

AN A-STABLE THREE-LEVEL METHOD FOR THE GALERKIN SOLUTION OF QUASILINEAR PARABOLIC PROBLEMS

I. FARAGÓ – A. GALÁNTAI

Abstract. Parabolic partial differential equations are often solved by semidiscrete Galerkin methods. These methods first make a finite element discretization in the space variables reducing the problem to solution of the Cauchy-problem of a system of ordinary differential equations. This problem is then solved by a highly stable difference or Runge-Kutta method.

In this paper we investigate the numerical solution of the semidiscretized Cauchy-problem for a class of nonlinear partial differential equations used in the modelling of chemical reactors and other areas.

1. Introduction

Consider the following initial-boundary value problem

$$(1) \quad \frac{\partial u}{\partial t} = Pu, \quad ((x, t) \in \Omega \times (0, T])$$

$$(2) \quad Bu = 0 \quad ((x, t) \in \Gamma \times (0, T])$$

$$(3) \quad u(x, 0) = u_0(x) \quad (x \in \bar{\Omega}).$$

where $x = (x_1, x_2, \dots, x_N) \in \Omega \subset R^N$, $t \in (0, T]$, $T > 0$, Ω is a bounded domain with a sufficiently smooth boundary. The operator P is defined by

$$(4) \quad Pu = \sum_{i,j=1}^N \frac{\partial}{\partial x_i} (F_{ij}(x) \frac{\partial u}{\partial x_j}) - F_0(u, x, t).$$

The operator B is representing some classical boundary condition and $u_0(x)$ is given function. Assume that the following conditions are satisfied.

- (i) There is only one generalized solution of the problem (1)-(3) in $H^1(\Omega)$ for arbitrary fixed $t \in (0, T]$ (see Ladizhenskaya [13] and Lions [7]).
- (ii) For arbitrary fixed $x \in \bar{\Omega}$ the matrix $[F_{ij}(x)]_{i,j=1}^N$ is symmetric and uniformly positive definite, that is there exist such positive numbers k_0, k_1 that for any vector $\xi = (\xi_1, \xi_2, \dots, \xi_N) \in R^N$ the inequality

$$(5) \quad k_0 \sum_{i=1}^N \xi_i^2 \leq \sum_{i,j=1}^N F_{ij}(x) \xi_i \xi_j \leq k_1 \sum_{i=1}^N \xi_i^2$$

holds.

- (iii) The function F_0 is continuous and uniformly Lipschitzian in its first variable, that is there exists a constant $L_0 > 0$ such that

$$(6) \quad |F_0(s_1, x, t) - F_0(s_2, x, t)| < L_0 |s_1 - s_2|$$

is satisfied for all $s_1, s_2 \in R$ and for all $(x, t) \in \bar{\Omega} \times [0, T]$.

We also assume that there exists a space V_h^l of finite elements which is a finite dimensional subspace of $H^1(\Omega)$ and it satisfies the approximation property that for given $l \in \mathbf{N}^+$ and an arbitrary $u \in H^1(\Omega) \cap H^{l+1}(\bar{\Omega})$ there is an element $\tilde{u} \in V_h^l$ such that

$$(7) \quad \|u - \tilde{u}\|_0 + h\|u - \tilde{u}\|_1 \leq ch^{l+1}\|u\|_{l+1},$$

where c is a positive constant and h is the maximal diameter of the discretization. (For the existence of such spaces of finite elements see Strang [8], Molchanov [14], Faragó [3].)

Let $V_h^l = \text{span}[\varphi, \dots, \varphi_n]$ and seek the approximate solution in the following form

$$(8) \quad u_n(x, t) = \sum_{i=1}^n \alpha_i(t) \varphi_i(x)$$

Then for the unknown vector $\alpha(t) = [\alpha_1(t), \dots, \alpha_n(t)]$ we have the Cauchy-problem of the form

$$(9) \quad M\alpha' + Q\alpha = F(\alpha, t) \quad (0 < t \leq T)$$

$$(10) \quad M\alpha(0) = \tilde{\alpha}_0,$$

where M and Q are positive definite matrices, $F : R^{n+1} \rightarrow R^n$ and the initial value $\tilde{\alpha}_0$ are also given (see [3], [8]). Hence the matrix M^{-1} exists and the problem (9)-(10) equivalent with the problem

$$(11) \quad \alpha' = A\alpha + f(\alpha, t)$$

$$(12) \quad \alpha(0) = \alpha_0.$$

It can be shown ([8], [5]) that the problem (11)-(12) satisfies the following two properties

- (iv) The matrix A has only negative eigenvalues and it is diagonalizable.
- (v) The map $f : R^{n+1} \rightarrow R^n$ is continuous and there exists a constant $L \geq 0$ such that

$$(13) \quad \|f(y_1, t) - f(y_2, t)\| \leq L\|y_1 - y_2\| \quad (y_1, y_2 \in R^n)$$

holds for all $t \in [0, T]$.

It is known ([1], [2], [11]) that the problem (11)-(12) can be considered as a stiff system. Therefore we need to choose a highly stable method to solve it.

In this paper we are going to investigate linear two-step methods. Two-step methods of the form (ς, σ) can be defined by their characteristic polynomials

$$(14) \quad \varsigma(\xi) = \sum_{s=0}^2 \varsigma_s \xi^s; \quad \sigma(\xi) = \sum_{s=0}^2 \sigma_s \xi^s,$$

where $\sigma(1) = 1$ is assumed. The method (ς, σ) is of order two if the coefficients satisfy the following conditions:

$$(15) \quad \begin{aligned} \varsigma_1 &= 1 - 2\varsigma_2, & \varsigma_0 &= -1 + \varsigma_2, \\ \sigma_1 &= \frac{1}{2} + \varsigma_2 - 2\sigma_2, & \sigma_0 &= \frac{1}{2} - \varsigma_2 + \sigma_2 \end{aligned}$$

DEFINITION 1. Consider the test problem

$$(16) \quad y' = \lambda y, \quad y(0) = y_0 \quad (\lambda \in \mathbb{C})$$

Let $z = \lambda\tau$ ($\tau > 0, \lambda \in \mathbb{C}$ and denote by y_m the numerical solution of (16) at the point $t_m = m\tau$ ($m \geq 0$). The set S of those values of z for which $\{y_m\}_{m=0}^{\infty}$ converges to 0 for all y_0 is said to be the region of absolute stability of the method.

DEFINITION 2. A method is said to be A_0 -stable if $(-\infty, 0) \subset S$. The method is said to be A -stable if $\mathbb{C}^- \subset S$, where $\mathbb{C}^- = \{z | \operatorname{Re}(z) < 0\}$.

Applying the method (14) to the problem (11)-(12) we obtain the recursion

$$(17) \quad \sum_{s=0}^2 \varsigma_s \alpha^{m+s} = \tau A \left(\sum_{s=0}^2 \sigma_s \alpha^{m+s} \right) + \tau \sum_{s=0}^2 \sigma_s f(\alpha^{m+s}, t_{m+s})$$

which is nonlinear systems of algebraic equations for the unknown vector α^{m+2} (α^m is the approximation of $\alpha(t)$ on the time level $t_m = m\tau$). In order to avoid the solution of this nonlinear system we linearize it using axtrapolation ([10]). Keeping the accuracy of the scheme we change t_{m+2} and α^{m+2} to $\overline{t_{m+2}}$ and α^{m+2} , where

$$(18) \quad \begin{aligned} \overline{t_{m+2}} &= t_m + (2\sigma_2 + \sigma_1)\tau, \\ \alpha^{m+2} &= (2\sigma_2 + \sigma_1)\alpha^{m+1} + (\sigma_0 - \sigma_2)\alpha^m \end{aligned}$$

are extrapolated values. Substituting (18) into the righthand side of (17) one obtains a linear algebraic system.

2. A special two step method and its stability

We derive a one-parameter class of two-step methods which are based on the extrapolation principle mentioned in the previous section. The method must be of order 2. We start with the standard two-step methods and choose the parameters as follows

$$(19) \quad \varsigma_2 = \frac{1}{2}, \quad \sigma_2 = \Theta \quad (\Theta \in R).$$

Then we have the following coefficients

$$(20) \quad \begin{aligned} \varsigma_0 &= -\frac{1}{2}, \quad \varsigma_1 = 0, \quad \varsigma_2 = \frac{1}{2} \\ \sigma_0 &= \Theta, \quad \sigma_1 = 1 - 2\Theta, \quad \sigma_2 = \Theta. \end{aligned}$$

If we apply this scheme to the problem (11)-(12) and use the notations

$$(21) \quad \alpha^{m+1,\Theta} = \Theta\alpha^{m+2} + (1 - 2\Theta)\alpha^{m+1} + \Theta\alpha^m,$$

$$(22) \quad \begin{aligned} f^{m+1,\Theta} &= \Theta f(\alpha^{m+2}, t_{m+2}) + (1 - 2\Theta)f(\alpha^{m+1}, t_{m+1}) + \\ &+ \Theta f(\alpha^m, t_m) \end{aligned}$$

then we obtain the recursion

$$(23) \quad \frac{\alpha^{m+2} - \alpha^m}{2\tau} = A\alpha^{m+1,\Theta} + f^{m+1,\Theta} \quad (m \geq 0).$$

The extrapolation formula(18) has now the form

$$(24) \quad t_{\overline{m+2}} = t_m + \tau = t_{m+1}, \quad \alpha^{\overline{m+2}} = \alpha^{m+1}$$

and $f^{m+1,\Theta} \sim f(t_{m+1}, \alpha^{m+1})$ for small τ 's. Thus we have the following linearized form of scheme (23)

$$(25) \quad \frac{\alpha^{m+2} - \alpha^m}{2\tau} = A\alpha^{m+1,\Theta} + f(t_{m+1}, \alpha^{m+1})$$

$(m = 0, 1, \dots)$.

Next we investigate the stability properties of the method (25). If we apply (25) to the test problem (16) then we get the difference equation

$$(26) \quad \frac{y^{m+2} - y^m}{2\tau} = \lambda y^{m+1,\Theta} \quad (m \geq 0)$$

which is equivalent to the recursion

$$(27) \quad (1 - 2\Theta z)y^{m+2} - 2(1 - 2\Theta)zy^{m+1} - (1 + 2\Theta z)y^m = 0$$

The characteristic equation of (27) has the form

$$(28) \quad \Pi(\xi, z) = (1 - 2\Theta z)\xi^2 - 2(1 - 2\Theta)z\xi - (1 + 2\Theta z) = 0.$$

Theorem 1. *The method (25) is A_0 -stable iff $\Theta > \frac{1}{4}$.*

Proof. The solution of (27) is tending to 0 for all y_0 if the zeros of the characteristic equation (28) lie in the disk $\{w \in \mathbb{C} : |w| < 1\}$. Hence it is enough to show that for all $z < 0$ this

condition is satisfied if and only if $\Theta > \frac{1}{4}$. If $z < 0$ the coefficients of (28) are real and we can use a special case of the Schur-Cohn criterion (Kobza [6]): A real polynomial of the form

$$(29) \quad a_2 x^2 + a_1 x + a_0 \quad (a_2 > 0)$$

has it both zeros in the open unit disk if and only if the coefficients satisfy the system

$$(30) \quad a_2 + a_1 + a_0 > 0, \quad a_2 - a_0 > 0, \quad a_2 - a_1 + a_0 > 0.$$

Using (30) one can show that the zeros of (28) lie in the open unit disk for all $z < 0$ if and only if $\Theta > \frac{1}{4}$.

Theorem 2. *The method (25) is A-stable if $\Theta > \frac{1}{4}$.*

Proof It is enough to show that for all $z \in \mathbb{C}^-$ the zeros of (28) lie in the open unit disk. Consider the Moebius-transformation

$$(31) \quad \zeta = (p+1)/(p-1)$$

which maps the open unit disk onto the half plane $\operatorname{Re}(p) < 0$, that is $|\zeta| < 1$ iff $\operatorname{Re}(p) < 0$. Furthermore, let be $q = -z$ and consider the equation

$$(32) \quad H(p, q) = (p-1)^2 \Pi((p+1)/(p-1), -q) = 0.$$

The method is A-stable if and only if $\operatorname{Re}(q) > 0$ implies $\operatorname{Re}(p(q)) < 0$ where $p(q)$ denotes any zero of (32) as a function of q . Similarly, if $q(p)$ denotes the zero of (32) as a function of p then the latter condition is equivalent to the condition, that $\operatorname{Re}(p) \geq 0$ implies $\operatorname{Re}(q(p)) \leq 0$. By elementary calculations one gets

$$(33) \quad H(p, q) = 4p + 2q(p^2 + 4\Theta - 1) = 0.$$

It is easy to see that for $p^2 = 1 - 4\Theta$ there is no zero of (33). Hence we can solve (33) in the form

$$(34) \quad q = -2p/(p^2 + 4\Theta - 1) = -2[\bar{p}|p|^2 + (4\Theta - 1)p]/|p^2 + 4\Theta - 1|^2.$$

For $\Theta > \frac{1}{4}$ and $Re(p) \geq 0$ we have

$$Re[\bar{p}|p|^2 + (4\Theta - 1)p] \geq 0$$

which implies $Re(q(p)) \leq 0$. Thus the theorem is proved.

It is noted that A-stability implies A_0 -stability. This is the reason for the assumption $\Theta > \frac{1}{4}$ of Theorem 2. It is also worth noting that the method (25) is I -stable for $\Theta > \frac{1}{4}$ which also implies the A-stability (see Wanner-Hairer-Norsett [9]).

3. The convergence of the method

We construct an error estimation for the global error from which the convergence of the algorithm follows.

Rewrite the method (25) in the form

$$(35) \quad \alpha^{m+2} - \alpha^m = 2\tau[A\alpha^{m+1,\Theta} + f(t_{m+1}, \alpha^{m+1})]$$

and let $\hat{\alpha}^m = \alpha(t_m)$, where $\alpha(t)$ is the exact solution of (11)-(12). The local error of the method at $t_m = m\tau$ is defined by

$$(36) \quad \begin{aligned} \hat{\alpha}^{m+2} - \hat{\alpha}^m &= 2\tau[A(\Theta\hat{\alpha}^{m+2} + (1 - 2\Theta)\hat{\alpha}^{m+1} + \Theta\hat{\alpha}^m) + \\ &\quad + 2\tau f(t_{m+1}, \hat{\alpha}^{m+1}) + T_m]. \end{aligned}$$

Introduction the notation $e_m = \alpha^m - \hat{\alpha}^m$ for the global error of the method at the point t_m we obtain

$$(37) \quad e_{m+2} - e_m = 2\tau A[\Theta e_{m+2} + (1 - 2\Theta)e_{m+1} + \Theta e_m] + R_m,$$

where

$$(38) \quad R_m = 2\tau[f(t_{m+1}, \alpha^{m+1}) - f(t_{m+1}, \hat{\alpha}^{m+1})] - T_m.$$

By simple calculation one has

$$(39) \quad (I - 2\Theta\tau A)e_{m+2} - 2(1 - 2\Theta)\tau Ae_{m+1} - (I + 2\Theta\tau A)e_m = R_m$$

and

$$(40) \quad \begin{aligned} e_{m+2} = & 2(1 - 2\Theta)(I - 2\Theta\tau A)^{-1}\tau A e_{m+1} + \\ & + (I - 2\Theta\tau A)^{-1}(I + 2\Theta\tau A)e_m + (I - 2\Theta\tau A)^{-1}R_m. \end{aligned}$$

Introduce the notations

$$(41) \quad E_m = \begin{bmatrix} e_{m+1} \\ e_m \end{bmatrix} \in R^{2n}, \quad G_m = \begin{bmatrix} (I - 2\Theta\tau A)^{-1}R_m \\ 0 \end{bmatrix} \in R^{2n}$$

and the block Frobenius matrix

$$\begin{aligned} \Phi(\tau A) = \\ = \begin{bmatrix} 2(1 - 2\Theta)(I - 2\Theta\tau A)^{-1}\tau A & (I - 2\Theta\tau A)^{-1}(I + 2\Theta\tau A) \\ I & 0 \end{bmatrix} \end{aligned}$$

The eigenvalues of $\Phi(\tau A)$ coincide with the zeros of the characteristic polynomial of (39). Recursion (40) takes the form

$$(42) \quad E_{m+1} = \Phi(\tau A)E_m + G_m \quad (m = 0, 1, \dots)$$

with the solution

$$(43) \quad E_m = [\Phi(\tau A)]^m E_0 + \sum_{i=0}^{m-1} [\Phi(\tau A)]^{m-i-1} G_i \quad (m \geq 0).$$

The triangle inequality implies

$$(44) \quad \|E_m\| \leq \|\Phi(\tau A)^m\| \|E_0\| + \sum_{i=0}^{m-1} \|\Phi(\tau A)^{m-i-1}\| \|G_i\|.$$

The term G_i may be estimated as follows

$$\|G_i\| = \|(I - 2\Theta\tau A)^{-1}R_i\| \leq \|(I - 2\Theta\tau A)^{-1}\| (2\tau L\|E_i\| + \|T_i\|).$$

Assume that there exists a constant $\gamma > 0$ such that for every $\tau > 0$ the inequality

$$(45) \quad \|(I - 2\Theta\tau A)^{-1}\| \leq \gamma$$

holds. It is also supposed that A is diagonalizable, that is $A = X^{-1}\Lambda X$ with $\Lambda = \text{diag}(\lambda_1, \dots, \lambda_n)$. Then we have

$$(46) \quad \Phi(\tau A) = \begin{bmatrix} X^{-1} & 0 \\ 0 & X^{-1} \end{bmatrix} \Phi(\tau \Lambda) \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix}$$

which implies

$$(47) \quad \Phi(\tau A)^m = \begin{bmatrix} X^{-1} & 0 \\ 0 & X^{-1} \end{bmatrix} \Phi(\tau \Lambda)^m \begin{bmatrix} X & 0 \\ 0 & X \end{bmatrix}$$

Since in the spectral norm $\|\cdot\|_2$ we have

$$(48) \quad \|\Phi(\tau \Lambda)^m\|_2 = \max_{1 \leq \mu \leq n} \|\Phi(\tau \lambda_\mu)^m\|_2$$

and the method (25) is A -stable we can use the uniform boundedness theorem of Gekeler [5]. This result guarantees the existence of constant $K > 0$ such that

$$(49) \quad \sup_{\xi \in \overline{\mathbb{C}^-}} \sup_{m \in \mathbb{N}} \|\Phi(\xi)^m\| \leq K.$$

From (49) the estimations $\|\Phi(\tau \Lambda)^m\|_2 \leq K$ and

$$(50) \quad \|\Phi(\tau A)^m\|_2 \leq K k_2(X) \quad (m \geq 0)$$

follow, where $k_2(X) = \|X^{-1}\|_2 \|X\|_2$ is the condition number of the matrix X . Using (50) and (45) we have

$$(51) \quad \begin{aligned} \|E_m\| &\leq K k_2(X) (\|E_0\| + \gamma \sum_{i=0}^{m-1} \|T_i\|) + \\ &+ \sum_{i=0}^{m-1} 2K k_2(X) \gamma \tau L \|E_i\|. \end{aligned}$$

We need a discrete version of the Gronwall-Bellman lemma : if $z(j) \geq 0$ ($j \geq 0$) and $x(i) \leq y(i) + \sum_{j=0}^{i-1} z(j)x(j)$ $i = 0, 1, \dots, m$; $n \in \mathbb{N}$ then

$$(52) \quad x(m) \leq y(m) + \sum_{i=0}^{m-1} z(i)y(i) \prod_{j=i+1}^{m-1} [1 + z(j)] \quad (m \in \mathbb{N}).$$

In our case we can chose

$$(53) \quad \begin{aligned} x(i) &= \|E_i\| \\ y(i) &= Kk_2(X)(\|E_0\| + \gamma \sum_{j=1}^{i-1} \|T_j\|) \\ z(i) &= 2Kk_2(x)\gamma\tau L = z^*. \end{aligned}$$

Using the monotonicity of $y(i)$ and the lemma we obtain

$$\|E_m\| \leq Kk_2(X)(\|E_0\| + \gamma \sum_{j=0}^{m-1} \|T_j\|)(1 + \sum_{i=0}^{m-1} z^*(1 + z^*)^{m-i-1}).$$

The inequality $1 + z^* \leq e^{z^*}$ implies that

$$1 + \sum_{i=0}^{m-1} z^*(1 + z^*)^{m-i-1} = (1 + z^*)^m \leq e^{mz^*}.$$

Hence we have proven

Theorem 3. *If the matrix A is diagonalizable, $\sigma(A) \in \mathbb{C}^-$ and (45) is satisfied then*

$$(54) \quad \|E_m\| \leq c_1 e^{c_2 m \tau} (\|E_0\| + \gamma \sum_{j=1}^{m-1} \|T_j\|),$$

where

$$(55) \quad c_1 = Kk_2(X), \quad c_2 = 2c_1\gamma L.$$

First we remark that the exponential part of the error constant is due to the nonlinear part of (11)-(12). Consequently for vanishing nonlinear part ($L = 0$) we obtain a sharp estimation (see [5]). In our application $\sigma(A) \in \mathbb{C}^-$ (condition (iv)) and what is more $\sigma(A) \subset \mathbb{R}^-$. If A is diagonalizable, then

$$(I - 2\Theta\tau A)^{-1} = X^{-1}(I - 2\Theta\tau\Lambda)^{-1}X$$

and

$$\|(I - 2\Theta\tau A)^{-1}\|_2 \leq k_2(X)\|(I - 2\Theta\tau\Lambda)^{-1}\|_2.$$

For $\Theta > \frac{1}{4}$ $\|(I - 2\Theta\tau\Lambda)^{-1}\|_2 \leq 1$. Hence one can choose $\gamma = k_2(X)$. If A is Hermitian, then $k_2(X) = 1$.

The convergence of (25) clearly follows from the inequality (54) and the fact that the methods under consideration are of order 2.

References

- [1] DOUGLAS J., DUPONT T., EWING R., Incomplete Iteration for Time-Stepping a Galerkin Method for a Quasi-Linear Parabolic Problem, SIAM J. Num. Anal. 19 (3) (1979).
- [2] DUPONT T., FAIRWEATHER G., JONSON J., Three-Level Galerkin Methods for Parabolic Equations, SIAM J. Num. Anal. 11 (2) (1974).
- [3] FARAGÓ I., Végés elemek módszere lineáris parabolikus típusú feladatok megoldására, Alk. Mat. Lapok, 11 (1985).
- [4] GEKELER E., A priori Error Estimates of Galerkin Backward Differentiation Methods in Time-Inhomogeneous Parabolic Problems, Num. Math 30 (1978).

- [5] GEKELER E., Discretization Methods for Stable Initial Value Problems, Springer, Berlin, 1984.
- [6] KOBZA J., Stability of the Second Derivative Linear Multistep Formulas, Acta Univ. Palackianae Olumucensis, Fac. Rer. Nat., 53 (1977)
- [7] LIONS J.L., Equations deffrentieles operationeles et problems aux limites, Springer, Berlin, 1961.
- [8] STRANG G., FIX G.J., An Analisis of the Finite Element Method, Prientice-Hall, Englewood Cliffs, 1973
- [9] WANNER G., HAIRER E., NORSETT S.P., When I -Stability Imples A -Stability, BIT 18 (1978)
- [10] ZLAMAL M., Unconditionally Stable Finite Element Schemes for Parabolic Equations, Topics in Numerical Analysis II., Proc. of the Royal Irish Academy, 1974
- [11] ZLAMAL M., Finite Element Multistep Discretizations of Parabolic Boundary Value Problems, Math. Comp., 29(130) (1975).
- [12] ZLATEV Z., THOMSEN P., Application of Backward Differentiation Methods to the finite Element Solution of Time-Dependent Problems, Int. J. for Num. Methods, 14 (1979)
- [13] LADISHENSKAYA O. Ya., SOLONNIKOV V. A. and URALTSEV N. N., Linejnije i kvazilejnije uravnenija parabolitseskogo tipa. (In Russian, Moscow 1967. Nauka)

- [14] MOLTSANOV I.I., NIKOLENKO L.D. and
NEZLINA A. Yn., Resenije metodom konetsik elementov
nekotorik klassov nelinejnik zadacs (In Russian) Preprint IK.
ANsssR 35, Kiev 1984

(Received June 28, 1986)

I. FARAGÓ

A. GALÁNTAI

University of Agriculture

Institute of Mathematics

H-2103, Gödöllő

HUNGARY