MATHEMATICS OF COMPUTATION, VOLUME 29, NUMBER 130
APRIL 1975, PAGES 350–359

Finite Element Multistep Discretizations of Parabolic Boundary Value Problems

By Miloš Zlámal

Abstract. The initial-boundary value problem for a linear parabolic equation in an infinite cylinder under the Dirichlet boundary condition is solved by applying the finite element discretization in the space dimension and A_0-stable multistep discretizations in time. No restriction on the ratio of the time and space increments is imposed. The methods are analyzed and bounds for the discretization error in the L_2-norm are given.

1. Introduction. The problem we are considering is the initial-boundary value problem

$$\begin{aligned} \partial u/\partial t &= Lu \quad && \text{for } (x, t) \in \Omega \times (0, \infty), \\ u &= 0 && \text{on } \Gamma \times (0, \infty), \\ u(x, 0) &= g(x) && \text{in } \Omega. \end{aligned} \tag{1.1}$$

Here,

$$\begin{aligned} Lu &= \sum_{i,j=1}^{N} \frac{\partial}{\partial x_i}\left(a_{ij}(x)\frac{\partial u}{\partial x_j}\right) - a(x)u, \\ a_{ij}(x) &= a_{ji}(x), \qquad \sum_{i,j=1}^{N} a_{ij}(x)\xi_i\xi_j \geqslant \alpha \sum_{i=1}^{N} \xi_i^2, \qquad \alpha = \text{const} > 0, \ a(x) \geqslant 0, \end{aligned} \tag{1.2}$$

and $x = (x_1, \cdots, x_N)$ is a point of a bounded domain Ω in Euclidean N-space R^N with a smooth boundary Γ.

Before formulating (1.1) in the weak variational form, let us introduce some notations. The norm $\|\cdot\|_{L_2(\Omega)}$ of the space $L_2(\Omega)$ and the scalar product are denoted by $\|\cdot\|_0$ and $(\cdot, \cdot)_0$, respectively. $H^m \equiv W_2^{(m)}(\Omega)$, $m = 0, 1, \cdots$, denotes the Sobolev space defined by

$$\|v\|_{H^m} = \left(\sum_{|j| \leqslant m} \|D^j v\|_0^2\right)^{1/2}.$$

Instead of $\|v\|_{H^m}$, we write $\|v\|_m$. H_0^1 is the closure of $\mathcal{D}(\Omega)$, the set of infinitely differentiable functions with compact support in Ω, in the norm $\|\cdot\|_1$. The energy norm $|v|_1$ is defined by $|v|_1^2 = a(v, v)$, where $a(v, w)$ is the energy bilinear functional

Received December 27, 1973.
AMS (MOS) subject classifications (1970). Primary 65N30.

$$a(v,w) = \int_\Omega \left[\sum_{i,j=1}^{N} a_{ij}(x) \frac{\partial v}{\partial x_i} \frac{\partial w}{\partial x_j} + a(x)vw \right] dx.$$

The weak form of (1.1) is to find for $t > 0$ the function $u \in H_0^1$ such that, besides the initial condition, it satisfies

$$(\dot{u}, \varphi)_0 + a(u, \varphi) = 0 \qquad \forall \varphi \in H_0^1. \tag{1.3}$$

A well-known approach for getting an approximate solution of the problem (1.1) consists in first applying the Galerkin principle to (1.3). Let S be a finite-dimensional subspace of H_0^1. The Galerkin solution is the function $U \in S$ which satisfies

$$(\dot{U}, \varphi)_0 + a(U, \varphi) = 0 \qquad \forall \varphi \in S. \tag{1.4}$$

The Galerkin formulation yields a system of ordinary differential equations in time. A suitable discretization in time will give a computable approximate solution of the problem (1.1). The choice of finite element subspaces for S and of Crank-Nicolson and other one step discretizations in time was considered in several papers published in recent years (see references in [9]). In [9], we chose for S finite-dimensional subspaces V_h^p of H_0^1 which have the following approximation property: for any $v \in H^{p+1} \cap H_0^1$, there exists a function $\hat{v} \in V_h^p$ such that

$$\|v - \hat{v}\|_j \leqslant Ch^{p+1-j} \|v\|_{p+1}, \qquad j = 0, 1, \tag{1.5}$$

C being a constant independent of the small positive parameter h and of the function v. Finite element subspaces constructed first for special domains, later for arbitrary curved domains (see [8], [2], [10], [11]), possess this property. The parameter h is, in general, the maximum diameter of all elements.

In this paper, we again choose the subspaces V_h^p for S, and we discretize (1.4) by a A_0-stable linear multistep method. A_0-stable linear multistep methods were introduced for ordinary differential equations by Cryer [3]. When we apply the multistep method (ρ, σ), where

$$\rho(\zeta) = \sum_{j=0}^{\nu} \alpha_j \zeta^j, \quad \alpha_\nu > 0, \qquad \sigma(\zeta) = \sum_{j=0}^{\nu} \beta_j \zeta^j,$$

to the scalar equation $\dot{x}(t) = -\lambda x(t)$, $x(0) = 1$, the approximate values x^n of $x(nk)$ (k is the time increment) are determined by $\Sigma_{j=0}^{\nu} \alpha_j x^{n+j} = -k\lambda \Sigma_{j=0}^{\nu} \beta_j x^{n+j}$. A_0-stability requires that $x^n \longrightarrow 0$ as $n \longrightarrow \infty$ for all positive λ. This is fulfilled iff all roots $\zeta_j(\tau)$, $j = 1, \cdots, \nu$, of the polynomial

$$p(\zeta) = \rho(\zeta) + \tau\sigma(\zeta) \tag{1.6}$$

satisfy $|\zeta_j(\tau)| < 1$ for every $\tau > 0$.

Denote by U^n the approximate values of U at the time level $t = nk$, $n = 0, 1, \cdots$, and assume that $U^0, U^1, \cdots, U^{\nu-1}$ are given. If we apply the scheme (ρ, σ) to (1.4), we get the recurrence relationship for U^n:

$$(1.7)\qquad \left(\sum_{j=0}^{\nu}\alpha_j U^{n+j},\varphi\right)_0 + ka\left(\sum_{j=0}^{\nu}\beta_j U^{n+j},\varphi\right)=0 \qquad \forall\varphi\in V_h^p,\ n=0,1,\cdots.$$

Besides A_0-stability, we require that the method (ρ, σ) be stable in the sense of Dahlquist and of the order $q \geqslant 1$ and that the roots of the polynomial $\sigma(\zeta)$ with modulus equal to one be simple. Under the assumption that the solution of (1.1) is smooth enough, we prove the following bound which is uniform for $\nu \leqslant n < \infty$ and which holds without any restriction on the ratio kh^{-1}:

$$\sup_{\nu\leqslant n<\infty}\|u^n - U^n\|_0 \leqslant C\left[\sum_{j=0}^{\nu-1}\|u^j - U^j\|_0 + (h^{p+1}+k^q)\log\frac{1}{k}\|g\|_m\right];$$

here u^n are the exact values $u(x, nk)$.

To see what computations the relationship (1.7) represents, let us choose a basis $v_1, \cdots, v_l$ of V_h^p (of course, in finite element subspaces, we do not choose an arbitrary basis). Let M be the so-called mass matrix, $M = \{(v_i, v_j)_0\}_{i,j=1}^l$, and K the stiffness matrix, $K = \{a(v_i, v_j)\}_{i,j=1}^l$. Both these matrices are positive definite. If $\mathbf{v} = (v_1, \cdots, v_l)^T$ (the superscript T denotes transposition) and $U^n = (\boldsymbol{\alpha}^n)^T\mathbf{v}$, where $\boldsymbol{\alpha}^n = (\alpha_1^n, \cdots, \alpha_l^n)^T$, then setting in (1.7) the basis functions $v_1, \cdots, v_l$ for φ, we get

$$(1.8)\qquad \sum_{j=0}^{\nu}(\alpha_j M + \beta_j kK)\boldsymbol{\alpha}^{n+j} = 0, \qquad n = 0, 1, \cdots.$$

Evidently, at every time step we have to solve a linear system with the same matrix $B = \alpha_\nu M + \beta_\nu kK$. This matrix is positive definite (from A_0-stability it follows $\beta_\nu > 0$; see [3, Theorem 3.1]) sparse, banded, and its condition number does not grow too fast. In the case of finite element subspaces, it follows that

$$(1.9)\qquad \operatorname{cond}(B) = O(kh^{-2}).$$

Multistep methods require the determination of starting values $U^0, \cdots, U^{\nu-1}$, and it is desirable that these values be calculated to an accuracy as high as the local accuracy of the method. This disadvantage of multistep methods can be overcome, at least for ν small ($\nu \leqslant 4$), by computing $U^0, \cdots, U^\nu$ by the Crank-Nicolson (i.e., trapezoidal) method or by the Calahan method (a third order one-step method, see [9]) and using a step sufficiently small with respect to the step k of the main method.

For simplicity, we restricted ourselves to the homogeneous problem (1.1). The generalization of (1.7) for the nonhomogeneous equation $\partial u/\partial t = Lu + F(x, t)$ is immediate:

$$\left(\sum_{j=0}^{\nu}\alpha_j U^{n+j},\varphi\right)_0 + ka\left(\sum_{j=0}^{\nu}\beta_j U^{n+j},\varphi\right) = k\left(\sum_{j=0}^{\nu}\beta_j F^{n+j}(x),\varphi\right)_0$$

$$\forall\varphi\in V_h^p,\ F^n(x) = F(x, nk).$$

The same bound for the error can be proved if t runs through a finite interval $\langle 0, T\rangle$. For the infinite interval $\langle 0, \infty)$, such a bound cannot, of course, be proved unless some assumption on the growth of $F_t(x, t)$ is imposed.

The exact solution of the problem (1.1) has the property that

$$\|u(x, t)\|_0 \leqslant e^{-\lambda_1 t}\|g\|_0 \tag{1.10}$$

for any $g \in L_2(\Omega)$. Here λ_1 is the smallest (positive) eigenvalue of the operator $-Lu$. Under the additional assumptions that the roots of the polynomial $\rho(\zeta)$ with modulus equal to one are real and the modulus of all roots of the polynomial $\sigma(\zeta)$ is less than one, we prove that scheme (1.7) preserves the asymptotic behavior characteristic of (1.1), again without placing any restriction on the ratio kh^{-1}: if $U^j \in L_2(\Omega)$, $j = 0, \cdots, \nu-1$, then $\|U^n\|_0$ decreases exponentially,

$$\|U^n\|_0 \leqslant Ce^{-\alpha_0 nk} \max_{0\leqslant j\leqslant \nu-1} \|U^j\|_0, \qquad \alpha_0 = \text{const} > 0, \qquad n \geqslant \nu. \tag{1.11}$$

The backward differentiation multistep methods (see [7, p. 242]) with the step number $\nu \leqslant 6$ possess all the above properties.

2. Preliminaries. For simplicity, we assume that

$$a_{ij}(x), a(x), g(x) \in C^\infty(\bar{\Omega}), \qquad \Gamma \in C^\infty. \tag{2.1}$$

We state some facts about the solution $u(x, t)$ of (1.1). It is of the form $\Sigma_{i=1}^\infty g_i e^{-\lambda_i t}\psi_i(x)$ where λ_i and $\psi_i(x)$ are (positive) eigenvalues and (orthonormal) eigenfunctions, respectively, of the problem

$$-L\psi = \lambda\psi, \qquad \psi|_\Gamma = 0, \tag{2.2}$$

and g_i are the Fourier coefficients of the initial value $g(x)$. Ladyženskaja [6, Chapter III, Section 17] showed that if $g \in H^m$ and

$$g|_\Gamma = Lg|_\Gamma = \cdots = L^{[(m-1)/2]}g|_\Gamma = 0, \tag{2.3}$$

then $u(x, t) \in H^m$ for $t \geqslant 0$. Conversely, if $u(x, t) \in H^m$ for $t \geqslant 0$ then $g \in H^m$ and (2.3) is satisfied. The proof is based on two inequalities. The first holds for any series $\Sigma_{i=1}^\infty g_i\psi_i(x)$:

$$\left\|\sum_{i=1}^\infty g_i\psi_i(x)\right\|_m^2 \leqslant C\sum_{i=1}^\infty \lambda_i^m g_i^2. \tag{2.4}$$

(In the sequel, C is a generic constant, not necessarily the same in any two places, which does not depend on h, k, n, l, τ, g.) Concerning the other, we need only the following consequence: if $g \in H^m$ and (2.3) is satisfied, then it holds that

$$\sum_{i=1}^\infty \lambda_i^m g_i^2 \leqslant C\|g\|_m^2. \tag{2.5}$$

3. Convergence. The main results of the paper are contained in the following

THEOREM. *Let the linear multistep method (ρ, σ) be stable in the sense of Dahlquist, A_0-stable and of the order $q \geqslant 1$, and let the roots of the polynomial $\sigma(\zeta)$ with modulus equal to one be simple. Let* (1.2) *and* (2.1) *hold and g satisfy* (2.3) *with $m = \max(p+1, 2q)$ (this requirement is equivalent to the assumption $u(x, t) \in H^m$ for $t \geqslant 0$). Then, for arbitrary h, k, the discretization error is bounded by*

$$\sup_{\nu \leqslant n < \infty} \|u^n - U^n\|_0 \leqslant C\left[\sum_{j=0}^{\nu-1} \|u^j - U^j\|_0 + (h^{p+1} + k^q)\log\frac{1}{k}\,\|g\|_m\right]. \tag{3.1}$$

If, in addition, the roots of the polynomial $\rho(\zeta)$ with modulus equal to one are real and the modulus of all roots of $\sigma(\zeta)$ is less than one, then (1.11) *holds for any $U^j \in L_2(\Omega)$, $j = 0, \cdots, \nu - 1$.*

Proof. We first write u^n in the form $u^n = \xi^n + \eta^n$ with $\eta^n \in V_h^p$ being the Ritz approximation of u^n, i.e., the orthogonal projection of u^n onto V_h^p with the energy norm $[a(\cdot, \cdot)]^{1/2}$ (several authors have used this decomposition; we learned it from Bramble, Thomée [1]). Hence,

$$a(\eta^n, \varphi) = a(u^n, \varphi) = (-Lu^n, \varphi)_0 = (-\dot{u}^n, \varphi)_0, \qquad \forall \varphi \in V_h^p, \tag{3.2}$$

and with respect to (1.5), we find (see, e.g., [10]) that

$$\|\xi^n\|_0 \leqslant Ch^{p+1}\|Lu^n\|_{p-1} \leqslant Ch^{p+1}\|u^n\|_{p+1}.$$

By means of (2.4) and (2.5), we immediately obtain

$$\|\xi^n\|_0 \leqslant Ch^{p+1}\|g\|_{p+1}. \tag{3.3}$$

Therefore it is sufficient to prove for $\epsilon^n = \eta^n - U^n$

$$\max_{\nu \leqslant n < \infty} \|\epsilon^n\|_0 \leqslant C\left[\sum_{j=0}^{\nu-1} \|\epsilon^j\|_0 + (h^{p+1} + k^q)\log\frac{1}{k}\,\|g\|_m\right]. \tag{3.4}$$

If we use (3.2), we see that

$$\left(\sum_{j=0}^{\nu} \alpha_j \eta^{n+j}, \varphi\right)_0 + ka\left(\sum_{j=0}^{\nu} \beta_j \eta^{n+j}, \varphi\right) = (\pi^n - \omega^n, \varphi)_0, \quad \forall \varphi \in V_h^p, \tag{3.5}$$

where

$$\pi^n = \sum_{j=0}^{\nu} (\alpha_j u^{n+j} - k\beta_j \dot{u}^{n+j}), \qquad \omega^n = \sum_{j=0}^{\nu} \alpha_j \xi^{n+j}.$$

Subtracting (1.7) from (3.5), we get

$$\left(\sum_{j=0}^{\nu} \alpha_j \epsilon^{n+j}, \varphi\right)_0 + ka\left(\sum_{j=0}^{\nu} \beta_j \epsilon^{n+j}, \varphi\right) = (\pi^n - \omega^n, \varphi)_0, \quad \forall \varphi \in V_h^p. \tag{3.6}$$

We write (3.6) in a matrix form. For this purpose, let $\mathbf{w}$ be the vector $\mathbf{w} = M^{-1/2}\mathbf{v}$ ($\mathbf{v}$ is the basis vector; see Introduction) and let us set $\epsilon^n = (\epsilon^n)^T\mathbf{w}$ (notice that $\epsilon^n = \eta^n - U^n \in V_h^p$). Since $(\mathbf{v}, \mathbf{v}^T)_0 = M$ and $a(\mathbf{v}, \mathbf{v}^T) = K$, we have $(\mathbf{w}, \mathbf{w}^T)_0 = I$ and $a(\mathbf{w}, \mathbf{w}^T) = M^{-1/2}KM^{-1/2}$. The matrix $S = M^{-1/2}KM^{-1/2}$ is symmetric and positive definite. Putting the components w_i $(i = 1, \cdots, l)$ of the vector $\mathbf{w}$ for φ in (3.6), we get

$$\sum_{j=0}^{\nu} (\alpha_j I + \beta_j kS)\epsilon^{n+j} = \mathbf{c}^n,$$

where

$$\mathbf{c}^n = (\pi^n - \omega^n, \mathbf{w})_0. \tag{3.7}$$

Denote

$$\delta_j(\tau) = \frac{\alpha_j + \beta_j\tau}{\alpha_\nu + \beta_\nu\tau}, \quad j = 0, \cdots, \nu \quad (\delta_\nu(\tau) \equiv 1), \quad \mathbf{d}^n = (\alpha_\nu I + \beta_\nu kS)^{-1}\mathbf{c}^n \tag{3.8}$$

(the matrix $\alpha_\nu I + \beta_\nu kS$ is positive definite since $\alpha_\nu > 0$, $\beta_\nu > 0$). Then

$$\sum_{j=0}^{\nu} \delta_j(kS)\epsilon^{n+j} = \mathbf{d}^n \tag{3.9}$$

and this difference equation will be solved in the way described by Henrici [5, pp. 242–244].

We define the coefficients $\gamma_l(\tau)$ $(l = 0, 1, \cdots)$ by

$$\frac{1}{\hat{p}(\zeta, \tau)} \equiv [\delta_\nu(\tau) + \delta_{\nu-1}(\tau)\zeta + \cdots + \delta_0(\tau)\zeta^\nu]^{-1} = \gamma_0(\tau) + \gamma_1(\tau)\zeta + \cdots. \tag{3.10}$$

Similarly as in [5, p. 242], we can prove the estimate

$$|\gamma_l(\tau)| \leqslant C, \quad \tau \geqslant 0, \quad l = 0, 1, \cdots \tag{3.11}$$

(we leave out the proof even when it is not a trivial matter). We also get the identity (see [5, (5–160), p. 243])

$$\delta_\nu(\tau)\gamma_l(\tau) + \delta_{\nu-1}(\tau)\gamma_{l-1}(\tau) + \cdots + \delta_0(\tau)\gamma_{l-\nu}(\tau) \equiv \begin{cases} 1, & l = 0, \\ 0, & l > 0. \end{cases} \tag{3.12}$$

Now we write (3.9) with $n - \nu - l$ instead of n, multiply by $\gamma_l(kS)$ and sum for $l = 0, 1, \cdots, n - \nu$. After some rearranging and using (3.12), we obtain (see [5, p. 243])

$$\begin{aligned} \epsilon^n = &- [\delta_{\nu-1}(kS)\gamma_{n-\nu}(kS) + \cdots + \delta_0(kS)\gamma_{n-2\nu+1}(kS)]\epsilon^{\nu-1} - \cdots \\ &- \delta_0(kS)\gamma_{n-\nu}(kS)\epsilon^0 + \sum_{l=0}^{n-\nu} \gamma_l(kS)\mathbf{d}^{n-\nu-l}. \end{aligned} \tag{3.13}$$

The coefficients $\delta_j(\tau)$ are bounded functions in the interval $\langle 0, \infty)$. Therefore, $\|\delta_j(kS)\| = \max_\Lambda |\delta_j(k\Lambda)|$ (by $\|\cdot\|$ we denote the Euclidean norm of a vector or a matrix) where Λ

are the eigenvalues of S. These are positive, consequently $\|\delta_j(kS)\| \leqslant \sup_{0<\tau}|\delta_j(\tau)| = 0(1)$. The coefficients $\gamma_l(\tau)$ are bounded by (3.11), hence $\|\gamma_l(kS)\| = 0(1)$, $l = 0, 1, \cdots$. Furthermore, the starting errors are bounded by

$$\|\epsilon^j\|_0 = \|u^j - U^j - \xi^j\|_0 \leqslant \|u^j - U^j\|_0 + \|\xi^j\|_0 \leqslant \|u^j - U^j\|_0 + Ch^{p+1}\|g\|_{p+1}.$$

Hence

$$\sum_{j=0}^{\nu-1}\|\epsilon^j\|_0 \leqslant \sum_{j=0}^{\nu-1}\|u^j - U^j\|_0 + Ch^{p+1}\|g\|_m.$$

Also $\|\mathbf{d}^n\| \leqslant \|(\alpha_\nu I + \beta_\nu kS)^{-1}\|\,\|\mathbf{c}^n\| \leqslant \alpha_\nu^{-1}\|\mathbf{c}^n\|$. Thus, we see that from (3.13) it follows

$$\|\epsilon^n\| \leqslant C\left(\sum_{j=0}^{\nu-1}\|\epsilon^j\| + \sum_{r=0}^{n-\nu}\|\mathbf{c}^r\|\right).$$

Since

$$\|\epsilon^n\|_0^2 = (\epsilon^n)^T(\mathbf{w}, \mathbf{w}^T)_0\epsilon^n = \|\epsilon^n\|^2,$$

we have

$$\|\epsilon^n\|_0 \leqslant C\left(\sum_{j=0}^{\nu-1}\|u^j - U^j\|_0 + h^{p+1}\|g\|_m + \sum_{r=0}^{n-\nu}\|\mathbf{c}^r\|\right). \tag{3.14}$$

We need a bound for $\|\mathbf{c}^r\|$. $\mathbf{c}^r$ is of the form (3.7). If $f \in L_2(\Omega)$ and $\hat{f} = \hat{\mathbf{f}}^T\mathbf{w} \in V_h^p$ is the orthogonal projection of f onto V_h^p with the norm $\|\cdot\|_0$, we easily find that $\hat{\mathbf{f}} = (f, \mathbf{w})_0$ and that $\|\hat{f}\|_0 \leqslant \|f\|_0$. Since $\|\hat{f}\|_0 = \|\hat{\mathbf{f}}\|$, we have $\|\hat{\mathbf{f}}\| \leqslant \|f\|_0$. Therefore,

$$\|\mathbf{c}^r\| \leqslant \|\pi^r\|_0 + \|\omega^r\|_0. \tag{3.15}$$

To estimate $\|\pi^r\|_0$, we use the assumption that the scheme (ρ, σ) is of the order q. It means that for any function $y(t) \in C^{(s)}$, $s \leqslant q + 1$, it holds that

$$\sum_{j=0}^{\nu}\alpha_j y(t + jk) - k\sum_{j=0}^{\nu}\beta_j\dot{y}(t + jk) = O\left(k^s \max_{0<\tau<\nu k}|y^{(s)}(t + \tau)|\right) \tag{3.16}$$

(it follows from the formula (5.178) in [5, p. 248]). Set $y(t) = e^{-\lambda_i t}$, $s = q + 1$. After dividing by $e^{-\lambda_i t}$, we get

$$\sum_{j=0}^{\nu}(\alpha_j + \beta_j k\lambda_i)e^{-jk\lambda_i} = O(k^{q+1}\lambda_i^{q+1}).$$

The Fourier coefficients π_i^r of π^r are evidently equal to

$$e^{-rk\lambda_i}g_i\sum_{j=0}^{\nu}(\alpha_j + \beta_j k\lambda_i)e^{-jk\lambda_i}.$$

Therefore,

$$(\pi_i^r)^2 \leqslant Ce^{-2rk\lambda_i} k^{2(q+1)} \lambda_i^{2(q+1)} g_i^2.$$

Since $e^{-2rk\lambda_i} k^2 \lambda_i^2 \leqslant e^{-rk\lambda_1} (e^{-\frac{1}{2}rk\lambda_i} k\lambda_i)^2 \leqslant r^{-2} e^{-rk\lambda_1}$ if $r \geqslant 1$ (due to $xe^{-\alpha x} \leqslant (e\alpha)^{-1} < (2\alpha)^{-1}$), we have $(\pi_i^r)^2 \leqslant Cr^{-2} e^{-rk\lambda_1} k^{2q} \lambda_i^{2q} g_i^2$ and

$$\|\pi^r\|_0^2 = \sum_{i=1}^{\infty} (\pi_i^r)^2 \leqslant Cr^{-2} e^{-rk\lambda_1} k^{2q} \|g\|_m^2, \qquad r \geqslant 1. \tag{3.17}$$

Concerning π^0, we use (3.16) with $s = q$ and we get

$$\|\pi^0\|_0^2 \leqslant Ck^{2q} \|g\|_m^2. \tag{3.18}$$

The estimates for $\|\omega^r\|_0$ can be obtained in a similar way. Set $z^r = \Sigma_{j=0}^{\nu} \alpha_j u^{r+j}$ and write z^r as the sum $x^r + y^r$, where y^r is the Ritz approximation of z^r. Then $\|x^r\|_0 \leqslant Ch^{p+1} \|z^r\|_{p+1}$. Since η^n is the Ritz approximation of u^n, we have $y^r = \Sigma_{j=0}^{\nu} \alpha_j \eta^{r+j}$ and $x^r = \Sigma_{j=0}^{\nu} \alpha_j \xi^{r+j} = \omega^r$, hence

$$\|\omega^r\|_0 \leqslant Ch^{p+1} \|z^r\|_{p+1}. \tag{3.19}$$

The Fourier coefficients z_i^r are equal to $\Sigma_{j=0}^{\nu} \alpha_j e^{-(r+j)k\lambda_i} g_i = e^{-rk\lambda_i} g_i \Sigma_{j=0}^{\nu} \alpha_j e^{-jk\lambda_i}$. Because of the consistency of the scheme (ρ, σ), it follows that $\Sigma_{j=0}^{\nu} \alpha_j = 0$. Therefore, $\Sigma_{j=0}^{\nu} \alpha_j e^{-jk\lambda_i} = O(k\lambda_i)$, consequently $(z_i^r)^2 \leqslant Ce^{-2rk\lambda_i} (k\lambda_i)^2 g_i^2$; by means of (3.19), we find in the same way as before that

$$\begin{aligned} &\|\omega^r\|_0^2 \leqslant Cr^{-2} e^{-rk\lambda_1} h^{2(p+1)} \|g\|_m^2, \qquad r \geqslant 1, \\ &\|\omega^0\|_0^2 \leqslant Ch^{2(p+1)} \|g\|_m^2. \end{aligned} \tag{3.20}$$

(3.17), (3.18), (3.20) and (3.15) give

$$\begin{aligned} &\|\mathbf{c}^r\| \leqslant Cr^{-1} e^{-\frac{1}{2}rk\lambda_1} (h^{p+1} + k^q) \|g\|_m, \qquad r \geqslant 1, \\ &\|\mathbf{c}^0\| \leqslant C(h^{p+1} + k^q) \|g\|_m. \end{aligned}$$

We come back to (3.14). If we find out that

$$\sum_{r=1}^{\infty} \|\mathbf{c}^r\| \leqslant C(h^{p+1} + k^q) \log \frac{1}{k} \|g\|_m,$$

the bound (3.1) is proved. For this purpose, it is sufficient to realize that

$$\sum_{r=1}^{\infty} r^{-1} e^{-\frac{1}{2}rk\lambda_1} = \sum_{r=1}^{\infty} r^{-1} (e^{-\frac{1}{2}k\lambda_1})^r = -\log(1 - e^{-\frac{1}{2}k\lambda_1}) = O\left(\log \frac{1}{k}\right).$$

To prove the second part of the Theorem, we need a better estimate of the coefficients $\gamma_l(\tau)$. According to our assumptions, the roots λ of $\rho(\zeta)$ with modulus equal to one are simple and real (i.e., $\lambda = \pm 1$). The associated roots $\zeta(\tau)$ of the polynomial $p(\zeta)$ defined by (1.6) have the expansion $\zeta(\tau) = \lambda + a_1 \tau + O(\tau^2)$, with $a_1 = -\sigma(\lambda)/\rho'(\lambda)$

$\neq 0$ (since $\sigma(\lambda) = 0$ means that λ would be a root of $p(\zeta)$). Therefore, $|\zeta(\tau)| = |1 + \lambda^{-1}a_1\tau + O(\tau^2)|$. The growth parameter $a = \lambda^{-1}a_1$ is different from zero and real, hence $|\zeta(\tau)| = [1 + 2a\tau + O(\tau^2)]^{1/2}$. As a must be negative (otherwise $|\zeta(\tau)| > 1$), we see that, for τ sufficiently small, $|\zeta(\tau)| \leqslant 1 - c\tau$, $c = \frac{1}{2}|a| > 0$. The other roots of $p(\zeta)$ have modulus less than one for $\tau \geqslant 0$. On the basis of these facts, we can prove that $|\gamma_l(\tau)| \leqslant C(1 - \frac{1}{2}c\tau)^l$ for $0 < \tau \leqslant \tau_1$ if τ_1 is a sufficiently small number. In the interval $\langle \tau_1, \infty)$, it follows easily (from the assumption that the modulus of all roots of $\sigma(\zeta)$ is less than one) that $|\gamma_l(\tau)| \leqslant C(1 - \vartheta)^l$, $0 < \vartheta < 1$. Hence

$$|\gamma_l(\tau)| \leqslant C[\max(1 - \tfrac{1}{2}c\tau, 1 - \vartheta)]^l, \qquad \tau > 0,$$

and, because the eigenvalues of the matrix kS are bounded from below by $k\lambda_1$, it holds for k sufficiently small that

$$\|\gamma_l(kS)\| \leqslant C(1 - \tfrac{1}{2}c\lambda_1 k)^l \leqslant Ce^{-\alpha_0 lk}, \quad \alpha_0 = \tfrac{1}{4}c\lambda_1 > 0. \tag{3.21}$$

To complete the proof, we set $U^n = (\mathbf{a}^n)^T\mathbf{w}$ and get (in the same way as we got (3.13))

$$\mathbf{a}^n = -[\delta_{\nu-1}(kS)\gamma_{n-\nu}(kS) + \cdots + \delta_0(kS)\gamma_{n-2\nu+1}(kS)]\,\mathbf{a}^{\nu-1} - \cdots - \delta_0(kS)\gamma_{n-\nu}(kS)\mathbf{a}^0.$$

The estimate (1.12) follows immediately from (3.21).

4. Some Remarks. 1. To get the estimate (1.9) for the condition number of the matrix $B = \alpha_\nu M + \beta_\nu kK$, we assume the following additional properties of the basis $\{v_1, \cdots, v_l\}$ of the space V_h^p: if $\varphi = \mathbf{x}^T\mathbf{v} \in V_h^p$, then

(a) $$ch^{-N}\|\varphi\|_0^2 \leqslant \|\mathbf{x}\|^2 \leqslant Ch^{-N}\|\varphi\|_0^2, \qquad c = \text{const} > 0,$$

(b) $$a(\varphi, \varphi) \leqslant Ch^{-2}\|\varphi\|_0^2.$$

The finite element subspaces used in applications possess these properties.

Let Λ be an eigenvalue of the matrix B and $\mathbf{x}$ the corresponding eigenvector. We have $\alpha_\nu M\mathbf{x} + \beta_\nu kK\mathbf{x} = \Lambda\mathbf{x}$; multiplying this equation by $\mathbf{x}^T$ and setting $\varphi = \mathbf{x}^T\mathbf{v}$, we get

$$\Lambda = \alpha_\nu \frac{\|\varphi\|_0^2}{\|\mathbf{x}\|^2} + \beta_\nu k \frac{a(\varphi, \varphi)}{\|\mathbf{x}\|^2}.$$

Hence

$$\Lambda_{\max} \leqslant C(1 + kh^{-2})h^N, \qquad \Lambda_{\min} \geqslant c_1 h^N, \qquad c_1 = \text{const} > 0,$$

and, if we exclude the uninteresting case $kh^{-2} \longrightarrow 0$, we have $\operatorname{cond}(B) = \Lambda_{\max}/\Lambda_{\min} = O(kh^{-2})$.

2. The backward differentiation methods with $\nu = q \leqslant 6$ are stable in the sense

of Dahlquist, the only root of $\rho(\zeta)$ with modulus equal to one is the principal root $\zeta = 1$ (see Cryer [4]), and they are A_0-stable (actually, they are $A(\alpha)$-stable with $90° \geqslant \alpha \geqslant 18°$, see [7, p. 242]). Further, the only root of $\sigma(\zeta)$ is zero. Hence, these methods fulfill all assumptions of the Theorem.

3. It is known (see [7, p. 243]) that the implicit R-stage Runge-Kutta methods of order $2R$ are A-stable. It will be shown elsewhere that if we discretize (1.4) by means of such a method, the error is bounded by

$$\sup_{1 \leqslant n < \infty} \|u^n - U^n\|_0 \leqslant \|u^0 - U^0\|_0 + C(h^{p+1} + k^{2R}) \log \frac{1}{k} \|g\|_m,$$

under the assumption that $g(x)$ satisfies (2.3) with $m = \max(p + 1, 4R)$.

Computing Center of the Technical University
Obránců míru 21
602 00 Brno, Czechoslovakia

1. J. H. BRAMBLE & V. THOMÉE, "Discrete time Galerkin methods for a parabolic boundary value problem." (To appear.)

2. P. G. CIARLET & P. A. RAVIART, "Interpolation theory over curved elements, with applications to finite element methods," *Comp. Meth. Appl. Mech. Eng.*, v. 1, 1972, pp. 217–249.

3. C. W. CRYER, "A new class of highly-stable methods: A_0-stable methods," *BIT*, v. 13, 1973, pp. 153–159.

4. C. W. CRYER, "On the instability of high order backward-difference multistep methods," *BIT*, v. 12, 1972, pp. 17–25.

5. P. HENRICI, *Discrete Variable Methods in Ordinary Differential Equations*, Wiley, New York-London, 1962. MR 24 #B1772.

6. O. A. LADYŽENSKAJA, V. A. SOLONNIKOV & N. N. URAL'CEVA, *Linear and Quasi-Linear Equations of Parabolic Type*, Transl. Math. Monographs, vol. 23, Amer. Math. Soc., Providence, R. I., 1968. MR 39 #3159b.

7. J. D. LAMBERT, *Computational Methods in Ordinary Differential Equations*, Wiley, London, 1973.

8. O. C. ZIENKIEWICZ, *The Finite Element Method in Engineering Science*, McGraw-Hill, London, 1971. MR 47 #4518.

9. M. ZLÁMAL, "Finite element methods for parabolic equations," *Math. Comp.*, v. 28, 1973, pp. 393–404.

10. M. ZLÁMAL, "Curved elements in the finite element method. I," *SIAM J. Numer. Anal.*, v. 10, 1973, pp. 229–240.

11. M. ZLÁMAL, "Curved elements in the finite element method. II," *SIAM J. Numer. Anal.*, v. 11, 1974, pp. 347–362.