A unified approach to a posteriori error estimation using element residual methods

Mark Ainsworth* and J. Tinsley Oden
Texas Institute for Computational Mechanics. The University of Texas at Austin. Austin, TX 78712, USA

Received September 11, 1991/Revised version received June 24, 1992

Summary. This paper deals with the problem of obtaining numerical estimates of
the accuracy of approximations to solutions of elliptic partial differential equations. It is shown that, by solving appropriate local residual type problems, one can
obtain upper bounds on the error in the energy norm. Moreover, in the special case
of adaptive \( h-p \) finite element analysis, the estimator will also give a realistic
estimate of the error. A key feature of this is the development of a systematic
approach to the determination of boundary conditions for the local problems. The
work extends and combines several existing methods to the case of full \( h-p \) finite
element approximation on possibly irregular meshes with elements of non-uniform
degree. As a special case, the analysis proves a conjecture made by Bank and
Weiser [Some A Posteriori Error Estimators for Elliptic Partial Differential

Mathematics Subject Classification (1991): 65N30

1. Introduction

In this work we address the problem of computing \( a \) posteriori error estimates for
approximations to elliptic boundary value problems. Although we have in mind
adaptive \( h-p \) finite element computations, the analysis also includes various other
types of approximation.

The error estimates are based on local residual problems similar in type to those
discussed in [5, 6, 11, 12]. There are, however, significant differences in the approach.

Although the boundary value problem which we are approximating may be
associated with an operator of the form

\[
L u = - V \cdot (a V u) + cu
\]
the local residual problem will always be Poisson's equation. This means that the error estimation analysis can be done independently of the analysis of the particular form of the operator \( L \). More importantly, error analysis routines can be developed which exploit properties of the Laplace operator to develop a very efficient routine for solving the local problems. The first main result is to show that the error estimator always overestimates the true error. Essentially, we do not need to make any further regularity assumptions on the true solution other than \( u \in H^1(\Omega) \).

In order to show that the estimator does not give an unduly pessimistic estimate, further assumptions are made on the source of the approximation \( u_h \). In particular, we assume \( u_h \) is obtained by an \( h-p \) adaptive finite element computation, possibly on irregular meshes with elements of differing shapes and non-uniform polynomial degree. A preliminary analysis reveals that the estimate can become very pessimistic unless the boundary conditions for the local problems are chosen carefully.

The question of boundary conditions is examined in detail with the result that a scheme proposed by Bank and Weiser [5] for piecewise affine approximation on triangular elements is extended to the case of full \( h-p \) approximation on irregular meshes. It is of interest to note that in determining the boundary conditions one works on the same element patches which occur in the related works of Babuška et al. [3,4]. As special cases of our results, we obtain the result conjectured in Bank and Weiser [5] that a certain error estimator always overestimates the true error, and provide theoretical support for the heuristic results of Kelly [9].

2. Notations and preliminaries

Let \( \Omega \) denote an open bounded Lipschitzian domain in \( \mathbb{R}^2 \) with a piecewise smooth boundary \( \partial \Omega \). The boundary consists of a finite number of smooth arcs meeting with internal angle \( \theta \in (0, 2\pi) \).

The Sobolev space \( H^m(\Omega) \), \( m \in \mathbb{Z}^+ \), is a Hilbert space defined as the completion of \( C^\infty(\Omega) \) in the Sobolev norm

\[
||u||_{m, \Omega} = \left\{ \sum_{|n| \leq m} \int_\Omega (|D^s u|^2 \, dx)^{1/2} \right\}^{1/2}
\]

where \( x = (x_1, x_2) \), \( x_i \in \mathbb{Z}^+ \), \( |x| = x_1 + x_2 \) and

\[
D^s u = \frac{\partial^{[x]} u}{\partial x_1^{x_1} \partial x_2^{x_2}}
\]

is the distributional derivative. \( H^m(\Omega) \) is equipped with the inner product

\[
(u, v)_{m, \Omega} = \sum_{|n| \leq m} \int_\Omega D^s u \cdot D^s v \, dx
\]

We use the notation \( H^0(\Omega) = L_2(\Omega) \) in the case \( m = 0 \). Let \( \mathcal{P} \) be a partition of \( \Omega \) into a collection of \( N = N(\mathcal{P}) \) subdomains \( \Omega_k \) with boundaries \( \partial \Omega_k \), \( 1 \leq K \leq N \), such that

- (i) \( N(\mathcal{P}) < \infty \)
- (ii) \( \Omega = \bigcup_{K=1}^N \Omega_k \), \( \Omega_k \cap \Omega_L = \emptyset \), \( K \neq L \)
(iii) \( \Omega_K \) are Lipschitzian with piecewise smooth boundaries \( \partial \Omega_K \).
(iv) \( \Gamma_{KL} = \partial \Omega_K \cap \partial \Omega_L \). \( 1 \leq K, L \leq N \) are sets consisting of a finite number \( p(K, L) \) of components such that 
\[
\Gamma_{KL} = \bigcup_{M=1}^{p(K, L)} \Gamma_{KL}^M \cup S_{KL}
\]
where \( \Gamma_{KL}^M \) are smooth arcs and \( S_{KL} \) is a set of isolated points such that 
\[
\bigcup_{M=1}^{p(K, L)} \Gamma_{KL}^M \cap S_{KL} = \emptyset \quad 0 \leq K, L \leq N
\]
We set \( \partial \Omega_K = \partial \Omega_K \cap \partial \Omega \). Notice that 
\[
\partial \Omega_K = \bigcup_{L, M} \Gamma_{KL}^M
\]
With these notations, it is possible to unambiguously characterize the boundary segments of the partition as
\[
E = E(\mathcal{P}) = \bigcup_{K, L=0, K > L, 1 \leq M \leq p(K, L)} \Gamma_{KL}^M
\]
The boundary segments lying on the interior of \( \Omega \) are denoted by
\[
E_I = E_I(\mathcal{P}) = \bigcup_{K, L=1, K > L, 1 \leq M \leq p(K, L)} \Gamma_{KL}^M
\]
The outward pointing unit normal vector on \( \Omega_K \) is denoted by \( n_K \). Let 
\[
\sigma_{KL} = -\sigma_{LK} = \begin{cases} +1, & K > L \\ -1, & K < L \end{cases}
\]
and define \( n(s) = \sigma_{KL} n_K(s) = \sigma_{LK} n_L(s) \), \( s \in \Gamma_{KL}^M \). That is, \( n \) points outward from the subdomain with the largest index.
Throughout, if \( v \) is some function defined on \( \Omega \), then its restriction to \( \Omega_K \) is denoted by
\[
v_K \equiv v|_{\Omega_K}, \quad 1 \leq K \leq N
\]
In addition to the global Sobolev spaces and norms above, we introduce the broken Sobolev spaces \( H^m(\mathcal{P}) \):
\[
H^m(\mathcal{P}) = \{ v \in L_2(\Omega); v_K \in H^m(\Omega_K); 1 \leq K \leq N \}
\]
equipped with the norm 
\[
\| v \|_{m, \mathcal{P}} = \left\{ \sum_{K=1}^{N} \| v_K \|_{m, \Omega_K}^2 \right\}^{1/2}
\]
Evidently \( H^m(\Omega) \subset H^m(\mathcal{P}) \) for \( m \in \mathbb{N} \) and 
\[
H^0(\mathcal{P}) = L_2(\mathcal{P}) = L_2(\Omega) = H^0(\Omega)
\]
Let $L_2(E)$ denote the space of classes of square integrable functions defined on $E$ with the norm

\[ \| \varphi \|_{0,E} = \left\{ \sum_{k,l=0}^{N} \int_{\Gamma_{KL}} |\varphi_{KL}^M|^2 ds \right\}^{1/2} \]

and inner product

\[ \langle \varphi, \chi \rangle_{0,E} = \sum_{k,l=0}^{N} \int_{\Gamma_{KL}} \varphi_{KL}^M \cdot \chi_{KL}^M ds \]

where $\varphi_{KL}^M = \varphi|_{\Gamma_{KL}}$ is the restriction of $\varphi$ to $\Gamma_{KL}$. Analogous inner product and norm are defined on $L_2(E)$. Throughout let $\gamma$ denote the various trace operators associated with mappings to the boundaries or segments of boundaries [1]. For any curve $\Gamma$, the space $H^{1/2}(\Gamma)$ is the completion of $C^\infty(\Gamma)$ in the norm on $H^1(\Gamma)$ given by

\[ |w|_{H^{1/2},\Gamma}^2 = \| w \|^2_{H^{1/2},\Gamma} + \| w \|_{0,\Gamma}^2 \]

where

\[ |w|_{H^{1/2},\Gamma}^2 = \int_{\Gamma} \frac{|w(x(s)) - w(x(t))|^2}{|x(s) - x(t)|^2} ds dt \]

It is well known that $\gamma \in \mathcal{L}(H^1(\Omega_K), H^{1/2}(\partial\Omega_K))$ and is surjective. Finally, we denote the space of continuous linear functionals on $H^{1/2}(\Gamma)$ by $H^{-1/2}(\Gamma)$.

3. Model problem

Consider the following boundary-value problem for given data $f \in L_2(\Omega)$ and $g \in L_2(\Gamma_N)$:

\[
\begin{align*}
Lu &= f \quad \text{in } \Omega \\
\gamma u &= 0 \quad \text{on } \Gamma_D \\
a \frac{\partial u}{\partial n} &= g \quad \text{on } \Gamma_N
\end{align*}
\]

where

\[ Lu = - \nabla \cdot (a \nabla u) + cu \]

and

\[ \Gamma_N \cap \Gamma_D = \emptyset, \quad \tilde{\Gamma}_N \cup \tilde{\Gamma}_D = \partial \Omega \]

We shall assume there exist constants $q, \bar{a}, \bar{c}$ and $\bar{c}$ such that the coefficients $a \in C^1(\tilde{\Omega})$ and $c \in C(\tilde{\Omega})$ satisfy

\[ 0 < q \leq a(x) \leq \bar{a}, \quad 0 < c \leq c(x) \leq \bar{c} \quad \text{for } x \in \tilde{\Omega} \]
The important case of \( c \equiv 0 \) is therefore excluded but will be discussed in the concluding remarks.

Let
\[
V(\Omega) = \{ v \in H^1(\Omega); \gamma v = 0 \text{ on } \Gamma_D \}
\]

Then \( u \in V(\Omega) \) is the weak solution to (3.1) if

(3.4) \[ a(u, v) = l(v) \quad \forall v \in V(\Omega) \]

where

(3.5) \[ a(u, v) = \int_\Omega (a \nabla u \cdot \nabla v + cuv) \, dx \]

and

(3.6) \[ l(v) = \int_\Omega f v \, dx + \int_{\Gamma_N} g v \, ds \]

Under the above hypotheses \( a: V(\Omega) \times V(\Omega) \to \mathbb{R} \) is a continuous and coercive bilinear form and \( l: V(\Omega) \to \mathbb{R} \) is a continuous linear form. The Lax-Milgram theorem guarantees the existence of a unique solution to (3.4).

It is convenient to break the global forms \( l(\cdot) \) and \( a(\cdot, \cdot) \) into sums of contributions from each subdomain \( \Omega_K \) of the partition \( \mathcal{D} \):

(3.7) \[ a_K(u, v) = \int_{\Omega_K} (a \nabla u \cdot \nabla v + cuv) \, dx \]

(3.8) \[ l_K(v) = \int_{\Omega_K} f v \, dx + \int_{\partial \Omega_K \cap \Gamma_N} g v \, ds \]

so that

(3.9) \[ a(u, v) = \sum_{K=1}^{N} a_K(u, v) \]

and

(3.10) \[ l(v) = \sum_{K=1}^{N} l_K(v) \]

We shall use the notation \( \| \cdot \|_E \) to denote the energy norm

(3.11) \[ \| v \|_E = \sqrt{a(v, v)} \]

and

(3.12) \[ \| v \|_{E, K} = \sqrt{a_K(v, v)} \]

so that

(3.13) \[ \| v \|_E^2 = \sum_{K=1}^{N} \| v \|_{E, K}^2 \]

Finally, we introduce the space \( V_K, 1 \leq K \leq N \)

(3.14) \[ V_K = \{ v \in H^1(\Omega_K); \gamma v = 0 \text{ on } \partial \Omega_K \cap \Gamma_D \} \]
and

\[(3.15) \quad V(\mathcal{P}) = \{ v \in H^1(\mathcal{P}) : \gamma v = 0 \text{ on } \Gamma_n \}\]

so that, equivalently,

\[(3.16) \quad V(\mathcal{P}) = \prod_{k=1}^{N} V_k\]

Let \( H(\text{div. } \Omega) \) denote the space

\[(3.17) \quad H(\text{div. } \Omega) = \{ q \in [L^2(\Omega)]^2 : \text{div } q \in L^2(\Omega) \}\]

which is equipped with the norm

\[(3.18) \quad \| q \|_{H(\text{div. } \Omega)}^2 = \| \text{div } q \|_{L^2(\Omega)}^2 + \| q \|_{L^2(\Omega)}^2\]

Further, define the subspace \( \mathcal{Q} \subset H(\text{div. } \Omega) \) to be

\[(3.19) \quad \mathcal{Q} = \left\{ q \in H(\text{div. } \Omega) : \frac{1}{\partial} v n \cdot q ds = 0 \ \forall v \in V(\Omega) \right\}\]

Finally, denote the space of continuous linear functionals on \( V(\mathcal{P}) \) and which vanish on \( V(\Omega) \) by \( \mathcal{M} \).

**Lemma 3.1.** A continuous linear functional \( \mu \) on \( V(\mathcal{P}) \) vanishes on \( V(\Omega) \) (i.e. \( \mu \in \mathcal{M} \)) if and only if there exists \( q \in \mathcal{Q} \) such that

\[(3.20) \quad \mu(v) = \sum_{k=1}^{N} \int_{\partial \Omega} v_K n_k \cdot q ds\]

**Proof.** The result follows in a similar way to Lemma 1 in [14]. \( \square \)

Let \( \alpha_{KL}^M : \Gamma_{KL}^M \to \mathbb{R} \) be such that for \( s \in \Gamma_{KL}^M \)

\[(3.21) \quad \alpha_{KL}^M(s) + \alpha_{LK}^M(s) = 1, \quad 0 \leq K, L \leq N, \quad 1 \leq M \leq \rho(K, L)\]

For any \( v \in H^2(\mathcal{P}), [v], \langle v \rangle_x, \left[ \frac{\partial v}{\partial n} \right]_{x} \) and \( \langle \frac{\partial v}{\partial n} \rangle_x \in L^2(\mathcal{P}) \) are defined as follows (with the convention \( v_0 \equiv 0 \)) for \( s \in \Gamma_{KL}^M \) and for all \( \Gamma_{KL}^M \in \mathcal{E}:

\[(3.22) \quad \begin{aligned}
[v]_{\Gamma_{KL}^M} &= \sigma_{KL} v_K + \sigma_{LK} v_L \\
\left[ \frac{\partial v}{\partial n} \right]_{\Gamma_{KL}^M} &= \sigma_{KL} v_K + \sigma_{LK} v_L \\
\langle v \rangle_{x} \big|_{\Gamma_{KL}^M} &= \alpha_{KL}^M v_K + \alpha_{LK}^M v_L \\
\left\langle \frac{\partial v}{\partial n} \right\rangle_x \big|_{\Gamma_{KL}^M} &= \alpha_{KL}^M v_K + \alpha_{LK}^M v_L
\end{aligned}\]
In other words, \([v]\) is the difference or *jump* in \(v\) between neighboring subdomains whilst \(\langle v \rangle_z\) is a linear combination or *weighted average* of \(v\) between neighboring subdomains.

A key result in the development is the following generalization of a result (Perceuil-Wheeler) found in [13]:

**Lemma 3.2.** Let \(v \in H^2(\mathcal{P})\) and \(w \in H^1(\mathcal{P})\). Then

\[
\sum_{K=1}^{N} \int_{\partial K} \frac{\partial v}{\partial n} \gamma w_K \, ds = \sum_{K=1}^{N} \int_{\partial K} \left[ \frac{\partial v}{\partial n} \right] \langle w \rangle_z \, ds + \left[ \frac{\partial w}{\partial n} \right] \langle \gamma v \rangle_{1-z}.
\]

**Proof.** Owing to the regularity \(v \in H^2(\mathcal{P})\), it follows that \(n \cdot \nabla v \in L^2(\partial \Omega_K)\) (see [8, p. 9, Remark 1.1]) and therefore the duality pairings may be treated as Lebesgue integrals over the boundaries of the subdomains. We have

\[
\sum_{K=1}^{N} \int_{\partial K} \frac{\partial v}{\partial n} \gamma w_K \, ds = \sum_{K=1}^{N} \sum_{L=0}^{N} \sum_{M=1}^{\rho(K,L)} \int_{r_{KL}} \frac{\partial v}{\partial n} \gamma w_K \, ds
\]

\[
= \sum_{K=0}^{N} \sum_{L>0}^{\rho(K,L)} \int_{r_{KL}} \frac{\partial v}{\partial n} \gamma w_K \, ds + \sum_{L=0}^{N} \sum_{K<L}^{\rho(L,K)} \int_{r_{KL}} \frac{\partial v}{\partial n} \gamma w_K \, ds
\]

\[
= \sum_{K=0}^{N} \sum_{L>0}^{\rho(K,L)} \int_{r_{KL}} \left( \frac{\partial v}{\partial n} \gamma w_K + \frac{\partial v}{\partial n} \gamma w_L \right) \, ds
\]

\[
= \sum_{K=0}^{N} \sum_{L>0}^{\rho(K,L)} \int_{r_{KL}} \left( \sigma_{KL} \frac{\partial v}{\partial n} \gamma w_K + \sigma_{LK} \frac{\partial v}{\partial n} \gamma w_L \right) \, ds
\]

Now

\[
\sigma_{KL} \frac{\partial v}{\partial n} = \left[ \frac{\partial v}{\partial n} \right]_{x_{KL}} + \sigma_{KL} \langle \frac{\partial v}{\partial n} \rangle_{1-z}
\]

and

\[
\gamma w_K = \sigma_{LK} \sigma_{KL} \langle w \rangle + \langle w \rangle_z
\]

so since \(\sigma_{KL} + \sigma_{LK} \equiv 0\) we obtain

\[
\sigma_{KL} \frac{\partial v}{\partial n} \gamma w_K + \sigma_{LK} \frac{\partial v}{\partial n} \gamma w_L = (\sigma_{KL} + \chi_{KL}) \left[ \frac{\partial v}{\partial n} \right] \langle w \rangle_z + (\sigma_{LK} + \chi_{KL}) \left[ \frac{\partial v}{\partial n} \right] \langle w \rangle_{1-z}
\]

and the result follows as claimed since \(\sigma_{KL} + \chi_{LK} \equiv 1\). \(\Box\)
4. A posteriori error bounds

Let $u_h$ be any approximation to the true solution of the model problem such that

$$u_h \in V(\Omega) \cap H^2(\mathcal{P})$$

(4.1)

Such an approximation might, for example, be obtained using a finite element discretization with $\Omega_K$ as the elements. Conversely, $u_h$ might simply be some guess at the true solution.

The problem of interest is that of numerically estimating the accuracy of the approximation. The error $e(x)$ in the approximation at $x$ is defined as $u(x) - u_h(x)$. We shall be specifically interested in obtaining bounds on the error measured in the energy norm, $\|e\|_E$.

Lemma 4.1. Let $J : V(\Omega) \to \mathbb{R}$ be defined as

$$J(v) = \frac{1}{2} a(v, v) - l(v) + a(u_h, v)$$

(4.2)

Then the error $e = u - u_h$ is the unique minimizer of $J$ over $V(\Omega)$. Moreover,

$$-\frac{1}{2} \|e\|_E^2 = J(e) = \inf_{v \in V(\Omega)} J(v)$$

(4.3)

Proof. First, notice that $e \in V(\Omega)$ since $u_h \in V(\Omega)$. Now

$$J(v) = \frac{1}{2} a(v, v) - l(v) + a(u_h, v) = \frac{1}{2} a(v, v) - a(u, v) + a(u_h, v)$$

$$= \frac{1}{2} a(v, v) - a(e, v)$$

and hence

$$J(e) = -\frac{1}{2} \|e\|_E^2$$

Let $\lambda \in \mathbb{R}$ and $\omega \in V(\Omega)$. Then

$$J(e) - J(e + \lambda \omega) = -\frac{1}{2} \lambda^2 a(\omega, \omega) \leq 0$$

and equality holds iff $\|\lambda \omega\|_E^2 = 0 \iff \lambda \omega = 0$. Thus $J(e) \leq J(v) \forall v \in V(\Omega)$. \qed

Let $J_\mathcal{P} : \mathcal{P} \to \mathbb{R}$ be defined as

$$J_\mathcal{P}(v) = \sum_{K=1}^{N} \left\{ \frac{1}{2} a_K(v, v) - l_K(v) + a_K(u_h, v) \right\}$$

(4.4)

That is, $J_\mathcal{P}$ is an extension of $J$ to $V(\mathcal{P}) \supset V(\Omega)$. Let $L_\mathcal{P} : \mathcal{P} \times \mathcal{M} \to \mathbb{R}$ denote the Lagrangian functional

$$L_\mathcal{P}(v, \mu) = J_\mathcal{P}(v) - \mu(v)$$

(4.5)

associated with the constraint $v \in V(\Omega)$, $\mu$ representing the Lagrange multiplier. It is instructive to compare the Lagrangian $L_\mathcal{P}$ with analogous functionals used in the analysis of primal-hybrid finite element methods [14].
Lemma 4.2. With the above assumptions and notations there follows

\[ -\frac{1}{2} \| e \|_E^2 = \inf_{v \in V(\Omega)} \sup_{\mu \in \mathcal{U}} \mathcal{L}_\varphi(v, \mu) \]

Proof. Let \( \Phi : V(\Omega) \to \mathbb{R} \) be the functional

\[ \Phi(v) = \sup_{\mu \in \mathcal{U}} \mathcal{L}_\varphi(v, \mu) \]

and observe that

\[ \Phi(v) = \begin{cases} J(v), & v \in V(\Omega) \\ +\infty, & \text{otherwise} \end{cases} \]

Therefore

\[ \inf_{v \in V(\Omega)} \sup_{\mu \in \mathcal{U}} \mathcal{L}_\varphi(v, \mu) = \inf_{v \in V(\Omega)} \Phi(v) = \inf_{v \in V(\Omega)} J(v) = -\frac{1}{2} \| e \|_E^2 \]

where the final step follows from Lemma 4.1. \( \square \)

Let \( v \in V(\Omega) \). Then applying Green's identity on each subdomain, we have

\[ a(u_h, v) = \sum_{K=1}^{N} \int_{\Omega_K} v Lu_h \, dx + \sum_{K=1}^{N} \sum_{L=0}^{K} \int_{\Gamma^M_{K,L}} \left\{ \left[ a \frac{\partial u_h}{\partial n} \right] \langle v \rangle_z + \gamma v_0 \right\} \, ds + \sum_{K=1}^{N} \int_{\Gamma^W_{K}} \left\{ \left[ a \frac{\partial u_h}{\partial n} \right] \langle v \rangle_z + \gamma v_0 \right\} \, ds \]

Applying Lemma 3.2 to the first term gives

\[ a(u_h, v) = \sum_{K=1}^{N} \int_{\Omega_K} v Lu_h \, dx + \sum_{K=1}^{N} \sum_{L=0}^{K} \int_{\Gamma^M_{K,L}} \left\{ \left[ a \frac{\partial u_h}{\partial n} \right] \langle v \rangle_z \right\} \, ds \]

since \( a_{K_0} \equiv 1 \), from (3.22) we obtain

\[ \left\{ \langle a \frac{\partial u_h}{\partial n} \rangle \right\}_{1-z} = 0, \quad \gamma v_0 = 0, \quad \langle v \rangle_z = v_K \quad \text{and} \]

\[ \left[ a \frac{\partial u_h}{\partial n} \right] = a \frac{\partial u_h}{\partial n} \quad \text{on } \Gamma^W_{K_0} \]
Now \( \gamma v = 0 \) on \( \Gamma^M_{k_0} \cap \Gamma_D \) and so

\[
J_u(v) = \sum_{k=1}^{N} \left\{ \frac{1}{2} a_k(v, v) - (f, v)_{0, K} + (Lu, v)_{0, K} \right\} + \sum_{K,L=1 \atop K \neq L}^{N} \int_{\Gamma^M_{kL}} \left\{ \left[ \frac{\partial u_h}{\partial n} \right] \langle v \rangle + \left[ a \frac{\partial u_h}{\partial n} \right] \right\}_1 \ ds \\
+ \sum_{K=1 \atop \rho(K,L)}^{N} \int_{\Gamma^M_{kL}} \left( a \frac{\partial u_h}{\partial n} - g \right) \gamma v \ ds \\
= \sum_{k=1}^{N} \left\{ \frac{1}{2} a_k(v, v) + (Lu - f, v)_{0, K} + \sum_{1 \leq M \leq \rho(K,0)} \int_{\Gamma^M_{kL}} \left( a \frac{\partial u_h}{\partial n} - g \right) \gamma v \ ds \\
+ \sum_{L=1 \atop \rho(K,L)}^{N} \int_{\Gamma^M_{kL}} \left[ a \frac{\partial u_h}{\partial n} \right] \gamma v \ ds \\
+ \sum_{K,L=1 \atop K \neq L}^{N} \int_{\Gamma^M_{kL}} \left( a \frac{\partial u_h}{\partial n} \right)_1 \ ds 
\right\}
\]

since

\[
\sum_{K,L=1 \atop K \neq L}^{N} \int_{\Gamma^M_{kL}} \left[ a \frac{\partial u_h}{\partial n} \right] \langle v \rangle \ ds = \sum_{K=1 \atop \rho(K,L)}^{N} \sum_{L=1}^{N} \sum_{M=1}^{\rho(K,L)} \int_{\Gamma^M_{kL}} a \frac{\partial u_h}{\partial n} \gamma v \ ds
\]

Let

\[
r_k(x) = f(x) - Lu_h(x), \ x \in \Omega_k
\]
denote the elementwise residual and define

\[
R_k = \begin{cases} 
  g - a \frac{\partial u_h}{\partial n} & \text{on } \Gamma_N \cap \Gamma^M_{k_0}, 1 \leq K \leq N, 1 \leq M \leq \rho(K,0) \\
  - \chi_{K L} \left[ a \frac{\partial u_h}{\partial n} \right] & \text{on } \Gamma^M_{kL}, 1 \leq K, L \leq N, 1 \leq M \leq \rho(K, L)
\end{cases}
\]

Then:

\[
J_u(v) = \sum_{K=1}^{N} \left( \frac{1}{2} a_k(v, v) - (r_K, v)_{0, K} + (Lu, v)_{0, K} \right) + \left( a \frac{\partial u_h}{\partial n} \right)_1 \ ds
\]

where \( J_{u,K} : V_K \to \mathbb{R} \) is given by

\[
J_{u,K}(v_K) = \frac{1}{2} a_k(v_K, v_K) - (r_K, v_K)_{0, K} - \int_{\Omega_k} R_K \gamma v_K \ ds
\]

Notice that it is unnecessary to define \( R_k \) on \( \Gamma_D \), since \( \gamma v_K = 0 \) on \( \Gamma_D \).
Roughly speaking, we may interpret this result as showing that $J_\varphi$ may be decomposed into a sum of local contributions from each of the subdomains $\Omega_K$, plus an extra term coupling the local contributions through the regularity $\nu \in V(\Omega)$. The significant feature, from our viewpoint, is the possibility of choosing the Lagrange multiplier $\mu \in \mathcal{M}$ in such a way that the coupling is severed. If this step could be achieved, then it would result in a sequence of local problems rather than a single global problem. This situation is highly desirable in practice since the computational cost of dealing with a sequence of local problems is much smaller than dealing with a single global problem.

In the following, we focus on the problem of constructing a suitable choice of $\mu$. Making use of the definitions (3.22) we find

$$
(4.14) \quad \left( \left\langle a \frac{\partial u_h}{\partial n} \right\rangle_{1-\varepsilon} , [v] \right)_{0,E} = \sum_{K=1}^{N} \left\{ \sum_{L=0}^{\sum_{1 \leq M \leq \rho(K,L)}} \int_{\Gamma_{KL}} a \left\langle \frac{\partial u_h}{\partial n} \right\rangle_{1-\varepsilon} v \, ds \right\}
$$

Noting that

$$
(4.15) \quad \partial \Omega_K = \bigcup_{1 \leq M \leq \rho(K,L)} \Gamma_{KL}^{M}
$$

we may formally define the linear functional $\zeta$ on $V_K$ by

$$
(4.16) \quad \zeta(v) = \sum_{L=0}^{\sum_{1 \leq M \leq \rho(K,L)}} \int_{\Gamma_{KL}} a \left\langle \frac{\partial u_h}{\partial n} \right\rangle_{1-\varepsilon} v_K \, ds
$$

Examining this expression one readily concludes that, in fact, $\zeta$ is a bounded linear functional on $L^2(\partial \Omega_K)$ and hence a fortiori $\zeta \in H^{-1/2}(\partial \Omega_K)$. Recalling that the Trace Operator

$$
H(\text{div}, \Omega_K) \ni q \mapsto n_K \cdot q \in H^{-1/2}(\partial \Omega_K)
$$

is surjective, we may construct $q \in H(\text{div}, \Omega_K)$ such that

$$
(4.17) \quad \int_{\partial \Omega_K} v_K n_K \cdot q \, ds = \sum_{L=0}^{\sum_{1 \leq M \leq \rho(K,L)}} \int_{\Gamma_{KL}} a \left\langle \frac{\partial u_h}{\partial n} \right\rangle_{1-\varepsilon} v_K \, ds
$$

Performing this procedure over each subdomain results in our having constructed $q \in H(\text{div}, \Omega)$ satisfying

$$
(4.18) \quad \sum_{K=1}^{N} \int_{\partial \Omega_K} v_K n_K \cdot q \, ds = \sum_{K=1}^{N} \sum_{L=0}^{\sum_{1 \leq M \leq \rho(K,L)}} \int_{\Gamma_{KL}} a \left\langle \frac{\partial u_h}{\partial n} \right\rangle_{1-\varepsilon} v_K \, ds
$$

$$
= \left( \left\langle a \frac{\partial u_h}{\partial n} \right\rangle_{1-\varepsilon} , [v] \right)_{0,E}
$$

Moreover, suppose $\nu \in V(\Omega)$, then

$$
(4.20) \quad \sum_{K=1}^{N} \int_{\partial \Omega_K} v_K n_K \cdot q \, ds = \int_{\partial \Omega} v n \cdot q \, ds
$$
and

\[(4.21) \quad \left( \left\langle a \frac{\partial u_h}{\partial n} \right\rangle_{1-z}, v \right\rangle_{0,E} = \frac{i}{\omega} \langle a \frac{\partial u_h}{\partial n} \rangle_{1-z} v ds = 0 \]

where the final quantity vanishes since \( \gamma v = 0 \) on \( \Gamma_D \) and \( \langle a (\partial u_h/\partial n) \rangle_{1-z} = 0 \) on \( \Gamma_N \) by definition. In view of the above developments, it follows that \( q \in \mathcal{Q} \). Applying Lemma 3.1 then shows that there exists \( \hat{\mu} \in \mathcal{M} \) such that

\[(4.22) \quad \hat{\mu}(v) = \left( \left\langle a \frac{\partial u_h}{\partial n} \right\rangle_{1-z}, v \right\rangle_{0,E} \]

Hence, returning to \( J_{\mathcal{P}} \), we find we may write

\[(4.23) \quad J_{\mathcal{P}}(v) = \sum_{K=1}^{N} J_{\mathcal{P},K}(v) = \hat{\mu}(v) \]

and, equally well,

\[(4.24) \quad L_{\mathcal{P}}(v, \mu) = \sum_{K=1}^{N} J_{\mathcal{P},K}(v) + \hat{\mu}(v) - \mu(v) \]

Therefore, we can obtain the desired decoupling in the case \( \mu = \hat{\mu} \).

**Lemma 4.3.** Let \( \hat{\mu} \in \mathcal{M} \) be constructed as above. Then

\[(4.25) \quad \| e \|_E^2 \leq \sup_{v \in V(\mathcal{P})} \text{sup} - 2L_{\mathcal{P}}(v, \hat{\mu}) \]

where

\[(4.26) \quad L_{\mathcal{P}}(v, \mu) = \sum_{K=1}^{N} J_{\mathcal{P},K}(v) \]

**Proof.** From Lemma 4.2, we have

\[-\frac{1}{2} \| e \|_E^2 = \inf_{v \in V(\mathcal{P})} \sup_{\mu \in \mathcal{M}} L_{\mathcal{P}}(v, \mu) \geq \sup_{v \in V(\mathcal{P})} \inf_{\mu \in \mathcal{M}} L_{\mathcal{P}}(v, \mu) \geq \inf_{v \in V(\mathcal{P})} L_{\mathcal{P}}(v, \hat{\mu}) \]

so \( \| e \|_E^2 \leq \sup_{v \in V(\mathcal{P})} - 2L_{\mathcal{P}}(v, \hat{\mu}) \). \( \Box \)

One serious drawback of this result as a means of estimating the error is the presence of the supremum term which either will not be attained by our computed choice of \( v \in V(\mathcal{P}) \) or the cost involved in calculating a suitable \( v \) will prove to be disproportionately expensive compared to the computational effort expended in obtaining \( u_h \) itself. This type of difficulty was successfully resolved in [2] using dual variational principles (see [7]).
Lemma 4.4. Let

\begin{equation}
W_K = \left\{ p \in H(\text{div}, \Omega_K) : \int_{\partial\Omega_K} n_K \cdot pv \, ds = \frac{1}{2} \int_{\Omega_K} R_K v \, ds \quad \forall v \in V_K \right\}
\end{equation}

In addition, let \( G_K : W_K \to \mathbb{R} \) be the functional

\begin{equation}
G_K(p) = -\frac{1}{2} \int_{\Omega_K} \frac{1}{a} p \cdot pdx - \frac{1}{2} \int_{\partial\Omega_K} \nabla \cdot (\nabla \cdot p + r_K)^2 \, dx
\end{equation}

where \( r_K \) is the element residual of (4.10). Further, let \( v_K^* \in V_K \) be the solution of the local problem

\begin{equation}
a_K(v_K^*, \omega) = (r_K, \omega)_{0,K} + \frac{1}{2} \int_{\partial\Omega_K} \gamma v \omega \, ds \quad \forall \omega \in V_K
\end{equation}

Then

\begin{equation}
(i) \quad p_K^* \overset{\text{def}}{=} a \nabla v_K^* \in W_K
\end{equation}

and

\begin{equation}
(ii) \quad \inf_{v_k \in V_K} J_{\mathcal{G}, K}(v_k) = J_{\mathcal{G}, K}(v_K^*) = G_K(p_K^*) = \sup_{p \in W_K} G_K(p)
\end{equation}

Proof. (i) The existence and uniqueness of \( v_K^* \) is a consequence of the Lax-Milgram Theorem. The strong form of (4.29) is

\(- \nabla \cdot (a \nabla v_K^*) + cv_K^* = r_K \quad \text{in} \ \Omega_K
\)

\(a \frac{\partial v_K^*}{\partial n_K} = R_K \quad \text{on} \ \partial\Omega_K \setminus \Gamma_D
\)

\(\gamma v_K^* = 0 \quad \text{on} \ \partial\Omega_K \cap \Gamma_D
\)

Let \( p_K^* = a \nabla v_K^* \). Then

\(\nabla \cdot p_K^* = cv_K^* - r_K \in L_2(\Omega_K)
\)

whilst

\(n_K \cdot p_K^* = a \frac{\partial v_K^*}{\partial n_K} = R_K \quad \text{on} \ \partial\Omega_K \setminus \Gamma_D
\)

Thus \( p_K^* \in W_K \).

(ii) That \( v_K^* \) is the minimizer of \( J_{\mathcal{G}, K} \) follows in a similar manner to Lemma 4.1. Moreover \( v_K^* \in V_K \) so

\(J_{\mathcal{G}, K}(v_K^*) = \frac{1}{2} a_K(v_K^*, v_K^*) - \frac{1}{2} \int_{\partial\Omega_K} R_K \gamma v_K^* \, ds
\)

\(= \frac{1}{2} a_K(v_K^*, v_K^*) - a_K(v_K^*, v_K^*) = -\frac{1}{2} a_K(v_K^*, v_K^*)
\)

Furthermore,

\(G_K(p_K^*) = -\frac{1}{2} \int_{\partial\Omega_K} a|\nabla v_K^*|^2 \, dx - \frac{1}{2} \int_{\partial\Omega_K} \frac{1}{c} (cv_K^*)^2 \, dx = -\frac{1}{2} a_K(v_K^*, v_K^*)
\)
Since $G_K$ is strictly concave and quadratic, it suffices to show $G_K$ is stationary at $p_K^*$.

Let

\[ q \in \left\{ q \in H(\text{div}, \Omega_K) : \frac{1}{\partial K} \nabla v \cdot q \, ds = 0 \ \forall v \in V_K \right\} \]

and let $\lambda \in \mathbb{R}$. Then $p_K^* + \lambda q \in W_K$ and

\[
\begin{align*}
\frac{d}{d\lambda} G_K(p_K^* + \lambda q) \big|_{\lambda = 0} &= -\int_{\partial K} q \cdot p_K^* \, dx - \int_{\partial K} \frac{1}{c} (\nabla \cdot q)(\nabla \cdot p_K^* + r_K) \, dx \\
&= -\int_{\partial K} \frac{q \cdot \nabla v_K^*}{c} \, dx - \int_{\partial K} (\nabla \cdot q)v_K^* \, dx \\
&= -\frac{1}{\partial K} \gamma v_K^* (n_K \cdot q) \, ds
\end{align*}
\]

This now vanishes owing to (4.32). \(\square\)

**Theorem 4.5.** Let $W_K$ be as in (4.27). Then

\[ \|e\|_{E}^2 \leq -2 \sum_{K=1}^{N} G_K(p) \ \forall p \in \prod_{K=1}^{N} W_K \]

**Proof.** From Lemmas 4.3 and 4.4 we have

\[
\|e\|_{E}^2 \leq \sup_{v \in V(\Omega)} \sum_{K=1}^{N} -2 J_{\rho,K}(v) = -2 \sum_{K=1}^{N} \inf_{v \in V_K} J_{\rho,K}(v_K) = -2 \sum_{K=1}^{N} \sup_{p \in W_K} G_K(p_K)
\]

\[
\leq -2 \sum_{K=1}^{N} G_K(p_K) \ \forall p_K \in W_K = -2 \sum_{K=1}^{N} G_K(p) \ \forall p \in \prod_{K=1}^{N} W_K \ \square
\]

The main result, Theorem 4.5, shows that computable error bounds can be obtained merely by constructing elements of the linear manifolds $W_K$. Obviously to obtain realistic estimates it is necessary to choose $p$ with some care.

5. Local element residual error estimator

In this section we propose a strategy for constructing $p$. Let $\rho_K : W_K \to \mathbb{R}$ and $A_K : W_K \to \mathbb{R}$ be defined as

\[ \rho_K^2(p) = \int_{\partial K} \frac{1}{c} \rho \cdot p \, dx \]

and

\[ A_K^2(p) = \int_{\partial K} \frac{1}{c} (\nabla \cdot p + r_K)^2 \, dx \]
Furthermore, let

\[(5.3) \quad \eta^2_K(p) = e^2_K(p) + A^2_K(p)\]

Then Theorem 4.5 may be rewritten as

\[(5.4) \quad \|e\|^2_K \leq \sum_{k=1}^{N} \eta^2_K(p) \quad \forall p \in \bigcap_{k=1}^{N} W_k\]

This result suggests that \(p\) should be chosen to minimize the local functionals \(\eta^2_K(p)\). Owing to the fact that the functionals and spaces are local this is a viable practical method. However, an alternative approach is pursued for reasons which will become clear. Rather than minimizing \(e^2_K + A^2_K\) we shall concentrate all our efforts on the term \(A^2_K\) alone. The strategy is therefore to choose \(p_K \in W_k\) such that

\[(5.5) \quad A_K(p_K) \leq A_K(q_K) \quad \forall q_K \in W_k\]

On subdomains \(\Omega_K: \partial \Omega_K \cap \Gamma_D \neq \emptyset\) it is always possible to choose \(p_K\) to give \(A_K(p_K) \equiv 0\). Specifically we choose \(p_K = \nabla \varphi_K\) where \(\varphi_K \in H^1(\Omega_K)\) satisfies

\[(5.6) \begin{cases} -\Delta \varphi_K = r_K & \text{in } \Omega_K \\ \frac{\partial \varphi_K}{\partial n} = R_K & \text{on } \partial \Omega_K \setminus \Gamma_D \\ \gamma \varphi_K = 0 & \text{on } \partial \Omega_K \cap \Gamma_D \end{cases} \]

The existence of a unique \(\varphi_K\) is again guaranteed by the Lax-Milgram Theorem.

For subdomains \(\Omega_K: \partial \Omega_K \cap \Gamma_D = \emptyset\), it is in general not possible to choose \(p_K\) such that \(A_K(p_K) = 0\). The following result quantifies this statement.

Lemma 5.1. Suppose \(\partial \Omega_K \cap \Gamma_D\) is empty. Let \(\varphi_K \in H^1(\Omega_K)\) be such that

\[(5.7) \begin{cases} -\Delta \varphi_K = r_K - c\delta_K & \text{in } \Omega_K \\ \frac{\partial \varphi_K}{\partial n_K} = R_K & \text{on } \partial \Omega_K \end{cases} \]

where

\[(5.8) \quad \delta_K = \left\{ \begin{array}{c} \int_{\Omega_K} r_K \, dx + \oint_{\partial \Omega_K} R_K \, ds \\ \oint_{\partial \Omega_K} c \, ds \end{array} \right\} \]

then

\[(5.9) \begin{cases} (i) \quad \nabla \varphi_K \in W_K \\ (ii) \quad A_K(\nabla \varphi_K) \leq A_K(q_K) \quad \forall q_K \in W_k \end{cases} \]

Proof. (i) First, observe that

\[\int_{\Omega_K} (r_K - c\delta_K) \, dx + \oint_{\partial \Omega_K} R_K \, ds = 0\]
It is well known that this compatibility condition is necessary and sufficient for the existence of a solution $\varphi_K \in H^1(\Omega_K)$ to (5.7). ($\varphi_K$ being unique up to the addition of an arbitrary constant). Now

$$\nabla \cdot (\nabla \varphi_K) = - r_K + c \delta_K \in L^2(\Omega_K)$$

and

$$n_K \cdot \nabla \varphi_K = \frac{\partial \varphi_K}{\partial n_K} = R_K \quad \text{on } \partial \Omega_K$$

so $\nabla \varphi_K \in W_K$.

(ii) Let $q_K \in W_K$ be given and define

$$\xi_K = \nabla \cdot q_K + r_K \in L^2(\Omega_K)$$

then

$$A^2_K(q_K) = \int_{\Omega_K} \frac{1}{c} \xi_K^2 \, dx$$

Moreover, since $q_K \in W_K$,

$$\int_{\Omega_K} \xi_K \, dx = \int_{\Omega_K} \frac{1}{c} n_K \cdot q_K \, ds + \int_{\Omega_K} r_K \, dx = \int_{\Omega_K} \frac{1}{c} \, dx$$

By the Cauchy-Schwarz Inequality,

$$\left( \int_{\Omega_K} c \, dx \right)^2 \leq \int_{\Omega_K} \xi_K \, dx \cdot \int_{\Omega_K} \frac{1}{c} \xi_K^2 \, dx$$

and, hence.

$$A^2_K(\nabla \varphi_K) = \delta^2_K \int_{\Omega_K} c \, dx \leq \int_{\Omega_K} \frac{1}{c} \xi_K^2 \, dx = A^2_K(q_K)$$

In view of this result, we choose $p_K = \nabla \varphi_K$ where $\varphi_K$ is any solution of the local problem (5.7). The local error estimator on subdomain $\Omega_K$ is taken as

$$\eta^2_K(\nabla \varphi_K) = \varepsilon^2_K(\nabla \varphi_K) + A^2_K(\nabla \varphi_K)$$

where $\varphi_K$ is the solution of (5.6) or (5.7) depending on whether or not $I_D$ intersects $\partial \Omega_K$.

**Theorem 5.2.** Let $\varphi_K$ be the solution of (5.6) or (5.7). Then

(5.11) \[ \| e \|^2_E \leq \sum_{K=1}^{N} \eta^2_K(\nabla \varphi_K) \]

**Proof.** Follows immediately from foregoing arguments and (5.4). \[ \square \]

The importance of this result is that no matter how we choose $\alpha(s)$ subject to (3.21), the resulting error estimator always gives an upper bound on the true error.

At this stage it is worthwhile to compare the result with other types of element residual methods. Almost all existing methods are associated with the specific choice $\alpha(s) \equiv \frac{1}{2}$.

A more fundamental difference between the method proposed here and existing methods is that the local problem involves only the Laplacian operator whilst
other methods have local problems based on the operator $L$. While it might seem more advantageous to have local problems based on $L$ this actually does not appear to be the case.

One requirement of error estimators which is of paramount importance is that they must be cheap to compute. The local problems are often solved using finite elements in an $h$, $p$, or $h-p$ mode. The main cost in solving by finite elements is the assembly of stiffness matrices and solution of the resulting matrix equation. By basing the local problems on the Laplacian one can assemble the stiffness matrix a priori on the reference element and keep this as data within the code. Secondly, one can easily construct an orthogonal basis for the $p$-version finite element approximation in the case of the Laplacian operator. This not only makes the solution of the linear system extremely cheap but allows one to increase the accuracy of the approximation of the local problem very simply. As a consequence, one can assume that the local problems have been solved exactly as indeed we shall do throughout our analysis.

Lemma 5.3. Let $\Phi_K$ be a solution of the local residual problem (5.7). Then

\begin{equation}
A_K(\nabla \Phi_K) \leq \left( \int_{\Omega_\epsilon} c e^2 \, dx \right)^{1/2} + \left( \int_{\Omega_\epsilon} c \, dx \right)^{-1/2} \left| \int_{\partial \Omega_\epsilon} \left( \frac{\partial \Phi}{\partial n_K} \right) _{1-z} \, ds \right|
\end{equation}

Proof. From Lemma 5.1 we have

\[ A_K^2(\nabla \Phi_K) = \delta_K \int_{\Omega_\epsilon} c \, dx \]

where

\[ \delta_K \int_{\Omega_\epsilon} c \, dx = \int_{\Omega_\epsilon} r_K \, dx + \int_{\partial \Omega_\epsilon} R_K \, ds \]

Now $r_K = f - Lu_h = L e_K$ so

\[ \delta_K \int_{\Omega_\epsilon} c \, dx = \int_{\Omega_\epsilon} \nabla \cdot (a \nabla e_K) \, dx + \int_{\Omega_\epsilon} c e_K \, dx + \int_{\partial \Omega_\epsilon} R_K \, ds \]

\[ = \int_{\partial \Omega_\epsilon} \left\{ -a \frac{\partial e_K}{\partial n_K} + R_K \right\} \, ds + \int_{\Omega_\epsilon} c e_K \, dx \]

On $\partial \Omega_K \setminus \partial \Omega$ we have

\begin{equation}
- a \frac{\partial e_K}{\partial n_K} + R_K = -a \frac{\partial u}{\partial n_K} + a \frac{\partial u_h}{\partial n_K} \bigg|_K - \sigma_{KL} \left[ a \frac{\partial u_h}{\partial n} \right] \\
= -a \frac{\partial u}{\partial n_K} + a \frac{\partial u_h}{\partial n_K} \bigg|_K - \sigma_{KL} \left( a \frac{\partial u_h}{\partial n} \right) \left[ \sigma_{KL} + a \frac{\partial u_h}{\partial n} \right]_L \\
= \sigma_{KL} \left( a \frac{\partial u_h}{\partial n} \right)_{1-z} - a \frac{\partial u}{\partial n} \\
= -\sigma_{KL} \left( a \frac{\partial e}{\partial n} \right)_{1-z}
\end{equation}
By the Triangle and Cauchy-Schwarz Inequalities,
\[
\frac{\partial}{\partial x}\int_{\Omega} c \, dx \leq \left| \int_{\Omega} c e \, dx \right| + \left| \int_{\Omega} a \frac{\partial e}{\partial n} \right|_{1-z} \cdot ds \leq \left( \int_{\Omega} c e \cdot \int_{\Omega} c e^2 \, dx \right)^{1/2} + \left| \int_{\Omega} a \frac{\partial e}{\partial n} \right|_{1-z} \cdot ds \]
and hence
\[
A_k(\nabla \varphi_k) \leq \left( \int_{\Omega} c e^2 \, dx \right)^{1/2} + \left( \int_{\Omega} c e \, dx \right)^{-1/2} \left| \int_{\Omega} a \frac{\partial e}{\partial n} \right|_{1-z} \cdot ds \]
Lemma 5.4. Let \( \psi \in H^1(Q_k) \) and let
\[
(5.14) \quad \psi_c = \int_{\Omega} c \psi \, dx \int_{\Omega} c \, dx
\]
Then there exists a constant \( C > 0 \) such that
\[
(5.15) \quad \| \psi - \psi_c \|_{0,\Omega} \leq Ch \| \psi \|_{1,\Omega}
\]
where \( h_k = \text{diam}(Q_k) \).
Proof. If \( \psi \) is constant then \( \psi_c = \psi \). The result follows from a standard application of the Bramble-Hilbert Lemma. \( \square \)

Lemma 5.5. There exists a constant \( C > 0 \) such that
\[
(5.16) \quad \epsilon_k(\nabla \varphi_k) \leq \left( \int_{\Omega} a |\nabla e|^2 \, dx \right)^{1/2} + Ch \left( \int_{\Omega} c e^2 \, dx \right)^{1/2}
\]
\[+ Ch^2 \left( \int_{\Omega} a \frac{\partial e}{\partial n} \right)^{1/2} \]
Proof. By the Triangle Inequality
\[
\epsilon_k(\nabla \varphi_k) = \left( \int_{\Omega} \frac{1}{4} |\nabla \varphi_k|^2 \, dx \right)^{1/2}
\]
\[\leq \left( \int_{\Omega} a |\nabla e|^2 \, dx \right)^{1/2} + \left( \int_{\Omega} \frac{1}{4} |\nabla \varphi_k - a \nabla e|^2 \, dx \right)^{1/2}
\]
Let \( a \nabla \psi = \nabla \varphi_k - a \nabla e \) then
\[- \nabla \cdot (a \nabla \psi) = - \nabla \varphi_k + \nabla \cdot (a \nabla e) = r_k - c \delta_k + \nabla \cdot (a \nabla e)
\]
\[= Le - c \delta_k + \nabla \cdot (a \nabla e) = c(e - \delta_k)
\]
While on \( \partial Q_k \setminus I_D \), using (5.13)
\[
a \frac{\partial \psi}{\partial n_k} = a \frac{\partial \varphi_k}{\partial n_k} - a \frac{\partial e_k}{\partial n_k} = R_k - a \frac{\partial e_k}{\partial n_k} = - \sigma_{KL} \left( \frac{\partial e}{\partial n} \right)_{1-z}
\]
Therefore
\[
\frac{1}{a_k} \int a |\nabla \psi|^2 \, dx = \frac{1}{a_k} \int a \frac{\partial \psi}{\partial n_k} \psi \, ds - \frac{1}{a_k} \int a \nabla \cdot (a \nabla \psi) \, dx
\]
\[
= - \frac{1}{a_k} \sigma \int a \frac{\partial e}{\partial n} \psi \, ds + \int \psi c(e - \delta_k) \, dx
\]
\[
= - \frac{1}{a_k} \sigma \int a \frac{\partial e}{\partial n} \psi \, ds + \int c \psi e \, dx - \psi \delta_k \int c \, dx
\]

Examining the proof of Lemma 5.3 we find
\[
\delta_k \int c \, dx = - \frac{1}{a_k} \sigma \int a \frac{\partial e}{\partial n} \psi \, ds + \int c \psi e \, dx
\]
and so
\[
\int a |\nabla \psi|^2 \, dx = - \frac{1}{a_k} \sigma \int a \frac{\partial e}{\partial n} \psi \, ds + \int c \psi e \, dx
\]
\[
\leq \left( \frac{1}{a_k} \int a \frac{\partial e}{\partial n} \right)^2 \int \psi - \psi_e \|_{0, a_k}
\]
\[
+ C \left( \int c \psi_e \, dx \right)^{1/2} \| \psi - \psi_e \|_{0, a_k}
\]

By Lemma 5.4 and standard Trace Inequalities
\[
\| \psi - \psi_e \|_{0, a_k} \leq Ch_k^{-1/2} \| \psi - \psi_e \|_{0, a_k} \leq Ch_k^{1/2} |\psi|_{1, a_k}
\]
and since a is bounded below by \( a > 0 \).
\[
|\psi|^2_{1, a_k} \leq \frac{1}{a_k} \int a |\nabla \psi|^2 \, dx
\]

Therefore
\[
\left( \int a |\nabla \psi|^2 \, dx \right)^{1/2} \leq Ch_k^{1/2} \left( \frac{1}{a_k} \int a \frac{\partial e}{\partial n} \right)^{1/2} + Ch_k \left( \int c \psi_e \, dx \right)^{1/2}
\]

Collecting these results gives (5.15). \( \Box \)

The following result complements Theorem 5.2.

**Theorem 5.6.** For any \( \mu_k > 0 \) there exists \( C > 0 \) such that

\[
\eta_k^2 (\nabla \varphi_k) \leq (1 + \mu_k) \| e \|_{2, a_k}^2 + C(1 + \mu_k^{-1}) \left\{ \| h_k \frac{\partial e}{\partial n} \|_{1-} \right\}^2
\]
\[
+ Ch_k^2 \| e \|_{0, a_k}^2 + \frac{1}{a_k} \sigma \int a \frac{\partial e}{\partial n} \psi \, ds + \int c \, dx \right\}^{1/2}
\]
Proof. Follows from previous lemmas and the elementary inequality

\[(a + b)^2 \leq (1 + \mu K) a^2 + (1 + \mu K^{-1}) b^2 \quad \forall \mu K > 0\]  

This result shows that a necessary condition in order for \(\sum_{k=1}^{N} \eta_{k}^2\) to realistically estimate the true error is that \(\langle \partial e / \partial n \rangle_{1 - \varepsilon}\) be sufficiently small. The controlling term in the right hand bound is

\[
(5.18) \quad \left| \frac{\bar{f}}{\partial u} \frac{\partial u}{\partial n} \right|_{1 - \varepsilon} ds \right|^2 \left( \int_{\Omega} c dx \right)^{-1/2}
\]

Notice that the denominator behaves like means \((\Omega K)^{-1}\) as \(h_k \to 0\). In particular there is a danger that the right hand side will blow up as \(h_k \to 0\), making the bound meaningless.

6. Flux splitting for finite element approximations

In this section we suppose \(u_h\) is an approximation obtained using the finite element method with \(\Omega_k\) corresponding to the elements. A related approach has been developed in [10]; but the types considered here of finite element schemes may be \(h, p\), or \(h-p\) versions, including \(k\)-irregular meshes [3].

Let \(\mathcal{F}(\mathcal{P})\) denote the set of unconstrained or proper nodes in the partition \(\mathcal{P}\), see [3] for details. Let \(A \in \mathcal{F}(\mathcal{P})\) and let \(L_A\) be the degree one basis function associated with node \(A\). We suppose \(L_A\) to be scaled so that it takes the value 1 at the node \(A\). Let \(S_A\) denote the patch of elements forming the support of \(L_A\)

\[
(6.1) \quad S_A = \{ \Omega_k : \Omega_k \cap \text{supp}(L_A) \text{ is non-empty} \}
\]

and let \(Q_A\) denote the boundary segments lying on the interior of the support of \(L_A\) \((\text{supp}^0 L_A)\)

\[
(6.2) \quad Q_A = \{ \Gamma_{KL} : \Gamma_{KL} \subseteq \text{supp}^0(L_A) \text{ or } \Gamma_{KL} \subseteq \Gamma_e \cap \text{supp}(L_A) \}
\]

In the following, we shall show how \(\gamma_{KL}\) may be constructed such that the scalar.

\[
(6.3) \quad A_k = \gamma_k \int_{\Omega} c dx = \int_{\Omega} r_k dx + \frac{1}{2} \int_{\Gamma} R_k ds
\]

vanishes on all elements \(\Omega_k: \partial \Omega_k \cap \Gamma_D = \emptyset\).

Let \(A \in \mathcal{F}(\mathcal{P})\). For each \(\Omega_k \in S_A\) define

\[
(6.4) \quad R^A_k = \int_{\Omega} r_k L_A dx
\]

and for each \(\Gamma_{KL} \in Q_A\) define

\[
(6.5) \quad \rho^A_{KL} = \begin{cases} 
\int_{\Gamma_{KL}} \left( a \frac{\partial u_k}{\partial n} \right) \gamma L_A ds & K, L \neq 0 \\
\int_{\Gamma_{KL}} \left( a \frac{\partial u_k}{\partial n} - g \right) \gamma L_A ds & L = 0
\end{cases}
\]
With each boundary segment $\Gamma_{KL} \in \Gamma_A$ we associate a constant $\lambda_{KL}^A$. Notice that both $\Gamma_{KL}$ and $\Gamma_{LK} \in \Gamma_A$ and it will usually be found that $\lambda_{KL}^A \neq \lambda_{LK}^A$.

**Theorem 6.1.** For each $A \in \mathcal{F}(\mathcal{D})$ there exist $\{\lambda_{jL}^A\}_{r \in \Omega}$ such that

1. $\lambda_{jL}^A + \lambda_{jL}^A = 1 \ \forall \Gamma_{jL} \in \Omega_A$
2. $\lambda_{j0}^A = 1 \ \forall \Gamma_{j0} \in \Omega_A$
3. $\sum_{L : \Gamma_{r} \in \Omega_A} \lambda_{KL}^A \rho_{KL}^A = R_k^A \ \forall \Omega_K \in \Omega_A$

**Proof.** Let $\mu_{jL}^A = \lambda_{jL}^A \rho_{jL}^A \ \Gamma_{jL} \in \Omega_A$. With this notation, (i)–(iii) become (since $\rho_{jL}^A = \rho_{jL}^A$)

1. $\mu_{jL}^A + \mu_{jL}^A = \rho_{jL}^A$
2. $\mu_{j0}^A = \rho_{j0}^A$
3. $\sum_{L : \Gamma_{r} \in \Omega_A} \mu_{KL}^A = R_k^A$

We may eliminate the unknowns $\mu_{jL}^A \ L < J$ by using the first two conditions to give, for each $\Omega_K \in \Omega_A$,

$$
\sum_{L : \Gamma_{r} \in \Omega_A} \mu_{KL}^A = \sum_{L : \Gamma_{r} \in \Omega_A} \mu_{LK}^A = R_k^A = \sum_{L : \Gamma_{r} \in \Omega_A} \rho_{LK}^A + \sum_{L : \Gamma_{r} \in \Omega_A} \rho_{K0}^A
$$

This represents a linear system of $|\Omega_A|$ equations in the unknowns $\mu_{jL}^A \ L > J > 0 \ \Gamma_{jL} \in \Omega_A$. Let $M_A$ denote the underlying matrix for this system.

We examine the null space $\ker(M_A^*)$ of $M_A^*$. Suppose $\xi \in \mathbb{R}^{|\Omega_A|}$ is such that $M_A^* \xi = 0$. First, notice that each column of $M_A$ (and therefore each row of $M_A^*$) has precisely two non-zero entries corresponding to each unknown $\mu_{jL}^A \ J > L > 0$ occurring once in the equations when $\Omega_K = \Omega_J$ and $\Omega_K = \Omega_L$. Moreover, these non-zero entries are $+1$ and $-1$. Therefore

$$
M_A^* \xi = 0 \iff \xi_K = \xi_J \ \forall \Gamma_{KL} \in \Omega_A \ K > J > 0
$$

Since supp$(L_A)$ is connected this is now equivalent to all of the components of $\xi$ being the same constant. That is

$$
\ker(M_A^*) = \text{span}\{\lambda\}
$$

where $\lambda = (1, 1, \ldots, 1) \in \mathbb{R}^{|\Omega_A|}$.

By the Fredholm Alternative, a necessary and sufficient condition for the existence of solutions to (6.9) is that the data be orthogonal to $\ker(M_A^*)$. That is

$$
\sum_{\alpha \in S_k} R_A^k = \sum_{\alpha \in S_k} \sum_{L : \Gamma_{r} \in \Omega_A} \rho_{LK}^A + \sum_{\alpha \in S_k, r \in \Omega_A} \rho_{K0}^A
$$

Now

$$
R_k^A = \int_{\Omega_k} r^A L_A \, dx
$$

$$
= \int_{\Omega_k} f L_A \, dx - \int_{\partial \Omega_k} (a \nabla u_k \cdot \nabla L_A + cu_k L_A) \, ds + \int_{\partial \Omega_k} a \frac{\partial u_k}{\partial n_k} L_A \, ds
$$
and, hence, since $\bigcup_{\Omega_e \in \mathcal{E}, \Omega_K = \text{supp}(L_A)}$, we have

$$\sum_{\alpha_e \in \mathcal{E}} R^A_{\alpha_e} = (f, L_A) - a(u_h, L_A) + \sum_{\alpha_e \in \mathcal{E}, \partial \alpha_e} \int_{\partial \alpha_e} a(\frac{\partial u_h}{\partial n}) L_A \, ds$$

Applying Lemma 3.2 and noting $[L_A] = 0$, $\langle L_A \rangle = L_A$ gives

$$\sum_{\alpha_e \in \mathcal{E}} \int_{\partial \alpha_e} a(\frac{\partial u_h}{\partial n}) L_A \, ds = \sum_{\alpha_e \in \mathcal{E}, \partial \alpha_e \in \mathcal{Q}, \Gamma_{\alpha_e}} \int_{\partial \alpha_e} a(\frac{\partial u_h}{\partial n}) L_A \, ds$$

Moreover, by the definition of the Galerkin finite element solution, there holds

$$a(u_h, L_A) = (f, L_A) + \sum_{\alpha_e \in \mathcal{E}} \sum_{r_k \in \mathcal{Q}, r_{\infty}} \int_{r_k} g L_A \, ds$$

Combining these results gives

$$\sum_{\alpha_e \in \mathcal{E}} R^A_{\alpha_e} = \sum_{\alpha_e \in \mathcal{E}, \partial \alpha_e \in \mathcal{Q}, \Gamma_{\alpha_e}} \rho^A_{K,L} + \sum_{\alpha_e \in \mathcal{E}, \partial \alpha_e \in \mathcal{Q}, \Gamma_{\alpha_e}} \int_{\partial \alpha_e} a(\frac{\partial u_h}{\partial n}) L_A \, ds$$

which in view of (6.5) gives (6.10). It therefore follows that there exists a solution of (i)–(iii). \(\square\)

Applying the Fredholm Alternative once again, we deduce that there will be an infinite number of solutions in the case of $\Gamma_N \cap \text{supp}^0(L_A)$ being empty. Otherwise the solution will be unique.

The constants $\lambda^A_{K,L}$ are used to construct the $x_{K,L}$'s used in the average $\langle \cdot \rangle_2$ and $\langle \cdot \rangle_{1-x}$. Specifically, for each element $\Omega_K \in \mathcal{P}$ and for each $\Gamma_{K,L} \in \mathcal{Q}_A$, we define

$$\lambda_{K,L}(s) = \sum_{\alpha_e \in \mathcal{E}(\mathcal{P})} \lambda^A_{K,L} L_A(s)$$

Some special cases of (6.11) for 0- and 1-irregular meshes in two dimensions are shown in Figs. 1 and 2.

**Theorem 6.2.** Let $x_{K,L}$ be constructed as in (6.11). Then

1. $x_{K,L}(s) + x_{K,L}(s) \equiv 1$, $s \in \Gamma_{K,L} \cap \Gamma_D = \emptyset$
2. $\int_{\partial \alpha_e} r_k dx + \int_{\partial \alpha_e} R_k ds = 0$, $\forall \Omega_K : \tilde{\Omega}_K \cap \Gamma_D = \emptyset$

**Proof.** Notice that $\sum_{\alpha_e \in \mathcal{E}(\mathcal{P})} L_A(x) \equiv 1$, $x \in \Omega_K : \tilde{\Omega}_K \cap \Gamma_D = \emptyset$.

1. Let $s \in \Gamma_{K,L}$; then

$$x_{K,L}(s) + x_{K,L}(s) = \sum_{\alpha_e \in \mathcal{E}(\mathcal{P})} \lambda^A_{K,L} L_A(s) + \sum_{\alpha_e \in \mathcal{E}(\mathcal{P})} \lambda^A_{K,L} L_A(s)$$

$$= \sum_{\alpha_e \in \mathcal{E}(\mathcal{P})} (\lambda^A_{K,L} + \lambda^A_{K,L}) L_A(s) \equiv 1$$

since $\lambda^A_{K,L} + \lambda^A_{K,L} \equiv 1$. 


Unified approach to error estimation

Fig. 1. Flux splitting for 0-irregular mesh ($\lambda^A_{KJ}$ are as in Theorem 6.1).

\[ x_{12}(s) = \lambda^A_{12} L_A(s) + \lambda^B_{12} L_B(s) \]
\[ x_{23}(s) = \lambda^A_{23} L_A(s) + \lambda^B_{23} L_B(s) \]
\[ x_{34}(s) = \lambda^A_{34} L_A(s) + \lambda^B_{34} L_B(s) \]
\[ x_{41}(s) = \lambda^A_{41} L_A(s) + \lambda^B_{41} L_B(s) \]

2.

\[ \int_{\partial x} R_K ds = - \sum_{J : I_{KJ} \in E} \sum_{r_{KJ}} x_{KJ} \left[ a \frac{d u_h}{dn} \right] ds + \sum_{r_{KJ} \in E} \left( g - a \frac{d u_h}{dn} \right) ds \]

\[ = - \sum_{J : I_{KJ} \in E} \sum_{r_{KJ} \in \mathcal{E}(\varphi)} \sum_{r_{KJ} \in \mathcal{Q}} \lambda^A_{KJ} L_A \left[ a \frac{d u_h}{dn} \right] ds + \sum_{r_{KJ} \in E} \sum_{r_{KJ} \in \mathcal{Q}} \rho^A_{KJ} \]

\[ = - \sum_{A \in \mathcal{E}(\varphi)} \sum_{r_{KJ} \in \mathcal{Q}} R^A_K - \sum_{A \in \mathcal{E}(\varphi)} \int_{\partial x} r_K L_A dx = - \int_{\partial x} r_K ds \]

and, hence, \[ A_K = \int_{\partial x} r_K dx + \int_{\partial x} R_K ds = 0. \]

7. Equivalence of estimator

While Theorem 5.2 assures that our estimate will always bound the error, it is of importance to examine whether the estimator provides an equivalent measure of the error. Theorem 5.6 suggests that unless care is taken to control the quantity $\langle \partial e/\partial n \rangle_{1-2}$, the bound can be very poor. The aim of this section is to show that the construction for $x$ in Sect. 6 does control this quantity effectively.
In addition to the previous assumptions we shall further assume that $\Omega_K$ are convex subdomains. Let

\begin{equation}
\mathcal{N}_K = \{ \Omega_L \in \mathcal{P} : \partial \Omega_L \cap \partial \Omega_K = \Gamma_{LK} \text{ is non-empty} \}
\end{equation}

Then $\mathcal{N}_K$ is the set of elements neighboring $\Omega_K$. Let

\begin{equation}
l_{KJ} = |\Gamma_{KJ}| = \text{length of } \Gamma_{KJ}
\end{equation}

and suppose that there is a fixed constant $\kappa > 0$ such that for all $\Omega_K \in \mathcal{P}$:

(i)

\begin{equation}
\frac{1}{\kappa} \leq \frac{l_{KJ}}{h_K} \leq \kappa \quad \forall J \in \mathcal{N}_K
\end{equation}

This means that the meshes should be \textit{locally} quasi-uniform.

Furthermore, we assume there is a fixed constant $M$ such that

(ii)

\begin{equation}
\text{card}(\mathcal{N}_K) \leq M \quad \forall \Omega_K \in \mathcal{P}
\end{equation}

where $\text{card}(\mathcal{N}_K)$ indicates the number of elements in the set $\mathcal{N}_K$. This assumption does not exclude the case of $k$-irregular meshes [3], but it does force the degree of irregularity ($k$) to be finite.

(iii) Finally, we assume that the number of edges meeting at any given node is uniformly bounded independently of $\mathcal{P}$.

The finite element approximation is assumed to be a piecewise polynomial on each element $\Omega_K$, but we shall not assume the polynomial degree is constant. However, we suppose the maximum degree $\bar{p}$ to be bounded above, independently of $\mathcal{P}$. This assumption excludes the $p$ or $h$-$p$ versions proper (but is satisfied for essentially every practical implementation of these versions).

These assumptions are placed on the regularity of the mesh. The following represents an assumption on the regularity of the true solution $u$:

There exists a piecewise polynomial $\pi$ on $\mathcal{P}$ of degree at most $\bar{p} + 1$ such that

\begin{equation}
\| u - \pi \|^2_E + \left\| h_K^{1/2} \left\langle a \frac{\partial}{\partial n}(u - \pi) \right\rangle_{\Gamma_{KJ}} \right\|^2_{1, -s} \leq C \| e \|^2_E
\end{equation}

for some constant $C$ independent of $\mathcal{P}$. This is similar to the \textit{saturation assumption} made in [5], but the present version is weaker in that we do not assume $C \to 0$ as $\mathcal{P}$ is refined. Finally, due to (7.3), we assume that the following \textit{local inverse estimates} for $v \in \mathcal{P}_k(\Omega_K)$ are valid

\begin{equation}
\begin{aligned}
|v|_{0, r_n} &\leq C \| v \|_{0, \Omega_K} \\
\left| \frac{\partial v}{\partial n_K} \right|_{0, r_n} &\leq C \| v \|_{1, \Omega_K} \\
|v|_{2, \Omega_K} &\leq C h_K^{-1/2} \| v \|_{1, \Omega_K}
\end{aligned}
\end{equation}
and that the trace inequality holds.

\[ (7.7) \quad \left| \frac{\partial v}{\partial n} \right|_{0, r_x}^2 \leq C \left( \frac{1}{I_{k,j}} |v|_{1, \alpha_x}^2 + h_k |v|_{2, \alpha_x}^2 \right). \]

**Lemma 7.2.** Under the previous assumptions

\[ \left\| h_k^{1/2} \left\langle \frac{\partial e}{\partial n} \right\rangle_{1-\varepsilon, 0,E} \right\| \leq C \| e \|_E \]

for some constant $C > 0$.

**Proof.** By the Triangle Inequality

\[ \left\| h_k^{1/2} \left\langle \frac{\partial e}{\partial n} \right\rangle_{1-\varepsilon, 0,E} \right\| \leq \left\| h_k^{1/2} \left\langle \frac{\partial}{\partial n} (u - \pi) \right\rangle_{1-\varepsilon, 0,E} \right\| + \left\| h_k^{1/2} \left\langle \frac{\partial}{\partial n} (\pi - u_h) \right\rangle_{1-\varepsilon, 0,E} \right\| \]

Now

\[ \left\| h_k^{1/2} \left\langle \frac{\partial}{\partial n} (\pi - u_h) \right\rangle_{1-\varepsilon, 0,E} \right\|^2 \leq C h_k \left( \left\| \frac{\partial}{\partial n} (\pi - u_h) \right\|^2_{0, r_x} + \left\| \frac{\partial}{\partial n} (\pi - u_h) \right\|^2_{0, r_x} \right) \]

Using the Trace Inequality gives

\[ h_k \left\| \frac{\partial}{\partial n} (\pi - u_h) \right\|^2_{0, r_x} \leq C h_k \left( I_{k,j}^{-1} |\pi - u_h|_{1, \alpha_x}^2 + h_k |\pi - u_h|_{2, \alpha_x}^2 \right) \]

Summing over all edges gives

\[ \left\| h_k^{1/2} \left\langle \frac{\partial}{\partial n} (\pi - u_h) \right\rangle_{1-\varepsilon, 0,E} \right\| \leq C \| \pi - u_h \|_E \]

Moreover

\[ \| \pi - u_h \|_E \leq \| \pi - u \|_E + \| e \|_E \]

Therefore using (7.5)

\[ \left\| h_k^{1/2} \left\langle \frac{\partial e}{\partial n} \right\rangle_{1-\varepsilon, 0,E} \right\| \leq C \left( \left\| h_k^{1/2} \left\langle \frac{\partial e}{\partial n} (u - \pi) \right\rangle_{1-\varepsilon, 0,E} \right\| + \| \pi - u \|_E \right) + C \| e \|_E \leq C \| e \|_E \]

**Theorem 7.3.** Let $\varphi_K$ be a solution of the problem

\[ -\Delta \varphi_K = r_K \quad \text{in} \quad \Omega_K \]

subject to

\[ \gamma \varphi_K = 0 \quad \text{on} \quad \partial \Omega_K \cap \Gamma_D \]

\[ \frac{\partial \varphi_K}{\partial n_K} = \begin{cases} -\zeta_{k,j} \left[ \frac{\partial u_K}{\partial n} \right] & \text{on} \quad \partial \Omega_K \setminus \Gamma_N \\ g - \frac{\partial u_K}{\partial n} & \text{on} \quad \partial \Omega_K \cap \Gamma_N \end{cases} \]
(7.8) \[ \| e \|_E^2 \leq \sum_{K=1}^{N} \varepsilon^2_K(V\varphi_K) \leq C \| e \|_E^2 \]

where \( h = \max h_K \) and

\[ \varepsilon^2_K(V\varphi_K) = \frac{1}{\alpha_i} \int |V\varphi_K|^2 \, dx. \]

Proof. From Theorem 6.2 we have \( \delta_K = 0 \) on \( \Omega_K \): \( \partial \Omega_K \cap \Gamma_D = \emptyset \) and so

\[ \eta^2_K(V\varphi_K) = \varepsilon^2_K(V\varphi_K) + \frac{\delta^2_K}{\partial_i} \int c(x) \, dx = \varepsilon^2_K(V\varphi_K) \]

Applying Theorem 5.2 gives

\[ \| e \|_E^2 \leq \sum_{K=1}^{N} \eta^2_K(V\varphi_K) = \sum_{K=1}^{N} \varepsilon^2_K(V\varphi_K) \]

By Lemma 5.5 we have

\[ \varepsilon^2_K(V\varphi_K) \leq (1 + C h^2_K) \| e \|_E^2 \, \| e \|_{E,K} + C \| h^{1/2}_{K} \left( \frac{\partial e}{\partial n} \right)_{1-a} \|_{0,\partial K}^2 \]

and by Lemma 7.2 we then obtain

\[ \varepsilon^2_K(V\varphi_K) \leq (1 + C h^2_K) \| e \|_E^2 \, \| e \|_{E,K} + C \| e \|_{E,K}^2 \]

Summing over all elements gives

\[ \sum_{K=1}^{N} \varepsilon^2_K(V\varphi_K) \leq C \| e \|_E^2. \]

8. Summary and examples

The foregoing analysis can be regarded as consisting of two main sections.

The culmination of the first part is Theorem 5.2 which states that the error estimator generated using the local element residual method should always provide an upper bound on the true error, so long as the boundary conditions do not entail any loss in flux, that is to say, condition (3.21) holds. The most common type of element residual method is to choose the symmetrical splitting factor \( \frac{1}{2} \). However, further analysis reveals that the estimator, while bounding the error, can be very pessimistic unless the boundary conditions for the local problem are chosen carefully (see comments which follow (5.18)). The upper bound property was conjectured by Bank and Weiser [5], on the basis of numerical experiment for the case of piecewise linear approximation on triangles.

The second part of the work then focuses on the determination of boundary conditions used in the local problems, and, in particular, on the choice of splitting which determines the boundary conditions. It is shown that there exists splittings which mean that the term previously leading to gross overestimation will now vanish. One by-product of this work is that the “equilibration” used by Kelly [9] in one dimension is extended to higher dimensions, more general operators and irregular meshes.
Throughout we have assumed that the operator is of the form
\[ Lu \equiv -\nabla \cdot (a(x) \nabla u) + c(x)u \]
with \( c > 0 \). However, if the boundary conditions are chosen as suggested so that Theorem 6.2 is valid, then the restriction \( c > 0 \) can be relaxed, meaning that both the theory and the method extend to problems with no absolute terms such as Poisson problems.

In the one-dimensional case, the splitting factors were given explicitly by Kelly [9]. Suppose that the element \( I_K \) is the interval \((x_K, x_{K+1})\). In this case we need only choose the splitting at each node. Let
\[ J_K = \left[ au'_k(x_K) \right] \]
and define
\[ \alpha_{k,K-1} = \frac{R^K_{K-1}}{J_K}. \]

Using the standard orthogonal projection property of finite element approximation one can easily show that
\[ \alpha_{k-1,K} = 1 - \alpha_{k,K-1} = \frac{R^K_{K-1}}{J_K}. \]

This choice of splitting then leads to the satisfaction of the condition of the condition \( A_K = 0 \). In order to illustrate the necessity of employing the equilibration procedure, we consider the simple problem of finding \( u \):
\[ -u'' + u = f \quad \text{on } (0, 1) \]
subject to
\[ u(0) = u(1) = 0 \]
The function \( f \) is chosen so that the true solution is of the form
\[ u(x) = x^7 + 10(1-x)^8 - x - 10(1-x) \]
We present results of approximating this problem on uniform meshes with elements of uniform degree. The results in Table 1 show the effectivity indices (ratio of estimated to true error) in the case of symmetrical splitting (\( \alpha = \frac{1}{2} \)). The results for the cases \( p = 2 \) and \( p = 4 \) are seen to be unsatisfactory owing to the poor approximation to the boundary flux obtained using a simple averaging between neighboring elements. In Table 2, we give the corresponding effectivity indices for the splitting described above.
For the case of Poisson's equation, Kelly attempted to satisfy the condition $A_k = 0$ by means of a global minimization of the functional $\sum A_k^2$ over the splittings $\alpha$ subject to the condition (3.21). It was found that the objective functional could be driven to zero to machine accuracy in each case. This comes as no surprise in view of Theorem 6.2 above. Numerical results given by Kelly [9] show that the constant appearing in (7.8) is close to unity.

Acknowledgements. The support of DARPA and ONR under Grants N00014-89-J-1451 and N00014-89-J-3109 is gratefully acknowledged.

References