

MICHEL CROUZEIX

**Étude de la stabilité des méthodes de Runge-Kutta appliquées
aux équations paraboliques**

Publications des séminaires de mathématiques et informatique de Rennes, 1974, fascicule S4

« Journées éléments finis », , p. 1-6

http://www.numdam.org/item?id=PSMIR_1974__S4_A3_0

© Département de mathématiques et informatique, université de Rennes, 1974, tous droits réservés.

L'accès aux archives de la série « Publications mathématiques et informatiques de Rennes » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

ETUDE DE LA STABILITE DES METHODES DE RUNGE-KUTTA
APPLIQUEES AUX EQUATIONS PARABOLIQUES.

par Michel CROUZEIX

Laboratoire d'Analyse Numérique
Tour 55-65 - 5ème étage
Université de Paris 6
4, place Jussieu
75230 - PARIS CEDEX 05

1. - Position du problème.

Nous désignerons par V et H deux espaces de Hilbert satisfaisant aux hypothèses : V est inclus dans H , V est dense dans H et l'injection canonique de V dans H est continue. Nous noterons $\| \cdot \|$ la norme dans V , $(\cdot, \cdot)$ et $|\cdot|$ respectivement le produit scalaire et la norme dans H . Nous identifierons H à son dual, ce qui nous permet d'identifier H à un sous-espace dense d'un dual V' de V et de noter encore par $(\cdot, \cdot)$ le produit scalaire mettant en dualité V' et V . Nous obtenons ainsi le triplet :

$$V \subset H \subset V'$$

Nous nous donnons maintenant un opérateur A linéaire continue de V dans V' , vérifiant l'hypothèse d'ellipticité : il existe une constante $\alpha > 0$ telle que :

$$\operatorname{Re} (Au, u) \geq \alpha \|u\|^2 \quad \forall u \in V$$

Nous nous intéressons à l'approximation du problème suivant : Etant donné $f(t)$ et g trouver $u(t) \in V$ telle que :

$$(1) \quad \begin{cases} u'(t) + A u(t) = f(t) & \forall t \in]0, T[\\ u(0) = g \end{cases}$$

La variable t intervenant ici sera appelée variable de temps.

2. - Le schéma d'approximation.

a) Approximation en espace.

Nous nous donnons une méthode d'éléments finis avec intégration numérique que nous décrirons abstraitement par : à chaque valeur d'un paramètre h , nous associerons :

- un espace de dimension finie $V_h \subset V$, servant à approcher les espaces V , H et V' .
- un produit scalaire $(\cdot, \cdot)_h$, approximation du produit scalaire $(\cdot, \cdot)$ de H .
- un opérateur $A_h \in \mathcal{L}(V_h, V_h)$, approximation de l'opérateur A .
- pour tout $t \in]0, T[$ un élément $f_h(t)$ de V_h , approximation de $f(t)$.

Nous supposerons qu'il existe deux constantes c et $\tilde{\alpha} > 0$, indépendantes de h , telles que :

$$(2) \quad \begin{cases} c|u_h| \leq |u_h|_h \leq \frac{1}{c} |u_h| \\ \operatorname{Re} (A_h u_h, u_h)_h \geq \tilde{\alpha} \|u_h\|^2 \end{cases} \quad \forall u_h \in V_h$$

où $|\cdot|_h$ désigne la norme associée au produit scalaire $(\cdot, \cdot)_h$.

Etant donné un opérateur $B_h \in \mathcal{L}(V_h, V_h)$, nous désignerons par $\|B_h\|_h$ la norme d'opérateur définie par :

$$(3) \quad \|B_h\|_h = \sup_{|v_h|_h \leq 1} |B_h v_h|_h$$

b) Approximation en temps.

Nous allons utiliser une méthode de Runge-Kutta (en général implicite) ; pour cela nous nous donnons q nombres réels $\tau_1, \tau_2, \dots, \tau_q$ et les formules de quadrature approchée :

$$(4) \quad \begin{cases} \int_0^{\tau_i} g(t) dt \approx \sum_{j=1}^q a_{ij} g(\tau_j) \\ \int_0^1 g(t) dt \approx \sum_{j=1}^q b_j g(\tau_j) \end{cases} \quad 1 \leq i \leq q$$

Nous nous donnons d'autre part un pas de discrétisation Δt et nous posons $t_n = n \Delta t$ et $t_{n,i} = t_n + \tau_i \Delta t$.

La solution $u(t)$ du problème (1) vérifie :

$$u(t_{n,i}) = u(t_n) - \int_{t_n}^{t_{n,i}} (Au(t) - f(t)) dt$$

Nous avons donc les équations approchées :

$$\begin{cases} u(t_{n,i}) \approx u(t_n) - \Delta t \sum_{j=1}^q a_{ij} (Au(t_{n,j}) - f(t_{n,j})) \\ u(t_{n+1}) \approx u(t_n) - \Delta t \sum_{j=1}^q b_j (Au(t_{n,j}) - f(t_{n,j})) \end{cases}$$

Nous sommes ainsi amenés à considérer le schéma d'approximation :

$u_h^0 \in V_h$ est donné ; u_h^{n+1} est défini à partir de u_h^n par les équations :

$$(5) \quad \begin{cases} u_h^{n,i} = u_h^n - \Delta t \sum_{j=1}^q a_{ij} (A_h u_h^{n,j} - f_h(t_{n,j})) \quad 1 \leq i \leq q \\ u_h^{n+1} = u_h^n - \Delta t \sum_{j=1}^q b_j (A_h u_h^{n,j} - f_h(t_{n,j})) \end{cases}$$

Nous allons d'abord chercher sous quelles conditions le schéma (5) admet une solution unique ; nous étudierons ensuite la stabilité de ce schéma.

Nous noterons $\mathcal{A}$ la matrice d'élément générique a_{ij} : $\mathcal{A} = (a_{ij}) \in \mathbb{R}^{q \times q}$

b le vecteur $\in \mathbb{R}^q$ d'élément générique b_i :

$$b = \begin{pmatrix} b_1 \\ \vdots \\ b_q \end{pmatrix}$$

e le vecteur $\in \mathbb{R}^q$ d'élément générique $e_i = 1$:

$$e = \begin{pmatrix} 1 \\ \vdots \\ 1 \end{pmatrix}$$

3 - Existence de la solution du schéma.

Théorème 1. Si les valeurs propres de $\mathcal{A}$ ont leurs parties réelles supérieures ou égales à zéro, alors le schéma (5) admet une solution unique.

Démonstration :

Posons : $U_h^n = \begin{pmatrix} u_h^{n,1} \\ \vdots \\ u_h^{n,q} \end{pmatrix} \quad \bar{U}_h^n = \begin{pmatrix} u_h^n \\ \vdots \\ u_h^n \end{pmatrix} \quad F_h^n = \begin{pmatrix} f_h(t_{n,1}) \\ \vdots \\ f_h(t_{n,q}) \end{pmatrix}$

et notons par $A_h \mathcal{A}$ la matrice appartenant à $\mathcal{L}(V_h, V_h)^{q \times q}$ d'élément générique $a_{ij} A_h$. Avec ces notations le schéma (5) s'écrit :

$$(6) \quad \begin{cases} (I + \Delta t A_h \mathcal{A}) U_h^n = \bar{U}_h^n + \Delta t \mathcal{A} F_h^n \\ u_h^{n+1} = u_h^n - \Delta t \sum_{j=1}^q b_j (A_h u_h^{n,j} - f_h(t_{n,j})) \end{cases}$$

Il existe une matrice $\mathcal{K}$ inversible et une matrice $\tilde{\mathcal{A}}$ triangulaire inférieure telle que : $\mathcal{A} = \mathcal{K}^{-1} \tilde{\mathcal{A}} \mathcal{K}$

Nous avons donc : $I + \Delta t A_h \mathcal{A} = \mathcal{K}^{-1} (I + \Delta t A_h \tilde{\mathcal{A}}) \mathcal{K}$

L'élément générique diagonal de la matrice triangulaire $I + \Delta t A_h \tilde{Q}$ est : $I + \Delta t \tilde{a}_{ii} A_h$ où $\operatorname{Re} \tilde{a}_{ii} \geq 0$. Cet élément est inversible puisque les valeurs propres de A_h ont leurs parties réelles strictement positives. L'opérateur $I + \Delta t A_h \tilde{Q}$ est donc inversible ; il en est donc de même pour l'opérateur $I + \Delta t A_h Q$, d'où le théorème.

Remarque : Dans tous les cas on peut démontrer l'existence et l'unicité d'une solution au schéma (5) sous l'hypothèse :

$$\Delta t \rho(Q) \rho(A_h) < 1$$

(où $\rho(T)$ désigne le rayon spectral de l'opérateur T)

4 - Résultats fondamentaux pour l'étude de la stabilité .

Notre outil fondamental d'étude de la stabilité sera le théorème suivant dû à J. von Neumann. On pourra en trouver une démonstration dans le livre de F. Riesz et B. Sz. Nagy.

Théorème 2. Soit H un espace de Hilbert, B un opérateur linéaire continu de H sur H et $f(z)$ une fonction holomorphe au voisinage du disque $\{z \in \mathbb{C} \mid |z| \leq \|B\|_{\mathcal{L}(H,H)}\}$, alors $f(B)$ vérifie :

$$\|f(B)\|_{\mathcal{L}(H,H)} \leq \sup_{|z| \leq \|B\|_{\mathcal{L}(H,H)}} |f(z)|$$

Nous considérons maintenant un opérateur A (en général non borné) linéaire sur H . Nous supposons que cet opérateur est fermé, que son domaine $D(A)$ est dense dans H et qu'il existe un nombre réel a tel que :

$$\operatorname{Re} (Au, u) \geq a \|u\|_H^2 \quad \forall u \in D(A)$$

Le théorème de von Neumann nous permet de démontrer :

Théorème 3. Soit $r(z)$ une fraction rationnelle de la variable z , bornée dans le demi plan $\{z \in \mathbb{C} \mid \operatorname{Re} z \geq a\}$, alors $r(A)$ est un opérateur linéaire continu de H sur H et on a :

$$\|r(A)\|_{\mathcal{L}(H,H)} \leq \sup_{\operatorname{Re} z \geq a} |r(z)|$$

Démonstration :

$$\text{Posons : } B = (I - (A - aI)) (I + (A - aI))^{-1}$$

$$\text{et } f(z) = r(a + \frac{1-z}{1+z})$$

Les hypothèses faites entraînent :

$$B \in \mathcal{L}(H,H) \quad , \quad \|B\|_{\mathcal{L}(H,H)} \leq 1$$

$f(z)$ est holomorphe dans un voisinage du disque $\{z \in \mathbb{C} \mid |z| \leq 1\}$

d'autre part : $r(A) = f(B)$

$$\text{et } \max_{|z| \leq 1} |f(z)| = \sup_{\operatorname{Re} z \geq a} |r(z)|$$

On en déduit le théorème 3 par application du théorème 2.

5 - Etude de la stabilité du schéma.

D'après (6) le schéma (5) s'écrit :

$$u_h^{n+1} = (I - \Delta t A_h^t b (I + \Delta t A_h Q)^{-1} e) u_h^n + \Delta t^t b (I + \Delta t A_h Q)^{-1} F_h^n$$

Posons :

$$(7) \quad \begin{cases} r(z) = 1 - z^t b(I+zQ)^{-1} e \\ f_h^n = {}^t b(I+\Delta t A_h Q)^{-1} F_h^n \end{cases}$$

Le schéma (5) s'écrit maintenant sous la forme condensée :

$$(8) \quad u_h^{n+1} = r(\Delta t A_h) u_h^n + \Delta t f_h^n$$

Définitions :

* Nous dirons que le schéma (5) est faiblement A-stable ssi :

$$\forall z \in \mathbb{C} \text{ vérifiant } \operatorname{Re} z \geq 0, \text{ on a } |r(z)| \leq 1$$

* Nous dirons que le schéma (5) est fortement A-stable ssi :

$$\begin{cases} \forall z \in \mathbb{C} \text{ vérifiant } \operatorname{Re} z > 0, \text{ on a } |r(z)| < 1 \\ \text{et } \lim_{z \rightarrow \infty} |r(z)| < 1 \end{cases}$$

* Nous dirons que le schéma (5) est consistant (avec le problème (1))

ssi : $\sum_{j=1}^q b_j = 1$
(c'est à dire ssi la formule $\int_0^1 g(t) dt \approx \sum_{j=1}^q b_j g(\tau_j)$ est exacte pour les fonctions $g(t) = \text{constante}$).

La stabilité du schéma (5) est donnée par le théorème :

Théorème 4.

a) Si le schéma (5) est faiblement A-stable, on a $\forall \Delta t \geq 0$:

$$\|r(\Delta t A_h)\|_h \leq 1$$

b) Si le schéma (5) est fortement A-stable, alors $\forall \Delta t_0 > 0$

il existe une constante $a > 0$ telle que $\forall \Delta t$ vérifiant $0 < \Delta t < \Delta t_0$:

$$\|r(\Delta t A_h)\|_h \leq e^{-a\Delta t}$$

c) Si le schéma (5) est consistant, il existe une constante

$\gamma > 0$ (ne dépendant que de Q et de b) telle que pour tout Δt vérifiant :

$$0 < \Delta t \rho \left[\left(\frac{A_h + A_h^*}{2} \right)^{-1} A_h^* A_h \right] \leq \gamma$$

on ait :

$$\|r(\Delta t A_h)\|_h \leq 1$$

Démonstration :

a) et b)

On a, d'après (2) et d'après la continuité de l'injection canonique de V dans H :

$$\operatorname{Re} (A_h u_h, u_h)_h \geq \tilde{\alpha} \|u_h\|^2 \geq C |u_h|_h^2 \quad \text{avec } C > 0$$

Nous déduisons donc du théorème 3 que :

$$\|r(\Delta t A_h)\|_h \leq \sup_{\operatorname{Re} z \geq C\Delta t} |r(z)|$$

On obtient ainsi immédiatement le a) du théorème 4. Sous l'hypothèse b) on peut démontrer l'existence d'une constante $a > 0$ telle que :

$$|r(z)| \leq e^{-a\Delta t} \quad \text{pour tout } z \text{ vérifiant } \operatorname{Re} z \geq C\Delta t,$$

d'où le résultat.

c) Le schéma étant consistant, on a $r(0) = 1$ et $r'(0) = -1$. $r(z)$ étant holomorphe au voisinage de 0, on en déduit l'existence d'une constante $\gamma > 0$ telle que :

$$|r(z)| \leq 1 \text{ pour tout } z \text{ vérifiant } |z - \frac{\gamma}{2}| \leq \frac{\gamma}{2}$$

La condition imposée à Δt entraîne :

$$\|(A_h - \frac{\gamma}{2} I)\|_h \leq \frac{\gamma}{2}$$

On en déduit le théorème 4 par application du théorème de von Neumann. ■

Associations maintenant au schéma (5) un schéma perturbé vérifiant :

$$(9) \quad v_h^{n+1} = r(\Delta t A_h) v_h^n + \Delta t f_h^n + \Delta t e_h^n$$

Sous les conditions a), b) ou c) du théorème 4, nous obtenons la majoration

$$|u_h^n - v_h^n|_h \leq |u_h^0 - v_h^0|_h + \Delta t \sum_{k=0}^{n-1} |e_h^k|_h$$

$$\text{d'où} \quad |u_h^n - v_h^n|_h \leq |u_h^0 - v_h^0|_h + t_n \max_{0 \leq k \leq n-1} |e_h^k|_h$$

Sous l'hypothèse b) du théorème 4 nous avons de plus :

$$|u_h^n - v_h^n|_h \leq e^{-at_n} |u_h^0 - v_h^0|_h + \Delta t \sum_{k=0}^{n-1} e^{-a(t_n - t_{k+1})} |e_h^k|_h$$

$$\text{Nous avons donc la majoration : } |u_h^n - v_h^n|_h \leq e^{-a t_n} |u_h^0 - v_h^0|_h + \frac{e^{a \Delta t}}{a} \max_{0 \leq k \leq n-1} |e_h^k|_h$$

6 - Exemple.

Le schéma défini par $q = 2$ et par :

$$(10) \quad \begin{cases} \tau_1 = \frac{2 + \sqrt{2}}{3 - \sqrt{3}} & \tau_2 = \frac{2 - \sqrt{2}}{3 - \sqrt{3}} \\ b_1 = \frac{4 - \sqrt{2} - \sqrt{6}}{8} & b_2 = \frac{4 + \sqrt{2} + \sqrt{6}}{8} \\ a = \frac{\sqrt{3}+1}{4\sqrt{6}} \begin{bmatrix} 2\sqrt{2}+1 & 2\sqrt{2}+3 \\ 2\sqrt{2}-3 & 2\sqrt{2}-1 \end{bmatrix} \end{cases}$$

est fortement A-stable et d'ordre 3. En faisant le changement de variable

$$\begin{aligned} v_h^{n,1} &= \frac{\sqrt{2}-1}{2\sqrt{2}} u_h^{n,1} + \frac{\sqrt{2}+1}{2\sqrt{2}} u_h^{n,2} \\ v_h^{n,2} &= \frac{\sqrt{2}+1}{2} u_h^{n,1} - \frac{\sqrt{2}-1}{2} u_h^{n,2} \end{aligned} \quad \text{et en posant } c = \frac{1}{2} + \frac{1}{2\sqrt{3}}$$

ce schéma s'écrit :

$$\begin{cases} (I + c \Delta t A_h) v_h^{n,1} = u_h^n + \frac{\sqrt{3}+1}{8\sqrt{3}} \Delta t ((2-\sqrt{2})f(t_{n,1}) + (2+\sqrt{2})f(t_{n,2})) \\ (I + c \Delta t A_h) v_h^{n,2} = u_h^n + 3(v_h^{n,1} - u_h^n) + \frac{\sqrt{3}+1}{4\sqrt{3}} \Delta t ((\sqrt{2}+1)f(t_{n,1}) - (\sqrt{2}-1)f(t_{n,2})) \\ u_h^{n+1} = u_h^n + (\frac{5}{\sqrt{3}} - 1)(v_h^{n,1} - u_h^n) - (\frac{2}{\sqrt{3}} - 1)(v_h^{n,2} - u_h^n) \end{cases}$$

Bibliographie :

- AXELSSON O. "A class of A-stable methods "
BIT,9, p.185-189 (1969)
- BUTCHER J.C. "Implicit Runge-Kutta processes "
Math.Comp.18, p.50-64 (1964)
"Integration processes based on Radau quadrature formulas"
Math.Comp.26, p.233-244 (1964)
- DAHLQUIST G. "A special stability problem for linear multistep methods"
BIT,3, p.27-43 (1963)
- GEAR C.W. "Numerical initial value problems in ordinary differential
equations" Prentice Hall Inc. (1971)
- LAMBERT J.D. "Computational methods in ordinary differential equations"
John Wiley and Sons, London N-Y. (1973)
- RIESZ F. et Sz.NAGY B. "Leçons d'Analyse fonctionnelle"
Budapest (1952)
- RAVIART P.A. "Cours de troisième cycle"
Université de Paris 6 (1971-1972)
- STRANG G. & FIX G. "An analysis of the finite element method"
Prentice Hall (1973)
- CIARLET P. & RAVIART P.A. "La méthode des éléments finis et les problèmes
aux limites elliptiques" à paraître