RAIRO. ANALYSE NUMÉRIQUE

B. MERCIER

Stabilité et convergence des méthodes spectrales polynômiales. Application à l'équation d'advection

RAIRO. Analyse numérique, tome 16, n° 1 (1982), p. 67-100

<http://www.numdam.org/item?id=M2AN_1982__16_1_67_0>

R A I R O Analyse numerique/Numerical Analysis
(vol 16, nº 1, 1982, p 67 a 100)

STABILITÉ ET CONVERGENCE DES MÉTHODES SPECTRALES POLYNÔMIALES. APPLICATION A L'ÉQUATION D'ADVECTION (*)

par B MERCIER (1)

Communique par P G CIARLET

Abstract — *We consider linear evolution equations of the hyperbolic type We first study the space semidiscretization, when the solution is restricted to be a polynomial of degree N We compare the stability and convergence results obtained for the Galerkin method, where the test functions are choosen in the same space as the approximate solution, and the Tau method which is a kind of least square method*

We then study the effect of time discretization, and show that the stability and convergence results obtained for the semidiscretized case extend easily to the Crank-Nicholson scheme

Resume — *Nous etudions l'approximation de la solution d'equations d'evolution lineaire de type hyperbolique par des polynomes de degre N Nous comparons les resultats de stabilite et de convergence obtenus pour la methode de Galerkin, ou les fonctions tests sont choisies dans le même espace que la solution approchee et la methode Tau qui correspond a une methode de moindres carres*

Nous montrons ensuite que ces resultats sont toujours valables lorsque l on utilise le schema de Crank-Nicholson pour discretiser le temps

Finalement nous donnons des estimations sur les valeurs propres de la matrice du probleme approche et des conditions suffisantes de stabilite pour des schemas de Runge-Kutta explicites

PLAN

(*) Reçu le 12 novembre 1980
(1) Commissariat a l'energie atomique, Centre d'Etudes de Limeil, Villeneuve St Georges, France

R A I R O Analyse numerique/Numerical Analysis, 0399-0516/1982/67/$5 00

INTRODUCTION

Le present travail a pour but de clarifier et de compléter certaines idées présentées par Gottlieb et Orszag dans leur livre sur les méthodes spectrales [1]

Il nous semble en particulier que la notion de stabilite algebrique qu'ils ont développée ne sert qu'à masquer certaines de leurs idées qui sont plus fécondes qu'il n'y paraît

Nous avons essayé de presenter les résultats généraux dans le cadre le plus abstrait possible, par exemple au § 1 Cependant ces résultats font appel à des hypothèses spécifiques qui ont bien des chances de n'être vérifiées que dans des exemples bien particuliers comme celui étudié au § 2

Cette présentation a tout de même l'avantage de bien mettre en évidence que *stabilité et consistance entraînent convergence*, et d'alléger les notations

Il en est de même pour le § 3, où l'on étudie la discrétisation du temps à l'aide du schéma de Crank-Nicholson

Enfin, au § 4, on donne quelques estimations sur les valeurs propres, qui permettent d'énoncer au § 5 quelques résultats de stabilité pour les schémas de Runge-Kutta explicites

La fin du § 5 est consacrée à la possibilité d'utiliser la transformée de Fourier rapide pour rendre les méthodes performantes sur le plan informatique

1. CADRE ABSTRAIT

1.1. Description des deux méthodes

Soit V un espace de Hilbert complexe muni du produit scalaire $(.,.)_\omega$, et $L : U \subset V \to V$, un opérateur linéaire non borné de domaine U.

Soit $f(t) \in V$, $t \geqslant 0$, et $u^0 \in U$ donné, on considère le problème d'évolution :

$$\left.\begin{aligned} &\frac{du}{dt} + Lu = f \\ &u(0) = u^0 \end{aligned}\right\}. \tag{1.1}$$

Soit $(U_N)_{N\in\mathbb{N}}, (V_N)_{N\in\mathbb{N}}$ deux familles d'espaces de dimension finie vérifiant

— $U_N \subset U, V_N \subset V$

— $\dim U_N = \dim V_N = N$

— *L est un isomorphisme de U_N sur V_N.*

Soit $f_N(t) \in V_N$ une approximation de $f(t)$ et $u_N^0 \in U_N$ une approximation de u^0 ; on considère les deux types de problèmes approchés suivants :

Méthode de Galerkin

Trouver $u_N(t) \in U_N$ vérifiant :

$$\left.\begin{aligned} &\left(\frac{du_N}{dt} + Lu_N - f_N, v_N\right)_\omega = 0\,, \quad \forall v_N \in U_N\,, \\ &u_N(0) = u_N^0 \end{aligned}\right\}. \tag{1.2}$$

Méthode Tau

Trouver $u_N(t) \in U_N$ vérifiant :

$$\left.\begin{aligned} &\left(\frac{du_N}{dt} + Lu_N - f_N, v_N\right)_\omega = 0\,, \quad \forall v_N \in V_N\,, \\ &u_N(0) = u_N^0 \end{aligned}\right\}. \tag{1.3}$$

La seule différence (mais elle est fondamentale) réside dans le choix (U_N ou V_N) de l'espace décrit par les fonctions tests.

On supposera par la suite que U_N ne contient pas d'élément orthogonal à V_N et réciproquement. Plus précisément ;

$$w_N \in U_N \quad \text{et} \quad (w_N, v_N)_\omega = 0\,, \quad \forall v_N \in V_N \Rightarrow w_N = 0\,, \tag{1.4}$$

$$v_N \in V_N \quad \text{et} \quad (v_N, w_N)_\omega = 0\,, \quad \forall w_N \in U_N \Rightarrow v_N = 0\,. \tag{1.5}$$

Dans ces conditions, on peut réécrire les problèmes (1.2) et (1.3) sous la forme suivante :

$$\frac{d}{dt}(\Pi_N u_N) + Lu_N = f_N \tag{1.6}$$

où $\Pi_N : U_N \to V_N$ est :

— l'inverse de la restriction à V_N de la projection sur U_N dans le cas de la méthode de Galerkin. Autrement dit Π_N vérifie :

$$(u_N - \Pi_N u_N, v_N)_\omega = 0\,, \quad \forall v_N \in U_N\,;$$

— la restriction à U_N de la projection sur V_N dans le cas de la méthode Tau.

On supposera par la suite que $U_{N-1} \subset U_N$, $V_{N-1} \subset V_N$, $N \geqslant 0$, et de plus que $U_{N-1} \subset V_N$, de sorte que Π_N coïncide avec l'identité sur U_{N-1}.

1.2. Stabilité pour la méthode de Galerkin

Soit $p_N \in V_{N+1}$ *orthogonal* à V_N. On suppose qu'il existe un produit scalaire $(.,.)_a$ indépendant de N tel que, $\forall u_N, w_N \in U_N$

$$(\Pi_N u_N, w_N)_a = (u_N, w_N)_a + \delta_N (u_N, p_N)_\omega \overline{(w_N, p_N)_\omega} \tag{1.7}$$

où $\delta_N > 0$ dépend éventuellement de N.

On suppose également qu'il existe deux normes notées $\| \cdot \|_b$ et $\| \cdot \|_c$ telles que

$$\mathrm{Re}\,(Lw_N, w_N)_a \geqslant \alpha \| w_N \|_b^2\,, \quad \forall w_N \in U_N \tag{1.8}$$

où $\alpha > 0$ est indépendant de N, et que

$$\mathrm{Re}\,(f, g)_a \leqslant \| f \|_b \| g \|_c\,, \quad \forall f, g \in V\,. \tag{1.9}$$

On a alors le résultat de *stabilité* suivant :

THÉORÈME 1.1 : *Soit $u_N(t)$ la solution du problème (1.2) et $\tilde{u}_N(t) \in U_N$ quelconque. On pose*

$$\tilde{f}_N = \frac{d}{dt}(\Pi_N \tilde{u}_N) + L\tilde{u}_N \tag{1.10}$$

et $g_N = f_N - \tilde{f}_N$, $w_N = u_N - \tilde{u}_N$; *sous les hypothèses* (1.7) (1.8) *et* (1.9), *et si de plus* $w_N(0) \in U_{N-1}$, *on a la majoration*

$$\| w_N(t) \|_a^2 \leqslant \| w_N(0) \|_a^2 + \frac{1}{\alpha} \int_0^t \| g_N(s) \|_c^2 \, ds \,. \tag{1.11}$$

Démonstration : Posons $b_N(t) = (w_N(t), p_N)_\omega$.
(On remarque que $b_N(0) = 0$ puisque $w_N(0) \in U_{N-1} \subset V_N$ et que p_N est orthogonal à V_N.)

Comme $\frac{d}{dt} \Pi_N w_N = \Pi_N \frac{dw_N}{dt} = \Pi_N \dot{w}_N$, on voit que d'après l'hypothèse (1.7) :

$$\left(\frac{d}{dt} \Pi_N w_N, w_N \right)_a = (\Pi_N \dot{w}_N, w_N)_a = (\dot{w}_N, w_N)_a + \delta_N \dot{b}_N \bar{b}_N \,. \tag{1.12}$$

D'autre part, en additionnant (1.6) et (1.10) on obtient

$$\frac{d}{dt} (\Pi_N w_N) + L w_N = g_N$$

qui donne après multiplication par w_N pour le produit scalaire $(.,.)_a$

$$\left(\frac{d}{dt} (\Pi_N w_N), w_N \right)_a + (L w_N, w_N)_a = (g_N, w_N)_a \,,$$

c'est-à-dire, en prenant la partie réelle et en appliquant (1.12) et (1.8)

$$\mathrm{Re}\, (\dot{w}_N, w_N)_a + \delta_N \,\mathrm{Re}\, (\dot{b}_N \bar{b}_N) + \alpha \| w_N \|_b^2 \leqslant \mathrm{Re}\, (g_N, w_N)_a \,,$$

qui équivaut à

$$\frac{1}{2} \frac{d}{dt} \| w_N(t) \|_a^2 + \frac{\delta_N}{2} \frac{d}{dt} | b_N(t) |^2 + \alpha \| w_N \|_b^2 \leqslant \mathrm{Re}\, (g_N, w_N)_a \,.$$

D'où, en appliquant (1.9)

$$\frac{d}{dt} \| w_N(t) \|_a^2 + \delta_N \frac{d}{dt} | b_N(t) |^2 \leqslant \frac{1}{\alpha} \| g_N(t) \|_c^2$$

on en déduit la majoration (1.11) en intégrant entre 0 et t et en utilisant le fait que $\delta_N > 0$ et que $b_N(0) = 0$. #

Remarque 1.1 : Le résultat précédent s'interprète comme un résultat de

stabilité. Si l'on choisit en effet $\tilde{u}_N(t) = 0$, et si l'on suppose que $u_N^0 \in U_{N-1}$ ce que l'on fera par la suite, on obtient :

$$\| u_N(t) \|_a^2 \leqslant \| u_N^0 \|_a^2 + \int_0^t \| f_N(s) \|_c^2 \, ds \, .$$

(Pour obtenir un résultat de stabilité avec $u_N^0 \in U_N$, il suffirait que δ_N soit majoré indépendamment de N.) #

1.3 Estimation d'erreur abstraite pour Galerkin

On suppose que $f(t)$ et $u(t)$ sont suffisamment réguliers. On choisit alors $\tilde{u}_N(t) \in U_{N-1}$ approchant $u(t)$ au mieux, et tel que $\tilde{u}_N(0) = u_N^0$.

Posons

$$\varepsilon_f(N) = \max_{0 \leqslant s \leqslant t} \| f(s) - f_N(s) \|_c$$

$$\varepsilon_1(N) = \max_{0 \leqslant s \leqslant t} \left\| \frac{d}{dt}(u - \tilde{u}_N)(s) \right\|_c$$

$$\varepsilon_2(N) = \max_{0 \leqslant s \leqslant t} \| L(u - \tilde{u}_N)(s) \|_c$$

$$\varepsilon_3(N) = \| u(t) - \tilde{u}_N(t) \|_a \, .$$

On a alors *l'estimation d'erreur* suivante

$$\| u(t) - u_N(t) \|_a \leqslant \varepsilon_3(N) + \left(\frac{t}{\alpha}\right)^{1/2} (\varepsilon_f(N) + \varepsilon_2(N) + \varepsilon_1(N)) \, . \quad (1.13)$$

En effet, d'après l'inégalité triangulaire

$$\| u(t) - u_N(t) \|_a \leqslant \| u(t) - \tilde{u}_N(t) \|_a + \| w_N(t) \|_a \quad (1.14)$$

où d'après la majoration (1.11)

$$\| w_N(t) \|_a \leqslant \left(\frac{1}{\alpha} \int_0^t \| g_N(s) \|_c^2 \, ds\right)^{1/2} ; \quad (1.15)$$

or d'après l'inégalité triangulaire

$$\| g_N(s) \|_c \leqslant \| f(s) - f_N(s) \|_c + \| f(s) - \tilde{f}_N(s) \|_c \quad (1.16)$$

et, comme $\tilde{u}_N \in U_{N-1}$,

$$f(s) - \tilde{f}_N(s) = \frac{d}{dt}(u - \tilde{u}_N)(s) + L(u - \tilde{u}_N)(s) \tag{1.17}$$

d'où le résultat annoncé (1.13) en combinant (1.14) à (1.17). #

Remarque 1.2 : Si l'on a l'analogue continu de (1.8) :

$$\mathrm{Re}\,(Lw, w)_a \geqslant \alpha \| w \|_b^2 , \quad \forall w \in U ,$$

alors on montre que le problème (1.1) est bien posé pour la norme $\| \cdot \|_a$.

En effet, on a alors, en multipliant (1.1) par u pour le produit scalaire $(\cdot, \cdot)_a$:

$$\left(\frac{du}{dt}, u\right)_a + (Lu, u)_a = (f, u)_a ,$$

d'où

$$\frac{1}{2}\frac{d}{dt}\| u(t) \|_a^2 + \alpha \| u \|_b^2 \leqslant \frac{1}{2\alpha}\| f \|_c^2 + \frac{\alpha}{2}\| u \|_b^2 ,$$

et en intégrant de 0 à t :

$$\| u(t) \|_a^2 \leqslant \| u(0) \|_a^2 + \int_0^t \| f(s) \|_c^2 \, ds .$$

En particulier si $f = 0$, $\| u(t) \|_a$ est *décroissante* avec t, et ceci limite déjà le degré d'arbitraire dans le choix du produit scalaire $(\cdot, \cdot)_a$. #

1.4. Stabilité de la méthode Tau

On suppose qu'il existe un produit scalaire $(\cdot, \cdot)_a$ et des normes $\| \cdot \|_b$ et $\| \cdot \|_c$ telles que

$$\mathrm{Re}\,(\Pi_N w_N, Lw_N)_a \geqslant \alpha \| w_N \|_b^2 , \quad \forall w_N \in U_N , \tag{1.18}$$

où $\alpha > 0$ est indépendant de N, et que

$$\mathrm{Re}\,(f, g)_a \leqslant \| f \|_c \| g \|_b . \tag{1.19}$$

On dénote par L^* l'adjoint de L pour le produit scalaire $(\cdot, \cdot)_a$ vérifiant

$$(f, Lu)_a = (L^* f, u)_a ;$$

on a alors le résultat de stabilité.

THÉORÈME 1.2 : *Soit* $u_N(t)$ *la solution du problème* (1.3) *et* $\tilde{u}_N(t) \in U_N$ *quelconque. On pose*

$$\tilde{f}_N = \frac{d}{dt}(\Pi_N \tilde{u}_N) + L\tilde{u}_N , \tag{1.20}$$

$w_N = u_N - \tilde{u}_N$ *et* $g_N = f_N - \tilde{f}_N$; *sous les hypothèses* (1.18) *et* (1.19) *on a la majoration* :

$$\| Lw_N(t) \|_a^2 \leqslant \| Lw_N(0) \|_a^2 + \frac{1}{\alpha} \int_0^t \| L^* g_N(s) \|_c^2 \, ds . \tag{1.21}$$

Démonstration : En additionnant les relations (1.6) et (1.20), on obtient :

$$\frac{d}{dt}(\Pi_N w_N) + Lw_N = g_N ,$$

que l'on multiplie par $L\dot{w}_N$ pour le produit scalaire $(.,.)_a$. En remarquant que

$$\frac{d}{dt} \Pi_N w_N = \Pi_N \frac{dw_N}{dt} \equiv \Pi_N \dot{w}_N ,$$

il vient

$$(\Pi_N \dot{w}_N, L\dot{w}_N)_a + (Lw_N, L\dot{w}_N)_a = (g_N, L\dot{w}_N)_a$$

dont on prend la partie réelle. Comme $\dot{w}_N \in U_N$, on peut appliquer (1.18) et on obtient

$$\alpha \| \dot{w}_N \|_b^2 + \frac{1}{2} \frac{d}{dt} \| Lw_N(t) \|_a^2 \leqslant \mathrm{Re}\,(g_N, L\dot{w}_N)_a .$$

Mais

$$\begin{aligned} \mathrm{Re}\,(g_N, L\dot{w}_N)_a &= \mathrm{Re}\,(L^* g_N, \dot{w}_N)_a \\ &\leqslant \frac{1}{2\alpha} \| L^* g_N \|_c^2 + \frac{\alpha}{2} \| \dot{w}_N \|_b^2 \end{aligned}$$

d'où

$$\frac{d}{dt} \| Lw_N(t) \|_a^2 \leqslant \frac{1}{\alpha} \| L^* g_N \|_c^2$$

qui entraîne (1.21) par intégration entre 0 et t. #

Remarque 1.3 : La majoration établie au théorème précédent s'interprète encore comme un résultat de stabilité. Si l'on choisit en effet $\tilde{u}_N(t) = 0$, la

majoration (1.21) montre que

$$\| Lu_N(t) \|_a^2 \leqslant \| Lu_N(0) \|_a^2 + \frac{1}{\alpha} \int_0^t \| L^* f_N(s) \|_c^2 \, ds$$

et que $u_N(t)$ est borné dans une norme *plus forte* que dans le cas Galerkin, mais sous des conditions de régularité *plus sévères* que pour la condition initiale et le second membre. #

1.5. Estimation d'erreur abstraite pour la méthode Tau

On suppose encore que $f(t)$ et $u(t)$ sont assez réguliers. On choisit $\tilde{u}_N(t) \in U_{N-1}$ approchant $u(t)$ au mieux, et tel que $\tilde{u}_N(0) = u_N^0$.

On a alors

$$\| L(u - u_N)(t) \|_a \leqslant \| L(u - \tilde{u}_N)(t) \|_a + \| Lw_N(t) \|_a .$$

Pour estimer $\| Lw_N(t) \|_a$ d'après (1.21) il suffit d'estimer

$$\| L^* g_N(s) \|_c \leqslant \| L^*(f - f_N)(s) \|_c + \| L^*(f - \tilde{f}_N)(s) \|_c$$

où l'on remarque que

$$L^*(f - \tilde{f}_N)(s) = L^* \frac{d}{dt}(u - \tilde{u}_N) + L^* L(u - \tilde{u}_N)$$

si l'on pose donc

$$\begin{aligned} \varepsilon_f(N) &= \max_{0 \leqslant s \leqslant t} \| L^*(f - f_N)(s) \|_c \\ \varepsilon_1(N) &= \max_{0 \leqslant s \leqslant t} \left\| L^* \frac{d}{dt}(u - \tilde{u}_N) \right\|_c \\ \varepsilon_2(N) &= \max_{0 \leqslant s \leqslant t} \| L^* L(u - \tilde{u}_N) \|_c \\ \varepsilon_3(N) &= \max_{0 \leqslant s \leqslant t} \| L(u - u_N)(s) \|_a \end{aligned} \tag{1.22}$$

on a une estimation d'erreur abstraite analogue à (1.13)

$$\| L(u - u_N)(t) \|_a \leqslant \varepsilon_3(N) + \left(\frac{t}{\alpha}\right)^{1/2} (\varepsilon_f(N) + \varepsilon_1(N) + \varepsilon_2(N)) . \tag{1.23}$$

qui est une estimation d'erreur dans une norme différente (plus forte) que pour la méthode de Galerkin. #

2. APPLICATION A L'ÉQUATION D'ADVECTION

2.1. Choix des espaces approchés

Soit I l'intervalle $]-1, +1[$, et $\omega : I \to \mathbb{R}^+$ une *fonction poids* donnée. On introduit le produit scalaire

$$(f, g)_\omega = \int_I f(x)\,\overline{g(x)}\,\omega(x)\,dx$$

et on appelle $L^2_\omega(I)$ l'espace des fonctions f mesurables telles que

$$\| f \|_\omega \equiv (f,f)^{1/2} < +\infty .$$

On supposera que le poids ω est continûment dérivable sur I, et que $L^2_\omega(I)$ contient les fonctions constantes.

Nous considérons *l'équation hyperbolique linéaire* [1]

$$\left.\begin{aligned} &\frac{\partial u}{\partial t} + \frac{\partial u}{\partial x} = f, \quad x \in I, t \geqslant 0, \\ &u(-1, t) = 0, \quad t \geqslant 0, \\ &u(x, 0) = u^0(x), \quad x \in I . \end{aligned}\right\} \tag{2.1}$$

Ce problème est un cas particulier du problème abstrait (1.1) à condition de poser

$$V = L^2_\omega(I)$$

$$U = \left\{ w \in L^2_\omega(I) : \frac{\partial w}{\partial x} \in L^2_\omega(I), w(-1) = 0 \right\}$$

et

$$Lu \equiv \frac{\partial u}{\partial x} .$$

Soit $\mathbb{P}_N$ l'espace des polynômes de degré $\leqslant N$, on pose

$$U_N = \{ w_N \in \mathbb{P}_N : w_N(-1) = 0 \}$$
$$V_N = \mathbb{P}_{N-1}$$

et on vérifie bien que L est un isomorphisme de U_N sur V_N.

---

[1] Appelée équation d'advection.

Nous allons étudier la convergence des solutions des problèmes approchés (1.2) et (1.3) lorsque $N \to \infty$, et pour cela *vérifier les hypothèses* des théomes 1.1 et 1.2.

2.2. Rappel des propriétés des polynômes orthogonaux ([1])

On rappelle que pour toute fonction poids ω positive il existe une famille de polynômes $(p_n)_{n\in\mathbb{N}}$ à coefficients réels telle que

$$\begin{cases} p_n \in \mathbb{P}_n\,, \\ \text{le coefficient de } x^n \text{ dans } p_n \text{ est strictement positif}\,, \\ (p_n, p_m)_\omega = \delta_{nm}\,. \end{cases}$$

Ces polynômes vérifient une relation de récurrence du type

$$xp_n = \alpha_n\, p_{n+1} + \beta_n\, p_n + \gamma_n\, p_{n-1}\,, \quad n \geqslant 1\,, \tag{2.2}$$

où $\alpha_n > 0$.

Les zéros de p_n séparent les zéros de p_{n+1}, et le polynôme p_n a n racines distinctes sur I.

Il en résulte que $p_n(1) > 0$, $\forall n$, et que

$$p_n(-1)\, p_{n+1}(-1) < 0\,, \quad \forall n \geqslant 0\,.$$

Soient x_j^N les racines du polynôme p_N et w_j^N les coefficients tels que la formule d'intégration numérique

$$\int_I f(x)\,\omega(x)\,dx \simeq \sum_{j=1}^{N} w_j^N\, f(x_j^N)\,. \tag{2.3}$$

Soit exacte pour $f \in \mathbb{P}_{N-1}$; on rappelle (2) que les x_j^N étant les racines du polynôme orthogonal p_N, la formule (2.3) est en fait exacte pour $f \in \mathbb{P}_{2N-1}$, et appelée *formule de Gauss* à N points.

Soient maintenant $\xi_0^N = -1$ et $(\xi_j^N)_{1\leqslant j\leqslant N}$ les racines du polynôme

$$p_N(-1)\, p_{N+1} - p_{N+1}(-1)\, p_N\,,$$

([1]) *Cf.* Laurent [2], pp. 63-66.
([2]) *Cf.* Davis-Rabinowitz [3].

et ω_j^N les coefficients tels que la formule d'intégration numérique

$$\int_I f(x)\,\omega(x)\,dx \simeq \sum_{j=0}^{N} \omega_j^N\, f(\xi_j^N) \tag{2.4}$$

soit exacte pour $f \in \mathbb{P}_N$.

Avec le choix particulier des ξ_j^N qui a été effectué, la formule (2.4) est en fait exacte pour $f \in \mathbb{P}_{2N}$ et appelée formule de Gauss-Radau à $(N+1)$ points.

On vérifie alors immédiatement les hypothèses (1.4) et (1.5) sur les espaces U_N et V_N. En effet pour $v_N \in V_N$ et $w_N \in U_N$

$$(v_N, w_N)_\omega = \sum_{j=1}^{N} w_j^N\, v_N(x_j^N)\, w_N(x_j^N)\,.$$

Par conséquent $(w_N, v_N)_\omega = 0$, $\forall v_N \in V_N \Leftrightarrow w_N(x_j^N) = 0$ qui entraîne que w_N est nul puisque nul en $(N+1)$ points distincts. Inversement $(v_N, w_N)_\omega = 0$, $\forall w_N \in U_N \Leftrightarrow v_N(x_j^N) = 0$ qui entraîne que v_N est nul, car un polynôme de degré $N-1$ nul en N points est nul.

2.3. Vérification des hypothèses dans le cas Galerkin

On va voir qu'il est crucial de choisir

$$(f, g)_a = \int_I f\bar{g}\omega_1\,dx \equiv (f, g)_{\omega_1}\,, \quad \text{où} \quad \omega_1(x) = (1-x)\,\omega(x)\,.$$

on rappelle que pour la méthode de Galerkin, l'opérateur Π_N intervenant dans (1.6) est caractérisé par

$$(u_N - \Pi_N u_N, v_N)_\omega = 0\,, \quad \forall v_N \in U_N\,. \tag{2.5}$$

LEMME 2.1 : *L'opérateur* $\Pi_N : U_N \to V_N$ *caractérisé par* (2.5) *vérifie la condition* (1.7) *avec*

$$\delta_N = -\alpha_N \frac{p_{N+1}(-1)}{p_N(-1)} > 0\,.$$

Démonstration : Soit u_N, $w_N \in U_N$ donnés. On pose $a_N = (u_N, p_N)_\omega$ et $b_N = (w_N, p_N)_\omega$.

Soit $l_N = u_N - \Pi_N u_N$; U_N étant de dimension N, son orthogonal dans $\mathbb{P}_N$ est de dimension 1. Or on vérifie que

$$\varphi_N = \sum_{n=0}^{N} p_n(-1)\, p_n \in \mathbb{P}_N$$

est orthogonal à U_N. Par conséquent l_N est proportionnel à φ_N. Soit

$$u_N - \Pi_N u_N = \tau_N \varphi_N .$$

En prenant le produit scalaire avec p_N, et en remarquant que $(p_N, \Pi_N u_N)_\omega = 0$, il vient

$$a_N = (u_N, p_N)_\omega = \tau_N p_N(-1) ,$$

qui fournit ainsi une évaluation de τ_N.

On a alors

$$(\Pi_N u_N, w_N)_a = (u_N, w_N)_a - \tau_N(\varphi_N, w_N)_a .$$

Or

$$(\varphi_N, w_N)_a \equiv (\varphi_N, (1-x) w_N)_\omega = -(\varphi_N, (1+x) w_N)_\omega \tag{2.6}$$

puisque $(\varphi_N, w_N)_\omega = 0$, par définition de φ_N.

Posons $\tilde{w}_N = w_N - b_N p_N \in \mathbb{P}_{N-1}$. On remarque, que $(1+x)\, \tilde{w}_N \in U_N$ et, par conséquent

$$(\varphi_N, (1+x) w_N)_\omega = \bar{b}_N(\varphi_N, (1+x) p_N)_\omega . \tag{2.7}$$

La relation de récurrence (2.2) permet d'écrire

$$\begin{aligned}(\varphi_N, x p_N)_\omega &= (\varphi_N, \alpha_N p_{N+1} + \beta_N p_N + \gamma_N p_{N-1})_\omega \\ &= \beta_N p_N(-1) + \gamma_N p_{N-1}(-1) .\end{aligned}$$

Donc

$$\begin{aligned}(\varphi_N, (1+x) p_N)_\omega &= (1+\beta_N) p_N(-1) + \gamma_N p_{N-1}(-1) \\ &= -\alpha_N p_{N+1}(-1)\end{aligned} \tag{2.8}$$

où l'on utilise à nouveau la relation de récurrence.

Finalement (2.6), (2.7) et (2.8) montrent que

$$(\varphi_N, w_N)_a = \bar{b}_N \alpha_N p_{N+1}(-1)$$

d'où

$$(\Pi_N u_N, w_N)_a = (u_N, w_N)_a + \delta_N a_N \bar{b}_N$$

qui est bien le résultat à démontrer, où $\delta_N = -\alpha_N \dfrac{p_{N+1}(-1)}{p_N(-1)}$. #

Le deuxième avantage du produit scalaire

$$(\cdot, \cdot)_a \equiv (\cdot, \cdot)_{\omega_1}$$

est que, le poids ω_1 étant nul en $x = 1$, on peut intégrer facilement par parties.

On suppose que le poids

$$\omega_2 \equiv -\omega_1' > 0$$

autrement dit que le poids ω_1 est *décroissant*.

On vérifie alors facilement l'hypothèse (1.8) :

LEMME 2.2 : *Soit* $\|\cdot\|_{\omega_2}$ *la norme associée au produit scalaire*

$$(f, g)_{\omega_2} \equiv \int_I f\bar{g}\omega_2 \, dx ,$$

on a

$$\mathrm{Re}\left(\frac{\partial w}{\partial x}, w\right)_{\omega_1} \geqslant \frac{1}{2} \| w \|_{\omega_2}^2 , \quad \textit{pour} \quad w \in U .$$

Démonstration : Comme $\mathrm{Re}\left(\dfrac{\partial w}{\partial x}\bar{w}\right) = \dfrac{1}{2}\dfrac{\partial}{\partial x}|w|^2$, on a, par intégrations par parties :

$$\mathrm{Re}\int_I \frac{\partial w}{\partial x}\bar{w}\omega_1 \, dX = |\omega_1 |w|^2 |_{-1}^{+1} - \frac{1}{2}\int_I |w|^2 \omega_1' \, dX ,$$

d'où le résultat puisque $w \in U \Rightarrow w(-1) = 0$, que $\omega_1(1) = 0$ et que $\omega_2 = -\omega_1'$. #

Enfin, le résultat suivant indique comment choisir $\|\cdot\|_c$ pour que la condition (1.9) soit vérifiée.

LEMME 2.3 : *Soit* $\omega_3 = \omega_1^2 \omega_2^{-1}$, *on a la majoration*

$$\mathrm{Re}\,(f, u)_{\omega_1} \leqslant \| f \|_{\omega_3} \| u \|_{\omega_2} , \quad f, u \in L^2_{\omega_1}(I) .$$

Démonstration : On écrit $f\bar{u}\omega_1 = (f\omega_1 \omega_2^{-1/2})(\bar{u}\omega_2^{1/2})$ et on applique l'inégalité de Schwarz. #

2.4. Estimation d'erreur dans le cas Galerkin-Tchebycheff

On a donc vérifié les hypothèses (1.7), (1.8), (1.9) du théorème 1.1 qui entraînent donc la validité de l'estimation d'erreur (1.13). Il reste à choisir $\tilde{u}_N$ et à évaluer les quantités $\varepsilon_f(N)$, $\varepsilon_1(N)$, $\varepsilon_2(N)$ et $\varepsilon_3(N)$, ce que l'on va faire dans le cas particulier où le poids $\omega(x) \equiv (1 - x^2)^{-1/2}$ est le poids de Tchebycheff.

Dans ce cas, on a

$$\omega_1 = \left(\frac{1-x}{1+x}\right)^{1/2} \quad \text{et} \quad \omega_2 = -\omega_1' = (1+x)^{-1}(1-x^2)^{-1/2} > 0\,.$$

Enfin, on vérifie que

$$\omega_3 = (1-x)(1-x^2)^{1/2}\,,$$

de sorte que $\omega_1 \leqslant \omega$ et $\omega_3 \leqslant \omega$, et l'on peut majorer les normes avec poids ω_1 et ω_3 par la norme avec le poids de Tchebycheff ω.

On va choisir pour $\tilde{u}_N(t) \in U_{N-1}$ *l'interpolé* de $u(t)$ aux N points de Gauss-Radau correspondant au poids de Tchebycheff.

D'après Canuto-Quarteroni [4], si $u(t) \in H^s_\omega(I)$ et $f(t) \in H^{s'}_\omega(I)$, on a

$$\begin{aligned}
\| u(t) - \tilde{u}_N(t) \|_\omega &\leqslant CN^{-s} \| u(t) \|_{s,\omega} \\
\| L(u - \tilde{u}_N)(t) \|_\omega &\leqslant CN^{2-s} \| u(t) \|_{s,\omega} \\
\| f(t) - f_N(t) \|_\omega &\leqslant CN^{-s'} \| f(t) \|_{s',\omega}\,.
\end{aligned}$$

D'autre part, en remarquant que $\dfrac{\partial \tilde{u}_N}{\partial t}(t)$ est l'interpolé de $\dfrac{\partial u}{\partial t}(t)$, on a, puisque

$$\frac{\partial u}{\partial t}(t) \in H^\sigma_\omega(I) \quad \text{avec} \quad \sigma = \min(s', s-1)$$

$$\left\| \frac{\partial u}{\partial t}(t) - \frac{\partial \tilde{u}_N}{\partial t}(t) \right\|_\omega \leqslant CN^{-\sigma} \left\| \frac{\partial u}{\partial t}(t) \right\|_{\sigma,\omega}\,.$$

Finalement, si l'on suppose que $u \in L^\infty(0, t; H^s_\omega(I))$ et $f \in L^\infty(0, t; H^{s'}_\omega(I))$, on aura

$$\begin{aligned}
\varepsilon_1(N) &= 0(N^{-\sigma})\,, \qquad \varepsilon_2(N) = 0(N^{2-s})\,, \\
\varepsilon_3(N) &= 0(N^{-s})\,, \qquad \varepsilon_f(N) = 0(N^{-s'})\,.
\end{aligned}$$

En général, c'est le terme en $0(N^{2-s})$ qui sera prépondérant et pour qu'il y ait une estimation d'erreur, il est donc nécessaire que $u(t) \in H^{2+\varepsilon}_\omega(I)$, et en pratique

que la solution $u(t)$ soit au moins C^1. La convergence, quant à elle, aura lieu si $u(t) \in H^1_\omega(I)$ en appliquant des arguments de densité (choisir $\tilde{u}_N(t)$ égal à l'interpolé d'un régularisé de $u(t)$).

Remarque 2.1 : On pourrait appliquer les résultats précédents dans le cas d'un poids uniforme $\omega \equiv 1$. Dans ce cas $\omega_1 = 1 - x, \omega_2 = 1$ et $\omega_3 = (1 - x)^2$. En remarquant que ω_1 et ω_3 sont majorés par le poids de Tchebycheff, on obtient des évaluations analogues des quantités :

$$\varepsilon_f(N), \varepsilon_1(N), \varepsilon_2(N) \quad \text{et} \quad \varepsilon_3(N). \quad \#$$

Remarque 2.2 : On a signalé à la remarque 1.1, que l'hypothèse $u_N^0 \in U_{N-1}$ n'était pas nécessaire si δ_N était majoré indépendamment de N. Or ceci est vrai si ω est le poids de Tchebycheff puisqu'alors $\delta_N = 1/2, \forall N \geqslant 1$. C'est aussi vrai dans le cas d'un poids uniforme (propriétés des polynômes de Legendre). #

Commentaire : Gottlieb et Orszag ont établi le résultat démontré au lemme 2.1 dans le cas particulier du poids de Tchebycheff $(1 - x^2)^{-1/2}$ (*cf.* [1], pp. 95-98) et s'en sont servis pour démontrer la stabilité de l'équation sans second membre. Les estimations d'erreur établies ici semblent démentir la conjecture de Orszag et Jayne [1], p. 134.

2.5. Vérification des hypothèses dans le cas de la méthode Tau

Il faut d'abord choisir le produit scalaire $(.,.)_a$, et comme on va le voir, il est crucial de le choisir égal à

$$(f, g)_{\omega_1} = \int_I fg\omega_1 \, dX$$

où $\omega_1 = (1 - x)\,\omega$, comme pour la méthode de Galerkin.

En supposant encore $\omega_2 = -\omega_1'$ décroissant, on vérifie alors l'hypothèse (1.18) :

LEMME 2.4 : *Soit* $\omega_2 = -\omega_1'$, *on a, pour* $w_N \in U_N$

$$\mathrm{Re}\,(\Pi_N w_N, Lw_N)_{\omega_1} = \alpha \parallel w_N \parallel^2_{\omega_2} + N \mid b_N \mid^2$$

où $b_N = (w_N, p_N)_\omega$.

Démonstration : On a, par définition de Π_N :

$$w_N = b_N p_N + \Pi_N w_N .$$

Soit c_N le coefficient de x^N dans p_N ; on voit que le coefficient de x^N dans $s_N = (1 - x)\dfrac{\partial w_N}{\partial x}$ est $- Nb_N c_N$, de sorte que

$$s_N = - Nb_N p_N + q_{N-1}$$

où $q_{N-1} \in \mathbb{P}_{N-1}$, de sorte que $q_{N-1} = \Pi_N s_N$.

Il en résulte que

$$\begin{aligned}\left(\Pi_N w_N, \frac{\partial w_N}{\partial x}\right)_{\omega_1} &= (\Pi_N w_N, s_N)_\omega = (w_N - b_N p_N, s_N)_\omega \\ &= (w_N, s_N)_\omega - b_N(p_N, s_N)_\omega \\ &= (w_N, s_N)_\omega + N \mid b_N \mid^2 .\end{aligned}$$

Il suffit pour conclure de remarquer que, en intégrant par parties,

$$\mathrm{Re}\,(w_N, s_N)_\omega = \mathrm{Re}\left(w_N, \frac{\partial w_N}{\partial x}\right)_{\omega_1} = \frac{1}{2} \| w_N \|^2_{\omega_2} . \quad \#$$

Comme dans le cas Galerkin, on voit qu'il faut choisir

$$\| \cdot \|_c = \| \cdot \|_{\omega_3}$$

avec $\omega_3 = \omega_2^{-1} \omega_1^2$ pour que (1.19) soit satisfait.

2.6. Estimation d'erreur pour la méthode Tau

Montrons tout d'abord, en intégrant par parties que

$$\left(f, \frac{\partial u}{\partial x}\right)_{\omega_1} \equiv - \left(\frac{\partial f}{\partial x}, u\right)_{\omega_1} + (f, u)_{\omega_2}$$

de sorte que

$$\| L^* f \|_{\omega_1} \leqslant \| Lf \|_{\omega_1} + \| f \|_{\omega_2} . \tag{2.9}$$

Dans le cas du poids de Tchebycheff $\omega(x) = (1 - x^2)^{-1/2}$, on a

$$\omega_2 = (1 + x)^{-1}(1 - x^2)^{-1/2}$$

qui n'est donc pas majoré par $\omega(x)$, et on ne peut donc appliquer les résultats de Canuto-Quarteroni [4].

Dans l'estimation d'erreur (1.23), c'est le terme $\varepsilon_2(N)$ défini en (1.22) qui est dominant. D'après (2.9) ce terme contient déjà la norme de $u(t) - \tilde{u}_N(t)$

dans l'espace $H^2_\omega(I)$ qui est seulement en $0(N^{4-s})$ si $u(t) \in H^s_\omega(I)$ c'est-à-dire qu'il n'y a d'estimation d'erreur que si $u(t) \in H^s_\omega(I)$ avec $s > 4$, et convergence que si $u(t) \in H^2_\omega(I)$.

La méthode Tau donnera donc éventuellement convergence dans une norme plus forte que la méthode de Galerkin, mais les conditions de régularité assurant la convergence de la méthode sont a priori beaucoup plus sévères que pour la méthode de Galerkin.

Commentaires bibliographiques

Gottlieb et Orszag ont établi le résultat démontré au lemme 2.4 dans le cas particulier du poids de Tchebycheff [1], pp. 99-100, et en ont déduit un résultat de stabilité pour l'équation sans second membre.

3. DISCRÉTISATION A L'AIDE DU SCHÉMA DE CRANK-NICHOLSON

3.1. Introduction

On revient au cadre abstrait du § 1.

Les équations (1.2) et (1.3) ne sont que semi-discrétisées en espace. Pour une résolution effective, il faut aussi une discrétisation du temps. Nous allons étudier l'application du schéma de Crank-Nicholson.

Soit $\Delta t > 0$ fixé, on cherche maintenant pour $k = 1, 2, ..., \nu$:

$$u_N^k \in U_N$$

vérifiant

$$((\partial u_N)^{k+1/2} + Lu_N^{k+1/2} - f_N^{k+1/2}, v_N)_\omega = 0, \quad \forall v_N \in U_N, \tag{3.1}$$

avec

$$(\partial u_N)^{k+1/2} \equiv \frac{1}{\Delta t}(u_N^{k+1} - u_N^k)$$
$$u_N^{k+1/2} \equiv \tfrac{1}{2}(u_N^{k+1} + u_N^k)$$
$$f_N^{k+1/2} \equiv f_N((k + 1/2)\,\Delta t)$$

dans le cas de la méthode de Galerkin, et

$$((\partial u_N)^{k+1/2} + Lu_N^{k+1/2} - f_N^{k+1/2}, v_N)_\omega = 0, \quad \forall v_N \in V_N, \tag{3.2}$$

dans le cas de la méthode Tau.

Nous allons étudier les résultats de stabilité et de convergence correspondants.

3.2. Stabilité pour Galerkin

On procède de façon analogue au cas semi-discrétisé étudié au § 1.2.

THÉORÈME 3.1 : *Soit* u_N^k, $k = 1, ..., \nu$, *la solution du problème* (3.1) *et* $\tilde{u}_N^k \in U_N$ *quelconque. On pose*

$$\begin{aligned} \tilde{f}_N^{k+1/2} &= (\partial \tilde{u}_N)^{k+1/2} + \frac{\partial}{\partial x} \tilde{u}_N^{k+1/2} \\ g_N^{k+1/2} &= f_N^{k+1/2} - \tilde{f}_N^{k+1/2}, \quad \text{et} \quad w_N^k = u_N^k - \tilde{u}_N^k . \end{aligned} \tag{3.3}$$

Sous les hypothèses (1.7), (1.8) *et* (1.9), *et si de plus* $w_N^0 \in U_{N-1}$, *on a la majoration* :

$$\| w_N^\nu \|_a^2 \leqslant \| w_N^0 \|_a^2 + \sum_{k=0}^{\nu-1} \Delta t \, \| g_N^{k+1/2} \|_c^2 . \tag{3.4}$$

Démonstration : Posons $b_N^k = (w_N^k, p_N)_\omega$. (Noter que $w_N^0 \in U_{N-1} \Rightarrow b_N^0 = 0$.) Comme

$$(\partial \Pi_N w_N)^{k+1/2} = \Pi_N (\partial w_N)^{k+1/2}$$

on voit que l'hypothèse (1.7) entraîne

$$((\partial \Pi_N w_N)^{k+1/2}, w_N^{k+1/2})_a = ((\partial w_N)^{k+1/2}, w_N^{k+1/2})_a + \delta_N (\partial b_N)^{k+1/2} \overline{b_N^{k+1/2}}$$

et par conséquent

$$\begin{aligned} \mathrm{Re}\,((\partial \Pi_N w_N)^{k+1/2}, w_N^{k+1/2})_a = \frac{1}{2\Delta t}(\| w_N^{k+1} \|_a^2 - \| w_N^k \|_a^2) + \\ + \frac{1}{2\Delta t}(| b_N^{k+1} |^2 - | b_N^k |^2) . \end{aligned} \tag{3.5}$$

En remarquant que (3.1) peut s'écrire, de façon analogue à (1.6) :

$$(\partial \Pi_N u_N)^{k+1/2} + L u_N^{k+1/2} = f_N^{k+1/2}$$

on obtient, par addition avec (3.3) :

$$(\partial \Pi_N w_N)^{k+1/2} + L w_N^{k+1/2} = g_N^{k+1/2} .$$

En effectuant le produit scalaire $(.,.)_a$ de cette équation par $w_N^{k+1/2}$, on obtient en prenant la partie réelle, et en utilisant (3.5).

On applique alors les hypothèses (1.8) et (1.9), et on somme de $k = 0$ à $\nu - 1$ pour obtenir :

$$\| w_N^\nu \|_a^2 + | b_N^\nu |^2 + \alpha \sum_{k=0}^{\nu-1} \Delta t \| w_N^{k+1/2} \|_b^2 \leqslant$$

$$\leqslant \| w_N^0 \|_a^2 + | b_N^0 |^2 + \frac{1}{\alpha} \sum_{k=0}^{\nu-1} \Delta t \| g_N^{k+1/2} \|_c^2$$

d'où l'on tire la majoration désirée.

Remarque 3.1 : En choisissant $\tilde{u}_N^k = 0$, $\forall k \geqslant 0$, et en supposant $u_N^0 \in U_{N-1}$, le résultat précédent s'interprète comme un résultat de stabilité pour $\| u_N^k \|_a$. #

3.3. Estimation d'erreur pour Galerkin

On choisit $\tilde{u}_N^k \in U_{N-1}$ aussi proche que possible de $u^k \equiv u(k\,\Delta t)$. En utilisant l'inégalité triangulaire on a

$$\| u^k - u_N^k \|_a \leqslant \| u^k - \tilde{u}_N^k \|_a + \| w_N^k \|_a ,$$

de sorte que d'après la majoration (3.4) il suffit de majorer

$$\| g_N^{k+1/2} \|_c \leqslant \| f^{k+1/2} - f_N^{k+1/2} \|_c + \| f^{k+1/2} - \tilde{f}_N^{k+1/2} \|_c .$$

Pour majorer le dernier terme, on remarque que

$$f^{k+1/2} - \tilde{f}_N^{k+1/2} = \left(\frac{du}{dt} + Lu \right) ((k + 1/2)\,\Delta t) - [(\partial \tilde{u}_N)^{k+1/2} + L\tilde{u}_N^{k+1/2}] .$$

Soit u_I *l'interpolée en temps* de u (c'est-à-dire la fonction continue et affine par morceaux en t, telle que u et u_I coïncident à tous les instants $k\,\Delta t$, $k = 0, 1, ..., \nu$).

On sait que

$$\left(\frac{du_I}{dt} + Lu_I \right) ((k + 1/2)\,\Delta t) = (\partial u_I)^{k+1/2} + Lu_I^{k+1/2}$$

de sorte que

$$f^{k+1/2} - \tilde{f}_N^{k+1/2} = \left(\frac{d}{dt} + L \right) (u - u_I)\,((k + 1/2)\,\Delta t) +$$

$$+ [(\partial(u_I - \tilde{u}_N))^{k+1/2} + L(u_I - \tilde{u}_N)^{k+1/2}] .$$

Posons

$$\eta(\Delta t) = \max_k \left\| \left(\frac{d}{dt} + L\right)(u - u_I)\left((k + 1/2)\,\Delta t\right) \right\|_c$$

$$\varepsilon_1(N) = \max_k \| \partial(u_I - \tilde{u}_N)^{k+1/2} \|_c$$

$$\varepsilon_2(N) = \max_k \| L(u_I - \tilde{u}_N)^{k+1/2} \|_c$$

$$\varepsilon_3(N) = \max_k \| u^k - \tilde{u}_N^k \|_a$$

$$\varepsilon_f(N) = \max_k \| f^{k+1/2} - f_N^{k+1/2} \|_c .$$

On a l'estimation d'erreur abstraite

$$\| u^k - u_N^k \|_a \leqslant \varepsilon_3(N) + \left(\frac{\nu\,\Delta t}{\alpha}\right)^{1/2} (\varepsilon_f(N) + \eta(\Delta t) + \varepsilon_1(N) + \varepsilon_2(N)) .$$

Par rapport à (1.13) cette estimation introduit en plus le terme $\eta(\Delta t)$ qui est l'erreur de discrétisation en t, qui est en $O(\Delta t^2)$ pour le schéma de Crank-Nicholson.

Les autres termes sont du même ordre que dans le cas semi-discret (*cf.* § 2.4).

Remarque 3.2 : *Schéma rétrograde* : C'est le schéma obtenu en remplaçant $u_N^{k+1/2}$ par u_N^{k+1} et $f_N^{k+1/2}$ par f_N^{k+1} dans l'équation (3.1).

On établit un résultat de stabilité analogue au théorème 3.1, en remarquant (seul point qui change dans la démonstration) que

$$\mathrm{Re}\,((\partial w_N)^{k+1/2}, w_N^{k+1})_a = \frac{1}{\Delta t}\,\mathrm{Re}\left(w_N^{k+1} - w_N^k, \frac{w_N^{k+1} + w_N^k}{2} + \frac{w_N^{k+1} - w_N^k}{2}\right)_a$$

$$\geqslant \frac{1}{2\,\Delta t}(\| w_N^{k+1} \|_a^2 - \| w_N^k \|_a^2)$$

et de même

$$\mathrm{Re}\,((\partial b_N)^{k+1/2}\,\overline{b}_N^{k+1}) \geqslant \frac{1}{2\,\Delta t}(|\, b_N^{k+1}\,|^2 - |\, b_N^k\,|^2)$$

cela étant, l'erreur due à la discrétisation du temps sera en $O(\Delta t)$ au lieu de $O(\Delta t^2)$. #

3.4. Stabilité et convergence de Crank-Nicholson pour la méthode Tau

De façon analogue au § 1.4, et sous les mêmes hypothèses (1.18) et (1.19), on établirait le résultat de stabilité

$$\| Lw_N^v \|_a^2 \leqslant \| Lw_N^0 \|_a^2 + \sum_{k=0}^{v} \Delta t \, \| L^* g_N^{k+1/2} \|_c^2$$

avec $w_N^k = u_N^k - \tilde{u}_N^k$ où $\tilde{u}_N^k \in U_N$ est donné

$$g_N^{k+1/2} = f_N^{k+1/2} - \tilde{f}_N^{k+1/2}$$
$$\tilde{f}_N^{k+1/2} = (\partial \tilde{u}_N)^{k+1/2} + L\tilde{u}_N^{k+1/2} .$$

On obtient alors une estimation d'erreur abstraite analogue à (1.23), mais contenant en plus un terme $\eta(\Delta t)$ mesurant l'erreur de discrétisation en t, et qui est en $O(\Delta t^2)$.

Les inconvénients cités au § 2.6 sont toujours présents. #

4. ESTIMATIONS SUR LES VALEURS PROPRES

4.1. Problèmes aux valeurs propres

Il faut noter que le spectre de l'opérateur L peut être vide si l'opérateur L n'est pas un opérateur normal (*). (Ceci est justement le cas pour l'exemple considéré au § 2.) En revanche la matrice du problème approché, que nous expliciterons au § 5, a toujours un spectre non vide dans le plan complexe. En vue d'étudier la stabilité de schémas explicites, nous allons donner quelques renseignements sur les valeurs propres du problème approché, qui s'écrit :

Trouver $u_N \in U_N$ *et* $\lambda_N \in \mathbb{C}$ *tels que*

$$(Lu_N, v_N)_\omega = \lambda_N (u_N, v_N)_\omega , \quad \forall v_N \in U_N , \tag{4.1}$$

dans le cas Galerkin, et

$$(Lu_N, v_N)_\omega = \lambda_N (u_N, v_N)_\omega , \quad \forall v_N \in V_N , \tag{4.2}$$

pour la méthode Tau. En utilisant l'opérateur Π_N défini au § 1.1, on peut réécrire les problèmes (4.1) et (4.2) sous la forme

$$Lu_N = \lambda_N \Pi_N u_N . \tag{4.3}$$

(*) *Cf.* Kato, chapitre 3.

4.2. Estimation sur les valeurs propres dans le cas Galerkin

THÉORÈME 4.1 : *Soit* $\lambda_N \in \mathbb{C}$ *une valeur propre du problème* (4.1) *alors, sous les hypothèses* (1.7) *et* (1.8) *on a*

$$\mathrm{Re}\,(\lambda_N) > 0\,.$$

Démonstration : Posons $a_N \equiv (u_N, p_N)_\omega$.
Sous l'hypothèse (1.7), (4.3) entraîne

$$(Lu_N, u_N)_a = \lambda_N(\Pi_N u_N, u_N)_a = \lambda_N[(u_N, u_N)_a + \delta_N a_N \bar{a}_N]\,,$$

d'où, en utilisant (1.8)

$$\alpha \| u_N \|_b^2 \leqslant \mathrm{Re}\,(Lu_N, u_N)_a = \mathrm{Re}\,(\lambda_N)\,[\| u_N \|_a^2 + \delta_N \mid a_N \mid^2]$$

qui équivaut à

$$\mathrm{Re}\,(\lambda_N) \geqslant \frac{\alpha \| u_N \|_b^2}{\| u_N \|_a^2 + \delta_N \mid a_N \mid^2} > 0\,. \tag{4.4}$$

4.3. Cas de la méthode Tau

Soit Π_N la projection sur V_N, on suppose que l'on a $\forall w_N \in U_N$

$$\mathrm{Re}\,(\Pi_N w_N, Lw_N)_a = \alpha \| w_N \|_b^2 + \beta_N \mid b_N \mid^2 \tag{4.5}$$

où $b_N \equiv (w_N, p_N)_\omega$. (Cette hypothèse est plus forte que (1.18) mais correspond plus précisément à ce que l'on a démontré au lemme 2.4, avec $\beta_N = N$.)

THÉORÈME 4.2 : *Soit* $\lambda_N \in \mathbb{C}$ *une valeur propre du problème* (4.2), *alors sous l'hypothèse* (4.5), *on a* :

$$\mathrm{Re}\,(\lambda_N) > 0\,.$$

Démonstration : La relation (4.3) entraîne immédiatement :

$$(Lu_N, Lu_N)_a = \lambda_N(\Pi_N u_N, Lu_N)$$

d'où le résultat désiré puisque

$$\mathrm{Re}\,(\Pi_N u_N, Lu_N) > 0$$

d'après l'hypothèse (4.5). #

Orientation : Le fait que les valeurs propres des problèmes approchés aient leur partie réelle positive rend possible leur résolution par des schémas de Runge-Kutta explicites. Il y aura cependant une condition de stabilité qui dépendra du rayon spectral correspondant.

4.4. Majoration du rayon spectral dans le cas Tchebycheff

Plaçons-nous pour simplifier dans le cas du § 2, où $Lu \equiv \partial u/\partial x$ choix qui correspond à l'équation d'advection (2.1).

Nous choisirons ω égal au poids de Tchebycheff $(1 - x^2)^{-1/2}$. D'après Canuto-Quarteroni, on sait que l'on a l'inégalité inverse :

$$\| Lu_N \|_\omega \leqslant CN^2 \| u_N \|_\omega , \quad \forall u_N \in \mathbb{P}_N . \tag{4.6}$$

Par conséquent, en choisissant $v_N = u_N$ dans (4.1), on a

$$\lambda_N = \frac{(Lu_N, u_N)_\omega}{\| u_N \|_\omega^2}$$

et par conséquent

$$| \lambda_N | \leqslant \frac{\| Lu_N \|_\omega}{\| u_N \|_\omega} \leqslant CN^2 \tag{4.7}$$

qui donne une majoration du rayon spectral pour Galerkin.

Dans le cas de la méthode Tau, on utilisera en premier lieu le résultat suivant qui est valable pour tout poids ω.

LEMME 4.1 : *Soient ω_j^N les coefficients de la formule de Gauss-Radau* (2.4) *à $(N + 1)$ points, et ω_j^{N-1} ceux de la formule à N points.*

Soient $u_N \in U_N$ et $\lambda_N \in \mathbb{C}$ solutions de (4.2), *on a la majoration*

$$| \lambda_N | \leqslant \left(1 - \frac{\omega_0^N}{\omega_0^{N-1}} \right)^{-1/2} \frac{\| Lu_N \|_\omega}{\| u_N \|_\omega} . \tag{4.8}$$

Démonstration : 1° Soit $p \in \mathbb{P}_{N-1}$, quelconque. La formule de Gauss-Radau à N points étant exacte pour les polynômes de degré $2N - 2$, on a puisque $\xi_0^{N-1} = -1$:

$$\int_I | p |^2 \omega \, dX = \sum_{j=0}^{N-1} \omega_j^{N-1} \, | p(\xi_j^{N-1}) |^2 \geqslant \omega_0^{N-1} \, | p(-1) |^2 .$$

Par conséquent, $\forall p \in \mathbb{P}_{N-1}$:

$$| p(-1) |^2 \leqslant \frac{1}{\omega_0^{N-1}} \| p \|_\omega^2 . \tag{4.9}$$

2° Choisissons $v_N \in \mathbb{P}_{N-1}$ tel que

$$v_N(\xi_j^N) = u_N(\xi_j^N), \quad 1 \leqslant j \leqslant N .$$

La formule de Gauss-Radau à $N+1$ points étant exacte pour les polynômes de degré $\leqslant 2N$, on a

$$\| u_N \|_\omega^2 = \sum_{j=1}^{N} \omega_j^N | u_N(\xi_j^N) |^2 = \sum_{j=1}^{N} \omega_j^N u_N(\xi_j^N) \overline{v}_N(\xi_j^N) = (u_N, v_N)_\omega \tag{4.10}$$

et

$$\| v_N \|_\omega^2 = \omega_0^N | v_N(-1) |^2 + \sum_{j=1}^{N} \omega_j^N | v_N(\xi_j^N) |^2$$

c'est-à-dire

$$\| v_N \|_\omega^2 = \omega_0^N | v_N(-1) |^2 + \| u_N \|_\omega^2 .$$

En utilisant (4.9), il vient

$$\| v_N \|_\omega \leqslant \left(1 - \frac{\omega_0^N}{\omega_0^{N-1}}\right)^{-1/2} \| u_N \|_\omega . \tag{4.11}$$

3° D'après (4.2), on a

$$| \lambda_N | = \frac{| (Lu_N, v_N)_\omega |}{| (u_N, v_N)_\omega |} .$$

Or, d'après l'inégalité de Schwarz et (4.11)

$$| (Lu_N, v_N)_\omega | \leqslant \| Lu_N \|_\omega \| v_N \|_\omega \leqslant \left(1 - \frac{\omega_0^N}{\omega_0^{N-1}}\right)^{-1/2} \| Lu_N \|_\omega \| u_N \|_\omega .$$

D'où (4.8), en utilisant (4.10). #

Application : Dans le cas où ω est le poids de Tchebycheff $(1-x^2)^{-1/2}$, on a $\omega_0^N \equiv \dfrac{\pi}{2N+1}$, d'où

$$\left(1 - \frac{\omega_0^N}{\omega_0^{N-1}}\right)^{-1/2} = (2N+1)^{1/2} .$$

D'après l'inégalité inverse (4 8), on a toujours

$$\frac{\| Lu_N \|_\omega}{\| u_N \|_\omega} \leqslant CN^2$$

d'où

$$| \lambda_N | \leqslant CN^2(2N + 1)^{1/2} = O(N^{5/2})$$

qui donne une majoration du rayon spectral pour la méthode Tau #

5. IMPLÉMENTATION DES METHODES PRÉCÉDENTES

5.1. Formulation en termes de méthodes de collocation

Dans le cas de l'équation d'advection (2 1) étudiée au § 2, il est fondamental sur le plan numérique de remarquer que la méthode de Galerkin et la méthode Tau sont équivalentes à des méthodes de collocation utilisant les points d'une formule d'intégration numérique appropriée

THÉORÈME 5 1 (*Méthode de Galerkin*) *Soit* $u_N(t) \in U_N$ *vérifiant*

$$\left(\frac{\partial u_N}{\partial t} + \frac{\partial u_N}{\partial x} - f_N, v_N\right)_\omega = 0\,, \quad \forall v_N \in U_N\,, \tag{5 1}$$

où $f_N \in \mathbb{P}_{N-1}$, *alors on a*

$$\left(\frac{\partial u_N}{\partial t} + \frac{\partial u_N}{\partial x}\right)(\xi_j^N) = f_N(\xi_j^N)\,, \quad 1 \leqslant j \leqslant N \tag{5 2}$$

où $(\xi_j^N)_{0 \leqslant j \leqslant N}$ *sont les points de la formule de Gauss-Radau à* $N + 1$ *points Réciproquement, si* $u_N(t) \in U_N$ *vérifie* (5 1) *il vérifie aussi* (5 2)

Démonstration La relation (5 1) entraîne immédiatement

$$\sum_{j=1}^{N} \omega_j^N \left(\frac{\partial u_N}{\partial t} + \frac{\partial u_N}{\partial x} - f_N\right)(\xi_j^N)\, v_N(\xi_j^N) = 0$$

d'où le résultat puisque l'on peut choisir $v_N(\xi_k^N) = \delta_{jk}$ #

De même, en utilisant le fait que la formule de Gauss à N points est exacte pour les polynômes de degré $\leqslant 2N - 1$, on démontrerait que

THÉORÈME 5 2 (*Méthode Tau*) *Soit* $u_N(t) \in U_N$ *vérifiant*

$$\left(\frac{\partial u_N}{\partial t} + \frac{\partial u_N}{\partial x} - f_N, v_N\right)_\omega = 0\,, \quad \forall v_N \in \mathbb{P}_{N-1}\,, \tag{5 3}$$

alors, on a

$$\left(\frac{\partial u_N}{\partial t} + \frac{\partial u_N}{\partial x}\right)(x_j^N) = f_N(x_j^N), \quad 1 \leqslant j \leqslant N \tag{5.4}$$

où les $(x_j^N)_{1 \leqslant j \leqslant N}$ *sont les points de la formule de Gauss* (2.3) *à N points.* #

(La propriété réciproque est également vraie.)

On notera que, entre (5.2) et (5.4), seuls les points de collocation changent.

5.2. Choix d'une base pour U_N pour la méthode de Galerkin

Soit $(p_N^n)_{1 \leqslant n \leqslant N}$ une base de U_N.

On pose

$$u_N(x, t) = \sum_{n=1}^{N} y_n(t)\, p_N^n(x),$$

et

$$v_N(x) = \sum_{m=1}^{N} z_m\, P_N^m(x).$$

Les z_m étant quelconques, on voit que (5.1) équivaut à

$$\sum_{n=1}^{N} (p_N^n, p_N^m)_\omega \frac{d}{dt} y_n + \sum_{n=1}^{N} \left(\frac{d}{dx} p_N^n, p_N^m\right)_\omega y_n = (f_N, p_N^m)_\omega \tag{5.5}$$

alors que (5.2) équivaut à

$$\sum_{n=1}^{N} p_N^n(\xi_j^N) \frac{d}{dt} y_n + \sum_{n=1}^{N} \left(\frac{d}{dx} p_N^n\right)(\xi_j^N)\, y_n = f_N(\xi_j^N). \tag{5.6}$$

Les systèmes différentiels (5.5) et (5.6) sont bien entendu équivalents, mais pas identiques.

Soit y le vecteur colonne de composantes $(y_n)_{1 \leqslant n \leqslant N}$, ils sont du type

$$M \frac{d}{dt} y + Ky = b \tag{5.7}$$

où M et K sont des matrices $N \times N$.

Pour résoudre ce système différentiel à l'aide de schémas explicites, il y a intérêt à choisir une base p_N^n telle que la matrice M soit *diagonale*.

Dans ce but, il faudrait utiliser une base orthogonale dans le cas de la formulation (5.5). Or celles-ci ne sont pas simples à construire à cause de la condition aux limites contenue implicitement dans U_N.

Au contraire, il suffit de prendre la base des polynômes d'interpolation de Lagrange associés aux points $(\xi_j^N)_{0 \leqslant j \leqslant N}$ pour rendre la matrice M correspondant au système (5.5) diagonale.

Les polynômes de base $p_N^n \in U_N$ seront donc définis par

$$p_N^n(\xi_j^N) = \delta_{jN}, \quad 1 \leqslant j \leqslant N,$$

et par conséquent, les composantes (y_n) de u_N sur la base seront tout simplement les valeurs de u_N aux points (ξ_n^N) :

$$u_N(x) = \sum_{n=1}^{N} u_N(\xi_n^N)\, p_N^n(x).$$

La matrice K sera définie par

$$K_{jn} = \left(\frac{d}{dx} p_N^n\right)(\xi_j^N).$$

Soit λ une valeur propre de K, on a

$$Ky = \lambda y$$

et par conséquent

$$\sum_{n=1}^{N} \left(\frac{d}{dx} p_N^n\right)(\xi_j^N)\, y_n = \lambda \sum_{n=1}^{N} p_N^n(\xi_j^N)\, y_n$$

c'est-à-dire, en posant

$$u_N(x) = \sum_{n=1}^{N} y_n\, p_N^n(x)$$

$$\left(\frac{d}{dx} u_N\right)(\xi_j^N) = \lambda u_N(\xi_j^N).$$

Par conséquent, en multipliant par $\omega_j \bar{v}_N(\xi_j^N)$ et en sommant en j, où $v_N \in U_N$ est quelconque, on obtient :

$$\left(\frac{d}{dx} u_N, v_N\right)_\omega = \lambda (u_N, v_N)_\omega.$$

Par conséquent, les valeurs propres de K sont solutions du problème aux valeurs propres (4.1) étudié au § 4, et on a ainsi directement la propriété

$$\mathrm{Re}\,(\lambda) > 0,$$

qui permet d'envisager l'utilisation de schémas de Runge-Kutta explicites, qui seront stables pour Δt assez petits.

Dans le cas particulier où ω est le poids de Tchebycheff et *d'un schéma de Runge-Kutta explicite d'ordre* 4 (*cf.* Lambert [5]), d'après la majoration du rayon spectral de K, on a la stabilité pour $\Delta t \leqslant \Delta t_{\max} = O(N^{-2})$. #

Remarque 5.1 : Au vu de l'inégalité (4.4), on voit que les schémas de Runge-Kutta explicites d'ordre inférieur à 4 seront également stables pour Δt assez petit, mais il faudrait évaluer l'ordre de

$$\operatorname*{Inf}_{u_N \in U_N} \frac{\alpha \| u_N \|^2_{\omega_2}}{\| u_N \|^2_{\omega_1} + \delta_N | a_N |^2}$$

pour connaître l'ordre du pas de temps maximal admissible dans ces méthodes. (Les schémas de Runge-Kutta ont des domaines de stabilité assez différents suivant l'ordre (*cf.* Lambert, *loc. cit.*) : l'ordre 4 apparaît le plus adapté à des matrices ayant des valeurs propres proches de l'axe imaginaire.)

Remarque 5.2 : *Cas de la méthode Tau* : Pour la méthode Tau on procédera de la même façon et on aura ainsi une alternative analogue dans le choix de la base de U_N.

Les valeurs propres de la matrice du système différentiel correspondant seront solutions du problème aux valeurs propres (4.2) et on aura ainsi des résultats analogues pour la stabilité des méthodes de Runge-Kutta explicites (avec un pas de temps $\Delta t_{\max} = O(N^{-5/2})$) dans le cas où ω est égal au poids de Tchebycheff, et d'une méthode d'ordre 4).

Remarque 5.3 : Le schéma de Crank-Nicholson étudié au § 3 est quant à lui *inconditionnellement stable* mais *implicite.*

Il est immédiat de vérifier que (3.1) équivaut à

$$((\partial u_N)^{k+1/2} + L u_N^{k+1/2})(\xi_j^N) = f_N^{k+1/2}(\xi_j^N) \tag{5.8}$$

En posant

$$u_N^k(x) = \sum_{n=0}^{N} y_n^k \, p_N^n(x)$$

on voit que (5.8) équivaut à

$$\sum_{n=1}^{N} p_N^n(\xi_j^N)(\partial y_n)^{k+1/2} + \sum_{n=1}^{N} \left(\frac{d}{dx} p_N^n\right)(\xi_j^N) \, y_n^{k+1/2} = f_N^{k+1/2}(\xi_j^N)$$

c'est-à-dire à

$$M(\partial y)^{k+1/2} + Ky^{k+1/2} = b^{k+1/2}$$

équivalent au système linéaire :

$$\left(M + \frac{\Delta t}{2} K\right) y^{k+1} = \left(M - \frac{\Delta t}{2} K\right) y^k + \Delta t b^{k+1/2} \tag{5.9}$$

de matrice $M + \frac{\Delta t}{2} K$.

Sans expériences numériques critiques, il est difficile de savoir s'il vaut mieux utiliser le schéma de Crank-Nicholson ou des schémas de Runge-Kutta explicites dans le cas de l'équation d'advection.

Le premier a l'avantage de permettre l'utilisation de pas de temps beaucoup moins petits, ce qui peut être déterminant pour un problème nécessitant la prise en compte de nombreux modes (N grand).

Les autres ont l'avantage d'être plus précis en temps, et de ne pas nécessiter la résolution d'un système linéaire puisque M est diagonale. #

Remarque 5.4 : La formulation en termes de collocation (5.2) au lieu de (5.1) *se prête beaucoup mieux* aux généralisations à des coefficients non constants ou à des problèmes non linéaires.

En revanche, il ne faut pas croire que l'on puisse choisir n'importe comment les points de collocation dans la formule (5.2) et définir ainsi une nouvelle méthode numérique. Par exemple si l'on prend des ξ_j^N *équidistants*, on risque de ne pas pouvoir calculer la matrice K en principe, car le problème de l'interpolation de Lagrange sur de tels points est *mal posé.*

Quant à des résultats de stabilité pour de telles méthodes ils semblent très douteux.

Remarque 5.5 : A ce stade de la description, le poids ω est quelconque. La seule contrainte pour avoir la stabilité est que $\omega_1 = (1 - x)\,\omega$ soit décroissant.

A un poids donné correspondra un jeu de points de collocation donné (les points de Gauss-Radau associés au poids). La réciproque est bien entendu fausse.

5.3. Utilisation de la Transformation de Fourier Rapide

Supposons pour fixer les idées que l'on utilise le schéma de Runge-Kutta le plus simple pour résoudre le système différentiel (5.7), à savoir

$$M \frac{y^{k+1} - y^k}{\Delta t} + Ky^k = b^k$$

d'où

$$y^{k+1} = y^k - \Delta t M^{-1}(Ky^k - b^k).$$

La matrice M étant diagonale, ce qui coûte le plus cher dans le calcul de y^{k+1} à partir de y^k c'est le calcul du produit de la matrice K par le vecteur y^k.

En effet ce calcul implique *en général* N^2 *multiplications et additions, soit* $O(N^2)$ *opérations.*

Pour pouvoir envisager le calcul de solutions pas très régulières, donc impliquant l'utilisation d'un nombre de modes assez élevé ($N \geqslant 30$), il est *fondamental* de pouvoir accélérer cette étape du calcul.

C'est ici que le choix du poids de Tchebycheff joue un rôle capital. En effet les points de Gauss-Radau associés au poids de Tchebycheff sont donnés par

$$\xi_j^N = \cos\left(\frac{2j' + 1}{2N + 1}\pi\right) \quad \text{où} \quad j' = N - j \quad \text{et} \quad j = 0, \ldots, N.$$

Soit $y \in \mathbb{R}^N$ donné et $u_N \in U_N$ la fonction telle que

$$\begin{cases} u_N(\xi_j^N) = y_j, & 1 \leqslant j \leqslant N, \\ u_N(\xi_0^N) = 0. \end{cases}$$

Introduisons les polynômes de Tchebycheff définis par

$$t_n(\cos\theta) = \cos n\theta$$

(qui sont orthogonaux mais pas orthonormés pour le poids $(1 - x^2)^{-1/2}$).

Supposons que l'on ait

$$u_N(x) = \sum_{n=0}^{N} a_n t_n(x)$$

les coefficients a_n vérifieront

$$u_N(\xi_j^N) = \sum_{n=0}^{N} a_n t_n(\xi_j^N), \quad 0 \leqslant j \leqslant N,$$

c'est-à-dire, en posant $\theta_j \equiv \dfrac{2j + 1}{2N + 1}\pi$ et

$$\left.\begin{aligned} z_j &= u_N(\xi_{j'}^N) \\ z_{2N+1-j} &= z_j \end{aligned}\right\} \text{pour} \quad j' = N - j, \ldots \; j = 0, \ldots, N$$

on obtient

$$z_j = \sum_{n=0}^{N} a_n \cos n\theta_j ,$$

qui est valable pour $j = 0, ..., 2N$.

En posant $b_0 = a_0$ et

$$\left.\begin{aligned} b_n &= \frac{1}{2} a_n \\ b_{2N+1-n} &= -\frac{1}{2} a_n \end{aligned}\right\} \text{ pour } \quad n = 1, ..., N$$

on peut réécrire ceci sous la forme

$$z_j = \sum_{n=0}^{2N} b_n e^{in\theta_j}$$

en effet, on a bien

$$\begin{aligned} z_j &= a_0 + \frac{1}{2} \sum_{n=1}^{N} a_n (e^{in\theta_j} - e^{i(2N+1-n)\theta_j}) \\ &= a_0 + \frac{1}{2} \sum_{n=1}^{N} a_n (e^{in\theta_j} + e^{-in\theta_j}) \\ &= \sum_{n=0}^{N} a_n \cos n\theta_j . \end{aligned}$$

En posant $q = e^{\frac{2i\pi}{2N+1}}$, on a

$$z_j = \sum_{n=0}^{2N} q^{n(j+1/2)} b_n .$$

Par conséquent on peut passer des z_j aux $c_n \equiv q^{n/2} b_n$ en utilisant la transformée de Fourier rapide à $2N + 1$ points, car

$$z_j = \sum_{n=0}^{2N} q^{nj} c_n$$

n'est autre que la formule de la transformée de Fourier discrète (*cf.* Auslander-Tolimieri [6]).

Le calcul des c_n et donc des a_n, à partir des z_j ne requiert que $O(N \log_2 N)$ opérations.

Connaissant alors les coefficients a_n du développement de u_N en polynômes de Tchebycheff, on en déduira les coefficients a'_n du développement de $\partial u_N/\partial x$ par la formule ([1]) :

$$a'_n = 2\,S_n \quad \text{pour} \quad 1 \leqslant n \leqslant N\,, \quad \text{et} \quad a'_0 = S_0$$

où S_n est défini par récurrence

$$S_n = S_{n+2} + (n+1)\,a_{n+1} \quad n = N-1\,, N-2, \ldots, 0\,,$$

et

$$S_N = S_{N+1} = 0\,.$$

Le calcul des a'_n à partir des a_n coûte donc $O(N)$ opérations.

Enfin, on calcule les $\dfrac{\partial u_N}{\partial x}(\xi_j^N)$ en utilisant la transformée de Fourier inverse, et on aura $(Ky)_j = \dfrac{\partial u_N}{\partial x}(\xi_j^N)$.

Au total, l'évaluation de Ky coûte $O(N \log_2 N)$ opérations.

Il en résulte un gros avantage pour les schémas de discrétisation *explicites*. Si l'on veut néanmoins utiliser des méthodes implicites du type de Crank-Nicholson, il sera préférable de résoudre, à chaque pas de temps, le système linéaire (5.9), à l'aide d'une *méthode itérative* (du type gradient ou gradient conjugué, avec ou sans opérateur auxiliaire ([2])) ne nécessitant que l'évaluation du produit de la matrice $M + \dfrac{\Delta t}{2} K$ par un vecteur donné, à chaque itération.

On notera que la présence de la condition aux limites $u_N(-1) = 0$ empêche l'utilisation de la transformée de Fourier Rapide dans le cas de la méthode Tau.

Remarque 5.6 : Il existe un autre choix possible que Gauss-Radau ou Gauss pour les points de collocation dans (5.2) ou (5.4), il s'agit des points de Gauss-Lobatto, qui, dans le cas du poids de Tchebycheff, sont donnés par

$$x_j^N = \cos\frac{j\pi}{N}\,, \quad 0 \leqslant j \leqslant N\,.$$

On ne sait pas démontrer de résultat de stabilité analogue à ceux démontrés aux § 1 et 2, pour la méthode de collocation ainsi définie. Cela étant, cette méthode se prête à l'utilisation de la transformée de Fourier Rapide à $2N$ points au lieu de $2N+1$, ce qui peut constituer un avantage déterminant, les transformées de Fourier Rapide à 2^n points étant plus courantes que les transformées de Fourier rapides à 3^n points.

([1]) *Cf.* Gottlieb, Orszag [1], p. 117.
([2]) *Cf.* Glowinski-Lions-Tremolières [7].

6. CONCLUSION

Les deux méthodes, Galerkin et Tau, que l'on a étudiées à la fois d'un point de vue théorique et numérique, se comportent de façon très différente quand on les applique à l'équation d'advection.

La méthode de Galerkin a l'avantage d'être stable dans les conditions les moins sévères, et de se prêter à l'utilisation de la transformée de Fourier Rapide. Deux avantages qui sont déterminants sur le plan numérique

Par rapport aux méthodes de différence finie, elle est beaucoup plus précise pour les solutions régulières, et pas tellement plus coûteuse

Pour les solutions non régulières (par exemple discontinues) il suffit de lisser la solution initiale, pour obtenir des résultats convenables, ne se dégradant pas au cours du temps, contrairement à ce qui se passe dans les schémas de différences finies dissipatifs

Ces conclusions optimistes ne concernent que le cas d'opérateurs linéaires à coefficients constants.

Nos essais numériques actuels ne nous permettent pas d'être aussi optimistes pour les problèmes d'évolution non linéaire avec formation de chocs comme l'équation de Burgers, qui semblent nécessiter l'introduction de procédures de lissage.

REFERENCES

1 D GOTTLIEB, S A ORSZAG, *Numerical Analysis of Spectral methods*, SIAM Regional conferences # 26, 1977

2 P J LAURENT, *Approximation et Optimisation*, Hermann, Paris, 1972

3 P J DAVIS, P RABINOWITZ, *Methods of Numerical Integration*, Academic Press, 1975

4 C CANUTO, A QUARTERONI, a paraître

5 J D LAMBERT, *Computational Methods in Ordinary Differential Equations*, John Wiley & Sons, New York, 1973

6 L AUSLANDER, R TOLIMIERI, *Is computing with the finite Fourier Transform pure or applied Mathematics ?* Bull (New Series) AMS, 1,6 (1979) 847-898

7 R GLOWINSKI, J L LIONS, R TREMOLIERES, *Analyse Numérique des Inéquations Variationnelles*, Dunod, 1976