédent, en prenant pour diagonale R la diagonale D de N elle-même.

Elle est suffisante : si $\rho(U) < 1$

$$N^{-1} = (I-U)^{-1}D^{-1} = (I+U + \dots\dots U^r + \dots)D^{-1}$$

ce qui montre que N^{-1} existe et est ≥ 0 : N est une M-matrice (point 11) du théorème précédent)

6-1-4 - Définition

Soit N une matrice (k,k) et ψ un vecteur ≥ 0. On posera

$$\mathscr{D}(N,\psi) = \{u \geq 0 \qquad Nu \leq \psi\}$$

Nous pouvons donner une autre caractérisation des M-matrices parmi les matrices de la classe Z :

6-1-5 - Proposition *Soit* $N \in Z$. *Pour que* N *soit une* M-*matrice il faut et il suffit que, pour tout* $\psi \geq 0$

$\mathscr{D}(N,\psi)$ *soit borné dans* $\mathbb{R}_+^k$

α) *la condition est nécessaire.* Si N est une M-matrice, de diagonale D,

$N = D(I-U)$ avec $\rho(U) < 1$

Soit $\psi \geq 0$ et $u \in \mathcal{D}(N,\psi)$:

$$D(I-U)u \leq \psi$$

d'où

$$u \leq (I + U + \ldots\ldots U^r + \ldots)D^{-1}\psi$$

ce qui montre que, pour tout $\psi \geq 0$, $\mathcal{D}(N,\psi)$ est borné.

β) *la condition est suffisante* Soit $N \in Z$, tel que $\mathcal{D}(N,\psi)$ est borné pour tout $\psi \geq 0$. Notons encore D, la diagonale (positive) de N, et posons

$$U = I - D^{-1}N \geq 0$$

Si $\rho(U)$ était supérieur ou égal à 1, U admettrait, d'après le théorème de Perron-Frobenius sur les matrices non négatives, [55], une valeur propre ρ égale à son rayon spectral, donc supérieure ou égale à 1, à laquelle on peut associer un vecteur propre x ($x \neq 0$) non négatif ($x \geq 0$)

$$Ux = \rho x$$

de sorte que

$$Nx = D(I-U)x = D(1-\rho)x \leq 0$$

Ainsi $\mathcal{D}(N,0)$ contient λx pour tout $\lambda > 0$: $\mathcal{D}(N,0)$ n'est pas borné : la conclusion étant en contradiction avec l'hypothèse, on conclut que $\rho(U) < 1$, c'est-à-dire, d'après le lemme (6-1-3) que N est une M-matrice.

Remarque Si N est une M-matrice, $N^{-1}\psi$ *est* *__le plus grand élément__ de* $\mathcal{D}(N,\psi)$

6 - 2 - MINORANTES. RECHERCHE de M-MINORANTES

Soit A un opérateur linéaire sur $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$) et soit la décomposition en somme directe :

$$\mathbb{K}^n = W_1 + \ldots\ldots + W_k$$

Soit p une norme vectorielle (régulière de taille k) construite sur les W_i (cf 3-1-3).

Pour tout x dans $\mathbb{K}^n$ nous noterons $x_i = P_i x$ la projection de x dans W_i, P_i désignant l'opérateur de projection. A est alors décomposée en "blocs"

$$A_{ij} \underset{df}{=} P_i A P_j \quad (1 \leq i,j \leq k) \qquad \text{(cf 4-1-4)}$$

$$A = \sum_{1 \leq i,j \leq k} A_{ij}$$

En toute rigueur, A_{ij} est un opérateur linéaire sur $\mathbb{K}^n$, appliquant $\mathbb{K}^n$ dans W_i, et de noyau $V_j \underset{df}{=} \underset{r \neq j}{+} W_r$ mais nous l'identifierons parfois avec sa restriction à W_j.

Nous utiliserons la notation

$$A = \begin{pmatrix} A_{ii} \\ \\ A_{ij} \end{pmatrix}_k$$

pour distinguer les blocs diagonaux des blocs hors-diagonaux. Nous noterons, de la même façon, toute matrice B de type (k,k) et d'éléments (b_{ij}) :

$$B = \begin{pmatrix} b_{ii} \\ \\ b_{ij} \end{pmatrix}_k$$

6-2-1 - Définition

Nous appellerons <u>minorante de A</u> (relativement à p) toute matrice N réelle de type (k,k) telle que

$$\forall x \in \mathbb{K}^n \qquad Np(x) \leq p(Ax)$$

Nous noterons

$$S_{ij}\,(A_{ij}) = \sup_{\substack{x_j \in W_j \\ x_j \neq 0}} \left\{ \frac{p_i(A_{ij}x_j)}{p_j(x_j)} \right\}$$

$$s_{ii}\,(A_{ii}) = \inf_{\substack{x_i \in W_i \\ x_i \neq 0}} \left\{ \frac{p_i(A_{ii}x_i)}{p_i(x_i)} \right\}$$

Δ désignera la bloc-diagonale de A

$$\Delta_{ii} = A_{ii} \qquad (i = 1,2 \ldots k) \qquad \text{soit}$$

$$\Delta = \begin{pmatrix} A_{ii} & \\ & \\ 0 & \end{pmatrix}_k$$

et $m(\Delta)$ est la matrice, (k,k), diagonale

$$m(\Delta) = \begin{pmatrix} s_{ii}(A_{ii}) & \\ & \\ 0 & \end{pmatrix}_k$$

On a alors

$$\forall x \in \mathbb{K}^n \qquad m(\Delta)p(x) \leq p(\Delta x)$$

Ainsi $m(\Delta)$ est une minorante de Δ relativement à p; c'est d'ailleurs <u>la plus grande minorante de Δ</u> .

Si M désigne la norme vectorielle sur $\mathcal{M}_{nn}$ issue de p, la matrice

$$\tilde{N}(A) = m(\Delta) - M(A-\Delta)$$

est une minorante de A. En effet

$$\forall x \in \mathbb{K}^n : \tilde{N}(A)p(x) = m(\Delta)p(x) - M(A-\Delta)p(x) \leq p(\Delta x) - p\big((A-\Delta)x\big) \leq p\big((\Delta+A-\Delta)x\big) = p(Ax)$$

$\tilde{N}(A)$ s'écrit

$$\tilde{N}(A) = \begin{pmatrix} s_{ii}(A_{ii}) & \\ & \\ -S_{ij}(A_{ij}) & \end{pmatrix}_k$$

Plusieurs auteurs (Ostrowski, Fiedler, Ptàk, Feingold, Varga; cf[40][41][19][17]) utilisent cette minorante particulière de A et obtiennent divers résultats sous l'hypothèse que $\tilde{N}(A)$est une M-matrice. Les raisonnements invoqués ne nécessitent, en fait, que de pouvoir disposer d'une minorante qui soit une M-matrice. D'où premièrement, la :

6-2-2 - Définition

On appelle M-minorante de A (relativement à p) toute minorante de A (relativement à p) qui est une M-matrice.

et ensuite, le problème de caractériser les matrices A admettant une M-minorante relativement à une norme vectorielle p régulière donnée. Nous avons pu résoudre entièrement ce problème :

Rappelons que Z désigne l'ensemble des matrices réelles (k,k) de diagonale positive et de hors diagonale ≤ 0. On notera Zm l'ensemble des Z-minorantes de A c'est-à-dire l'ensemble des minorantes de A appartenant à Z. Une condition nécessaire évidente pour que A admette une M-minorante est que Zm soit non vide. Or, on a d'abord la

6-2-3 - Proposition *Pour que* Zm *soit non vide, il faut et il suffit que tous les* A_{ii} *soient réguliers*(*) *(c'est-à-dire que* Δ *soit régulier)*

La condition est nécessaire : soit $N = (n_{ij})_{1\leq i,j\leq k}$ appartenant à Zm :

$$\forall x \in \mathbb{K}^n \qquad Np(x) \leq p(Ax) \qquad (1)$$

soit x_i quelconque, non nul dans W_i ; (1) permet d'écrire :

$$n_{ii}p_i(x_i) \leq p_i(A_{ii}x_i) \qquad \text{d'où}$$

$$n_{ii} \leq \operatorname*{Inf}_{\substack{x_i \in W_i \\ x_i \neq 0}} \left\{ \frac{p_i(A_{ii}x_i)}{p_i(x_i)} \right\} = s_{ii}(A_{ii}) \qquad (2)$$

$n_{ii} > 0$ ($N \in$ Zm) entraîne donc que $s_{ii}(A_{ii})$ est lui-même positif,

(*) Par abus de langage nous considérerons A_{ii} comme opérateur de W_i dans W_i

d'où A_{ii} non singulier ($i = 1,2...k$)

La condition est suffisante : Si tous les A_{ii} sont réguliers, Δ et $m(\Delta)$ sont réguliers. La minorante $\tilde{N}(A)$ est de diagonale $m(\Delta)$ positive; c'est une Z-minorante : Zm est non vide.

Nous supposerons donc dorénavant Δ non singulier; nous pouvons alors définir :

$$J = I - \Delta^{-1} A \qquad \text{de sorte que}$$

$$A = \Delta\,(I-J)$$

$$J = \begin{pmatrix} 0 & \\ -A_{ii}^{-1} & A_{ij} \end{pmatrix}_k \qquad \text{est la matrice de}$$

Jacobi par blocs attachée à A et relative à la décomposition en blocs donnée. L'on a

$$\tilde{N}(A) = m(\Delta) - M(\Delta J)$$

6-2-4- Proposition *La matrice de type* (k,k)

$$N(A) = m(\Delta)\ [I-M(J)]$$

est une minorante de A.

En effet

$$\forall x \in \mathbb{K}^n : \quad N(A)p(x) = m(\Delta)[p(x)-M(J)p(x)] \leq m(\Delta)[p(x)-p(Jx)] \leq m(\Delta)p[(I-J)x]$$
$$\leq p(\Delta(I-J)x) = p(Ax)$$

Remarques 1- N(A) s'écrit

$$N(A) = \begin{pmatrix} s_{ii}(A_{ii}) & \\ & \\ -s_{ii}(A_{ii})\ S_{ij}(A_{ii}^{-1}A_{ij}) & \end{pmatrix}_k$$

et l'inégalité

$$s_{ii}(A_{ii})S_{ij}(A_{ii}^{-1}\ A_{ij}) \leq S_{ij}(A_{ij})$$

entraîne que l'on a :

$$\tilde{N}(A) \leq N(A)$$

2 - Soit B une matrice quelconque (n,n). D'où l'inégalité :

$$N(A)M(B) \leq M(AB)$$

En effet :

$$M(AB) = M\big(\Delta(I-J)B\big) \geq m(\Delta)M[(I-J)B]$$
$$\geq m(\Delta)[M(B) - M(JB)]$$
$$\geq m(\Delta)[M(B) - M(J)M(B)]$$
$$= m(\Delta)[I-M(J)]M(B) = N(A)M(B)$$

d'où la triple inégalité

$$\tilde{N}(A)M(B) \leq N(A)M(B) \leq M(AB) \leq M(A)M(B)$$

3 - dans le cas de la norme vectorielle-type $p(x) = |x|$ tous les blocs sont (1,1). On a :

$$A = \begin{pmatrix} a_{ii} & \\ a_{ij} & \end{pmatrix}_n \qquad J = \begin{pmatrix} & 0 \\ -\dfrac{a_{ij}}{a_{ii}} & \end{pmatrix}_n \qquad m(\Delta) = \begin{pmatrix} |a_{ii}| & \\ & 0 \end{pmatrix}_n$$

$$M(J) = |J| = \begin{pmatrix} & 0 \\ \left|\dfrac{a_{ij}}{a_{ii}}\right| & \end{pmatrix}_n \qquad \text{et} \qquad N(A) = \tilde{N}(A) = \begin{pmatrix} |a_{ii}| & \\ -|a_{ij}| & \end{pmatrix}_n$$

6-2-5 - Proposition *Les matrices* $N = D - DU$, *de diagonale* D, *appartenant à* Zm *sont données par*

$$0 < D \leq m(\Delta) \qquad (3)$$
$$M(J) \leq U \qquad (4)$$

Ces équations sont nécessaires : (3) est l'écriture matricielle de (2). De plus, la ième ligne de (1) permet d'écrire, pour $x = x_i + x_j$ avec

$$x_i = -A_{ii}^{-1} A_{ij} x_j \qquad (0 \neq x_j \in W_j)$$

$$-n_{ij} \geq \frac{p_i(A_{ii}^{-1} A_{ij} x_j)}{p_j(x_j)} \; n_{ii}$$

d'où

$$\frac{-n_{ij}}{n_{ii}} \geqslant S_{ij}(A_{ii}^{-1} \quad A_{ij}) \qquad \text{d'où} \qquad (4)$$

Ces équations sont suffisantes : soit $N = D - DU$ telle que (3) et (4) soient vérifiées. On a

$$(I-U)p(x) \leqslant [I-M(J)]p(x) \leqslant [m(\Delta)]^{-1}p(Ax) \leqslant D^{-1}p(Ax)$$

d'où

$$Np(x) = D(I-U)p(x) \leqslant p(Ax) \quad \text{et } N \in Zm$$

Enonçons alors le :

6-2-6 - Théorème 4 *Pour que* A *admette une* M-*minorante relativement à* p, *il faut et il suffit que la minorante de* A :

$$N(A) = m(\Delta)[I-M(J)]$$

soit elle-même une M-*matrice, c'est-à-dire que*

$$\rho\big(M(J)\big) < 1 \qquad (5)$$

c'est-à-dire encore que J *soit contractante relativement à* p

Montrons que cette condition, évidemment suffisante, est nécessaire. Avec les notations précédentes, supposons que la minorante $N = D - DU$ ($0 < D \leqslant m(\Delta)$; $0 < M(J) \leqslant U$) soit une M-matrice. Le rayon spectral de U est donc inférieur à l'unité, d'où

$$\rho\big(M(J)\big) \leqslant \rho(U) < 1$$

et par conséquent N(A) est elle-même une M-matrice.

6-2-7 - Conséquence Soit $A = \Delta(I-J)$ de bloc diagonale Δ régulière d'où $\tilde{N}(A) = m(\Delta)-M(\Delta J)$ et $N(A)=m(\Delta)[I-M(J)]$

Si $\tilde{N}(A)$ est une M-matrice N(A) l'est à fortiori (Théorème 3 ci-dessus). Mais, inversement, N(A) peut être une M-matrice sans que $\tilde{N}(A)$ le soit, comme dans l'exemple suivant :

$$A = \left|\begin{array}{cc|cc} 6 & 5 & 2.1 & 4.0 \\ 7 & 6 & 2.5 & 4.7 \\ \hline 9.6 & 4.7 & 9 & 10 \\ 8.6 & 4.2 & 8 & 9 \end{array}\right|$$

En utilisant le partitionnement indiqué et la norme du max sur chaque bloc, il vient

$$m_\infty(\Delta) = \begin{vmatrix} 1/_{13} & 0 \\ 0 & 1/_{19} \end{vmatrix} \quad M_\infty(\Delta J) = \begin{vmatrix} 0 & 7.2 \\ 14.3 & 0 \end{vmatrix} \quad M_\infty(J) = \begin{vmatrix} 0 & 0.6 \\ 0.8 & 0 \end{vmatrix}$$

$$\text{d'où } \rho\left(M_\infty(J)\right) = \sqrt{0,48} < 1$$

$N(A)$ est une M-matrice; cependant

$$\rho\left(\left[m_\infty(\Delta)\right]^{-1} M_\infty(\Delta J)\right) = \sqrt{25431.12} > 150 > 1$$

$\tilde{N}(A)$ n'est pas une M-matrice

Remarque On a vu que dans le cas de la norme vectorielle type $p(x) = |x|$ $\left(\text{d'où } M(A) = |A| \text{ : tous les blocs sont } (1,1)\right)$ il vient

$$\tilde{N}(A) = N(A) \begin{pmatrix} |a_{ii}| & \\ & \\ -|a_{ij}| & \end{pmatrix}_n$$

Or, dire que $N(A)$ est une M-matrice, c'est, dans la terminologie introduite par Ostrowski, dire que A est une <u>H-matrice</u> :

6 - 3 - RAPPELS et REMARQUES sur les H-MATRICES. LES G-MATRICES.

Rappelons d'abord les

6-3-1 - Définitions

La matrice $A = \begin{pmatrix} a_{ii} & \\ & \\ a_{ij} & \end{pmatrix}_n$ est dite <u>H-matrice</u> ([43],[44]) si la matrice

$$N(A) = \begin{pmatrix} |a_{ii}| & \\ & \\ -|a_{ij}| & \end{pmatrix}$$

est une M-matrice, ce qui impose d'ailleurs aux éléments diagonaux a_{ii} de A d'être tous non nuls. Dire que N(A) est une M-matrice, c'est dire (cf6-1) que la matrice

$$|J| = \begin{pmatrix} & 0 & \\ & & \\ & \left|\frac{a_{ij}}{a_{ii}}\right| & \end{pmatrix}_n$$

est de rayon spectral inférieur à l'unité :

$$\rho(|J|) < 1 \tag{1}$$

A est dite <u>à diagonale dominante</u> si

$$|a_{ii}| > \sum_{j \neq i} |a_{ij}| \qquad (i = 1,2.. n) \tag{2}$$

Il est bien connu que la classe des matrices à diagonale dominante est une sous-classe de la classe des H-matrices. En effet la condition (2) s'écrit aussi

$$S_{\infty\infty}(|J|) < 1 \tag{3}$$

où $S_{\infty\infty}$ est la norme de matrice issue de la norme du max sur $\mathbb{K}^n$ (*) et l'on a

$$\rho(|J|) \leq S_{\infty\infty}(|J|) < 1 \tag{4}$$

Il est bien connu, également, que le résultat qui assure la non-singularité de toute H-matrice (cf [43]), généralise par conséquent le théorème dit d'Hadamard selon lequel toute matrice à diagonale dominante est non-singulière ([63])

(*) à savoir $S_{\infty\infty}(B) = \operatorname{Max}_i \{ \sum_j |b_{ij}| \}$. La condition $S_{\infty\infty}(B) < 1$ s'écrit aussi $|B|\, e < e$ où e est le vecteur de $\mathbb{R}^n$ ayant toutes ses composantes égales à l'unité et où $<$ désigne (entre vecteurs de $\mathbb{R}^n$) l'inégalité <u>stricte</u> composante à composante.

On a en effet, si D désigne la diagonale (non singulière) de la H-matrice A,

$$A = D(I-J)$$

J étant de rayon spectral inférieur à l'unité, I-J est non singulière, d'où le résultat.

Citons enfin un résultat classique concernant les H-matrices :

6-3-2 - Théorème *Si* A *est une* H-*matrice, les méthodes de Jacobi et Gauss Seidel par point correspondantes pour résoudre un système linéaire $Ax = b$ convergent vers la solution unique* $x = A^{-1}b$

Dans un article ([29] en russe) V.V. Gudkov prouve le résultat suivant

6-3- 2 - Théorème *Soit* $A = (a_{ij})$ *une matrice* (n,n) *et soient*

$$R_1 = \sum_{j=2}^{n} |a_{1j}|$$

$$R_i = \sum_{j=1}^{i-1} |a_{ij}| \frac{R_j}{a_{jj}} + \sum_{j=i+1}^{n} |a_{ij}| \qquad (i = 2,3 \ldots n-1)$$

$$R_n = \sum_{j=1}^{n-1} |a_{nj}| \frac{R_j}{a_{jj}}$$

Si l'on a

$$R_i < |a_{ii}| \qquad (i = 1,2 \ldots n) \qquad (5)$$

alors A est non singulière.

La condition (5) sera dite "condition G" (pour Gudkov) et l'on désignera par G-matrice toute matrice vérifiant la condition G. Démontrons alors la

6-3-3 - Proposition *La classe des* G-*matrices contient celle* *des matrices à diagonale dominante et est contenue dans celle des* H-*matrices.*

La condition G impose aux éléments diagonaux de A d'être tous non nuls : la matrice

$$J = \begin{pmatrix} & 0 & \\ & & \\ & \dfrac{-a_{ij}}{a_{ii}} & \end{pmatrix}_n$$

existe, dont nous notons α_{ij} les éléments. Effectuons le changement de variable

$$S_i = \frac{R_i}{|a_{ii}|}$$

il vient :

$$\begin{cases} S_1 = \sum\limits_{j=2}^{n} |\alpha_{1j}| \\ S_i = \sum\limits_{j=1}^{i-1} |\alpha_{ij}|\ S_j + \sum\limits_{j=i+1}^{n} |\alpha_{ij}| \qquad (i = 2,3, \dots n-1) \\ S_n = \sum\limits_{j=1}^{n-1} |\alpha_{nj}|\ S_j \end{cases}$$

Notons : s le vecteur de composantes $(S_1 \dots S_n)$; L la triangulaire inférieure de J; U sa triangulaire supérieure, e le vecteur de composantes égales à 1. Le système s'écrit

$$s = |L|\ s + |U| e$$

d'où

$$s = (I - |L|)^{-1} |U| e$$

et la condition G s'écrit

$$s < e \quad \text{soit}$$

$$(I - |L|)^{-1} |U| e < e$$

soit encore, puisque $(I - |L|)^{-1} = I + |L| + \dots |L|^{n-1}$ est non négative :

$$S_{\infty\infty}\left((I - |L|)^{-1} \; |U|\right) < 1 \qquad (6)$$

Ainsi

La condition G *exprime que la matrice de Gauss-Seidel* $(I-|L|)^{-1}|U|$ *associée à la matrice de Jacobi* $|J| = |L| + |U|$ *est de norme* $S_{\infty\infty}$ *inférieure à l'unité.*

Dire que A est à diagonale dominante, c'est dire que

$$S_{\infty\infty}(|J|) < 1$$

soit

$$|J|e < e$$

ce qui s'écrit $(|J| = |L| + |U|)$:

$$|U|e < (I - |L|)e \qquad (7)$$

Or $(I-|L|)^{-1}$ est une triangulaire inférieure non négative de diagonale unité : elle conserve l'inégalité stricte entre vecteurs réels

$$a < b \implies (I - |L|)^{-1}a < (I - |L|)^{-1}b$$

On tire donc de (7)

$$(I - |L|)^{-1}|U|e < e$$

Ce qui est la condition G : la première inclusion est démontrée : *la classe des* G*-matrices contient celle des matrices à diagonale dominante.*

Mais alors, la condition G

$$S_{\infty\infty}\left((I- |L|)^{-1} \; |U|\right) < 1$$

entraîne que l'on a

$$\rho\left((I-|L|)^{-1} \; |U|\right) < 1 \qquad (8)$$

Or, d'après le théorème de Stein-Rosenberg ([64] p 73) appliqué à la matrice non négative $|J| = |L| + |U|$, l'inégalité (8) implique que l'on a :

$$\rho[(I-|L|)^{-1}|U|] < \rho\,(|L| + |U|) = \rho(|J|) < 1$$

La seconde inclusion est démontrée : *la classe des* H*-matrices contient celle des* G*-matrices.*

6-3-4 - Exemples

$$\text{Soit } A = \begin{vmatrix} -10 & 5 & -2 \\ 7 & -10 & -5 \\ -8 & 3 & 10 \end{vmatrix} \qquad \text{d'où } |J| = \begin{vmatrix} 0 & 0.5 & 0.2 \\ 0.7 & 0 & 0.5 \\ 0.8 & 0.3 & 0 \end{vmatrix}$$

A n'est pas à diagonale dominante mais vérifie la condition G :

$$S_1 = 0.5 + 0.2 = 0.7 < 1$$

$$S_2 = 0.7 \times 0.7 + 0.5 = 0.99 < 1$$

$$S_3 = 0.8 \times 0.7 + 0.3 \times 0.99 < 1$$

De même, toute matrice B conduisant à la matrice

$$|J| = \begin{vmatrix} 0 & 5\ 10^{-4} & 5\ 10^{-4} \\ 1/_6 & 0 & 1/_3 10^{-3} \\ 600 & 700 & 0 \end{vmatrix}$$

est une H-matrice car elle vérifie la condition G

$$S_1 = 5\ 10^{-4} + 5\ 10^{-4} = 10^{-3} < 1$$

$$S_2 = \frac{1}{6}\ 10^{-3} + \frac{1}{3}\ 10^{-3} = \frac{1}{2}\ 10^{-3} < 1$$

$$S_3 = 600.\ 10^{-3} + \frac{700}{2}\ 10^{-3} = 0.95 < 1$$

or

$$S_{\infty\infty}(J) = 600 + 700 = 1300$$

B n'est pas à diagonale dominante!

6 - 4 - BLOC H-MATRICES, BLOC G-MATRICES et MATRICES à BLOC-DIAGONALE DOMINANTE

Revenons au propos du paragraphe 6-2. J étant la matrice de Jacobi par blocs associée à A :

$$J = \begin{pmatrix} & 0 & \\ & & \\ & -A_{ii}^{-1}\ A_{ij} & \end{pmatrix}_k$$

nous noterons L la bloc triangulaire inférieure de J et U sa bloc triangulaire supérieure :

$$L = \begin{pmatrix} 0 & \\ \begin{cases} 0 \text{ si } j > i \\ -A_{ii}^{-1}\ A_{ij} \text{ si } j < i \end{cases} \end{pmatrix}_k \quad U = \begin{pmatrix} 0 & \\ \begin{cases} -A_{ii}^{-1}\ A_{ij} & \text{si } j > i \\ 0 & \text{si } j < i \end{cases} \end{pmatrix}_k$$

Nous sommes alors amenés à poser les

6-4-1 - Définitions

Relativement à la norme vectorielle p, on dira que

* A est une <u>bloc-H-matrice</u> si N(A) est une M-matrice, c'est-à-dire (6-2-6) si J est contractante relativement à p :

$$\rho\big(M(J)\big) < 1$$

* A est une <u>bloc-G-matrice</u> si

$$S_{\infty\infty}\big([I-M(L)]^{-1}M(U)\big) < 1 \quad \text{(condition G)}$$

* A est à <u>bloc diagonale dominante</u> si N(A) est à diagonale dominante, ce qui s'écrit :

$$\sum_{j\neq i} S_{ij}(A_{ii}^{-1}A_{ij}) < 1 \qquad (i = 1,2 \ldots\ldots k) \quad (*)$$

(*) D. Feingold et R.S. Varga [17] donnent une définition plus restrictive de la bloc-diagonale-dominance, à savoir

$$\sum_{j\neq i} S_{ij}(A_{ij}) < s_{ii}(A_{ii}) \qquad (i = 1,2\ldots\ k)$$

[autrement dit : dominance diagonale de $\tilde{N}(A)$]

c'est-à-dire

$$S_{\infty\infty}\big(M(J)\big) < 1$$

Il est clair que la classe des blocs-H-matrices contient celle des blocs-G-matrices qui contient à son tour la classe des matrices à bloc-diagonale dominante : cela se démontre comme en (6-3-3), en remplaçant $|L|$ par $M(L)$ et $|U|$ par $M(U)$. Il est bien clair également que lorsque la norme vectorielle utilisée est $p(x) = |x|$, on a $M(L) = |L|$, $M(U) = |U|$ tous les blocs sont (1,1) et les définitions précédentes coïncident avec celles de H-matrice, de G-matrice ou de matrice à diagonale dominante.

Remarque : Soit S une norme vectorielle sous-multiplicative et régulière sur $\mathcal{M}_{nn}$. D'où (4-2-2) une décomposition en blocs (blocs diagonaux carrés) de toute matrice $A \in \mathcal{M}_{nn}$.

Soit $A \in \mathcal{M}_{nn}$, de bloc diagonale Δ régulière, d'où :

$$A = \Delta(I - J)$$

Si $\rho\big(S(J)\big) < 1$, alors on sait (4-2-4) exister une norme vectorielle p sur $\mathbb{K}^n$ générant M sur $\mathcal{M}_{nn}$ telle que

$$\forall B \in \mathcal{M}_{nn} \qquad M(B) \leq S(B)$$

d'où en particulier

$$M(J) \leq S(J) \qquad \text{d'où enfin}$$

$$\rho\big(M(J)\big) \leq \rho\big(S(J)\big) < 1$$

et A est, relativement à p, une bloc H-matrice.

6-4-2 - Exemple type de bloc-H-matrice

Soit la matrice dite "du potentiel"

$$A = \left|\begin{array}{ccccc} H & I & & & \\ I & H & I & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & I \\ & & & I & H \end{array}\right| \quad (\lambda k, \lambda k)$$

où I est la matrice unité (λ,λ) et

$$H = \begin{vmatrix} -4 & 1 & & & \\ 1 & -4 & 1 & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & 1 & -4 \end{vmatrix}_{(\lambda,\lambda)}$$

d'où, relativement aux blocs mis en évidence :

$$J = \begin{vmatrix} 0 & -H^{-1} & & \\ -H^{-1} & 0 & & \\ & & & -H^{-1} \\ & & -H^{-1} & 0 \end{vmatrix}_{(\lambda k,\lambda k)}$$

Les valeurs propres de H sont :

$$\mu_p = 4-2\cos\frac{p\pi}{\lambda+1} \quad (p = 1,\ 2,\ \ldots\ \lambda)$$

$$2 < \mu_p < 6$$

Le rayon spectral de H^{-1} est donc inférieur à 1/2. Il existe donc, d'après un théorème de Householder-Ostrowski, rappelé plus haut (4-3) une norme ϕ sur $\mathbb{R}^\lambda$ telle que

$$\gamma = \operatorname*{Sup}_{\substack{x\in\mathbb{R}^\lambda \\ x\neq 0}} \left\{\frac{\phi(H^{-1}x)}{\phi(x)}\right\} \underset{df}{=} S_{\varphi\varphi}(H^{-1}) < 1/_2$$

D'où, en prenant la même norme φ sur chacun des blocs :

$$M(J) = \begin{vmatrix} 0 & \gamma & & \\ \gamma & 0 & \ddots & \\ & \ddots & \ddots & \gamma \\ & & \gamma & 0 \end{vmatrix}_{(k,k)}$$

de rayon spectral

$$\rho\big(M(J)\big) \leqslant S_{\infty\infty}\big(M(J)\big) = 2\gamma < 1$$

Ainsi, la matrice classique du potentiel est à bloc-diagonale dominante pour la norme vectorielle construite avec la norme φ; c'est donc, relativement à cette norme vectorielle, une bloc-H-matrice.

De même que celle des H-matrices (6-3-1) la classe des blocs-H-matrices est une classe de matrices non singulières :

6-4-3 - Proposition *Toute bloc* H-*matrice* A *est non singulière. On a alors :*

$$[N(A)]^{-1} \geq M(A^{-1})$$

et le module minimum des valeurs propres de N(A) *n'est pas supérieur au module minimum des valeurs propres de* A.

En effet : $\forall\, x \in \mathbb{K}^n \qquad N(A)p(x) \leq p(Ax)$

N(A) est une M-matrice : elle admet une inverse non négative d'où :

$$\forall\, x \in \mathbb{K}^n \qquad p(x) \leq [N(A)]^{-1} p(Ax) \quad (1)$$

D'où les implications suivantes qui prouvent la non singularité de A :

$Ax = 0 \Rightarrow p(Ax) = 0 \Rightarrow p(x) = 0 \Rightarrow x=0$

Alors, (1) peut s'écrire, de façon équivalente

$$\forall\, y \in \mathbb{K}^n \qquad p(A^{-1}y) \leq [N(A)]^{-1}\, p(y)$$

ce qui montre que $[N(A)]^{-1}$ est une majorante de A^{-1} relativement à p. D'où, puisque $M(A^{-1})$ est la plus petite des majorantes de A^{-1}, l'inégalité annoncée :

$$M(A^{-1}) \leq [N(A)]^{-1} \qquad (2)$$

D'où la même inégalité sur les rayons spectraux de ces matrices non-négatives.

$$\rho(A^{-1}) \leq \rho\big(M(A^{-1})\big) \leq \rho([N(A)]^{-1})$$

D'où en notant $\sigma(A) = \dfrac{1}{\rho(A^{-1})}$ le module minimum des valeurs propres de A

$$\sigma(A) \geq \sigma\big(N(A)\big)$$

cette inégalité entre rayons spectraux est symétrique de celle établie en (4-1-3)

<u>Application</u> : La matrice du potentiel, puisqu'elle est (6-4-2) une bloc H-matrice, est non singulière.

<u>Exemple</u> : (emprunté à [3] p 109)

soit $A = \begin{vmatrix} 3 & 1 & 1 \\ 1 & 3 & 1 \\ 1 & 1 & 3 \end{vmatrix}$ et sur $\mathbb{R}^3$ la norme vectorielle $p(x) = |x|$

d'où

$$N(A) = \begin{vmatrix} 3 & -1 & -1 \\ -1 & 3 & -1 \\ -1 & -1 & 3 \end{vmatrix} \quad \text{et } [N(A)]^{-1} = \begin{vmatrix} 0.5 & 0.25 & 0.25 \\ 0.25 & 0.5 & 0.25 \\ 0.25 & 0.25 & 0.5 \end{vmatrix}$$

or

$$A^{-1} = \begin{vmatrix} 0.4 & -0.1 & -0.1 \\ -0.1 & 0.4 & -0.1 \\ -0.1 & -0.1 & 0.4 \end{vmatrix} \quad \text{d'où } M(A^{-1}) = \begin{vmatrix} 0.4 & 0.1 & 0.1 \\ 0.1 & 0.4 & 0.1 \\ 0.1 & 0.1 & 0.4 \end{vmatrix}$$

l'inégalité (2) est vérifiée.

6 - 5 - B-MATRICES et RECHERCHE de B-MINORANTES

Soit N une matrice réelle (k,k). On rappelle que, pour tout ψ pris dans $\mathbb{R}^k_+$, on a posé (cf 6-1-4) :

$$\mathcal{D}(N,\psi) \underset{df}{=} \{u \in \mathbb{R}^k_+ \quad Nu \leq \psi\}$$

la définition suivante nous sera utile :

<u>6-5-1 - Définition</u>

N est dite <u>B-matrice</u>, si, quel que soit ψ dans $\mathbb{R}^k_+$, $\mathcal{D}(N,\psi)$ est une partie <u>bornée</u> de $\mathbb{R}^k_+$

Avec cette définition, la proposition (6-1-5) s'énonce ainsi

<u>"les B-matrices de la classe Z sont les M-matrices".</u>

Nous allons caractériser la classe des B-matrices.

6-5-2 - Caractérisation (N.Gastinel) (*)

Soit N *une matrice réelle* (k,k)*. Les propositions suivantes sont équivalentes*

a) est une B-matrice

b) le cône $\mathcal{D}(N,0) = \{u \in \mathbb{R}^k_+ \quad Nu \leq 0\}$ *est réduit à* {0}

c) il existe $v>0$ *dans* $\mathbb{R}^k$ *tel que* $N^t v > 0$

L'équivalence de b) et c) est la conséquence de théorèmes généraux de dualité. Bornons-nous à démontrer l'équivalence de a) et b)

1) a => b   Montrons en effet que $\overline{b} \Rightarrow \overline{a}$

Soit N une matrice réelle (k,k) telle que $\mathcal{D}(N,0) \neq \{0\}$: il existe $u \geq 0$ $(u \neq 0)$ tel que $Nu \leq 0$. Ainsi $\mathcal{D}(N,0)$ contient λu pour tous les $\lambda > 0$ et n'est donc pas borné : N n'est pas une B-matrice.

2) b => a. Montrons que $\overline{a} \Rightarrow \overline{b}$

Soit N une matrice (k,k) qui ne soit pas une B-matrice : il existe $\psi \geq 0$ et une suite $(u_n)_{n \in N}$, non bornée de vecteurs non négatifs tels que

$$Nu_n \leq \psi \quad \left(\text{ie :} \quad u_n \in \mathcal{D}(N,\psi)\right) \quad (1)$$

Il n'y a pas de restriction à supposer (quitte à prendre une sous-suite de la suite des u_n) que pour une certaine norme ϕ sur $\mathbb{R}^k$ on a

$$\phi(u_{n+1}) > \phi(u_n)$$

$$\lim_{n \to \infty} \phi(u_n) = \infty$$

Posons $V_n = \dfrac{u_n}{\phi(u_n)}$. La suite des Vn est une suite de vecteurs non négatifs situés sur la frontière de la boule-unité de ϕ. (1) s'écrit

$$N\, v_n \leq \frac{\psi}{\phi(u_n)} \qquad (2)$$

On peut extraire de la suite des v_n une sous suite convergente; soit v sa limite, et $(v_{r_1}, \ldots\ldots v_{r_n} \ldots)$ cette sous suite;

(*) voir également [21] où la classe des B-matrices est aussi définie on a

$$v \geq 0$$

$$\phi(v) = 1$$

Par hypothèse $\phi(u_{r_n})$ tend vers l'infini avec n. (2) permet alors d'écrire

$$Nv \leq 0$$

Ce qui montre que $\mathcal{D}(N,0)$ n'est pas réduit à zéro puisque contenant v, vecteur non négatif et de norme ϕ égale à l'unité.

Soit alors p une norme vectorielle régulière de taille k sur $\mathbb{K}^n$, générant M sur $\mathcal{M}_{nn}$

6-5-3 - Définition

Soit $A \in \mathcal{M}_{nn}$. On appellera B-minorante de A (relativement à p) toute minorante de A (relativement à p) qui est une B-matrice.

Il semble alors naturel d'essayer de plonger la classe des blocs H-matrices dans une classe plus vaste, définie comme la classe des matrices admettant une B-minorante relativement à p. En effet, certains résultats vrais pour les blocs-H-matrices sont encore vrais pour cette nouvelle classe ; par exemple :

6-5- 4 - Proposition *Soit* $A \in \mathcal{M}_{nn}$. *Si* A *admet une* B-*minorante,* A *est non-singulière.*

En effet soit N une B-minorante de A :

$$\begin{cases} \forall\, x \in \mathbb{K}^n & Np(x) \leq p(Ax) \\ \mathcal{D}(N, 0) = \{0\} \end{cases}$$

Soit alors x tel que $Ax = 0$ d'où $Np(x) \leq 0$

Ainsi $p(x) \in \mathcal{D}(N,0) = \{0\}$

D'où $p(x) = 0$

d'où $x = 0$

Cependant, le problème de la caractérisation des matrices qui admettent (relativement à une norme vectorielle régulière donnée) une B-minorante, n'a pu être résolu dans sa généralité. Néammoins, on peut établir le résultat suivant :

6-5-5 - Proposition *Si tous les blocs sont carrés (donc de même type) toute* B-*minorante de* A *est déduite d'une* M-*matrice par permutation de lignes*

En effet, soit N une B-minorante de A.

a) toute colonne de N contient un élément >0

sinon soit N_j une colonne ≤ 0 de N. e_j désignant le $j^{\text{ème}}$ vecteur de base, on a

$$N\, e_j = N_j \leq 0$$

$\mathcal{D}(N,0)$, contenant $e_j \neq 0$, n'est pas réduit à $\{0\}$ et N ne serait pas une B-matrice, conclusion absurde.

Soit n_{ij} un élément positif de la $j^{\text{ème}}$ colonne de N. N minore A :

$$Np(x) \leq p(Ax)$$

d'où, pour $x = x_j \in W_j - \{0\}$

$$n_{ij}\, p_j(x_j) \leq p_i(A_{ij}x_j)$$

$$0 < n_{ij} \leq \operatorname*{Inf}_{\substack{x_j \in W_j \\ x_j \neq 0}} \left\{ \frac{p_i(A_{ij}x_j)}{p_j(x_j)} \right\}$$

ce qui montre que A_{ij}, carrée par hypothèse, est non-singulière. D'où, pour $x = x_j + x_r \quad (x_j \in W_j ;\ x_r \in W_r)$

$$n_{ij}p_j(x_j) + n_{ir}p_r(x_r) \leq p_i(A_{ij}x_j + A_{ir}x_r)$$

soit pour $x_j = -A_{ij}^{-1}\; A_{ir}x_r$

$$n_{ir} \leq -n_{ij} \sup_{\substack{x_r \in W_r \\ x_r \neq 0}} \left\{ \frac{p_j(A_{ij}^{-1}\; A_{ir}x_r)}{p_r(x_r)} \right\} \leq 0$$

D'où

b) *si une ligne de* N *contient un élément* > 0, *elle en contient un un seul* . D'où

c) N *contient un élément positif et un seul par colonne*

[conséquence directe de a et b]

D'où

d) chaque ligne de N *contient au moins un élément positif*

En conclusion :

N *contient un élément positif et un seul par ligne et par colonne*

Il existe donc une matrice de permutation P telle que

$$N = PM$$

où M est de la classe Z (diagonale positive, hors diagonale ≤ 0)

Il est alors facile de voir que pour que N soit une B-matrice, il faut et il suffit que M soit une M-matrice, en effet :

$$\mathcal{D}(N,\psi) = \{u \in \mathbb{R}_+^k \ Nu \leq \psi\} = \{u \in \mathbb{R}_+^k \quad Mu \leq P^t\psi\} = \mathcal{D}(M, P^t\psi)$$

et dire que $\mathcal{D}(N,\psi) = \mathcal{D}(M,P^t\psi)$ est, pour tout ψ, une partie bornée de $\mathbb{R}_+^k$ c'est (6-1-5) dire que M est une M-matrice.

D'où la :

6-5-6 - Proposition *Soit* p *une norme vectorielle régulière de taille* k *sur* $\mathbb{K}^n$ ($\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$) *telle que tous les sous-espaces* W_i *sur lesquels elle est construite soient de même dimension* λ *(d'où* $n = k\lambda$*)*

Pour que $A \in \mathcal{M}_{nn}$ *admette une* B-*minorante relativement à* p, *il faut et il suffit que la condition suivante soit réalisée.*

On peut choisir un bloc et un seul par ligne et colonne (de blocs) qui soit non singulier. Soit $A_{ij(i)}$ *ce bloc* $(i = 1,2 \ldots. k)$

Alors la minorante $N = (n_{ij})$, *définie par*

$$n_{ij(i)} = s_{ij(i)}\left(A_{ij(i)}\right) \underset{df}{=} \operatorname*{Inf}_{\substack{x \in W_{j(i)} \\ x \neq 0}} \left\{ \frac{p_i(A_{ij(i)}x)}{p_{j(i)}(x)} \right\} > 0$$

$$n_{ij} = -s_{ij(i)}\left(A_{ij(i)}\right) S_{j(i)j}\left(A_{ij(i)}^{-1} A_{ij}\right) \qquad \left(\text{si } j \neq j(i)\right)$$

est une permutée de M *-matrice.*

6 - 5 - 7 - Exemples

1) Prenons $k = 2$ d'où

$$A = \begin{vmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{vmatrix}$$

les quatre sous-matrices étant carrées de même type

α) Recherche d'une M-minorante

Pour que A admette une M-minorante, c'est-à-dire soit une bloc-H-matrice, il est d'abord nécessaire que A_{11} et A_{22} soient non singulières. D'où

$$N(A) = \begin{vmatrix} s_{11}(A_{11}) & -s_{11}(A_{11})\, S_{12}(A_{11}^{-1}A_{12}) \\ -s_{22}(A_{22})S_{21}(A_{22}^{-1}\, A_{21}) & s_{22}(A_{22}) \end{vmatrix}$$

et

$$M(J) = \begin{vmatrix} 0 & S_{12}(A_{11}^{-1}A_{12}) \\ S_{21}(A_{22}^{-1}\, A_{21}) & 0 \end{vmatrix}$$

la condition nécessaire et suffisante pour que A soit une bloc H-matrice est alors:

$$\rho\big(M(J)\big) < 1 \quad \text{soit, ici,}$$

$$S_{12}(A_{11}^{-1}\, A_{12})\; S_{21}(A_{22}^{-1}A_{21}) < 1 \qquad (1)$$

β) recherche d'une B-minorante

Il n'y a ici (car $k = 2$) qu'une autre possibilité : les matrices A_{12} et A_{21} sont non singulières et la matrice minorante:

$$N_1(A) = \begin{vmatrix} -s_{12}(A_{12})\ S_{21}(A_{12}^{-1}A_{11}) & s_{12}(A_{12}) \\ & \\ s_{21}(A_{21}) & -s_{21}(A_{21})\ S_{12}(A_{21}^{-1}\ A_{22}) \end{vmatrix}$$

est une permutée de M-matrice, d'où la condition

$$S_{12}(A_{21}^{-1}\ A_{22})\ S_{21}(A_{12}^{-1}\ A_{11}) < 1 \qquad (2)$$

Ainsi :

Pour que A admette une B-minorante, il faut et il suffit que (1) ou (2) soient vérifiées. La B-minorante est alors respectivement N(A) ou $N_1(A)$

Remarque : Lorsque les quatre sous-matrices A_{11}, A_{12}, A_{21}, A_{22} sont non singulières, on a

$$A_{11}^{-1}\ A_{12}\quad A_{12}^{-1}\ A_{11}\ A_{21}^{-1}\ A_{22}\quad A_{22}^{-1}\ A_{21} = I \qquad \text{d'où}$$

$$S_{11}(\qquad\qquad) = 1 \qquad \text{d'où}$$

$$1 \leqslant S_{12}(A_{11}^{-1}A_{12})\ S_{21}(A_{12}^{-1}A_{11})\ S_{12}(A_{21}^{-1}\ A_{22})\ S_{21}(A_{22}^{-1}\ A_{21}) \qquad \text{soit}$$

$$S_{12}(A_{11}^{-1}A_{12})\ S_{21}(A_{22}^{-1}\ A_{21}) \geqslant \frac{1}{S_{12}(A_{21}^{-1}\ A_{22})\ S_{21}(A_{12}^{-1}\ A_{11})}$$

ce qui montre que, lorsqu'elles peuvent être l'une et l'autre testées (A_{11} A_{12} A_{21} A_{22} régulières) les conditions (1) et (2) s'excluent mutuellement. L'une des deux est alors toujours vérifiée, sauf dans le cas limite où

$$\begin{cases} S_{12}(A_{11}^{-1}\ A_{12})\ S_{21}(A_{22}^{-1}\ A_{21}) = 1 \\ S_{12}(A_{21}^{-1}\ A_{22})\ S_{21}(A_{12}^{-1}\ A_{11}) = 1 \end{cases}$$

2) Cas de la norme vectorielle type $p(x) = |x|$

C'est bien un cas où tous les blocs, puisqu'ils sont (1,1) sont de même type. Nous allons voir que, dans ce cas simple, on peut expliciter entièrement la forme de toute minorante d'une matrice A donnée.

Soit donc A une matrice (n,n). Quelles sont les matrices N telles que

$$\forall x \in \mathbb{K}^n \qquad N\,|x| \leq |Ax| \tag{3}$$

On notera n_i^t et a_i^t les ièmes lignes de N et A, (3) s'écrit :

$$\forall x \in \mathbb{K}^n \qquad n_i^t\,|x| \leq |a_i^t\,x| \qquad i = 1,2 \ldots n$$

Le problème de la détermination de l'ensemble des minorantes de A se ramène donc au suivant :

soit $a \in \mathbb{K}^n$. Caractériser Q_a, la partie constituée des éléments s de $\mathbb{K}^n$ tels que

$$\forall x \in \mathbb{K}^n \qquad s^t\,|x| \leq |a^t\,x| \tag{4}$$

On posera

$$r_j = \begin{vmatrix} -|a_1| \\ +|a_j| \\ -|a_n| \end{vmatrix} \qquad (j = 1, 2 \ldots n)$$

α) <u>tous les r_j appartiennent</u> à Q_a:

$$r_j^t|x| = |a_j|\;|x_j| - \sum_{k\neq j} |a_k|\;|x_k| \leq |a_j x_j| - \Big|\sum_{k\neq j} a_k x_k\Big| \leq \Big|\sum_{i=1}^n a_i x_i\Big| = |a^t x|$$

β) <u>pour tout s dans</u> Q_a

il existe $j \in \{1,2 \ldots n\}$ et $\alpha \in [01]$: $\quad s \leq \alpha r_j \qquad (5)$

<u>et réciproquement, si s vérifie (5) il appartient à</u> Q_a. <u>Autrement dit</u> :

$$Q_a = \{s \in \mathbb{K}^n \quad \exists\, j \in \{1,2 \ldots n\},\ \exists\, \alpha \in [01] \quad s \leq \alpha r_j\}$$

1) si $s \leq \alpha r_j \qquad (\alpha \in [01])$

$s^t\,|x| \leq \alpha r_j^t|x| \leq \alpha\,|a^t\,x| \leq |a^t\,x| \qquad$ d'où $s \in Q_a$

2) si $s \in Q_a$

α) si $s \leq 0$ s vérifie trivialement (5) [il suffit en effet de prendre j quelconque et $\alpha = 0$]

β) si $s \nleq 0$ s a une composante s_j, positive
(4) permet d'écrire, pour $x = e_j$, $j^{\text{ème}}$ vecteur de base :

$$s_j \leq |a_j|$$

d'où l'on conclut; puisque s_j est > 0, que a_j n'est pas nul.

soit $x = x_j e_j + x_k e_k$ avec

$$x_j = \frac{-a_k \, x_k}{a_j} \qquad (a_j \neq 0)$$

(4) permet d'écrire

$$s_j \left|\frac{a_k}{a_j}\right| + s_k \leq 0$$

d'où

$$\begin{cases} s_k \leq -s_j \left|\dfrac{a_k}{a_j}\right| & k = 1,2 \ldots n; \quad k \neq j \\ s_j \leq |a_j| & \end{cases}$$

soit :

$$s \leq \frac{s_j}{|a_j|} \, r_j \qquad (6)$$

Posons $\alpha = \dfrac{s_j}{|a_j|}$. (6) s'écrit

$$s \leq \alpha \, r_j \qquad (\alpha \; [01])$$

Remarque $Q_a = Q_{|a|}$

Revenant au problème initial, nous sommes alors en mesure d'expliciter toute minorante de A (relativement à la norme vectorielle $p(x) = |x|$).
On posera

$$r_{ij} = \left| \begin{array}{c} -|a_{i1}| \\ \vdots \\ +|a_{ij}| \\ \vdots \\ -|a_{in}| \end{array} \right. \qquad (i,j = 1,2 \ldots n)$$

et toute minorante N de A peut être ainsi définie : pour tous les i (de 1 à n) il existe j(i) appartenant à $\{1,2 \ldots\ldots n\}$ et $\alpha_i \in [01]$ tel que

$$n_i \leqslant \alpha_i \, r_{ij(i)} \qquad \text{soit}$$

$$N \leqslant \begin{vmatrix} \alpha_1 & & \\ & \ddots & \\ & & \alpha_n \end{vmatrix} \begin{vmatrix} r^t_{1j(1)} \\ \vdots \\ r^t_{nj(n)} \end{vmatrix}$$

Donc <u>toute minorante N est définie par $N \leqslant DS$ où $0 \leqslant D \leqslant I$ et où S est l'une des n^n matrices obtenues en changeant dans la matrice $|A|$, le signe d'un élément arbitraire par ligne</u> (*)

A quelle condition existe-t-il une B-minorante? Si N est une B-matrice, tout $M \geqslant N$ est également une B-matrice : cela résulte du point c) de (6-5-2)

Ainsi, pour qu'existe une B-minorante , il faut et il suffit que l'une des matrices S soit une B-matrice : elle sera alors B-minorante. Dire que l'une des matrice S est une B-matrice, c'est, à cause de la répartition des signes des éléments de S, dire qu'elle est de la forme

$$S = PM$$

où M est une M-matrice et P une matrice de permutation. L'application $i \to j(i)$ est alors une bijection de $\{1,2\ldots n\}$, et l'on se trouve bien décrire un cas particulier de la proposition (6-5-6)

En d'autres termes, soit, sur $\mathcal{M}_{nn}$, l'opérateur N :

$$B = \begin{pmatrix} b_{ii} & \\ & b_{ij} \end{pmatrix}_n \longrightarrow N(B) = \begin{pmatrix} |b_{ii}| & \\ & -|b_{ij}| \end{pmatrix}_n$$

(*) il est clair que les matrices S sont elles-mêmes des minorantes de A (D = I)

<u>Pour que $A \in \mathcal{M}_{nn}$ admette une B-minorante relativement à la norme vectorielle $p(x) = |x|$, il faut et il suffit qu'il existe une matrice de permutation P telle que N(PA) soit une M-matrice (*). Alors $P^tN(PA)$ est une B-minorante de A.</u>

Exemple :

$$A = \begin{vmatrix} 2 & -1 & 4 \\ 3 & 1 & -1 \\ 1 & -5 & 2 \end{vmatrix} \quad \text{et soit } P = \begin{vmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \end{vmatrix}$$

d'où

$$PA = \begin{vmatrix} 3 & 1 & -1 \\ 1 & -5 & 2 \\ 2 & -1 & 4 \end{vmatrix} \qquad N(PA) = \begin{vmatrix} 3 & -1 & -1 \\ -1 & 5 & -2 \\ -2 & -1 & 4 \end{vmatrix}$$

PA est bien une H-matrice puisque de diagonale dominante d'où la B-minorante de A relativement à la norme vectorielle $p(x) = |x|$

$$S = P^tN(PA) = \begin{vmatrix} -2 & -1 & 4 \\ 3 & -1 & -1 \\ -1 & 5 & -2 \end{vmatrix}$$

6 - 6 - APPLICATION : ESTIMATION (en norme vectorielle) de l'APPROXIMATION OBTENUE sur la SOLUTION d'un SYSTEME LINEAIRE

Soit $Ax = b$ un système linéaire, de solution inconnue x

$A \in \mathcal{M}_{nn}$, espace des matrices (n,n) sur $\mathbb{K}$ ($\mathbb{R}$ ou $\mathbb{C}$), $b \in \mathbb{K}^n$

Soit $\overset{*}{x}$ une "solution approchée" de x (c'est-à-dire ici un vecteur parfaitement quelconque de $\mathbb{K}^n$!)

On posera

$\varepsilon = x^* - x$ (vecteur, inconnu, d'erreur)

$\rho = Ax^* - b = A(x^* - x) = A\varepsilon$ (vecteur résidu; connu)

Soit p une norme vectorielle régulière de taille k sur $\mathbb{K}^n$. Tout le problème consiste à estimer

$$p(\varepsilon) = p(x^* - x)$$

(*) i.e : PA est une H-matrice

On peut procéder ainsi : Soit N une minorante de A relativement à p :

$$\forall\, y \in \mathbb{K}^n \qquad Np(y) \leq p(Ay)$$

D'où, puisque $\qquad A\varepsilon = \rho$,

$$Np(\varepsilon) \leq p(\rho)$$

Cette inégalité impose au vecteur inconnu $p(\varepsilon)$ d'être dans la partie $\mathcal{D}(N, p(\rho))$: rappelons que pour tout $\psi \in \mathbb{R}^k_+$ on a posé (6-1-4)

$$\mathcal{D}(N,\psi) = \{u \in \mathbb{R}^k_+ \quad Nu \leq \psi\}$$

Ainsi

6-6-1 - Proposition *La connaissance d'une minorante* N *de* A *relativement à* p *permet l'estimation d'erreur suivante :*

$$p(\varepsilon) \in \mathcal{D}(N, p(\rho)) \qquad (1)$$

Pour que cette estimation ait un intérêt numérique, il convient que $\mathcal{D}(N, p(\rho))$ soit une partie bornée de $\mathbb{R}^k_+$ quel que soit ρ dans $\mathbb{K}^n$. Or p, étant régulière, est surjective (3-1-4). Pour que l'estimation (1) ait un sens numérique, il est donc nécessaire et suffisant que $\mathcal{D}(N,\psi)$ soit borné quel que soit $\psi \geq 0$ dans $\mathbb{R}^k$; c'est-à-dire, par définition, que la minorante N soit une B-matrice (6-5-1) :

6-6-2 - Proposition *Pour que l'estimation d'erreur* (1) *ait un sens numérique, il faut et il suffit que* N *soit une* B-*minorante*

Ainsi l'estimation (1) n'est utilisable que si A admet une B-minorante. Comme nous manquons d'une caractérisation des matrices ayant cette propriété, nous supposerons, dorénavant que la minorante N utilisée est une Z-minorante (diagonale positive, hors diagonale ≤ 0)

On sait (6-2-3) qu'une condition nécessaire et suffisante pour qu'il puisse en être ainsi est que la bloc diagonale Δ de A soit non singulière. On peut alors établir la

6-6-3 - Proposition *Si* A *admet une* Z *minorante* N *on a pour tout* $\psi \in \mathbb{R}_+^k$

$$\mathcal{D}(N,\psi) \supset \mathcal{D}(N(A),\ \psi) \qquad (2)$$

En effet, soit $N = D(I - U)$ une Z-minorante de A, de diagonale (positive) D. On a, d'après (6-2-5)

$$0 < D \leq m(\Delta)$$

$$M(J) \leq U$$

soit u appartenant à $\mathcal{D}(N(A),\psi)$;

$$N(A)u \leq \psi \qquad \text{soit}$$

$$m(\Delta)[I - M(J)]u \leq \psi$$ soit encore, puisque $m(\Delta)$ est une diagonale positive

$$[I-M(J)]u \leq [m(\Delta)]^{-1}\psi$$

d'où

$$[I-U]\psi \leq [I-M(J)]u \leq [m(\Delta)]^{-1}\psi \leq D^{-1}\psi \qquad \text{soit}$$

$$Nu = D(I - U)u \leq \psi$$

u appartient à $\mathcal{D}(N,\psi)$ et l'inclusion (2) est démontrée.

Alors dire que A admet une M-minorante c'est dire (Th 4 (6-2-6)) que N(A) est une M-matrice, ou encore que A est une bloc-H-matrice D'où la

6-6-4 - Proposition *Si tous les blocs diagonaux de* A *sont non-singuliers, la meilleure* Z*-minorante (au sens de l'inclusion des parties* $\mathcal{D}$*) que l'on puisse utiliser dans l'estimation (1) est* N(A).

Pour que cette estimation définisse, quel que soit $\rho \in \mathbb{K}^n$, *une partie* $\mathcal{D}(N(A),\ p(\rho))$ *bornée dans* $\mathbb{R}_+^k$ *il faut et il suffit que* A *soit une bloc-*H*-matrice.*

Remarques 1 - Ainsi se trouve caractérisée d'une autre façon la classe des blocs-H-matrices

2 - Si A est une bloc-H-matrice, on a l'inclusion suivante

$$\mathcal{D}\,(N(A),\ p(\rho)) \subset \{u \in \mathbb{R}_+^k\ ;\ u \leq [N(A)]^{-1}\ p(\rho)\}$$

ce qui montre que la partie $\mathcal{D}(N(A),\ p(\rho))$ considérée tend vers $\{0\}$ si ρ tend vers zéro.

3 - Bien entendu, si p est une norme sur $\mathbb{K}^n$ l'estimation (1) se ramène à la majoration bien connue

$$s_{pp}(A)\ p(\varepsilon) \leq p(\rho) \qquad \text{où}$$

$$s_{pp}(A) = \operatorname*{Inf}_{\substack{x \in \mathbb{K}^n \\ x \neq 0}} \left\{ \frac{p(Ax)}{p(x)} \right\}$$

estimation qui n'exige, pour être bornée, que la non singularité de A ($s_{pp}(A) > 0$)

4 - L'estimation (1) reste alors, dans sa généralité, assez académique. Donnons en, néanmoins un

<u>Exemple numérique</u> Soit le système linéaire (4,4)

$$\left|\begin{array}{cccc} 6 & 5 & 2.1 & 4.0 \\ 7 & 6 & 2.5 & 4.7 \\ \hdashline 9.6 & 4.7 & 9 & 10 \\ 8.6 & 4.2 & 8 & 9 \end{array}\right| \left|\begin{array}{c} \S_1 \\ \S_2 \\ \S_3 \\ \S_4 \end{array}\right| = \left|\begin{array}{c} -0.9 \\ -1.2 \\ 3.9 \\ 3.4 \end{array}\right| \qquad (3)$$

dont la solution exacte est $x = (1,\ -1,\ 1,\ -1)$. Prenons

$x^* = (0.9,\ -0.9,\ 0.9,\ -0.9)$ d'où

$\varepsilon = (-0.1,\ 0.1,\ -0.1,\ 0.1)$; $= (-0.81,\ -1.08,\ 3.51,\ 3.06)$

Soit p la norme vectorielle de taille 2 sur $\mathbb{R}^4$:

$$x = (\S_1,\ \S_2,\ \S_3,\ \S_4) \longrightarrow p(x) = (\mathrm{Max}\{|\S_1|, |\S_2|\}, \mathrm{Max}\{|\S_3|, |\S_4|\})$$

La matrice A du système linéaire (3) est, relativement à p, une bloc H-matrice (6-2-7). L'estimation d'erreur

$$N(A)\ p(\varepsilon) \leq p(\rho)$$

donne ici, en posant $p(\varepsilon) = \begin{vmatrix} \alpha \\ \beta \end{vmatrix}$

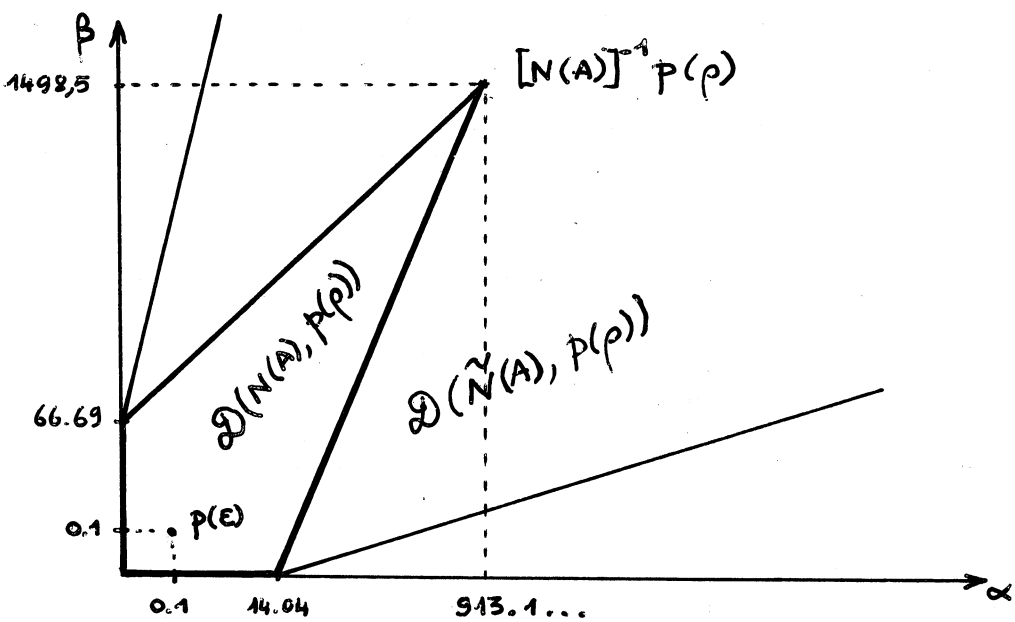
$$\begin{vmatrix} 1/_{13} & \\ & 1/_{19} \end{vmatrix} \times \begin{vmatrix} 1 & -0.6 \\ -0.8 & 1 \end{vmatrix} \begin{vmatrix} \alpha \\ \beta \end{vmatrix} \leqslant \begin{vmatrix} 1.08 \\ 3.51 \end{vmatrix}$$

Cependant que l'estimation :

$$\tilde{N}(A)p(\varepsilon) \leqslant p(\rho) \qquad \text{s'écrit}$$

$$\begin{vmatrix} 1/_{13} & -0.72 \\ -14.3 & 1/_{19} \end{vmatrix} \begin{vmatrix} \alpha \\ \beta \end{vmatrix} \leqslant \begin{vmatrix} 1.08 \\ 3.51 \end{vmatrix}$$

D'où, tous calculs faits, les localisations suivantes :



On vérifie que $p(\varepsilon)$ appartient à $\mathcal{D}(N(A), p(\rho))$, qui est contenue dans $\mathcal{D}(\tilde{N}(A), p(\rho))$.

Bien que définissant une localisation de $p(\varepsilon)$ numériquement désastreuse, $\mathcal{D}(N(A), p(\rho))$ est une partie bornée de $\mathbb{R}_+^2$. Or (6-2-7) $\tilde{N}(A)$ n'est pas une M-matrice : il se trouve que $\mathcal{D}(\tilde{N}(A), p(\rho))$ n'est pas borné.

CHAPITRE 7 : UN THEOREME de CONVERGENCE des METHODES ITERATIVES CLASSIQUES par POINT ou BLOCS

L'objet de ce chapitre est d'utiliser des normes vectorielles pour l'élaboration d'un résultat assez général assurant la convergence des méthodes itératives classiques (Jacobi, Gauss Seidel, sur-relaxation), par point ou par blocs, associées à un système linéaire donné. Pour l'établir, nous mettons préalablement au point la version du théorème de Stein Rosenberg qui nous est nécessaire, un peu plus générale que la version habituelle, telle qu'elle est donnée par exemple dans Varga ([64] p. 73).

Le théorème de convergence obtenu (Th 5; (7-1-2)) assure, en fait, sous une condition numériquement accessible, que les matrices d'itération considérées sont contractantes relativement à la norme vectorielle p utilisée d'où, outre la convergence, la possibilité de mise en oeuvre de l'algorithme de Schroeder de contrôle de l'approximation obtenue en cours d'itération.

Dans le cas où A est, relativement à p, une bloc-H-matrice, le théorème 5 prend une forme particulière (Th 6 (7-2-1)) assurant la convergence des méthodes par blocs correspondantes de Jacobi et Gauss Seidel : ainsi le théorème 6 constitue-t-il l'extension, au cas des blocs du résultat classique, rappelé en (6-3-2) assurant la convergence des méthodes de Jacobi et Gauss Seidel par point lorsque la matrice du système linéaire est une H-matrice. C'est d'ailleurs ce résultat qui justifie effectivement la désignation de bloc-H-matrice que nous avons adoptée. Nous donnons un résultat supplémentaire relatif à la convergence de la sur-relaxation correspondante.

En utilisant la notion de décompositions plus ou moins fines, le paragraphe 7-3 développe différentes formes particulières du théorème 5, donnant des conditions suffisantes de convergence dont les unes semblent nouvelles, les autres étant connues (mais démontrées différemment).

Enfin, le paragraphe 7-4 constitue un exemple d'application numérique, traité in extenso, des résultats qui viennent d'être établis : la convergence de la relaxation simple (pour le problème du potentiel discrétisé classiquement) étant obtenue par l'application du théorème 5, on met en oeuvre l'algorithme de Schroeder correspondant, de contrôle de l'approximation obtenue, selon une "technique de grille" analogue à celle qui est utilisée pour l'itération principale et par conséquent bien adaptée à un calcul automatique.

7 - 1 - THEOREME :

$\mathbb{K}$ désigne $\mathbb{R}$ ou $\mathbb{C}$. Soit A une matrice (n,n) à éléments dans $\mathbb{K}$, b un vecteur donné de $\mathbb{K}^n$, d'où le système linéaire

$$Ax = b \qquad (1)$$

de solution inconnue x. Pour le résoudre, on se donne une matrice B de type (n,n), non singulière (numériquement facilement inversible) et l'on pose

$$T = I - BA \qquad (2)$$

d'où l'itération linéaire dans $\mathbb{K}^n$:

$$x_{r+1} = Tx_r + Bb \qquad (3)$$

Cette itération sera convergente à la condition (nécessaire et suffisante) que l'on ait pu choisir B de façon que T = I - BA soit de rayon spectral inférieur à 1. La limite x^* de cette itération vérifie la relation

$$x^* = Tx^* + Bb = (I - BA)x^* + Bb$$

soit

$$B\,Ax^* = Bb$$

soit encore, puisque B est non singulière

$$Ax^* = b$$

x est donc solution de (1)

Dans la terminologie de Varga [64] (3) n'est pas autre chose que la méthode itérative associée au "splitting" suivant de A

$$A = (B^{-1}) - (B^{-1} - A)$$

Nous ferons plusieurs choix pour B

1) si A est non singulière on peut poser

$$B = A^{-1}$$

Ce choix, parfaitement théorique, donne la matrice nulle pour matrice d'itération. La formule d'itération est remplacée par la forme explicite de la solution

$$x = A^{-1}b$$

(pour mémoire)

2) Soit Δ une matrice non singulière connue. Nous posons

$$B = \Delta^{-1}$$

Nous noterons J la matrice d'itération correspondante

$$J = I - \Delta^{-1}A \qquad (4)$$

et (J) le processus itératif associé :

$$(J) \qquad x_{r+1} = Jx_r + \Delta^{-1}b \qquad (5)$$

3) Soit L une matrice n'ayant pas de valeur propre égale à 1 : I - L est non singulière. Nous posons

$$U = J - L$$

d'où le choix suivant pour la matrice B

$$B = (I-L)^{-1}\Delta^{-1}$$

nous noterons $\mathcal{L}_1$ la matrice d'itération correspondante

$$\mathcal{L}_1 = I - BA = (I - L)^{-1}[I - L - \Delta^{-1}\Delta(I - L - U)] = (I - L)^{-1}U \qquad (6)$$

et (G.S) le processus itératif associé

$$(G.S) \qquad x_{r+1} = \mathcal{L}_1 x_r + (I - L)^{-1}\Delta^{-1}b \qquad (7)$$

4) Enfin soit ω un nombre réel. On peut poser, si $(I - \omega L)$ est non singulière

$$B = \omega(I - \omega L)^{-1}\Delta^{-1}$$

nous noterons alors $\mathcal{L}_\omega$ la matrice d'itération correspondante

$$\mathcal{L}_\omega = I - \omega(I - \omega L)^{-1}\Delta^{-1}\Delta(I - L - U) = (I - \omega L)^{-1} (\omega U + (1 - \omega)I) \qquad (8)$$

et (S.R) le processus itératif associé.

$$(S.R) \qquad x_{r+1} = \mathcal{L}_\omega x_r + \omega(I - \omega L)^{-1}\Delta^{-1}b \qquad (9)$$

Remarques : 1) Auchangement de notation $\Delta^{-1} = B$ près, le procédé (J) est aussi général que celui défini par l'itération (3)

2) Dans le cas particulier où Δ est la matrice bloc-diagonale (supposée non singulière) de A pour une décomposition en blocs donnée (blocs diagonaux carrés) J est alors la matrice classique de Jacobi par blocs associée à A (cf 6-2). Si de plus L est la bloc-triangulaire inférieure de J, U est alors

la bloc triangulaire supérieure de J : L et U sont de bloc diagonale nulle, I - ωL est pour tout ω, une bloc-triangulaire inférieure de bloc-diagonale unité : elle est inversible et (S.R) est alors le procédé classique de sur-relaxation par blocs associé au système linéaire (1). Pour $\omega = 1$, on retrouve l'itération (G.S) de Gauss-Seidel par blocs définie par la matrice d'itération $\mathcal{L}_1$.

Enfin, si tous les blocs sont (1,1), on retrouve les méthodes de Jacobi, Gauss Seidel et sur-relaxation par point associées au système linéaire (1)

Revenons à la formulation générale des procédés (J) (G-S) et (S.R). Nous allons établir, au Théorème 5, un résultat prouvant, sous une condition simple ne portant que sur J (en réalité sur L et U), que les matrices J, $\mathcal{L}_1$ et $\mathcal{L}_\omega$ sont contractantes relativement à une norme vectorielle p donnée sur $\mathbb{K}^n$.

On en déduira (cf chapitre 5) la convergence des méthodes évoquées et la possibilité de leur appliquer l'algorithme de Schroeder de contrôle des approximations obtenues.

Pour établir ce résultat, nous allons préalablement donner une formulation un peu particulière du théorème de Stein-Rosenberg :

R.S. Varga dpnne, dans [64] p.89 un résultat qui peut s'énoncer ainsi :

Théorème : *Soient* F *et* G *deux matrices* (n,n) *telles que :* F^{-1} *existe et est* ≥ 0; $G \geq 0$. *Posons* $H = F - G$. *Si* H^{-1} *existe et est non-négative, alors*

$$\rho(F^{-1}G) = \frac{\rho(H^{-1}G)}{1+\rho(H^{-1}G)} < 1$$

D'où le

Corollaire : *Soient* F *et* G *deux matrices* (n,n) *telles que :* F^{-1} *existe et est* ≥ 0; $G \geq 0$. *Posons* $H = F - G$, *les propositions suivantes sont équivalentes :*

(1) H^{-1} existe et est ≥ 0

$$(2) \quad \rho(F^{-1}G) < 1$$

En effet : (1) implique (2) : c'est le théorème précédent; l'implication inverse est triviale : si $\rho(F^{-1}G) < 1$, $(I - F^{-1}G)^{-1}$ existe et s'écrit

$$(I - F^{-1}G)^{-1} = I + (F^{-1}G) + \ldots\ldots (F^{-1}G)^r + \ldots\ldots \geq 0$$

D'où $$(I - F^{-1}G)^{-1} \geq 0$$

ce qui prouve que $H^{-1} = (F - G)^{-1} = (I - F^{-1}G)^{-1}F^{-1}$ existe et est ≥ 0

D'où la formulation suivante, donnée uniquement dans un but utilitaire du

7-1-1 - Théorème de Stein-Rosenberg

Soient Q *et* R *deux matrices non négatives. Les deux propositions suivantes sont équivalentes :*

$$(1) \quad \rho(R + Q) < 1$$

$$(2) \quad \rho(R) < 1 \text{ et } \rho([I - R]^{-1}Q) < 1$$

Si elles sont vérifiées on a alors :

$$\rho[(I - R)^{-1}Q] \leq \rho(R + Q) < 1$$

(La formulation classique suppose R et Q triangulaires inférieures et supérieures respectivement et strictement, de sorte que R + Q est supposée de diagonale nulle. Nous ne faisons aucune telle restriction ici).

α) Chacune des deux propositions (1) et (2) contient la condition $\rho(R) < 1$: (2) explicitement; quant à (1) :

$$\rho(R) \leq \rho(R + Q) < 1$$

Par conséquent $(I - R)^{-1}$ existe et est non négative :

$$(I - R)^{-1} = I + R + \ldots\ldots R^r + \ldots \geq 0$$

β) Alors, par application double du corollaire précédent

$$\rho\left[(I - R)^{-1}Q\right] < 1 \underset{\substack{F=I-R \\ G=Q}}{\iff} [I - R - Q]^{-1} \text{ existe et est } \geq 0 \underset{\substack{F=I \\ G=R+Q}}{\iff} \rho(R + Q) < 1$$

L'équivalence est démontrée.

γ) on suppose ces conditions réalisées. Le théorème de Varga cité plus haut permet d'écrire

$$\rho\left[(I - R)^{-1}Q\right] = \frac{\rho[(I - R - Q)^{-1}Q]}{1+\rho[(I - R - Q)^{-1}Q]} \; ; \; \rho(R + Q) = \frac{\rho[(I - R - Q)^{-1}(R+Q)]}{1+\rho[(I - R - Q)^{-1}(R+Q)]}$$

or

$0 < [I - R - Q]^{-1}Q \leqslant [I - R - Q]^{-1}(R + Q)$, d'où la même inégalité sur les rayons spectraux de ces matrices non négatives :

$$a \underset{df}{=} \rho\left([I - R - Q]^{-1}Q\right) \leqslant \rho\left([I - R - Q]^{-1}(R + Q)\right) \underset{df}{=} b$$

La fonction $\frac{x}{1+x}$ étant non décroissante pour $x \geqslant 0$, on en déduit l'inégalité attendue :

$$\frac{a}{1+a} = \rho\left[(I - R)^{-1}Q\right] \leqslant \rho(R + Q) = \frac{b}{1+b}$$

Nous sommes alors en mesure d'établir le résultat général suivant

7-1-2 - Théorème 5 *Soit* p *une norme vectorielle régulière de taille* k *sur* $\mathbb{K}^n$ *et* $\mathcal{M}$ *la norme vectorielle de matrices qu'elle génère. Soient* L *et* U *deux matrices* (n,n). *Posons*

$$\rho \underset{df}{=} \rho\left(\mathcal{M}(L) + \mathcal{M}(U)\right)$$

Si $\rho<1$ *alors, relativement à* p,

* $J = L + U$ *est contractante, car elle admet la majorante convergente* $\mathcal{M}(L) + \mathcal{M}(U)$

* $\mathcal{L}_1 = (I - L)^{-1}U$ *existe et est contractante, car elle admet pour majorante la matrice* $[I-\mathcal{M}(L)]^{-1}\mathcal{M}(U)$ *qui est convergente*

* *Plus précisément, la matrice "de sur-relaxation"*

$$\mathcal{L}_\omega = (I - \omega L)^{-1}(\omega U + (1-\omega)I)$$

existe et est contractante pour tous les ω *tels que*

$$0 < \omega < \frac{2}{1+\rho} \qquad (*)$$

cette dernière condition définissant en fait le lieu des ω *tels que la matrice*

(*) intervalle qui, puisque ρ est inférieur à l'unité, contient $]0,1]$ et est contenu dans $]0,2[$

$$\overline{\mathscr{L}_\omega} \underset{df}{=} [I - |\omega| M(L)]^{-1} [|\omega| M(U) + |1-\omega| I]$$

qui est une majorante de $\mathscr{L}_\omega$*, soit convergente.*

(1) Le premier point se justifie aisément :

$$M(J) = M(L + U) \leqslant M(L) + M(U)$$

d'où :

$$\rho\big(M(J)\big) \leqslant \rho\big(M(L) + M(U)\big) = \rho < 1$$

J, par conséquent, est contractante relativement à p.

(2) Avec l'hypothèse $\rho < 1$, il est clair que l'intervalle $]0 \; \frac{2}{1+\rho}[$ contient l'unité : le second point du théorème est cas particulier du troisième, que nous allons démontrer maintenant :

(3) Montrons qu'avec les hypothèses :

$$\rho < 1 \quad ; \quad 0 < \omega < \frac{2}{1+\rho}$$

la matrice ωL est contractante. En effet

$$\rho\big(M(\omega L)\big) \leqslant \rho\big(M(\omega L) + M(\omega U)\big) = |\omega|\rho < \frac{2\rho}{1+\rho} < 1$$

(4) Alors, par application du Lemme (5-2-2), on en déduit

α) l'existence de $(I - \omega L)^{-1}$, donc de $\mathscr{L}_\omega$

β) l'inégalité

$$M[(I-\omega L)^{-1}] \leqslant [I - M(\omega L)]^{-1} = [I - |\omega| M(L)]^{-1}$$

D'où

$$M(\mathscr{L}_\omega) = M[(I-\omega L)^{-1}(\omega U + (1-\omega)I)] \leqslant M[(I-\omega L)^{-1}] M(\omega U + (1-\omega)I)$$
$$\leqslant [I - |\omega| M(L)]^{-1} [|\omega| M(U) + |1-\omega| I] \underset{df}{=} \overline{\mathscr{L}_\omega}$$

Ainsi $\overline{\mathscr{L}_\omega}$ est bien une majorante de $\mathscr{L}_\omega$ relativement à p.

Posons $R = M(\omega L) = |\omega| \, M(L)$

$Q = |\omega| M(U) + |1-\omega| I$

Alors, d'après la formulation que nous avons donnée du théorème de Stein-Rosenberg, dire que $\overline{\mathscr{L}_\omega}$ est convergente, c'est donc dire que :

$$\rho(R+Q) = \rho\Big(|\omega| [M(L) + M(U)] + |1-\omega| I\Big) < 1$$

c'est-à-dire

$$|\omega|\rho + |1-\omega| < 1$$

et il est facile de vérifier que cette dernière condition est équivalente à

$$\begin{cases} \rho < 1 \\ 0 < \omega < \dfrac{2}{1+\rho} \end{cases}$$

Ce sont bien les hypothèses que nous avons adoptées.

7-1-3 - *Remarques :*

1) Le résultat reste bien entendu valide lorsque est une norme habituelle sur $\mathbb{K}^n$

2) En explicitant une majorante convergente pour diverses matrices d'itération, le théorème précédent non seulement permet de prouver la convergence des méthodes itératives correspondantes mais de plus autorise une mise en oeuvre effective de l'algorithme de Schroeder destiné à contrôler l'approximation obtenue au cours de l'itération (cf 5-3). Nous traiterons d'ailleurs in extenso au paragraphe (7-4) un exemple numérique illustrant ceci

3) Soit S une norme vectorielle régulière, sous multiplicative, de taille k^2 sur l'algèbre $\mathcal{M}_{nn}$ des matrices carrées (n,n) (cf 4-2-1). Soient L et U deux matrices (n,n) telles que

$$\rho\big(S(L) + S(U)\big) < 1$$

D'après (4-2-4) il existe sur $\mathbb{K}^n$ une norme vectorielle p de taille k, régulière et générant M sur $\mathcal{M}_{nn}$ telle que

$$\forall B \in \mathcal{M}_{nn} \qquad M(B) \leq S(B)$$

d'où

$$M(L) + M(U) \leq S(L) + S(U)$$

et la même inégalité sur les rayons spectraux de ces matrices non négatives.

$$\rho = \rho\big(M(L) + M(U)\big) \leq \rho\big(S(L) + S(U)\big) < 1$$

Ainsi L et U vérifient, relativement à p, les hypothèses du théorème précédent = Même si on ne peut que difficilement expliciter la norme vectorielle ρ (relativement à laquelle les matrices $J = L + U$, $\mathcal{L}_1$ et $\mathcal{L}_\omega$ $\left(0 < \omega < \frac{2}{1+\rho}\right)$ sont contractantes) on est néanmoins assuré, grâce au théorème 5, de la convergence de ces matrices et par conséquent de celle des méthodes itératives correspondantes.

Avant de montrer quelles formes particulières peut prendre le théorème 5, donnons en un exemple type d'utilisation

7-1-4 - Exemple type : Méthodes par point pour la matrice du potentiel

La matrice de Jacobi par point relative à la matrice dite du potentiel (cf 6-4-2) s'écrit

$$J = \frac{1}{4}\begin{vmatrix} E+E^t & I & & \\ I & E+E^t & I & \\ & \ddots & \ddots & I \\ & & I & E+E^t \end{vmatrix}_{(\lambda k,\lambda k)}$$

où I est la matrice unité (λ,λ) et

$$E = \begin{vmatrix} 0 & & & \\ 1 & \ddots & & \\ & \ddots & \ddots & \\ & & 1 & 0 \end{vmatrix}_{(\lambda,\lambda)}$$

Prenons pour L la triangulaire inférieure de J, d'où $U = L^t$ la triangulaire supérieure de J

$$L = \frac{1}{4}\begin{vmatrix} E & 0 & & \\ I & E & \ddots & \\ & \ddots & \ddots & 0 \\ & & I & E \end{vmatrix}_{(\lambda k,\lambda k)} \qquad U = L^t$$

D'où selon les blocs mis en évidence :

$$M_\infty(L) = \frac{1}{4}\begin{vmatrix} 1 & 0 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 0 \\ & & 1 & 1 \end{vmatrix}_{(k,k)} \qquad \text{et } M_\infty(U) = [M_\infty(L)]^t$$

D'où

$$M_\infty(L) + M_\infty(U) = \frac{1}{4}\begin{vmatrix} 2 & 1 & & \\ 1 & 2 & 1 & \\ & \ddots & \ddots & 1 \\ & & 1 & 2 \end{vmatrix}_{(k,k)}$$

Il vient

$$\rho \underset{df}{=} \rho\left(M_\infty(L)+M_\infty(U)\right) = \frac{1}{4}\left[2+2\cos\left(\frac{\pi}{k+1}\right)\right] = \frac{1+\cos\left(\frac{\pi}{k+1}\right)}{2} < 1$$

D'où

1°/ la convergence des méthodes de Jacobi et Gauss-Seidel par point pour la matrice du potentiel (qui n'est pas à diagonale dominante)

2°/ la convergence de la sur-relaxation par point correspondante pour tous les ω pris dans la plage

$$0 < \omega < \frac{2}{1+\rho} = \frac{4}{3+\cos\left(\frac{\pi}{k+1}\right)}$$

Cet exemple est typique de l'utilisation effective de telles techniques numériques. Voir 7-4.

Revenons au théorème 5 (7-1-2) et montrons quelles formes il peut prendre dans le cas des méthodes classiques de Jacobi, Gauss-Seidel et de sur-relaxation par blocs : Δ est la bloc-diagonale (supposée non singulière) de A pour une décomposition donnée en blocs (blocs-diagonaux carrés); L et U sont les blocs triangulaires respectivement inférieures et supérieures de $J = I - \Delta^{-1}A$. Le théorème 5 admet alors les résultats suivants (qui peuvent éventuellement être démontrés directement) comme cas particuliers.

7 - 2 - CAS des BLOCS-H-MATRICES

Il est clair que, dans le cas où la décomposition en blocs des méthodes itératives considérées coïncide avec celle définie par la norme vectorielle M, on a

$$M(J) = M(L + U) = M(L) + M(U) \qquad (10)$$

et la condition du théorème 5 s'écrit ici

$$\rho\big(M(J)\big) < 1$$

ce qui revient à dire (6-4-1) que A <u>est une bloc H-matrice relativement à p.</u> d'où le

7-2-1 - Théorème **6.** *Soit* p *une norme vectorielle régulière de taille* k *sur* $\mathbb{K}^n$, *générant la norme vectorielle* M *de matrice.*

Si A *est, relativement à* p, *une bloc-H-matrice, les matrices de Jacobi, Gauss-Seidel et de sur-relaxation* ($0 < \omega < \dfrac{2}{1+\rho\big(M(J)\big)}$) *selon les blocs correspondants sont contractantes relativement à* p, *et les méthodes itératives correspondantes pour résoudre un système linéaire* $Ax = b$ *convergent vers la solution unique* $x = A^{-1}b$

En effet, puisque A est une bloc-H-matrice, elle est non singulière (6-4-3) et le système linéaire $Ax = b$ a pour solution unique $x = A^{-1}b$. Le théorème précédent assure effectivement que les matrices d'itération par blocs considérées sont contractantes; les méthodes itératives correspondantes sont convergentes et ne peuvent converger que vers la solution unique $x = A^{-1}b$.

Remarques 1) Avec la terminologie employée ici, on peut énoncer de la façon suivante un résultat démontré dans [41] :
"Si $\tilde{N}(A)$ est une M-matrice, le procédé par blocs de Jacobi correspondant est convergent"

D'après le théorème 6 ci-dessus on a de plus la convergence de la méthode par blocs de Gauss-Seidel et de sur-relaxation $\left(0<\omega<\dfrac{2}{1+\rho\big(M(J)\big)}\right)$

sous l'hypothèse <u>moins restrictive</u> (6-2-7) que N(A) est une M-matrice, i.e que A est une bloc-H-matrice.

2) Comme cas particuliers du théorème 6 on a celui où A est une bloc-G-matrice ou celui, encore plus restrictif, où A est à bloc-diagonale dominante (cf 6-4-1)

3) Considérons le cas des méthodes par point (tous les blocs sont (1,1)).En utilisant la norme vectorielle type $p(x) = |x|$ le théorème 6 redonne en particulier le résultat classique, cité en (6-3-2), que nous rappelons ici :

<u>Si A est une H-matrice, les méthodes de Jacobi et Gauss-Seidel par point correspondantes pour résoudre un système linéaire $Ax = b$ convergent vers la solution unique $x = A^{-1}b$</u>

Le théorème 6 précise de plus que la sur-relaxation par point converge pour tous les ω pris dans plage

$$0 < \omega < \frac{2}{1+\rho(|J|)}$$

Ainsi le théorème 6 constitue l'extension aux méthodes par blocs du résultat classique cité ci-dessus. Il illustre d'une nouvelle façon la classe des blocs-H-matrices.

<u>7-2-2 - Exemples numériques</u>

1) L'exemple 1) de 5-2-4 met en évidence une matrice T de Jacobi par blocs contractante. D'où, par application du théorème 6, outre la convergence de T, celle de la matrice de sur-relaxation par bloc associée, pour les ω pris dans la plage

$$0 < \omega < \frac{2}{1+\rho\big(M_\infty(T)\big)}$$

2) <u>Exemple-type</u> : On a vu en (6-4-2) que la matrice du potentiel

est une bloc-H-matrice relativement à une norme vectorielle conduisant à

$$M(J) = \begin{vmatrix} 0 & \gamma & & & \\ \gamma & 0 & \gamma & & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & \gamma \\ & & & \gamma & 0 \end{vmatrix} \quad (k,k)$$

où γ peut être aussi près que l'on veut (par excès) de la quantité

$$\rho(H^{-1}) = \frac{1}{4-2\cos\left(\frac{\pi}{\lambda+1}\right)}$$

D'où

$$\rho(M(J)) = 2\,\gamma\,\cos\left(\frac{\pi}{k+1}\right) \qquad \text{et}$$

$$\frac{2}{1+\rho(M(J))} \lesssim \frac{2}{1+\dfrac{2\cos\left(\frac{\pi}{k+1}\right)}{4-2\cos\left(\frac{\pi}{\lambda+1}\right)}} = \frac{2\left(2-\cos\left(\frac{\pi}{\lambda+1}\right)\right)}{2-\cos\left(\frac{\pi}{\lambda+1}\right)+\cos\left(\frac{\pi}{k+1}\right)}$$

Finalement pour la matrice du potentiel décomposé en k^2 blocs (λ,λ) on peut conclure à la convergence de la méthode de Jacobi par blocs et à celle de la sur-relaxation par blocs correspondante pour tous les ω pris dans la plage

$$0 < \omega < \frac{2\left(2-\cos\left(\frac{\pi}{\lambda+1}\right)\right)}{2-\cos\left(\frac{\pi}{\lambda+1}\right)+\cos\left(\frac{\pi}{k+1}\right)} \underset{df}{=} \omega_M$$

Remarques $1 < \omega_M < 2$

$$\lim_{k=\lambda\to\infty} (\omega_M) = 1$$

7-2-3 - *Remarque* : Soit A une matrice (n,n) décomposée en k^2 blocs A_{ij}. Les blocs diagonaux A_{ii} sont supposés carrés (de type $(\lambda i,\lambda i)$; $\sum_{i=1}^{k} \lambda i = n$) et non singuliers, d'où J la matrice de Jacobi par blocs correspondante. Désignons par $\phi_\infty(A_{ii}^{-1}\ A_{ij})$ le module maximum des éléments de la matrice $A_{ii}^{-1}\ A_{ij}$ (de type $(\lambda i, \lambda j)$). Le point 3) du paragraphe (7-1-3) peut prendre ici la forme particulière suivante : si la matrice (k,k) :

$$S(J) = \begin{vmatrix} 0 & \sqrt{\lambda_1\lambda_2}\ \phi_\infty(A_{11}^{-1}\ A_{12}) & \cdots\cdots & \sqrt{\lambda_1\lambda_k}\ \phi_\infty(A_{11}^{-1}\ A_{1k}) \\ & & & \\ & & & \\ \sqrt{\lambda_k\lambda_1}\ \phi_\infty(A_{kk}^{-1}\ A_{k1}); & \sqrt{\lambda_k\lambda_2}\phi_\infty(A_{kk}^{-1}\ A_{k2}) & \cdots\cdots\cdots\cdots & 0 \end{vmatrix}$$

est de rayon spectral inférieur à l'unité, les procédés itératifs par blocs de Jacobi et Gauss Seidel correspondants pour résoudre un système linéaire de matrice A sont convergents.

En effet S étant une norme vectorielle régulière et sous-multiplicative (4-2-3), on sait exister une norme vectorielle p régulière sur $\mathbb{K}^n$ relativement à laquelle A est une bloc-H-matrice. (6-4-1, *Rem.*) D'où le résultat par application du théorème 6 précédent.

Comme exemple numérique, se reporter au point 2) du paragraphe (5-2-4) où l'on prouve que la matrice de Jacobi par blocs T' qui y est donnée vérifie

$$\rho\ S(T') < 1$$

D'où, outre la convergence de T', celle de la matrice de Gauss-Seidel par blocs associée
(On pourrait d'ailleurs prouver un peu plus : à savoir la convergence de la matrice de sur-relaxation associée à T', pour tous les ω pris dans la plage

$$0 < \omega < \frac{2}{1+0.2\sqrt{6}+0.5} = \frac{2}{1,99..}$$

Remarque : Au théorème 6, la décomposition en blocs des méthodes itératives étudiées est supposée coïncider avec celle sur laquelle est construite la norme vectorielle utilisée, ce qui nous assure de la validité de la relation

$$M(L + U) = M(L) + M(U)$$

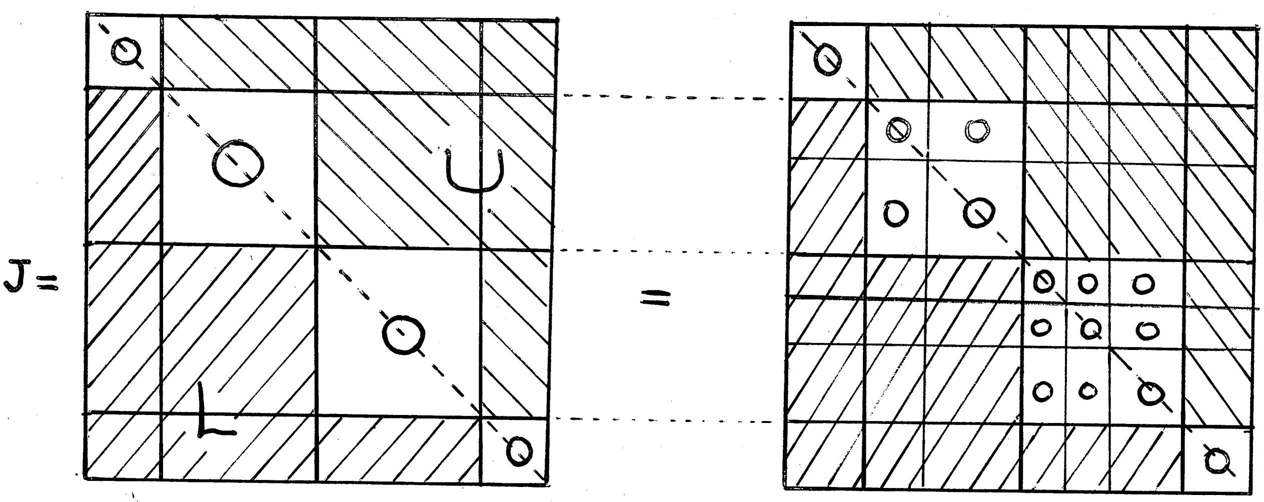
Nous allons voir que cette relation peut être conservée dans un cadre plus général

7 - 3 - AUTRES FORMES du THEOREME 5

De façon générale, nous noterons $\mathcal{D}_J$ la décomposition en blocs des procédés itératifs étudiés et $\mathcal{D}_M$ la décomposition en blocs relative à la norme vectorielle M de matrice utilisée (rappelons que M est supposée issue d'une norme vectorielle régulière sur $\mathbb{K}^n$)

7 -3 - 1 - Définition.

Nous dirons que $\mathcal{D}_M$ est <u>plus fine</u> que $\mathcal{D}_J$ si $\mathcal{D}_M$ peut être obtenue à partir de $\mathcal{D}_J$ en décomposant à leur tour les blocs de la décomposition $\mathcal{D}_J$ (tout en conservant, pour les blocs diagonaux, la propriété d'être carrés). Illustrons ceci par un schéma :



J et la décomposition $\mathcal{D}_J$ qu'elle définit

La même matrice J, décomposée selon une décomposition en blocs $\mathcal{D}_M$ plus fine que $\mathcal{D}_J$

Il est clair que si $\mathcal{D}_M$ est plus fine que $\mathcal{D}_J$, alors la propriété

$$M(L + U) = M(L) + M(U)$$

reste vraie. D'où par application du théorème 5, le

7-3-2- Corollaire 1 *Soit* $J = L + U$ *une matrice de Jacobi par blocs (matrice de bloc-diagonale nulle) définissant la décomposition en blocs* $\mathcal{D}_J$

Soit M *une norme vectorielle de matrice issue d'une norme vectorielle* p *régulière sur* $\mathbb{K}^n$ *et relative à une décomposition en blocs* $\mathcal{D}_M$ *plus fine que* $\mathcal{D}_J$.

Alors si J *est contractante relativement à* $p(\rho \underset{df}{=} \rho\left(M(J)\right) < 1)$ *on a les mêmes conclusions qu'au théorème 5 (7-1-2).*

Appliquons le corollaire précédent en utilisant la norme vectorielle type $p(x) = |x|$ qui génère la norme vectorielle de matrice $M(A) = |A|$

Cette norme vectorielle correspond à la décomposition $\mathcal{D}_M$ <u>la plus fine possible</u> (tous les blocs sont (1,1). Elle est donc plus fine que toute décomposition $\mathcal{D}_J$ d'où le résultat suivant, d'ailleurs connu partiellement.

7-3-3- Corollaire 2 *Soit* $J = L + U$ *une matrice de Jacobi par point ou blocs. Si*

$$\rho(|J|) < 1$$

on a les mêmes conclusions qu'au théorème 5.

(pour la norme vectorielle $M(.) = |.|$)

Remarque : Ainsi que nous l'avons déjà noté, le théorème 5 et ses diverses formulations particulières, en explicitant une majorante convergente pour diverses matrices d'itération non seulement permettent de prouver la convergence de méthodes itératives classiques (Jacobi, Gauss-Seidel, sur-relaxation) dans des cas d'utilisation pratique, mais de plus autorisent une mise en oeuvre effective de l'algorithme de Schroeder destiné à contrôler l'approximation obtenue au cours de l'itération :

Par exemple nous savons (Th 5) que la condition

$$\rho\big(M(L) + M(U)\big) < 1$$

entraîne, outre la convergence de la matrice $\mathcal{L}_1 = (I - L)^{-1}U$ que l'on a

$$M(\mathcal{L}_1) \leq [I - M(L)]^{-1}M(U) \underset{df}{=} \overline{\mathcal{L}}_1$$

$$\rho(\overline{\mathcal{L}}_1) < 1$$

On peut donc prendre $\overline{\mathcal{L}}_1$ pour matrice B dans la mise en oeuvre de la méthode de Schroeder parallèle à l'itération principale (de Gauss-Seidel) :

$$x_{r+1} = (I - L)^{-1}Ux_r + h.$$

Ce choix présente deux avantages :

1) l'itération d'erreur est parfaitement explicitée :

$$[I - M(L)]z_{r+1} = M(U)z_r$$

ce qui ne serait pas le cas si on voulait prendre $M\big((I-L)^{-1}U\big)$ pour matrice B : en effet, si L et U sont supposées connues, $\mathcal{L}_1 = (I-L)^{-1}U$ et à fortiori $M[(I - L)^{-1}U]$ doivent être considérées comme inaccessibles.

2) mais surtout, l'itération d'erreur est elle-même une itération de Gauss-Seidel (*)

Ainsi, dans des cas d'utilisation habituels (matrices de grande taille, très creuses, provenant de la dicrétisation de problèmes différentiels) la mise en oeuvre de l'itération d'erreur s'effectuera, de même que pour l'itération principale, selon une "technique de grille" bien adaptée à un calcul automatique. Nous traitons in-extenso, au paragraphe suivant, un exemple-type d'utilisation de ces techniques.

(*) J. Schroeder le signale : il vaut mieux utiliser, si elle est convergente, la matrice $(I-|L|)^{-1}|U|$ comme matrice d'itération d'erreur plutôt que la matrice $|(I-L)^{-1}U|$, plus difficilement accessible. Cette remarque [58] est dondée sur l'inégalité $|(I - L)^{-1}U| \leq (I - |L|)^{-1}|U|$, cas particulier, pour la norme vectorielle $M(.) = |.|$ utilisée, de l'inégalité

$$M[(I - L)^{-1}U] \leq [I - M(L)]^{-1} M(U)$$

7 - 4 - EXEMPLE-TYPE de la RELAXATION SIMPLE (METHODE de GAUSS-SEIDEL par POINT) pour le PROBLEME du POTENTIEL

L'équation de Laplace $\Delta u = 0$ à l'intérieur du carré unité, u étant donné sur les bords, conduit par une discrétisation à $\lambda.k$ points intérieurs à un système linéaire dont la matrice A est la matrice dite "du potentiel" déjà explicitée (cf 6-4-2)

$$A = \begin{vmatrix} H & I & & \\ I & H & I & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & I \\ & & & I & H \end{vmatrix} \quad (\lambda.k, \lambda k)$$

où I est la matrice unité de type (λ,λ) et

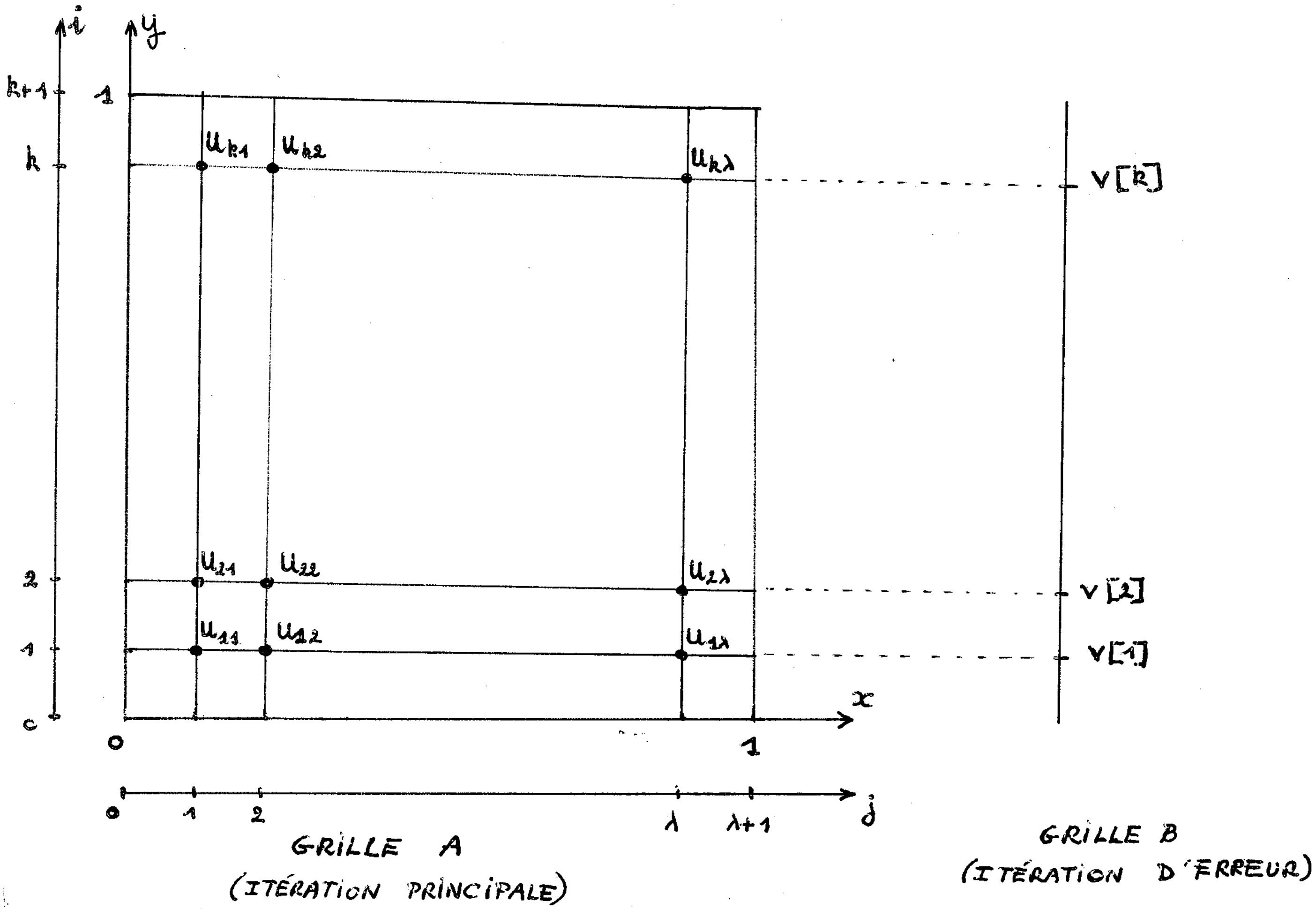
$$H = \begin{vmatrix} -4 & 1 & & \\ 1 & -4 & 1 & \\ & \ddots & \ddots & \ddots & \\ & & \ddots & \ddots & 1 \\ & & & 1 & -4 \end{vmatrix} \quad (\lambda,\lambda)$$

Soit J la matrice de Jacobi <u>par point</u> associée à **A**, L la triangulaire inférieure de J, U = J - L la triangulaire supérieure de J.
On a vérifié, au paragraphe (7-1-4), à titre d'exemple d'application du Théorème 5 (7-1-2) que la condition

$$\rho\left(M_\infty(L) + M_\infty(U)\right) < 1$$

était satisfaite. Ceci nous assure donc la convergence de la relaxation simple et autorise, d'après ce qui précède, la mise en oeuvre suivante de la méthode de Schroeder :

Considérons les deux grilles



Avec la norme vectorielle choisie, on obtiendra, à chaque pas de l'itération principale, une majoration (en norme du max) de l'erreur actuelle sur chacune des lignes de la grille (A). Sur la grille (B), V[1], V[2], V[k]portent ces majorations.

1) <u>itération principale</u>

Il est bien connu que la technique de grille (pour l'itération principale) est donnée par l'affectation algol

```
(A) u[i,j]:=(u[i-1,j]+ u[i+1,j]+u[i,j-1]+u[i,j+1])/4 ;
```

Pour chaque i (de 1 à k) on active l'instruction (A) pour tous les j de 1 à λ. Un balayage de toute la grille correspond à un pas de l'itération principale.

2) Choix de la matrice B d'itération d'erreur

D'après ce qui précède, on prend :

$$B = [I - M_\infty(L)]^{-1} M_\infty(U)$$

avec

$$M_\infty(L) = \frac{1}{4}\begin{vmatrix} 1 & 0 & & \\ 1 & \ddots & \ddots & \\ & \ddots & \ddots & 0 \\ & & 1 & 1 \end{vmatrix} \quad (k,k) \qquad\qquad (cf\ 7\text{-}1\text{-}4)$$

et $M_\infty(U) = [M_\infty(L)]^t$

3) Calcul du vecteur z_{r_o}

Au rang r_o (que nous fixons) de l'itération principale nous déterminons [cf(5-3-4)] z_{r_o} par la résolution du système linéaire de type (k,k)

$$(I - B)z_{r_o} = p(\rho_{r_o})$$

soit

$$[I - M_\infty(L) - M_\infty(U)]\,z_{r_o} = [I - M\ (L)]\,p(\rho_{r_o})$$

soit

$$\begin{vmatrix} 2 & -1 & & \\ -1 & 2 & \ddots & \\ & \ddots & \ddots & -1 \\ & & -1 & 2 \end{vmatrix} z_{r_o} = \begin{vmatrix} 3 & & & \\ -1 & \ddots & & \\ & \ddots & 3 & \\ & & -1 & \end{vmatrix} p(\rho_{r_o}) \quad (k,k)$$

système tridiagonal symétrique facile à résoudre. La seule contrainte est le calcul du vecteur $p(\rho_{r_o})$, norme vectorielle de la différence des "vecteurs grille A" au rang r_o+1 et r_o

4) Itération d'erreur : Elle sécrit matriciellement

$[I - M_\infty(L)]z_{r+1} = M_\infty(U)z_r$, soit

$$\begin{vmatrix} 3 & & \\ -1 & \ddots & \\ & \ddots & \ddots \\ & -1 & 3 \end{vmatrix} z_{r+1} = \begin{vmatrix} 1 & 1 & \\ & \ddots & \ddots \\ & & \ddots & 1 \\ & & & 1 \end{vmatrix} z_r \qquad (k,k)$$

V[1], ... V[k] sont les composantes du vecteur z_r d'itération d'erreur. D'où parallèlement à l'itération principale, une technique de grille pour l'itération d'erreur :

```
             V[1]:=  (V[1] + V[2])/3;
puis         V[2]:=  (V[1] + V[2] + V[3])/3;
         (B) v[i]:=  (v[i-1] + v[i] + v[i+1])/3;
           v[k-1]:=  (v[k-2]+ v[k-1]+ v[k])/3;
             v[k]:=  (v[k-1]+ v[k])/3;
```

Comme on l'a déjà vu, un balayage complet de la grille (A) correspond à un pas de l'itération principale. Un balayage complet de la grille B correspond à un pas de l'itération d'erreur. Les deux balayages sont conduits parallèlement : on active l'instruction (B) lorsque l'instruction (A) a été activée (pour le même i) pour tous les j de 1 à λ.

Si ε[i,j] désigne l'erreur (inconnue) au noeud (i,j) de la grille A, on obtiendra donc la majoration d'erreur suivante

$$V[i] \geq \operatorname*{Max}_{j=1}^{\lambda} \{ |\varepsilon[i,j]| \} \qquad i = 1,2 \;....\; k$$

<u>Résultats numériques</u> :

La solution exacte choisie est

$$u(x,y) = \cos x \operatorname{sh} y$$

ce qui nous permet de connaître l'erreur exacte (*)

(*) en fait la solution exacte du système linéaire ne coïncide pas avec la solution u = cosx shy du problème différentiel. Dans tous les exemples donnés la différence est inférieure à 10^{-5} : en limitant les tableaux de résultats à 4 décimales, nous supprimons cette distinction.

Pour différentes valeurs du couple (λ, k) on donne les majorations obtenues dans plusieurs cas : r_o fixé à priori ou obtenu par itération préliminaire; on vérifie que, à r_o égal, cette dernière technique donne des majorations moins fines. D'ailleurs, tant au point de vue de la finesse des majorations obtenues qu'à celui du temps de calcul nécessaire, la meilleure estimation que l'on puisse obtenir à un rang q donné de l'itération principale est de fixer à priori $r_o = q$. En effet, du fait de l'inégalité

$$\rho(T) \leqslant \rho(B)$$

où T est la matrice d'itération principale et B celle d'itération d'erreur, la finesse des majorations obtenues se dégrade au cours de l'itération d'erreur (cf (5-5-3)) λ = k = 10 (Grille à 100 points)

itération n°	ligne n°	$p(\varepsilon_r)$ norme vectorielle de l'erreur exacte	$r_0 = 60$ (fixé a priori)	$r_0 = 27$ (fixé a priori)	$r_0 = 27$ (obtenu par itération préliminaire)
27	1	0,0284		0,0538	0,1220 min
	2	0,0530		0,1015	0,2355
	3	0,0722		0,1396	0,3332
	4	0,0848		0,1654	0,4070
	5	0,0900 max		0,1771 max	0,4490
	6	0,0876		0,1741	0,4536 max
	7	0,0782		0,1571	0,4138
	8	0,0629		0,1280	0,3473
	9	0,0434		0,0897	0,2463
	10	0,0217 min		0,0457 min	0,1263
60	1	0,0019	0,0036	0,0142	0,0366
	2	0,0036	0,0068	0,0268	0,0689
	3	0,0048	0,0093	0,0367	0,0945
	4	0,0056	0,0110	0,0433	0,1115
	5	0,0058 max	0,0116 max	0,0462 max	0,1190 max
	6	0,0056	0,0113	0,0453	0,1166
	7	0,0049	0,0101	0,0408	0,1051
	8	0,0039	0,0082	0,0332	0,0855
	9	0,0027	0,0057	0,0232	0,0600
	10	0,0013 min	0,0029 min	0,0118 min	0,0306 min
80	1	0,0003	0,0015	0,0063	0,0162
	2	0,0006	0,0029	0,0118	0,0306
	3	0,0009	0,0041	0,0162	0,0419
	4	0,00105	0,0048	0,0192	0,0494
	5	0,00109 max	0,0051 max	0,0204 max	0,0527 max
	6	0,00104	0,0050	0,0200	0,0516
	7	0,0009	0,0045	0,0180	0,0465
	8	0,0007	0,0036	0,0147	0,0378
	9	0,0004	0,0025	0,0103	0,0265
	10	0,0002 min	0,0013 min	0,0052 min	0,0135 min

$\lambda = k = 20$ (Grille à 400 points)

itération n°	ligne n°	$p(\varepsilon_r)$ Norme vectorielle de l'erreur exacte max et min	$r_0 = 200$ (fixé à priori)	$r_0 = 100$ (fixé a priori)	$r_0 = 83$ (fixé a priori)	$r_0 = 83$ (obtenu par une itération préliminaire)
83	11 20	0,1217 0,0174			0,2423 max 0,0354 min	0,5670 max 0,0896
100	10 20	0,0833 0,0115		0,1659 max 0,0237 min	0,2005 max 0,1344 min	0,4639 0,0705
200	10 20	0,0088 0,0011	0,0176 max 0,0024 min	0,0542 max 0,0076 min	0,0656 max 0,0092 min	0,1525 max 0,0216 min
300	10 20	0,0009 0,0001	0,0057 max 0,0008 min	0,0177 max 0,0025 min	0,0214 max 0,0030 min	0,0498 max 0,0070 min

$\lambda = k = 25$ (Grille à 625 points)

itération n°	ligne n°	$p(\varepsilon_r)$ Norme vectorielle de l'erreur (max et min)	$r_0 = 220$ (fixé à priori)	$r_0 = 180$ (fixé a priori)	$r_0 = 122$ (fixé à priori)	$r_0 = 122$ (obtenu par une itération préliminaire)
122	13 25	0,1302 0,0154			0,2597 max 0,0311	0,5996 0,0786
180	13 25	0,0557 0,0062		0,1113 max 0,0129 min	0,1701 max 0,0198 min	0,3931 max 0,0470 min
220	13 25	0,0310 0,0034	0,0619 max 0,0071	0,0831 max 0,0096 min	0,1270 max 0,0147 min	0,2936 max 0,0344 min

$\lambda = k = 30$ (grille à 900 points)

N° itération	N° ligne	$p(E_r)$ norme vectorielle de l'erreur max et min	$r_0 = 169$ (fixé a priori)	$r_0 = 169$ (obtenu par une itération préliminaire)
169	16 1	0,1357 0,0133	0,2704 max 0,0268 min	0,6371 0,0547 min
200	16 30	0,0985 0,0097	0,2304 max 0,0230 min	0,5371 max 0,0568
300	15 30	0,0352 0,0033	0,1379 max 0,0135 min	0,3204 max 0,0318 min

$\lambda = k = 40$ (Grille à 1600 points)

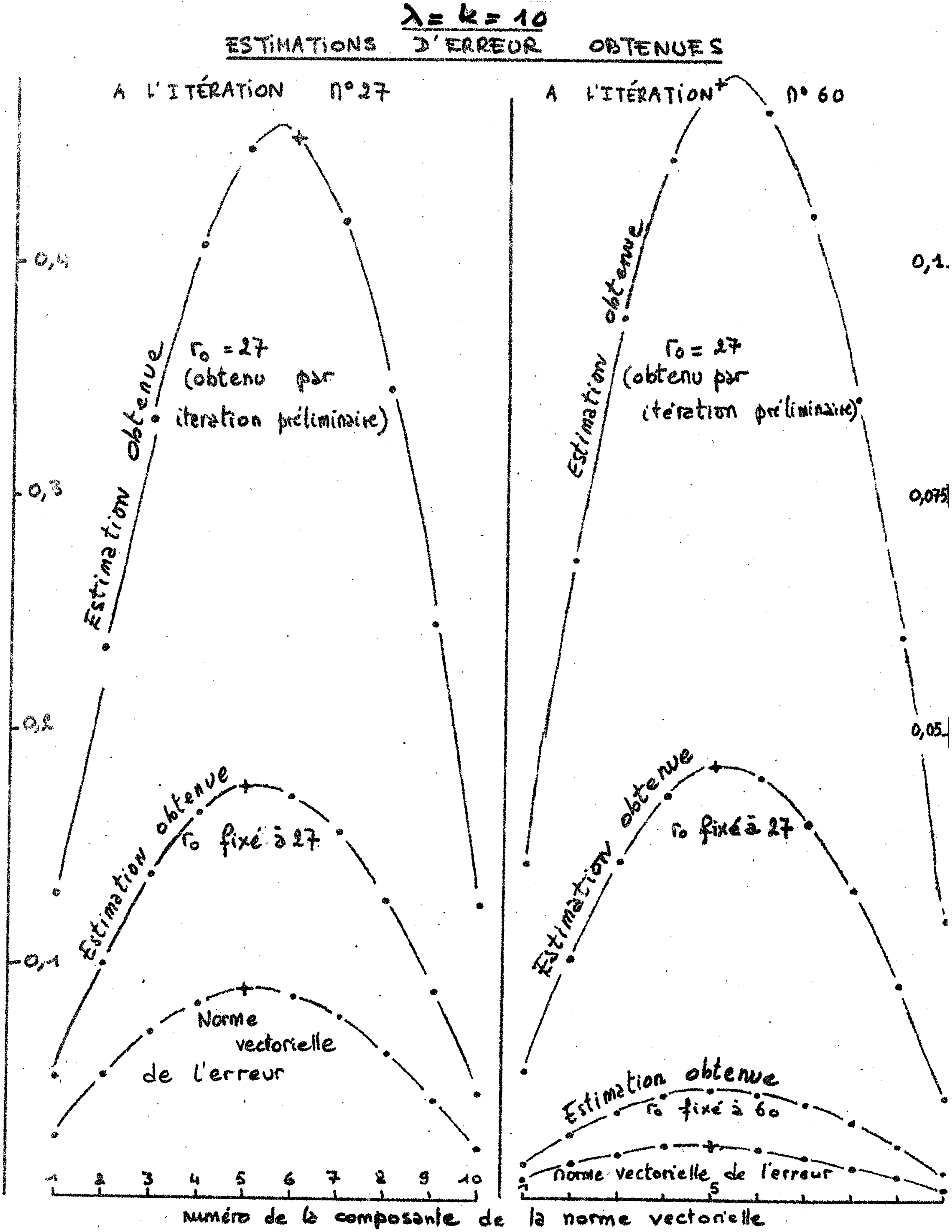
N° itération	N° ligne	$p(E_r)$ norme vectorielle de l'erreur (max et min)	$r_0 = 100$ (fixé a priori)
100	26 1	0,4388 0,0201	0,8127 0,0486 min
160	24 1	0,3024 0,0173	0,6798 0,0448 min
220	22 1	0,2107 0,0139	0,5720 max 0,0403 min

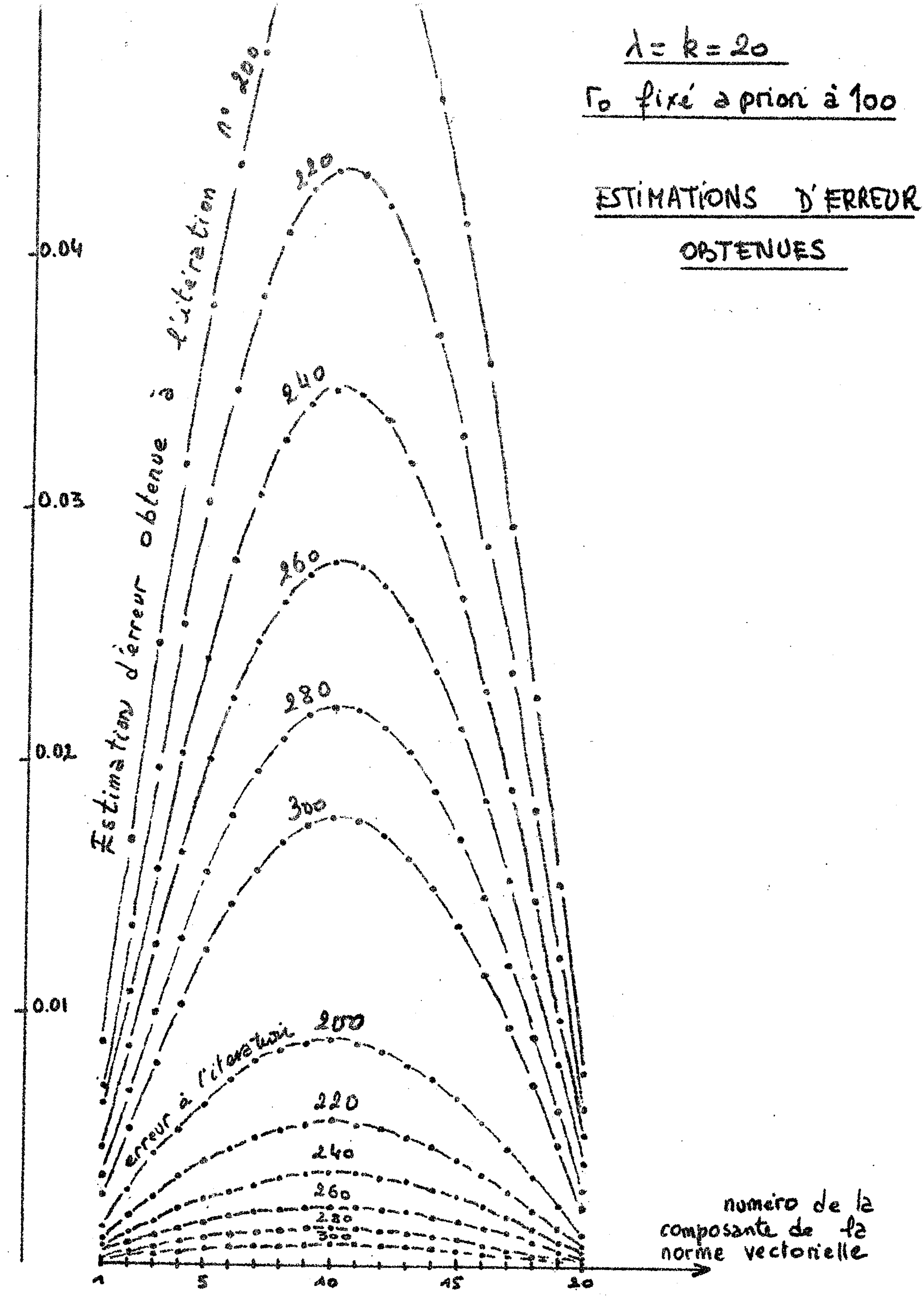
r_0, recherché par une itération préliminaire n'a pu être obtenu au bout de 8 minutes de calcul (programme ALGOL sur IBM 7044)

$\lambda = k = 50$ (Grille à 2500 points)

N° itération	N° ligne	$p(E_r)$ norme vectorielle de l'erreur max et min	$r_0 = 20$ (fixé a priori)
20	42 1	0,8524 0,0138	3,2024 0,2406 min
80	36 1	0,5992 0,0175	3,5055 0,1975 min
120	34 1	0,5037 0,0168	3,3619 0,1889 min

idem


λ = k = 10
ESTIMATIONS D'ERREUR OBTENUES
A L'ITÉRATION n° 27
A L'ITÉRATION n° 60
0,4
0,3
0,2
0,1
0,1.
0,075
0,05
Estimation obtenue
$r_0 = 27$ (obtenu par iteration préliminaire)
Estimation obtenue
$r_0 = 27$ (obtenu par itération préliminaire)
Estimation obtenue
r_0 fixé à 27
Estimation obtenue
r_0 fixé à 27
Norme vectorielle de l'erreur
Estimation obtenue
r_0 fixé à 60
Norme vectorielle de l'erreur
1 2 3 4 5 6 7 8 9 10
1 5
numéro de la composante de la norme vectorielle



$\lambda = k = 20$
r_0 fixé a priori à 100
ESTIMATIONS D'ERREUR OBTENUES
Estimation d'erreur obtenue à l'itération n° 200.
220
240
260
280
300
erreur à l'itération 200
220
240
260
280
300
0.04
0.03
0.02
0.01
1
5
10
15
20
numero de la composante de la norme vectorielle

ANNEXE

A 1 : *Une formulation plus générale du problème étudié en (6-6) :*

$\mathbb{K} = \mathbb{R}$ ou $\mathbb{C}$. Soit p une norme vectorielle régulière, de taille k sur $\mathbb{K}^n$, générant la norme vectorielle M, régulière et sous multiplicative, de taille k^2 sur l'espace $\mathcal{M}_{nn}$ des matrices carrées (n,n) à éléments dans $\mathbb{K}$.

Soit le système linéaire

$$Ay = b$$

où $A \in \mathcal{M}_{nn}$; y et $b \in \mathbb{K}^n$; y est la solution (inconnue). Soit y^* une solution approchée. D'où

$$\rho = Ay^* - b = A(y^* - y) \underset{df}{=} A\varepsilon$$

Peut-on estimer la norme vectorielle $p(\varepsilon)$ *du vecteur (inconnu) d'erreur* ε, *connaissant la norme vectorielle* $p(\rho)$ *du résidu* ρ?

Une réponse positive a été donnée au paragraphe (6-6) par l'intermédiaire de l'inégalité :

$$N(A)\, p(\varepsilon) \leq p(\rho)$$

qui impose au vecteur $p(\varepsilon)$ d'être dans la partie (connue) de $\mathbb{R}_+^k$ définie par :

$$\mathcal{D}(N(A), p(\rho)) \underset{df}{=} \{u \in \mathbb{R}_+^k \quad N(A)u \leq p(\rho)\}$$

cette partie étant bornée quel que soit ρ si et seulement si A est une bloc H-matrice.

Le chapitre précédent suggère une manière de procéder plus générale. Soit B une matrice (n,n). D'où

$$[I - M(I-BA)]p(x) \leq p(x) - p[(I-BA)x] \leq p(BAx) \leq M(B)p(Ax)$$

soit :

$$\forall x \in \mathbb{K}^n \quad [I - M(I-BA)]p(x) \leqslant M(B)p(Ax) \qquad (1)$$

et en particulier, pour $x = \varepsilon$

$$[I - M(I-BA)]p(\varepsilon) \leq M(B)\, p(\rho) \qquad (2)$$

Or, en posant $G = B^{-1}$, $A = G - H$ | $I - BA = I - G^{-1}A = G^{-1}(G-A) = G^{-1}H$

$I - BA = G^{-1}H$

$A = G - H$ splitting de A

inégalité qui impose au vecteur $p(\varepsilon)$ d'être dans la partie (connue) de $\mathbb{R}_+^k$ définie par

$$\{u \in \mathbb{R}_+^k \quad [I - M(I-BA)]u \leqslant M(B)p(\rho)\}$$

et l'on cherchera, bien entendu, à choisir B de façon que cette partie de $\mathbb{R}_+^k$ soit <u>bornée</u> quel que soit ρ (c'est-à-dire, puisque p, régulière, est surjective, quel que soit $p(\rho)$ dans $\mathbb{R}_+^k$)

<u>Exemples</u>

1 - Supposons A non singulière. Si nous prenons $B = A^{-1}$, (1) devient

$$\forall\, x \in \mathbb{K}^n \qquad p(x) \leqslant M(A^{-1})p(Ax) \tag{3}$$

d'où

$$p(\varepsilon) \leqslant M(A^{-1})p(\rho) \tag{4}$$

(pour mémoire, car sans intérêt pratique)

2 - Soit Δ une matrice non singulière, d'où $J = I - \Delta^{-1}A$ L étant une matrice sans valeur propre unité on décomposera J en L + U.

Prenons alors $B = \Delta^{-1}$. Il vient

$$\forall\, x \in \mathbb{K}^n \qquad [I - M(J)]p(x) \leqslant M(\Delta^{-1})\ p(Ax) \tag{5}$$

d'où

$$[I - M(J)]p(\varepsilon) \leqslant M(\Delta^{-1})\ p(\rho) \tag{6}$$

Remarque : Si Δ est la bloc diagonale de A pour la décomposition en blocs définie par la norme vectorielle M, on a (cf 6-2)

$$M(\Delta^{-1}) = [m(\Delta)]^{-1} \quad \text{(matrice diagonale positive)}$$

en effet

$$s_{ii}(A_{ii}^{-1}) = \operatorname*{Sup}_{x_i \in W_i} \left\{ \frac{\phi_i(A_{ii}^{-1}\, x_i)}{\phi_i(x_i)} \right\} = \frac{1}{\operatorname*{Inf}_{\substack{x_i \in W_i \\ x_i \neq 0}} \left\{ \frac{\phi_i(A_{ii}\, x_i)}{\phi_i(x_i)} \right\}} = \frac{1}{s_{ii}(A_{ii})}$$

et (5) est alors équivalente à

$$\forall\, x \in \mathbb{K}^n \quad m(\Delta)[I - M(J)]p(x) \leqslant p(Ax) \qquad (7)$$

Or on a posé (cf 6-2) $\quad N(A) = m(\Delta)[I - M(J)]$

On retrouve donc ainsi le cas, étudié au chapitre 6, de l'utilisation de la minorante $N(A)$.

3 - Prenons $B = (I - L)^{-1}\Delta^{-1}$ d'où

$I - BA = I - (I-L)^{-1}\Delta^{-1}\Delta(I-L-U) = (I-L)^{-1}(I - L - I + L + U) = (I - L)^{-1}U$

d'où

$$\forall\, x \in \mathbb{K}^n \quad (I - M[(I-L)^{-1}U])p(x) \leqslant M[(I-L)^{-1}\Delta^{-1}]p(Ax) \qquad (8)$$

4 - Enfin le choix de $B = \omega(I - \omega L)^{-1}\Delta^{-1}$ donne

$$I - BA = I - \omega(I - \omega L)^{-1}\Delta^{-1}\Delta(I - L - U) =$$

$$(I - \omega L)^{-1}[I - \omega L - \omega I + \omega L + \omega U] = \left(I - \omega L\right)^{-1}(\omega U + (1 - \omega)\, I)$$

et l'inégalité correspondante :

$$\forall\, x \in \mathbb{K}^n \quad I - M\left((I - \omega L)^{-1}(\omega U + (1 - \omega)I\right)\, p(x) \leqslant |\omega| M\left((I - \omega L)^{-1}\Delta^{-1}\right)p(Ax) \qquad (9)$$

(8) est évidemment cas particulier de (9) pour $\omega = 1$.

Remarque : Nous avons évidemment utilisé les mêmes "splittings" de A qu'au paragraphe 7-1.

Nous sommes ainsi amenés à poser la :

<u>A1 -1 - Définition</u> :

Pour tout $\nu \in \mathbb{R}_+^k$ on posera

$$\mathcal{H}(B,\nu) = \{u \in \mathbb{R}_+^k [I - M(I - BA)]u \leqslant M(B)\nu\}$$

$\mathcal{H}(B,\nu)$ est, pour tout ν, une partie convexe de $\mathbb{R}_+^k$ et le problème à résoudre est le suivant

Comment choisir la matrice B *pour que* $\mathcal{H}(B,\nu)$ *soit, pour tout* ν, *une partie bornée de* $\mathbb{R}^k_+$?

On sait répondre, grâce au

A1 -2 - Lemme *Soient* P *et* Q *2 matrices* $(k,k) \geq 0$. D'où, *pour tout* ν *dans* $\mathbb{R}^k_+$

$$G(\nu) \underset{df}{=} \{u \in \mathbb{R}^k_+ \quad (I - P)u \leq Q\nu\}$$

G(ν) est borné, pour tout ν de $\mathbb{R}^k_+$ si et seulement si P est convergente :

$$\rho(P) < 1.$$

α) *La condition est suffisante*

Si $\rho(P) < 1$, I - P est inversible et l'on a :

$$(I - P)^{-1} = I + P + \ldots\ldots P^r + \ldots\ldots \geq 0$$

soit $u \in G(\nu)$

$$(I - P)u \leq Q\nu$$

d'où par pré-multiplication par la matrice non-négative $(I - P)^{-1}$:

$$u \leq (I - P)^{-1}Q\nu$$

G(ν) est borné, quel que soit ν par $(I - P)^{-1}Q\nu$

β) *la condition est nécessaire*

Ab absurdo : si $\rho(P) \geq 1$ P, non négative, admet (Perron-Frobenius) un vecteur propre $\tau \geq 0$ ($\tau \neq 0$) associée à la valeur propre ρ

$$(I - P)\tau = \tau - \rho\tau = (1-\rho)\tau \leq 0 \leq Q\nu$$

Ainsi $\lambda\tau$ appartient, pour $\lambda > 0$ aussi grand que l'on veut, à G(ν) qui n'est donc pas borné.

La réponse au problème posé est donc la suivante

A1 -3 - Proposition. *Pour que* $\mathcal{H}(B\nu)$ *soit, quel que soit* ν, *une partie bornée de* $\mathbb{R}^k_+$, *il faut et il suffit que :*

$$\rho\Big(M(I-BA)\Big) < 1$$

c'est-à-dire que I - BA *soit contractante relativement à* p.

On retrouve donc un problème très classique : comment choisir B de façon à ce que la matrice d'itération I - BA soit contractante relativement à p.

A2 - Cas des "splittings" usuels

Dans le cas où l'on choisit $B = \Delta^{-1}$, l'inverse de la bloc diagonale de A pour la décomposition en blocs définie par M, on a vu que $I - BA = I - \Delta^{-1}A$ est alors la matrice de Jacobi par blocs associée à A.

Dire que J est contractante relativement à p, c'est dire que A est une bloc-H-matrice : ce cas particulier ($B = \Delta^{-1}$) redonne donc, évidemment, celui de la recherche d'une M-minorante, résolu en (6-2).

On note toujours L et U les blocs triangulaires respectivement inférieures et supérieures de J.

Alors, si A est une bloc-H-matrice, on sait, d'après le théorème 6 (7-2-1) que $\mathcal{L}_1 \underset{df}{=} (I - L)^{-1}U$ est contractante; de même

$\mathcal{L}_\omega \underset{df}{=} (I - \omega L)^{-1}(\omega U + (1-\omega)I)$, pour tous les ω pris dans la plage

$$0 < \omega < \frac{2}{1+\rho\big(M(J)\big)}$$

D'où la propriété suivante, qui ne fait qu'appliquer le résultat du théorème 6 à la question qui nous intéresse ici

A2 - 1 - Proposition *Soit* p *une norme vectorielle régulière sur* $\mathbb{K}^n$, *générant la norme vectorielle* M *de matrices. Soit* A *une matrice dont la bloc-diagonale* Δ *(pour la décomposition en blocs définie par* M*) est supposée non singulière, d'où*
$J = I - \Delta^{-1}A \underset{df}{=} L + U$

(*) *Dire que* $\mathcal{H}(\Delta^{-1}, \nu)$ *est, pour tout* ν, *une partie bornée de* $\mathbb{R}^k_+$ *c'est dire que* A *est une bloc-H-matrice relativement à* **p**.

(*) *Alors* $\mathcal{H}\big((I - L)^{-1}\Delta^{-1}, \nu\big)$ *est de même borné pour tout* ν;

(*) $\mathcal{H}\left(\omega(I - \omega L)^{-1}\Delta^{-1},\nu\right)$ *est borné pour tout* ν, *si* ω *est pris dans la plage*

$$0 < \omega < \frac{2}{1+\rho\left(M(J)\right)}$$

Remarques : 1 - En remplaçant, dans l'énoncé ci-dessus, la condition "A est une bloc-H-matrice" par son expression effective

$$\rho\left(M(J)\right) < 1$$

le résultat vrai si Δ est la bloc-diagonale de A pour une décomposition en blocs $\mathcal{D}_J$ moins fine que $\mathcal{D}_M$, décomposition définie par M (cf 7-2)

2 - De même que le théorème 6 et les corollaires donnés au chapitre 7 sont diverses formulations du théorème 5 (7-1-2), les résultats ci-dessus s'intègrent dans la formulation plus générale suivante, simple application du Théorème 5 à la question qui nous intéresse ici :

A2 - 2 - Proposition *Soit* p *une norme vectorielle régulière sur* $\mathbb{K}^n$, *et* M *la norme vectorielle de matrice issue de* p. *Soit* A *une matrice donnée,* B *une matrice que l'on choisit, d'où*

$$J = I - BA$$

décomposée arbitrairement en

$$J = L + U$$

On posera :

$$\rho = \rho\left(M(L) + M(U)\right)$$

Si $\rho<1$, *alors, pour tout* ν *pris dans* $\mathbb{R}^k_+$,

(*) $\mathcal{H}(B,\nu)$ *est borné dans* $\mathbb{R}^k_+$

(*) $(I - L)^{-1}$ *existe et* $\mathcal{H}\left((I - L)^{-1}B,\nu\right)$ *est borné dans* $\mathbb{R}^k_+$

(*) *pour tous les* ω *pris dans la plage*

$$0 < \omega < \frac{1}{1+\rho}$$

$(I - \omega L)^{-1}$ *existe et* $\mathcal{H}\left(\omega(I - L)^{-1}B,\nu\right)$ *est borné dans* $\mathbb{R}^k_+$

A3 - Applications

Nous étudions, dans ce paragraphe des propriétés des matrices A telles qu'il existe B pour laquelle I - BA est contractante relativement à une norme vectorielle p donnée. Ces propriétés généralisent celles des matrices admettant une M-minorante ou une B-Minorante.

A3 -1 - Proposition *Si* I-BA *est contractante alors*

$$u \in \mathcal{H}(B,\nu) \Rightarrow u \leq [I-M(I-BA)]^{-1}M(B)\nu$$

Posant $u_B(\nu) = [I-M(I-BA)]^{-1}M(B)\nu$

$u_B(\nu)$ *est, pour tout* ν *de* $\mathbb{R}^k_+$, *le plus grand élément de* $\mathcal{H}(B,\nu)$

En effet, si I - BA est contractante relativement à une norme vectorielle p donnée, $[I - M(I-BA)]^{-1}$ existe (5-2-2) et s'écrit

$$[I - M(I - BA)]^{-1} = I + M(I - BA) + \dots [M(I-BA)]^r + \dots \geq 0$$

Soit $u \in \mathcal{H}(B,\nu)$

$[I - M(I - BA)]u \leq M(B)\nu$ d'où par prémultiplication par $[I - M(I - BA)]^{-1}$, qui est non négative

$$u \leq [I - M(I - BA)]^{-1}M(B)\nu \underset{df}{=} u_B(\nu)$$

Ainsi $u_B(\nu)$ est bien un majorant de tout u pris dans $\mathcal{H}(B,\nu)$. Pour montrer que $u_B(\nu)$ est <u>le plus grand élément</u> de $\mathcal{H}(B,\nu)$ il suffit de montrer que $u_B(\nu) \in \mathcal{H}(B,\nu)$ ce qui est trivial.

A3 - 2 - Proposition *S'il existe* B *telle que* I-BA *est contractante alors* A *est non singulière*

En effet : de l'inégalité

$$\forall x \in \mathbb{K}^n \quad [I - M(I-BA)]p(x) \leq M(B)p(Ax)$$

résulte alors que

$$\forall x \in \mathbb{K}^n \quad p(x) \leq [I - M(I - BA)]^{-1}M(B)p(Ax) \qquad (1)$$

D'où la chaîne d'implications prouvant la non-singularité de A

$$Ax=0 \Rightarrow p(Ax)=0 \Rightarrow p(x)\leq 0 \Rightarrow p(x)=0 \Rightarrow x=0$$

D'ailleurs, d'après le lemme (5-2-2), si I-BA est contractante, on en déduit l'existence de

$$[I-(I-BA)]^{-1}$$

c'est-à-dire la non-singularité de BA : d'où, outre la non-singularité de A, celle de B : pour rendre I-BA contractante il est inutile de chercher B parmi les matrices singulières.

Remarque : Il résulte de (1) que, si I-BA est contractante, on a

$$M(A^{-1}) \leqslant [I-M(I-BA)]^{-1}M(B)$$

Cette dernière inégalité s'obtient aussi, lorsque I-BA est contractante, par prémultiplication par la matrice (non négative) $[I-M(I-BA)]^{-1}$ de chacun des membres de l'inégalité suivante, vraie quels que soient A inversible et B :

$$[I - M(I - BA)]M(A^{-1}) \leqslant M(B)$$

En effet

$$M(B) = M(A^{-1} - (A^{-1}-B)) \geqslant M(A^{-1})-M(A^{-1}-B) = M(A^{-1}) - M((I-BA)A^{-1})$$

$$\geqslant M(A^{-1}) - M(I - BA)M(A^{-1}) = [I - M(I-BA)]M(A^{-1})$$

A - 4 - CONCLUSION

Etant données une norme vectorielle p, régulière de taille k sur $\mathbb{K}^n$, générant la norme vectorielle M de matrices, et une matrice A de type (n,n), on peut poser les trois questions suivantes :

1) Quels sont les couples (N,M) de matrices telles que

$$\forall\, x \in \mathbb{K}^n \qquad Np(x) \leqslant Mp(Ax) \qquad (1)$$

2) A quelle condition sur A et p existe-t-il un couple (N,M) vérifiant (1) et tel que la partie suivante de $\mathbb{R}_+^k$

$$K(N,M,\nu) \underset{df}{=} \{u \in \mathbb{R}_+^k \qquad Nu \leqslant M\nu\}$$

soit <u>bornée</u>, quel que soit ν pris dans $\mathbb{R}_+^k$

Un tel couple, s'il existe, sera dit <u>utile</u>

3) Existe-t-il, parmi les couples vérifiant (1) un couple $(N_o M_o)$ <u>optimal</u> dans le sens où, quel que soit le couple (N,M) vérifiant (1) on ait

$$\forall \nu \in \mathbb{R}_+^k \qquad K(N_o, M_o, \nu) \subset K(N, M, \nu)$$

Nous ne savons pas apporter de réponse générale aux questions 1) et 3). Par contre, on peut répondre de la façon suivante à la question 2) :

Pour qu'existe un couple utile, il faut et il suffit que A soit non singulière

(cette condition est donc indépendante de p)

En effet

1) Si A est singulière, soit x^* un vecteur non nul de $\mathbb{R}^n$ tel que $Ax^* = 0$. On montre aisément que pour tout couple (N,M) vérifiant (1), $K(N,M,0)$ n'est pas borné puisqu'il contient $\lambda p(x^*)$ pour tous les λ positifs pris dans $\mathbb{R}$.

2) Si A est régulière, le couple $\left(I, M(A^{-1})\right)$ est un couple <u>utile</u>.

Remarque : Si (N,M) est un couple vérifiant (1) alors, pour toute diagonale D positive, et toute matrice de permutation P, le couple (DPN, DPM) vérifie de même (1) et l'on a

$$\forall \nu \in \mathbb{R}_+^k \qquad K(N,M,\nu) = K(DPN, DPM, \nu)$$

La recherche d'une Z-minorante s'inscrit dans le formalisme précédent pour $M = I$ et $N \in Z$. On a pu montrer que le couple $\left(N(A), I\right)$ est <u>optimal</u> parmi les couples (N,I) où N est une Z-minorante; et que, pour que $K\left(N(A), I, \nu\right)$ soit borné quel que soit ν dans $\mathbb{R}_+^k$, il était nécessaire et suffisant que $N(A)$ soit une M-matrice. (6-2)

La recherche d'une B-minorante n'impose que $M = I$ (6-5); cependant que le choix d'une matrice B rendant I-BA contractante (Annexe) revient à prendre $N = I-M(I-BA)$ et $M = M(B)$.

BIBLIOGRAPHIE

[1] F.L. BAUER *"On the field of values subordinated to a norm"*
Num Math 4, 103-113 (1962)

[2] F.L. BAUER *"Optimally scaled matrices"*
Num Math 5, 73-87 (1963)

[3] F.L. BAUER *"Theory of norms"*
Technical report n° CS 75 (August 1967)
Standford University

[4] F.L. BAUER - A.S. HOUSEHOLDER
"Absolute norms and characteristics roots"
Num Math 3, 241-246 (1961)

[5] F.L. BAUER - J.STOER - C.WITZGALL
"Absolute and monotonic norms"
Num Math 3, 257-264 (1961)

[6] F.L. BAUER - A.S. HOUSEHOLDER
"Some inequalities involving the euclidean condition of a matrix"
Num Math 2, 308-311 (1960)

[7] L. BOLLIET - N. GASTINEL - P.J. LAURENT
"Algol, manuel pratique"
Hermann (1964)

[8] N. BOURBAKI *"Integration"*
Ch. II Espaces de Riesz

[9] L. COLLATZ *"Funktional analysis und numerische mathematik"*
Springer-Verlag (1964)

[10] P. CAMION - AJ. HOFFMAN
"On the nonsingularity of complex matrices"
Pac. J. of Math Vol 17 n° 2 (1966)

[11] P. CAMION *"Une méthode de calcul des rayons spectraux des matrices carrées réelles non négatives ..."*
I C C research report N 65-3 (Septembre 1965)

[12] E. DURAND *"Solutions numériques des équations algébriques"*
(Tomes I et II) Masson (1960)

[13] N.A. DZERKO, A.M. PFEFFER
"Bounds for the spectral radius of a matrix"
Math Comput USA (1965) n° 89 p 62 à 67

[14] KY FAN *"Note on M-matrices"*
Quart J. Math Oxford 2nd Ser. 11 43-49 (1960)

[15] KY FAN *"Inequalities for M-matrices"*

[16] KY FAN *"Topological proofs for certain theorems on matrices with non-negative elements"*
Monatsh für Math 62, 219-237 (1958)

[17] D.FEINGOLD - R.S. VARGA
"Block diagonally dominant matrices and generalization of the Gershgorin circle theorem"
Pac. J. of Math 12 n°4 1241-49 (1962)

[18] M. FIEDLER - V. PTAK
"On matrices with non positive off diagonal elements and positive principal minors
Czec. Math J. 12(87) 382-400 (1962)

[19] M. FIEDLER - V. PTAK
"Generalized norms of matrices and the location of the spectrum"
Czec. Math J. 12 (87) 558-571 (1962)

[20] M. FIEDLER *"On inverting partitionned matrices"*
Czec. Math Journal 13 574-586 (1963)

[21] M. FIEDLER - V. PTAK
"Some generalization of positive definiteness and monotonicity"
Num. Math 9 163-172 (1966)

[22] M. FIELDER - V. PTAK
"Some results on matrices of class K and their application to the convergence rate of iteration procedures"
Czec. Math J. 16 (91) (1966) 260 - 273

[23] M. FIEDLER - V. PTAK
"Diagonally dominant matrices"
Czec. Math Journal 17 (92) 420-433 (1967)

[24] C. FOUSSAT - C. de POLIGNAC
"Borne supérieure de l'erreur dans une itération linéaire"
Projet de fin d'études - Laboratoire de Calcul de Grenoble (1966)

[25] F.R. GANTMACHER

"Théorie des matrices - Tomes I et II"
Dunod (1966)

[26] N. GASTINEL *"Matrices du second degré et normes générales en analyse numérique linéaire"*
These SDIT (1962)

[27] N. GASTINEL *"Analyse numérique linéaire"*
Hermann (1966)

[28] W. GAUTSCHI *"The asymptotic behaviour of powers of matrices"*
I 20, 127-140 (1953)
II 375-379

[29] V.V. GUDKOV *"On a certain test for non-singularity of matrices"*
(Russian and English suunmaries)
Latvian Math Yearbook (1965) 385-390
Izdat Zinatne Riga (1966)
Math Review # 1323 (February 1967)

[30] HARDY, LITTLEWOOD POLYA

"Inequalities"

[31] HENRICI *"Bounds for iterates, inverses, spectral variation and fieds of values of non-normal matrices"*
Num Math 4, 24-40 (1962)

[32] A.J. HOFFMANN *"On the non singularity of real partitionned matrices"*
ICC Bulletin Vol 4 7-17 (1965)

[33] A.S. HOUSEHOLDER

"The theory of matrices in numerical analysis"
Blaisdell Pub New York (1964)

[34] A.S. HOUSEHOLDER
"The approximate solution of matrix problems"
Journal of the ACM Vol 5 n° 3 (1958)

[35] A.S. HOUSEHOLDER
"Some numérical methods for solving systems of linear equations"

[36] KANTOROVITCH, VULIK, PINSKER
"Analyse fonctionnelle dans les espaces ordonnés"
(en Russe) Moscou (1950)

[37] KUHN TUCKER *"Linear inequalities and related systems"*
Princeton University Press 1956

[38] LIUSTERNIK SOBOLEV
"Elements of functionnal analysis"
UNGAR Pub Co New York 1961

[39] M. MARCUS H.MINC
"A survey of matrix theory inequalities"
Allyn and Bacon Boston (1964)

[40] A.OSTROWSKI *"Metrical properties of operators matrices and matrices partitionned into blocks"*
J.of Math Anal and Appl 2, 161-209 (1961)

[41] A. OSTROWSKI *"Iterative solution of linear systems of functional equations"*
J. of Math Anal and Appl 2,351-369 (1961)

[42] A.OSTROWSKI *"Uber normen von matrizen"*
Math Zeischr. Bd 63 2-18 (1955)

[43] A. OSTROWSKI *"Uber die determinanten mit uberwiegender Haupt-diagonale"*
Comment Math Helv. 10-1 69-96 (1937)

[44] A. OSTROWSKI *"Determinanten mit uberwiegender Hauptdiagonale und die Konvergenz von linearen iterations prozessen"*
Comment Math Helv. 30-3 175-209 (1954)

[45] A. OSTROWSKI *"Solutions of equations and systems of equations"*
Academic Press New York and London (1960)

[46] H.S. PRICE *"Monotone and Oscillation matrices applied to finite difference approximations"*
Report n° 189 Gulf Research (Avril 1965)

[47] F. ROBERT *"Normes vectorielles de vecteurs et de matrices"*
Revue de l'AFIRO n° 4 (1964)

[48] F. ROBERT *"Sur les normes vectorielles régulières sur un espace de dimension finie"*
CRAS T. 261 5173-5176 (Mai 1965)

[49] F. ROBERT *"Eléments propres d'une norme sur $\mathbb{R}^n$"*
Congrès de l'AFIRO LIlle (1966)

[50] F. ROBERT *"Calcul du rapport maximal de deux normes sur $\mathbb{R}^n$"*
Revue Française d'Informatique et de recherche Opérationnelle n° 5 (1967)

[51] F. ROBERT *"M-minorantes"*
Colloque international du CNRS - Besançon (Sept. 1966)

[52] F. ROBERT *"Remarque sur le conditionnement d'une matrice"*
Colloque International du CNRS, Besançon (Sept. 1966)

[53] F. ROBERT *"Recherche d'une M-matrice parmi les minorantes d'un opérateur linéaire"*
Num. Math 9 189-199 (1966)

[54] F. ROBERT *"Normes Vectorielles à valeurs dans un espace de Yosida régulier"*
CRAS T. 264 p. 1005-1008 (Juin 1967)

[55] F. ROBERT *"Espaces vectoriellement normés"*
Congrès de l'AFIRO - Nancy (1967)

[56] F. ROBERT *"Bloc-H-matrices et convergence des méthodes itératives classiques par blocs"*
A paraître

[57] H.H. SCHAEFFER *"Topological vector spaces"*
Macmillan Company 203-250 (1966)

[57] J.SCHROEDER *"Computing error bounds in solving linear systems"*
MRC Techn. report 242 University of Wisconsin (1961)

[59] J.SCHROEDER *"Das Iterations verfahren bei allgemeinerem Abstandsbegriff"*
Math Zeischr Bd 66 111-116 (1956)

[60] J. STOER *"On the characterization of least upper bounds norms in matrix space "*
Num Math 6 302-314 (1964)

[61] J. STOER - C. WITZGALL
"Transformations by diagonal matrices in normed spaces"
Num Math 4 158-171 (1962)

[62] R. TARRAL *"Le jus de pomme; contribution à l'étude de son utilisation diététique"*
Thèse Lescuyer, Lyon (1962)

[63] O. TAUSSKY *"A reccuring theorem on determinants"*
American Math Monthly 56 n° 10 (1949)

[64] R.S. VARGA *"Matrix iterative analysis"*
Prentice Hall (1962)

[65] R.S. VARGA *"Minimal Gershgorin sets"*
Pac J. of Math 15 n° 2 (1965)

[66] (B.LEVINGER) R.S. VARGA
"Minimal Gershgorin sets II"
Pac. J. of Math 17 n° 2 (1966)

[67] K. YOSIDA *"Functional Analysis"*
Springer Verlag (1966)

*

* *

Table des matières

VU

Grenoble, le

Le Président de la Thèse

VU

Grenoble, le

Le Doyen de la Faculté des Sciences

Vu, et permis d'imprimer,

Le Recteur de l'Académie de GRENOBLE