Received December 2003
Revised January 2005
Accepted April 2005

Isoparametric finite element approximation for a boundary flux problem

Andrey B. Andreev
Department of Applied Informatics, Technical University, Gabrovo, Bulgaria, and
Todor D. Todorov
Department of Mathematics, Technical University, Gabrovo, Bulgaria

Abstract

Purpose – To study and to analyze a second order finite-element boundary-flux approximation using isoparametric numerical integration.

Design/methodology/approach – The numerical finite-element integration is the main method used in this research. Since a domain with curved boundary is considered we apply an isoparametric approach. The lumped flux formulation is another method of approach in this paper.

Findings – This research study presents a careful analysis of the combined effect of the numerical integration and isoparametric FEM on the boundary-flux error. Some L_2-norm estimates are proved for the approximate solutions of the problem under consideration.

Research limitations/implications – The authors offer a general study within the framework of the boundary-flux approximation theory, which completes the results of published works in this scientific field of research.

Practical implications – A useful application is to employ appropriate quadrature formulae without violating the precision of the boundary-flux FEM. The lumped mass approximation is also an important practical approach to the problem in question.

Originality/value – The paper presents an entire investigation in FE boundary-flux approximation theory, in particular, elements of arbitrary degree and domains with curved boundaries. The work is addressed to the possible related fields of interest of postgraduate students and specialists in fluid mechanics and numerical analysis.

Keywords Finite element analysis, Boundary-elements methods

Paper type Research paper

1. Introduction

Calculation of derivatives (flux or stresses) of finite element solutions to boundary value problems has many important applications such as the heat of the mass transfer, potential flow, plate stability, etc. Computing of boundary-flux is based on the idea proposed by Wheeler (1973) and developed by Carey (1982) and Carey *et al.* (1985). Different methods for boundary-flux approximations have been tested in a series of numerical experiments on linear and nonlinear elliptic problems (Carey *et al.*, 1985).

The standard procedure of differentiating of the approximate solution at an arbitrary boundary point in the finite element will give an asymptotic error $O(h^{n-(1/2)})$ for trial functions of degree n (see, for example, the work of Barret and Elliot (1987)). Higher order approximations to the boundary-flux than those that arise from the straightforward differentiation of the Galerkin solution are presented in many papers (Douglas *et al.*, 1974; Lazarov and Pehlivanov, 1989; Pehlivanov *et al.*, 1992). More

International Journal of Numerical Methods for Heat & Fluid Flow
Vol. 16 No. 1, 2006
pp. 46-66
© Emerald Group Publishing Limited
0961-5539
DOI 10.1108/09615530610636955

recent study on superconvergent boundary-flux approximation has been presented by Carey (2002).

This paper deals with a finite element method for planar second-order boundary-flux problem on a bounded domain with curved boundary. Curved elements are used in the boundary layer for getting good approximation of the boundary. Quadrature formulae are used for computing integrals in the discrete problem. The isoparametric approach to the problem in question requires a careful study of the combined effect of the numerical integration and isoparametric elements on the boundary-flux error.

Drawing a presentation of L_2-error estimate of the effect of the numerical integration has been the underlying purpose of this paper. Furthermore, we aimed at analyzing the lumped mass flux formulation and presenting some consequent algorithmic aspects.

The paper is organized as follows. Section 2 presents the boundary-flux problem, whereas Section 3 precisely defines the isoparametric finite element transformations and the numerical quadrature schemes. Next, the corresponding discrete formulations are introduced and an error analysis for isoparametric triangular finite elements of degree $n \geq 2$ is developed in Section 4. Section 5 is devoted to the lumped mass flux formulation. Numerical tests confirming the theoretical results are presented in Section 6. The closing section contains some concise generalizations.

2. Problem formulation

Let $\Omega \subset \mathbf{R}^2$ be a bounded curved domain with Lipschitz-continuous boundary Γ. Consider the Dirichlet problem

$$\mathscr{P}\begin{cases} \text{find a function } u \text{ satisfying} \\ Lu = f \text{ in } \Omega, \\ u = 0 \text{ on } \Gamma \end{cases},$$

where

$$Lu = -\sum_{i,j=1}^{2} \frac{\partial}{\partial x_j}\left(a_{ij}\frac{\partial u}{\partial x_i}\right),$$

is a linear elliptic operator. Suppose that the matrix $A = (a_{ij}(x))_{i,j\in\{1,2\}}$ is uniformly positive definite in Ω and the coefficients $a_{ij} = a_{ji},\ i,j = 1,2$ belong to $C^1(\Omega)$. Therefore, the operator L is strongly elliptic.

Standard notations for the Sobolev spaces (Adams, 1975) and associated norms and seminorms are used throughout this consideration. Let $\mathbf{V} = H_0^1(\Omega)$, where

$$H_0^1(\Omega) = \{v \in H^1(\Omega) | v = 0 \text{ on } \Gamma\}.$$

We associate the usual bilinear form

$$a(u,v) = \int_\Omega \sum_{i,j=1}^{2} a_{ij}(x)\frac{\partial u}{\partial x_i}\,\frac{\partial v}{\partial x_j}\,\mathrm{d}x, \quad u,v \in \mathbf{V}$$

and the linear functional

$$(f, v) = \int_\Omega fv\,\mathrm{d}x, \quad v \in \mathbf{V}$$

with the problem ($\mathscr{P}$).

Since L is strongly elliptic the bilinear form $a(\cdot\,,\cdot)$ is coercive on $\mathbf{V}\times\mathbf{V}$. Furthermore, the boundedness of a_{ij} on $\overline{\Omega}$ implies that $a(\cdot\,,\cdot)$ is continuous on $H^1(\Omega)$.

The weak formulation of the problem ($\mathscr{P}$) is

$$\mathscr{P}_W \begin{cases} \text{find a function } u \in \mathbf{V} \text{ such that} \\ a(u, v) = (f, v), \quad \forall v \in \mathbf{V}. \end{cases}$$

We use the usual hypotheses for the smoothness of the weak solution.

C1 The boundary Γ is piecewise C^{n+1}.

C2 The right hand side $f \in W^{n,\infty}(\Omega)$ and the weak solution $u \in H^{n+1}(\Omega)$, $n \geq 2$.

C3 The coefficients $a_{ij} \in W^{n,\infty}(\Omega)$.

Let us define the vector function

$$\underline{\sigma} = -A^{\mathrm{t}}(\nabla u)^{\mathrm{t}},$$

where u is the solution of ($\mathscr{P}_W$) and "t" is the sign for transposition.

The normal flux across boundary Γ is defined by

$$q = \underline{\sigma}\cdot\underline{n} = -\sum_{i,j=1}^{2} a_{ij}(x)\frac{\partial u}{\partial x_i}\cos(\underline{n}, x_j), \quad x \in \Gamma,$$

where $\underline{n}$ is the outward normal vector to the boundary Γ. Multiplying the equality in ($\mathscr{P}$) by a function $v \in H^1(\Omega)$ and using integration by parts we get the following relation for the flux q.

$$-\langle q, v\rangle = a(u, v) - (f, v), \quad \forall v \in H^1(\Omega), \tag{1}$$

where $\langle\cdot\,,\cdot\rangle$ denotes the inner product on the boundary, i.e.

$$\langle q, v\rangle = \int_\Gamma qv\,\mathrm{d}s.$$

Corresponding Sobolev spaces and associated norms for the functions defined on the boundary are connected with this notation. Using the interpolation of the functional spaces for m integer, we get the corresponding Sobolev spaces for noninteger m (Adams, 1975).

3. Isoparametric finite element method and numerical integration

The point of interest is in the approximation of the flux of the solution of ($\mathscr{P}_W$) by an isoparametric finite element method of Lagrangian type (Ciarlet, 1978; Ciarlet and Raviart, 1972a). To that end, for each $0 < h \leq 1$ let τ_h be a triangulation of Ω by triangular finite elements isoparametric equivalent to one finite element $(\hat{K}, \hat{P}, \hat{\Sigma})$ called finite element of reference:

$\hat{K} = \{(\hat{x}_1, \hat{x}_2) | \quad \hat{x}_1 \geq 0, \ \hat{x}_2 \geq 0, \ \hat{x}_1 + \hat{x}_2 \leq 1\}$ is the canonical 2-simplex;

$\hat{P} = P_n(\hat{K})$, where P_n is the space of all polynomials of degree, not exceeding *n;*

$\hat{\Sigma} = \{\hat{x} = (\hat{x}_1, \hat{x}_2) | \quad \hat{x}_1 = i/n; \ \hat{x}_2 = j/n; \ i + j \leq n; \ i,j \in \mathbf{N} \cup \{0\}\}$ is the set of all Lagrangian interpolation nodes of order n.

Define the edges of the finite element of reference $\hat{K}$

$$\hat{K}_i = \{\hat{x} \in \hat{K} | \quad \hat{x}_i = 0\}, \quad i = 1, 2, 3,$$

where $\hat{x}_3 = 1 - \hat{x}_1 - \hat{x}_2$.

Let $(\hat{T}, \hat{P}_{\hat{T}}, \hat{\Sigma}_{\hat{T}})$ be the one-dimensional finite element of reference corresponding to $\hat{K}$:

$\hat{T} = \{\hat{t} | 0 \leq \hat{t} \leq 1\}$ is the interval [0,1];

$\hat{P}_{\hat{T}} = P_n(\hat{T})$;

$\hat{\Sigma}_{\hat{T}} = \{\hat{t} = i/n, | \, i = 0, 1, 2, \ldots, n\}$ is the set of all Lagrangian interpolation nodes of order n.

An arbitrary finite element $K \in \tau_h$ is defined by $K = F_K(\hat{K})$, where $F_K \in \hat{P}^2$ is an invertible transformation.

We use not only straight elements but also isoparametric elements with one curved side for getting good approximation of the boundary Γ. Thus we obtain a perturbed domain $\Omega_h = \cup_{K \in \tau_h} K$ of the domain Ω with boundary Γ_h.

Denote the boundary layer of τ_h by

$$B_h = \{K \in \tau_h | K \text{ has more than one node on the boundary}\}.$$

Let $K_3 = F_K(\hat{K}_3)$. We input the map $\chi : \hat{T} \rightarrow \hat{K}_3$ defined by $\chi(\hat{t}) = (\hat{t}, 1 - \hat{t})$. Thus we obtain the map $\chi_K : \hat{T} \rightarrow K_3$ determined by $\chi_K = F_K \circ \chi$.

To use optimal finite elements the validity of the hypotheses with respect to any used triangulation τ_h is needed.

T1 Only the elements of B_h have curved edge.

T2 The edges K_i, $i = 1, 2$ are straight edges for all elements K in any used triangulation τ_h.

T3 If $K \in B_h$ then meas $(K_3 \cap \Gamma_h) \neq 0$.

T4 Any considered triangulation is n-regular in the sense of Ciarlet and Raviart (1972b).

T5 The triangulation τ_h is *quasi*-uniform for any considered h so that the usual inverse inequalities hold (Ciarlet, 1978).

Let $P_K = \{p : K \rightarrow \mathbf{R} | \quad p = \hat{p} \circ F_K^{-1}, \ \hat{p} \in \hat{P}\}$. Then the finite element space $\mathbf{V}_h$ is defined by

$$\mathbf{V}_h = \{v \in C(\Omega_h) | v|_K \in P_K,\ \ K \in \tau_h\},$$

associated with a triangulation τ_h. It is well known that $\mathbf{V}_h \subset H_0^1(\Omega_h)$.

Let $\tilde{\Omega}$ be a bounded open set satisfying $\Omega \subset \tilde{\Omega}$, $\Omega_h \subset \tilde{\Omega}$ for all considered triangulations τ_h. Suppose that every function from $H_0^1(\Omega)$ ($H_0^1(\Omega_h)$) is extended by zero outside of Ω (Ω_h) to $\mathbf{R}^2$ in a continuous way. We shall also use the space $\mathbf{W}_h = \mathbf{V} + \mathbf{V}_h$. Define the approximating bilinear form

$$A_h(u, v) = \int_{\tilde{\Omega}} \sum_{i,j=1}^{2} \tilde{a}_{ij}(x) \frac{\partial u}{\partial x_i} \frac{\partial v}{\partial x_j}\, \mathrm{d}x, \ \ u, v \in \mathbf{W}_h, \tag{2}$$

where $\tilde{a}_{ij}(x) \in W^{n,\infty}(\tilde{\Omega})$ are continuous extensions of the coefficients $a_{ij}(x)$ to $\tilde{\Omega}$. The scalar product in the spaces $L_2(\Omega)$, $L_2(\Omega_h)$ and $L_2(\tilde{\Omega})$ will be written by one and the same denotation $(\cdot\,,\,\cdot)$. Suppose that the bilinear forms (2) are uniformly $\mathbf{W}_h$-elliptic, i.e. there exists a constant $\tilde{\beta} > 0$ independent of the spaces $\mathbf{W}_h$, such that for all h sufficiently small and $\forall v \in \mathbf{W}_h$

$$\tilde{\beta} \|v\|_{1,\tilde{\Omega}}^2 \leq A_h(v, v).$$

Define the discrete problem corresponding to the problem $(\mathscr{P}_W)$

$$\tilde{\mathscr{P}}_h \begin{cases} \text{find } \ u_h^* \in \mathbf{V}_h \ \text{ that} \\ A_h\left(u_h^*, v\right) = (\tilde{f}, v), \quad \forall v \in \mathbf{V}_h, \end{cases}$$

where $\tilde{f} \in W^{n,\infty}(\tilde{\Omega})$ is an extension of the right hand side f to $\tilde{\Omega}$.

The finite element approximation $(\tilde{\mathscr{P}}_h)$ of the problem $(\mathscr{P}_W)$ has matrices whose elements involve integrals which, except in very simple examples, must be evaluated by numerical integration or quadrature rules.

To evaluate integrals over finite elements $\hat{T}$ and $\hat{K}$ numerically, quadrature formulas

$$\int_{\hat{T}} \hat{\varphi}(\hat{t}) \mathrm{d}\hat{t} \cong \hat{\mathscr{I}}(\hat{\varphi}) = \sum_{i=1}^{N} \hat{\nu}_i \hat{\varphi}(\hat{d}_i), \quad \hat{\varphi} \in C(\hat{T}),\ \ \hat{d}_i \in \hat{T}, \tag{3}$$

$$\int_{\hat{K}} \hat{\varphi}(\hat{x}) \mathrm{d}\hat{x} \cong \hat{I}(\hat{\varphi}) = \sum_{i=1}^{L} \hat{\omega}_i \hat{\varphi}(\hat{b}_i), \quad \hat{\varphi} \in C(\hat{K}),\ \ \hat{b}_i \in \hat{K} \tag{4}$$

are used. Denote the set of the nodes of quadrature formulas (3) and (4) by $\mathcal{N}(\hat{T})$ and $\mathcal{N}(\hat{K})$, respectively.

Assume that the following hypotheses concerning the quadrature formulas hold.

Q1. All the coefficients of the quadrature formulas (3) and (4) are strictly positive.

Q2. The set $\mathcal{N}(\hat{T})$ ($\mathcal{N}(\hat{K})$) contains $\hat{P}_n(\hat{T}) -$ ($\hat{P}_n(\hat{K})-$) unisolvent subset.

Some properties of the quadrature formulas (3) and (4) will be determined and applied in the next section.

The quadrature formula on the edge K_3, $K \in B_h$ corresponding to equation (3) is

$$\int_{K_3} \varphi(x)\mathrm{d}l \cong \mathscr{I}_{K_3}(\varphi) = \hat{\mathscr{I}}(|D\chi_K|\hat{\varphi}), \tag{5}$$

where $\hat{\varphi}(\hat{t}) = \varphi\big(\chi_K^{-1}(x)\big)$, $x \in K_3$ and $|\underline{w}(w_1, w_2)| = \sqrt{w_1^2 + w_2^2}$, $\quad \underline{w} \in \mathbf{R}^2$.

The quadrature formula over the finite element K for continuous φ is

$$\int_K \varphi(x)\mathrm{d}x \cong I_K(\varphi) = \hat{I}(J(F_K)\hat{\varphi}), \tag{6}$$

where $J(F_K)$ is the Jacobian of F_K.

The integrals over Γ_h and Ω_h will be computed element by element using equations (5) and (6), respectively. Thus we obtain the approximate bilinear and linear forms

$$a_h(v, w) = \sum_{K \in \tau_h} \sum_{i,j=1}^{2} I_K\left(a_{ij}\frac{\partial v}{\partial x_i}\frac{\partial w}{\partial x_j}\right),$$

$$(v, w)_h = \sum_{K \in \tau_h} I_K(vw), \quad \forall v, w \in \mathbf{V}_h.$$

Accounting of the fact that $a_h(\cdot\,,\,\cdot)$ is uniformly $\mathbf{V}_h$-elliptic (see Theorem 4.4.2 by Ciarlet (1978)), we define the approximate problem

$$\mathscr{P}_h \begin{cases} \text{find } u_h \in \mathbf{V}_h \text{ such that} \\ a_h(u_h, v) = (f, v)_h, \quad \forall v \in \mathbf{V}_h \end{cases}$$

obtained by numerical integration.

Further, we shall apply the construction of n-regular isoparametric triangulation τ_h presented by Lenoir (1986). Consider a finite element space associated with a triangulation τ_h by

$$\mathscr{V}_h = \{v_h \in C(\Omega_h) | v_h|_K \in P_K,\ v_h = 0 \text{ at the corners of } \Omega,\ \mathrm{K} \in \tau_h\}.$$

For any function $v \in H^1(\Omega)$ we obtain a function $v_h = \Pi_h(v \circ \Phi_h)$, $v_h \in \mathscr{V}_h$, by means of an invertible mapping $\Phi_h : \Omega_h \to \Omega$ constructed by Lenoir (1986) and by means of an interpolation operator Π_h over the whole triangulation τ_h. We input the restriction $\phi_h : \Gamma_h \to \Gamma$ of the map Φ_h, and the restriction of the space $\mathscr{V}_h$ on the boundary Γ_h

$$S_h = \{w_h | w_h = v_h|_{\Gamma_h}, \quad v_h \in \mathscr{V}_h\}.$$

Then it is possible to define the approximation $g_h \in S_h$ of any function $g \in C(\Gamma)$ by $g_h = \pi_h(g \circ \phi_h)$, where π_h is an interpolation operator on the whole boundary Γ_h.

Assume that the finite element solution u_h of the problem $(\mathscr{P}_h)$ is already found. Then the approximate flux across Γ_h can be constructed as a function $q_h \in S_h$ such that

$$-\langle q_h, v_h \rangle_h = a_h(u_h, v_h) - (f, v_h)_h, \quad \forall v_h \in \mathscr{V}_h, \tag{7}$$

where for the approximate inner product on Γ_h we use

$$\langle q_h, v_h \rangle_h = \sum_{K \in B_h} \mathscr{I}_{K_3}(q_h v_h), \quad \forall v_h \in \mathbf{V}_h.$$

The identity (7) leads to a system of equations for the unknown values of q_h at some points on Γ_h and, consequently, on Γ. This procedure has been proposed by Carey *et al.* (1985) for the consistent case.

The essential purpose of the present paper is to analyze the isoparametric case when the numerical quadratures are applied to compute the inner products in equation (7).

Notations $C, C_1, C_2, \ldots,$ are reserved for generic positive constants, which may vary with the context.

4. Error estimate for the boundary-flux

We associate the error functionals with quadrature schemes considered in the previous section. Let

$$E_K(\varphi) = \int_K \varphi \, \mathrm{d}x - I_K(\varphi).$$

Then the total quadrature error is

$$E(u, v) = \sum_{K \in \tau_h} E_K(uv), \quad u, v \in \mathscr{V}_h.$$

Moreover, if

$$\hat{E}(\hat{\varphi}) = \int_{\hat{K}} \hat{\varphi} \, \mathrm{d}\hat{x} - \hat{I}(\hat{\varphi}),$$

then

$$E_K(\varphi) = \hat{E}(J(F_K)\hat{\varphi}).$$

Similarly, according to the quadrature formulas (3) and (5), we define

$$\hat{\mathscr{E}}(\hat{\varphi}) = \int_{\hat{T}} \hat{\varphi} \, \mathrm{d}\hat{t} - \hat{\mathscr{I}}(\hat{\varphi}),$$

and

$$\mathscr{E}_K(\varphi) = \hat{\mathscr{E}}(|D\chi_K|\hat{\varphi}), \ \ \varphi \in S_h, \ \ K \in B_h,$$
$$\mathscr{E}(v, w) = \sum_{K \in B_h} \mathscr{E}_K(vw), \quad v, w \in S_h.$$

Suppose that the numbering of the nodes of each element $K \in \tau_h$ is such that $J(F_K)(\hat{x}) > 0$, $\forall \hat{x} \in \hat{K}$.

The space S_h is provided with a norm

$$\|w\|_h = \sqrt{\langle w, w \rangle_h}, \quad \forall w \in S_h.$$

If a map $F(x)$ is k-times differentiable, we denote the kth Fréchet derivative of $F(x)$ by $D^k F(x)$. Let $\mathscr{L}_n(\mathbf{R}^2; \mathbf{R}^2)$ is the space of the continuous n-linear mappings from $(\mathbf{R}^2)^n$ to

$\mathbf{R}^2$ and $\hat{K}$, K are bounded subsets of $\mathbf{R}^2$. For estimating the Fréchet derivatives and Jacobians we need the following seminorms

$$|F|_{n,\infty,\hat{K}} = \sup_{\hat{x}\in\hat{K}} \|D^n F(\hat{x})\|_{\mathscr{L}_n(\mathbf{R}^2;\mathbf{R}^2)}, \ |F^{-1}|_{n,\infty,K} = \sup_{x\in K} \|D^n F^{-1}(x)\|_{\mathscr{L}_n(\mathbf{R}^2;\mathbf{R}^2)}, \ n = 0,1,2,\ldots$$

for arbitrary sufficiently smooth transformation $F:\hat{K}\to K$ with sufficiently smooth inverse transformation F^{-1}. Further, we shall proceed by omitting the index $\mathscr{L}_n(\mathbf{R}^2;\mathbf{R}^2)$ of the norms of the Fréchet derivatives and write only $\|\cdot\|$ instead of $\|\cdot\|_{\mathscr{L}_n(\mathbf{R}^2;\mathbf{R}^2)}$.

Lemma 1. The norms $\|\cdot\|_h$ and $\|\cdot\|_{0,\Gamma_h}$ are uniformly equivalent on the space S_h, i.e. there exists constants c_1, $c_2 > 0$, independent of h, such that

$$c_1\|v\|_h \le \|v\|_{0,\Gamma_h} \le c_2\|v\|_h, \quad \forall v \in S_h. \tag{8}$$

Proof. First we argue that the map

$$\hat{p} \in \hat{P}(\hat{T}) \to \left(\sum_{i=1}^{N} \hat{\nu}_i \hat{p}^2(\hat{d}_i)\right)^{1/2} \in \mathbf{R}^+$$

is a norm on $P_n(\hat{T})$. It is sufficient to note that $\hat{p}(\hat{d}_i) = 0$, $1 \le i \le N$, implies according to **Q1** and **Q2** that $\hat{p} = 0$, since the set $\mathscr{N}(\hat{T})$ is $P_n(\hat{T})$-unisolvent and $N \ge n+1$. From the equivalence of all norms on the finite dimensional space $P_n(\hat{T})$, we infer the existence of the constants c_1, $c_2 > 0$, independent of h, such that

$$c_1\left(\sum_{i=1}^{N} \hat{\nu}_i \hat{p}^2(\hat{d}_i)\right)^{1/2} \le \|\hat{p}\|_{0,\hat{T}} \le c_2\left(\sum_{i=1}^{N} \hat{\nu}_i \hat{p}^2(\hat{d}_i)\right)^{1/2}, \quad \forall \hat{p} \in P_n(\hat{T}). \tag{9}$$

Denote the edge K_3 of the element $K \in B_h$ by T_K. We obtain

$$c_1(\mathscr{I}(p^2))^{1/2} \le \|p\|_{0,T_K} \le c_2(\mathscr{I}(p^2))^{1/2} \tag{10}$$

from equations (5), (9) and

$$C\left(\inf_{\hat{t}\in\hat{T}} \|D\chi_K\|\right)^{1/2} |\hat{f}|_{0,\hat{T}} \le |f|_{0,T_K} \le C(|D\chi_K|_{0,\infty,\hat{T}})^{\frac{1}{2}} |\hat{f}|_{0,\hat{T}}.$$

The restriction $v_h|_{T_K}$, $v_h \in S_h$ belongs to $P_n(T_K)$ for all $K \in B_h$. Since

$$|v_h|_{0,\Gamma_h} = \left(\sum_{K\in B_h} |v_h|^2_{0,T_K}\right)^{1/2}$$

the inequalities just obtained in equation (10) imply equation (8). □

The following lemma is devoted to estimates of the total quadrature errors in Ω_h and on Γ_h.

Lemma 2. Let the hypotheses **T1-T5** and **Q1-Q2** hold and quadrature schemes over reference finite elements $\hat{T}$ and $\hat{K}$ be such that

$$\hat{\mathscr{E}}(\hat{\varphi}) = 0, \quad \hat{\varphi} \in P_{2n-1}(\hat{T}), \;\; n \geq 2, \tag{11}$$

$$\hat{E}(\hat{\varphi}) = 0, \quad \hat{\varphi} \in P_{2n-2}(\hat{K}), \;\; n \geq 2. \tag{12}$$

If $f \in H^n(\Omega)$ and $q \in H^n(\Gamma)$ the following estimates hold for all $v_h \in \mathscr{V}_h$

$$|E(f, v_h)| \leq Ch^n \|f\|_{n,\Omega} \|v_h\|_{1,\Omega_h}, \tag{13}$$

$$|\mathscr{E}(q_I, v_h)| \leq Ch^{n-\frac{1}{2}} \|\tilde{u}\|_{n+1,\tilde{\Omega}} \|v_h\|_{0,\Gamma_h}, \tag{14}$$

where $q_I = \pi_h(q \circ \phi_h)$ is a standard S_h-interpolant of q and $\tilde{u} \in H^{n+1}(\tilde{\Omega})$ is an extension of the weak solution u to $\tilde{\Omega}$.

Proof. The estimate (13) is a consequence of the proof of Theorem 4 by Ciarlet and Raviart (1972b) (see also Theorems 4.1.5 and 4.4.5 by Ciarlet (1978)).Similar argument can be applied to the estimate (14). If we denote $\eta = q_I v_h$, the error of the quadrature formula is

$$|\mathscr{E}(q_I, v_h)| = |\mathscr{E}(\eta, 1)| \leq \sum_{K \in B_h} |\mathscr{E}_K(\eta)| = \sum_{K \in B_h} |\hat{\mathscr{E}}(|D\mathscr{F}_K|\hat{\eta})|, \tag{15}$$

where $\mathscr{F}_K$ is the restriction $F_K|_{K_3}$. For the seminorms of this mapping it follows

$$|\{|D\mathscr{F}_K|\}|_{i,\infty,\hat{T}} \begin{cases} \leq \hat{C}h^{i+1} & \text{if } \; i \leq n, \\ = 0 & \text{if } \; i > n. \end{cases}$$

The linear functional $\hat{\mathscr{E}}(\hat{\varphi})$ is bounded for $\hat{\varphi} \in W^{2n,1}(\hat{T})$. It vanishes according to equation (11) for polynomials of degree $2n - 1$. Thus, by the Bramble-Hilbert lemma (Ciarlet, 1978)

$$|\hat{\mathscr{E}}(|D\mathscr{F}_K|\hat{\eta})| \leq \hat{C}\||D\mathscr{F}_K|\hat{\eta}|_{2n,\hat{T}} \leq \hat{C}\sum_{\substack{i+j=2n,\\ i \leq n}} |\{|D\mathscr{F}_K|\}|_{i,\infty,\hat{T}} |\hat{\eta}|_{j,\hat{T}} \leq \hat{C}\sum_{\substack{i+j=2n,\\ i \leq n}} h^{i+1} |\hat{\eta}|_{j,\hat{T}}. \tag{16}$$

In the case considered we have $\hat{q}_I$, $\hat{v}_K \in P_n(\hat{T})$ $\left(v_h|_{T_K} = \hat{v}_K \circ \chi_K^{-1}\right)$. From the Leibniz rule and the inverse inequalities (Ciarlet, 1978, pp. 140-3) it follows that

$$|\hat{\eta}|_{j,\hat{T}} \leq \hat{C}\sum_{\substack{l+m=j,\\ n \leq j \leq 2n}} |\hat{q}_I|_{l,\hat{T}} |\hat{v}_K|_{m,\hat{T}} \leq Ch^{j-n-\frac{3}{2}} \|q_I\|_{n-\frac{1}{2},T_K} \|v_h\|_{0,T_k}.$$

This inequality along with equations (15) and (16) lead to the estimate

$$|\mathscr{E}(q_I, v_h)| \leq Ch^{n-\frac{1}{2}} \left(\sum_{K \in B_h} \|q_I\|^2_{n-\frac{1}{2},T_K} \right)^{1/2} \|v_h\|_{0,\Gamma_h}.$$

Denote a standard $\mathscr{V}_h$-interpolant of u by u_I. Using the imbedding theorems and the fact that $u_{I|K} \in P_n(K)$ we have

$$|\mathscr{E}(q_I, v_h)| \leq Ch^{n-\frac{1}{2}}\left(\sum_{K\in B_h}\|u_I\|^2_{n,K}\right)^{1/2}\|v_h\|_{0,\Gamma_h}.$$

First, we prove that if the triangulations τ_h satisfy the conditions **T1-T5**, then for every $\tilde{v} \in H^{n+1}(\tilde{\Omega}), \tilde{v}|_\Omega = v$

$$\left(\sum_{K\in\tau_h}\|v_I\|^2_{m,K}\right)^{1/2} \leq C\|\tilde{v}\|_{n+1,\tilde{\Omega}}, \quad m = 0, 1, 2, \ldots, n+1. \tag{17}$$

Indeed, the cases $m = 0$ and $m = 1$ are obvious. For $m = 2, \ldots, n$ we can write

$$\|v_I\|_{m,K} \leq \|\tilde{v} - v_I\|_{m,K} + \|\tilde{v}\|_{m,K}.$$

Assume that the isoparametric finite elements $(\hat{K}, \hat{P}, \hat{\Sigma})$ are optimal (Ciarlet and Raviart, 1972a, b), i.e. the family they make satisfies the standard interpolation estimates for any function $\tilde{v} \in H^{n+1}(K)$. Then

$$\left(\sum_{K\in\tau_h}\|v_I\|^2_{m,K}\right)^{1/2} \leq Ch^{n+1-m}\|\tilde{v}\|_{n+1,\Omega_h} + \|\tilde{v}\|_{n+1,\Omega_h} \leq C\|\tilde{v}\|_{n+1,\Omega_h},$$

$\forall\tilde{v} \in H^{n+1}(\tilde{\Omega})$. Using equation (17) for $m = n$, we obtain the estimate (14). □

The next lemma is an important tool in isoparametric technique.

Lemma 3. The following estimate holds

$$\|v_h\|_{1,\Omega_h} \leq C\|v_h \circ \Phi_h^{-1}\|_{1,\Omega}, \quad \forall v_h \in \mathscr{V}_h. \tag{18}$$

Proof. The key point of the proof is that (Lenoir, 1986)

$$\left|J\left(\Phi_h^{-1}\right)\right|_{0,\infty,\Omega} = O(1) \text{ and } |\Phi_h|_{1,\infty,\Omega_h} = O(1). \tag{19}$$

First we prove that

$$\|v_h\|_{0,\Omega_h} \leq C\|v_h \circ \Phi_h^{-1}\|_{0,\Omega}, \quad \forall v_h \in \mathscr{V}_h. \tag{20}$$

Changing the variables, we obtain

$$\begin{aligned}\|v_h\|^2_{0,\Omega_h} &= \int_{\Omega_h} v_h^2\,\mathrm{d}x \\ &= \int_\Omega \left(v_h \circ \Phi_h^{-1}\right)^2 J\left(\Phi_h^{-1}\right)\mathrm{d}x \\ &\leq \left|J\left(\Phi_h^{-1}\right)\right|_{0,\infty,\Omega}\left\|v_h \circ \Phi_h^{-1}\right\|^2_{0,\Omega}.\end{aligned}$$

It remains to apply equation (19) for completing the proof of equation (20).

As a second step we prove that

$$|v_h|_{1,\Omega_h} \leq C|v_h \circ \Phi_h^{-1}|_{1,\Omega}, \quad \forall v_h \in \mathscr{V}_h.$$

Changing the variables once more we get

$$\begin{aligned}|v_h|^2_{1,\Omega_h} &= \int_{\Omega_h} \nabla v_h \cdot \nabla v_h \,\mathrm{d}x \\ &= \int_{\Omega} D\big(v_h \circ \Phi_h^{-1}\big) D\Phi_h \cdot D\big(v_h \circ \Phi_h^{-1}\big) D\Phi_h J\big(\Phi_h^{-1}\big)\,\mathrm{d}x \\ &\leq C|\Phi_h|^2_{1,\infty,\Omega_h}\big|J\big(\Phi_h^{-1}\big)\big|_{0,\infty,\Omega_h}\big|v_h \circ \Phi_h^{-1}\big|_{1,\Omega}.\end{aligned}$$

Taking into account that equation (19) holds we get equation (18). □

Lemma 4. Let u, u_h^{*}, and u_h, be the solutions of the problems $(\mathscr{P}_W)$, $(\tilde{P}_h)$ and (P_h), respectively. Let for some integer $n \geq 2$, the quadrature formula (4) satisfy the assumption (12) of Lemma 2, and hypotheses **C1**-**C3**, **T4** and **T5** hold. Then

$$\Big|A_h\Big(u_h^{*}, v_h\Big) - a_h(u_h, v_h)\Big| \leq Ch^n(\|u\|_{n+1,\Omega} + \|f\|_{n,\Omega})|v_h|_{1,\Omega_h}, \quad \forall v_h \in \mathscr{V}_h. \tag{21}$$

Proof. Adding and subtracting some terms in the left hand side of equation (21) we obtain

$$\begin{aligned}A_h\Big(u_h^{*}, v_h\Big) - a_h(u_h, v_h) = &\Big\{A_h\Big(u_h^{*}, v_h\Big) - A_h(u \circ \Phi_h, v_h)\Big\} \\ &+ \{A_h(u \circ \Phi_h, v_h) - A_h(u_h, v_h)\} \\ &+ \{A_h(u_h, v_h) - a_h(u_h, v_h)\} \underset{=}{def}\, \mathfrak{A}_1 + \mathfrak{A}_2 + \mathfrak{A}_3.\end{aligned} \tag{22}$$

We estimate each term in the right hand side of equation (22).

Using Lemma 3 we have

$$|\mathfrak{A}_1| \leq C\Big\|u_h^{*} - u \circ \Phi_h\Big\|_{1,\Omega_h}|v_h|_{1,\Omega_h} \leq C\Big\|u - u_h^{*} \circ \Phi_h^{-1}\Big\|_{1,\Omega}|v_h|_{1,\Omega_h}.$$

Applying Theorem 3 by Lenoir (1986) we obtain

$$|\mathfrak{A}_1| \leq Ch^n\|u\|_{n+1,\Omega}|v_h|_{1,\Omega_h}, \quad \forall v_h \in \mathscr{V}_h. \tag{23}$$

Let $\tilde{u} \in H^{n+1}(\tilde{\Omega})$ and $\tilde{f} \in H^n(\tilde{\Omega})$ be sufficiently smooth extensions of the solution u of the problem $(\mathscr{P}_W)$ and the right hand side f of the problem $(\mathscr{P})$ to $\tilde{\Omega}$. Then (Vanmaele and Ženišek, 1993)

$$|\tilde{f}|_{n,\tilde{\Omega}} \leq C\|f\|_{n,\Omega}, \;\; \|\tilde{u}\|_{n+1,\tilde{\Omega}} \leq C\|u\|_{n+1,\Omega}. \tag{24}$$

Using the triangle inequality for any $v_h \in \mathscr{V}_h$ we obtain (Ciarlet and Raviart, 1972b)

$$
\begin{aligned}
|\mathfrak{A}_2| &\le C\|u\circ\Phi_h - u_h\|_{1,\Omega_h}|v_h|_{1,\Omega_h}\\
&\le C(\|u\circ\Phi_h - \tilde{u}\|_{1,\tilde{\Omega}} + \|\tilde{u} - u_h\|_{1,\tilde{\Omega}})|v_h|_{1,\Omega_h}\\
&\le C(\|D\Phi_h - I\|_{0,\infty,\Omega_h}\|u\|_{1,\Omega} + \|\tilde{u} - u_h\|_{1,\tilde{\Omega}})|v_h|_{1,\Omega_h},\\
&\le Ch^n\left(\|\tilde{u}\|_{n+1,\tilde{\Omega}} + \sum_{i,j=1}^{2}\|\tilde{a}_{ij}\|_{n,\infty,\tilde{\Omega}}\|\tilde{u}\|_{n+1,\tilde{\Omega}} + \|\tilde{f}\|_{n,\tilde{\Omega}}\right)|v|_{1,\Omega_h},
\end{aligned}
$$

because of the estimate $\|D\Phi_h - I\|_{0,\infty,\Omega_h} = O(h^n)$ (Lenoir, 1986). It follows

$$
|\mathfrak{A}_2| \le Ch^n(\|u\|_{n+1,\Omega} + \|f\|_{n,\Omega_h})|v_h|_{1,\Omega_h}, \quad \forall v_h \in \mathscr{V}_h \tag{25}
$$

from equation (24).

By analogy with the inequality (13) (see also Ciarlet and Raviart (1972b) and Theorem 4.4.4 by Ciarlet (1978)) it is easily seen that

$$
\begin{aligned}
|\mathfrak{A}_3| &\le \sum_{K\in\tau_h}\left|E_K\left(\sum_{i,j=1}^{2}\tilde{a}_{ij}\frac{\partial u_h}{\partial x_i}\frac{\partial v_h}{\partial x_j}\right)\right|\\
&\le Ch^n\left(\sum_{K\in\tau_h}\|u_h\|^2_{n+1,K}\right)^{1/2}|v_h|_{1,\Omega_h}, \quad \forall v_h \in \mathscr{V}_h.
\end{aligned}
\tag{26}
$$

Let us continue with the application of the inequalities (17), (24) and the inverse inequality

$$
\begin{aligned}
\left(\sum_{K\in\tau_h}\|u_h\|^2_{n+1,K}\right)^{1/2} &\le \left(\sum_{K\in\tau_h}\|u_h - u_I\|^2_{n+1,K}\right)^{1/2} + \left(\sum_{K\in\tau_h}\|u_I\|^2_{n+1,K}\right)^{1/2}\\
&\le Ch^{-n}\left(\sum_{K\in\tau_h}\left\{\|\tilde{u} - u_h\|^2_{1,K} + \|\tilde{u} - u_I\|^2_{1,K}\right\}\right)^{1/2}\\
&\quad + C\|\tilde{u}\|_{n+1,\tilde{\Omega}} \quad \le C\|\tilde{u}\|_{n+1,\tilde{\Omega}} \le C\|u\|_{n+1,\Omega}.
\end{aligned}
$$

Substituting this inequality in equation (26), we obtain

$$
|\mathfrak{A}_3| \le Ch^n\|u\|_{n+1,\Omega}|v_h|_{1,\Omega_h}, \quad \forall v_h \in \mathscr{V}_h. \tag{27}
$$

Thus the equality (22) and inequalities (23), (25) and (27) prove the estimate (21). □

Introduce the following scalar product

$$
< q_h, v_h >_h = \int_{\Gamma_h} q_h v_h \,\mathrm{d}s, \;\; q_h \in S_h, \;\; v_h \in \mathscr{V}_h.
$$

Lemma 5. The following interpolation error estimates hold

$$\|q\circ\phi_h - q_I\|_{0,\Gamma_h} \leq Ch^{n-\frac{1}{2}}\|u\|_{n+1,\Omega}, \tag{28}$$

$$< q\circ\phi_h - q_I, v_h >_h \leq Ch^{n-\frac{1}{2}}\|u\|_{n+1,\Omega}\|v_h\|_{0,\Gamma_h}, \quad \forall v_h \in \mathscr{V}_h. \tag{29}$$

Proof. Using the Bramble-Hilbert lemma we get

$$\|q\circ\phi_h - q_I\|_{0,\Gamma_h} \leq Ch^{n-\frac{1}{2}}\|q\|_{n-\frac{1}{2},\Gamma}.$$

Then we have from the imbedding theorems (Adams, 1975)

$$\|q\|_{n-\frac{1}{2},\Gamma} \leq \|u\|_{n+1,\Omega},$$

which proves the required inequality (28). The estimate (29) is a consequence of

$$< q\circ\phi_h - q_I, v_h >_h \leq \|q\circ\phi_h - q_I\|_{0,\Gamma_h}\|v_h\|_{0,\Gamma_h}. \qquad \square$$

Introduce the consistent case flux problem $(\tilde{\mathscr{F}}_h)$ corresponding to $(\tilde{\mathscr{P}}_h)$

$$\tilde{\mathscr{F}}_h \begin{cases} \text{find } \overset{*}{q_h} \in \mathscr{V}_h \text{ such that} \\ - < \overset{*}{q_h}, v >_h = A_h\left(\overset{*}{u_h}, v\right) - (f_h, v), \quad \forall v \in \mathscr{V}_h \end{cases},$$

where $f_h = \Pi_h(f\circ\Phi_h)$.

Lemma 6. Let $\overset{*}{q_h}$ be the solution of the problem $(\tilde{\mathscr{F}}_h)$ and q_I be the interpolant of the weak solution q of equation (1). Suppose that the conditions of Lemma 4 are fulfilled. Then the following estimate holds

$$< \overset{*}{q_h} - q_I, v_h >_h \leq C\left(h^n\|v_h\|_{1,\Omega_h} + h^{n-\frac{1}{2}}\|v_h\|_{0,\Gamma_h}\right)\|u\|_{n+1,\Omega}, \quad \forall v_h \in \mathscr{V}_h. \tag{30}$$

Proof. By the triangle inequality, we have

$$\left|< \overset{*}{q_h} - q_I, v_h >_h\right| \leq \left|< \overset{*}{q_h} - q\circ\phi_h, v_h >_h\right| + |< q\circ\phi_h - q_I, v_h >_h|. \tag{31}$$

The second term in the right hand side of equation (31) can be estimated by Lemma 5 – the inequality (29). It is obvious that $\left(\left(v_h\circ\Phi_h^{-1}\right)|_\Gamma\right)(x) = \left(v_h\circ\phi_h^{-1}\right)(x)$, $x \in \Gamma$.Then we obtain

$$\left|< \overset{*}{q_h} - q\circ\phi_h, v_h >_h\right| \leq \left|< \overset{*}{q_h}, v_h >_h - \langle q, v_h\circ\phi_h^{-1}\rangle\right| + \left|\langle q, v_h\circ\phi_h^{-1}\rangle - < q\circ\phi_h, v_h >_h\right|. \tag{32}$$

Using the theory of approximation of the boundary condition presented by Lenoir (1986), for the second term of the right hand side of equation (32), we get

$$\left|\langle q, v_h\circ\phi_h^{-1}\rangle - < q\circ\phi_h, v_h >_h\right| \leq Ch^{n-\frac{1}{2}}\|q\|_{n-\frac{1}{2},\Gamma}\|v_h\|_{0,\Gamma_h} \leq Ch^{n-\frac{1}{2}}\|u\|_{n+1,\Omega}\|v_h\|_{0,\Gamma_h}. \tag{33}$$

Let us consider the first term of the right hand side of the inequality (32)

$$\left| < \overset{*}{q}_h, v_h >_h - \langle q, v_h \circ \phi_h^{-1} \rangle \right| \leq \left| \int_{\Omega_h} f_h v_h \, \mathrm{d}x - \int_\Omega f\left(v_h \circ \Phi_h^{-1}\right) \mathrm{d}x \right| + \left| a\left(u, v_h \circ \Phi_h^{-1}\right) - A_h\left(\overset{*}{u}_h, v_h\right) \right|. \tag{34}$$

The approximation of the domain gives the estimate (see Lemma 8 by Lenoir (1986))

$$\left| \int_{\Omega_h} f_h v_h \, \mathrm{d}x - \int_\Omega f\left(v_h \circ \Phi_h^{-1}\right) \mathrm{d}x \right| \leq C h^n \|f\|_{n,\Omega} \|v_h\|_{0,\Omega_h}. \tag{35}$$

It remains to estimate the last term in equation (34). For the sake of simplicity and for notational convenience we consider the *a*-forms with constant coefficients. A more detailed but simple analysis gives the same order of convergence when the bilinear forms have variable coefficients and the hypothesis **C3** holds.

Estimate

$$\begin{aligned}
&\left| a\left(u, v_h \circ \Phi_h^{-1}\right) - A_h\left(\overset{*}{u}_h, v_h\right) \right| \\
&\quad \leq \left| \int_{\Omega_h} \nabla(u \circ \Phi_h) D\Phi_h^{-1} \cdot \nabla v_h D\Phi_h^{-1} J(\Phi_h) \, \mathrm{d}x - \int_{\Omega_h} \nabla(u \circ \Phi_h) D\Phi_h^{-1} \cdot \nabla v_h J(\Phi_h) \, \mathrm{d}x \right| \\
&\qquad + \left| \int_{\Omega_h} \nabla(u \circ \Phi_h) D\Phi_h^{-1} \cdot \nabla v_h J(\Phi_h) \, \mathrm{d}x - \int_{\Omega_h} \nabla(u \circ \Phi_h) \cdot \nabla v_h J(\Phi_h) \, \mathrm{d}x \right| \\
&\qquad + \left| \int_{\Omega_h} \nabla(u \circ \Phi_h) \cdot \nabla v_h J(\Phi_h) \, \mathrm{d}x - \int_{\Omega_h} \nabla(u \circ \Phi_h) \cdot \nabla v_h \, \mathrm{d}x \right| \\
&\qquad + \left| \int_{\Omega_h} \nabla(u \circ \Phi_h) \cdot \nabla v_h \, \mathrm{d}x - \int_{\Omega_h} \nabla \overset{*}{u}_h \cdot \nabla v_h \, \mathrm{d}x \right| \\
&\quad \leq \left| D\Phi_h^{-1} - I \right|_{0,\infty,\Omega} \left(\left| D\Phi_h^{-1} \right|_{0,\infty,\Omega} + 1 \right) |J(\Phi_h)|_{0,\infty,\Omega_h} \|u\|_{1,\Omega} |v_h|_{1,\Omega_h} \\
&\qquad + |J(\Phi_h) - 1|_{0,\infty,\Omega_h} \|u\|_{1,\Omega} |v_h|_{1,\Omega_h} + \left\| u \circ \Phi_h - \overset{*}{u}_h \right\|_{1,\Omega_h} |v_h|_{1,\Omega_h}.
\end{aligned}$$

Taking into account the relations (Lenoir, 1986):

$$\left| D\Phi_h^{-1} - I \right|_{0,\infty,\Omega} = O(h^n), \quad |J(\Phi_h) - 1|_{0,\infty,\Omega_h} = O(h^n),$$

$$\left| \Phi_h^{-1} \right|_{1,\infty,\Omega} = O(1), \quad |J(\Phi_h)|_{0,\infty,\Omega_h} = O(1),$$

as well as the error estimate (see Theorem 3 by Lenoir (1986))

$$\left| u \circ \Phi_h - \overset{*}{u}_h \right|_{1,\Omega_h} \leq C \left| u - \overset{*}{u}_h \circ \Phi_h^{-1} \right|_{1,\Omega} \leq C h^n \|u\|_{n+1,\Omega},$$

we obtain

$$\left| a\left(u, v_h \circ \Phi_h^{-1}\right) - A_h\left(\overset{*}{u}_h, v_h\right) \right| \leq C h^n \|u\|_{n+1,\Omega} \|v_h\|_{1,\Omega_h}, \quad \forall v_h \in \mathscr{V}_h. \tag{36}$$

Substituting the inequalities (35) and (36) in equation (34) we find

$$\left| < \overset{*}{q_h}, v_h >_h - \langle q, v_h \circ \phi_h^{-1} \rangle \right| \leq Ch^n \|u\|_{n+1,\Omega} \|v_h\|_{1,\Omega_h}.$$

Combining the latter estimate and inequalities (31)-(33), we prove equation (30). □

Lemma 7. For any $v_h \in \mathscr{V}_h$, there exists an element $\tilde{v}_h \in \mathscr{V}_h$, such that $\tilde{v}_h = v_h$ on Γ_h and

$$\|\tilde{v}_h\|_{1,\Omega_h} \leq Ch^{-\frac{1}{2}} \|v_h\|_{0,\Gamma_h}. \tag{37}$$

Proof. It is enough to construct $\tilde{v}_h \in \mathscr{V}_h$, such that $\tilde{v}_h = v_h$ on Γ_h and $\tilde{v}_h = 0$ for all internal nodes of the triangulation τ_h.

Let us introduce the Hilbert space (Adams, 1975)

$$H^{1/2}(\Gamma_h) = \{v \in L_2(\Gamma_h) : \exists u \in H^1(\Omega_h) \text{ such that } \mathrm{tr}(u) = v \text{ on } \Gamma_h\},$$

provided with the norm

$$\|v\|_{1/2,\Gamma_h} = \inf\{\|u\|_{1,\Omega_h} : u \in H^1(\Omega_h); \ \mathrm{tr}(u) = v \text{ on } \Gamma_h\}.$$

This space is dense in $L_2(\Gamma_h)$. Having in mind that $\tilde{v}_h = 0$ at any internal node of Ω_h and the space $\mathscr{V}_h$ consists of piecewise polynomials, it is evident that

$$\|\tilde{v}_h\|_{1,\Omega_h} \leq C \|\tilde{v}_h\|_{1/2,\Gamma_h}.$$

Using the inverse inequality

$$\|\tilde{v}_h\|_{1/2,\Gamma_h} \leq Ch^{-1/2} \|\tilde{v}_h\|_{0,\Gamma_h} \leq Ch^{-1/2} \|v_h\|_{0,\Gamma_h},$$

we obtain the estimate (37). □

The following theorem contains the main result concerning boundary-flux error estimates.

Theorem 1. Let the conditions of Lemmas 2 and 4 be fulfilled. Then the following error estimates hold

$$\|q_h - q_I\|_h \leq Ch^{n-\frac{1}{2}} (\|u\|_{n+1,\Omega} + \|f\|_{n,\Omega}), \tag{38}$$

$$\|q \circ \phi_h - q_h\|_{0,\Gamma_h} \leq Ch^{n-\frac{1}{2}} (\|u\|_{n+1,\Omega} + \|f\|_{n,\Omega}). \tag{39}$$

Proof. First we prove the inequality (38). For any function $v_h \in \mathscr{V}_h$ we have

$$\begin{aligned} \langle q_h - q_I, v_h \rangle_h &= \langle q_h, v_h \rangle_h - \langle q_I, v_h \rangle_h \\ &= (f_h, v_h)_h - a_h(u_h, v_h) - < \overset{*}{q_h}, v_h >_h + < q_h^{;}, v_h >_h - < q_I, v_h >_h + \mathscr{E}(q_I, v_h) \\ &= \mathscr{E}(q_I, v_h) - E(f_h, v_h) + \left\{ A_h\left(\overset{*}{u_h}, v_h\right) - a_h(u_h, v_h) \right\} + < \overset{*}{q_h} - q_I, v_h >_h. \end{aligned} \tag{40}$$

The four terms in the right hand side of equation (40) are estimated by equations (13), (14), (21) and (30), respectively. Consequently, it follows the inequality

$$\langle q_h - q_I, v_h\rangle_h \le Ch^{n-\frac{1}{2}}\|u\|_{n+1,\Omega}\|v_h\|_{0,\Gamma_h} + Ch^n(\|u\|_{n+1,\Omega} + \|f\|_{n,\Omega})\|v_h\|_{1,\Omega_h}.$$

For the second term in the right hand side we use the estimate (37) of Lemma 7 with $v_h = \tilde{v}_h$. We obtain that

$$\langle q_h - q_I, v_h\rangle_h \le Ch^{n-\frac{1}{2}}(\|u\|_{n+1,\Omega} + \|f\|_{n,\Omega})\|v_h\|_{0,\Gamma_h}$$

because $v_h|_{\Gamma_h} = \tilde{v}_h|_{\Gamma_h}$.

Finally, applying the norm equivalence for $\|v_h\|_{0,\Gamma_h}$ and $\|v_h\|_h$ and choosing $v_h|_{\Gamma_h} = q_h - q_I$ we derive the desired estimate (38).

We easily get the estimate (39). It is sufficiently to combine equation (38) with the estimate (28). Then

$$\begin{aligned}\|q\circ\phi_h - q_h\|_{0,\Gamma_h} &\le \|q\circ\phi_h - q_I\|_{0,\Gamma_h} + \|q_I - q_h\|_{0,\Gamma_h}\\ &\le \|q\circ\phi_h - q_I\|_{0,\Gamma_h} + C\|q_I - q_h\|_h\\ &\le Ch^{n-\frac{1}{2}}(\|u\|_{n+1,\Omega} + \|f\|_{n,\Omega}).\end{aligned}$$

□

5. Lumped mass boundary-flux

The lumped mass formulation is often the most practical form. For instance, when the heat and fluid flow application codes are concerned. Lumped flux formulations are examined by Carey *et al.* (1985), Lazarov and Pehlivanov (1989) and Pehlivanov *et al.* (1992). This approach is appropriate for various eigenvalue problems (Andreev and Todorov, 1999, 2004). In the case of lumped flux approach the integrals $< \cdot, \cdot >_h$ are evaluated using quadrature formula with quadrature nodes coincident with the element nodes. Applying such a type formula a diagonal coefficient matrix results and hence the flux q_h is determined explicitly.

The estimate (38) implies that

$$\|q_h - q_I\|_{0,\Gamma_h} = [(Q_h - Q)^t M(Q_h - Q)]^{1/2} = O\left(h^{n-\frac{1}{2}}\right),$$

where $Q = (q(a_1),\ q(a_2),\ \ldots, q(a_{d_h})),\ a_i \in \Gamma_h,\ i = 1, 2, \ldots, d_h,\ d_h = \dim(S_h)$ is the vector containing the exact values of the boundary-flux at the nodes on Γ and $Q_h = (q_h(a_1),\ q_h(a_2),\ \ldots, q_h(a_{d_h}))$ is the corresponding approximating vector obtained by the numerical integration. Here M is a mass matrix obtained by the inner product $< \cdot, \cdot >_h$ on the boundary Γ_h.

Problem (7) leads to the system of linear equations $MQ_h = F$ for the vector Q_h. The right hand side F is a known vector.

We lump the mass matrix into a diagonal form $\overline{M}$ and find the approximate values at the grid nodes from the system $\overline{M}Q_h = F$ with diagonal matrix $\overline{M}$. We confine to the case $n = 2$. Consider a quadrature formula giving diagonal matrix $\overline{M}$

$$\int_{\hat{T}} \hat{\varphi}(\hat{t})\,\mathrm{d}\hat{t} \sim \hat{\mathscr{I}}(\hat{\varphi}) = \frac{1}{6}\left(\hat{\varphi}(0) + 4\hat{\varphi}\left(\frac{1}{2}\right) + \hat{\varphi}(1)\right). \qquad (41)$$

The formula (41) represents the Simpson rule and it is exact for all polynomials belonging to $P_3(\hat{T})$. The quadrature formula over the finite element of reference $\hat{K}$ is

$$\int_{\hat{K}} \hat{\varphi}(\hat{x})\,\mathrm{d}\hat{x} \sim \hat{I}(\hat{\varphi}) = \frac{1}{120}\Big\{3[\hat{\varphi}(0,0)+\hat{\varphi}(1,0)+\hat{\varphi}(0,1)] + 8\Big[\hat{\varphi}\Big(\frac{1}{2},0\Big)+\hat{\varphi}\Big(\frac{1}{2},\frac{1}{2}\Big)+\hat{\varphi}\Big(0,\frac{1}{2}\Big)\Big] + 27\hat{\varphi}\Big(\frac{1}{3},\frac{1}{3}\Big)\Big\}. \tag{42}$$

This formula is also exact for all polynomials from P_3 (Ciarlet, 1978, p. 184). The numerical integration for the isoparametric case using equation (42) is analyzed by Andreev and Todorov (1999), where lumped mass approximation for second-order eigenvalue problem is considered.

Both formulas (41) and (42) are related to each other in the following sense:

- they are equivalent with respect to the precision, i.e. they are exact for all polynomials of degree three; and
- both quadrature formulas have common set of nodes on the hypotenuse of the finite element of reference.

Theorem 2 gives an error estimate in the lumped mass case.

Theorem 2. Let us keep the conditions of Theorem 1 and let the integrals in equation (7) be evaluated by quadrature formulas (41) and (42), respectively. Let also the hypotheses **T1-T5** and **C1-C3** are valid for $n = 2$ and the approximate boundary-flux is computed by the formula (7). Then

$$\|q - q_h\|_h \leq Ch^{\frac{3}{2}}(\|u\|_{3,\Omega} + \|f\|_{2,\Omega}), \tag{43}$$

Proof. It is easy to verify the validity of the hypotheses **Q1** and **Q2**. Consequently, we have the norm equivalence equation (8). We obtain

$$|E(f, v_h)| \leq Ch^2\|f\|_{2,\Omega}\|v_h\|_{1,\Omega_h},$$
$$|\mathscr{E}(q_I, v_h)| \leq Ch^{\frac{3}{2}}\|u\|_{3,\Omega}\|v_h\|_{0,\Gamma_h}, \quad \forall v_h \in \mathscr{V}_h.$$

from Lemma 2. Proceeding as in the case without lumping, we obtain the estimates (21), (28) and (30) with $n = 2$. But, as far as the lumped mass matrix case is concerned, the approximate inner product $\langle u, v\rangle_h$ deals only with the values of the functions u and v at the nodes of the quadrature formula, which at the same time are nodes of the finite elements. Then

$$\|q - q_h\|_h = \|q_I - q_h\|_h.$$

It remains to apply the estimate (38) with $n = 2$ in order to prove the theorem. □

6. Numerical tests

The point of discussion inhere is the boundary-flux calculation using numerical integrations. At the beginning we consider some algorithmic aspects.

Denote: the set of nodes of the triangulation τ_h by $\mathscr{N}_h$, the set of nodes belonging to Γ_h by $\mathscr{N}_{Bh}$. Define $\mathscr{N}_{Ih} = \mathscr{N}_h \backslash \Gamma_h$. Let $\{\varphi_i\}$, $i = 1, 2, \ldots, \mathrm{card}(\mathscr{N}_h)$ associated with $a_i \in \mathscr{N}_h$ be the nodal basis in $\mathbf{V}_h$. Define the spaces

$$\mathbf{V}_{Bh} = \mathrm{Span}\{\varphi_i\}_{i:a_i\in\mathcal{N}_{Bh}},$$
$$\mathbf{V}_{Ih} = \mathrm{Span}\{\varphi_i\}_{i:a_i\in\mathcal{N}_{Ih}}.$$

We shall use the vectors and matrices

$$\begin{aligned}
A &= (a_h(\varphi_i, \varphi_j))_{i,j:a_i,a_j\in\mathcal{N}_{Ih}},\\
\underline{u}_h &= (u_h(a_i))_{i:a_i\in\mathcal{N}_{Ih}},\\
\underline{q}_h &= (q_h(a_i))_{i:a_i\in\mathcal{N}_{Bh}},\\
C &= (a_h(\varphi_i, \varphi_j))_{i,j:a_i\in\mathcal{N}_{Ih},a_j\in\mathcal{N}_{Bh}},\\
M &= (\langle\varphi_i, \varphi_j\rangle_h)_{i,j:a_i,a_j\in\mathcal{N}_{Bh}},\\
\underline{F}_I &= ((f_h, \varphi_i),\ \ i : a_i \in \mathcal{N}_{Ih}),\\
\underline{F}_B &= ((f_h, \varphi_i),\ \ i : a_i \in \mathcal{N}_{Bh}).
\end{aligned}$$

Write a matrix form of problem (7)

$$-\begin{pmatrix} 0 & 0 \\ 0 & M \end{pmatrix}\begin{pmatrix} 0 \\ \underline{q}_h \end{pmatrix} = \begin{pmatrix} A & \mathrm{C} \\ C^t & 0 \end{pmatrix}\begin{pmatrix} \underline{u}_h \\ 0 \end{pmatrix} - \begin{pmatrix} \underline{F}_I \\ \underline{F}_B \end{pmatrix} \tag{44}$$

There are different approaches for solving problem (44). We choose the case with a lumped mass matrix. Initially, we eliminate $\underline{u}_h$ from equation (44) making complete Cholesky factorization of the matrix $A = L_A L_A^t$, where L_A is a lower triangle matrix. Then

$$-M\underline{q}_h = C^t L_A^{-t} L_A^{-1}\underline{F}_I - \underline{F}_B \underset{=}{\overset{def}{}} \underline{\tilde{F}}_B.$$

Since we consider the method with lumped mass, it is not necessary to construct the matrix M. It is enough to compile the vector $\underline{m}$ by $m_i = M_{ii}^{-1}$, $i = 1, 2, \ldots, d_h$. We compute the solution $\underline{q}_h$ from $\underline{q}_h = \underline{m}*\underline{\tilde{F}}_B$, where the multiplication "*" is defined by $\underline{v}*\underline{w} = (v_1w_1, v_2w_2, \ldots, v_nw_n) \in \mathbf{R}^n$, $\forall v, w \in \mathbf{R}^n$.

The above algorithm enables us to improve the accuracy of calculation and for decreasing the necessary computer resource. Andreev and Todorov (2004) showed that for the similar problem with different dimensions, the lumped mass technique gives the best results among many different approaches. Moreover, it is proved that the iterative solutions of systems such as equation (44) are stable, i.e. the used block splitting is applicable.

Continue with numerical examples. Consider a model problem

$$-\Delta u = 12xy \ \text{ in } \ \Omega, \ \ u = 0 \ \text{ on } \ \Gamma, \tag{45}$$

where Ω is a quarter of the unit disc and $\Gamma = \Gamma_1 \cup \Gamma_2 \cup \Gamma_3$ (Figure 1). The exact solution of problem (45) is

$$u(x, y) = xy - x^3y - xy^3.$$

Then the flux across the boundary is

$$q = \begin{cases} y - y^3, & 0 \le y \le 1 \text{ on } \Gamma_1, \\ x - x^3, & 0 \le x \le 1 \text{ on } \Gamma_2, \\ 2x\sqrt{1-x^2}, & 0 \le x \le 1 \text{ on } \Gamma_3. \end{cases}$$

Further, we illustrate the rate of convergence α arising from the considered finite element method. Using the approximate solution on two different meshes, we have

$$\begin{cases} \|q - q_h\|_h = Ch^\alpha, \\ \|q - q_{h/2}\|_{h/2} = C\left(\frac{h}{2}\right)^\alpha. \end{cases}$$

Then

$$\alpha = \frac{\log \dfrac{\|q - q_h\|_h}{\|q - q_{h/2}\|_{h/2}}}{\log 2}.$$

Example 1. For problem (45) we use the conditions of Theorem 2. The triangulations consist of the 6-node isoparametric elements. The quadrature formula (41) gives the lumped mass matrix and the integrals over the elements are evaluated by the 7-node isoparametric integration equation (42) (Vanmaele and Ženišek, 1993). An initial triangulation for solving problem (45) is shown in Figure 1. □

Example 2. Here we find a solution of problem (45) on the basis of the 10-node isoparametric elements. Keeping the optimal rate of convergence for the approximate boundary flux, we make use of the following Radon quadrature formula (Hammer *et al.*, 1956)

$$\int_{\hat{K}} \hat{\varphi}(\hat{x})\,\mathrm{d}\hat{x} \sim I_{\hat{K}}(\hat{\varphi}) = \frac{9}{80}\varphi(\hat{\omega}_7) + \frac{155+\sqrt{15}}{2,400}[\varphi(\hat{\omega}_1) + \varphi(\hat{\omega}_2) + \varphi(\hat{\omega}_6)] + \frac{155-\sqrt{15}}{2,400}[\varphi(\hat{\omega}_3) + \varphi(\hat{\omega}_4) + \varphi(\hat{\omega}_5)], \quad \hat{\varphi} \in C(\hat{K}) \tag{46}$$

where

$$\hat{\omega}_1 = (\hat{\zeta}_3, \hat{\zeta}_1), \quad \hat{\omega}_2 = (\hat{\zeta}_1, \hat{\zeta}_3), \quad \hat{\omega}_3 = (\hat{\zeta}_4, \hat{\zeta}_2), \quad \hat{\omega}_4 = (\hat{\zeta}_2, \hat{\zeta}_4),$$

$$\hat{\omega}_5 = (\hat{\zeta}_2, \hat{\zeta}_2), \quad \hat{\omega}_6 = (\hat{\zeta}_3, \hat{\zeta}_3), \quad \hat{\omega}_7 = (\hat{\zeta}_5, \hat{\zeta}_5),$$

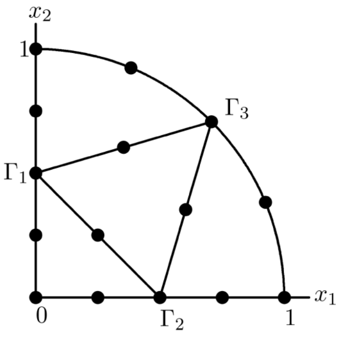

Figure 1.
An initial decomposition of the domain Ω by 6-node finite elements

and

$$\hat{\zeta}_1 = \frac{9-2\sqrt{15}}{21},\ \ \hat{\zeta}_2 = \frac{6-\sqrt{15}}{21},\ \ \hat{\zeta}_3 = \frac{6+\sqrt{15}}{21},\ \ \hat{\zeta}_4 = \frac{9+2\sqrt{15}}{21},\ \ \hat{\zeta}_5 = \frac{1}{3}.$$

The quadrature formula (46) is exact for all polynomials of degree five.

For computing the approximate scalar products on the boundary we use a Gauss quadrature formula exact for all polynomials of degree five

$$\int_{\hat{T}} \hat{\varphi}(\hat{x})\,\mathrm{d}\hat{x} \sim I_{\hat{T}}(\hat{\varphi}) = \frac{1}{18}\left(8\varphi\left(\frac{1}{2}\right) + 5\varphi\left(\frac{1-\sqrt{\frac{3}{5}}}{2}\right) + 5\varphi\left(\frac{1+\sqrt{\frac{3}{5}}}{2}\right)\right),$$

where $\hat{\varphi} \in C(\hat{T})$.

The computational process in Example 2 is more complicated because the mass matrix is not diagonal. Therefore, we make complete Cholesky factorization for inversion of this matrix. In this case we need greater computer resources. □

The results obtained in both examples are presented in the comparative Table I. The confirmation of the theoretical achievements is performed (estimates (39) and (43)).

7. Concluding remarks

The obtained convergence results for the isoparametric boundary flux enable us to conclude that:

- We have proved the optimal order of convergence subject to the hypotheses that are presupposed. These hypotheses are not too restrictive regarding the isoparametric approach.
- The precision of the quadrature formulas presented in equations (11) and (12) is crucial for proving the order of convergence. In this respect, the one-dimensional Lobatto quadrature formulas could be mentioned.
- Although the investigations on the lumped mass approximation are performed for quadratic triangular finite elements, they could be applied to more general elliptic systems and other types of finite elements. It is necessary to combine the appropriate quadrature formula, satisfying the conditions **Q1** and **Q2** with nodes coinciding with the nodes of the finite element.

	$\|\|q - q_h\|\|_h$	
	6-node finite elements	10-node finite elements
4 elements	0.8351214066	0.0609879166
16 elements	0.2811585814	0.0106510992
α	1.57060	2.51752
16 elements	0.2811585814	0.0106510992
64 elements	0.0950408902	0.0018609964
α	1.56476	2.51686
64 elements	0.0950408902	0.0018609964
256 elements	0.0334027457	0.0003253089
α	1.50858	2.51619
256 elements	0.0334027457	0.0003253089
1,024 elements	0.0118019900	0.0000571134
α	1.50093	2.50991

Table I. Asymptotic rate of convergence α obtained in the examples

- The method hereby presented, could be used in the cases when two subdomains have curved interface.

References

Adams, R.A. (1975), *Sobolev Spaces*, Academic Press, New York, NY.

Andreev, A.B. and Todorov, T.D. (1999), "Lumped mass approximation for an isoparametric finite element eigenvalue problem", *Sib. J. Numer. Math.*, Vol. 2 No. 4, pp. 295-308.

Andreev, A.B. and Todorov, T.D. (2004), "An isoparametric finite element approximation of a Steklov eigenvalue problem", *IMA J. Numer. Anal.*, Vol. 24, pp. 309-22.

Barret, J.W. and Elliot, C.M. (1987), "Total flux estimates for solutions of elliptic equations", *IMA J. Numer. Anal.*, Vol. 7, pp. 129-48.

Carey, G.F. (1982), "Derivative calculation from finite element solutions", *J. Comput. Meth. Appl. Mech. Eng.*, Vol. 35, pp. 1-14.

Carey, G.F. (2002), "Some further properties of the superconvergent flux projection", *CNME*, Vol. 18 No. 4, pp. 241-50.

Carey, G.F., Chow, S.S. and Seager, M.R. (1985), "Approximate boundary flux calculations", *J. Comput. Meth. Appl. Mech. Eng.*, Vol. 50, pp. 107-20.

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

Ciarlet, P.G. and Raviart, P.A. (1972a), "Interpolation theory over curved elements, with applications to finite element methods", *Comp. Meth. Appl. Mech. Eng.*, Vol. 1, pp. 217-49.

Ciarlet, P.G. and Raviart, P.A. (1972b), "The combined effect of curved boundaries and numerical integration in isoparametric finite element method", in Aziz, A.K. (Ed.), *Math. Foundation of the FEM with Applications to PDE*, Academic Press, New York, NY, pp. 409-74.

Douglas, J., Dupont, T. and Wheeler, M.F. (1974), "A Galerkin procedure for approximating the flux on the boundary for elliptic and parabolic boundary value problems", *RAIRO, Modelization Math. Anal. Numer.*, Vol. 8, pp. 47-59.

Hammer, P.C., Marlowe, O.J. and Stroud, A.H. (1956), "Numerical integration over simplexes and cones", *Math. Tables Aids Comput.*, Vol. 10, pp. 130-7.

Lazarov, R.D. and Pehlivanov, A.I. (1989), "Local superconvergence analysis of the approximate boundary flux calculations", *Proc. Conf. EQUADIFF'7, Prague, Teubner – Texte zur Mathematik, BSB Teubner, Leipzig*, pp. 275-9, d. 118.

Lenoir, M. (1986), "Optimal isoparametric finite elements and error estimates for domains involving curved boundaries", *SIAM J. Numer. Anal.*, Vol. 23 No. 3, pp. 562-80.

Pehlivanov, A.I., Lazarov, R.D., Carey, G.F. and Chow, S.S. (1992), "Superconvergence analysis of approximate boundary flux calculations", *Numer. Math.*, Vol. 63, pp. 483-501.

Vanmaele, M. and Ženišek, A. (1993), "External finite element approximations of eigenvalue problems", *Math. Model Numer. Anal.*, Vol. 27, pp. 565-89.

Wheeler, J.A. (1973), "Simulation of heat transfer from a warm pipeline buried permafrost", paper presented at 74th National Meeting of the American Institute of Chemical Engineers, New Orleans, LA.

Corresponding author
Todor D. Todorov can be contacted at paralaxview@yahoo.com