JOSEF NEDOMA

The finite element solution of elliptic and parabolic equations using simplicial isoparametric elements

RAIRO. Analyse numérique, tome 13, n° 3 (1979), p. 257-289

<http://www.numdam.org/item?id=M2AN_1979__13_3_257_0>
THE FINITE ELEMENT SOLUTION
OF ELLIPTIC AND PARABOLIC EQUATIONS
USING SIMPLICIAL ISOPARAMETRIC ELEMENTS (*)

par Josef NEDOMA (1)

Abstract. — Error bounds introduced in [7] were given for fully discretized approximate solutions of parabolic equations by the finite element method. For time discretization the A-stable linear v-step methods (for \( v = 1 \) or 2) were used. In this paper the \( A_0 \)-stable linear v-step methods for any \( v \) are used for time discretization. It is known that A-stable methods for \( v = 1, 2 \) are included in the class of \( A_0 \)-stable methods. The consideration for the elliptic equations is similar to the parabolic equations. Hence, the error bounds for elliptic equations are formulated in this paper too.

Résumé. — On a donné en [7] des majorations de l’erreur pour des approximations complètement discrètes d’équations paraboliques par la méthode des éléments finis. On utilisait des méthodes linéaires A-stables à \( v \) pas (\( v = 1 \) ou 2) pour la discrétisation en temps. Dans cet article, on utilise des méthodes linéaires \( A_0 \)-stables à \( v \) pas, \( v \) quelconque, pour la discrétisation en temps. On sait que les méthodes A-stables pour \( v = 1, 2 \) sont incluses dans la classe des méthodes \( A_0 \)-stables. Les développements étant semblables dans les cas elliptiques et paraboliques, on énonce également dans cet article les majorations d’erreurs pour les équations elliptiques.

1. CONSTRUCTION OF THE FINITE ELEMENT SPACE. NOTATION

We consider the \( k \)-regular family \( \{ K \}_h \) of simplicial isoparametric finite elements \( K \) introduced by Ciarlet and Raviart [3]. Hence, the simplicial element \( K \in \{ K \}_h \) is the image of the unit \( n \)-simplex \( \hat{K} \) \((\hat{K} \) is the closed convex hull of a set \( \hat{\Sigma} = \bigcup_{i=1}^{N} \{ \hat{a}_i \} \)) through the unique mapping \( F_K : \hat{K} \rightarrow \mathbb{R}^n \) (the mapping \( F_K \) is supposed to be a \( C^{k+1} \)-diffeomorphism) such that \( F_K \in \bar{P}^n \), \( F_K (\hat{a}_i) = a_i \) \((\bar{P} \subset C^{k+1}(\hat{K}) \) is a finite dimensional space of functions defined on \( \hat{K} \) with

(*) Reçu septembre 1978.
(1) Technical University, Laboratoř počítacích strojů V.U.T., Brno, Tchécoslovaquie.
dim $P = N$ such that $\hat{S}$ is $\hat{P}$-unisolvent and $\hat{P} \supset \hat{P}(1)$, where for any integer $r \geq 0$, $\hat{P}(r)$ is the space of restrictions to $\hat{K}$ of all polynomials of degree $\leq r$ in $n$ variables $\hat{x}_1, \ldots, \hat{x}_n$ and there exist constants $c_i$, $0 \leq i \leq k + 1$, independent of $h$ such that for all $h$:

$$\sup \max_{\hat{x} \in K} |D^a F_{K}(\hat{x})| \leq c_1 h^i, \quad 1 \leq i \leq k + 1$$

and

$$0 < \frac{1}{c_0} h^n \leq |J_K(\hat{x})| \leq c_0 h^n,$$

where $\alpha = (\alpha_1, \ldots, \alpha_n)$. $|\alpha| = \alpha_1 + \ldots + \alpha_n$ and $J_K(\hat{x})$ is the Jacobian of the mapping $F_K$ at the point $x \in \hat{K}$.

To every element $K$ there is associated the finite dimensional space $P_K$ (with $\dim P_K = N$) of functions

$$P_K = \{ p_K : K \to R; p_K = p^*(F_K^{-1}), \forall p^* \in \hat{P} \}. \quad (1.3)$$

The $K$-interpolate $\pi_K u$ of a given function $u : K \to R$ is the unique function which satisfies

$$\pi_K u \in P_K, \quad \pi_K u(a_i) = u(a_i), \quad 1 \leq i \leq N. \quad (1.4)$$

For a $k$-regular family $\{ K \}_h$ of finite elements the following interpolation theorem is true (see Ciarlet and Raviart [3], theorem 2, p. 429).

**Lemma 1.1** (interpolation theorem): Let a $k$-regular family $\{ K \}_h$ of simplicial elements such that $\hat{P}(k) \subset \hat{P}$ be given. Let

$$k > \frac{n}{2} - 1. \quad (1.5)$$

Then for any integer $i$ such that $0 \leq i \leq k + 1$, there exists a constant $c$ independent of $h$ such that for any $K \in \{ K \}_h$ and for any function $u \in H^{k+1}(K)$ we have

$$|u - \pi_K u|_{1,K} \leq ch^{k+1-i} \|u\|_{k+1,K}. \quad (1.6)$$

Here the following notation is used:

The norm and the scalar product in the space $L^2(A)$ is denoted by $\|\cdot\|_{0,A}$ and $(\cdot, \cdot)_{0,A}$ respectively.

$H^m(A) \equiv W^{m}(A)$, $m = 0, 1 \ldots$ is a Sobolev space with the norm

$$\|v\|_{m,A} = \left( \sum_{i=0}^{m} |v|_{i,A}^2 \right)^{1/2}, \quad \text{where} \quad |v|_{i,A} = (\sum_{|a|=i} \|D^a v\|_{0,A}^2)^{1/2}.$$
In the sequel we mean by $\Omega$ a bounded domain in $\mathbb{R}^n$ with a sufficiently smooth boundary $\partial \Omega$.

Using the way described by Ciarlet and Raviart [3] we define a $k$-regular triangulation $\mathcal{C}_h$ of $\Omega$. Let $\Omega_h$ be the union of a finite number of simplicial elements $K$. Every element $K = F_K(\overline{K})$ is determined by $N$ points $a_i, K$. We suppose that all points $a_i, K$ belong to $\overline{\Omega}$. The family of elements constructed in this way is called a triangulation of $\Omega$ and is denoted by $\mathcal{C}_h$. We say that a triangulation $\mathcal{C}_h$ of $\Omega$ is $k$-regular if:

a) the family of all elements from which the triangulation is formed is $k$-regular;

b) the geometrical shape of any “face” $\Delta$ of a given element $K \in \mathcal{C}_h$ must be completely determined by those points $a_i, K$ which belong to $\Delta$;

c) for the boundary elements (i.e. for elements $K \subseteq \partial \Omega$) of the triangulation $\mathcal{C}_h$ we have

$$\max_{y' \in \Delta} |\psi_h(y') - \psi(y')| \leq ch^{k+1},$$

(1.7)

where $c$ is a constant independent of $h$ and the notation is that of figure.

To a given $k$-regular triangulation $\mathcal{C}_h$ there is associated the finite dimensional space $V_h$ of functions $v$ defined by

$$V_h = \{ v \in C^0(\overline{\Omega}_h); v_K \in P_K, \forall K \in \mathcal{C}_h, v = 0 \text{ on } \partial \Omega_h \},$$

(1.8)

where $v_K$ is the restriction of the function $v$ to the set $K$. 

vol. 13, n° 3, 1979
Next, to any function $v$ defined on $\Omega$ or on $\Omega_h$ we may associate its unique interpolate $\pi_h v$, which satisfies

$$\pi_h v = \pi_K v, \quad \forall K \in \mathcal{G}_h. \quad (1.9)$$

In our paper we suppose that $\hat{P} \equiv \hat{P}(k)$. This restriction is not essential. It enables us to give simpler proofs.

In the sequel we use the following notation:

- $H^1_0(A)$ is the closure of the set $C_0^\infty(A)$ (i.e. of the set of infinitely differentiable functions with compact support in $A$) in the norm $\| \cdot \|_{1,A}$.
- $H^{-1}(A)$ is the space dual to $H^1_0(A)$ (with dual norm).
- $L^\infty(H^m(A))$ is the space of all functions $\varphi(x, t), x = (x_1, \ldots, x_n) \in A, t \in [0, T]$ such that $\varphi(x, t) \in H^m(A), \forall t \in [0, T]$ and the function $\|\varphi(x, t)\|_{m,A}$ is bounded for almost all $t \in [0, T]$.

Let $\Phi(x)$ be any function defined on the element $K$. Then the function $\Phi(F_K(\hat{x}))$ is defined on $\hat{K}$. In the sequel we will denote it by $\Phi^*(\hat{x})$.

In the sequel the constants independent of $h$ will be denoted by $c$. The notation is generic, i.e. $c$ will not denote necessarily the same constant in any two places.

### 2. ISOPARAMETRIC INTEGRATION

In the same way as in Ciarlet and Raviart [3] let us suppose that we have at our disposal a quadrature formula of degree $d$ over the reference set $\hat{K}$. In other words

$$\int_K \varphi(\hat{x}) d\hat{x} \text{ is approximated by } \sum_r \hat{\omega}_r \varphi(\hat{b}_r) \quad (2.1)$$

for some specified points $\hat{b}_r \in \hat{K}$ and weights $\hat{\omega}_r$, which will be assumed once and for all to satisfy

$$\hat{\omega}_r > 0. \quad (2.2)$$

This assumption is by no means necessary but it yields simpler proofs. Concerning $\hat{b}_r$ we suppose that for every $r$, $\hat{b}_r$ either lies inside $\hat{K}$ or it coincides with some of the points $\hat{a}_r$. With the quadrature scheme (2.1) we associate the error

$$\hat{E}(\varphi) = \int_K \varphi(\hat{x}) d\hat{x} - \sum_r \hat{\omega}_r \varphi(\hat{b}_r). \quad (2.3)$$
Using the standard formula for change of variables in multiple integrals, we find that

\[ \int_K \varphi(x) \, dx \text{ is approximated by } \sum_r \omega_{r,K} \varphi(b_{r,K}), \]  

(2.4)

where

\[ \omega_{r,K} = \hat{\omega}_r J_K(b_r), \quad b_{r,K} = F_K(b_r). \]  

(2.5)

We may, and will, assume that \( J_K(\hat{x}) > 0, \forall \hat{x} \in \hat{K}. \) We see that the quadrature scheme (2.1) over the reference set \( \hat{K} \) induces the quadrature scheme (2.4) over the element \( K \), a circumstance which is called by Ciarlet and Raviart [3] “isoparametric numerical integration”. With the scheme (2.4) we associate the error

\[ E_K(\varphi) = \int_K \varphi(x) \, dx - \sum_r \omega_{r,K} \varphi(b_{r,K}) \]  

(2.6)

so that we have

\[ E_K(\varphi) = \hat{E}(\varphi^* J_K) \quad \text{and} \quad \hat{E}(\varphi^*) = E_K(\varphi J_K^{-1}). \]  

(2.7)

In the sequel we will denote

\[ E(\varphi) = \sum_{K \in E_h} E_K(\varphi) \text{ for any function } \varphi. \]  

(2.8)

Now, we derive two theorems concerning isoparametric numerical integration. Before, we give some technical lemmas.

**Lemma 2.1:** Let \( D^\beta \varphi_i = O(h|\beta| + x^\gamma) \) for \( i = 1, \ldots, s, |\beta| = 0, \ldots, |\alpha|. \) Then

\[ D^\alpha(\varphi_1 \varphi_2 \ldots \varphi_s) = O(h^{|\alpha| + \gamma_1 + \ldots + \gamma_s}). \]  

(2.9)

The proof is trivial using the mathematical induction.

**Lemma 2.2:** For polynomials \( r, s \) on the reference set \( \hat{K} \) the following inequalities are true

\[ \max_{\hat{K}} |D^{\alpha} r| \leq c_1 |r|_{\alpha, \hat{K}}, \]  

(2.10)

\[ |r|_{j, \hat{K}}^2 \leq c_2 |r|_{j, \hat{K}}^2 \quad \text{for } j \geq 0, \]  

(2.11)

\[ |rs|_{i, \hat{K}}^2 \leq c_3 \sum_{j=0}^i |r|_{j, \hat{K}}^2 |s|_{i-j, \hat{K}}^2, \]  

(2.12)

where \( c_1, c_2, c_3 \) are constants.
The proof follows from Zlámal’s paper [12], p. 356 and from lemma 3 in [7].

**Lemma 2.3:** Let $\mathcal{G}_h$ be a $k$-regular triangulation of $\Omega$. Let $J_K^{(r,p)}$ be a cofactor of the Jacobian $J_K$. Then

\begin{align}
D^a J_K &= O(h^{|a| + n}), \\
D^a J_K^{(r,p)} &= O(h^{|a| + n - 1}), \\
D^a \left( \frac{1}{J_K} \right) &= O(h^{|a| - n}).
\end{align}

For the proof see Lemma 5 in [7].

**Lemma 2.4:** Let $\tau^* \in H^{k+1}(\hat{K})$, $\tau \in H^{k+1}(K)$, $K \in \mathcal{G}_h$, $\mathcal{G}_h$ be a $k$-regular triangulation of $\Omega$. Then there exists a constant $c$ independent of $h$ such that

\begin{equation}
|\tau^*|_{k+1,K} \leq ch^{-(n/2)+k+1} \|\tau\|_{k+1,K}.
\end{equation}

Lemma is an immediate consequence of Lemma 1 from [3], p. 427.

**Lemma 2.5:** Let $\varphi \in H^s(\hat{K})$, where $s > n/2$ and let $\pi_{s-1} \varphi$ be a polynomial of degree $s - 1$ which uniquely interpolates the function $\varphi$ on $\hat{K}$. Then there exists a constant $c$ such that

\begin{equation}
|\varphi - \pi_{s-1} \varphi|_{j,K} \leq c |\varphi|_{s,K} \quad \text{for} \quad j = 0, \ldots, s.
\end{equation}

Lemma follows from Bramble and Hilbert paper [2], p. 812.

**Lemma 2.6:** Let $\psi(\hat{x}) \in H^s(\hat{K})$, where

\begin{equation}
s > \frac{n}{2},
\end{equation}

$\tau(\hat{x})$ be a polynomial of degree $\leq r$, where

\begin{equation}
r \leq s,
\end{equation}

$\delta(\hat{x}) \in C^s(\hat{K})$ be a function such that

\begin{equation}
D^\alpha \delta = O(h^{|\alpha| + s}) \quad \text{for} \quad 0 \leq |\alpha| \leq s, \forall s \quad \text{some int.}
\end{equation}

Let $d$ be the order of a quadrature formula on the reference set $\hat{K}$ such that

\begin{equation}
d > \frac{n}{2} - 1.
\end{equation}
Then there exists a constant $c$ such that

$$|\hat{\mathcal{E}}(\delta\psi\tau)|^2 \leq ch^{2s}\left\{ (h^{2s}\|\psi\|^2_{0,\mathcal{K}} + \|\psi\|^2_{0,\mathcal{K}})\|\tau\|^2_{0,\mathcal{K}} + h^{2(d+1)}\left( \sum_{i=0}^{s-1} h^{-2i}\|\tau\|^2_{i,\mathcal{K}} \right)\right\}.$$  \hspace{1cm} (2.22)

When supposing, in addition, that $\psi(\tilde{x})$ is a polynomial of degree $\leq r$, then there exists a constant $c$ such that

$$|\hat{\mathcal{E}}(\delta\psi\tau)|^2 \leq ch^{2s}\left\{ h^{2s}\|\psi\|^2_{0,\mathcal{K}}\|\tau\|^2_{i,\mathcal{K}} + h^{2(d+3)}\sum_{i=1}^r h^{-2i}\|\psi\|^2_{i,\mathcal{K}}\sum_{i=1}^r h^{-2i}\|\tau\|^2_{i,\mathcal{K}} \right\}.$$  \hspace{1cm} (2.23)

**Proof:** Evidently

$$|\hat{\mathcal{E}}(\delta\psi\tau)| \leq |\hat{\mathcal{E}}((\delta - \pi_{s-1}\delta)(\psi - \pi_{s-1}\psi)\tau)| + |\hat{\mathcal{E}}((\delta - \pi_{s-1}\delta)\pi_{s-1}\psi\tau)|$$
$$+ |\hat{\mathcal{E}}(\pi_{s-1}\delta(\psi - \pi_{s-1}\psi)\tau)| + |\hat{\mathcal{E}}(\pi_{s-1}\delta\pi_{s-1}\psi\tau)|.$$  \hspace{1cm} (2.24)

From (2.3), from the first Sobolev theorem and from lemma 2.5 it follows

$$|\hat{\mathcal{E}}((\delta - \pi_{s-1}\delta)(\psi - \pi_{s-1}\psi)\tau)|$$
$$\leq c\sup_{\mathcal{K}}|\delta - \pi_{s-1}\delta|\sup_{\mathcal{K}}|\psi - \pi_{s-1}\psi|\max_{\mathcal{K}}|\tau|$$
$$\leq c\|\delta - \pi_{s-1}\delta\|_{s,\mathcal{K}}\|\psi - \pi_{s-1}\psi\|_{s,\mathcal{K}}\max_{\mathcal{K}}|\tau|$$
$$\leq c|\delta|_{s,\mathcal{K}}\|\psi\|_{s,\mathcal{K}}\max_{\mathcal{K}}|\tau|. $$

Hence, from (2.20) and from lemma 2.2 we get

$$|\hat{\mathcal{E}}((\delta - \pi_{s-1}\delta)(\psi - \pi_{s-1}\psi)\tau)|^2 \leq ch^{2(s+r)}\|\psi\|^2_{0,\mathcal{K}}\|\tau\|^2_{0,\mathcal{K}}.$$  \hspace{1cm} (2.25)

Similarly we obtain

$$|\hat{\mathcal{E}}((\delta - \pi_{s-1}\delta)\pi_{s-1}\psi\tau)|^2 \leq ch^{2(s+r)}\|\psi\|^2_{0,\mathcal{K}}\|\tau\|^2_{0,\mathcal{K}},$$  \hspace{1cm} (2.26)

$$|\hat{\mathcal{E}}(\pi_{s-1}\delta(\psi - \pi_{s-1}\psi)\tau)|^2 \leq ch^{2(r+s)}\|\psi\|^2_{0,\mathcal{K}}\|\tau\|^2_{0,\mathcal{K}}.$$  \hspace{1cm} (2.27)

Evidently

$$|\pi_{s-1}\varphi|_{i,\mathcal{K}} \leq |\pi_{s-1}\varphi - \varphi|_{i,\mathcal{K}} + |\varphi|_{i,\mathcal{K}} \leq c(|\varphi|_{s,\mathcal{K}} + |\varphi|_{i,\mathcal{K}}), \quad 0 \leq i \leq s.$$
Hence
\begin{align}
|\pi_{s-1} \psi|_{i, \mathcal{K}} &\leq c (|\psi|_{s, \mathcal{K}} + |\psi|_{i, \mathcal{K}}), \quad 0 \leq i \leq s, \\
|\pi_{s-1} \delta|_{i, \mathcal{K}} &\leq c (|\delta|_{s, \mathcal{K}} + |\delta|_{i, \mathcal{K}}) \leq c (h^{s+\mathcal{K}} + h^{i+\mathcal{K}}) \leq ch^{i+\mathcal{K}}.
\end{align}

Let us remember that the inequality (2.29) is true also for \( i > s \) since \( |\pi_{s-1} \delta|_{i, \mathcal{K}} = 0 \). From the Bramble-Hilbert lemma (see [1]), from lemma 2.2 and from (2.29) we get
\begin{align}
|\hat{E}(\pi_{s-1} \delta \pi_{s-1} \psi \tau)|^2 &\leq c \left| \pi_{s-1} \delta \pi_{s-1} \psi \tau \right|_{d+1, \mathcal{K}}^2 \\
&\leq c \sum_{j=0}^{d+1} \left| \tau \pi_{s-1} \psi \right|_{j, \mathcal{K}}^2 \left| \pi_{s-1} \delta \right|_{d+1-j, \mathcal{K}}^2 \\
&\leq c h^{2(d+1)} \sum_{j=0}^{2s} h^{-2j} \left| \tau \pi_{s-1} \psi \right|_{j, \mathcal{K}}^2 \\
&\leq c h^{2(d+1)} \sum_{j=0}^{2s} h^{-2j} \sum_{i=0}^{j} \left| \tau \right|_{i, \mathcal{K}}^2 \left| \pi_{s-1} \psi \right|_{i, \mathcal{K}}^2.
\end{align}

It is easy to verify that
\begin{align}
\sum_{j=0}^{2s} h^{-2j} \sum_{i=0}^{j} \left| \tau \right|_{i, \mathcal{K}}^2 \left| \pi_{s-1} \psi \right|_{i, \mathcal{K}}^2 = \left( \sum_{j=0}^{s} h^{-2j} \left| \tau \right|_{j, \mathcal{K}}^2 \right) \left( \sum_{j=0}^{s} h^{-2j} \left| \pi_{s-1} \psi \right|_{j, \mathcal{K}}^2 \right).
\end{align}

Hence, from (2.30) and from (2.28) it follows
\begin{align}
|\hat{E}(\pi_{s-1} \delta \pi_{s-1} \psi \tau)|^2 &\leq c h^{2(d+1)} \left( \sum_{j=0}^{s} h^{-2j} \left| \tau \right|_{j, \mathcal{K}}^2 \right) \left( h^{-2(s-1)} \left| \psi \right|_{s-1, \mathcal{K}}^2 + \sum_{j=0}^{s-1} h^{-2j} \left| \psi \right|_{j, \mathcal{K}}^2 \right) \left( \sum_{j=0}^{s} h^{-2j} \left| \pi_{s-1} \psi \right|_{j, \mathcal{K}}^2 \right).
\end{align}

Substituting from (2.25), (2.26), (2.27) and from (2.31) into (2.24) we get (2.22). From (2.22), it follows
\begin{align}
\left| \hat{E} \left( \delta \frac{\partial \psi}{\partial x_i} \frac{\partial \tau}{\partial x_j} \right) \right| &\leq c h^{2d+1} \left( h^{2s} \left| \psi \right|_{s+1, \mathcal{K}}^2 + \left| \psi \right|_{s+1, \mathcal{K}}^2 \right) \left| \tau \right|_{s+1, \mathcal{K}}^2 \\
+ h^{2(d+1)} \left( \sum_{i=0}^{s} h^{-2i} \left| \tau \right|_{i+1, \mathcal{K}}^2 \right) \left( h^{-2(s-1)} \left| \psi \right|_{s+1, \mathcal{K}}^2 + \sum_{i=0}^{s-1} h^{-2i} \left| \psi \right|_{i+1, \mathcal{K}}^2 \right).
\end{align}

If \( \psi(\hat{x}) \) is a polynomial of degree \( \leq r (r \leq s) \) then
\begin{align}
\left| \psi \right|_{s+1, \mathcal{K}}^2 \leq c \left| \psi \right|_{0, \mathcal{K}}^2 \quad \text{and} \quad \left| \psi \right|_{s+1, \mathcal{K}}^2 = 0.
\end{align}
Evidently

\[
\sum_{i=0}^r h^{-2i} |\tau|^2_{i+1, K} = \sum_{i=1}^{r+1} h^{-2(i-1)} |\tau|^2_{i, K} = h^2 \sum_{i=1}^r h^{-2i} |\tau|^2_{i, K},
\]  
\(2.34\)

\[
\sum_{i=0}^{s-1} h^{-2i} |\psi|^2_{i+1, K} = \sum_{i=1}^s h^{-2(i-1)} |\psi|^2_{i, K} = h^2 \sum_{i=1}^r h^{-2i} |\psi|^2_{i, K}.
\]  
\(2.35\)

Substituting from (2.33)-(2.35) into (2.32) we get (2.23).

Now we can formulate two theorems concerning isoparametric integration.

**Theorem 2.1:** Let \(\mathcal{C}_h\) be a \(k\)-regular triangulation of the set \(\Omega\) where

\[ k > \frac{n}{2} - 1. \]  
\(2.36\)

Let \(v \in V_h(\Omega_h)\) and \(\varphi \in H^m(\Omega_h)\), where

\[ m = \max \left( \left\lfloor \frac{n}{2} \right\rfloor + 1, k \right). \]  
\(2.37\)

Let the quadrature formula given on the reference set \(\tilde{K}\) be of degree

\[ d \geq \max(1, 2k-2). \]  
\(2.38\)

Then there exists a constant \(c\) such that

\[ |E(\varphi, v)| \leq c h^k \| \varphi \|_{m, \Omega_h} \| v \|_{1, \Omega_h}. \]  
\(2.39\)

If, in addition, \(\varphi \in H^{k+1}(\Omega_h)\) then there exists a constant \(c\) such that

\[ |E(\varphi, v)| \leq c h^{k+1} \| \varphi \|_{k+1, \Omega_h} \left( \sum_{K \in \mathcal{S}_h} \| v \|_{2, K}^2 \right)^{1/2}. \]  
\(2.40\)

**Proof:** Obviously

\[ E(\varphi, v) = \sum_{K \in \mathcal{S}_h} E_K(\varphi, v) = \sum_{K \in \mathcal{S}_h} \tilde{E}(J_K \varphi^* v^*). \]  
\(2.41\)

It is easy to verify that \(m > n/2, k \leq m \leq k+1, \max(1, 2k-2) \geq k\) and that

\[ D^s(J_K) = O(h^{s+1}). \]  
Hence, we may apply Lemma 2.6 for \(\psi = \varphi^*, s = m, \tau = v^*, r = k, \delta = J_K, \mathcal{X} = n\) and \(d \geq \max(1, 2k-2)\). From (2.22) we get

\[ |\tilde{E}(J_K \varphi^* v^*)|^2 \leq c h^{2n} \left\{ h^{2m} \| \varphi^* \|_{m, \mathcal{K}}^2 + \| \varphi^* \|_{m, \mathcal{K}}^2 \right\} \| v^* \|_{0, \mathcal{K}}^2 \]

\[ + h^2 (d+1) \left( \sum_{i=0}^k h^{-2i} |v^*|^2_{i, K} \right) \left( h^{-2(m-1)} \| \varphi^* \|_{m, \mathcal{K}}^2 + \sum_{i=0}^{m-1} h^{-2i} |\varphi^*|^2_{i, K} \right). \]
Hence from Lemma 2.4 (notice that a $k$-regular family is a $k'$-regular family for any $k' \leq k$) it follows

$$
|\tilde{E}(J_K \varphi^* v^*)|^2 \leq c \left\{ h^{2m} \left\| v \right\|_{0,K}^2 \left\| \varphi \right\|^2_{m,K} + h^{2(d+1)} \left\| \varphi \right\|^2_{m,K} h^n \sum_{i=0}^k h^{-2i} \left\| v^* \right\|_{i,K}^2 \right\}. \quad (2.42)
$$

In the same manner we may apply the inequality (2.22) for $s=k+1$ assuming $\varphi \in H^{k+1}(\Omega_h)$. Then we obtain

$$
|\tilde{E}(J_K \varphi^* v^*)|^2 \leq c \left\{ h^{2(k+1)} \left\| v \right\|_{0,K}^2 \left\| \varphi \right\|^2_{k+1,K} + h^{2(d+1)} \left\| \varphi \right\|^2_{k+1,K} h^n \sum_{i=0}^k h^{-2i} \left\| v^* \right\|_{i,K}^2 \right\}. \quad (2.43)
$$

From lemma 2.2 and from lemma 2.4 we get for $k \geq 1$:

$$
\sum_{i=0}^k h^{-2i} \left\| v^* \right\|_{i,K}^2 = \left\| v^* \right\|_{0,K}^2 + \sum_{i=1}^k h^{-2i} \left\| v^* \right\|_{i,K}^2 \leq \left\| v^* \right\|_{0,K}^2 + c \left\| v^* \right\|^2_{1,K} h^{-2k}
\leq ch^{-n} \left( \left\| v \right\|_{0,K}^2 + h^{-2k+2} \left\| v \right\|_{1,K}^2 \right)
$$

Hence

$$
\sum_{i=0}^k h^{-2i} \left\| v^* \right\|_{i,K}^2 \leq ch^{-n-2k+2} \left\| v \right\|_{0,K}^2 \quad \text{for } k \geq 1. \quad (2.44)
$$

Similarly

$$
\sum_{i=0}^k h^{-2i} \left\| v^* \right\|_{i,K}^2 \leq ch^{-n-2k+4} \left\| v \right\|^2_{2,K} \quad \text{for } k \geq 2. \quad (2.45)
$$

Substituting from (2.44) into (2.42) and observing that $m \geq k$ and $d \geq 2k-2$, we obtain

$$
|\tilde{E}(J_K \varphi^* v^*)|^2 \leq ch^{2k} \left\| \varphi \right\|_{m,K}^2 \left\| v \right\|_{1,K}^2 \quad \text{for } k \geq 1. \quad (2.46)
$$

Substituting from (2.45) into (2.43) and observing that $d \geq 2k-2$, we get

$$
|\tilde{E}(J_K \varphi^* v^*)|^2 \leq ch^{2(k+1)} \left\| \varphi \right\|_{k+1,K}^2 \left\| v \right\|^2_{2,K} \quad \text{for } k \geq 2. \quad (2.47)
$$

Substituting from (2.44) into (2.43) and observing that $d \geq 1$, we obtain

$$
|\tilde{E}(J_K \varphi^* v^*)|^2 \leq ch^{4} \left\| \varphi \right\|^2_{2,K} \left\| v \right\|_{1,K}^2 \quad \text{for } k = 1. \quad (2.48)
$$
From (2.47) and (2.48) we see that
\[
|\mathcal{E}(J_{JK} \Phi^* v^*)|^2 \leq ch^2 (k+1) \|\Phi\|_{k+1,K}^2 \|v\|_{2,K}^2 \quad \text{for} \ k \geq 1. \tag{2.49}
\]
From (2.46), (2.49), (2.41) and from the Schwarz inequality the inequalities (2.39) and (2.40) follow.

**Theorem 2.2:** Let $\mathcal{T}_h$ be a $k$-regular triangulation of the set $\Omega$, where $k > n/2 - 1$. Let $\Phi \in V_h(\Omega)$, $v \in V_h(\Omega)$, $b \in C^{k+1}(\Omega_h)$ and $\Phi \in H^{k+1}(\Omega_h)$ be any function such that $\pi_h \Phi \in V_h(\Omega)$. Let the quadrature formula given on the reference set $K$ be of a degree $d \geq \max(1, 2k-2)$.

Then there exists a constant $c$ such that
\[
\left| \mathcal{E}(b \frac{\partial \Phi}{\partial x_i} \frac{\partial v}{\partial x_j}) \right| \leq c \left[ h^k \|\Phi\|_{k+1,K} + \|\Phi - \Phi_{0,\Omega}\|_{2,K} \right] \left( \sum_{K \in \mathcal{T}_h} \|v\|_{2,K}^2 \right)^{1/2}. \tag{2.50}
\]

If, in addition, $b \in C^{k+2}(\Omega_h)$ then there exists a constant $c$ such that
\[
\left| \mathcal{E}(b \frac{\partial \Phi}{\partial x_i} \frac{\partial v}{\partial x_j}) \right| \leq ch \left[ h^k \|\Phi\|_{k+1,K} + \|\Phi - \Phi_{0,\Omega}\|_{2,K} \right] \left( \sum_{K \in \mathcal{T}_h} \|v\|_{2,K}^2 \right)^{1/2}. \tag{2.51}
\]

**Proof:** Obviously
\[
\mathcal{E}(b \frac{\partial \Phi}{\partial x_i} \frac{\partial v}{\partial x_j}) = \sum_{k \in \mathcal{T}_h} \mathcal{E}(J_{JK} b^* \left( \frac{\partial \Phi}{\partial x_i} \right)^* \left( \frac{\partial \Phi}{\partial x_j} \right)^*).
\]

From the rule on differentiation of the composite function it follows
\[
\mathcal{E}(J_{JK} b^* \left( \frac{\partial \Phi}{\partial x_i} \right)^* \left( \frac{\partial \Phi}{\partial x_j} \right)^*) = \sum_{r=1}^n \mathcal{E} \left( \gamma_{i,j} b^* \frac{\partial \Phi}{\partial x_i} \frac{\partial v}{\partial x_j} \right), \tag{2.53}
\]
where
\[
\gamma_{i,j} = b^* \frac{J_{K}(r_i, r_j)}{J_{K}} \quad (J_{K}(r_i, r_j), J_{K}(r_j, r_i) \text{ are cofactors of } J_{K}).
\]

From lemma 2.3 and from lemma 2.1 we get $D^s(\gamma_{ij}) = O(h^{s+1} \gamma_{ij}^{n-2})$. Hence we may apply lemma 2.6 for $\psi = \Phi^*$, $s = k + q$ (if $b \in C^{k+1}$ or $q = 2$ if $b \in C^{k+2}$), $\tau = v^*$, $r = k$, $\delta = \gamma_{ij}$, $\mathcal{H} = n-2$ and $d \geq \max(1, 2k-2)$. From (2.23) we get
\[
\left| \mathcal{E} \left( \gamma_{ij} \frac{\partial \Phi^*}{\partial x_i} \frac{\partial v^*}{\partial x_j} \right) \right|^2 \leq ch^2 (n-2) \left\{ h^2 (k+q) \|\Phi^*\|^2_{0,k} \|v^*\|^2_{0,k} + h^2 (d+3) \sum_{i=1}^k h^{-2i} \|\Phi^*\|^2_{0,k} \sum_{i=1}^k h^{-2i} \|v^*\|^2_{0,k} \right\}. \tag{2.54}
\]
From lemma 2.2, lemma 2.4 and from the interpolation theorem (see lemma 1.1) it follows

\[ \sum_{i=1}^{k} h^{-2i} |\varphi^*|_{i,K}^2 \]

\[ \leq c \left\{ \sum_{i=1}^{k} h^{-2i} |\varphi^* - (\pi_h \Phi)^*|_{i,K}^2 + \sum_{i=1}^{k} h^{-2i} |(\pi_h \Phi)^* - \Phi^*|_{i,K}^2 + \sum_{i=1}^{k} h^{-2i} |\Phi^*|_{i,K}^2 \right\} \]

\[ \leq ch^{-n} \left\{ h^{-2k} \|\varphi - \pi_h \Phi\|_{0,K}^2 + \sum_{i=1}^{k} \|\pi_h \Phi - \Phi\|_{i,K}^2 \sum_{i=1}^{k} \|\Phi^2\|_{i,K} \right\} \]

\[ \leq ch^{-n} \left\{ h^{-2k} ( \|\varphi - \Phi\|_{0,K}^2 + \|\pi_h \Phi - \Phi\|_{0,K}^2) + \|\pi_h \Phi - \Phi\|_{k+1,K}^2 + \|\Phi^2\|_{k,K} \right\} \]

Hence

\[ \sum_{i=1}^{k} h^{-2i} |\varphi^*|_{i,K}^2 \leq ch^{-n} \left\{ h^{-2k} \|\varphi - \Phi\|_{0,K}^2 + \|\Phi^2\|_{k+1,K} \right\}. \quad (2.55) \]

Evidently

\[ \|\varphi^*\|_{0,K}^2 \leq ch^{-n} \|\varphi\|_{0,K}^2, \quad |v^*|_{1,K}^2 \leq ch^{-n+2} \|v\|_{1,K}^2. \quad (2.56) \]

Substituting from (2.56), (2.55) and from (2.44) into (2.54) for \( q=1 \) and observing that \( d^2k - 2 \), we similarly obtain

\[ \|\varphi^*\|_{0,K}^2 \leq ch^{-n} \|\varphi - \Phi\|_{0,K}^2 + \|\Phi^2\|_{k+1,K} \] for \( k \geq 2 \). \quad (2.58)

Substituting from (2.56), (2.55) and from (2.45) into (2.54) for \( q=2 \) and observing that \( d \geq 2k - 2 \), we similarly obtain

\[ \|\varphi^*\|_{0,K}^2 \leq ch^{-n} \|\varphi - \Phi\|_{0,K}^2 + \|\Phi^2\|_{k+1,K} \] for \( k \geq 2 \). \quad (2.58)
Substituting from (2.56), (2.55) and from (2.44) into (2.54) for \( k = 1, q = 2 \) and observing that \( d \), we get
\[
\left| E \left( \gamma_{ij} \frac{\partial \phi^*}{\partial x_i} \frac{\partial v^*}{\partial x_j} \right) \right|^2 \leq ch^{2n-4} \{ h^6 h^{-n} \| \phi \|_{0,k}^2 h^{-n+2} \| v \|_{1,k}^2 \}
+ h^2 h^{-n} (h^{-2} \| \phi - \Phi \|_{0,k}^2 + \| \Phi \|_{2,k}^2 h^{-n} \| v \|_{1,k}^2) \}
\leq c \{ h^4 (\| \phi - \Phi \|_{0,k}^2 + \| \Phi \|_{2,k}^2) \| v \|_{1,k}^2 \}
+ h^4 (h^{-2} \| \phi - \Phi \|_{0,k}^2 + \| \Phi \|_{2,k}^2) \| v \|_{1,k}^2 \}.
\]
Hence
\[
\left| E \left( \gamma_{ij} \frac{\partial \phi^*}{\partial x_i} \frac{\partial v^*}{\partial x_j} \right) \right|^2 \leq ch^2 \{ \| \phi - \Phi \|_{0,k}^2 + h^2 \| \Phi \|_{2,k}^2 \| v \|_{1,k}^2 \}. \quad (2.59)
\]
From (2.58) and (2.59) we see that
\[
\left| E \left( \gamma_{ij} \frac{\partial \phi^*}{\partial x_i} \frac{\partial v^*}{\partial x_j} \right) \right|^2 \leq ch^2 \{ \| \phi - \Phi \|_{0,k}^2 + h^2 \| \Phi \|_{2,k}^2 \| v \|_{1,k}^2 \} \quad \text{for } k \geq 1. \quad (2.60)
\]
From (2.57), (2.60), (2.53), (2.52) and from the Schwarz inequality the inequalities (2.50) and (2.51) follow.

3. APPROXIMATE SOLUTION OF THE ELLIPTIC PROBLEMS

Let \( \Omega \) be a bounded domain in \( \mathbb{R}^n \) with sufficiently smooth boundary \( \partial \Omega \). We study the elliptic problem
\[
\begin{align*}
- lu &= f(x), & x \in \Omega, \\
u(x) &= 0 & \text{on } \partial \Omega,
\end{align*}
\]
where \( f \) is a sufficiently smooth function and
\[
l = \sum_{i,j=1}^n \frac{\partial}{\partial x_j} \left( g_{ij}(x) \frac{\partial}{\partial x_i} \right). \quad (3.2)
\]
We suppose that the functions \( g_{ij}(x) \) are sufficiently smooth and
\[
g_{ij}(x) = g_{ji}(x). \quad (3.3)
\]
About the differential operator \( l \) we suppose that it is strongly elliptic, i.e. there exists a constant \( g_1 > 0 \) such that
\[
\sum_{i,j=1}^n g_{ij}(x) \xi_i \xi_j \geq g_1 \sum_{i=1}^n \xi_i^2, \quad \forall x \in \overline{\Omega}, \quad (\xi_1, \ldots, \xi_n) \in \mathbb{R}^n. \quad (3.4)
\]
The variational formulation of the elliptic problem is:

Find a function \( u(x) \in H_0^1(\Omega) \) such that

\[
\begin{align*}
\langle a(u, v) \rangle &= \langle f, v \rangle_{0, \Omega}, \forall v \in H_0^1(\Omega),
\end{align*}
\]

where

\[
a(u, v) = \int_{\Omega} \sum_{i, j=1}^{n} g_{ij}(x) \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} dx.
\]

We extend the functions \( g_{ij}(x), f(x) \) to a greater set \( \tilde{\Omega} \supset \Omega \) so that the conditions (3.3) and (3.4) are satisfied (with positive constants \( G_1 \)). In this way we obtain the functions \( G_{ij}(x), F(x) \). We denote

\[
L = \sum_{i, j=1}^{n} \frac{\partial}{\partial x_j} \left( G_{ij}(x) \frac{\partial}{\partial x_i} \right)
\]

Let \( \mathcal{C}_h \) be a \( k \)-regular triangulation of the set \( \Omega \) and Let \( V_h \) be the corresponding finite element space. The union of the elements \( K \) from \( \mathcal{C}_h \) forms a set \( \Omega_h \) which, in general, differs from \( \Omega \). We suppose that

\[
\Omega_h \subset \tilde{\Omega},
\]

for all sufficiently small \( h \) and formulate the following discrete problem

Find a function \( u_d(x) \in V_h \) such that

\[
\begin{align*}
\langle a_h(u_d, v) \rangle &= \langle F, v \rangle_{0, \Omega}, \forall v \in V_h,
\end{align*}
\]

where

\[
a_h(u_d, v) = \int_{\Omega_h} \sum_{i, j=1}^{n} G_{ij}(x) \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j} dx.
\]

Since it is either too costly or simply impossible to evaluate exactly the integrals \( \langle ..., \rangle_{0, \Omega_h}, a_h(\cdot, \cdot) \), we must now take into account the fact that approximate integration is used for their computation. For this purpose we use the isoparametric numerical integration, i.e. in agreement with (2.4) we replace

\[
(\varphi, \psi)_{0, \Omega_h} \approx (\varphi, \psi)_h, \quad a_h(\varphi, \psi) \approx A_h(\varphi, \psi)
\]

where

\[
(\varphi, \psi)_h = \sum_{K \in \mathcal{C}_h} \sum_r \omega_{r,K} \varphi(b_{r,K}) \psi(b_{r,K})
\]
Let us note that form (2.6) and (2.8) it follows

\begin{equation}
(\varphi, \psi)_h - (\varphi, \psi)_h = E(\varphi \psi),
\end{equation}

\begin{equation}
a_h (\varphi, \psi) - A_h (\varphi, \psi) = E \left( \sum_{i, j=1}^{n} G_{ij} (x) \frac{\partial \varphi}{\partial x_i} \frac{\partial \psi}{\partial x_j} \right).
\end{equation}

Evidently \( b_r \in \overline{\Omega} \) for sufficiently small \( h \) (remember that \( b_r \) are supposed to lie inside \( \hat{K} \) or coincide with some of the points \( \hat{a}_i \)). Hence \( F(b_r, K) = f(b_r, K) \), \( G_{ij}(b_r, K) = g_{ij}(b_r, K) \). Therefore from (3.12) we see that \((F, v)_h = (f, v)_h\). In such a way we come to the following fully discrete problem:

Find a function \( u_h(x) \in V_h \) such that

\[ A_h(u_h, v) = (f, v)_h, \quad \forall v \in V_h \]  \hspace{1cm} (3.16)

Let the functions \( \varphi_1, \ldots, \varphi_s \) form the basis of the space \( V_h \). Denoting

\[ \gamma = [\gamma_1, \ldots, \gamma_s]^T, \]  \hspace{1cm} (3.17)

\[ K_h = \{ A_h(\varphi_i, \varphi_j) \}^t_{i,j=1}, \]  \hspace{1cm} (3.18)

\[ F_h = [(f, \varphi_1)_h, \ldots, (f, \varphi_s)_h]^T \]  \hspace{1cm} (3.19)

the system (3.16) can be written in the form

\[ K_h \gamma = F_h. \]  \hspace{1cm} (3.20)

**4. APPROXIMATE SOLUTION OF THE PARABOLIC PROBLEMS**

We study the parabolic problem

\[ g(x) \frac{\partial w}{\partial t} - lw = f(x, t) \quad \text{for} \quad x \in \Omega \quad \text{and} \quad t \in (0, T), \]  \hspace{1cm} (4.1)

\[ w(x, t) = 0 \quad \text{for} \quad x \in \partial \Omega \quad \text{and} \quad t \in (0, T), \]  \hspace{1cm} (4.1)

\[ w(x, 0) = w_0(x) \in L^2(\Omega), \]

where \( g(x) \) and \( f(x, t) \) are sufficiently smooth functions,

\[ g(x) \geq g_0 (\text{= Const.}) > 0 \]  \hspace{1cm} (4.2)
and the differential operator $l$ defined by (3.2) satisfies the conditions (3.3) and (3.4) with sufficiently smooth functions $g_{ij}(x)$. Similarly as in the elliptic case we come to the variational formulation of the parabolic problem (see [8]):

Find a function $w(x, t)$ such that

$$
\begin{align*}
    w & \in L^\infty (H_0^1(\Omega)), \\
    \frac{\partial w}{\partial t} & \in L^\infty (H^{-1}(\Omega)), \\
    \left( g \frac{\partial w}{\partial t}, v \right)_{0, \Omega} + a(w, v) &= (f, v)_0, \\
    \forall v \in H_0^1(\Omega) \text{ and } t \in (0, T), \\
    w(x, 0) &= w_0(x) \in L^2(\Omega).
\end{align*}
$$

(4.3)

where the bilinear form $a(., .)$ is given by (3.6).

Let us denote by $G(x)$ a sufficiently smooth extension of the function $g(x)$ to a greater set $\Omega$ satisfying (4.2) (with some positive constant $G_0$). First, in the same way as in the elliptic case, we discretize this problem for every $t \in (0, T)$ by the finite element method with respect to $x$. Then we use isoparametric numerical integration. In such a way we come to the following fully semidiscrete problem:

Find a function $w_s(x, t)$ such that

$$
\begin{align*}
    w_s, \frac{\partial w_s}{\partial t} & \in V_h, \forall t \in (0, T), \\
    \left( g \frac{\partial w_s}{\partial t}, v \right)_h + a_h(w_s, v) &= (f, v)_h, \\
    \forall v \in V_h \text{ and } t \in (0, T), \\
    w_s(x, 0) &= w_{s0}(x),
\end{align*}
$$

(4.4)

where $w_{s0}(x)$ is an approximation of $w_0(x)$.

Replacing $v$ in (4.4) by the basic functions $\varphi_i$ we come, to the conclusion that the problem (4.4) is represented by the system of ordinary differential equations with an unknown vector function of paramètre $t$:

$$
M_h \gamma'(t) + K_h \gamma(t) = F_h.
$$

(4.5)

where

$$
\gamma(t) = [\gamma_1(t), \ldots, \gamma_s(t)]^T,
$$

(4.6)

$$
M_h = \{ (g \varphi_i, \varphi_j)_h \}_{i, j=1}^s.
$$

(4.7)
This suggests the way how to discretize the problem (4.4) with respect to \( t \). We solve the mentioned system of ordinary differential equations by \( v \)-step \( A_0 \)-stable method of order \( q \). We divide the time interval \((0, T)\) into a finite number of equal parts \( \Delta t \). We introduce the notation

\[
\Phi^m = \Phi^m(x) = \Phi(x, m \Delta t), \quad m = 0, 1, \ldots
\]

(4.10)

for any function \( \Phi(x, t) \).

According to (4.4) and to the described way of the time discretization we define the following fully discrete problem

Find a function \( w_h(x, t) \) such that

\[
\begin{aligned}
    w_h & \in V_h \quad \text{for} \quad t = \Delta t, 2 \Delta t, \ldots, T, \\
    \left( g \sum_{j=0}^v \alpha_j w^{m+j}_h, v \right)_h + \Delta t A_h \left( \sum_{j=0}^v \beta_j w^{m+j}_h, v \right)_h \\
    &= \Delta t \left( \sum_{j=0}^v \beta_j f^{m+j}_h, v \right)_h, \quad \forall v \in V_h \quad \text{and} \quad m = 0, 1, \ldots
\end{aligned}
\]

(4.11)

From (4.5) we can see that the system in (4.11) is represented by the linear system of algebraic equations

\[
\sum_{j=0}^v (\alpha_j M_h + \Delta t \beta_j K_h) \gamma^{m+j} = \Delta t \sum_{j=0}^v \beta_j F^{m+j}_h
\]

(4.12)

i.e. by the system

\[
(\alpha_v M_h + \Delta t \beta_v K_h) \gamma^{m+v} = \Delta t \beta_v F^{m+v}_h \\
+ \sum_{j=0}^{v-1} [\Delta t \beta_j (F^{m+j}_h - K_h \gamma^{m+j}) - \alpha_j M_h \gamma^{m+j}].
\]

(4.13)

5. RITZ APPROXIMATIONS

Let \( U \) be a function from \( H^1(\Omega) \). The function \( \eta \in V_h(\Omega_h) \) such that

\[
a_h(\eta, v) = -(LU, v)_{0, \Omega}, \quad \forall v \in V_h
\]

(5.1)
is called the Ritz approximation of the function $U$. The function $\eta_d \in V_h(\Omega)$ such that

$$A_h(\eta_d, v) = -(LU, v)_h, \quad \forall v \in V_h$$

is called the Ritz discrete approximation of the function $U$.

From the Green theorem it follows

$$a_h(\eta, v) = a_h(U, v), \quad \forall v \in V_h,$$

i.e. the function $\eta$ is an orthogonal projection onto $V_h$ of the function $U$ in the energy norm given by the bilinear form $a_h(\cdot, \cdot)$. This is the reason why we use the name Ritz approximation. From the proof of theorem 1 in [7] the following theorem follows:

**Theorem 5.1 (theorem on the Ritz approximation):** Let $\mathcal{C}_h$ be a $k$-regular triangulation of the set $\Omega$, $k > n/2 - 1$ and let

$$\Omega_h \subseteq \Omega \quad \text{for all } h.$$  \hspace{1cm} (5.4)

Let $U \in H^{k+1}(\bar{\Omega})$ be any function such that

$$U = 0 \text{ on } \partial \Omega$$

and let $\eta$ be the Ritz approximation of the function $U$.

Then there exists a constant $c$ (independent of $h$) such that

$$|u - \eta|_{1, \Omega_h} \leq ch^k \|U\|_{k+1, \Omega_h}.$$  \hspace{1cm} (5.6)

If, in addition, $U \in H^{k+2}(\bar{\Omega})$, then there exists a constant $c$ such that

$$\|U - \eta\|_{0, \Omega_h} \leq ch^{k+1} \|U\|_{k+2, \Omega}.$$  \hspace{1cm} (5.7)

**Remark:** From (5.6) and (5.7) it follows immediately

$$\|U - \eta\|_{1, \Omega_h} \leq ch^k \|U\|_{k+2, \Omega}$$

provided $U \in H^{k+2}(\bar{\Omega})$.

We are going to derive the similar theorem for the Ritz discrete approximation. Before, we formulate two lemmas.

**Lemma 5.1.** Let $\mathcal{C}_h$ be a $k$-regular triangulation of $\Omega(k > (n/2) - 1)$. Let $v \in H^1(\bar{\Omega})$ and

$$v(y', y_h) = 0 \text{ on } \partial \Omega_h$$

\hspace{1cm} (5.9)
(for notation see figure). Then there exists a constant $c$ such that
\[ \| v \|_{0, \Omega} - \Omega \leq c h^{k+1} \| v \|_{1, \Omega}. \] (5.10)

The proof follows from [7] (see lemma 1 and note 1).

We introduce the notation
\[ \| v \|_h^2 = (g(x)v, v)_h, \quad \| v \|_h^2 = A_h(v, v), \] (5.11)
where the forms $(\cdot, \cdot)_h$, $A_h(\cdot, \cdot)$ are defined in (3.12) and (3.13).

**Lemma 5.2.** Let $\mathcal{G}_h$ be a $k$-regular triangulation of $\Omega (k > (n/2) - 1)$. Then there exist positive constants $c_1$ and $c_2$ such that:

(a) $c_1 \| v \|_{0, \Omega} \leq \| v \|_h, \quad \forall v \in V_h$ (5.12)

provided the quadrature formula on the reference set $\hat{\mathcal{K}}$ is of a degree $d \geq 2k$,

(b) $c_2 \| v \|_{1, \Omega} \leq \| v \|_h, \quad \forall v \in V_h$ (5.13)

provided the quadrature formula on the reference set $\hat{\mathcal{K}}$ is of a degree $d \geq 2k - 2$.

For the proof see [7] (Theorem 5).

**Theorem 5.2** (Theorem on the Ritz discrete approximation): Let $\mathcal{G}_h$ be a $k$-regular triangulation of the set $\Omega$, $k > (n/2) - 1, (5.4)$ be satisfied and $h < 1$. Let $U \in H^{m+2}(\Omega)$, where $m = \max \{(n/2) + 1, k\}$ be any function such that $U = 0$ on $\partial \Omega$ and the quadrature formula given on the reference set $\hat{\mathcal{K}}$ be of a degree $d \geq \max(1, 2k - 2)$. Let $\eta_d$ be the Ritz discrete approximation of the function $U$. Then there exists a constant $c$ such that
\[ \| U - \eta_d \|_{1, \Omega} \leq c h^{k} \| U \|_{m+2, \Omega}. \] (5.14)

If, in addition, $U \in H^{k+3}(\Omega)$ then there exist constants $c_1, c_2$ such that
\[ \| U - \eta_d \|_{0, \Omega} \leq c h^{k+1} \| U \|_{k+3, \Omega}, \] (5.15)
\[ \| U - \eta_d \|_h \leq c h^{k+1} \| U \|_{k+3, \Omega}. \] (5.16)

**Proof:** Evidently
\[ \| U - \eta_d \|_{1, \Omega} \leq \| U - \eta \|_{1, \Omega} + \| \eta - \eta_d \|_{1, \Omega}. \] (5.17)

where $\eta$ is the Ritz approximation of the function $U$.

From (5.13) (lemma 5.2 may be applied since $\eta - \eta_d \in V_h$), from (3.15), (5.1), (5.2) and from (3.14) it follows
\[ \| \eta - \eta_d \|_{1, \Omega} \leq c A_h(\eta - \eta_d, \eta - \eta_d) = c \left\{ A_h(\eta, \eta) - A_h(\eta_d, \eta - \eta_d) \right\} \]
\[ = c \left\{ a_h(\eta) - E \left( \sum_{i,j=1}^{n} G_{ij} \frac{\partial \eta}{\partial x_i} \frac{\partial (\eta - \eta_d)}{\partial x_j} \right) - A_h(\eta_d, \eta - \eta_d) \right\} \]
\[ c \left\{ - (LU, \eta - \eta_d) - E \sum_{i,j=1}^{n} G_{ij} \frac{\partial \eta}{\partial x_i} \frac{\partial (\eta - \eta_d)}{\partial x_j} \right\} \]

Hence, using the inequality (2.39) for \( \varphi = LU, v = \eta - \eta_d \) and the inequality (2.50) for \( \varphi = \eta, v = \eta - \eta_d \) and \( \Phi = U \) (notice that \( \pi_h U \in V_h \)) we get

\[ |\eta - \eta_d|_{1,h}^2 \leq c \left\{ h^k \|LU\|_{m+1,h} + h^k \|U\|_{k+1,h} + \|\eta - U\|_{0,h} \right\} \|\eta - \eta_d\|_{1,h}. \quad (5.18) \]

We notice at this point that because of the assumption (5.4) there exists a constant \( c \) independent of \( h \) such that

\[ \|v\|_{1,h} \leq c \|v\|_{1,h}, \quad \forall v \in V_h \quad (5.19) \]

(see Ciarlet and Raviart [3], p. 455).

Therefore from (5.18) and from the theorem on the Ritz approximation [see (5.7)] it follows

\[ |\eta - \eta_d|_{1,h}^2 \leq c h^k \|U\|_{m+2,h} \|\eta - \eta_d\|_{1,h}. \]

Hence

\[ |\eta - \eta_d|_{1,h} \leq c h^k \|U\|_{m+2,h}. \quad (5.20) \]

From (5.17), (5.6) and (5.20) we get

\[ |U - \eta_d|_{1,h} \leq c h^k \|U\|_{m+2,h}. \quad (5.21) \]

Evidently

\[ \|U - \eta_d\|_{0,h} \leq \|U - \eta\|_{0,h} + \|\eta - \eta_d\|_{0,h}. \]

Therefore from the theorem on the Ritz approximation, from (5.19) and from (5.20) it follows

\[ |U - \eta_d|_{0,h} \leq c h^k \|U\|_{m+2,h}. \quad (5.22) \]

The inequalities (5.21) and (5.22) imply (5.14).

We prove now the inequality (5.15). We give the proof for \( n \leq 3 \); the proof for \( n > 3 \) can be achieved by using a smoothing procedure, following an idea of Strang [9].

R.A.I.R.O. Analyse numérique/Numerical Analysis
Let us denote
\[ z = \begin{cases} \eta - \eta_d & \text{for } x \in \Omega_h, \\ 0 & \text{for } x \in \Omega - \Omega_h. \end{cases} \] (5.23)

Let \( y \) be the solution of the homogeneous Dirichlet problem
\[ -ly = z \quad \text{in } \Omega, \quad y = 0 \quad \text{on } \partial \Omega. \] (5.24)

If \( \partial \Omega \) is smooth enough then \( y \in H^1_0(\Omega) \cap H^2(\Omega) \) and
\[ \| y \|_{2,\Omega} \leq c \| z \|_{0,\Omega} \leq c \| z \|_{0,\Omega} = c \| z \|_{0,\Omega}. \]
i. e. :
\[ \| y \|_{2,\Omega} \leq c \| \eta - \eta_d \|_{0,\Omega}. \] (5.25)

Using the Calderon theorem we extend the function \( y \) from \( \Omega \) onto \( \tilde{\Omega} \). In this way we obtain a function \( \tilde{y} \in H^2(\tilde{\Omega}) \) such that \( \| \tilde{y} \|_{2,\Omega} \leq c \| y \|_{2,\Omega} \). Therefore from (5.25) it follows
\[ \| \tilde{y} \|_{2,\Omega} \leq c \| \eta - \eta_d \|_{0,\Omega}. \] (5.26)

Using simple calculation we get
\[ \| \eta - \eta_d \|_{0,\Omega}^2 = \int_{\Omega_{\eta} - \Omega} (\eta - \eta_d)(z + L\tilde{y}) \, dx - \int_{\Omega_h} (\eta - \eta_d) L\tilde{y} \, dx. \]

The Green theorem \((\eta - \eta_d = 0 \text{ on } \partial\Omega_h)\) yields
\[ - \int_{\Omega_h} (\eta - \eta_d) L\tilde{y} \, dx = a_h(\eta - \eta_d, \tilde{y}). \]

Hence
\[ \| \eta - \eta_d \|_{0,\Omega}^2 \leq \left| \int_{\Omega_{\eta} - \Omega} (\eta - \eta_d)(z + L\tilde{y}) \, dx \right| + \left| a_h(\eta - \eta_d, \tilde{y}) \right|. \] (5.27)

The Schwarz inequality gives
\[ \left| \int_{\Omega_{\eta} - \Omega} (\eta - \eta_d)(z + L\tilde{y}) \, dx \right| \leq \| \eta - \eta_d \|_{0,\Omega_{\eta} - \Omega} \| z + L\tilde{y} \|_{0,\Omega_{\eta} - \Omega}. \] (5.28)

Using (5.26) we get
\[ \| z + L\tilde{y} \|_{0,\Omega_{\eta} - \Omega} \leq \| z \|_{0,\Omega_{\eta}} + \| L\tilde{y} \|_{0,\Omega_{\eta}} \]
\[ \leq \| \eta - \eta_d \|_{0,\Omega_{\eta}} + c \| \tilde{y} \|_{2,\Omega_{\eta}} \leq c \| \eta - \eta_d \|_{0,\Omega_{\eta}}. \] (5.29)

vol. 13, n° 3, 1979
From (5.10) and (5.20) it follows
\[ \| \eta - \eta_d \|_{0, \Omega_0} \leq c h^{k+1} \| \eta - \eta_d \|_{1, \Omega_0} \leq c h^{2k+1} \| U \|_{m+2, \Omega}. \]

Therefore from (5.29) and (5.28) we get
\[ \left| \int_{\Omega_0} (\eta - \eta_d) (z + L \tilde{y}) \, dx \right| \leq c h^{k+1} \| U \|_{m+2, \Omega} \| \eta - \eta_d \|_{0, \Omega_0}. \quad (5.30) \]

Evidently
\[ a_h(\eta - \eta_d, \tilde{y}) = a_h(\eta - U, \tilde{y} - \pi_h \tilde{y}) + a_h(\eta - U, \pi_h \tilde{y}) \]
\[ + a_h(U - \eta_d, \tilde{y} - \pi_h \tilde{y}) + a_h(U - \eta_d, \pi_h \tilde{y}). \]

From (5.3) (we know that \( \pi_h \tilde{y} \in V_h \)) it follows that \( a_h(\eta - U, \pi_h \tilde{y}) = 0 \). Hence
\[ |a_h(\eta - \eta_d, \tilde{y})| \leq \left| a_h(\eta - U, \tilde{y} - \pi_h \tilde{y}) \right| \]
\[ + |a_h(U - \eta_d, \tilde{y} - \pi_h \tilde{y})| + |a_h(U - \eta_d, \pi_h \tilde{y})|. \quad (5.31) \]

From the Schwarz inequality, from (5.6) and from the interpolation theorem (see Lemma 1) we get
\[ |a_h(\eta - U, \tilde{y} - \pi_h \tilde{y})| \leq c |\eta - U|_{1, \Omega_0} |\tilde{y} - \pi_h \tilde{y}|_{1, \Omega_0} \leq c h^{k} \| U \|_{k+1, \Omega_0} \| \tilde{y} \|_{2, \Omega_0}. \]

This and (5.26) imply
\[ |a_h(\eta - U, \tilde{y} - \pi_h \tilde{y})| \leq c h^{k+1} \| U \|_{k+1, \Omega_0} \| \eta - \eta_d \|_{0, \Omega_0}. \quad (5.32) \]

Similarly, using (5.21), we get
\[ |a_h(U - \eta_d, \tilde{y} - \pi_h \tilde{y})| \leq c h^{k+1} \| U \|_{m+2, \Omega} \| \eta - \eta_d \|_{0, \Omega_0}. \quad (5.33) \]

From the Green theorem, from (3.15), (5.2) and from (3.14) we get
\[ a_h(U - \eta_d, \pi_h \tilde{y}) = a_h(U, \pi_h \tilde{y}) - a_h(\eta_d, \pi_h \tilde{y}) \]
\[ = -(LU, \pi_h \tilde{y})_{0, \Omega_0} - A_h(\eta_d, \pi_h \tilde{y}) - E \left( \sum_{i, j=1}^{n} G_{ij} \frac{\partial \eta_d}{\partial x_i} \frac{\partial \pi_h \tilde{y}}{\partial x_j} \right) \]
\[ = -(LU, \pi_h \tilde{y})_{0, \Omega_0} + (LU, \pi_h \tilde{y})_h - E \left( \sum_{i, j=1}^{n} G_{ij} \frac{\partial \eta_d}{\partial x_i} \frac{\partial \pi_h \tilde{y}}{\partial x_j} \right) \]
\[ = - E(LU, \pi_h \tilde{y}) - E \left( \sum_{i, j=1}^{n} G_{ij} \frac{\partial \eta_d}{\partial x_i} \frac{\partial \pi_h \tilde{y}}{\partial x_j} \right). \]

Hence
\[ |a_h(U - \eta_d, \pi_h \tilde{y})| \leq |E(LU, \pi_h \tilde{y})| + \left| E \left( \sum_{i, j=1}^{n} G_{ij} \frac{\partial \eta_d}{\partial x_i} \frac{\partial \pi_h \tilde{y}}{\partial x_j} \right) \right|. \quad (5.34) \]
Evidently for $K \in \mathcal{S}_h \left\| \pi_h \tilde{y} \right\|_{2, \kappa} \leq \left\| \pi_h \tilde{y} - \tilde{y} \right\|_{2, \kappa} + \left\| \tilde{y} \right\|_{2, \kappa}$. Hence the interpolation theorem implies $\left\| \pi_h \tilde{y} \right\|_{2, \kappa} \leq c \left\| \tilde{y} \right\|_{2, \kappa}$ and from (5.26) we get

$$\left( \sum_{K \in \mathcal{T}_h} \left\| \pi_h \tilde{y} \right\|_{2, \kappa}^2 \right)^{1/2} \leq c \left\| \eta - \eta_d \right\|_{0, \Omega}.$$  \hspace{1cm} (5.35)

Therefore from (2.40) (we apply theorem 2.1 for $\varphi = LU$ and $v = \pi_h \tilde{y}$) it follows

$$\left| E(LU \pi_h \tilde{y}) \right| \leq c h^{k+1} \left\| LU \right\|_{k+1, \Omega} \left\| \eta - \eta_d \right\|_{0, \Omega}.$$ Hence

$$\left| E(LU \pi_h \tilde{y}) \right| \leq c h^{k+1} \left\| U \right\|_{k+3, \Omega} \left\| \eta - \eta_d \right\|_{0, \Omega}.$$ \hspace{1cm} (5.36)

From (2.51) (we apply theorem 2.2 for $b = G_{ij}$, $\varphi = \eta_d$, $\Phi = U$, $v = \pi_h \tilde{y}$) and from (5.35) it follows

$$\left| E \left( \sum_{i, j=1}^n G_{ij} \frac{\partial \eta_d}{\partial x_i} \frac{\partial \pi_h \tilde{y}}{\partial x_j} \right) \right| \leq c h^k \left( \left\| U \right\|_{k+1, \Omega} + \left\| \eta_d - U \right\|_{0, \Omega} \right) \left\| \eta - \eta_d \right\|_{0, \Omega}.$$ Hence, from (5.22) observing that $m \leq k + 1$ we get

$$\left| E \left( \sum_{i, j=1}^n G_{ij} \frac{\partial \eta_d}{\partial x_i} \frac{\partial \pi_h \tilde{y}}{\partial x_j} \right) \right| \leq c h^{k+1} \left\| U \right\|_{k+3, \Omega} \left\| \eta - \eta_d \right\|_{0, \Omega}.$$ \hspace{1cm} (5.37)

From (5.34), (5.36) and (5.37) it follows

$$\left| a_h(U - \eta_d, \pi_h \tilde{y}) \right| \leq c h^{k+1} \left\| U \right\|_{k+3, \Omega} \left\| \eta - \eta_d \right\|_{0, \Omega}.$$ \hspace{1cm} (5.38)

Substituting from (5.32), (5.33) and (5.38) into (5.31) we get

$$\left| a_h(\eta - \eta_d, \tilde{y}) \right| \leq c h^{k+1} \left\| U \right\|_{k+3, \Omega} \left\| \eta - \eta_d \right\|_{0, \Omega}.$$ \hspace{1cm} (5.39)

From (5.27), (5.30) and (5.39) it follows

$$\left\| \eta - \eta_d \right\|_{0, \Omega}^2 \leq c h^{k+1} \left\| U \right\|_{k+3, \Omega} \left\| \eta - \eta_d \right\|_{0, \Omega}.$$ Hence

$$\left\| \eta - \eta_d \right\|_{0, \Omega} \leq c h^{k+1} \left\| U \right\|_{k+3, \Omega}.$$ \hspace{1cm} (5.40)

Therefore, from the trivial inequality

$$\left\| U - \eta_d \right\|_{0, \Omega} \leq \left\| U - \eta \right\|_{0, \Omega} + \left\| \eta - \eta_d \right\|_{0, \Omega}$$

and from the Theorem on the Ritz approximation the inequality (5.15) follows.
From (5.11), (3.12) and from (1.2) it follows
\[ \| U - \eta_d \|_h^2 = \sum_{K \in \mathcal{T}_h} \sum_r \partial_r J_K (\hat{b}_r) g^* (\hat{b}_r) [U^* (\hat{b}_r) - \eta_d^* (\hat{b}_r)]^2 \leq ch^* \sum_{K \in \mathcal{T}_h} \sum_r [U^* (\hat{b}_r) - \eta_d^* (\hat{b}_r)]^2. \]

Hence
\[ \| U - \eta_d \|_h^2 \leq ch^* \sum_{K \in \mathcal{T}_h} [\max (U^* - \eta_d^*)]^2 \tag{5.41} \]

Evidently
\[ \max_{K} |U^* - \eta_d^*| \leq \max_{K} |U^* - \pi_k U^*| + \max_{K} |\pi_k U^* - \eta_d^*|. \tag{5.42} \]

From the Bramble-Hilbert lemma and from lemma 2.4 [see (2.17)] it follows
\[ \max_{K} |U^* - \pi_k U^*| \leq c |U^*|_{k+1, K} \leq ch^{-(n/2) + k + 1} \| U \|_{k+1, K}. \tag{5.43} \]

From lemma 2.2 [see (2.10)], from lemma 2.5 [see (2.17)], from lemma 2.4 and from the evident inequality \( \| U^* - \eta_d^* \|_{0, K} \leq \| U - \eta_d \|_{0, K} \) we get
\[ \max_{K} |\pi_k U^* - \eta_d^*| \leq c \| \pi_k U^* - \eta_d^* \|_{0, K} \]
\[ \leq c [ \| \pi_k U^* - U^* \|_{0, K} + \| U^* - \eta_d^* \|_{0, K}] \]
\[ \leq c [ \| U^* \|_{k+1, K} + \| U^* - \eta_d^* \|_{0, K}] \]
\[ \leq ch^{-n/2} [h^{k+1} \| U \|_{k+1, K} + \| U - \eta_d \|_{0, K}]. \]

Hence, from (5.43) and (5.42) it follows
\[ \max_{K} |U^* - \eta_d^*| \leq ch^{-n/2} [h^{k+1} \| U \|_{k+1, K} + \| U - \eta_d \|_{0, K}]. \]

Therefore, (5.41) implies
\[ \| U - \eta_d \|_h^2 \leq c \sum_{K \in \mathcal{T}_h} [h^{2(k+1)} \| U \|_{k+1, K} + \| U - \eta_d \|_{0, K}^2] \]
\[ = c [h^{2(k+1)} \| U \|_{k+1, \Omega_h}^2 + \| U - \eta_d \|_{0, \Omega_h}^2]. \]

This and (5.15) imply (5.16).

6. ERROR ESTIMATE FOR ELLIPTIC PROBLEMS

Theorem 6.1: Let \( u \) be the solution of the elliptic problem (3.1) with sufficiently smooth functions \( f, g_{ij} \) satisfying the conditions (3.3) and (3.4). Let \( \mathcal{G}_h \) be a
k-regular \((k > (n/2)-1)\) triangulation of the set \(\Omega\) with a sufficiently smooth boundary \(\partial \Omega\). Let the quadrature formula on the reference set \(K\) be of a degree \(d \geq \max(1, 2k - 2)\). Then the fully discrete problem (3.16) has a unique solution \(u_h(x)\) and there exists a constant \(c\) independent of \(h\) and \(u\) such that

\[
\|u - u_h\|_{1, \Omega \cap \Omega_h} \leq c h^k \|u\|_{m+2, \Omega},
\]

(6.1)

\[
\|u - u_h\|_{0, \Omega \cap \Omega_h} \leq c h^{k+1} \|u\|_{k+3, \Omega},
\]

(6.2)

where \(m = \max([n/2]+1, k)\).

Proof: We know that the problem (3.16) is represented by the system (3.20). Hence, to prove the existence and the uniqueness it suffices to show that the matrix \(K_h\) defined by (3.18) is positive definite, i.e. that

\[
m^T K_h m > 0
\]

(6.3)

for any nonzero vector \(m = [m_1, \ldots, m_s]^T\).

From (3.18), (5.11) and (5.13) we may apply lemma 5.2 since \(\sum_{i=1}^{s} m_i \varphi_i \in V_h\), it follows

\[
m^T K_h m = A_h \left( \sum_{i=1}^{s} m_i \varphi_i, \sum_{i=1}^{s} m_i \varphi_i \right) = \left| \sum_{i=1}^{s} m_i \varphi_i \right|^2 \geq c \left| \sum_{i=1}^{s} m_i \varphi_i \right|_{1, \Omega_h}^2 > 0
\]

and (6.3) is proved.

Let us suppose that \(u \in H^{m+2}(\Omega)\). By the Calderon theorem there exists an extension \(\tilde{u}\) of the function \(u\) onto \(\bar{\Omega}\) such that

\[
\|\tilde{u}\|_{m+2, \bar{\Omega}} \leq c \|u\|_{m+2, \Omega}.
\]

(6.4)

Let us denote

\[
\tilde{f} = -L \tilde{u}.
\]

(6.5)

Evidently the function \(\tilde{f}\) is an extension of the function \(f\).

Hence

\[
(f, v)_h = (\tilde{f}, v)_h = -(L \tilde{u}, v)_h.
\]

(6.6)

Substituting from (6.6) into (3.16) we get

\[
A_h(u_h, v) = -(L \tilde{u}, v), \quad \forall v \in V_h.
\]

(6.7)

Therefore from (5.2) we can see that the function \(u_h\) is the Ritz discrete approximation of the function \(\tilde{u}\). Since \(\tilde{u} = u = 0\) on \(\partial \Omega\) we may apply the
Theorem on the Ritz discrete approximation for \( \eta_d = u_h \) and \( U = \tilde{u} \). From (5.14), (5.15) and (6.4) we get the estimates (6.1) and (6.2).

**Remark:** The results formulated in theorem 6.1 represent a generalization of the results which have been obtained for \( H^1 \) norm by Ciarlet [4] and by Zlámal [12] for special cases. In the case of selfadjoint operator \( l \) they also improve the results which have been obtained for \( H^1 \) and \( L_2 \) norm by Ciarlet and Raviart [3]. They are similar to those obtained by Ženíšek [15] for \( C^m \)-elements.

### 7. ERROR ESTIMATE FOR PARABOLIC PROBLEMS

**Theorem 7.1:** Let \( w \) be the solution of the parabolic problem (4.1) with sufficiently smooth functions \( f, g \), \( u \), \( g \) satisfying the conditions (3.3), (3.4) and (4.2). Let \( \mathcal{G}_h \) be a \( k \)-regular (\( k \) is a positive integer such that \( k > n/2 - 1 \)) triangulation of the set \( \Omega \) with a sufficiently smooth boundary \( \partial \Omega \). Let the quadrature formula on the reference set \( \tilde{K} \) be of a degree \( d \geq \max(1, 2k-2) \) and let exist a positive constant \( c_1 \) independent of \( v \) and \( h \) such that

\[
\left\| v \right\|_{0, \Omega, h}^2 \leq (g(x), v, h) \quad \forall v \in V_h. \tag{7.1}
\]

Let a given \( v \)-step time discretization method \((\rho, \sigma)\) of an order \( q (\geq 1) \) be \( A_0 \)-stable. Besides \( A_0 \)-stability, we require that the method \((\rho, \sigma)\) be stable in the sense of Dahlquist and that the roots of the polynomial \( \sigma(\xi) \) with modulus equal to one be simple. Then the fully discrete problem (4.11) has one and only one solution \( w_h \) and there exists a constant \( c \) independent of \( t \) and \( h \) such that

\[
\left\| w^s - w_h^s \right\|_{0, \Omega, h} \leq c \left\{ h^{k+1} \left[ \sup_{(0, T)} \| w \|_{k+3, \Omega} + \sup_{(0, T)} \left\| \frac{\partial w}{\partial t} \right\|_{k+3, \Omega} \right] \right. \\
+ \Delta t^q \sup_{(0, T)} \left\| \frac{\partial^q}{\partial t^q} \frac{1}{w} \right\|_{k+1, \Omega} + \sum_{j=0}^{v-1} \left\| \eta_j^s - \nu_j^s \right\|_h \right\}, \quad \forall s \,(s\Delta t < T). \tag{7.2}
\]

where \( \eta_d \) is the Ritz discrete approximation of the Calderon extension \( \tilde{w} \) of the function \( w \).

**Remark 1:** From (7.2) we see that the \( L_2 \)-norm of the error is of a magnitude of the order \( \Delta t^q \) with respect to \( t \) and of the order \( h^{k+1} \) with respect to \( x \).

**Remark 2:** Evidently

\[
\left\| w^m - w_h^m \right\|_h \leq \left\| w^m - \eta_d^m \right\|_h + \left\| \eta_d^m - w_h^m \right\|_h.
\]
Hence, from (7.27) and from the theorem on the Ritz discrete approximation [see (5.16)] it follows

\[ \| w^s - w^s_h \|_h \leq c \left \{ \begin{array}{c} h^{k+1} \left [ \sup_{(0, T)} \| w \|_{k+1, \Omega} + \sup_{(0, T)} \left \| \frac{\partial w}{\partial t} \right \|_{k+1, \Omega} \right ] \\ + \Delta t^4 \sup_{(0, T)} \left \| \frac{\partial^{q+1} w}{\partial t^{q+1}} \right \|_{k+1, \Omega} + \sum_{j=0}^{v-1} \| \eta_j^* - w_j^* \|_h \end{array} \right \} \]

i.e. in (7.2) the norm \( \| . \|_{0, \Omega} \cap \Omega \) may be replaced by the norm \( \| . \|_h \).

**Remark 3:** From the Lemma 5.2 [see (5.12)] we can see that the condition (7.1) is satisfied for example in case that the quadrature formula on the reference set \( \tilde{K} \) for evaluation of the form \((.,.)\) is of a degree \( d \geq 2k \). Nevertheless this condition is not necessary. Using the quadrature formula

\[ \int_{\tilde{K}} \varphi(\tilde{x}) \, d\tilde{x} \approx \frac{\text{mes } \tilde{K}}{n} \left [ \varphi(0, \ldots, 0) + \varphi(1, 0, 0, \ldots, 0) + \ldots + \varphi(0, \ldots, 0, 1) \right ] \]

(which is of degree 1) in case of 1-regular triangulation (i.e. \( k = 1 \)-linear elements) it can be proved that (7.1) is satisfied, too.

**Proof of the theorem 7.1:** We know that the problem (4.11) is represented by the linear system of algebraic equations (4.13). Hence, to prove the existence and the uniqueness it suffices to show that the matrix \( \alpha_v M_h + \Delta t \beta_v K_h \) is positive definite. In the previous part we have proved that \( K_h \) is positive definite. In the case of \( A_0 \)-stable methods \( \alpha_v > 0, \beta_v > 0 \). Hence it is sufficient to prove that

\[ m^T M_h m \geq 0 \]  

(7.3)

for any nonzero vector \( m = [m_1, m_2, \ldots, m_s]^T \).

From (4.7), (5.11) and (7.1) it follows

\[ m^T M_h m = \left ( \sum_{i=1}^{s} m_i \varphi_i, \sum_{i=1}^{s} m_i \varphi_i \right )_h = \left \| \sum_{i=1}^{s} m_i \varphi_i \right \|_h^2 \geq c \left \| m_i \varphi_i \right \|_{0, \Omega}^2 > 0 \]

and (7.3) is proved. More, the matrix \( M_h \) is positive definite. Let us suppose that \( w \in H^{k+3}(\Omega), \forall t \in (0, T) \). By the Calderon theorem there exist extensions \( \tilde{w}, \tilde{w}_t \) of the functions \( w, \partial w/\partial t \) onto \( \tilde{\Omega} \) such that

\[ \| \tilde{w} \|_{k+3, \tilde{\Omega}} \leq c \| w \|_{k+3, \Omega}, \]  

(7.4)

where \( c \) is a constant independent of \( h \) and \( t \). Denote

\[ \tilde{f} = -L \tilde{w} + G(x) \tilde{w}_t. \]  

(7.5)

vol. 13, n° 3, 1979
Evidently the function $\tilde{f}$ is an extension of the function $f$. Obviously

$$
\|\tilde{w}^m - \eta^m_d\|_{0, \Omega} \leq \|\tilde{w}^m - \eta^m\|_{0, \Omega} + \|\eta^m_d - w_h^m\|_{0, \Omega},
$$

where $\eta_d$ is the Ritz discrete approximation of the function $\tilde{w}$. Since $\tilde{w}(x, t) = w(x, t) = 0$ on $\partial \Omega$ for every $t \in (0, T)$ we may apply the theorem on the Ritz discrete approximation. From (5.15) and (7.4) we get

$$
\|\tilde{w}^m - \eta^m_d\|_{0, \Omega} \leq c h^{k+1} \|w^m\|_{k+3, \Omega}, \quad \forall t \in (0, T),
$$

where $c$ is a constant independent of $t$ and $h$. From (5.2) and (7.5) it follows

$$
A_h(\eta^m_d, v) = -(L \tilde{w}^m, v)_h = (\tilde{f} - G(x) \tilde{w}_t, v)_h, \quad \forall v \in V_h.
$$

Hence

$$
A_h(\eta^m_d, v) = (f^m, v)_h - \left(\frac{\partial w^m}{\partial t}, v\right)_h, \quad \forall v \in V_h.
$$

Therefore from (4.11) we get for any $v \in V_h$

$$
\left(\sum_{j=0}^{\nu} \alpha_j (\eta^m_d + j - w_h^{m+j}), v\right)_h + \Delta t A_h\left(\sum_{j=0}^{\nu} \beta_j (\eta^m_d + j - w_h^{m+j}), v\right)_h
$$

$$
= \left(\sum_{j=0}^{\nu} \alpha_j w_h^{m+j}, v\right)_h - \left(\sum_{j=0}^{\nu} \alpha_j (w_h^{m+j} - \eta^m_d), v\right)_h - \left(\sum_{j=0}^{\nu} \alpha_j w_h^{m+j}, v\right)_h
$$

$$
+ \Delta t \left(\sum_{j=0}^{\nu} \beta_j f_h^{m+j}, v\right)_h - \Delta t \left(\sum_{j=0}^{\nu} \beta_j \frac{\partial w_h^{m+j}}{\partial t}, v\right)_h
$$

$$
+ \left(\sum_{j=0}^{\nu} \alpha_j w_h^{m+j}, v\right)_h - \Delta t \left(\sum_{j=0}^{\nu} \beta_j f_h^{m+j}, v\right)_h.
$$

Hence

$$
\left(\sum_{j=0}^{\nu} \alpha_j (\eta^m_d + j - w_h^{m+j}), v\right)_h + \Delta t A_h\left(\sum_{j=0}^{\nu} \beta_j (\eta^m_d + j - w_h^{m+j}), v\right)_h
$$

$$
=(g (\pi^m - \omega^m), v)_h, \quad \forall v \in V_h.
$$

where

$$
\pi^m = \sum_{j=0}^{\nu} \alpha_j w_h^{m+j} - \Delta t \beta_j \frac{\partial w_h^{m+j}}{\partial t},
$$

$$\omega^m = \sum_{j=0}^{\nu} \alpha_j (w_h^{m+j} - \eta^m_d + j).
$$
We write (7.9) in a matrix form. For this purpose, let \( v \) be the vector \( v = M_h^{1/2} \phi \), where \( \phi = [\phi_1, \ldots, \phi_s]^T \) is the vector of the basis functions. Let us set \( \eta_d^m - w_h^m = (e^m)T v \) (notice that \( \eta_d^m - w_h^m \in V_h \)). Since \( (g \phi, \phi^T)_h = M_h \) and \( A_h(\phi, \phi^T) = K_h \) we have \( (g v, v^T)_h = I \) and \( A_h(v, v^T) = M_h^{-1/2} K_h M_h^{-1/2} \). The matrix \( S_h = M_h^{-1/2} K_h M_h^{-1/2} \) is symmetric and positive definite. Putting the components \( v_i (i = 1, \ldots, s) \) of the vector \( v \) for \( v \) in (7.9) we get

\[
\sum_{j=0}^{v} (\alpha_j I + \Delta t \beta_j S_h) e^{m+j} = c_h^m,
\]

(7.12)

where

\[
c_h^m = (g (\pi^m - \omega^m), v)_h.
\]

(7.13)

Denote

\[
\delta_j(\tau) = \frac{\alpha_j + \beta_j \tau}{\alpha_v + \beta_v \tau}, \quad d_h^m = (\alpha_v I + \Delta t \beta_v S_h)^{-1} c_h^m
\]

(7.14)

the matrix \( \alpha_v I + \Delta t \beta_v S_h \) is positive definite). Then

\[
\sum_{j=0}^{v} \delta_j(\Delta t S_h) e^{m+j} = d^m
\]

and this difference equation will be solved in the way described by Zlámal [10], pp. 355-356 who used the ideas given in Henrici [5], pp. 242-244 for ordinary differential equations. From Zlámal's result (see [10], pp. 355-356) we get

\[
\| e^m \| \leq c \left( \sum_{j=0}^{v-1} \| e^j \| + \sum_{j=0}^{m-v} \| c_h^j \| \right)
\]

(7.15)

(by \( \| . \| \) we denote the Euclidean norm of a vector or of a matrix). Since

\[
\| \eta_d^m - w_h^m \|_h^2 = (g (\eta_d^m - w_h^m), \eta_d^m - w_h^m)_h = (g e^m T v, v^T e^m)_h = e^m (g v, v^T)_h e^m = \| e^m \|^2
\]

we get from (7.15):

\[
\| \eta_d^m - w_h^m \|_h \leq c \left( \sum_{j=0}^{v-1} \| \eta_d^m - w_h^m \|_h + \sum_{j=0}^{m-v} \| c_h^j \| \right).
\]

(7.16)

Let \( \phi \in L_2 (\Omega) \) be any function and let \( \psi \in V_h \) be its orthogonal projection onto \( V_h \) in the norm \( \| . \|_h \), i.e.:

\[
(g (\phi - \psi), v)_h = 0, \quad \forall v \in V_h.
\]

(7.17)
Then
\[ 0 \leq \| \varphi - \psi \|_h^2 = (g(\varphi - \psi), \varphi - \psi)_h = (g(\varphi - \psi), \varphi)_h - (g(\varphi - \psi), \varphi)_h = \| \varphi \|_h^2 - \| \psi \|_h^2. \]

Hence
\[ \| \psi \|_h \leq \| \varphi \|_h. \]  \hspace{1cm} (7.18)

Putting \( \psi = \psi^T v = v^T \psi \) we get
\[ \| \psi \|_h^2 = (g(\psi, \psi)_h = (g(\psi^T v, v^T \psi)_h = \psi^T (g v^T, v)_h = \| \psi \|^2 \]  \hspace{1cm} (7.19)

and
\[ (g(\varphi, v)_h = (g(\varphi, v), v)_h = (g(\psi^T v, v^T \psi)_h = (g(\psi^T g v, v^T)_h = \psi. \]

Therefore from (7.19) and from (7.18) it follows
\[ \| (g(\varphi, v)_h \leq \| \varphi \|_h. \]

Hence, from (7.13) we get
\[ \| \psi \|_h \leq \| \psi \|_h. \]

Substituting this inequality into (7.16) we obtain
\[ \| \eta^T - w^T \|_h \leq c \left( \sum_{j=0}^{\nu-1} \| \eta_j^T - w_j^T \|_h + \sum_{j=0}^{m-\nu} \| \pi_j - \omega_j \|_h \right) \quad (7.20) \]

Evidently
\[ \| \pi^j - \omega_j \|_h \leq \| \pi^j \|_h + \| \omega_j \|_h. \]  \hspace{1cm} (7.21)

To estimate \( \| \pi^j \|_h \) we use the assumption that the scheme \((\rho, \sigma)\) is of order \(q\). It means that for any function \( y(t) \in C^s, s \leq q + 1 \), it holds
\[ \sum_{j=0}^{\nu} A_j y(t + j \Delta t) - \Delta t \sum_{j=0}^{\nu} B_j y(t + j \Delta t) = O(\Delta t^{s+1}|y^{(s)}(t + \tau)|) \]

Hence, from (7.10) and from the first Sobolev theorem it follows
\[ \| \pi \|_h \leq c \Delta t^{q+1} \sup_{(0, T)} \left| \frac{\partial^{q+1} w}{\partial t^{q+1}} \right| \leq c \Delta t^{q+1} \sup_{(0, T)} \left| \frac{\partial^{q+1} w}{\partial t^{q+1}} \right|_{k+1, \Omega} \quad (7.22) \]

provided that the function \( \partial^{q+1} w(x, t)/\partial t^{q+1} \) is continuous for every \( x \in \Omega \). From (3.12) we have
\[ \| \pi^j \|_h = (g(\pi^j, \pi^j)_h = \sum_{K \in \mathcal{V}_h} \sum_{r} \omega_{K,r} (g(b_{r,K})(\pi^j(b_{r,K}))^2. \]

R.A.I.R.O. Analyse numérique/Numerical Analysis
Hence, (7.22) implies
\[ \| \pi^j \|_h^2 \leq c \Delta t^2 (q+1) \left( \sup_{(0,T)} \left\| \frac{\partial^{q+1} u}{\partial t^{q+1}} \right\|_{k+1, \Omega} \right)^2 \sum_{K \in \mathcal{T}_h} \sum_{r} \omega_{r,K} \] (7.23)

From (2.5) and from (1.2) we get
\[ \sum_{K \in \mathcal{T}_h} \sum_{r} \omega_{r,K} = \sum_{K \in \mathcal{T}_h} \sum_{r} \phi_r J_K (\hat{t}_r) \leq \sum_{K \in \mathcal{T}_h} \max_{r} J_K (\hat{t}) \sum_{r} \phi_r \]

\[ = \sum_{K \in \mathcal{T}_h} \max_{r} J_K (\hat{t}) \int_{K} d\hat{\chi} \leq \sum_{K \in \mathcal{T}_h} \frac{\kappa}{\min_{r} J_K (\hat{t})} \int_{K} J_K (\hat{t}) d\hat{\chi} \]

\[ \leq c_0^2 \sum_{K \in \mathcal{T}_h} \int_{K} J_K (\hat{t}) d\hat{\chi} = c_0^2 \sum_{K \in \mathcal{T}_h} \text{mes } K = c_0^2 \text{ mes } \Omega_h \leq c_0^2 \text{ mes } \Omega. \]

Hence, (7.23) implies
\[ \| \pi^j \|_h \leq c \Delta t^{q+1} \sup_{(0,T)} \left\| \frac{\partial^{q+1} u}{\partial t^{q+1}} \right\|_{k+1, \Omega}. \] (7.24)

A simple calculation gives
\[ \omega^m = \sum_{j=0}^{\nu} \alpha_j (w^{m+j} - \eta_d^{m+j}) \]
\[ = \sum_{j=1}^{\nu} \gamma_j [w^{m+j} - \eta_d^{m+j} - (w^{m+j-1} - \eta_d^{m+j-1})] + \gamma_0 (w^m - \eta_d^m), \]

where \( \gamma_j = \sum_{i=j}^{\nu} \alpha_i. \) From the consistency of the scheme (\( \rho, \sigma \)) it follows that \( \gamma_0 = 0. \)

Hence
\[ \| \omega^m \|_h \leq c \sum_{j=1}^{\nu} \| (w^{m+j} - w^{m+j-1}) - (\eta_d^{m+j} - \eta_d^{m+j-1}) \|_h \]
\[ \leq c \sum_{j=1}^{\nu} \| (\tilde{w}^{m+j} - \tilde{w}^{m+j-1}) - (\eta_d^{m+j} - \eta_d^{m+j-1}) \|_h. \] (7.25)

Evidently \( \eta_d^{m+j} - \eta_d^{m+j-1} \) is the Ritz discrete approximation of the function \( \tilde{w}^{m+j} - \tilde{w}^{m+j-1}. \) We may apply the Theorem on the Ritz discrete approximation. From (5.16), (7.25) and from the Calderon theorem we get for \( (m+j-1) \Delta t \leq \mathcal{H}_j < (m+j) \Delta t: \)
\[ \| \omega^m \|_h \leq c h^{k+1} \sum_{j=1}^{\nu} \| \tilde{w}^{m+j} - \tilde{w}^{m+j-1} \|_{k+3, \Omega} \]
\[ \leq c h^{k+1} \sum_{j=1}^{\nu} \| w^{m+j} - w^{m+j-1} \|_{k+3, \Omega} \leq c h^{k+1} \Delta t \sum_{j=1}^{\nu} \left\| \frac{\partial w^m}{\partial t} \right\|_{k+3, \Omega}. \]
Hence

\[ \| \omega^m \|_h \leq c h^{k+1} \Delta t \sup_{(0, T)} \left\| \frac{\partial w}{\partial t} \right\|_{k+3, \Omega} \]  

(7.26)

From (7.21), (7.24) and (7.26) we get

\[
\sum_{j=0}^{m-v} \| \pi^j - \omega^j \|_h \leq c \sum_{j=0}^{m-v} \Delta t^{q+1} \sup_{(0, T)} \left\| \frac{\partial q^{+1} w}{\partial t^{q+1}} \right\|_{k+1, \Omega} + h^{k+1} \Delta t \sup_{(0, T)} \left\| \frac{\partial w}{\partial t} \right\|_{k+3, \Omega} \sum_{j=0}^{m-v} \Delta t
\]

\[
= c T \left[ \Delta t^{q} \sup_{(0, T)} \left\| \frac{\partial q^{+1} w}{\partial t^{q+1}} \right\|_{k+1, \Omega} + h^{k+1} \sup_{(0, T)} \left\| \frac{\partial w}{\partial t} \right\|_{k+3, \Omega} \right] \sum_{j=0}^{m-v} \Delta t
\]

Hence, the inequality (7.20) implies

\[
\| \eta^m_n - \psi_n^m \|_h \leq c \left\{ \Delta t^{q} \sup_{(0, T)} \left\| \frac{\partial q^{+1} w}{\partial t^{q+1}} \right\|_{k+3, \Omega} + h^{k+1} \sup_{(0, T)} \left\| \frac{\partial w}{\partial t} \right\|_{k+3, \Omega} + \sum_{j=0}^{v-1} \| \eta^j_n \psi_n^j \|_h \right\} . \]  

(7.27)

From (7.6), (7.7), (7.1) and from (7.27) we get (7.2).

REFERENCES


R.A.I.R.O. Analyse numérique/Numerical Analysis


