Internat. J. Math. & Math. Sci.
Vol. 2 No. 4 (1979) 651-667

UNIFORM STABILITY OF LINEAR MULTISTEP METHODS IN GALERKIN PROCEDURES FOR PARABOLIC PROBLEMS

ECKART GEKELER

Der Universitat Stuttgart
7 Stuttgart 80
Pfaffenwaldring 57 West Germany

(Received May 3, 1978 and in revised form, April 3, 1979)

ABSTRACT. Linear multistep methods are considered which have a stability region S and are D-stable on the whole boundary $\partial S \subset S$ of S. Error estimates are derived which hold uniformly for the class of initial value problems $Y' = AY + B(t)$, $t > 0$, $Y(0) = Y_o$, with normal matrix A satisfying the spectral condition $Sp(\Delta tA) \subset S$, Δt time step, Sp(A) spectrum of A. Because of this property, the result can be applied to semidiscrete systems arising in the Galerkin approximation of parabolic problems. Using known results of the Ritz theory in elliptic boundary value problems error bounds for Galerkin multistep procedures are then obtained in this way.

KEY WORDS AND PHRASES. *Numerical solution of ordinary differential equations, A-priori error estimates of linear multistep methods, Linear multistep methods with region of absolute stability, Uniform boundedness of powers of Frobenius matrices.*

1980 MATHEMATICS SUBJECT CLASSIFICATION CODES. *65L05, 65M10, 65M15.*

1. INTRODUCTION.

If a linear parabolic initial boundary value problem is descretized in the space dimensions by the finite element method then the resulting semidiscrete problem is an initial value problem for a system of ordinary differential equations:

$$M(\Delta x)U_t + K(\Delta x)U = P(t), \; t > 0, \; U(0) = U_0, \qquad (1)$$

see e.g. Strang and Fix [19]. In the nomenclature of matrix structural analysis, $P(t)$ is the external load vector, $M(\Delta x)$ the mass matrix, and $K(\Delta x)$ the stiffness matrix; cf. e.g. Bathe and Wilson [2] and Przemieniecki [15]. In the present communication both matrices are supposed to be independent of the time t, real symmetric and positive definite. They depend together with their dimension on the small parameter Δx which is, in general, the maximum diameter of all elements in the finite element subspace; see e.g. [2, 15, 19]. The condition number of $M(\Delta x)$, $\text{cond}(M(\Delta x)) = \|M(\Delta x)\|_2 \|M(\Delta x)^{-1}\|_2$, depends not on Δx but for $L(\Delta x) = M(\Delta x)^{-1/2} \cdot K(\Delta x)M(\Delta x)^{-1/2}$ we find $\text{cond}(L(\Delta x)) \sim 1/\Delta x^j$ with $j = 2$ or $j = 4$ respectively if the elliptic operator $\mathcal{L}$ in the analytic problem is of order two or four, cf.e.g. Strang and Fix [19, ch. 5]. These properties follow in a natural way since the finite-dimensional operator L must be an approximation to the analytical operator $\mathcal{L}$. As a consequence the problem (1) becomes very stiff if a small mesh width Δx is chosen.

Linear multistep methods were frequently proposed for the solution of stiff problems, cf. e.g. Lambert [10]. Their application to systems of the form (1) was studied for instance by Descloux [5], Zlamal [23, 24], and Gekeler [6, 7]. Apart from the drawback to need a special starting procedure, (one-stage) multistep methods have two advantageous properties in comparison with multistage (one-step) methods (i.e. Runge-Kutta methods etc.):

- The order of the discretization error is not negatively affected if the mesh width Δx becomes small.

- The linear system of equations to be solved in every time step of an implicit multistep method has the simple form $(M(\Delta x) + \kappa\Delta t K(\Delta x))Y = C$, Δt time step.

The aim of this paper is to show that the stability of linear multistep approximations to (1), too, remains unaffected by the space discretization if the below defined spectral condition is satisfied. This condition implies no restriction of the relation between Δt and Δx if the multistep method is A_0-stable (definition below).

In the last section we apply our estimates to Galerkin multistep discretizations of parabolic initial boundary problems and show that the order of consistence is the order of convergence in this class of numerical approximations. The results improve some of our error bounds derived in [6] where exponential stability was not yet obtained.

An other goal of the present contribution was to obtain error estimates in a form which is applicable to multistep methods in systems of second order. In a subsequent paper we use our results and consider Galerkin multistep discretizations of hyperbolic initial boundary value problems. The error estimates deduced there correspond to a high degree to those established here for parabolic problems.

2. UNIFORM STABILITY.

To introduce linear multistep methods let

$$\rho(\zeta) = \Sigma_{\kappa=0}^{k}\alpha_\kappa\zeta^{k-\kappa},\ \alpha_0 > 0,\ \sigma(\zeta) = \Sigma_{\kappa=0}^{k}\beta_\kappa\zeta^{k-\kappa},\ \beta_0 \geqslant 0,$$

be two real polynomials without common roots (including zero). Let Δt be a small increment of time t, $Y_n = Y(n\Delta t)$, and let the shift operator T be defined by $(TY)(t) = Y(t+\Delta t)$, $T^\kappa = TT^{\kappa-1}$. Then a linear k-step method $\langle\rho,\sigma\rangle$ for the initial value problem

$$Y' = AY + B(t),\ t > 0,\ Y(0) = Y_0, \tag{2}$$

is formally defined by

$$\rho(T)V_n = \Delta t A\sigma(T)V_n + \Delta t\sigma(T)B_n, \qquad n = 0,1,\dots . \tag{3}$$

Here the starting values $V_0,\dots,V_{k-1}$ are assumed to be known. If we apply the scheme $<\rho,\sigma>$ to (1) then we obtain

$$M(\Delta x)\rho(T)V_n + \Delta t K(\Delta x)\sigma(T)V_n = \Delta t\sigma(T)P_n, \qquad n = 0,1,\dots . \tag{3'}$$

By the linearity of the multistep method this relation is equivalent to (3) if we set $A = -L(\Delta x)$ and replace $M(\Delta x)^{1/2}V_n$ by V_n again.

DEFINITION 1. (Cf. Stetter [17, Def. 4.1.7], Lambert [10, ch. 2].) Let $0 < \delta$ be a fixed constant and $0 < \Delta t < \delta$. A linear multistep method $<\rho,\sigma>$ is consistent if there exists a positive integer q called the order of $<\rho,\sigma>$ such that the truncation error (or defect)

$$d_{<\rho,\sigma>}(\Delta t,w)(t) = \rho(T)w(t) - \Delta t\sigma(T)w'(t)$$

satisfies

$$|d_{<\rho,\sigma>}(\Delta t,w)(t)| \leqslant \kappa\Delta t^{q+1},$$

for all $w \in C^{q+1}_{\mathbb{R}}(\mathbb{R})$, where κ depends on t, δ, and w but not on Δt.

A method $<\rho,\sigma>$ is consistent if and only if the following conditions are fulfilled:

$$\rho(1) = 0,\ \rho'(1) = \sigma(1), \tag{4}$$

see e.g. Lambert [10, p.30]. The following estimation of the truncation error is due to Dahlquist [4, ch.4]; see also Lambert [10, § 3.3].

LEMMA 1. If the linear multistep method $<\rho,\sigma>$ is consistent of order q then

$$\| d_{<\rho,\sigma>}(\Delta t,W)(t)\| \leqslant \kappa_c \max_{t\leq\tau\leq t+k\Delta t} \| W^{(q+1)}(\tau)\| \,\Delta t^{q+1},$$

for all $W \in C^{q+1}_{\mathbb{R}^m}(\mathbb{R})$, where κ_c depends on the data of $<\rho,\sigma>$ but not on t, Δt, W, and the norm.

Let now $\hat{\mathbb{C}}$ be the complex plane extended by the point $\zeta = \infty$ in the usual way, let

$$\pi(\zeta,\eta) \equiv \rho(\zeta) - \eta\sigma(\zeta), \ \pi(\zeta,\infty) = \sigma(\zeta),$$

be the characteristic polynomial of the method $<\rho,\sigma>$ and let Sp(A) be the spectrum of the matrix A. As well-known, the concept of absolute stability plays a fundamental role in the numerical solution of stiff differential equations. However, there exists no method $<\rho,\sigma>$ of order $q > 2$ which is A-stable in the sense of Dahlquist, see e.g. Stetter [17, Def. 2.3.13, Th. 4.6.6]. Therefore Odeh and Liniger [13] weakened the requirement of A-stability in the following way.

DEFINITION 2. A method $<\rho,\sigma>$ is A_G-stable if $G \neq \emptyset$ and

$$\eta \in G \subset \mathbb{C} \text{ implies } |\zeta_\nu| < 1 \text{ for all roots } \zeta_\nu \text{ of } \pi(\zeta,\eta) .$$

A method $<\rho,\sigma>$ is said to be D-stable if no root of the polynomial $\pi(\zeta,0) = \rho(\zeta)$ has modulus greater than one and every root with modulus one is simple. It is strongly D-stable if it is D-stable and $\zeta = 1$ is the only root of $\rho(\zeta)$ with modulus one. In a D-stable method $<\rho,\sigma>$ the roots ζ_ν of $\rho(\zeta)$ with modulus one are called essential roots of $\rho(\zeta)$, the quantity $\chi_\nu = \sigma(\zeta_\nu)/(\zeta_\nu\rho'(\zeta_\nu))$ associated with an essential root ζ_ν of $\rho(\zeta)$ is called the growth parameter of ζ_ν. A D-stable and consistent method $<\rho,\sigma>$ is A_G-stable if all $\mathrm{Re}\chi_\nu > 0$, cf. e.g. [17, Th. 4.6.4]. Especially, every consistent and strongly D-stable method $<\rho,\sigma>$ is A_G-stable with G including the negative real line in a neighborhood of the origin since the essential root $\zeta_1 = 1$ has the growth parameter $\chi_1 = 1$.

If a method $<\rho,\sigma>$ is A_G-stable then the numerical approximation of (2) obtained by (3) is exponentially stable for every fixed matrix A and fixed time step Δt if $\mathrm{Sp}(\Delta tA) \subset G$, cf. e.g. [17, Th. 4.6.3]. But in the present communication we consider an entire class of initial value problems (2) thus $\mathrm{Sp}(\Delta tA)$ can be arbitrary close to the whole boundary ∂G of G where - by a continuity argument - roots of $\pi(\zeta,\eta)$ can have modulus one. (Moreover, if we admit an arbitrary small Δt then G must contain the negative real line in a neighborhood of the origin, i.e., $0 \in \partial G$.) Therefore we must require that $<\rho,\sigma>$ is D-stable on ∂G - and not only in the origin. The next definition was also used by Nevanlinna [12].

DEFINITION 3. The stability region S of a method $\langle\rho,\sigma\rangle$ consists of those $\eta \in \hat{\mathbb{C}}$ for which all roots of $\pi(\zeta,\eta)$ satisfy $|\zeta_\nu| \leq 1$ and those of modulus one are simple.

Obviously, an A_G-stable method $\langle\rho,\sigma\rangle$ has the stability region $S = G$. With this notation the below needed spactral condition can now be written as follows.

DEFINITION 4. The method $\langle\rho,\sigma\rangle$, Δt, and the matrix A fulfil the spectral condition if

(i) $\langle\rho,\sigma\rangle$ has a stability region $S \neq \emptyset$.

(ii) S is closed in $\hat{\mathbb{C}}$.

(iii) $Sp(\Delta t A) \subset S$.

A method $\langle\rho,\sigma\rangle$ is called A_0-stable if it is A_G-stable with G containing the open negative real line $\mathbb{R}_-$. Thus, for instance, $\langle\rho,\sigma\rangle$ fulfils the first two conditions of Def. 4 with $S = \bar{\mathbb{R}}_-$ if it is A_0-stable, D-stable in the origin and in $\zeta = \infty$ the latter meaning that all roots of $\sigma(\zeta)$ have modulus not greater than one and those of modulus one are simple.

We now consider the Frobenius matrix $F_\pi(\eta)$ associated with the characteristic polynomial $\pi(\zeta,\eta) = \rho(\zeta) - \eta\sigma(\zeta) = \Sigma_{\kappa=0}^{k}\gamma_\kappa(\eta)\zeta^{k-\kappa}$, $\gamma_0(\eta) \neq 0$,

$$F_\pi(\eta) = \begin{bmatrix} 0 & 1 & & & 0 \\ & \ddots & \ddots & & \\ & & \ddots & \ddots & \\ 0 & & & 0 & 1 \\ -\gamma_0(\eta)^{-1}\gamma_k(\eta) & \cdots & \cdots & \cdots & -\gamma_0(\eta)^{-1}\gamma_1(\eta) \end{bmatrix}$$

For a matrix A with regular $\gamma_0(A) = \alpha_0 I - \beta_0 A$ the block matrix $\underset{\sim}{F}_\pi(A)$ is obtained if in $F_\pi(\eta)$ the scalar η is replaced by the matrix A. The following lemma represents the basic tool of this paper. It was proved in an entirely different way by Zlamal [23, p. 355] for A_0-stable methods which are D-stable in $\zeta = 0$ and $\zeta = \infty$.

LEMMA 2. Let $\langle\rho,\sigma\rangle$ be a multistep method with stability region S. If S is closed then

$$\Xi_{<\rho,\sigma>} \equiv \sup_{\eta\in S}\sup_{n\in\mathbb{N}}\| F_\pi(\eta)^n \| < \infty$$

$\|\cdot\|$ denoting the Euclidean norm (spectral norm).

PROOF. (i) By assumption there exists for every fixed $\eta \in S$ a lub-norm $\|\cdot\|_\eta$ such that $\|F_\pi(\eta)\|_\eta \leq 1$, see e.g. Stoer and Bulirsch [18, Th. 6.8.2]. Hence $\|F_\pi(\eta)^n\|_\eta \leq 1$, and by the norm equivalence theorem (cf. e.g. Ortega and Rheinboldt [14, Th. 2.2.1]) we obtain

$$\|F_\pi(\eta)^n\| \leq c(\eta)\|F_\pi(\eta)^n\|_\eta \leq c(\eta) < \infty$$

where $c(\eta)$ depends only on η and the dimension k. Therefore the assertion follows for every fixed $\eta \in S$.

(ii) The matrix $F_\pi(\eta)$ has the characteristic polynomial $\pi(\zeta,\eta)$, i.e., the eigenvalues of $F_\pi(\eta)$ are the roots of $\pi(\zeta,\eta)$. Hence the eigenvalues of $F_\pi(\eta)$ are the branches of the algebraic variety ζ defined by $\pi(\zeta,\eta) = 0$. This algebraic variety has the unique finite pole at $\eta = \alpha_0 / \beta_0 > 0$ which cannot lie in S. Moreover, a simple calculation shows that ζ can have only a finite set $\mathfrak{E}$ of exceptional points or branching points η^*, i.e., points in the complex plane where some eigenvalues coalesce or - in other words - where $F_\pi(\eta^*)$ has multiple eigenvalues. Now, we have by (i) that $c(\eta^*) < \infty$ for every $\eta^* \in S \cap \mathfrak{E}$ but by construction of the norm $\|\cdot\|_\eta$ (cf. [18] and the explicit Jordan decomposition of $F_\pi(\eta)$ given below) we find $\lim_{\eta\to\eta^*}c(\eta) = \infty$ for $\eta^* \in S \cap \mathfrak{E}$. Thus we must prove the assertion of the lemma for $\eta \to \eta^*$ in a separated way. For simplicity we show the boundedness of $\|F_\pi(\eta)^n\|$ near η^* here only for a finite η^* in which the algebraic variety ζ has only one confluenting cycle of branches. The assertion is proved analogeously if $\eta^* = \infty \in S$ replacing $1/\eta$ by η since in this case the method $<\rho,\sigma>$ must be implicit, i.e., $\beta_0 \neq 0$.

(iii) Consider a fixed $\eta^* \in S \cap \mathfrak{E}$. By assumption all confluenting branches $\zeta_\nu(\eta)$ - where without loss of generality $\nu = 1,\dots,m \leq k$ - satisfy

$|\zeta_\nu(\eta)| < 1$ in a small disc $\mathfrak{D}$ with center η^*. By Kato [8, 2.1.7] we may write $\zeta_\nu(\eta)$ in $\mathfrak{D}$ as a Puiseux series,

$$\zeta_\nu(\eta) = \Sigma_{\mu=0}^{\infty}\phi_\mu[\omega^\nu(\eta - \eta^*)^{1/m}]^\eta, \qquad \nu = 1,\dots,m, \tag{5}$$

where $\phi_0 = \zeta_1(\eta^*) = \dots = \zeta_m(\eta^*)$ and $\omega = \exp\{2\pi i/m\}$. $\psi(\xi) = \Sigma_{\mu=0}^{\infty}\phi_\mu\xi^\mu$ represents an analytic function in $\mathfrak{D}$ with $|\psi(\xi)| < 1$ in a sufficiently small disc $\tilde{\mathfrak{D}}$ with center 0 and radius θ. Hence we can estimate the coefficients $\phi_\mu^{(n)}$ of $\zeta_\nu(\eta)^n$ as the coefficients of $\psi(\xi)^n$ by Cauchy's estimate (see Ahlfors [1, p. 98]),

$$|\phi_\mu^{(n)}| \leqslant \max_{\xi\in\partial\tilde{\mathfrak{D}}}|\psi(\xi)^n|/\theta^\mu \underset{n\to\infty}{\to} 0, \qquad \mu = 0,\dots . \tag{6}$$

(iv) By the norm equivalence theorem it suffices to show that the elements $f_{ij}^{(n)}(\eta)$ of $F_\pi(\eta)^n$ are uniformly bounded in a neighborhood of η^* with exception of η^*. Let $F_\pi(\eta) = Q(\eta)Z(\eta)Q(\eta)^{-1}$ be the Jordan canonical decomposition. Then, for $\eta \notin \mathfrak{E}$, $Q(\eta)$ is the matrix of the column eigenvectors of $F_\pi(\eta)$, and it has the form

$$Q = \begin{bmatrix} 1 & \text{———} & 1 \\ \zeta_1 & \cdots\cdots & \zeta_k \\ \vdots & \cdots & \vdots \\ \zeta_1^{k-1} & \cdots\cdots & \zeta_k^{k-1} \end{bmatrix}, \quad \det(Q) = \Pi_{\mu>\nu}(\zeta_\mu - \zeta_\nu), \tag{7}$$

where the argument η is omitted. Consequently, using Cramer's rule we obtain after some simple calculations

$$f_{ij}^{(n)}(\eta) = \det(Q_{ij}^{(n)}(\eta))/\det(Q(\eta)) . \tag{8}$$

Here $Q_{ij}^{(n)}(\eta)$ is obtained from $Q(\eta)$ by replacing the jth row by

$$(\zeta_1(\eta)^{n+i-1},\dots,\zeta_k(\eta)^{n+i-1}) . \tag{9}$$

The denominator of (8) satisfies by (7) $\lim_{\eta\to\eta^*}\det(Q(\eta)) = 0$ with exactly the rate of convergence $m(m-1)/2m = (m-1)/2$ since by assumption exactly m roots ζ_ν coalesce in η^*. (We assume that $\phi_1 \neq 0$ in (5) otherwise the proof follows in a slightly modified way.)

(v) By (6) and (9) the assertion of the lemma is now proved for $\eta \to \eta^*$ when we show that the numerator in (8) converges to zero at the same rate as the denominator. To this aim we set briefly $\Omega_\nu = \omega^\nu(\eta - \eta^*)^{1/m}$ then, substituting (5), $\det(Q_{ij}^{(n)}(\eta))$ can be written in the form

$$\det(Q_{ij}^{(n)}(\eta)) = \mathcal{O}((\eta - \eta^*)^{(m-1)^2/m})$$
$$+ \det\Big(\Big(\sum_{\mu=0}^{(m-1)^2-1} \Phi_\mu^{(n)}\Omega_1^\mu, \ldots, \sum_{\mu=0}^{(m-1)^2-1} \Phi_\mu^{(n)}\Omega_m^\mu, \Theta_{m+1}^{(n)}(\eta), \ldots, \Theta_k^{(n)}(\eta)\Big)\Big)$$

where $\|\Theta_\nu^{(n)}(\eta)\|_\infty \leq 1$ by assumption for $\nu = m+1, \ldots, k$, and

$$\Phi_\mu^{(n)} = (\phi_\mu^{(0)}, \ldots, \phi_\mu^{(j-1)}, \phi_\mu^{(n+i-1)}, \phi_\mu^{(j+1)}, \ldots, \phi_\mu^{(k-1)})^T$$

with the notations of (iii). Accordingly, as a determinant is zero if two columns coincide up to a scalar factor,

$$\det(Q_{ij}^{(n)}(\eta)) = \mathcal{O}((\eta - \eta^*)^{(m-1)^2/m})$$
$$+ \sum_{\substack{\mu_1, \ldots, \mu_m \\ 0 \leq \mu_1 \leq (m-1)^2-1 \\ \mu_r \neq \mu_s,\, r \neq s}} \Omega_1^{\mu_1} \cdots \Omega_m^{\mu_m} \det((\Phi_{\mu_1}^{(n)}, \ldots, \Phi_{\mu_m}^{(n)}, \Theta_{m+1}^{(n)}(\eta), \ldots, \Theta_k^{(n)}(\eta))) .$$

This proves the desired result since

$$\Omega_1^{\mu_1} \cdots \Omega_m^{\mu_m} = \mathcal{O}((\eta - \eta^*)^{\Sigma_{1=1}^m \mu_1/m}) = \mathcal{O}((\eta - \eta^*)^{m(m-1)/2m})$$

if $0 \leq \mu_1 \leq (m-1)^2-1$ and $\mu_r \neq \mu_s$ for $r \neq s$.

By means of Lemma 2 an error bound is now easily derived. The error $E_n = Y_n - V_n$ of the method (3) in the problem (2) fulfils the relation

$$\rho(T)E_n - \Delta t A \sigma(T) E_n = d_{<\rho,\sigma>}(\Delta t, A)_n , \qquad n = 0, 1, \ldots, \tag{10}$$

by definition of the truncation error $d_{<\rho,\sigma>}(\Delta t, Y)$. Introducing the block vectors

$$\underset{\sim}{E}_n = (E_{n-k+1}, \ldots, E_n)^T ,$$
$$\underset{\sim}{d}_{<\rho,\sigma>}(\Delta t, Y)_n = (0, \ldots, 0, (\alpha_0 I - \beta_0 \Delta t A)^{-1} d_{<\rho,\sigma>}(\Delta t, Y)_{n-k})^T ,$$

the Jordan canonical decomposition of A, $A = X\Lambda X^{-1}$, and the block diagonal matrix $\underset{\sim}{X} = (X, \ldots, X)$ we find that (10) is equivalent to

$$\underset{\sim}{E}_n = \underset{\sim}{F}_\pi(\Delta t A)\underset{\sim}{E}_{n-1} + \underset{\sim}{d}_{<\rho,\sigma>}(\Delta t, Y)_n, \qquad n = k,\ldots,$$

or

$$\underset{\sim}{E}_n = \underset{\sim}{X}\underset{\sim}{F}_\pi(\Delta t\Lambda)\underset{\sim}{X}^{-1}\underset{\sim}{E}_{n-1} + \underset{\sim}{d}_{<\rho,\sigma>}(\Delta t, Y)_n, \qquad n = k,\ldots \quad . \tag{11}$$

Therefore Lemma 2 immediately yields

THEOREM 1. If the matrix A is diagonable and the spectral condition is satisfied then

$$\|E_n\| \leq \Xi_{<\rho,\sigma>}\|X\|\,\|X^{-1}\|\,(\Sigma_{\kappa=0}^{k-1}\|E_\kappa\| + \Sigma_{\nu=k}^{n}\|\underset{\sim}{d}_{<\rho,\sigma>}(\Delta t, Y)_\nu\|)\,.$$

It should be emphasized that this error bound holds uniformly for the class of normal matrices A satisfying the spectral condition.

3. UNIFORM EXPONENTIAL STABILITY.

The regions $S_\mu \subset S$ of μ-exponential stability considered in this section are here introduced in a slight modification of Stetter [17, Def. 2.3.15 and § 4.6.2].

DEFINITION 4. A method $<\rho,\sigma>$ has the region of μ-exponential stability S_μ, $\mu \geq 0$, if the associated Frobenius matrix $F_\pi(\eta)$ satisfies $\mathrm{spr}(F_\pi(\eta)) \leq 1 - 4\mu$ for all $\eta \in S_\mu$, spr(F) denoting the spectral radius of F.

For two (m,m)-matrices P, Q we write $P \leq Q$ if $W^H PW \leq W^H QW$ holds for all $W \in \mathbb{C}^m$, $W^H = \bar{W}^T$. The following matrix theorem of Kreiss [9] is quoted here in a somewhat shortened form using a modification of Widlund [20].

KREISS' MATRIX THEOREM. Let $\mathfrak{F}$ denote a family of real or complex (k,k)-matrices. Then the following statements are equivalent:

(i) $$\Xi_1 \equiv \sup_{F\in\mathfrak{F}}\sup_{n\in\mathbb{N}}\|F^n\| < \infty$$

(ii) There is a constant $\Xi_2 > 0$ depending only on Ξ_1 and the dimension k, and for every matrix $F \in \mathfrak{F}$ a positive definite hermitean matrix H with

$$\Xi_2^{-1} I \leq H \leq \Xi_2 I \qquad \text{(I identity)}$$

such that

$$F^H HF \leq (1 + \mathrm{spr}(F))H/2\ .$$

As already used, we have $\underset{\sim}{F}_\pi(\Delta tA) = \underset{\sim}{X}\underset{\sim}{F}_\pi(\Delta t\Lambda)\underset{\sim}{X}^{-1}$ for every diagonable matrix $A = X\Lambda X^{-1}$ with regular $\alpha_0 I - \beta_0 \Delta tA$ and, moreover, $\|F_\pi(\Delta t\Lambda)^n\| = \max_{1\leq\kappa\leq k}\|F_\pi(\Delta t\lambda_\kappa)^n\|$ is true. Hence the following corollary is an immediate consequence of Lemma 2 and the Kreiss' Matrix Theorem.

COROLLARY. Let the matrix A be diagonable and let the spectral condition be satisfied. If $Sp(\Delta tA) \subset S_\mu$ then there is a constant Ξ_2 depending only on $\Xi_{<\rho,\sigma>}$ and the dimension k, and for every $\underset{\sim}{F}_\pi(\Delta tA)$ a positive definite hermitean matrix $\underset{\sim}{H}$ such that

$$\Xi_2^{-1}\underset{\sim}{I} \leq \underset{\sim}{H} \leq \Xi_2\underset{\sim}{I}$$

and

$$\underset{\sim}{F}_\pi(\Delta tA)^H(\underset{\sim}{X}^{-1})^H\underset{\sim}{H}\underset{\sim}{X}^{-1}\underset{\sim}{F}_\pi(\Delta tA) \leq (1 -2\mu)(\underset{\sim}{X}^{-1})^H\underset{\sim}{H}\underset{\sim}{X}^{-1}.$$

Let now $\|\underset{\sim}{E}_n\|_H^2 = \underset{\sim}{E}_n^H(\underset{\sim}{X}^{-1})^H\underset{\sim}{H}\underset{\sim}{X}^{-1}\underset{\sim}{E}_n$ then, multiplying the error equations (11) from left by $\underset{\sim}{H}^{1/2}\underset{\sim}{X}^{-1}$ we obtain

$$\|\underset{\sim}{E}_n\|_H \leq \|\underset{\sim}{F}_\pi(\Delta tA)\underset{\sim}{E}_{n-1}\|_H + \|\underset{\sim}{d}_{<\rho,\sigma>}(\Delta t,Y)_n\|_H, \quad n = k,\ldots,$$

and

$$\Xi_2^{-1/2}\|X\|^{-1}\|\underset{\sim}{E}\| \leq \|\underset{\sim}{E}\|_H \leq \Xi_2^{1/2}\|X^{-1}\|\,\|\underset{\sim}{E}\| \quad (\|.\| \text{ Euclidean norm}).$$

Accordingly, a recursive estimation by means of the Corollary yields

THEOREM 2. Let the matrix A be diagonable and the spectral condition be satisfied. If $Sp(\Delta tA) \subset S_\mu$, $\mu \geq 0$, then

$$\|Y_n - V_n\| \leq \Xi_2\|X\|\,\|X^{-1}\|\Big[\exp\{-(\mu/\Delta t)n\Delta t\}\Sigma_{\kappa=0}^{k-1}\|Y_\kappa - V_\kappa\| + \Sigma_{\nu=k}^{n}\exp\{-(\mu/\Delta t)(n-\nu)\Delta t\}\|\underset{\sim}{d}_{<\rho,\sigma>}(\Delta t,Y)_\nu\|\Big],$$

and

$$\|\underset{\sim}{d}_{<\rho,\sigma>}(\Delta t,Y)_\nu\| \leq \|(\alpha_0 I-\beta_0\Delta tA)^{-1}\|\kappa_c\,|||Y^{(q+1)}|||_\nu\Delta t^{q+1},$$

$|||Y|||_\nu = \max_{0\leq t\leq\nu\Delta t}\|Y(t)\|$, if $<\rho,\sigma>$ is of order q.

The next lemma shows that for a fixed matrix A the exponential growing factor $\mu/\Delta t$ can be estimated more exactly.

LEMMA 3. If the growth parameters χ_κ of a D-stable method $<\rho,\sigma>$ satisfy

Re $\chi_\kappa > 0$ and if all eigenvalues λ_μ of the matrix A satisfy $\lambda_\mu \leqslant -\alpha < 0$ then, for Δt sufficiently small,

$$\mu/\Delta t \geqslant \alpha \min_\kappa \mathrm{Re}\chi_\kappa/8 > 0 \ .$$

PROOF. For simplicity let ζ_κ, $\kappa = 1,\ldots,i$, $i \leqslant k$, be the essential roots of the polynomial $\rho(\zeta)$ and let $\zeta_\kappa(\eta)$ be the corresponding roots of $\pi(\zeta,\eta)$, $\zeta_\kappa(\eta) \to \zeta_\kappa$, $\eta \to 0$. Then there exists a non-empty interval such that

$$\mathrm{spr}(F_\pi(\eta)) = \max_{1\leq\kappa\leq i} |\zeta_\kappa(\eta)|, \ \delta_1 \leqslant \eta \leqslant 0, \ \delta_1 < 0 \ . \tag{13}$$

But $\zeta_\kappa(\eta) = \zeta_\kappa[1 + \chi_\kappa\eta + \mathcal{O}(\eta^2)]$ as follows by substituting this expansion in $\pi(\zeta,\eta) = 0$. Therefore, $|\zeta_\kappa(\eta)|^2 = 1 + 2\mathrm{Re}\chi_\kappa\eta + \mathcal{O}(|\eta|^2)$ and there is a non-empty interval such that

$$|\zeta_\kappa(\eta)| \leqslant 1 + (\mathrm{Re}\chi_\kappa\eta/2), \ \kappa = 1,\ldots,i, \ \delta_2 \leqslant \eta \leqslant 0, \ \delta_2 < 0 \ , \tag{14}$$

since $\mathrm{Re}\chi_\kappa > 0$ by assumption. By (13) and (14) we obtain

$$\mathrm{spr}(F_\pi(\eta)) \leqslant 1 + \eta\min_{1\leq\kappa\leq i}(\mathrm{Re}\chi_\kappa/2), \ \max\{\delta_1,\delta_2\} \leqslant \eta \leqslant 0 \ ,$$

and thus

$$\mathrm{spr}(F_\pi(\Delta t\lambda_\mu)) \leqslant 1 - \Delta t\alpha\min_{1\leq\kappa\leq i}(\mathrm{Re}\chi_\kappa/2) \ ,$$

for $\Delta t\alpha \leqslant \max\{|\delta_1|, |\delta_2|\}$. Now the assertion follows because we can set here

$$\mu = [1 - \max_{1\leq\mu\leq m}\mathrm{spr}(F_\pi(\Delta t\lambda_\mu))]/4 \ .$$

4. GALERKIN MULTISTEP PROCEDURES IN PARABOLIC PROBLEMS.

To be brief we first list up some necessary notations: $\Omega \subset \mathbb{R}^p$ bounded domain;

$\|f\|_s = (\int_\Omega \Sigma_{0\leq|\sigma|\leq s}|D^\sigma f(x)|^2dx)^{1/2}$, $s \in \mathbb{N}$, Sobolev norm with the standard multi-index notation;

$C_0^\infty(\Omega)$ set of real-valued functions $f \in C^\infty(\Omega)$ with compact support in Ω;

$W_0^s(\Omega)$ closure of $C_0^\infty(\Omega)$ with respect to $\|\cdot\|_s$;

$W^s(\Omega)$ closure of $C^\infty(\Omega)$ with respect to $\|\cdot\|_s$;

$\mathcal{H} \subset W^s(\Omega)$ closed subspace with $S_0^s(\Omega) \subset \mathcal{H}$;

$a: \mathcal{H} \times \mathcal{H} \ni (u,v) \to a(u,v) \in \mathbb{R}$ symmetric bilinear form over $\mathcal{H}$ satisfying

$$0 < \alpha \|v\|_s^2 \leqslant a(v,v) \qquad \text{for all } 0 \neq v \in \mathcal{H} ;$$

$(f,g) = \int_\Omega f(x)g(x)dx$.

Then, following Schultz [16], see also Strang and Fix [19, ch. 7], Mitchell and Wait [11, § 6.3], a large class of parabolic initial boundary value problems with homogeneous boundary conditions can be written in the weak form

$$\begin{aligned} &(\frac{\partial u}{\partial t}(\cdot,t),v) + a(u(\cdot,t),v) = (b(\cdot,t),v) && \text{for all } v \in \mathcal{H},\ t > 0 , \\ &(u(\cdot,0) - b_0,v) = 0 && \text{for all } v \in \mathcal{H}. \end{aligned} \tag{15}$$

Let $\mathcal{S} \subset \mathcal{H}$ be a finite-dimensional subspace which is defined by an explicitely known set of basis functions $s_1,\ldots,s_m$. The Galerkin approximation to the solution $u(\cdot,t)$ of (15) is a function $u_G(\cdot,t) \in \mathcal{S}$ which satisfies (15) for all $v \in \mathcal{S}$. Accordingly, substituting in (15)

$$u_G(x,t) = U(t)^T \underset{\sim}{s}(x),\ \underset{\sim}{s}(x) = (s_1(x),\ldots,s_m(x)),\ U(t) \in \mathbb{R}^m ,$$

we obtain a system of the form (1) where, as well-known,

$$M = ((s_\mu,s_\nu))_{\mu,\nu=1}^m,\ K = (a(s_\mu,s_\nu))_{\mu,\nu=1}^m . \tag{16}$$

The Galerkin multistep approximation $u_\Delta(\cdot,n\Delta t) = u_\Delta(\cdot)_n \in \mathcal{S}$, $n = k,\ldots,$ to the problem (15) is given by

$$u_\Delta(x,t) = V(t)^T \underset{\sim}{s}(x) , \qquad t = n\Delta t , \tag{17}$$

and

$$\begin{aligned} \rho(T)(u_\Delta(\cdot)_n,s_\mu) + \Delta t\sigma(T)a(u_\Delta(\cdot)_n,s_\mu) &= \Delta t\sigma(T)(b(\cdot)_n,s_\mu), \\ \mu = 1,\ldots,m,\ n &= k,\ldots, \end{aligned} \tag{18}$$

where $u_\Delta(\cdot)_0,\ldots,u_\Delta(\cdot)_{k-1}$ are assumed to be known. It is easily shown that (17) and (18) are equivalent to (3') and (16).

An estimation of $u - u_G$ usually needs the Ritz projection $u_R \in \mathcal{S}$ of u defined by

$$a(u(\cdot,t) - u_R(\cdot,t),v) = 0 \qquad \text{for all } v \in \mathcal{S},\ t > 0 ,$$

see e.g. Strang and Fix [19, § 7.2]. Hence we estimate the error $u - u_\Delta$ via the decomposition

$$u - u_\Delta = (u - u_R) + (u_R - u_\Delta) \tag{19}$$

directly. However, it is not the aim of the present contribution to discuss the vast field of error bounds in the Ritz theory of elliptic boundary value problems. Therefore we make the following assumption on the bilinear form a, the boundary of Ω, and the subspace $\mathcal{S}$, cf. Ciarlet [3, Th. 3.2.5], Strang and Fix [19, Th. 3.7], Schultz [16], and Zlamal [21, 22].

ASSUMPTION A. The Ritz projection $\tilde{u}_R \in \mathcal{S}$ of the solution $\tilde{u}$ of the elliptic boundary value problem

$$a(\tilde{u},v) = (f,v) \qquad \text{for all } v \in \mathcal{H}$$

satisfies for a fixed $l \in \mathbb{N}$

$$\|\tilde{u} - \tilde{u}_R\|_0 \leqslant \kappa_R \Delta x^l \|\tilde{u}\|_l$$

where κ_R depends not on the sufficiently smooth right side f and the small parameter Δx.

Under Assumption A we obtain in case the solution u of the parabolic problem (15) is sufficiently smooth

$$\|u(\cdot,t) - u_R(\cdot,t)\|_0 \leqslant \kappa_R \Delta x^l \|u(\cdot,t)\|_l , \qquad t > 0 ,$$

since $u(\cdot,t)$ can be viewed as the solution of the elliptic problem

$$a(u(\cdot,t),v) = (b(\cdot,t) - u_t(\cdot,t),v) \quad \text{for all } v \in \mathcal{H} ,$$

$u_t = \partial u/\partial t$. Further, let

$$(u_R - u_\Delta)(x,t) = \Sigma_{\mu=1}^m \varepsilon_\mu(t) s_\mu(x) , \qquad t = n\Delta t ,$$

then, writing $E(t) = M^{1/2}(\varepsilon_1(t),\ldots,\varepsilon_m(t))^T$ we obtain

$$\|(u_R - u_\Delta)(\cdot,t)\|_0 = \|E(t)\| , \ t = n\Delta t , \ \|\cdot\| \text{ Euclid norm} . \tag{20}$$

In order to deduce an estimation of $u_R - u_\Delta$ we first observe that by (18) and Def. 1

$$\rho(T)((u - u_\Delta)(\cdot)_n, s_\mu) + \Delta t\sigma(T)a((u - u_\Delta)(\cdot)_n, s_\mu)$$
$$= (d_{<\rho,\sigma>}(\Delta t, u(\cdot))_n, s_\mu)$$

or, by definition of the Ritz projection,

$$\rho(T)((u_R - u_\Delta)(\cdot)_n, s_\mu) + \Delta t\sigma(T)a((u_R - u_\Delta)(\cdot)_n, s_\mu)$$
$$= d_{<\rho,\sigma>}(\Delta t, u(\cdot))_n, s_\mu) + \rho(T)((u_R - u)(\cdot)_n, s_\mu)$$
$$\mu = 1,\dots,m, \quad n = 0,1,\dots \quad .$$

This relation is equivalent to

$$\rho(T)E_n + \Delta t L\sigma(T)E_n = D_n , \qquad n = 0,1,\dots, \tag{21}$$

where $L = M^{-1/2}KM^{-1/2}$ with the notations of (16), and the vector $D_n \in \mathbb{R}^m$ has the form

$$D_n = M^{-1/2}((\Psi_n, s_1),\dots,(\Psi_n, s_m))^T$$

with

$$\Psi_n = d_{<\rho,\sigma>}(\Delta t, u(\cdot))_n + \rho(T)(u_R - u)(\cdot)_n \in \mathcal{H}.$$

Hence, using the W^0-projection of Ψ_n onto $\mathcal{S}$ it is easily shown that

$$\| D_n \| \leqslant \| d_{<\rho,\sigma>}(\Delta t, u(\cdot))_n + \rho(T)(u_R - u)(\cdot)_n \|_0 .$$

If the method $<\rho,\sigma>$ is consistent then $\rho(\zeta)$ has the root $\zeta = 1$, i.e., $\rho(\zeta) = \tilde\rho(\zeta)(\zeta - 1)$. Accordingly,

$$\| \rho(T)(u_R - u)(\cdot)_n \|_0 = \| \tilde\rho(T)(T - 1)(u_R - u)(\cdot)_n \|_0$$
$$\leqslant \kappa_\rho \Delta t \max_{n\Delta t \leq t \leq (n+k)\Delta t} \| (u_R - u)_t(\cdot, t) \|_0$$

and

$$\| D_n \| \leqslant \| d_{<\rho,\sigma>}(\Delta t, u(\cdot))_n \|_0 + \kappa_\rho \Delta t \max_{0 \leq t \leq (n+k)\Delta t} \| (u - u_R)_t(\cdot, t) \|_0 .$$

But $(u - u_R)_t = u_t - (u_t)_R$ as $(u_R)_t = (u_t)_R$ in the present case of a time-homogeneous bilinear form a. We therefore obtain under Assumption A

$$\| D_n \| \leqslant \| d_{<\rho,\sigma>}(\Delta t, u(\cdot))_n \|_0 + \kappa_\rho \kappa_R \Delta t \Delta x^1 |\!|\!| u_t |\!|\!|_{1,n+k} , \tag{22}$$

$|\!|\!| u |\!|\!|_{1,n} = \max_{0 \leq t \leq n\Delta t} \| u(\cdot, t) \|_1$, since u_t can be viewed as solution of an elliptic problem, too.

Finally we can apply Theorem 2 to the error equation (21) and estimate the defect D_n by (22). Then, by means of (20) we obtain an error bound for the second term on the right side of (19). We summarize our result in the following theorem.

THEOREM 3. If the spectral condition is satisfied and $Sp(\Delta t M^{-1/2} K M^{-1/2}) \subset S_\mu$, $\mu \geqslant 0$, then

$$\| u_R(\cdot)_n - u_\Delta(\cdot)_n \|_0 \leqslant \Xi_2 \big(\exp\{-(\mu/\Delta t) n\Delta t\} \Sigma_{\kappa=0}^{k-1} \| u_R(\cdot)_\kappa - u_\Delta(\cdot)_\kappa \|_0 + \Sigma_{\nu=k}^{n} \exp\{-(\mu/\Delta t)(n-\nu)\Delta t\} \| \underset{\sim}{D}_\nu \| \big) .$$

If the solution u of (15) is sufficiently smooth, Assumption A is fulfilled, and the method $<\rho,\sigma>$ is consistent of order q then

$$\| \underset{\sim}{D}_\nu \| \leqslant \alpha_0^{-1} (\kappa_c \Delta t^{q+1} ||| \partial^{q+1} u / \partial t^{q+1} |||_{0,\nu} + \kappa_\rho \kappa_R \Delta t \Delta x^1 ||| u_t |||_{1,\nu}) .$$

REFERENCES

1. Ahlfors, L.V. Complex Analysis, McGraw-Hill, New York (1953).

2. Bathe, K.J. and Wilson, E.L. Numerical Methods in Finite Element Analysis, Prentice-Hall, Englewood Cliffs, N.J. (1976).

3. Ciarlet, Ph.G. The Finite Element Method for Elliptic Problems, North-Holland, Amsterdam (1978).

4. Dahlquist, G. Stability and Error Bounds in the Numerical Integration of Ordinary Differential Equations, Transactions Royal Inst. Techn., Stockholm, Nr. 130 (1959).

5. Descloux, J. On the Numerical Integration of the Heat Equation, Numer. Math. 15 (1970) 371-381.

6. Gekeler, E. Linear Multistep Methods and Galerkin Procedures for Initial Boundary Value Problems, SIAM J. Numer. Anal. 13 (1976) 536-548.

7. Gekeler, E. A-priori Error Estimates of Galerkin Backward Differentiation Methods in Time-inhomogeneous Parabolic Problems, Numer. Math. 30 (1978) 369-383.

8. Kato, T. Perturbation Theory for Linear Operators, Springer-Verlag, Berlin-Heidelberg-New York (1966).

9. Kreiss, H.O. Über die Stabilitätsdefinition für Differenzengleichungen, die partielle Differentialgleichungen approximieren, BIT 2 (1962) 153-181.

10. Lambert, J.D. Computational Methods in Ordinary Differential Equations, John Wiley, New York (1973).

11. Mitchel, A.R. and Wait, R. The Finite Element Method in Partial Differential Equations, John Wiley, New York (1977).

12. Nevanlinna, O. On the Numerical Integration of Nonlinear Initial Value Problems by Linear Multistep Methods, BIT 17 (1977) 58-71.

13. Odeh, F. and Liniger, W. A Note on Unconditional Fixed-h Stability of Linear Multistep Formulae, Computing 7 (1971) 240-253.

14. Ortega, J.M. and Rheinboldt, W.C. Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, New York-London (1970).

15. Przemieniecki, J.S. Theory of Matrix Structural Analysis, McGraw-Hill, New York (1968).

16. Schultz, M.H. L^2 Error Bounds for the Rayleigh-Ritz-Galerkin method, SIAM J. Numer. Anal. 8 (1971) 737-748.

17. Stetter, H.J. Analysis of Discretization Methods for Ordinary Differential Equations, Springer-Verlag, Berlin-Heidelberg-New York (1973).

18. Stoer, J. and Bulirsch, R. Einführung in die Numerische Mathematik II, Springer-Verlag, Berlin-Heidelberg-New York (1973).

19. Strang, G. and Fix, G.J. An Analysis of the Finite Element Method, Prentice-Hall, Englewood Cliffs, N.J. (1973).

20. Widlund, O.B. Stability of Parabolic Difference Schemes in the Maximum Norm, Numer. Math. 8 (1966) 186-202.

21. Zlamal, M. Curved Elements in the Finite Element Method I, SIAM J. Numer. Anal. 10 (1973) 229-240.

22. Zlamal, M. Curved Elements in the Finite Element Method II, SIAM J. Numer. Anal. 11 (1974) 347-362.

23. Zlamal, M. Finite Element Multistep Discretizations of Parabolic Boundary Value Problems, Mathem. Comput. 29 (1975) 350-359.

24. Zlamal, M. Finite Element Multistep Methods for Parabolic Equations, in: Finite Elemente und Differenzenverfahren (J.Albrecht, L.Collatz eds.) ISNM Bd. 28, Birkhäuser-Verlag, Basel (1975).