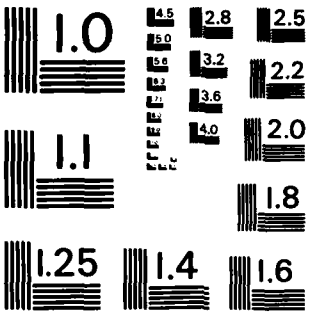
AD-A148 345 A FEEDBACK FINITE ELEMENT METHOD WITH A POSTERIORI ERROR ESTIMATION PART 1.. (U) MARYLAND UNIV COLLEGE PARK LAB FOR NUMERICAL ANALYSIS I BABUSKA ET AL. OCT 84 1/1

UNCLASSIFIED BN-1031 N00014-77-C-0623 F/G 12/1 NL

END

FILMED

DTIC



MICROCOPY RESOLUTION TEST CHART
NATIONAL BUREAU OF STANDARDS - 1963 - A

INSTITUTE FOR PHYSICAL SCIENCE
AND TECHNOLOGY

15

Laboratory for Numerical Analysis

Technical Note BN-1031

A FEEDBACK FINITE ELEMENT METHOD WITH A POSTERIORI ERROR ESTIMATION

PART I. THE FINITE ELEMENT METHOD AND SOME BASIC PROPERTIES OF THE A POSTERIORI ERROR ESTIMATOR

by

I. Babuška
Institute for Physical Science and Technology
University of Maryland

A. Miller
Centre for Mathematical Analysis
The Australian National University

October 1984

UNIVERSITY OF MARYLAND

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) A Feedback Finite Element Method with A Posteriori Error Estimation - Part I. The Finite Element Method and Some Basic Properties of the A Posteriori Error Estimator		5. TYPE OF REPORT & PERIOD COVERED Final lifeof the contract 6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) I. Babuška A. Miller		8. CONTRACT OR GRANT NUMBER(s) ONR N00014-77-C-0623
9. PERFORMING ORGANIZATION NAME AND ADDRESS Institute for Physical Science and Technology University of Maryland College Park, MD 20742		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Department of the Navy Office of Naval Research Arlington, VA 22217		12. REPORT DATE October 1984 13. NUMBER OF PAGES 68
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Approved for public release: distribution unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

20. ABSTRACT (Continue on reverse side if necessary and identify by block number) This paper is the first in a series of three in which we discuss some theoretical and practical aspect of a feedback finite element method for solving systems of linear second order elliptic partial differential equations (with particular interest in classical linear elasticity). In this first part we introduce some nonstandard finite element spaces, though based on the usual square bilinear elements, permit local mesh refinement. The algebraic structure of these spaces and their approximation properties are analyzed. An "equivalent estimator" for the H^1 finite element error is developed.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

S N 0102-LF-014-6601

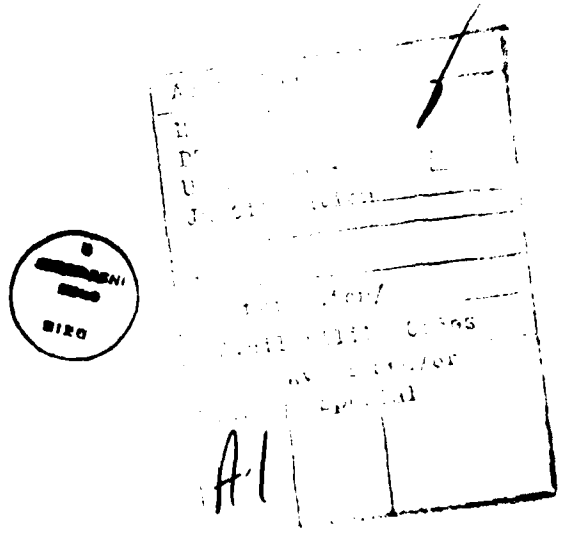
SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

A FEEDBACK FINITE ELEMENT METHOD WITH A POSTERIORI ERROR ESTIMATION

PART I. THE FINITE ELEMENT METHOD AND SOME BASIC PROPERTIES OF THE A POSTERIORI ERROR ESTIMATOR

I. Babuška

A. Miller



Technical Note BN-1031

October 1984

This research was partially supported by ONR Contract N00014-77-C-0623.

Abstract

This paper is the first in a series of three in which we discuss some theoretical and practical aspect of a feedback finite element method for solving systems of linear second order elliptic partial differential equations (with particular interest in classical linear elasticity). In this first part we introduce some nonstandard finite element spaces, though based on the usual square bilinear elements, permit local mesh refinement. The algebraic structure of these spaces and their approximation properties are analysed. An "equivalent estimator" for the H_n^1 finite element error is developed. In the second paper we shall discuss the asymptotic properties of this estimator. In the third paper we shall also report on some computational experience with the FEARS program which uses this estimator as part of a feedback loop to control mesh refinement and some of its programing features.

§0. Introduction

The practical success or failure of many finite element computations often depends critically on the user's choices of finite element mesh and element type. As a simple illustration of this, consider the boundary value problems that arise in classical plane linear elasticity. For such problems it is well known that in the neighbourhood of certain critial boundary points (e.g. angular boundary points, or points where the boundary conditions change between specified tractions and specified displacements), the stresses exhibit some form of singular behaviour. Unless such critical points are handled carefully, the resulting finite element solution may have disappointing accuracy.

One way in which such critical points may be treated is to employ an appropriate mesh refinement strategy in the neighbourhood of the point. Broadly speaking, two kinds of refinement strategy can be identified. On the one hand, there are a priori refinement techniques which grade the mesh in a manner governed either by earlier experience with similar kinds of problems or by the results of some a priori analysis of the nature of the singularity. For many problems in linear elasticity an asymptotic representation of the solution in the vicinity of the critical point is available. Using such representations it is often possible to derive a sequence of graded meshes which can be shown to converge in the energy norm at an optimal rate with respect to the number of degrees-of-freedom of the resulting discrete system. From a practical point of view however, a weakness of such a priori methods is that the analysis or experience they are based upon is asymptotic in nature and is seldom discriminating enough to tell whether a singularity, though present in theory, is going to cause significant problems at the level of accuracy that one is working. As an example consider the stress singularity

$Kr^{-1/2}g(\theta)$ typically associated with cracks in plane elasticity. The potential of this singularity to affect the accuracy of a finite element approximation will depend on K. Usually an a priori analysis can give little insight into the value of K. If K is small enough (compared to the overall level of accuracy desired), then no harm will be done by employing, say, a uniform mesh near the crack tip. Indeed, were a refined mesh used, the extra degrees of freedom would not lead to any significant improvement in accuracy over the uniform case. On the other hand, any (fixed) mesh refined near the crack tip will give a far from optimal mesh as $K \to \infty$. (Optimal here indicates a sense of minimum error for a given number of degrees of freedom). This kind of phenomenon becomes more pronounced when, as occurs in most practical problems, there are a number of critical boundary points in the region of interest, and a decision must be made on how the corresponding refinements should be "weighted."

The other kind of refinement strategy referred to earlier is a posteriori in character. In this approach an initial finite element solution is calculated using some mesh. This solution is then examined in some fashion, and based upon this examination a refinement of the initial mesh is decided upon. Using this new mesh, a new finite element solution is computed. This process can obviously be iterated until some stopping criterion is satisfied. This kind of feedback technique could conceivably avoid the problems mentioned above, since the computational significance of the singularities present may be able to be ascertained during the feedback process. Everything, of course, depends on the a posteriori examination carried out after each step and the refinement decision arising from it.

So far we have only mentioned mesh refinement necessitated by some form of singular behaviour of the solution. However in many other situations

proper mesh refinement can also be crucial. As a typical example consider the use of curved elements for solving problems in elasticity. In general such elements cannot exactly represent rigid body motion. Because of this, proper mesh refinement may again be essential for satisfactory results, even though the solution is very smooth. It would seem very difficult to predict the pattern of such refinement a priori.

In this series of papers we shall describe and provide an analysis of one such feedback approach. The approach we shall deal with has been implemented in a practical form in the FEARS program [see [1]]. This algorithm is based upon an a posteriori examination which involves the calculation of an a posteriori error estimator for the energy (or similar) norm of the error. This estimator is composed of elementwise error "indicators", and the refinement decision at each stage is made on the basis of the distribution of these indicators. The theoretical analysis of this method is far from complete. At the moment there are many cojectures, etc., which though convincingly demonstrated by many numerical experiments can either not be proved rigorously, or not be proved in the generality that practical experience would indicate they hold.

Our theoretical analysis will concentrate upon the error indicators and estimators, in particular, upon clarifying in what sense they "estimate" the energy norm of the error of the finite element solution. We shall show under quite weak assumptions that the estimator, E say, is an equivalent estimator for the energy norm $\|e\|$ of the error e, that is

$$C^{-1}E \leq \|e\| \leq CE \tag{1}$$

for some constant $C > 0$ uniformly over large class of meshes; and moreover under some further restrictions that E is also an asymtotically exact estimator for $\|e\|$, that is

$$\|e\| = E(1+o(\|e\|)) \tag{2}$$

as $\|e\| \to 0$.

The properties (1) and especially (2) suggest that the estimator can be reliably used for stopping the refinement process once some desired accuracy is achieved. Experience with FEARS confirms this.

To be considered worthwhile the process of succesive construction of meshes should lead to a sequence of meshes whose rate of convergence is comparable to that of the theoretically optimal mesh grading (at least when dealing with practical problems). They should also have other "optimal" properties. When a feedback process has such optimal properties, it is called an adaptive process. (For more about this see [2] [3] [4].) Experience indicates that the FEARS program implements an adaptive process.

In §1.1 of this paper we shall describe the kinds of boundary value problems that we wish to consider (essentially those related to linear elasticity). §1.2 contains a description of the finite element discretization to be employed. In §1.3-1.5 we give some properties of the corresponding finite element spaces for use later in the paper. In particular, we introduce the important concept of a K-mesh in §1.4. Although we describe a rather general set up in §1.1 and §1.2, the subsequent analysis is carried out in a more restricted setting. This has been done for the sake of clarity and simplicity of notation. The analysis of the general case can be done analogously. In §2 we derive an equivalent estimator for the energy norm of the finite element error. An important step in this is the basic error estimate of §2.1. Using this result the error is able to be localized to a small number of elements. Some technical lemmas are proved in §2.2 and §2.3, while §2.4 contains the main results of this section.

In the second paper of this series we shall deal with the asymptotic

exactness of the estimator, as well as discussing the overall performance of the algorithm.

The third part will deal more specifically with design of the FEARS program and analyse its performance in the light of the developed theory.

1. Formulation of the problem and its finite element solution

1.1. We shall consider the boundary value problem

$$L_i(u) \equiv -\sum_{j,k,\ell=1}^{2} D_k(a_{ijk\ell}D_\ell u_j) + \sum_{j=1}^{2} c_{ij}u_j = f_i \quad \text{in } \Omega$$

$$u_i = 0 \quad \text{on } \partial^0\Omega, \tag{1.1}$$

$$\sum_{k=1}^{2} \sigma_{ik}(u)n_k \equiv \sum_{k=1}^{2} \Big(\sum_{j,\ell=1}^{2} a_{ijk\ell}D_\ell u_j\Big)n_k = t_i \quad \text{on } \partial^1\Omega \quad (i = 1,2)$$

on a bounded domain $\Omega \subset \mathbb{R}^2$ whose boundary $\partial\Omega$ is made up of two disjoint parts $\partial^0\Omega \neq \emptyset$ and $\partial^1\Omega$; $\hat{n} = (n_1,n_2)$ is the outward pointing unit normal on $\partial\Omega$. This problem can be cast in a variational form. If $\partial^0\Omega$ is sufficiently smooth (piecewise smooth, say) we may define the trace of $H^1(\Omega)$ functions on $\partial^0\Omega$. Let us write

$$H = \{(v_1,v_2): v_i \in H^1(\Omega)\}$$

$$H_0 = \{(v_1,v_2): v_i \in H^1(\Omega),\ v_i = 0 \text{ on } \partial^0\Omega\}.$$

and H_0 are Hilbert spaces with respect to the norm $\|v\|_{1,\Omega} = \Big(\sum_{i=1}^{2} \|v_i\|_{1,\Omega}^2\Big)^{1/2}$, with $\|v_i\|_{1,\Omega}$ being the usual (scalar) $H^1(\Omega)$ Sobolev norm. The boundary value problem (1.1) can now be posed as: Find $u \in H_0$ such that

$$b(u,v) = \int_\Omega \sum_{i=1}^{2} f_i v_i \, dx + \int_{\partial^1\Omega} \sum_{i=1}^{2} t_i v_i ds, \quad \forall v \in H_0 \tag{1.2}$$

where the bilinear form $b: H_0 \times H_0 \to \mathbb{R}$ is defined by

$$b(w,z) = \int_\Omega \Big(\sum_{i,j,k,\ell=1}^{2} a_{ijk\ell}D_\ell w_j D_k z_i + \sum_{i,j=1}^{2} c_{ij}w_j z_i\Big)dx.$$

We shall assume that Ω can be represented as a collection of transformed unit squares. To this end we make the following assumption on Ω: There is a finite number of <u>subdomains</u> $\Omega_d \subseteq \Omega (d = 1,\ldots,N)$ such that

(i) $\Omega_d \cap \Omega_g = \emptyset \qquad (d,g = 1,\ldots,N;\ d \neq g)$

(ii) $\bar{\Omega} = \bigcup_{d=1}^{N} \bar{\Omega}_d$

(iii) There is an invertible transformation $\phi_d: \bar{\Omega}_d \to [0,1]^2$ which, together with is inverse ϕ_d^{-1}, is sufficiently smooth (bounded derivatives of all orders, say)

(iv) $\Omega_d = \phi_d^{-1}((0,1)^2)$.

The image under ϕ_d^{-1} of the closed edges and corners of $[0,1]^2$ will be referred to as the <u>edges</u> and <u>corners</u> respectively of Ω_d.

(v) Ω_d and $\Omega_e (d \neq e)$ can only have a single edge or a single corner in common (or else $\bar{\Omega}_d$ and $\bar{\Omega}_e$ are disjoint)

(vi) $\partial^0\Omega$ is the union of a number of (complete) edges of subdomains. We shall write $\partial_d^0(0,1)^2 = \phi_d(\partial^0\Omega \cap \partial\Omega_d)$ and $\partial_d^1(0,1)^2 = \phi_d(\partial^1\Omega \cap \partial\Omega_d)$.

An edge common to Ω_d and Ω_e will be called an <u>interface</u> between Ω_d and Ω_e.

(vii) If Γ is an interface between Ω_d and Ω_e, then ϕ_d and ϕ_e must "agree" on Γ in the following sense: if s is an arclength measured along Γ, and s_d and s_e are arclengths measured along $\phi_d(\Gamma)$ and $\phi_e(\Gamma)$ respectively, then regarding s_d and s_e as functions of s, either $s_d(s) = s_e(s)$ or $s_d(s) = 1 - s_e(s)$.

Figure 1 shows a possible partitioning of a circle into subdomains. For a suitable choice for the mappings ϕ_d in the case, see [5]. If Γ is an

interface between Ω_d and Ω_e, then we can naturally identify $\Omega_d(\Gamma)$ and $\Omega_e(\Gamma)$ as illustrated in Fig. 1. By virtue of (vii) above this identification takes a particularly simple form when expressed in terms of local arclength.

Figure 1. Partitioning of the circle into subdomains.

This representation of Ω induces a natural correspondence between H and the set M of N tuples of the form

$$V = (V^{(1)},\ldots,V^{(d)},\ldots,V^{(N)})$$

where

(i) $V^{(d)} = (V_1^{(d)},V_2^{(d)})$, $V_i^{(d)} \in H^1((0,1)^2)$ $(i = 1,2,\ d = 1,\ldots,N)$

(ii) if Γ is an interface between Ω_d and Ω_e $(d,e = 1,\ldots,N;\ d\neq e)$

$$v_j^{(d)} \circ \phi_d = v_j^{(e)} \circ \phi_e \quad \text{on } \Gamma . \qquad (j = 1,2)$$

This correspondence is defined by:

$$v = (v_1,v_2) \in H \to V = (\ldots,(v_1 \circ \phi_d^{-1}, v_2 \circ \phi_d^{-1}),\ldots) \in M,$$

and

$$W = (W^{(1)},\ldots,W^{(N)}) \in M \to w = (w_1,w_2) \in H$$

where

$$w_i = W_i^{(d)} \circ \phi_d \quad \text{on } \Omega_d, \qquad (i = 1,2, \quad d = 1,\ldots,N)$$

($w_i \in H^1(\Omega)$ by virtue of the interface continuity requirement (ii) above). Let $M_0 \subseteq M$ be the set of tuples satisfying the additional condition

(iii) If Γ is an edge of Ω_d and $\Gamma \subseteq \partial^0\Omega$ then

$$v_j^{(d)} \circ \phi_d = 0 \quad \text{on } \Gamma. \qquad (j = 1,2).$$

Clearly $M_0 \leftrightarrow H_0$ under the above natural correspondence. M and M_0 are Hilbert spaces with respect the norm

$$\|V\|_{1,(\Omega_1,\ldots,\Omega_d)} = \Big(\sum_{d=1}^{N} \sum_{i=1}^{2} \|v_i^{(d)}\|^2_{1,(0,1)^2}\Big)^{1/2}.$$

The problem (1.2) can now be reformulated as: Find $U \in M_0$ such that

$$B(U,V) = \sum_{d=1}^{N} \Big(\int_{(0,1)^2} \sum_{i=1}^{2} F_i^{(d)} v_i^{(d)} \, dx + \int_{\partial_d^1(0,1)^2} \sum_{i=1}^{2} T_i^{(d)} v_i^{(d)} \, ds \Big), \qquad \forall V \in M_0 \qquad (1.3a)$$

where the bilinear form $B: M_0 \times M_0 \to \mathbb{R}$ is defined by

$$B(W,Z) = \sum_{d=1}^{N} \int_{(0,1)^2} \Big(\sum_{i,j,k,\ell=1}^{2} A_{ijk\ell}^{(d)} D_\ell W_j^{(d)} D_k Z_i^{(d)} + \sum_{i,j=1}^{2} C_{ij}^{(d)} W_j^{(d)} Z_i^{(d)} \Big) dx. \qquad (1.3b)$$

In (1.3) we have used the notation

$$A_{ijk\ell}^{(d)} = \frac{1}{|E^{(d)}|} \sum_{s,t=1}^{2} (a_{ijst} \circ \phi_d^{-1}) E_{tk}^{(d)} E_{s\ell}^{(d)},$$

$$C_{ij}^{(d)} = \frac{1}{|E^{(d)}|} (c_{ij} \circ \phi_d^{-1}),$$

$$F_i^{(d)} = \frac{1}{|E^{(d)}|} (f_i \circ \phi_d^{-1}),$$

each defined on $(0,1)^2$; while on $\partial_d^1 (0,1)^2$

$$T_i^{(d)} = \frac{1}{E^{(d)}} (t_i \circ \phi_d^{-1}).$$

We have set $E_{ij}^{(d)} = D_i(\phi_d)_j$, $|E^{(d)}| =$ determinant $(E_{ij}^{(d)}: i,j = 1,2)$, and $E^{(d)} = \left(\sum_{i=1}^{2} \left(\frac{d}{ds} (\phi_d)_i\right)^2 \right)^{1/2}$ with $\frac{d}{ds}$ denoting differentiation with respect to arclength along $\partial_d^1 (0,1)^2$.

With regard to the coefficients and input data of (1.2) we shall suppose that the $a_{ijk\ell}$, c_{ij} and f_i are sufficiently smooth on each Ω_d separately (say, bounded derivatives of all orders), and that the t_i are sufficiently smooth on each subdomain edge contained in $\partial^1\Omega$.

We assume the symmetries

$$a_{ijk\ell} = a_{ji\ell k} \qquad (i,j,k,\ell = 1,2),$$

so ensuring that $b(\cdot,\cdot)$, is a symmetric bilinear form. Additionally, we shall suppose that the bilinear form $b(\cdot,\cdot)$ is coercive over H_0, that is, there exists $\alpha > 0$ such that

$$|b(w,w)| \geq \alpha \|w\|^2_{1,\Omega} \qquad \forall w \in H_0. \tag{1.4a}$$

Further we will assume that for some $C > 0$

$$\sum_{i,j=1}^{2} a_{ijkk}\xi_i\xi_j \geq C(\xi_1^2 + \xi_2^2), \qquad k = 1,2; \quad \xi \in \mathbb{R}^2. \tag{1.4b}$$

By virtue of our assumptions on the mappings ϕ_d, the above properties transfer naturally to the transformed system. Let us explicitly note the symmetry condition $B(U,V) = B(V,U)$ $\forall V,W \in M_0$, the coercivity condition

$$|B(W,W)| \geq \alpha \|W\|^2_{1,(\Omega_1,\ldots,\Omega_d)} \quad \forall W \in M_0,$$

and

$$\sum_{i,j=1}^{2} A_{ijkk}\xi_i\xi_j \geq C(\xi_1^2+\xi_2^2), \qquad k = 1,2; \quad \xi \quad \mathbb{R}^2.$$

Note that in the case when the $a_{ijk\ell}$ are discontinuous across an interface Γ between two subdomains, then the classical formulation (1.1) needs to be supplemented by an interface condition expressing the continuity of "tractions" across Γ. This condition is of course implicit in the variational formulations (1.2) and (1.3) of the problem.

The finite element approximation that we shall discuss is based upon the formulation (1.3) of the problem. If $\tilde{M}_0 \subset M_0$, the corresponding Galerkin approximation $\tilde{U} \in \tilde{M}_0$ to U is defined by

$$B(\tilde{U},V) = \sum_{d=1}^{N} \Big(\int_{(0,1)^2} \sum_{i=1}^{2} F_i^{(d)} v_i^{(d)} dx + \int_{\partial_d^1(0,1)} \sum_{i=1}^{2} T_i^{(d)} v_i^{(d)} ds \Big) \qquad \forall V \in \tilde{M}_0.$$

The coercivity of $B(\cdot,\cdot)$ ensures the existence and uniqueness of $\tilde{U}$. We have of course the projection property

$$\|\tilde{U}\|_{1,(\Omega_1,\ldots,\Omega_d)} \leq C\|U\|_{1,(\Omega_1,\ldots,\Omega_d)}. \tag{1.5}$$

The finite element error $U-\tilde{U}$ satisfies the usual orthogonality relation

$$B(U-\tilde{U},V) = 0 \qquad \forall V \in \tilde{M}_0 \tag{1.6a}$$

which leads to the standard kind of best approximation estimate

$$\|U-\tilde{U}\|_{1,(\Omega_1,\ldots,\Omega_d)} \leq C \inf_{V\in\tilde{M}_0} \|U-V\|_{1,(\Omega_1,\ldots,\Omega_d)}. \tag{1.6b}$$

The finite element approximation $\tilde{U}$ corresponds (under the natural correspondence between H and M) to an approximation $\tilde{u}$ to the solution u of (1.1), (1.2). In terms of the energy norms of the respective errors we have

$$(b(u-\tilde{u},u-\tilde{u}))^{1/2} = (B(U-\tilde{U},U-\tilde{U}))^{1/2}. \tag{1.7}$$

1.2. Suppose that a subdivision of Ω into subdomains $\Omega_1,\ldots,\Omega_N$ with corresponding mappings $\psi_1,\ldots,\psi_N$ satisfying the conditions of §1.1 has been decided upon. We shall now define what we mean by a mesh $\mathcal{D}$ on Ω.

A <u>mesh $\mathcal{D}$</u> is an N-tuple $(\mathcal{D}_1,\ldots,\mathcal{D}_N)$ where each $\mathcal{D}_d$ $(d = 1,\ldots,N)$ is a partition of $[0,1]^2$ into <u>closed</u> squares Δ (with edges parallel to the coordinate axes). Each $\mathcal{D}_d$ is called a <u>submesh</u> on Ω_d and either

(i) $\mathcal{D}_d = \mathcal{D}_d^0 = \{[0,1]^2\}$

or

(ii) $\mathcal{D}_d = \mathcal{D}_d^{(i)}$ is constructed from an existing submesh $\mathcal{D}_d^{(i-1)}$, by replacing any $\Delta \in \mathcal{D}_d^{(i-1)}$ by the four congruent

squares resulting from the simultaneous bisections of Δ in the two coordinate directions (see Fig. 2).

$$\mathcal{D}_1^{(0)} = \{[0,1]^2\} \quad \mathcal{D}_1^{(1)} = \{\Delta_1,\ldots,\Delta_4\} \quad \mathcal{D}_1^{(2)} = \{\Delta_2,\ldots,\Delta_8\}$$

Figure 2. The construction of a submesh on the subdomain Ω_1.

Each closed square $\Delta \in \mathcal{D}_d$ is called an element of $\mathcal{D}_d$. We shall use $|\Delta|$ to denote the length of a side of Δ. Clearly $|\Delta| = 2^{-s}$ with $s > 0$ an integer. Further we denote $h(\mathcal{D}_d) = \max_{\Delta \in \mathcal{D}_d} |\Delta|$.

A point P is called a node of $\mathcal{D}_d$ if either

(i) P is a vertex of an element of $\mathcal{D}_d$

or

(ii) there is an interface between Ω_d and Ω_e, and $\phi_e \circ \phi_d^{-1}(P)$ is a vertex of an element of $\mathcal{D}_e$.

Nodes P of $\mathcal{D}_d$ are classified as $\mathcal{D}$-proper if either

(i) $P \in (0,1)^2$ and whenever $P \in \Delta'$ for $\Delta' \in \mathcal{D}_d$, then P is a vertex of Δ'

or

(ii) $\phi_d^{-1}(P) \in \partial\Omega$

or

(iii) $\phi_d^{-1}(P)$ lies on an interface between Ω_d and Ω_e, P is the vertex of an element of $\mathcal{D}_d$, and $(\phi_e \circ \phi_d^{-1})(P)$ is the vertex of an element of $\mathcal{D}_e$.

element of $\mathcal{D}_e$.

If a node is not $\mathcal{D}$-proper, it is said to be $\mathcal{D}$-improper. The cases (i), (ii), (iii) above are clearly mutually exclusive, and we shall further classify the $\mathcal{D}$-proper nodes as interior, boundary or interface nodes depending upon whether (i), (ii) or (iii) applies. Fig. 3 shows an example of mesh with proper and improper nodes indicated.

Each of the straight line segments γ

Figure 3. A mesh on Ω ($\leftrightarrow$ denotes the natural identification of points across interfaces).

forming the boundary of an element of $\mathcal{D}_d$ will be called an edge. For definiteness an edge will be assumed to be closed, that is, it includes its endpoints. $|\gamma|$ will denote the length of an edge. An edge is called a primitive edge if it contains no nodes other than its endpoints.

We associate with a mesh $\mathcal{D}$ the finite element subspace $\tilde{M}(\mathcal{D})$ consisting of all tuples of the form

$$W = (W^{(1)},\ldots,W^{(N)})$$

where

(i) $W^{(d)} = (W_1^{(d)},W_2^{(d)})$, $W_j^{(d)} \in H^1((0,1)^2)$ $(j=1,2;\ d=1,\ldots,N)$

(ii) $W_j^{(d)}$ is bilinear on each element $\Delta \in \mathcal{D}_d$ $(j=1,2;\ d=1,\ldots,N)$

(iii) if Γ is an interface between Ω_d and Ω_e

$$W_j^{(d)} \circ \phi_d = W_j^{(e)} \circ \phi_e \quad \text{on } \Gamma, \quad (j=1,2;\ d,e=1,\ldots,N).$$

Clearly $\tilde{M}(\mathcal{D}) \subseteq M$. Define $\tilde{M}_0(\mathcal{D}) = \tilde{M}(\mathcal{D}) \cap M_0$.

The finite element subspaces we have introduced are non-standard since improper nodes are permitted. (Note however that the spaces remain conforming). These spaces permit local mesh refinement, yet maintain many of the desirable programming characteristics of the more usual square or rectangular elements.

Having defined a general framework, we shall in the analysis for the remainder of this paper only consider the particular case of one subdomain (N=1) and ϕ the identity mapping (so $\Omega = (0,1)^2$). This restriction is for notational simplicity only. Our results extend quite naturally to the case $N > 1$ and ϕ general, though with a considerable growth in notation. In the light of this simplification we shall from now on suppress the subdomain index and just write Ω, M, $\partial^0(0,1)^2$, etc. There is also no need now to distinguish between the original problem (1.2) and the transformed problem (1.3). For definiteness we shall from now on use the notation of the original problem.

To further contain notation, where no confusion is possible we shall not notationally distinguish between vector valued functions and their components. In such instances all operations, relations, etc., are to be understood in a componentwise sense.

1.3. On a matter of terminology we shall say that a line $[P_1,P_2]$ is a binary segment of another line $[Q_1,Q_2]$ if

$$P_1 = Q_1 + \frac{k}{2^m}(Q_2-Q_1)$$

$$m = 0,1,\ldots;\ k = 0,1,\ldots,2^m-1$$

$$P_2 = Q_1 + \frac{k+1}{2^m}(Q_2-Q_1).$$

Lemma 3.1.

(a) If Δ, Δ^* are distinct elements of $\mathcal{D}$ with $|\Delta| \leq |\Delta^*|$, then one and only one of the following holds:

(i) $\Delta \cap \Delta^* = \emptyset$

(ii) Δ and Δ^* share only one point. This point being a proper node (and hence also a vertex of both Δ and Δ^*).

(iii) $\Delta \cap \Delta^*$ is an edge of Δ. This edge is a binary segment of an edge of Δ^*.

(b) If Q is an improper node, then Q lies in the interior (i.e., is not an endpoint) of an edge of a unique element Δ^*. Furthermore, if Q is the vertex of $\Delta \in \mathcal{D}$ then $|\Delta| < |\Delta^*|$.

Proof. (a) These results follow readily by an induction based on the refinement process used to construct $\mathcal{D}$.

(b) Since Q is improper, there is an element Δ^* with $Q \in \Delta^*$ but Q is not a vertex of *. It is readily seen that * is unique. Case (iii) of (a) applies. Assuming that $| \; | \; | \; ^*|$ leads to a contradiction since Q is not a vertex of *. □

Lemma 3.2. Suppose $\mathcal{D}$ is a mesh and that $z \in L_2((0,1)^2)$ is bilinear on each $\Delta \in \mathcal{D}$. If z has a well defined limiting value at each interior node (i.e., z is continuous in these points) then $z \in \tilde{M}(\mathcal{D})$.

Proof. It suffices to show that $z \in C^0((0,1)^2)$. Let $x \in (0,1)^2$. If x lies in the interior of an element then z is obviously continuous at x. The other possibility is that x is a common boundary point of (at least) two elements, Δ and Δ^* say. There are two cases to consider here: (i) x is a vertex of Δ or Δ^*. In this case the hypothesis guarantees continuity. (ii) x is not a vertex of Δ or Δ^*. Case (iii) of Lemma 3.1(a) must then apply. Quite generally we may suppose that $|\Delta| \leq |\Delta^*|$, so $\Delta^* \cap \Delta$ is an edge, $[P_1,P_2]$ say, of Δ and is contained in an edge of Δ^*. The limit of z on $[P_1,P_2]$ from within Δ is a linear function, as is the limit from within Δ^*. But, by our hypothesis, at P_1 and P_2 these limits must agree. Thus they must agree througout $[P_1,P_2]$, and so in particular at x. The continuity of z at x follows. □

Theorem 3.3.

(a) For any proper node P there is a $\phi_P \in \tilde{M}(\mathcal{D})$ satisfying

$$\phi_P(Q) = \begin{cases} 1 & \text{if } Q = P \\ 0 & \text{if } Q \text{ is a proper node, } Q \neq P. \end{cases} \tag{3.1}$$

In addition $\phi_P \geq 0$ everywhere.

(b) If $\varphi \in \tilde{M}(\mathcal{D})$ and $\varphi(P) = 0$ for all proper nodes P, then $\varphi = 0$ (In particular, the ϕ_P are unique).

(c) $\{\phi_P: P \text{ a proper node}\}$ is a basis for $\tilde{M}(\mathcal{D})$. In fact, for any $\varphi \in \tilde{M}(\mathcal{D})$, $\varphi = \sum_P \varphi(P)\phi_P$ where $\sum_P$ denotes summation over all proper nodes P of $\mathcal{D}$.

Proof. (a) We shall construct a function $z \in L_2(0,1)^2$ which is biliear on each $\Delta \in \mathcal{D}$. Arrange the elements of $\mathcal{D}$ in order $\Delta_1,\ldots,\Delta_m$ where $|\Delta_j| > |\Delta_{j+1}|$ $(j = 1,\ldots,m-1)$. Suppose that $z > 0$ has been defined on each Δ_j for $j < n$. To define z on Δ_n it suffices to specify the limiting values of z from withing Δ_n at each vertex Q of Δ_n:

(i) If Q is a proper node, then set

$$z(Q) = \begin{cases} 1 & \text{if } Q = P \\ 0 & \text{if } Q \neq P \end{cases}$$

(ii) If Q is improper, then by Lemma 3.1(b), Q lies in the interior of an edge, $[P_1,P_2]$ say, of some unique element Δ^* with $|\Delta^*| > |\Delta_n|$. Thus $\Delta^* = \Delta_j$ for some $j < n$ and we may define $z(Q) = z^*(Q) > 0$ when $z^*(Q)$ is the limiting value of z at Q from within Δ^*.

Thus we are able to define $z > 0$ on Δ_n. It is clear from this method of construction that z has a well defined limiting value at each node of $\mathcal{D}$, and so by Lemma 3.2, $z \in \tilde{M}(\mathcal{D})$. Obviously if we set $\phi_P = z$, then $\phi_P > 0$ and (3.1) is satisfied.

(b) Again arrange the elements of $\mathcal{D}$ in order of nonincreasing size. Suppose that $\varphi = 0$ on all Δ; for $j < n$. Consider any vertex Q of Δ_n. If Q is proper then $\varphi(Q) = 0$. On the other hand, if Q improper, then again by Lemma 3.1(b), Q lies in the interior of an edge, $[P_1,P_2]$ say, of some unique element Δ^* with $|\Delta^*| > |\Delta_n|$. Thus $\Delta^* = \Delta_j$ with $j < n$. But $\varphi = 0$, on Δ_j, and so $\varphi(Q) = 0$. Thus $\varphi = 0$ at all vertices of Δ_n, and therefore $\varphi = 0$ on Δ_n.

(c) This follows readily from (a) and (b).

Corollary 3.4.

(a) $\{\phi_P: P \notin \partial^0(0,1)^2\}$ is a basis for $\tilde{M}_0(\mathcal{D})$.

(b) $\sum_P \phi_P = 1$, $|\phi_P| \leq 1$.

(c) On any $\Delta \in \mathcal{D}$, $|D_j\phi_P| \leq \frac{1}{|\Delta|}$. $(i = 1,2)$.

Proof. This follows readily from the Theorem 3.

Theorem 3.5.

(a) $\text{Supp}(\phi_P)$ = closure $\{x \in [0,1]^2: \phi_P(x) \neq 0\}$ is the union of a (whole) number of (closed) elements.

(b) $\text{Supp}(\phi_P)$ is "connected" in the sense that if Δ', $\Delta'' \subseteq$ supp ϕ_P then there exists a sequence of elements $\Delta' = \Delta_0,\ldots,\Delta_s = \Delta''$ such that:

(i) $\Delta_j \subseteq$ supp ϕ_P $(j = 0,\ldots,s)$

(ii) The pair Δ_j, Δ_{j+1} share an edge of the smaller of the pair $(j = 0,\ldots,s-1)$.

Proof. (a) This follows from the simple observation that if a bilinear function on an element Δ vanishes on $K \subseteq \Delta$, then either $K = \Delta$ or K is a (one dimensional) curve.

(b) Let supp $\phi_P = \bigcup_{k=1}^{r} \Delta_k$. Assume that $|\Delta_k| \geq |\Delta_{k+1}|$, $k = 1,\ldots,r-1$. It will suffice to show the result in the case $\Delta' = \Delta_1$.

First we show that one of the vertices of Δ_1 must be P. Suppose that this is not the case, and let Q be any vertex of Δ_1. If Q is proper, $\phi_P(Q) = 0$. So at least one vertex, Q say, of Δ_1 must be improper with $\phi_P(Q) \neq 0$. By Lemma 3.1(b), there therefore exists an element Δ^* with Q lying inside an edge of Δ^* and $|\Delta^*| > |\Delta_1|$. Since $Q \in \Delta^*$ with $\phi_P(Q) \neq 0$, it follows that $\Delta^* \subseteq$ supp ϕ_P. But this contradicts the maximality of Δ_1.

Next notice that any element with P as a vertex can be "connected" to Δ_1 either directly (if (iii) of Lemma 3.1(a) applies) or indirectly by way of one intermediate element (if (ii) of Lemma 3.1(a) applies).

We now prove the result by induction on k. Suppose that we can "connect" Δ_k to Δ_1 for all $k < j$, and consider Δ_{j+1}. If P is a vertex of Δ_{j+1}, then by what was said above, Δ_{j+1} can be "connected" to Δ_1. If P is not a vertex then at least one vertex, Q say, of Δ_{j+1} *must* be improper with $\phi_P(Q) \neq 0$. From Lemma 3.1(b) it follows that there exists $\Delta^* \in \mathcal{D}$ with $Q \in \Delta^*$, $|\Delta^*| > |\Delta_{j+1}|$ and $\Delta^* \cap \Delta_{j+1}$ an edge of Δ_{j+1}. Clearly $\Delta^* \subseteq \text{supp } \phi_P$, so $\Delta^* = \Delta_k$ for some $k < j$. Thus Δ_{j+1} can be "connected" to Δ_1 by appending Δ_{j+1} to the connecting chain for Δ_k. □

For any proper node P we shall refer to $\Omega_P = \text{Interior } (\text{supp}(\phi_P))$ as the star of P. We shall also make use of the following notation:

$$\mathcal{D}_P = \{\Delta \in \mathcal{D}: \Delta \subseteq \text{supp}(\phi_P)\},$$

$$\Gamma_P^* = \cup\{\gamma: \gamma \text{ and edge of some } \Delta \in \mathcal{D}_P, \ \gamma \not\subseteq \partial\Omega_P\},$$

$$\Gamma_P^1 = \partial\Omega_P \cap \partial^1(0,1)^2,$$

$$\Gamma_P^0 = \partial\Omega_P - \Gamma_P^1,$$

$$\Gamma_P = \Gamma_P^* \cup \Gamma_P^1.$$

The three tuple $\langle P, \mathcal{D}_P, \Gamma_P^0 \rangle$ will be called the <u>star tuple of P</u>. Note that the star uniquely determines Ω_P, Γ_P^*, Γ_P^1 and Γ_P. For instance

$$\Omega_P = \text{Int}\Big(\bigcup_{\Delta \in \mathcal{D}_P} \Delta\Big),$$

$$\Gamma_P^* = \bigcup_{\Delta \in \mathcal{D}_P} (\partial\Delta - \partial\Omega_P).$$

Furthermore observe that whether or not an edge contained in Γ_P is primitive can be determined from $\mathcal{D}_P$ alone.

Let us also introduce the notation

$$\Gamma^* = \cup\{\gamma: \gamma \text{ and edge of some } \Delta \in \mathcal{D},\ \gamma \not\subseteq \partial(0,1)^2\},$$

$$\Gamma = \Gamma^* \cup \partial^1\Omega.$$

Lemma 3.6.

(a) If $\Delta \in \mathcal{D}$ (and $\mathcal{D}$ contains more than one element), then $\Delta \in \mathcal{D}_P$ for some interior proper node P.

(b) $\Gamma^* = \bigcup_{P \text{ (all proper nodes)}} \Gamma_P^*$.

(c) $\partial^1\Omega = \bigcup_{P \text{ (all proper nodes)}} \Gamma_P^1$.

Proof. (a) An easy induction on the refinement process used to construct the mesh shows that for any element at least one vertex is an interior proper node.

(b) Clearly $\Gamma_P^* \subseteq \Gamma^*$, so $\bigcup_P \Gamma_P^* \subseteq \Gamma^*$. For the reverse inclusion: Suppose γ is an edge of Δ with $\gamma \not\subseteq \partial\Omega$. Let x be the midpoint of γ. By Corollary 4(b), $\sum_P \phi_P(x) = 1$. So for at least one proper node P, $\phi_P(x) \neq 0$. It follows that $\Delta \in \mathcal{D}_P$ and $\gamma \not\subseteq \partial\Omega_P$.

(c) Again it is obvious that $\bigcup_P \Gamma_P^1 \subseteq \partial^1\Omega$. To show the reverse. Let γ be an edge of $\Delta \in \mathcal{D}$. Then by (a) of the lemma, $\Delta \in \mathcal{D}_P$ for some proper node P. □

Suppose γ is a straight line segment which when extended to infinity in either direction divides R^2 into half planes $\pi^{(+)}$ and $\pi^{(-)}$. Let $x \in \gamma$

and suppose $v = (v_1, v_2)$ is a sufficiently smooth function defined in $S^+ = \pi^+ \cap \{y: |x-y| < \varepsilon\}$ and $S^- = \pi^- \cap \{y: |x-y| < \varepsilon\}$. Let $\hat{n}^+$ and $\hat{n}^{(-)}$ be the unit normals to γ pointing out of $\pi^{(+)}$ and $\pi^{(-)}$ respectively. We shall define $[\![v\cdot\hat{n}]\!]$ at x by

$$[\![v\cdot\hat{n}]\!] = \lim_{\substack{y\to x \\ y\in S^{(+)}}} v(y)\cdot\hat{n}^{(+)} + \lim_{\substack{y\to x \\ y\in S^{(-)}}} v(y)\cdot\hat{n}^{(-)}.$$

Lemma 3.7. Suppose $z \in C^0(\Omega_P)$ is bilinear on each $\Delta \in \mathcal{D}_P$. If $[\nabla z\cdot\hat{n}] = 0$ on each $\gamma \in \Gamma_P^*$ then z is bilinear in Ω_P.

Proof. Let Δ' be any element of $\mathcal{D}_P$. The function z is bilinear on Δ'. Extend this bilinear function to all of Ω_P, and call this extension z^*. We claim that $z = z^*$ on all elements of $\mathcal{D}_P$. To show this it suffices by Theorem 3.5(b) to show that if $z = z^*$ on Δ_1 and the pair Δ_1, Δ_2 share an edge, γ say, of the smaller of the pair, then $z = z^*$ on Δ_2. But this follows readily since (i) z is bilinear on Δ_2, (ii) z is continuous across γ, and (iii) $\gamma \in \Gamma_P^*$ and so the directional derivative of z normal to γ is continuous across γ. □

For any $\Delta \in \mathcal{D}$ define Ω_Δ, the influence region for Δ by

$$\Omega_\Delta = \text{Interior}\Big(\bigcup_{\substack{P \\ (\Delta\in\mathcal{D}_P)}} \text{supp}\ \phi_P\Big).$$

Introduce the notation $\mathcal{D}_\Delta = \{\Delta^* \in \mathcal{D}: \Delta^* \in \bigcup_{\substack{P \\ \Delta\in\mathcal{D}_P}} \text{supp}\ \phi_P\}$ and $\Gamma_\Delta = \{\gamma: \gamma$ an edge of some $\Delta^* \in \mathcal{D}_\Delta$, $\gamma \subseteq \partial^0\Omega\}$. The three tuple $\langle\Delta, \mathcal{D}_\Delta, \Gamma_\Delta\rangle$ will be called the influence tuple of Δ.

1.4. We shall now introduce a restriction on the "spread" of $\text{supp}(\phi_P)$. This restriction will be essential for our analysis later on.

For any set $S \subseteq \mathbb{R}^2$ let $|S| = \max_{i=1,2} \sup_{x,y \in S} |x_i - y_i|$.

Let $K > 1$. A mesh $\mathcal{D}$ is called a K-mesh if for each proper node P

$$|\text{supp}(\phi_P)| \leq K \min_{\Delta \in \mathcal{D}_P} |\Delta|. \tag{4.1}$$

For any $K > 1$ there are certainly meshes that are not K-meshes. For instance, for meshes of the type shown in Fig. 4(a), $\text{supp}(Q_P)$ is always the

Figure 4(a). Example of a mesh which is not a K-mesh.

shaded region, yet the minimum element size can be made arbitrarily small by continuing the refinement process sufficiently far. The K-mesh property is, in some sense, only a local property as it permits many natural forms of mesh grading. For instance, the grading pattern of Figs. 3(b) can be continued indefinitely without violating (4.1) with $K = 2$.

Figure 4(b). Example of a K-mesh with refinement.

[It is conjectured that the above definition of a K-mesh is equivalent to: There is $L > 1$ such that if $\Delta, \Delta' \in \mathcal{D}$ and $\Delta \cap \Delta' \neq \emptyset$ then

$$L^{-1} \leq |\Delta'|/|\Delta| \leq L. \tag{4.2}$$

By equivalence here we mean that if (4.1) holds then (4.2) holds with $L = L(K)$; and if (4.2) holds then (4.1) also holds with $K = K(L)$.]

In what follows we shall always assume that the mesh $\mathcal{D}$ is a K-mesh. We now state some important properties of K-meshes.

Lemma 4.1. Suppose $\mathcal{D}$ is a K-mesh, then there exist $C = C(K) > 0$ such that

(i) If $\Delta \in \mathcal{D}_P$, $\quad C|\Delta| \geq \max_{\Delta \in \mathcal{D}_P} |\Delta|$.

(ii) If P is a proper node, card $\mathcal{D}_P \leq C$.

(iii) If $\Delta \in \mathcal{D}$, card$\{P: \Delta \in \mathcal{D}_P\} \leq C$.

(iv) If γ is an edge of an element of $\mathcal{D}$, card$\{P: \gamma \subseteq \Gamma_P\} \leq C$.

(v) If P is a proper node, $\text{card}\{Q: \mathcal{D}_P \cap \mathcal{D}_Q \neq \emptyset\} \leq C$.

(vi) If $\Delta^* \in \mathcal{D}$, $\text{card}\{\Delta: \Delta^* \in \mathcal{D}_\Delta\} \leq C$.

(vii) If γ is an edge of an element of $\mathcal{D}$, then there are at most C nodes on γ.

Proof.

(i) $\max_{\Delta \in \mathcal{D}_P} |\Delta| \leq |\text{supp } \phi_P| \leq K \min_{\Delta \in \mathcal{D}_P} |\Delta|$.

(ii) $$\text{card}(\mathcal{D}_P) \leq \frac{|\text{supp } \phi_P|^2}{\min_{\Delta \in \mathcal{D}_P} |\Delta|^2} \leq K^2.$$

(iii) If $\Delta \in \mathcal{D}_P$, then $|\text{supp } \phi_P| \leq K \min_{\Delta^* \in \mathcal{D}_P} |\Delta^*| \leq K|\Delta|$. So $\bigcup_{P\ (\Delta \in \mathcal{D}_P)} (\text{supp } \phi_P) \subseteq Q_\Delta$ where Q_Δ is a square with centre at the center of Δ and side $2(K - 1/2)\ |\Delta|$.

Now if $\Delta' \in \bigcup_{P\ (\Delta \in \mathcal{D}_P)} \mathcal{D}_P$, then Δ' and Δ are elements of the same $\mathcal{D}_P$ for at least one proper node P. So

$$|\Delta'| > \frac{1}{K} |\text{supp } \phi_P| > \frac{1}{K} |\Delta|.$$

Thus $$\text{card}\Big(\bigcup_{P\ (\Delta \in \mathcal{D}_P)} \mathcal{D}_P\Big) \leq \frac{|Q_\Delta|^2}{(\frac{1}{K}|\Delta|)^2} \leq 4K^2(K - \tfrac{1}{2})^2.$$

But certainly $\text{card}\Big(\bigcup_{P\ (\Delta \in \mathcal{D}_P)} \mathcal{D}_P\Big) \geq \frac{1}{4} \text{card}(\{P: \Delta \in \mathcal{D}_P\})$ since P must be the vertex of at least one element of $\mathcal{D}_P$ and, of course, no element has more than four vertices. Thus

$$\text{card}\{P: \Delta \in \mathcal{D}_P\} \leq 16\, K^2(K - \tfrac{1}{2})^2.$$

(iv) γ can be an edge of at most two elements. By (iii) each of these elements is contained in $\mathcal{D}_P$ for at most C proper nodes P. Thus γ can be contained in Γ_P for at most $2C$ proper nodes P.

(v) By (ii) there are at most C elements in $\mathcal{D}_P$, while, by (iii), each such element in turn belongs to a $\mathcal{D}_Q$ for no more than C proper nodes Q. Thus $\mathcal{D}_P \cap \mathcal{D}_Q \neq \emptyset$ for at most C^2 proper nodes Q.

(vi) By (iii), Δ^* is contained in supp ϕ_P for at most C proper nodes P. For each such node in turn, supp ϕ_P can contain at most C elements Δ. Thus, $\text{card}\{\Delta: \Delta^* \in \mathcal{D}_\Delta\} \leq C^2$.

(vii) If γ is on $\partial(0,1)^2$ then the endpoints of γ are the only nodes on γ. Otherwise suppose γ is an edge of Δ. Since $\sum_P \phi_P = 1$, at, say, the midpoint of γ, then for at least one proper node P, $\phi_P > 0$ at the midpoint of γ. In fact it follows that $\phi_P > 0$ everywhere on γ, except perhaps at an endpoint. Any node on γ must then be a vertex of an element of $\mathcal{D}_P$. By (ii) the number of such elements is bounded. □

We would expect that many stars Ω_P are essentially identical except for a translation and scaling. Indeed, as we shall see in Theorem 3, there are in fact only a finite number (depending on K) of stars up to translation and rescaling. To set the scene for the result let us consider an arbitrary proper node P. Define an affine transformation $\tau_P: R^2 \to R^2$,

$$\tau_P(x) = \frac{1}{\left(\max_{\Delta \in \mathcal{D}_P} |\Delta|\right)} (x - P).$$

It is readily verified that $\tau_P(P) = 0$ and $\max_{\Delta \in \mathcal{D}_P} |\tau_P(\Delta)| = 1$. We shall consider τ_P to map the star of P onto some "standard" star. The mapping τ_P will be referred to as the star transformation of P. Let us set

$$\Omega_0 = \tau_P(\Omega_P),$$

$$_0 = \{\tau_P(\Delta): \Delta \in \mathcal{D}_P\},$$

$$\Gamma_0^* = \tau_P(\Gamma_P^*),$$

$$\Gamma_0^0 = \tau_P(\Gamma_P^0),$$

$$\Gamma_0^1 = \tau_P(\Gamma_P^1),$$

$$\Gamma_0 = \Gamma_0^* \cup \Gamma_0^1,$$

$$\phi_0 = \phi_P \circ \tau_P^{-1}.$$

The two-tuple $\langle \mathcal{D}_0, \Gamma_0^0 \rangle$ will be called the standard star tuple of P. With a harmless abuse of notation we shall refer to members of $\mathcal{D}_0$ as elements and their (closed) sides as edges. Notice that Ω_0, Γ_0^*, Γ_0^1 and Γ_0 can be reconstructed from $\langle \mathcal{D}_0, \Gamma_0^0 \rangle$ in much the same way as Ω_P, Γ_P^*, Γ_P^1 and Γ_P can be expressed in terms of star tuple of P. Somewhat less obvious is the following result.

Lemma 4.2. ϕ_0 is completely determined by $\langle \mathcal{D}_0, \Gamma_0^0 \rangle$.

Proof. This is clearly the same as saying that if $\mathcal{D}$ is a mesh and P is a proper node of $\mathcal{D}$, then ϕ_P is completely determined by $\langle P, {}_P, \Gamma_P^0 \rangle$. To show this, we shall construct a function $z \in \tilde{M}(\mathcal{D})$ using only our knowledge of $\langle P, \mathcal{D}_P, \Gamma_P^0 \rangle$, and then prove $z = \phi_P$:

Arrange the elements of $\mathcal{D}_P$ in order of nonincreasing size, $\Delta_1,\Delta_2,\ldots,\Delta_m$, say. Suppose that z has been defined on each Δ; for $j < n$. To define z on Δ_n it suffices to specify the limiting values of z from within Δ_n at each vertex Q of Δ_n:

(I) If $Q = P$, then $z(Q) = 1$.

(II) If $Q \neq P$:

(i) If $Q \in \partial\Omega_P$, then $z(Q) = 0$.

(ii) If $Q \in \Omega_P$ and Q is proper node (in the context of the $\langle P,\mathcal{D}_P,\Gamma_P^0\rangle$ this just means that whenever $Q \in \Delta^*$ for $\Delta^* \in \mathcal{D}_P$ then Q is a vertex of Δ^*) then $z(Q) = 0$.

(iii) If $Q \in \Omega_P$ and Q is improper (just as in (ii), this can be determined from $\mathcal{D}_P$ alone), then by Lemma 3.1(b) Q lies in the interior of an edge $[P_1,P_2]$, say, of some unique element Δ^* with $|\Delta^*| > |\Delta_n|$. In our present case $\Delta^* \in \mathcal{D}_P$ and so $\Delta^* = \Delta_j$ for some $j < n$ and we may define $z(Q) = z^*(Q)$ where $z^*(Q)$ is the limiting value of z at Q from within Δ^*.

This enables us to define z on each $\Delta \in \mathcal{D}_P$. Define z on the remaining elements of $\mathcal{D}$ to be identically zero. By the above construction z has a well defined limiting value at each node of $\mathcal{D}$, and so by Lemma 3.2, $z \in \tilde{M}_{(\mathcal{D})}$. However z satisfies (3.1), and so by Theorem 3.3, $z = \phi_P$. □

Theorem 4.3. There are at most $C = C(K)$ standard star tuples $\langle\mathcal{D}_0,\Gamma_0^0\rangle$.

Proof. It will suffice to show that there are at most C possibilities for $\mathcal{D}_0$, since once $\mathcal{D}_0$ is known there are only C possibilities for Γ_0^0. ($\mathcal{D}_0$ has at most C elements and each such element has no more than four edges. Γ_0^0 must be the union of some of these edges.) Since

$$\left| \bigcup_{\Delta \in \mathcal{D}_0} \Delta \right| \leq K \min_{\Delta \in \mathcal{D}_0} |\Delta| \leq K,$$

$\bigcup_{\Delta \in \mathcal{D}_0} \Delta$ must be contained in a square with centre 0 and side 2K.

Let m be an integer such that $2^m > K$. Let Q be a square with centre 0 and side 2^{m+1}. Define a uniform grid on Q of grid size 2^{-m}. We claim that each $\Delta \in \mathcal{D}_0$ consists of a (whole) number of grid squares. If we can show this, then the theorem will follow at once by simple combinatoric considerations.

Firstly note that for any $\Delta \in \mathcal{D}_0$, certainly $|\Delta| = 2^{-\xi}$ for some integer $\xi \geq 0$. Moreover,

$$|\Delta| \geq \min_{\Delta' \in \mathcal{D}_0} |\Delta'| \geq \frac{1}{K} \left| \bigcup_{\Delta' \in \mathcal{D}_0} \Delta' \right| \geq 2^{-m}$$

and, of course,

$$|\Delta| \leq \max_{\Delta' \in \mathcal{D}_0} |\Delta'| = 1.$$

In particular, if $(0,0)$ is a vertex of Δ, then Δ must consist of a whole number of grid squares. To prove this result for an arbitrary element of $\mathcal{D}_0$, it will suffice by Theorem 3.5(b) to show that whenever the pair $\Delta^*, \Delta^{**} \in \mathcal{D}_0$ are connected in the sense of Theorem 3.5(b) and Δ^* consists of a whole number of grid squares, then so also Δ^{**}. But this follows readily on recalling (iii) of Lemma 3.1. □

We also shall need standard forms for the influence tuples. To this end for each $\Delta \in \mathcal{D}$ define an affine transformation $\tau_\Delta : \mathbb{R}^2 \to \mathbb{R}^2$

$$\tau_\Delta(x) = \frac{1}{|\Delta|} \left(x - x_\Delta + \left(\frac{1}{2}, \frac{1}{2}\right) \right)$$

where x_Δ is the centre of Δ. It is easily seen that $\tau_\Delta(\Delta) = [0,1]^2$. Just as for the case of star tuples we shall consider τ_Δ as mapping the influence tuple of Δ onto some standards reference tuple. Let us set

$$\Omega_* = \tau_\Delta(\Omega_\Delta),$$

$$\mathcal{D}_* = \{\tau_\Delta(\Delta^*): \Delta^* \in \mathcal{D}_\Delta\},$$

$$\Gamma_* = \tau_\Delta(\Gamma_\Delta).$$

The two tuple $\langle \mathcal{D}_*, \Gamma_* \rangle$ will be called the standard influence tuple of Δ, and clearly Ω_* can be reconstructed from the standard tuple:

$$\Omega_* = \mathrm{Int}\Big(\bigcup_{\Delta \in \mathcal{D}_*} \Delta \Big).$$

Akin to Theorem 4.3 we have

Theorem 4.4. There are at most $C = C(K)$ standard influence tuples $\langle \mathcal{D}_*, \Gamma_* \rangle$.

Proof. It will suffice to show that there are at most C possibilities for $\mathcal{D}_*$, since once $\mathcal{D}_*$ is known there are only C possibilities for Γ_*. (By (ii) and (iii) of Lemma 4.1, $\mathcal{D}_*$ has at most C elements; Γ_* must be some union of the edges of these elements.)

To prove the claim for $\mathcal{D}_*$, we apply an almost identical argument to that used in Theorem 4.3. □

1.5. For later use we shall need to know something about the approximation properties of the finite element subspaces $\tilde{M}(\mathcal{D})$ and $\tilde{M}_0(\mathcal{D})$. In this section we shall prove some results in this direction. The methods of proof are more or less standard, the only real difference arising from the fact that the star

Ω_p may spread further than just those elements adjacent to P. The K-mesh assumption however restricts this spread and allows us to retrieve much of the standard theory.

Lemma 5.1. Let $\langle \mathcal{D}_*, \Gamma_* \rangle$ be a standard influence tuple. There is a constant $C = C(\mathcal{D}_*, \Gamma_*) > 0$ such that if $z \in H^r(\Omega_*)$ $(r = 1,2)$ and $z = 0$ on Γ^* then

$$\inf_{q \in Q} \|z-q\|_{r,\Omega_*} \leq C|z|_{r,\Omega_*} \tag{5.1}$$

where Q is the set of all polynomials of degree $r-1$ on Ω_* which vanish on Γ_*. By $\|\cdot\|_{r,\Omega}(|\cdot|_{r,\Omega})$ we denote the usual Sobolev norm (seminorm).

Proof. Suppose not, then we can find $z^{(n)}$ $(n = 1,2,\ldots)$

$$\inf_{q \in Q} \|z^{(n)}-q\|_{r,\Omega_*} > n|z^{(n)}|_{r,\Omega_*}. \tag{5.2}$$

Without loss of generality we may as well suppose that,

$$\inf_{q \in Q} \|z^{(n)}-q\|_{r,\Omega_*} = 1 \tag{5.3}$$

(by taking suitable multiples of the original $z^{(n)}$ if necessary) and

$$\|z^{(n)}\|_{r,\Omega_*} < 2 \tag{5.4}$$

(by adding a suitable $q^n \in Q$ to each of the original $z^{(n)}$ if necessary).

From (5.4) we can conclude using Rellich's Lemma that a subsequence of the $z^{(n)}$ converges in $H^{r-1}(\Omega_*)$. We can suppose that this subsequence is the entire sequence (by deleting members of the original sequence if necessary). But from (5.2), $\lim_{n\to\infty} |z^{(n)}|_{r,\Omega_*} = 0$. Thus $z^{(n)}$ converges in $H^r(\Omega_*)$. Let $z^{(\infty)}$ be the limit. Obviously, $|z^{(\infty)}|_{r,\Omega_*} = 0$, and as Ω_* is connected

(this fact follows readily from Theorem 3.5(b)), the only possibility is for $z^{(\infty)}$ to be a polynomial of degree $r-1$ on Ω_*. Moreover, $z^{(\infty)} = 0$ on Γ_* (since taking traces is a continuous operation in $H^r(\Omega^*)$). Thus $z^{(\infty)} \in Q$. But this contradicts the limiting form of (5.3). □

Theorem 5.2. There is a constant $C > 0$ such that for any $z \in H^r(\Omega) \cap H_0$ $(r = 1,2)$ and any K-mesh $\mathcal{D}$, there exists a function $\pi z \in \tilde{M}_0(\mathcal{D})$ such that

$$|z-\pi z|_{s,\Delta} \leq C|\Delta|^{r-s}|z|_{r,\Omega_\Delta}, \qquad (s = 0,1; \quad \Delta \in \mathcal{D}). \tag{5.1}$$

Proof. Let $p(t)$ be a polynomial in one variable satisfying $\int_0^1 p(t)dt = 1$ and $\int_0^1 tp(t)dt = 0$. For any $\varepsilon > 0$ define

$$\phi_\varepsilon = \begin{cases} p(t/\varepsilon) & t \in [0,\varepsilon] \\ p(-t/\varepsilon) & t \in [-\varepsilon,0] \\ 0 & |t| > \varepsilon. \end{cases}$$

For any proper node P of $\mathcal{D}$ set

$$\varepsilon_P = \frac{1}{2} \min_{\tilde{\Delta} \in E_P} |\tilde{\Delta}|$$

and $S_P = \{x \in \Omega: |x_i - P_i| < \varepsilon_P, \ i = 1,2\}$. Our choice of ε_P guarantees $S_P \subseteq \Omega_P$. For any $y \in H^1(\Omega)$ let

$$Y_P = \int_{S_P} y(x)\phi_{\varepsilon_P}(x_1-P_1)\Phi_{\varepsilon_P}(x_2-P_2)\,dx \,/\, \int_{S_P} dx \tag{5.2}$$

and define

$$\pi y = \sum_{\substack{P \\ P \notin \partial^0\Omega}} Y_P \, \phi_P.$$

Clearly (Corollary 3.4(a)) $\pi y \in \tilde{M}_0(\mathcal{D})$.

We shall now prove (5.1). Suppose $\Delta \in \mathcal{D}$ and let $\langle \mathcal{D}_*, \Gamma_* \rangle$ be the standard influence tuple of Δ, and, as usual, write $\Omega_* = \text{Interior}\left(\bigcup_{\Delta \in \mathcal{D}_*} \Delta\right)$. Then

$$|\pi y|_{s,\Delta} \leq \sum_{\substack{P \\ P \nmid \partial^0 \Omega}} |Y_P| |\phi_P|_{s,\Delta}$$

$$\leq C \sum_{\substack{P \\ (P \nmid \partial^0 \Omega \\ \Delta \in \mathcal{D}_P)}} |Y_P| |\Delta|^{1-s}$$

by Corollary 3.4 (b) and (c). However, from (5.2)

$$|Y_P| \leq \varepsilon_P^{-1} \|y\|_{0,S_P} \leq C|\Delta|^{-1} \|y\|_{0,S_P}$$

whenever $\Delta \in \mathcal{D}_P$. Thus

$$|\pi y|_{s,\Delta} \leq C \sum_{\substack{P \\ (P \nmid \partial^0 \Omega \\ \Delta \in \mathcal{D}_P)}} \|y\|_{0,S_P} |\Delta|^{-s} \leq C|\Delta|^{-s} \|y\|_{0,\Omega_\Delta} \tag{5.3}$$

by (iii) of lemma 4.1 and since $S_P \subseteq \Omega_P \subseteq \Omega_\Delta$ when $\Delta \in \mathcal{D}_P$. Upon rescaling Δ to the unit square, (5.3) becomes

$$|\pi y \circ \tau_\Delta^{-1}|_{s,(0,1)^2} \leq C \|y \circ \tau_\Delta^{-1}\|_{0,\Omega_*}$$

and so

$$|y \circ \tau_\Delta^{-1} - \pi y \circ \tau_\Delta^{-1}|_{s,(0,1)^2} \leq C \|y \circ \tau_\Delta^{-1}\|_{r,\Omega_*}. \tag{5.4}$$

If ζ is a polynomial of degree $r-1$ on Ω_Δ which vanishes on Γ_Δ, then ζ can trivially be extended to all of Ω. Direct calculation shows that $\pi\zeta = \zeta$. Thus

$$|z \circ \tau_\Delta^{-1} - \pi z \circ \tau_\Delta^{-1}|_{s,(0,1)^2} = \inf_\zeta |(z+\zeta) \circ \tau_\Delta^{-1} - \pi(z+\zeta) \circ \tau_\Delta^{-1}|_{s,(0,1)^2}$$
$$\leq \inf_\zeta \|(z+\zeta) \circ \tau_\Delta^{-1}\|_{r,\Omega_*} \tag{5.5}$$

by (5.4). But clearly as ζ ranges over all possibilities $\zeta \circ \tau_\Delta^{-1}$ ranges over all polynomials of degree $r-1$ on Ω_* which vanish on Γ_*. Thus by the result of Lemma 5.1,

$$\inf_\zeta \|z \circ \tau_\Delta^{-1} - \zeta \circ \tau_\Delta^{-1}\|_{r,\Omega_*} \leq C|z \circ \tau_\Delta^{-1}|_{r,\Omega_*}. \tag{5.6}$$

Combining (5.5) and (5.6), and rescaling back to Δ gives the desired result (5.1) upon noting that the constant in (5.6) can be taken to be mesh independent by Theorem 4.4. □

Corollary 5.3. There is a constant $C > 0$ such that for any K-mesh $\mathcal{D}$,

(i) If $z \in H^r(\Omega) \cap M_0$ $(r = 1,2)$.

$$\|z-\pi z\|_{s,\Omega} \leq Ch(\mathcal{D})^{r-s}\|z\|_{r,\Omega}, \qquad (s = 1,2)$$

(ii) If $z \in H^1(\Omega) \cap M_0$

$$\sum_P \|\phi_P(z-\pi z)\|^2_{1,\Omega_P} \leq C\|z\|^2_{1,\Omega}.$$

Proof.

(i) It follows directly from the theorem that

$$\|z-\pi z\|_{s,\Omega} \leq Ch(\mathcal{D})^{r-s} \Big(\sum_{\Delta} \|z\|^2_{r,\Omega_\Delta} \Big)^{1/2}$$

$$\leq Ch(\mathcal{D})^{r-s} \|z\|_{r,\Omega}$$

by (vi) of Lemma 4.1.

(ii) If $\Delta \in \mathcal{D}_p$, then the theorem and Corollary 3.4(b) and (c) shows

$$\|\phi_p(z-\pi z)\|_{1,\Delta} \leq C\Big(\max_{i=1,2} \sup_{x\in\Delta} |D_i\phi_p| \|z-\pi z\|_{0,\Delta} + \sup_{x\in\Delta} |\phi_p| |z-\pi z|_{1,\Delta} \Big).$$

On the other hand, if $\Delta \notin \mathcal{D}_p$ then obviously

$$\|\phi_p(z-\pi z)\|_{1,\Delta} = 0.$$

Fixing Δ, squaring and summing over all proper nodes P gives

$$\sum_{P} \|\phi_p(z-\pi z))\|^2_{1,\Delta} \leq C\|z\|^2_{1,\Omega_\Delta}$$

by (iii) of Lemma 4.1. The result now follows, just as in (i), from (vi) of Lemma 4.1. □

We shall also need a lower bound on the approximation power of $M(\mathcal{D})$. To this end we prove the following lemma.

Lemma 5.4. Suppose that $z \in H^1(\Omega)$ and that there is an open disc $D \subseteq \Omega$ where

(i) $z \in C^\infty(\overline{D})$,

(ii) $$\max_{i=1,2} \inf_{x\in\overline{D}} |D_{ii}z(x)| > 0. \qquad (5.7)$$

Suppose that the mesh $\mathcal{D}$ satisfies for some $\kappa > 0$

$$\min_{\substack{\Delta\in\mathcal{D}\\ \Delta\subseteq\overline{D}}} |\Delta| \;>\; \kappa h(\mathcal{D}). \tag{5.8}$$

Then there are constants C, h_0 such that if $h(\mathcal{D}) < h_0$,

$$\inf_{v\in\tilde{M}(\mathcal{D})} \|z-v\|_{1,\Omega} \;>\; Ch(\mathcal{D}).$$

Proof. Certainly

$$\inf_{v\in\tilde{M}(\mathcal{D})} \|z-v\|^2_{1,\Omega} \;>\; \sum_{\Delta\in\mathcal{D}} \inf_{\substack{w\\ (w \text{ bilinear on } \Delta)}} \|z-w\|^2_{1,\Delta}$$

$$> \sum_{\substack{\Delta\in\mathcal{D}\\ \Delta\subseteq\overline{D}}} \inf_{\substack{w\\ (w \text{ bilinear on } \Delta)}} \|z-w\|^2_{1,\Delta}. \tag{5.9}$$

On each element $\Delta \subseteq \overline{D}$, we can expand z as a Taylor series about the centre, $\bar{x}$ say, of Δ,

$$z(x) \;=\; \bar{w}(x) + \tfrac{1}{2}\sum_{i=1}^{2} D_{ii}z(\bar{x})\,(x_i-\bar{x}_i)^2 + r(x) \qquad (x \in \Delta)$$

where $\bar{w}$ is bilinear on Δ and

$$|r(x)| \;<\; C|x-\bar{x}|^3, \qquad |D_i r(x)| \;<\; C|x-\bar{x}|^2 \qquad (i = 1,2)$$

for some constant C, independent of the particular Δ under consideration. Writing

$$\bar{z}(x) \;=\; \bar{w}(x) + \tfrac{1}{2}\sum_{i=1}^{2} D_{ii}z(\bar{x})(x_i-\bar{x}_i)^2$$

we obtain using the finite dimension of the classes of functions involved

$$\inf_{\substack{w \\ w \text{ bilinear on } \Delta}} \|\bar{z}-w\|_{1,\Delta} > C\Big(\sum_{i=1}^{2} |D_{ii} z(\bar{x})|\Big)|\Delta|^2$$

$$> C^*|\Delta|^2$$

where $C^* > 0$ is independent of Δ by (5.7). Hence

$$\inf_{\substack{w \\ (w \text{ bilinear on } \Delta)}} \|z-w\|_{1,\Delta} > \inf_{\substack{w \\ (w \text{ bilinear on } \Delta)}} \|\bar{z}-w\|_{1,\Delta} - \|r(x)\|_{1,\Delta}$$

$$> C^*|\Delta|^2 - C|\Delta|^3$$

$$> C|\Delta|^2$$

for $h(\mathcal{D}) < h_0$, and hence $|\Delta|$, sufficiently small. Therefore (5.9) enables us to say

$$\inf_{v \in \tilde{M}(\mathcal{D})} \|z-v\|^2_{1,\Omega} > C \sum_{\substack{\Delta \in \mathcal{D} \\ \Delta \in \bar{D}}} |\Delta|^4$$

$$> Ch(\mathcal{D})^2 \sum_{\substack{\Delta \in \mathcal{D} \\ \Delta \in \bar{D}}} |\Delta|^2 \qquad (5.10)$$

using (5.8). Note however, that for $h(\mathcal{D})$ small enough, $D^* \subseteq \bigcup_{\substack{\Delta \in \mathcal{D} \\ \Delta \subseteq \bar{D}}} \Delta$ where D^* is a disc concentric with D but of radius half that of D. Thus $\sum_{\substack{\Delta \in \mathcal{D} \\ \Delta \subseteq \bar{D}}} |\Delta|^2 >$ area D^*, and the lemma follows at once from (5.10).

Let us remark that the hypotheses of the lemma are not very demanding at all.

2. The A-posteriori Error Estimate

2.1. For each proper node P of $\mathcal{D}$, define

$$M_P = \{v \in H^1(\Omega_P): v = 0 \text{ on } \Gamma_P^0\}.$$

(Clearly by extending functions in M_P by zero to the remainder of Ω we can consider $M_P \subseteq H_0$). We shall associate the following local subproblem with the proper node P: Find $\eta_P \in M_P$ such that

$$b(\eta_P, v) = b(u-\tilde{u}, v), \qquad v \in M_P. \tag{1.1}$$

After an integration by parts the right hand side of (1.1) can be written out explicitly in terms of the input data of (1.1.1):

$$\begin{aligned} b(u-\tilde{u}, v) = & \sum_{\Delta \in \mathcal{D}_P} \int_\Delta (f-L(\tilde{u}))v \, dx - \int_{\Gamma^*} \sigma(\tilde{u})\cdot\hat{n}\, v \, ds \\ & + \int_{\partial'\Omega} (t-\sigma(\tilde{u})\cdot\hat{n})v \, ds. \end{aligned} \tag{1.2}$$

Note also that as long as $P \notin \partial^0(0,1)^2$, then $\phi_P \in \tilde{M}_0(\mathcal{D})$, and that therefore

$$b(u-\tilde{u}, \phi_P) = 0 \tag{1.3}$$

by the usual finite element error orthogonality relation (1.1.5b).

We are now able to state and prove the fundamental error estimate of this paper. This result will allow us to estimate the global finite element error $\|u-\tilde{u}\|_{1,\Omega}$ in terms of the solution η_P of the local subproblems (1.1). The basic ideas behind this estimate were first presented in [6].

Theorem 1.1. There is a constant $C > 0$ such that for any K-mesh

$$C^{-1}\Big(\sum_P \|\eta_P\|^2_{1,\Omega_P}\Big)^{1/2} \le \|u-\tilde{u}\|_{1,\Omega} \le C\Big(\sum_P \|\eta_P\|^2_{1,\Omega_P}\Big)^{1/2} \tag{1.4}$$

where $\sum_P$ denotes a summation over all proper nodes P of $\mathcal{D}$.

Proof. (i) The right hand inequality: By the coercivity of B (see 1.1.4c), and the finite element error orthogonality relation (1.1.6a), we have for any $V \in \tilde{M}_0(\mathcal{D})$

$$\begin{aligned}\|u-\tilde{u}\|^2_{1,\Omega} &\le C\, b(u-\tilde{u},u-\tilde{u}) \\ &\le C\, b(u-\tilde{u},u-\tilde{u}-V) \\ &\le C\, b\Big(u-\tilde{u},\Big(\sum_P \phi_P\Big)(u-\tilde{u}-V)\Big)\end{aligned}$$

since $\sum_P \phi_P = 1$ by Corollary 1.3.4(b). Thus,

$$\|u-\tilde{u}\|^2_{1,\Omega} \le C \sum_P b(u-u,\phi_P(u-\tilde{u}-V)).$$

But $\phi_P(u-\tilde{u}-V) \in M_P$ (certainly $\phi_P(u-\tilde{u}-V) \in H^1(\Omega_P)$; for the trace behaviour note that if $P \notin \partial(0,1)^2$ then $\phi_P = 0$ on $\partial\Omega_P$, whereas if $P \in \partial\Omega$ then $u-\tilde{u}-V = 0$ on $\Gamma_P^0 \cap \partial^0(0,1)^2$ while $\phi_P = 0$ on $\Gamma_P^0 \cap \Omega$). Thus

$$b(u-\tilde{u},\phi_P(u-\tilde{u}-V)) = b(\eta_P,\phi_P(u-\tilde{u}-V))$$

and we have

$$\begin{aligned}\|u-u\|^2_{1,\Omega} &\le C \sum_P |b(\eta_P,\phi_P(u-\tilde{u}-V))| \\ &\le C\Big(\sum_P \|\eta_P\|^2_{1,\Omega_P}\Big)^{1/2}\Big(\sum_P \|\phi_P(u-\tilde{u}-V)\|^2_{1,\Omega_P}\Big)^{1/2}\end{aligned}$$

$$\leq C \left(\sum_P \|\eta_P\|^2_{1,\Omega_P}\right)^{1/2} \|u-\tilde{u}\|_{1,\Omega}$$

on making the particular choice of V described in Corollary 1.5.3(ii).

The left hand inequality: Let us partition the set of proper nodes of $\mathcal{D}$ into disjoint classes χ_j $(j = 1,\ldots,J(\mathcal{D}))$ say, with the property that whenever $P,P' \in \chi_j$, then $\Omega_P \cap \Omega_{P'} = \emptyset$. It follows from Lemma 1.4.1(v) that we may always ensure that $J(\mathcal{D}) \leq C$, for some mesh independent constant $C > 0$. Now

$$b(u-\tilde{u}, \sum_P \eta_P) = \sum_P b(u-\tilde{u},\eta_P)$$

$$= \sum_P b(\eta_P,\eta_P)$$

$$\geq C \sum_P \|\eta_P\|^2_{1,\Omega_P}$$

by the coercivity of B over $M_P \subset M_0$ (1.1.4c). But

$$|b(u-\tilde{u}, \sum_P \eta_P)| \leq C\|u-\tilde{u}\|_{1,\Omega} \left\| \sum_{j=1}^{J} \sum_{P\in\chi_j} \eta_P \right\|_{1,\Omega}$$

$$\leq C \|u-\tilde{u}\|_{1,\Omega} \sum_{j=1}^{J} \left(\sum_{P\in\chi_j} \|\eta_P\|^2_{1,\Omega_P}\right)^{1/2}$$

$$\leq C \|u-\tilde{u}\|_{1,\Omega} \sum_{j=1}^{J} \left(\sum_P \|\eta_P\|^2_{1,\Omega_P}\right)^{1/2}$$

$$\leq C \|u-\tilde{u}\|_{1,\Omega} \left(\sum_P \|\eta_P\|^2_{1,\Omega_P}\right)^{1/2}.$$

The left hand inequality now follows. □

The significance of this result is that within the class of K-meshes the ratio $\theta = \left(\sum_P \|\eta_P\|^2\right)^{1/2} / \|u-\tilde{u}\|_{1,\Omega}$ is bounded above and below (away from zero), independently of the global mesh refinement pattern (as long as the K-mesh property is maintained). Note also that Theorem 1 demands no assumptions on the regularity of the solution u (other than it lie in M_0, of course).

However the practicality (1.4) will depend on whether the solutions of the local subproblems (1.1) can in some sense be effectively estimated. This is the matter that shall concern us for the remainder of this section. Our main result will be Theorem 4.5 which will turn out to estimate the η_P's in terms of the "jump residuals" $[\![\sigma(\tilde{u})\cdot\hat{n}]\!]$ across the interelement boundaries.

2.2. In this and the following section we establish some preliminary results which will be the basis of our effective estimation of $\|\eta_P\|_{1,\Omega}$. We shall be working initially with a version of the subproblem (1.1) "standardized" to the standard star tuple $\langle \mathcal{D}_0, \Gamma_0^0 \rangle$. We now describe this standardized problem: Let

$$M_0 = \{u \in H^1(\Omega_0): \ u = 0 \ \text{ on } \ \Gamma_0^0\}$$

and consider a continuous bilinear form $\Lambda_0: M_0 \times M_0 \to R$ which satisfies the conditions

$$\left.\begin{aligned} |\Lambda_0(u,v)| &\leq \beta \|u\|_{1,\Omega_0} \|v\|_{1,\Omega_0} \\ |\Lambda_0(u,u)| &\geq \beta^{-1} \|u\|^2_{1,\Omega_0} \end{aligned}\right\} \quad u,v \in M_0$$

for some $\beta > 0$. Corresponding to any $R_0 \in L_2(\Omega_0)$ and $r_0 \in L_2(\Gamma_0)$ let $\eta_0 \quad M_0$ be the solution of

$$A_0(\eta_0,v) = \int_{\Omega_0} R_0 v \, dx + \int_{\Gamma_0} r_0 v \, ds, \qquad v \in M_0. \tag{2.1}$$

Let $\varphi_0 \in M_0$ be a function with the properties

(i) $\varphi_0 > 0$ on Ω_0, (2.2)

(ii) for some $\delta > 0$, $\varphi_0 \geq \delta$ on some disc contained in Ω_0.

Finally let $G_0 \subseteq L_2(\Omega_0)$ and $g_0 \subseteq L_2(\Gamma_0)$ be finite dimensional subspaces.

Lemma 2.1. There is a constant $C = C(\mathcal{D}_0,\Gamma_0^0) > 0$ such that for any $v \in M_0$

$$\|v\|_{0,\Gamma_0} \leq C|v|_{1,\Omega_0},$$

and provided $\Gamma_0^0 \neq \emptyset$

$$\|v\|_{0,\Omega_0} \leq C|v|_{1,\Omega_0}.$$

Proof. The first estimate is a standard trace result, while the second is a Poincaré-type bound on the L_2 norm in terms of the H_1 semi-norm. □

Lemma 2.2. There is a constant $C = C(\beta,\mathcal{D}_0,\Gamma_0^0) > 0$ such that

$$\|\eta_0\|_{1,\Omega_0} \leq C(\|R_0\|_{0,\Omega_0} + \|r_0\|_{0,\Gamma_0}).$$

Proof.

$$\begin{aligned}
\|\eta_0\|^2_{1,\Omega_0} &\leq C\, A_0(\eta_0,\eta_0) \\
&\leq C \left| \int_{\Omega_0} R_0\eta_0 \, dx + \int_{\Gamma_0} r_0\eta_0 \, ds \right| \\
&\leq C(\|R_0\|_{0,\Omega_0}\|\eta_0\|_{0,\Omega_0} + \|\eta_0\|_{0,\Gamma_0}\|r_0\|_{0,\Gamma_0}).
\end{aligned}$$

By Lemma 2.1, $\|\eta_0\|_{0,\Gamma_0} \leq C\|\eta_0\|_{1,\Omega_0}$, and of course $\|\eta_0\|_{0,\Omega_0} \leq \|\eta_0\|_{1,\Omega_0}$. The lemma follows immediately. □

Lemma 2.3. Suppose G_0 and g_0 are finite-dimensional subspaces of $L_2(\Omega_0)$ and $L_2(\Gamma_0)$ respectively, then there is a constant $C = C(\beta,\mathcal{D}_0,\Gamma_0,G_0,g_0) > 0$ such that if $R_0 \in G_0$ and $r_0 \in g_0$ then

$$\|\eta_0\|_{1,\Omega_0} \geq C(\|R_0\|_{0,\Omega_0} + \|r_0\|_{0,\Gamma_0}).$$

Proof. Suppose the result does not hold, then there are sequences $R_0^{(n)}$ and $r_0^{(n)}$, $(n = 1,\ldots)$ from G_0 and g_0 respectively, and corresponding solutions $\eta_0^{(n)}$ of (2.1) such that

$$\|\eta_0^{(n)}\|_{1,\Omega_0} < \frac{1}{n}(\|R_0^{(n)}\|_{0,\Omega_0} + \|r_0^{(n)}\|_{0,\Gamma_0}). \tag{2.3}$$

We may without loss of generality assume that

$$\|R_0^{(n)}\|_{0,\Omega_0} + \|r_0^{(n)}\|_{0,\Gamma_0} = 1. \tag{2.4}$$

In particular $\|R_0^{(n)}\|_{0,\Omega_0} \leq 1$ and $\|r_0^{(n)}\|_{0,\Gamma_0} \leq 1$. By the finite dimensions of G_0 and g_0 we may suppose that

$$R_0^{(n)} \to R \in G_0 \quad \text{in} \quad L_2(\Omega_0)$$

and

$$r_0^{(n)} \to r \in g_0 \quad \text{in} \quad L_2(\Gamma_0).$$

Let η be the solution of (2.1) corresponding to R and r. By Lemma 2.2 applied to $(\eta-\eta_0^{(n)})$,

$$\|\eta-\eta_0^{(n)}\|_{1,\Omega_0} \leq C(\|R-R_0^{(n)}\|_{0,\Omega_0} + \|r-r_0^{(n)}\|_{0,\Gamma_0}) \to 0;$$

but by (2.3), $\eta_0^{(n)} \to 0$ in $\|\cdot\|_{1,\Omega_0}$, and so we conclude that $\eta = 0$.

Thus for any $v \in M_0$

$$0 = A_0(\eta,v) = \int_{\Omega_0} Rv\, dx + \int_{\Gamma_0} rv\, ds.$$

But this can only mean that $R = 0$ and $r = 0$. However this clearly contradicts the limiting case of (2.4). □

Lemma 2.4. Let G_0 and g_0 be as in Lemma 2.3, then there is a constant $C = C(\Omega_0,\Gamma_0,G_0,g_0,\varphi_0)$ such that if $R_0 \in G_0$, $r_0 \in g_0$ and $p \in P_0$ (the set of constant functions on Ω_0)

$$\|R_0\|_{0,\Omega_0} \leq C(\|r_0\|_{0,\Gamma_0} + \|R_0-p\|_{0,\Omega_0} + |\int_{\Omega_0} R_0\, 0\, dx + \int_{\Gamma_0} r_0\, 0\, ds|).$$

Proof. We shall again argue by contradiction. So suppose the result is not true, then there are sequences $R_0^{(n)}$, $r_0^{(n)}$ and $p^{(n)}$ from G_0, g_0 and P_0 respectively, such that

$$\|R_0^{(n)}\|_{0,\Omega_0} > n(\|r_0^{(n)}\|_{0,\Gamma_0} + \|R_0^{(n)} - p^{(n)}\|_{0,\Omega_0} + |\int_{\Omega_0} R_0^{(n)}\varphi_0\, dx + \int_{\Gamma_0} r_0^{(n)}\varphi_0\, ds|) \tag{2.5}$$

where we can suppose that

$$\|R_0^{(n)}\|_{0,\Omega_0} = 1. \tag{2.6}$$

By the finite dimension of G_0 we may as well assume that $R_0^{(n)} \to R \in G_0$ in $L_2(\Omega_0)$. But by (2.5), $\|R_0^{(n)} - p^{(n)}\|_{0,\Omega_0} \to 0$. Thus $p^{(n)} \to R$ in $L_2(\Omega_0)$. But P_0 is certainly a closed subspace of $L_2(\Omega_0)$ (since it is finite dimensional), and so $R \in P_0$, i.e., R is a constant on Ω_0. Now (2.5) also shows that $r_0^{(n)} \to 0$ in $L_2(\Gamma_0)$ and that

$$\left|\int_{\Omega_0} R_0^{(n)}\varphi_0 \, dx + \int_{\Gamma_0} r_0^{(n)}\varphi_0 \, ds\right| \to 0 \quad \text{as} \quad n \to \infty.$$ Thus we must have

$$0 = \lim_{n\to\infty} \int_{\Omega_0} R_0^{(n)}\varphi_0 \, dx = \int_{\Omega_0} R\varphi_0 \, dx.$$

The only way this can happen when φ_0 has the properties (2.2) and R is a constant, is for $R = 0$. But this contradicts the limiting case of (2.6). □

Lemma 2.5. There is a constant $C = C(\Omega_0, \mathcal{D}_0) > 0$ such that for any $z \in \mathcal{B}_0 = \{z \in C^0(\Omega_0)$: z bilinear on each $\Delta \in \mathcal{D}_0\}$,

$$\inf_{p\in\mathcal{D}_0} \|Dz-p\|_{0,\Omega_0} \leq C\left\{\left(\sum_{\Delta\in\mathcal{D}_0} |Dz|_{1,\Delta}^2\right)^{1/2} + \left(\int_{\Gamma_0^0} [\![\nabla z\cdot\hat{n}]\!]^2 \, ds\right)^{1/2}\right\} \tag{2.7}$$

where Dz denotes any (element-by-element) zeroth, first or second order derivative of z.

Proof. Let $M = \{z \in \mathcal{B} : Dz \in P_0\}$. Clearly the quotient space $\mathcal{B}_0/M$ is finite dimensional, and (2.7) will follow if we can show that both sides of (2.7) define norms on $\mathcal{B}_0/M$. In both cases it clearly suffices to show that if the respective quantity in (2.7) vanishes then $z \in M$:

(i) The left side: If $\inf_{p\in P_0} \|Dz-p\|_{0,\Omega_0} = 0$, then Dz is a constant on Ω_0 since P_0 is finite dimensional and hence closed.

(ii) The right side: If $\left(\sum_{\Delta \in \mathcal{D}_0} |Dz|^2_{1,\Delta} \right)^{1/2} = 0$, then Dz is constant on each $\Delta \in \mathcal{D}_0$. On the other hand, if $\int_{\Gamma_0^0} [\![\nabla z \cdot \underset{\sim}{n}]\!]^2 \, ds = 0$, then $[\![\nabla z \cdot \tilde{n}]\!] = 0$ on Γ_0^0, and so by Lemma 1.3.7 z is bilinear on Ω_0. In particular, $Dz \in C^0(\Omega_0)$. The net result being that $Dz \in P_0$. □

We shall now relax the requirements of Lemmas 2.3 and 2.4 that R_0 and r_0 lie in the finite dimensional spaces G_0 and g_0. Instead we shall require that they can be approximated from within these spaces.

Lemma 2.6. Let G_0 and g_0 be as in Lemma 2.3, and suppose that

$$\inf_{R^* \in G_0} \|R_0 - R^*\|_{0,\Omega_0} + \inf_{r^* \in g_0} \|r_0 - r^*\|_{0,\Gamma_0} < \varepsilon.$$

Then

(i) there is a constant $C = C(\beta, \Omega_0, \Gamma_0, G_0, g_0) > 0$ such that

$$C^{-1}(\|R_0\|_{0,\Omega_0} + \|r_0\|_{0,\Gamma_0} - \varepsilon) < \|\eta_0\|_{1,\Omega_0} < C(\|R_0\|_{0,\Omega_0} + \|r_0\|_{0,\Gamma_0}) \tag{2.8}$$

(ii) there is a constant $C = C(\Omega_0, \mathcal{D}_0, \Gamma_0, G_0, g_0, \psi_0) > 0$ such that if

$$\int_{\Omega_0} R_0 \varphi_0 \, dx + \int_{\Gamma_0} r_0 \varphi_0 \, ds = 0 \tag{2.9}$$

then

$$\|R_0\|_{0,\Omega_0} < C(\|r_0\|_{0,\Gamma_0} + \inf_{p \in P_0} \|R_0 - p\|_{0,\Omega_0} + \varepsilon). \tag{2.10}$$

Proof. Choose $R^* \in G_0$ and $r^* \in g_0$ such that

$\|R_0-R^*\|_{0,\Omega_0} + \|r_0-r^*\|_{0,\Gamma_0} < 2\varepsilon$. Let η^* be the solution of (2.1) corresponding to r^* and R^*. Applying Lemma 2.2 to the difference $(\eta_0-\eta^*)$ we get

$$\|\eta_0-\eta^*\|_{1,\Omega_0} < C(\|R_0-R^*\|_{0,\Omega_0} + \|r_0-r^*\|_{0,\Gamma_0}) < C\varepsilon.$$

Using Lemma 2.3,

$$\begin{aligned}\|R_0\|_{0,\Omega_0} + \|r_0\|_{0,\Gamma_0} &< \|R^*\|_{0,\Omega_0} + \|r^*\|_{0,\Gamma_0} + 2\varepsilon \\ &< C\|\eta^*\|_{1,\Omega_0} + 2\varepsilon \\ &< C(\|\eta_0\|_{1,\Omega_0} + \varepsilon),\end{aligned}$$

proving the left side of (2.8). The right side follows from Lemma 2.2.

For (2.10) we have

$$\|R_0\|_{0,\Omega_0} < \|R^*\|_{0,\Omega_0} + 2\varepsilon$$

while Lemma 2.4 gives for any $p \in P_0$

$$\begin{aligned}\|R^*\|_{0,\Omega_0} &< C(\|r^*\|_{0,\Gamma_0} + \|R^*-p\|_{0,\Omega_0} + |\int_{\Omega_0} R^*\phi_0\, dx + \int_{\Gamma_0} r^*\phi_0\, ds|) \\ &< C(\|r_0\|_{0,\Gamma_0} + \|R_0-p\|_{0,\Omega_0} + |\int_{\Omega_0} R_0\phi_0\, dx + \int_{\Gamma_0} r_0\phi_0\, ds| + \varepsilon) \\ &< C(\|r_0\|_{0,\Gamma_0} + \|R_0-p\|_{0,\Omega_0} + \varepsilon)\end{aligned}$$

by virtue of the orthogonality relation (2.9). □

2.3. In the last section we considered a "standardized" form of the local subproblems (1.1). In this section we extend those results to the actual star tuple $\langle P, {}_P, \Gamma_P^0 \rangle$.

For any proper node P with standard star tuple $\langle \mathcal{D}_0, \Gamma_0^0 \rangle$, let $\tau_p: \mathbb{R}^2 \to \mathbb{R}^2$ be the star transformation of P (see §1.4), that is

$$\tau_\rho(x) = \frac{1}{\rho}(x-P)$$

where $\rho = \rho_P = \max_{\Delta \in \mathcal{D}_P} |\Delta|$. (Notice that $\rho < 1$.)

Corresponding to the finite dimensional spaces G_0 and g_0 of §2.2 let us define

$$G_P = \{v \in L_2(\Omega_P): v \circ \tau_P^{-1} \in G_0\} \subset L_2(\Omega_P),$$

$$g_P = \{v \in L_2(\Gamma_P): v \circ \tau_P^{-1} \in \tau_P^{-1} g_0\} \subset L_2(\Gamma_P).$$

At this stage it is convenient to consider a more general version of (1.1): Find $\eta \in M_P$ such that

$$b(\eta, v) = \int_{\Omega_P} Rv \, dx + \int_{\Gamma_P} rv \, ds, \qquad \forall v \in M_P \tag{3.1}$$

where $R \in L_2(\Omega_P)$ and $r \in L_2(\Gamma_P)$.

Lemma 3.1. Set

$$\varepsilon = \rho_P \inf_{\bar{R} \in G_P} \|R-\bar{R}\|_{0,\Omega_P} + \rho_P^{1/2} \inf_{\bar{r} \in g_P} \|r-\bar{r}\|_{0,\Gamma_P} \tag{3.2}$$

then

(i) there is a constant $C = C(\mathcal{D}_0,\Gamma_0,G_0,g_0)$ such that

$$C^{-1}(\rho_P\|R\|_{0,\Omega_P} + \rho_P^{1/2}\|r\|_{0,\Gamma_P}-\varepsilon) \leq \|\eta\|_{1,\Omega_P} \leq C(\rho_P\|R\|_{0,\Omega_P} + \rho_P^{1/2}\|r\|_{0,\Gamma_P})$$

(ii) there is a constant $C = C(\mathcal{D}_0,\Gamma_0^0,G_0,g_0)$ such that if

$$\int_{\Omega_P} R\phi_P\,dx + \int_{\Gamma_P} r\phi_P\,ds = 0$$

then

$$\rho_P\|R\|_{0,\Omega_P} \leq C(\rho_P^{1/2}\|r\|_{0,\Gamma_P} + \rho_P \inf_{p\in P_0(\Omega_P)}\|R-p\|_{0,\Omega_P} + \varepsilon)$$

where $P_0(\Omega_P)$ is the space of constant functions on Ω_P.

Proof. It is clear that $M_0 = \{v \in H^1(\Omega_0): v \circ \tau_P \in M_P\}$. Define a bilinear form $\Lambda_0: M_0 \times M_0 \to \mathbb{R}$ by

$$\Lambda_0(u,v) = b(u \circ \tau_P,\ v \circ \tau_P).$$

It follows from the boundedness of $b(\cdot,\cdot)$ on $M_P \times M_P$

$$|\Lambda_0(u,v)| = |b(u \circ \tau_P,\ v \circ \tau_P)|$$

$$\leq C\|u \circ \tau_P\|_{1,\Omega_P}\|v \circ \tau_P\|_{1,\Omega_P}$$

$$\leq C(\rho^2\|u\|^2_{0,\Omega_0} + |u|^2_{1,\Omega_0})^{1/2}(\rho^2\|v\|^2_{0,\Omega_0} + |v|^2_{1,\Omega_0})^{1/2}$$

$$\leq C\|u\|_{1,\Omega_0}\|v\|_{1,\Omega_0}.$$

In addition, by the coercivity of $b(\cdot,\cdot)$ over M_P,

$$|\Lambda_0(u,u)| = |b(u \circ \tau_p,\ u \circ \tau_p)| > \alpha \|u \circ \tau_p\|^2_{1,\Omega_p}$$

$$> C|u|^2_{1,\Omega_0}$$

$$> C\|u\|^2_{1,\Omega_0}$$

using lemma 2.1, noting that $\Gamma_0^0 \neq \phi$. It is readily verified by a change of variable of integration that for any $v \in M_0$

$$\Lambda_0(\eta \circ \tau_p, v) = b(\eta, v \circ \tau_p^{-1}) = \int_{\Omega_p} R(v \circ \tau_p^{-1})\, dx + \int_{\Gamma_p} r(v \circ \tau_p^{-1}) ds$$

$$= \int_{\Omega_0} (\rho^2 R \circ \tau_p) v\, dx + \int_{\Gamma_0} (\rho r \circ \tau_p) v\, ds.$$

So we are in a situation covered by the analysis of §2.2 with $\eta_0 = \eta \circ \tau_p$, $R_0 = \rho^2 R \circ \tau_p$ and $r_0 = \rho r \circ \tau_p$. Furthermore, for any $\bar{R} \in G_p$ and $\bar{r} \in g_p$, we have $\rho^2 \bar{R} \circ \tau_p \in G_0$ and $\rho \bar{r} \circ \tau_p \in g_0$ with

$$\|\rho^2 R \circ \tau_p - \rho^2 \bar{R} \circ \tau_p\|_{0,\Omega_0} + \|\rho r \circ \tau_p - \rho \bar{r} \circ \tau_p\|_{0,\Gamma_0}$$

$$= \rho \|R - \bar{R}\|_{0,\Omega_p} + \rho^{1/2} \|r - \bar{r}\|_{0,\Gamma_p}.$$

By virtue of (3.2) we are now able to apply Lemma 2.6 to infer from (2.8)

$$C^{-1}(\|\rho^2 R \circ \tau_p\|_{0,\Omega_0} + \|\rho r \circ \tau_p\|_{0,\Gamma_0} - \varepsilon) \leq \|\eta \circ \tau_p\|_{1,\Omega_0}$$

$$\leq C(\|\rho^2 R \circ \tau_p\|_{0,\Omega_0} + \|\rho r \circ \tau_p\|_{0,\Gamma_0}). \tag{3.3}$$

Upon rescaling back to the actual star Ω_p,

$$\|\eta \circ \tau_P\|^2_{1,\Omega_0} = |\eta \circ \tau_P|^2_{0,\Omega_0} + |\eta \circ \tau_P|^2_{1,\Omega_0} = \frac{1}{\rho^2} |\eta|^2_{0,\Omega_P} + |\eta|^2_{1,\Omega_P}$$

$$> \|\eta\|^2_{1,\Omega_P};$$

while by Lemma 2.1 we also have the opposing bound

$$\|\eta \circ \tau_P\|_{1,\Omega_0} \leq C|\eta \circ \tau_P|_{1,\Omega_0} \leq C|\eta|_{1,\Omega_P} \leq C\|\eta\|_{1,\Omega_P}.$$

Rescaling the other terms in (3.3) leads at once to (i) of the lemma.

To show (ii) note that $\phi_0 = \phi_P \circ \tau_P$ satisfies (2.2) and that

$$\int_{\Omega_0} (\rho^2 R \circ \tau_P)\phi_0 + \int_{\Gamma_0} (\rho r \circ \tau_P)\phi_0 = \int_{\Omega_P} R\phi_P \, dx + \int_{\Gamma_P} r\phi_P \, ds = 0.$$

Part (ii) of Lemma 2.6 then gives

$$\|\rho^2 R \circ \tau_P\|_{0,\Omega_0} \leq C(\|\rho r \circ \tau_P\|_{0,\Gamma_0} + \inf_{p \in \mathcal{P}_0} \|\rho^2 R \circ \tau_P - P\|_{0,\Omega_0} + \varepsilon),$$

where, since ϕ_0 is determined by the standard tuple $\langle \Omega_0, \mathcal{D}_0, \Gamma_0 \rangle$, the constant C of Lemma 2.6 (ii) is in fact $C = C(\Omega_0, \mathcal{D}_0, \Gamma_0, G_0, g_0)$. Rescaling to Ω_P gives the result. □

2.4. We now return to the task of effectively estimating $(\sum_P \|\eta_P\|^2_{1,\Omega_P})^2$. The local subproblems (1.1) fit into the pattern of §2.3 if we set (see (1.2))

$$R = (f - L(\tilde{u})) \qquad \text{on each } \Delta$$

and

$$r = \begin{cases} -[\![\sigma(\tilde{u})\cdot\hat{n}]\!] & \text{on } \Gamma^* \\ t - \sigma(\tilde{u})\cdot\hat{n} & \text{on } \partial^1(0,1)^2. \end{cases} \tag{4.1}$$

To complete the framework required for applying the results of §2.3 we define

$$G_0 = \{v \in L_2(\Omega_0): \ v \text{ is constant on each } \Delta \in \mathcal{D}_0\},$$

$$g_0 = \{v \in L_2(\Gamma_0): \ v \text{ is linear on each primitive edge } \gamma \subseteq \Gamma_0\}.$$

It is clear that G_0 and g_0 are finite dimensional and that they are determined completely by the standard star tuple $\langle \mathcal{D}_0, \Gamma_0^0 \rangle$.

Lemma 4.1:

(i) $$\sum_{\Delta \in \mathcal{D}} |\Delta|^2 \|\tilde{u}\|^2_{2,\Delta} = o(1) \quad \text{as} \quad h(\mathcal{D}) \to 0.$$

(ii) $$\sum_P \rho_P^2 \inf_{\bar{R} \in G_P} \|R - \bar{R}\|^2_{0,\Omega_P} = o(h(\mathcal{D})^2) \quad \text{as} \quad h(\mathcal{D}) \to 0.$$

(iii) $$\sum_P \rho_P \inf_{\bar{r} \in g_P} \|r - \bar{r}\|^2_{0,\Gamma_P} = o(h(\mathcal{D})^2) \quad \text{as} \quad h(\mathcal{D}) \to 0.$$

(iv) There exists $r^* \in L_2(\Gamma)$ which is linear on each primitive edge $\gamma \in \Gamma$, which agrees with r at all nodes lying on Γ, and which satisfies

$$\sum_\Delta |\Delta| \|r - r^*\|^2_{\partial\Delta \cap \Gamma} = o(h(\mathcal{D})^2) \quad \text{as} \quad h(\mathcal{D}) \to 0.$$

Proof. To prove (i),

$$\sum_{\Delta \in \mathcal{D}} |\Delta|^2 \|\tilde{u}\|^2_{2,\Delta} \leq h^2(\mathcal{D}) \|\tilde{u}\|^2_{1,\Omega} + \sum_{\Delta \in \mathcal{D}} |\Delta|^2 |\tilde{u}|^2_{2,\Delta}.$$

By the projection property (1.1.5) of the finite element solution the first terms obviously tends to 0. For the second term

$$\sum_{\Delta\in\mathcal{D}} |\Delta|^2 |\tilde{u}|^2_{2,\Delta} \leq C \sum_{\Delta\in\mathcal{D}} \inf_{\substack{v \\ (v \text{ linear on } \Delta)}} |\tilde{u}-v|^2_{1,\Delta}$$

$$\leq C \sum_{\Delta\in\mathcal{D}} \left(\|u-\tilde{u}\|^2_{1,\Delta} + \inf_{\substack{v \\ (v \text{ linear on } \Delta)}} \|u-v\|^2_{1,\Delta} \right)$$

$$\leq C\left(\|u-\tilde{u}\|^2_{1,\Omega} + \sum_{\Delta} \inf_{\substack{v \\ (v \text{ linear on } \Delta)}} \|u-v\|^2_{1,\Delta} \right)$$

which certainly tends to zero as $h(\mathcal{D}) \to 0$.

Now turn to (ii) of the lemma. On each element $\Delta \in \mathcal{D}_P$, R is a sum of terms each of the form $\lambda(x)\xi(x)$ where λ has bounded derivatives of all orders on Ω and $\xi = 1, \tilde{u}, D_i\tilde{u}$ or $D_{ij}\tilde{u}$ $(i,j = 1,2)$. We can find a constant function θ on Δ such that

$$\|\lambda\xi-\theta\|_{0,\Delta} \leq C|\Delta|\|\lambda\xi\|_{1,\Delta}$$

$$\leq C|\Delta|\|\xi\|_{1,\Delta}$$

$$\leq C|\Delta|\left(\|\tilde{u}\|_{2,\Delta} + |\Delta|\right).$$

Thus

$$\inf_{\bar{R}\in G_P} \|R-\bar{R}\|_{0,\Omega_P} = C\Big\{ \sum_{\Delta\in\mathcal{D}_P} \left(|\Delta|^2 \|\tilde{u}\|^2_{2,\Delta} + |\Delta|^4 \right) \Big\}^{1/2},$$

and so

$$\sum_P \rho_P^2 \inf_{\bar{R}\in G_P} \|R-\bar{R}\|^2_{0,\Omega_P} \leq C \sum_P \rho_P^2 \sum_{\Delta\in\mathcal{D}_P} \left(|\Delta|^2 \|\tilde{u}\|^2_{2,\Delta} + |\Delta|^4 \right)$$

$$\leq C\, h(\mathcal{D})^2 \Big(\sum_{\Delta\in\mathcal{D}} |\Delta|^2 \|\tilde{u}\|^2_{2,\Delta} + h(\mathcal{D})^2 \Big)$$

where we have made use of the fact that each element $\Delta \in \mathcal{D}$ belongs to a $\mathcal{D}_P$ for at least one, but no more than $C = C(K)$, proper nodes P (Lemmas 1.3.6 and 1.4.1). Part (ii) of the lemma now follows immediately on using (i).

For part (iii), observe that on each element edge γ making up Γ_P, r is a sum of terms of the form $\lambda(t)\xi(t)$ where λ has bounded derivatives of all orders (independent of $\mathcal{D}$) and $\xi = 1$ or a limiting value of $D_i\tilde{u}$ $(i = 1,2)$ on the edge. There is no loss of generality in assuming tha $D_i\tilde{u}$ is linear on this edge. We can find a linear function β on γ satisfying

$$\begin{aligned} \|\lambda\xi-\beta\|_{0,\gamma} &\leq C|\gamma|^2\|\lambda\xi\|_{2,\gamma} \\ &\leq C|\gamma|^2(\|\xi\|_{2,\gamma}) \\ &\leq C|\gamma|^2\Big(\sum_{i=1}^{2}\|D_i\tilde{u}\|_{1,\gamma} + |\gamma|^{1/2}\Big) \end{aligned}$$

where $D_i\tilde{u}$ is the limiting value on γ from within an element $\Delta \in \mathcal{D}_P$. We have

$$\|D_i\tilde{u}\|_{1,\gamma} \leq C|\Delta|^{-1/2}\|\tilde{u}\|_{2,\Delta}$$

and so

$$\|\lambda\xi-\beta\|_{0,\gamma} \leq C|\Delta|^{1/2}\big(|\Delta|\|\tilde{u}\|_{2,\Delta} + |\Delta|^2\big). \tag{4.2}$$

Therefore

$$\inf_{r\in g_P}\|r-\bar{r}\|_{0,\Gamma_P} \leq C\rho_P^{1/2}\Big\{\sum_{\Delta\in\mathcal{D}_P}|\Delta|^2\|\tilde{u}\|_{2,\Delta}^2 + |\Delta|^4\Big\}^{1/2}$$

and so, much as part (i), squaring and summing over all proper nodes

$$\sum_P \rho_P \inf_{r\in g_P} \|r-\bar{r}\|^2_{0,\Gamma_P} \leq C \sum_P \rho_P^2 \sum_{\Delta\in\mathcal{D}_P} (|\Delta|^2\|\tilde{u}\|^2_{2,\Delta} + |\Delta|^4)$$

$$\leq Ch(\mathcal{D})^2\Big(\sum_\Delta |\Delta|^2\|\tilde{u}\|^2_{2,\Delta} + h(\mathcal{D})^2\Big).$$

Again, on using (i), part (iii) of the lemma follows.

Part (iv) of the lemma likewise follows from (4.2) and (4.1), since we may as well choose β to interpolate $\lambda\xi$ at the endpoints of λ. □

We are now able to state and prove the main results of this paper concerning an equivalent estimator for the energy norm of the finite element error. Because of the coercivity of $b(\cdot,\cdot)$ there is no need to distinguish between equivalent estimators for $(b(u-\tilde{u},u-\tilde{u}))^{1/2}$, or $\|u-\tilde{u}\|_{1,\Omega}$. For definiteness we shall phrase our result in terms of the latter.

Theorem 4.2. There is a constant $C > 0$ such that for any K-mesh

$$\|u-\tilde{u}\|_{1,\Omega} \leq CE \tag{4.4a}$$

where

$$E = \Big\{\sum_{\Delta\in\mathcal{D}} \big[|\Delta|^2\|R\|^2_{0,\Delta} + |\Delta|\|r\|^2_{0,\partial\Delta\cap\Gamma}\big]^{1/2}.$$

If in addition for some $L > 0$

$$\|u-\tilde{u}\|_{1,\Omega} \geq Lh(\mathcal{D}) \tag{4.5}$$

then there are constants $C > 0$, $h_0 > 0$ such that for any K-mesh $\mathcal{D}$ with $h(\mathcal{D}) \leq h_0$

$$C^{-1}E \leq \|u-\tilde{u}\|_{1,\Omega}. \tag{4.4b}$$

Proof. Consider a particular proper node P and its star Ω_P. By Lemma 3.1

$$C^{-1}(E_P-\rho_P \inf_{\bar{R}\in G_P} \|R-\bar{R}\|_{0,\Omega_P} + \rho_P^{1/2} \inf_{\bar{r}\in g_P} \|r-\bar{r}\|_{0,\Gamma_P}) \leq \|\eta_P\|_{1,\Omega_P} \leq C E_P \qquad (4.6)$$

where

$$E_P^2 = \sum_{\Delta\in\mathcal{D}_P} [|\Delta|^2\|R\|^2_{0,\Delta} + |\Delta|\|r\|^2_{0,\partial\Delta\ \Gamma_P}].$$

Here we have used the fact that for any $\Delta \in \mathcal{D}_P$, $C^{-1}\rho_P \leq |\Delta| \leq \rho_P$ (see Lemma 1.4.1).

The constant in (4.6) depends only on the standard star tuple $\langle\mathcal{D}_0,\Gamma_0^0\rangle$ associated with the proper node P, but is otherwise independent of the mesh or the node P. However, by Theorem 1.4.3, for any K mesh there are at most $C = C(K) < \infty$ possible standard star tuples. We may therefore regard the constant C in (4.6) as uniform for all proper nodes P and all meshes of the specified class.

We claim that $C^{-1} E^2 \leq \sum_P E_P^2 \leq C E^2$. To see this we need only recall

(i) any element $\Delta \in \mathcal{D}$ belongs to a $\mathcal{D}_P$ for at least one but no more than $C = C(K)$ proper nodes P (see Lemmas 1.3.6 and 1.4.1)

(ii) any edge contained in Γ is contained in a Γ_P for at least one, but no more than $C = C(K)$ proper nodes P (see Lemma 1.4.1).

Thus squaring (4.6), summing over all proper nodes P, and using (ii) and (iii) of lemma 4.1 gives

$$C^{-1}(E^2-h(\mathcal{D})^2 o(1)) \leq \sum_P \|\eta_P\|^2_{1,\Omega_P} \leq C E^2. \qquad (4.7)$$

Theorem 1.1 allows us to conclude from (4.7)

$$C^{-1}(E^2-h(\mathcal{D})^2 o(1)) \leq \|u-\tilde{u}\|^2_{1,\Omega} \leq C E^2.$$

Making use of (4.5) gives the desired result. □

At first glance the estimate of (4.4a) may not seem too surprising since it bounds the error in terms of a "residual." Note however a simple integration by parts of (1.2) would only have yielded

$$\|u-\tilde{u}\|_{1,\Omega} \leq C E^*$$

where $(E^*)^2 = \sum_{\Delta} [\|R\|^2_{0,\Delta} + |\Delta|^{-1}\|r\|^2_{0,\partial\Delta\cap\Gamma}]$.

To obtain the extra powers of $|\Delta|$ in (4.4) it was essential that we were able to reduce the estimation of the global error $\|u-\tilde{u}\|_{1,\Omega}$ to the consideration of a number of local subproblems.

The estimate of (4.4) is sharp in the sense that uniformly for all K-meshes

$$0 < C^{-1} \leq E/\|u-\tilde{u}\|_{1,\Omega} \leq C \tag{4.8}$$

(at least for $h(\mathcal{D})$ sufficiently small). In contrast, for E^*, although we certainly have the estimate

$$C^{-1} \leq \frac{E^*}{\|u-\tilde{u}\|_{1,\Omega}},$$

no two-sided estimate similar to (4.8) can hold, since $E^*/E \to \infty$ and so

$$\frac{E^*}{\|u-\tilde{u}\|_{1,\Omega}} \to \infty \quad \text{as} \quad h(\mathcal{D}) \to 0.$$

Theorem 4.2 requires no extra assumptions on the regularity of the exact solution u other than (4.5). However this requirement is hardly a restriction at all being satisfied in all but trivial cases (see Lemma 1.5.4).

Although the estimate of Theorem 4.2 is sharp in the sense described above, it can be simplified even further. It turns out that the L_2 residual contribution can be dropped from E, giving an expression solely in terms of "line residuals," while still maintaining the two-sided estimate (4.4). This will be shown in Theorem 4.5. It is natural to ask whether the "reverse" simplification is valid, that is, can the "line residuals" be dropped from ε without affecting the estimate. It is not difficult to see that this cannot be true. Consider the simple case where $L(\cdot)$ is the Laplacian. Then for bilinear elements, $L(\tilde{u}) = 0$, and the L_2 residual becomes just

$$\Big(\sum_{\Delta} |\Delta|^2 \int_{\Delta} |f|^2 \, dx\Big)^{1/2} \geq Ch(\mathcal{D})\|f\|_{0,\Omega}$$

if the mesh $\mathcal{D}$ is quasiuniform. However, if the solution u has some singular behaviour so that $u \notin H^2(\Omega)$ then $\|u-\tilde{u}\|_{1,\Omega}$ must converge at a slower rate than $O(h(\mathcal{D}))$. So the L_2 residual alone cannot suffice for E in (4.5) in this case.

For proving Theorem 5 we shall need the following lemmas:

Lemma 4.3. There is a mesh independent constant $C > 0$ such that for any proper node P and any $z \in B(\Omega_P) = \{z \in C^0(\Omega_P)$: z bilinear on each $\Delta \in \mathcal{D}_P\}$,

$$\inf_{p \in P_0(\Omega_P)} \|Dz-p\|_{0,\Omega_P} \leq C\{\rho_P\Big(\sum_{\Delta \in \mathcal{D}_P} |Dz|^2_{1,\Delta}\Big)^{1/2} + \rho_P^s\Big(\int_{\Gamma_P^*} [\![\nabla z \cdot \hat{n}]\!]^2 \, ds\Big)^{1/2}\}$$

where Dz denotes any (element-by-element) zeroth ($s = 3/2$), first ($s = 1/2$) or second ($s = -1/2$) derivative of z.

Proof. This is just the rescaled version of Lemma 2.5, taking note again that there are only a finite number of standard star tuples. □

Lemma 4.4. Suppose $\gamma \in \Omega$ is a straight horizontal or vertical line segment, then for any $z \in \tilde{M}(\mathcal{D})$,

$$\int_\gamma [\![\nabla z \cdot \hat{n}]\!]^2 \, ds \leq C \int_\gamma [\![\sigma(z)\cdot\hat{n})]\!]^2 \, ds.$$

Proof. For definiteness suppose that γ is vertical and let superscript (-) and (+) indicate limits from the left and righthand sides of γ respectively. (See Fig. 5.) Since all the $a_{ijk\ell}$ are continuous in Ω,

Figure 5. The scheme of notation.

$$[\![\sigma(z)\cdot\hat{n}]\!] = \sum_{i,j,k,\ell} \{(a_{ijk\ell}D_\ell z_j n_k)^{(+)} + (a_{ijk\ell}D_\ell z_j n_k)^{(-)}\}$$

$$= \sum_{i,j,\ell} \{a_{ij1\ell}[-(D_\ell z_j)^{(+)} + (D_\ell z_j)^{(-)}]\}$$

since $n_2 = 0$, $n_1^{(+)} = -1$ and $n_1^{(-)} = 1$. But $D_2 z_j$ is continuous across γ, so

$$[\![\sigma(z)\cdot\hat{n}]\!] = \sum_{i,j} a_{ij11}\,[-(D_1 z_j)^{(+)} + (D_1 z_j)^{(-)}]$$

$$= a^{11}\,[\![\nabla z\cdot\hat{n}]\!]$$

where

$$a^{11} = \begin{pmatrix} a_{1111} & a_{1211} \\ a_{2111} & a_{2211} \end{pmatrix}.$$

But the matrix a^{11} is uniformly positive definite on Ω (see 1.1.4b) and so $\det(a^{11}) \geq C > 0$ on Ω. It follows then that on γ

$$|[\![\nabla z \cdot \hat{n}]\!]| \leq C|[\![\sigma(z)\cdot\hat{n}]\!]|,$$

and the lemma is proven in this case.

The case of γ horizontal is treated similarly. □

Theorem 4.5. Suppose that for some $L > 0$

$$\|u-\tilde{u}\|_{1,\Omega} \geq Lh(\mathcal{D})$$

then there are constants $C, h_0 > 0$ such that for any K-mesh $\mathcal{D}$ with $h(\mathcal{D}) \leq h_0$

$$C^{-1}E \leq \|u-\tilde{u}\|_{1,\Omega} \leq CE \tag{4.9a}$$

where

$$E = \Big\{ \sum_{\Delta} |\Delta| \|r\|^2_{0,\partial\Delta\cap\Gamma} \Big\}^{1/2}. \tag{4.9b}$$

Proof. First note that the left hand side of (4.9) follows trivially from (4.4b) since $E \leq \mathcal{E}$. To prove the right hand side it will suffice to show

$$\sum_{\Delta} |\Delta|^2 \|R\|^2_{0,\Delta} \leq E^2 + o(h(\mathcal{D})^2) \tag{4.10}$$

with the $o(1)$ term valid as $h(\mathcal{D}) \to \infty$.

For any interior proper node P we have (see (1.2), (1.3))

$$\int_{\Omega_P} R\phi_P \, dx + \int_{\Gamma_P} r\phi_P \, ds = 0$$

with R and r as in (4.1). Thus we may apply part (ii) of Lemma 3.1,

$$\rho_P \|R\|_{0,\Omega_P} \leq C(\rho_P^{1/2} \|r\|_{0,\Gamma_P} + \rho_P \inf_{p \in P_0(\Omega_P)} \|R-p\|_{0,\Omega_P} + \varepsilon_P), \qquad (4.11)$$

where $P_0(\Omega_P)$ is the space of constant functions on Ω_P and

$$\varepsilon_P = \rho_P \inf_{\bar{R} \in G_P} \|R-\bar{R}\|_{0,\Omega_P} + \rho_P^{1/2} \inf_{\bar{r} \in g_P} \|r-\bar{r}\|_{0,\Gamma_P}.$$

We shall concentrate on the term $\inf_{p \in P_0(\Omega_P)} \|R-p\|_{0,\Omega_P}$ for a while. As the coefficients of $L(\cdot)$ are smooth on Ω, let us write $\bar{L}(\cdot)$ for $L(\cdot)$ but with its coefficients replaced by their averages over Ω_P. If we then set

$$\bar{\bar{R}} = \sum_{\Delta \in \mathcal{D}_P} (F - \bar{L}(\tilde{u})),$$

we have

$$\inf_{p \in P_0(\Omega_P)} \|R-p\|_{0,\Omega_P} \leq \|R-\bar{\bar{R}}\|_{0,\Omega_P} + \inf_{p \in P_0(\Omega_P)} \|\bar{\bar{R}}-p\|_{0,\Omega_P}$$

$$\leq C\rho_P \|\tilde{u}\|_{2,\Omega_P} + \inf_{p \in P_0(\Omega_P)} \|\bar{\bar{R}}-p\|_{0,\Omega_P} \qquad (4.12)$$

$$\leq C\rho_P(\|\tilde{u}\|_{2,\Omega_P} + \|\Omega_P\|) + \inf_{p \in P_0(\Omega_P)} \| \sum_{\Delta \in \mathcal{D}_P} L(\tilde{u}) - p\|_{0,\Delta}$$

where $\|\tilde{u}\|_{2,\Omega_P}$ is to be understood in an element-by-element sense. Now $L(\tilde{u})$ is just a sum of terms of the form $\lambda(D\tilde{u})$ where λ is a constant and $D\tilde{u}$ denotes a zeroth, first or second derivative of the funtion $\tilde{u} \in S(\Omega_P)$. (The constant λ is bounded independently of P and $\mathcal{D}$). We may clearly apply Lemma 4.3 to obtain

$$\inf_{p \in P_0(\Omega_p)} \| \sum_{\Delta \in \mathcal{D}_p} \bar{L}(\tilde{u}) - p \|_{0,\Delta}$$

$$\leq C\Big(\rho_p \|\tilde{u}\|_{2,\Omega_p} + \rho_p^{-1/2} \Big(\int_{\Gamma_p^*} [\![\nabla\tilde{u} \cdot \hat{n}]\!]^2 \, ds\Big)^{1/2}\Big) \tag{4.13}$$

$$\leq C\Big(\rho_p \|\tilde{u}\|_{2,\Omega_p} + \rho_p^{-1/2} \Big(\int_{\Gamma_p^*} [\![\sigma(\tilde{u}) \cdot \hat{n}]\!]^2 \, ds\Big)^{1/2}$$

by Lemma 4.4.

Substituting (4.13) in (4.12), and then this in turn into (4.11) gives

$$\rho_p \|R\|_{0,\Omega_p} \leq C(\rho_p^{1/2} \|r\|_{0,\Gamma_p} + \varepsilon_p^*) \tag{4.14}$$

where

$$\varepsilon_p^* \leq \varepsilon_p + \rho_p^2(\|\tilde{u}\|_{2,\Omega_p} + |\Omega_p|).$$

Squaring (4.14) and summing over all interior proper nodes P gives

$$\sum_{\Delta \in \mathcal{D}} |\Delta|^2 \|R\|_{0,\Delta}^2 \leq C \sum_{\substack{P \text{ an interior} \\ \text{proper node}}} \rho_P^2 \|R\|_{0,\Omega_p}^2 \tag{4.15}$$

$$\leq C\Big(E^2 + \sum_P (\varepsilon_p)^2 + h(\mathcal{D})^2\Big(\sum_\Delta |\Delta|^2 \|\tilde{u}\|_{2,\Delta} + h(\mathcal{D})^2\Big)\Big)$$

where, much as in the proof of Theorem 4.2, we have used the following properties of our meshes:

(i) for $\Delta \in \mathcal{D}_p$, $C^{-1}\rho_p \leq |\Delta| \leq C\rho_p$ (see Lemma 1.4.1)

(ii) any element $\Delta \in \mathcal{D}$ belongs to $\mathcal{D}_p$ for at least one interior proper node P (see Lemma 1.3.6) and at most $C = C(K)$ proper nodes P (Lemma 1.4.1)

(iii) any edge contained in γ is contained in Γ_p for one more than $C = C(K)$ proper nodes P (Lemma 1.4.1). □

The desired result (4.10) follows at once from (4.15) and the assumed lower bound on $\|u-\tilde{u}\|_{1,\Omega}$ upon appealing to (i), (ii) and (iii) of Lemma 4.1.

A slightly more computationally convenient form of estimator than that given by Theorem 4.5 would be obtained if the integrals involved in $\|r\|^2_{0,\partial\Delta\cap\Gamma}$ could be replaced by discrete sums.

If g is a (sufficiently smooth) function defined on an edge, γ say, of an element, and if $Q_1,\ldots,Q_k$ are the nodes which lie on γ then let us write

$$\|g\|^*_{0,\gamma} = \Big(|\gamma| \sum_{j=1}^{k} |g(Q_j)|^2\Big)^{1/2}.$$

If Δ is an element and g is defined on $\partial\Delta \cap \Gamma$ then we shall write

$$\|g\|^*_{0,\partial\Delta\cap\Gamma} = \Big(\sum_{\substack{\gamma \text{ an edge of } \Delta \\ \gamma\subset\Gamma}} \|g\|^*_{0,\gamma}\Big)^2\Big)^{1/2}.$$

Corollary 4.6. Theorem 4.5 remains true if E is replaced by E^* where

$$E^* = \Big\{ \sum_{\Delta} |\Delta| (\|r\|^*_{0,\partial\Delta\cap\Gamma})^2 \Big\}^{1/2}.$$

Proof. Let r^* be as in (iv) of Lemma 4.1. Since any edge γ can have no more than $C = C(K)$ nodes on it (see Lemma 1.4.1)

$$C^{-1}\|r^*\|_{0,\gamma} \leq \|r\|^*_{0,\gamma} \leq C\|r^*\|_{0,\gamma}.$$

Thus

$$C^{-1}E^* \leq \Big(\sum_{\Delta} |\Delta| \|r^*\|^2_{0,\partial\Delta\cap\Gamma}\Big)^{1/2} \leq CE^*.$$

But

$$\sum_{\Delta} |\Delta| \|r^*\|^2_{0,\partial\Delta\cap\Gamma} \leq C \sum_{\Delta} |\Delta| \big(\|r\|^2_{0,\partial\Delta\cap\Gamma} + \|r-r^*\|^2_{0,\partial\Delta\cap\Gamma}\big)$$

$$\leq C\Big(E^2 + \sum_{\Delta} |\Delta| \|r-r^*\|^2_{0,\partial\Delta\cap\Gamma}\Big)$$

$$\leq C(E^2 + o(h(\mathcal{D})^2)$$

by (iv) of Lemma 4.1. Likewise

$$E^2 \leq C\Big(\sum_{\Delta} |\Delta| \|r^*\|^2_{0,\partial\Delta\cap\Gamma} + o(h(\mathcal{D})^2)\Big).$$

The corollary now follows by virtue of the assumed lower bound on $\|u-\tilde{u}\|_{1,\Omega}$ in the theorem. □

Since we are only concerned here with equivalent estimators for $\|u-\tilde{u}\|_{1,\Omega}$ there are many other modifications that can be made to E or E* without affecting their equivalent estimator property. Let us just mention one other such modification which we shall say a little more about in Part II. Using the same notation as above, define

$$\|g\|^{**}_{0,\gamma} = \Big(|\gamma| \max_{j=1,\ldots,k} |g(Q_j)|^2\Big)^{1/2}$$

$$\|g\|^{**}_{0,\partial\Delta\cap\Gamma} = \left\{ \max_{\substack{\gamma \\ (\gamma \text{ a vertical edge of } \Delta) \\ \gamma\subset\Gamma}} \big(\|g\|^{**}_{0,\gamma}\big)^2 + \max_{\substack{\gamma \\ (\gamma \text{ a horizontal edge of } \Delta) \\ \gamma\subset\Gamma}} \big(\|g\|^{**}_{0,\gamma}\big)^2 \right\}^{1/2}.$$

Now set

$$E^{**} = \{ \sum_{\Delta \in \mathcal{D}} |\Delta| (\|r\|^{**}_{0,\partial\Delta\cap\Gamma})^2 \}^{1/2}.$$

Since there are at most $C = C(K)$ nodes on an edge of any element, it is immediate that $C^{-1}E^* \leq E^{**} \leq CE^*$. Thus E^{**} is an equivalent estimator under the same conditions that apply to E^*.

REFERENCES

[1] C. Mesztenyi, W. Szymczak, FEARS User's Manual for Univac 1100. Tech. Note BN-991, Institute for Physical Science and Technology, University of Maryland, October 1982.

[2] W. C. Rheinboldt, Feedback Systems and Adaptivity for Numerical Computations, in "Adaptive Computational Methods for Partial Differential Equations," eds. I. Babuška, J. Chandra, J. E. Flaherty, SIAM Publications, Philadelphia, PA, 1983, pp. 3-19.

[3] I. Babuška, M. Vogelius, Feedback and Adaptive Finite Element Solution of One-Dimensional Boundary Value Problems, Numerische Mathem. 44, 1984, pp. 75-103.

[4] I. Babuška, A. Miller, M. Vogelius, Adaptive Method and Error Estimation for Elliptic Problems of Structural Mechanics, in "Adaptive Computational Method for Partial Differential Equations, eds. I. Babuška, J. Chandra, J. E. Flaherty, SIAM Publications, Philadelphia, PA, 1983, pp. 35-56.

[5] C. Mesztenyi, A. Miller, W. Szymczak, FEARS Details of Mathematical Formulation, Tech. Note BN-994, Institute for Physical Science and Technology, University of Maryland, December 1982.

[6] I. Babuška, W. C. Rheinboldt, Error estimates for adaptive finite element computations, SIAM, J. Numer. Anal. 15 (1978), pp. 736-754.

The Laboratory for Numerical Analysis is an integral part of the Institute for Physical Science and Technology of the University of Maryland, under the general administration of the Director, Institute for Physical Science and Technology. It has the following goals:

- To conduct research in the mathematical theory and computational implementation of numerical analysis and related topics, with emphasis on the numerical treatment of linear and nonlinear differential equations and problems in linear and nonlinear algebra.

- To help bridge gaps between computational directions in engineering, physics, etc. and those in the mathematical community.

- To provide a limited consulting service in all areas of numerical mathematics to the University as a whole, and also to government agencies and industries in the State of Maryland and the Washington Metropolitan area.

- To assist with the education of numerical analysts, especially at the postdoctoral level, in conjunction with the Interdisciplinary Applied Mathematics Program and the programs of the Mathematics and Computer Science Departments. This includes active collaboration with government agencies such as the National Bureau of Standards.

- To be an international center of study and research for foreign students in numerical mathematics who are supported by foreign governments or exchange agencies (Fulbright, etc.).

Further information may be obtained from Professor I. Babuška, Chairman, Laboratory for Numerical Analysis, Institute for Physical Science and Technology, University of Maryland, College Park, Maryland 20742.

END

FILMED

1-85

DTIC