ELSEVIER

Comput. Methods Appl. Mech. Engrg. 142 (1997) 1–88

Computer methods in applied mechanics and engineering

Computational Mechanics Advances

A posteriori error estimation in finite element analysis

Mark Ainsworth [a,*], J. Tinsley Oden [b]

[a] *Department of Mathematics, University of Leicester, Leicester, UK*

[b] *Texas Institute for Computational and Applied Mathematics, The University of Texas at Austin, Austin, TX 78712, USA*

Abstract

This monograph presents a summary account of the subject of a posteriori error estimation for finite element approximations of problems in mechanics. The study primarily focuses on methods for linear elliptic boundary value problems. However, error estimation for unsymmetrical systems, nonlinear problems, including the Navier–Stokes equations, and indefinite problems, such as represented by the Stokes problem are included. The main thrust is to obtain error estimators for the error measured in the energy norm, but techniques for other norms are also discussed.

Contents:

1. Introduction 1
 1.1. A posteriori error estimation: the setting 1
 1.2. Status and scope 2
 1.3. Notations 3
 1.4. Approximation spaces 5
 1.5. Model problem 7
 1.6. Properties of a posteriori error estimators 7
2. Estimators based on gradient recovery 9
 2.1. A priori and a posteriori error estimates 9
 2.2. Complementary variational principles 10
 2.3. Recovery operators 13
 2.4. The superconvergence property 15
 2.5. Examples of recovery based estimators 16
 2.6. Summary 20
3. Explicit a posteriori estimators 21
 3.1. Introduction 21
 3.2. A simple a posteriori error estimate 21
 3.3. A simple error estimator in the L_2 norm 23
 3.4. Equivalence of estimator 24
 3.5. The effect of numerical quadrature 29
 3.6. Error estimators for $W^1_p(\Omega)$ and $L_p(\Omega)$ 29
4. Implicit a posteriori estimators 32
 4.1. Introduction 32
 4.2. The subdomain residual method 32
 4.3. The element residual method 36
 4.4. Equivalence of estimator 39
 4.5. Performance of estimators 41
5. The equilibrated residual method 48
 5.1. Introduction 48
 5.2. A posteriori error analysis 49
 5.3. The equilibration principle 54
 5.4. Construction of equilibrated fluxes 56
 5.5. A posteriori error estimators 65
 5.6. The equilibrated residual method 66
 5.7. Treatment of the local spaces $\mathcal{B}^{(q)}$ 69
6. Applications 74
 6.1. Stokes and Oseen's equations 74
 6.2. Incompressible Navier–Stokes equations 81
 6.3. Variational inequalities 84
Acknowledgement 86
References 86

1. Introduction

1.1. A posteriori error estimation: the setting

Since the beginning of computer simulations of physical events, the presence of numerical error in calculations has been a principal source of concern. Numerical error is intrinsic in such simulations: the

*Corresponding author.

PII S0045-7825(96)01107-3

discretization process of transforming a continuum model of mechanical behavior into one manageable by digital computers cannot capture all of the information embodied in models characterized by partial differential equations or integral equations. What is the approximation error in such simulations? How can the error be measured, controlled and effectively minimized? These questions have confronted computational mechanicians, practitioners and theorists alike, since the earliest applications of numerical methods to problems in engineering and science.

Recent years have seen concrete advances toward the resolution of these questions made in the form of theories and methods of a posteriori error estimation, whereby the computed solution itself is used to somehow assess the accuracy. The remarkable success of some a posteriori error estimators has opened a new chapter in computational mathematics and mechanics that could revolutionize the subject. By effectively estimating error, the possibility of controlling the entire computational process through new adaptive algorithms emerges. Fresh criteria for judging the performance of algorithms become apparent. Most importantly, the analyst can use a posteriori error estimates as an independent measure of the quality of the simulation under study.

The present work is intended to provide an introduction to the subject of a posteriori error estimation for finite element approximations of boundary value problems in mechanics and physics. The treatment is by no means exhaustive, focusing primarily on elliptic partial differential equations and on the chief methods currently available. However, extensions to unsymmetrical systems of partial differential equations, nonlinear problems and indefinite problems are included. Our aim is to present a coherent summary of a posteriori error estimation methods.

1.2. Status and scope

The a priori estimation of errors in numerical methods has long been an enterprise of numerical analysts. Such estimates give information on the convergence and stability of various solvers and can give rough information on the asymptotic behaviour of errors in calculations as mesh parameters are appropriately varied. Traditionally, the practitioner using numerical simulations, while aware that errors exist, is rarely concerned with quantifying them. The quality of a simulation is generally assessed by physical or heuristic arguments based on the experience and judgment of the analyst. Frequently such arguments are later proved to be flawed.

Some of the earliest a posteriori error estimates used in computational mechanics were in the solution of ordinary differential equations. These are typified by predictor corrector algorithms in which the difference in solutions obtained by schemes with different orders of truncation error is used as rough estimates of the error. This estimate can in turn be used to adjust the time step. It is notable that the original a posteriori error estimation schemes for elliptic problems had many features that resemble those for ordinary differential equations.

Interest in a posteriori error estimation for finite element methods for two point elliptic boundary value problems really began with the pioneering work of Babuska and Rheinboldt [13]. A posteriori error estimation techniques were developed that delivered numbers η_K approximating the error in energy or an energy norm on each finite element K. These formed the basis of adaptive meshing procedures designed to control and minimize the error. During the period 1978–1983, a number of results for explicit error estimation techniques were obtained: we mention as representatives the work of Babuska and Rheinboldt [11,12].

The use of complementary energy formulations for obtaining a posteriori error estimates was put forward by de Veubeke [27]. However, the method failed to gain popularity being based on a global computation. The idea of solving element by element complementary problems together with the important concept of constructing *equilibrated boundary data* to obtain error estimates was advanced by Ladeveze and Leguillon [43]. Related ideas are found in the work of Kelly [39] and Stein and Ahmad [53].

In 1984, an important conference on adaptive refinement and error estimation was held in Lisbon (see [20]). At that meeting, several new developments in a posteriori error estimation were presented, including the *element residual method*. The method was described by Demkowicz et al. [29,30] and applied to a variety of problems in mechanics and physics. Essentially the same process was advanced

simultaneously by Bank and Weiser [23,22] who focused on the applications to scalar elliptic problems in two dimensions and provided a mathematical analysis of the method. The paper of Bank and Weiser [23] also involved a number of basic ideas that proved to be fundamental to certain theories of a posteriori error estimation including the saturation assumption and the equilibration of boundary data in the context of piecewise linear approximation on triangles.

During the early 1980s the search for effective adaptive methods led to a wide variety of ad hoc error estimators. Many of these were based on a priori or interpolation estimates, that provided crude but effective indications of features of error sufficient to drive adaptive processes. In this context, we mention the interpolation error estimates of Demkowicz et al. [28]. In computational fluid dynamics calculations these crude interpolation estimates proved to be useful for certain problems in inviscid flow (see [50]), where solutions featured surfaces of discontinuity, shocks, and rarefaction waves. Relatively crude error estimates are sufficient to locate regions in the domain in which discontinuities appear and these are satisfactory for use as a basis for certain adaptive schemes. However, when more complex features of the solution are present, such as boundary layers or shock-boundary layer interactions, these cruder methods are often disastrously inaccurate.

Zienkiewicz and Zhu [60] developed a simple error estimation technique that is effective for some classes of problem and types of finite element approximations. Their method falls into the category of *recovery based methods*: gradients of solutions obtained on a given mesh are smoothed and the smoothed solution is compared with the original solution to assess error. More recently, Zienkiewicz and Zhu [61,62] modified their approach leading to the *superconvergent patch recovery* method.

Extrapolation methods have been used effectively to obtain global error estimates for both h and p version of the finite element method. For example, by using sequences of hierarchical p version approximations, Szabo [55] obtained efficient a posteriori estimators for two dimensional linear elasticity problems.

By the early 1990s the basic techniques of a posteriori error estimation were established. Attention then shifted to the application to general classes of problem. Verfurth [56] obtained two-sided bounds and derived error estimates for the Stokes problem and the Navier–Stokes An important paper on explicit error residual methods for broad classes of boundary value problems, including nonlinear problems, was presented by Baranger and El-Amri [24]. Erikson et al. [32–34,38] derived a posteriori error estimates for both parabolic and hyperbolic problems.

Most studies have dealt with a posteriori error estimation for the h version of the finite element method. The element residual method is applicable to both p version finite elements and h-p versions finite element approximations. An extensive study of error residual methods is reported in the paper by Oden et al. [47]. These techniques were applied to non-uniform h-p meshes. Later, in a series of papers Ainsworth and Oden [6] produced extensions of the element residual method in conjunction with equilibrated boundary data. This was extended to elliptic boundary value problems, elliptic systems, variational inequalities and indefinite problems such as the Stokes problem and steady Navier–Stokes equations with small data.

The subject of a posteriori error estimation for finite element approximation has now reached maturity. The emphasis has now shifted from the development of new techniques to the study of robustness of existing estimators and identifying limits on their performance. Particularly noteworthy in this respect is the work of Babuska et al. [15,16] who conduct an extensive study of the performance and robustness of the main error estimation techniques applied to first order finite element approximation.

The literature on a posteriori error estimation for finite element approximation is vast. We have strongly resisted the temptation to produce an exhaustive survey. The availability of computer databases means that anyone can generate an up to date survey with minimal effort. Instead, the bibliography consists solely of key references and work having a direct influence on our exposition. Surveys of the earlier literature will be found in [45,46] and more recently in [35].

1.3. Notations

1.3.1. Sobolev spaces

Throughout, standard notations and conventions for function spaces are followed [1]. Let Ω be an open bounded domain in $\mathbb{R}^n$, where $n = 1, 2$ or 3, with boundary Γ.

Integer order spaces

The integer order Sobolev spaces $W^{m,p}(\Omega)$, $m \in \mathbb{Z}^+$, $1 \leqslant p \leqslant \infty$ are equipped with the norm $\|\cdot\|_{W^{m,p}(\Omega)}$ defined by

$$\|u\|_{W^{m,p}(\Omega)} = \left\{ \sum_{|\alpha| \leqslant m} \int_\Omega |D^\alpha u|^p \, \mathrm{d}x \right\}^{1/p} \qquad \text{if } 1 \leqslant p < \infty \tag{1.1}$$

and

$$\|u\|_{W^{m,\infty}(\Omega)} = \max_{|\alpha| \leqslant m} \|D^\alpha u\|_{L^\infty(\Omega)} \qquad \text{if } p = \infty \tag{1.2}$$

where

$$\|u\|_{L^\infty(\Omega)} = \operatorname*{ess\,sup}_{x \in \Omega} |u(x)|. \tag{1.3}$$

The space $W^{m,p}(\Omega)$ itself is the completion of $C^\infty(\overline{\Omega})$ in this norm and is therefore a Banach space. The space $W_0^{m,p}(\Omega)$ is the closure of $C_0^\infty(\overline{\Omega})$ in the norm on $W^{m,p}(\Omega)$ where $C_0^\infty(\overline{\Omega})$ consists of all functions which, together with their derivatives of all orders, are continuous and compactly supported on Ω. In the case $p = 2$, the notations $W^{m,2}(\Omega) = H^m(\Omega)$ and $W_0^{m,2}(\Omega) = H_0^m(\Omega)$ are used.

1.3.2. Partitions

The basic procedure in the finite element method is the partitioning of the computational domain Ω into a collection $\mathcal{P} = \{K\}$ of open subdomains or *elements*. Various sets of assumptions are made on the construction of the partition sufficient to ensure the convergence of the method. More generally, *families* $\mathcal{F} = \{\mathcal{P}\}$ of partitions are considered so that statements may be made concerning the convergence of the sequence of finite element approximations obtained on the partitions. In the present work, various versions of the finite element will be considered including adaptive methods. The partitions used for adaptive meshes are generally disallowed by the classical finite element theory but, nonetheless, must obey strict conditions. For convenience we formulate the particular assumptions on each notion of partition to be considered.

Partition

A *partition* $\mathcal{P}$ of Ω is a collection of elements K satisfying

($\mathcal{P}_1$) $\overline{\Omega} = \cup_{K \in \mathcal{P}} \overline{K}$

($\mathcal{P}_2$) each element K is a non-empty, open subset of Ω

($\mathcal{P}_3$) the intersection of each distinct pair $K, J \in \mathcal{P}$ is empty

($\mathcal{P}_4$) each element K is a triangle or convex quadrilateral

Proper partition

A *proper partition* $\mathcal{P}$ of Ω is a partition of Ω that satisfies the additional condition

($\mathcal{P}_5$) each side of an element K is either a subset of the boundary $\partial\Omega$ or a side of another element J in the partition.

The collection of element sides is denoted by $\partial\mathcal{P}$.

Non-degeneracy condition

Let K be any triangular element from a partition $\mathcal{P}$. Define the diameter h_K of the element by

$$h_K = \operatorname{diam}(K) \tag{1.4}$$

and let $h(\mathcal{P})$ be

$$h(\mathcal{P}) = \max_{K \in \mathcal{P}} h_K. \tag{1.5}$$

Define ρ_K by

$$\rho_K = \sup\{\mathrm{diam}(S) : S \text{ is a ball contained in } \overline{K}\}. \tag{1.6}$$

The partition is called *non-degenerate* or *shape regular* if there exists a constant γ_0 that is independent of $h(\mathcal{P})$ such that

$$\max_{K\in\mathcal{P}} \frac{h_K}{\rho_K} \leqslant \gamma_0. \tag{1.7}$$

The non-degeneracy condition does not require that the elements be of comparable size and permits highly refined meshes. Similar concepts may be defined for quadrilateral elements.

Regular family of partitions

A *regular family* $\mathcal{F}$ of partitions is a collection of proper, non-degenerate partitions $\{\mathcal{P}\}$ with the non-degeneracy constant γ_0 independent of $\mathcal{P}$: and, such that $h(\mathcal{P})$ approaches zero. This condition essentially rules out the use of many adaptive algorithms.

Quasi-uniform family

A family $\mathcal{F}$ of partitions is *quasi-uniform* if each partition $\mathcal{P}$ in $\mathcal{F}$ is regular and there exists a constant γ_1 such that for all partitions $\mathcal{P}$

$$\max_{K\in\mathcal{P}} \frac{h(\mathcal{P})}{h_K} \leqslant \gamma_1. \tag{1.8}$$

Locally quasi-uniform family

A family $\mathcal{F}$ of partitions is *locally quasi-uniform* if each partition $\mathcal{P}$ in $\mathcal{F}$ is proper and is composed of elements satisfying a non-degeneracy assumption with a constant γ_0 independent of $\mathcal{P}$.

Typically, we shall assume that the partitions are locally quasi-uniform. Such partitions can be highly refined and yet satisfy a local quasi-uniformity condition. For instance, let K be a element belonging to a locally quasi-uniform partition $\mathcal{P}$. The patch of elements surrounding K is defined by

$$\tilde{K} = \mathrm{int}\left\{\bigcup \overline{J} : \overline{J} \cap \overline{K} \text{ is non-empty}\right\}. \tag{1.9}$$

The non-degeneracy condition implies that there is a constant C depending only on γ_0 such that for any element J contained in the subdomain $\tilde{K}$ a local quasi-uniformity condition holds on the subdomain

$$\frac{1}{C} \leqslant \frac{h_K}{h_J} \leqslant C. \tag{1.10}$$

Moreover, notice that the number of elements contained in the subdomain $\tilde{K}$ must be uniformly bounded by a constant depending only on γ_0

$$\max_{K\in\mathcal{P}} \mathrm{card}\{J : J \subset \tilde{K}\} \leqslant C. \tag{1.11}$$

Equally well, the number subdomains containing a particular element is uniformly bounded by a constant depending on γ_0

$$\max_{K\in\mathcal{P}} \mathrm{card}\{\tilde{J} : K \subset \tilde{J}\} \leqslant C. \tag{1.12}$$

1.4. Approximation spaces

Reference elements

In both the mathematical analysis of finite element methods and in their application to specific problems, it is natural to consider each element in a partitioning of the domain (or *finite element mesh*) to

be the image of a standard *reference element* $\widehat{K}$. The reference element defines the element type while providing the template on which element computations are performed. For instance, in the case of a triangular element the reference element may be taken as

$$\widehat{K} = \{(\hat{x}, \hat{y}) : 0 \leqslant \hat{x} \leqslant 1; \quad 0 \leqslant \hat{y} \leqslant 1 - \hat{x}\} \tag{1.13}$$

or, in the case of quadrilateral elements

$$\widehat{K} = \{(\hat{x}, \hat{y}) : -1 \leqslant \hat{x} \leqslant 1; \quad -1 \leqslant \hat{y} \leqslant 1\}. \tag{1.14}$$

Polynomial spaces of degree $p \in \mathbb{N}$ are defined on the triangular and quadrilateral reference elements, respectively, by

$$\widehat{P}(p) = \operatorname{span}\{\hat{x}^j \hat{y}^k : 0 \leqslant j + k \leqslant p\} \tag{1.15}$$

and

$$\widehat{Q}(p) = \operatorname{span}\{\hat{x}^j \hat{y}^k : 0 \leqslant j, k \leqslant p\}. \tag{1.16}$$

In three dimensions, hexahedral, tetrahedral or prismatic reference elements are used with analogous polynomial spaces.

Finite element spaces

Let $\mathcal{P}$ be a locally quasi-uniform partitioning of a connected, bounded, polygonal domain Ω into triangular and quadrilateral elements. For simplicity, assume that there exists an invertible mapping $F_K : \widehat{K} \mapsto K$ that is affine for triangular elements and bilinear for quadrilateral elements. Each element is assigned a parameter $p_K \in \mathbb{N}$ controlling the degree of approximation on the element. A polynomial space P_K is thereby selected to be either $\widehat{Q}(p_K)$ or $\widehat{P}(p_K)$ as appropriate. The finite element space X consists of continuous piecewise polynomials

$$X = \{v \in C(\Omega) : v|_K = \hat{v} \circ F_K^{-1} \text{ for some } \hat{v} \in P_K \text{ for all } K \in \mathcal{P}\}. \tag{1.17}$$

If the polynomial degree p_K is non-uniform over the partition $\mathcal{P}$ then the continuity requirements may impose constraints on the approximation in the particular element which mean that the effective polynomial degree within the element is, in essence, reduced. When we speak of the polynomial degree p_K, it will be understood to mean the effective polynomial degree. Further, the subscript K will be often be omitted.

Approximation theory

Approximation theoretic results concerning approximation using piecewise polynomials on partitions will be required. A number of results concerning approximation of continuous functions on regular partitions are known and well documented in the literature [25]. On occasion, it will be necessary to approximate functions that may be discontinuous on partitions that are only supposed to be locally quasi-uniform. Such approximations have been considered by Clément [26] and Scott and Zhang [52]:

THEOREM 1.1. Suppose $\mathcal{P}$ be a locally quasi-uniform partitioning of the domain $\Omega \subset \mathbb{R}^2$. Let K be any element in the partition and denote

$$\tilde{K} = \operatorname{int}\left\{\bigcup \bar{J} : \bar{J} \cap \bar{K} \text{ is non-empty}\right\}. \tag{1.18}$$

Suppose that $0 \leqslant l \leqslant p + 1$ and m are integers and $r, s \in [1, \infty]$ are chosen so that the embedding $W^{l,r}(K) \hookrightarrow W^{m,s}(K)$ holds. Then, for any $v \in W^{l,r}(\tilde{K})$ there exists $\Pi_X v \in X$ such that

$$\|v - \Pi_X v\|_{W^{m,s}(K)} \leqslant C(\gamma_0, p) h_K^{l-m+2(1/s-1/r)} \|v\|_{W^{l,r}(\tilde{K})}. \tag{1.19}$$

1.5. Model problem

Let $\Omega \subset \mathbb{R}^2$ be a bounded domain with Lipschitz boundary $\partial\Omega$. Consider the model elliptic boundary value problem of finding the solution u of

$$-\Delta u + cu = f(x) \quad \text{in } \Omega \tag{1.20}$$

subject to the boundary conditions

$$\frac{\partial u}{\partial n} = g \quad \text{on } \Gamma_N \tag{1.21}$$

and

$$u = 0 \quad \text{on } \Gamma_D. \tag{1.22}$$

The data are assumed to be smooth, i.e. $f \in L_2(\Omega)$, $g \in L_2(\Gamma_N)$, c is a non-negative constant and the boundary segments Γ_N, Γ_D are assumed to be disjoint with $\overline{\Gamma}_N \cup \overline{\Gamma}_D = \partial\Omega$. The unit outward normal vector to $\partial\Omega$ is denoted by n and belongs to the space $[L^\infty(\partial\Omega)]^n$.

The variational form of this problem is to find $u \in V$ such that

$$B(u, v) = L(v) \quad \forall v \in V \tag{1.23}$$

where V is the space

$$V = \{v \in H^1(\Omega) : v = 0 \quad \text{on } \Gamma_D\} \tag{1.24}$$

and where

$$B(u, v) = \int_\Omega (\nabla u \cdot \nabla v + cuv)\, dx \tag{1.25}$$

and

$$L(v) = \int_\Omega fv\, dx + \int_{\Gamma_N} gv\, ds. \tag{1.26}$$

Suppose that $X \subset V$ is a finite element subspace. The finite element approximation of this problem is to find $u_X \in X$ such that

$$B(u_X, v_X) = L(v_X) \quad \forall v_X \in X. \tag{1.27}$$

The error $e = u - u_X$ belongs to the space V and satisfies

$$B(e, v) = B(u, v) - B(u_X, v) = L(v) - B(u_X, v) \quad \forall v \in V. \tag{1.28}$$

Moreover, the standard orthogonality condition for the error in the Galerkin projection holds

$$B(e, v_X) = 0 \quad \forall v_X \in X. \tag{1.29}$$

1.6. Properties of a posteriori error estimators

There are many techniques for error estimation. One can extrapolate approximate solutions obtained on sequences of progressively finer meshes or on sequences of meshes with shape functions of increasing order and then compare solutions to obtain an indication of the error. Such methods can be quite effective when data structures admit such multilevel computations. One popular method amongst the engineering community is to postprocess the approximation u_X to obtain more accurate representations of the gradient $G(u_X)$. One can then use the difference $G(u_X) - \nabla u_X$ as an estimate for the error. This type of approach can lead to surprisingly good error estimators and is discussed in Section 2. One of the weaknesses of the method and at the same time, one of its advantages, is that no use is made of the information from the original problem.

Other error estimators make use of the data for the problem and properties of the error in various ways. For instance, the approximation error satisfies the residual equation (1.28) and the orthogonality

condition (1.29). A residual equation similar to (1.28) may be obtained by integrating by parts over each element leading to

$$\int_K (\nabla e \cdot \nabla v + cev)\,\mathrm{d}x = \int_K rv\,\mathrm{d}x + \oint_{\partial K} v\,(n_K \cdot \nabla u - n_K \cdot \nabla u_X)\,\mathrm{d}s \tag{1.30}$$

where r is the *residual*

$$r = f + \Delta u_X - cu_X \tag{1.31}$$

and n_K is a unit exterior normal to the element boundary ∂K. Under suitable conditions, the solution to (1.30) may be bounded by

$$|||e|||_K \leqslant C_1 \|r\|_{L_2(K)} + C_2 \|n_K \cdot \nabla e\|_{L_2(\partial K)} \tag{1.32}$$

where C_1 and C_2 may depend upon the element size h_K and other mesh parameters. Replacing the flux on the element boundary by a suitable approximation leads to a bound on the error on the element K (apart from the constants C_1 and C_2). This type of estimator is referred to as an *explicit estimator* and is discussed in Section 3.

The presence of the constants C_1 and C_2 in the explicit a posteriori error estimators leads one to consider trying to solve an approximate local boundary value problem for the error of the form

$$\int_K (\nabla \phi_K \cdot \nabla v + c\phi_K v)\,\mathrm{d}x = \int_K rv\,\mathrm{d}x + \oint_{\partial K} v\,(g_K - n_K \cdot \nabla u_X)\,\mathrm{d}s \tag{1.33}$$

where g_K is an approximation to the boundary flux. The solution ϕ_K may be used as follows:

$$|||\phi_K|||^2 = \int_K (|\nabla \phi_K|^2 + c\phi_K^2)\,\mathrm{d}x \tag{1.34}$$

to provide a measure of the error content in the approximation associated with element K. The approach raises a number of issues:

- the infinite dimensional space containing the error must be approximated by an appropriate finite dimensional subspace
- the boundary flux $n_K \cdot \nabla u$ must be approximated in some effective way
- if $c = 0$ the error residual problem may have no solution unless the condition

$$\int_K r\,\mathrm{d}x + \oint_{\partial K} (g_K - n_K \cdot \nabla u_X)\,\mathrm{d}s = 0 \tag{1.35}$$

is satisfied.

The general process just described is an example of an *implicit error residual method*. It is said to be implicit because the error residual problem must be solved over each element to determine the *error estimator* $|||\phi_K|||$. Implicit estimators are the subject of Section 4. Section 5 deals with implicit error estimators in which the boundary fluxes are specially constructed so that the local problem is well posed. The resulting estimators may be shown to possess several very powerful properties.

If η_K is a local error estimator on element K then the global error estimate η is usually taken to be

$$\eta = \left\{ \sum_{K \in \mathcal{P}} \eta_K^2 \right\}^{1/2}. \tag{1.36}$$

A major property demanded of all successful error estimators is that positive constants C_1, C_2 exist such that

$$C_1 |||e||| \leqslant \eta \leqslant C_2 |||e||| \tag{1.37}$$

where $|||e|||$ is the global error in the energy norm. Then η tends to zero at the same *rate* as the true error. The quality of an estimator is often judged by global *effectivity indices*

$$\frac{\eta}{|||e|||} \tag{1.38}$$

or *local effectivity indices*

$$\frac{\eta_K}{|||e|||_K}. \tag{1.39}$$

These indices can be used to measure the quality of an estimator when the exact error or a good approximation of it are known. Naturally, one hopes that effectivity indices close to unity can be obtained, but global effectivity indices of 2.0–3.0 or even higher are often regarded as acceptable in many engineering applications.

Throughout, the ideas are presented for the simple model problem in the plane. For the most part, the analysis may be easily extended to three dimensions. Therefore, we shall comment on higher dimensions only in cases where the extension is not immediately apparent. The extension of the results to more general problems may be less straightforward. Therefore, in Section 6 applications to more complicated problems are given including problems with side constraints (such as the Stokes' problem); unsymmetric operators (Oseen's equations); non-linearities (the incompressible Navier–Stokes' equations); and unilateral constraints (the obstacle problem).

Essentially, all of the results are already known in the literature. However, it is hoped that by presenting the results in a single notation the interrelation between different techniques will be more apparent. Furthermore, in many cases the presentation is much simpler than the original references. Sections 2, 3 and 4 may be read independently. Section 5 is also largely independent of the earlier sections, although the reader might find it helpful to first read Section 4.

2. Estimators based on gradient recovery

A particularly simple approach to error estimation consists of obtaining a continuous approximation to the gradient by postprocessing the gradient of the finite element approximation. These often rather crude methods can result in surprisingly good approximations to the true gradient. The difference between the postprocesses approximation and the direct approximation is then used as an estimate of the error, often quite successfully. The current section follows [4] and attempts to provide a simple framework for analyzing such recovery based estimators.

2.1. A priori and a posteriori error estimates

Consider the model problem in Section 1.5. Typically, the error in the finite element approximation may be bounded a priori by an estimate of the form

$$|||e||| \leqslant Ch^p \|u\|_{H^{p+1}(\Omega)} \tag{2.1}$$

where C is a constant independent of h and u; and $|||\cdot|||$ is the energy norm for the problem. The a priori estimate reveals the rate of convergence but is of limited use if one requires a numerical estimate of the accuracy. The difficulty is that either the constant C is unknown or if bounds are found on C, then the estimate will generally be extremely pessimistic. Equally well, the higher derivatives of the true solution u are unknown.

One way to improve the prospects of finding a reasonable estimate of the discretization error is to use the finite element approximation itself. Error estimators can be easily developed using heuristic arguments as follows. Suppose the coefficient $c \equiv 0$; then the expression for the true error is

$$|||e|||^2 = \int_\Omega |\nabla u - \nabla u_X|^2 \, \mathrm{d}\boldsymbol{x}. \tag{2.2}$$

If the true gradient were known then it would be a relatively easy matter to substitute it into this expression and to calculate the true error exactly. Intuitively, a reasonable error estimator should be obtained using an approximation to the gradient in place of ∇u.

A technique that is popular with the engineering community is the averaging method. The gradient of the finite element approximation provides a (discontinuous) approximation to the true gradient. This may be used to construct an approximation at each node by averaging contributions from each of the elements surrounding the node. These values may be interpolated to obtain a continuous approximation over the whole domain. While the method is apparently rather crude, it can be astonishingly effective.

The case $c \not\equiv 0$ is dealt with by arguing that the dominant term in the error is the component containing the derivatives, and so it should be enough to estimate this dominant part only. In effect, the absolute term is simply ignored.

The intuitive argument is appealing but does little to provide confidence in the resulting estimator. Several estimators actually used in practice are based on replacing ∇u by a quantity which is believed to be a good approximation. Part of the reason for the popularity of such methods is that frequently, a suitable gradient (or stress) recovery module is already implemented in the finite element code.

Conversely, it is found that some rigorously analyzed estimators obtained in quite different ways fall within the framework of corresponding to a particular choice of recovered gradient. The next section is devoted to developing a general result for analyzing estimators falling within this framework.

2.2. Complementary variational principles

The error in the finite element approximation is the solution of a boundary value problem is analogous to (1.23). In fact, replacing $u = e + u_X$ in (1.23) and rearranging gives

$$e \in V : B(e, v) = L(v) - B(u_X, v) \quad \forall v \in V. \tag{2.3}$$

The function u_X is regarded as being known explicitly since we envisage using u_X itself in obtaining estimates of the error. Equally well, e is the solution to a variational problem

$$e \in V : \quad J(e) \leqslant J(w) \quad \forall w \in V \tag{2.4}$$

where J is the quadratic functional

$$J(w) = \frac{1}{2} B(w, w) - L(w) + B(u_X, w). \tag{2.5}$$

There is a unique solution to (2.4) since $B(\cdot,\cdot)$ and $L(\cdot)$ are bounded on V. Notice that using (1.23) gives

$$\begin{aligned} J(e) &= \frac{1}{2} B(e, e) - L(e) + B(u_X, e) \\ &= \frac{1}{2} B(e, e) - B(u, e) + B(u_X, e) \\ &= -\frac{1}{2} B(e, e) = -\tfrac{1}{2} |||e|||^2. \end{aligned} \tag{2.6}$$

This result in conjunction with (2.4) gives

$$|||e|||^2 = -2J(e) \geqslant -2J(w) \quad \forall w \in V. \tag{2.7}$$

An interesting consequence of (2.7) is that for any $w \in V$ we can calculate a lower bound on the error $|||e|||$. In general, the lower bound will be poor, or even trivial, unless w is chosen suitably. The best choice is $w = e$.

In practice, we are interested in finding an upper bound on the error. An alternative variational principle is associated with the *primal* variational problem (2.4). This *complementary* variational principle may be used in a similar manner to the primal principle with the important difference that an upper bound is obtained. For instance, consider Poisson's equation in $\mathbb{R}^2$:

$$-\nabla^2 u = f \quad \text{in } \Omega; \qquad u = 0 \quad \text{on } \partial\Omega. \tag{2.8}$$

The primal problem for the error is

$$e \in V : J(e) \leqslant J(w) \quad \forall w \in V \tag{2.9}$$

where

$$J(w) = \frac{1}{2} \int_\Omega |\nabla w|^2 \, d\mathbf{x} - \int_\Omega f w \, d\mathbf{x} + \int_\Omega \nabla u_X \cdot \nabla w \, d\mathbf{x}. \tag{2.10}$$

The complementary problem is to find $\boldsymbol{p}$ such that

$$p \in W : \mathcal{G}(p) \geqslant \mathcal{G}(q) \quad \forall\, q \in W \tag{2.11}$$

where $\mathcal{G}$ is the quadratic functional

$$\mathcal{G}(q) = -\frac{1}{2} \int_\Omega |q - \nabla u_X|^2 \, \mathrm{d}x \tag{2.12}$$

and W is the set

$$W = \{q \in H(\mathbf{div}, \Omega) : \nabla \cdot q + f = 0 \text{ in } \Omega\} \tag{2.13}$$

with

$$H(\mathbf{div}, \Omega) = \{q \in L_2(\Omega) \times L_2(\Omega) : \mathbf{div}\, q \in L_2(\Omega)\}. \tag{2.14}$$

It may be shown that the unique solution of the complementary problem is $p = \nabla u$ and

$$-2\mathcal{G}(\nabla u) = |||e|||^2. \tag{2.15}$$

Combining these results gives

$$|||e||| \leqslant \sqrt{-2\mathcal{G}(q)} \quad \forall\, q \in W. \tag{2.16}$$

Therefore, to obtain a computable upper bound on $|||e|||$, we need only make a suitable choice of q to substitute into the functional $\mathcal{G}(q)$. The best choice is $q = \nabla u$. However, the constraint condition (2.13) on the choice of functions is the main drawback, making the construction of feasible functions q awkward.

One possibility is to obtain a suitable q by means of a finite element discretization of the complementary problem as suggested by Aubin and Burchard [9] and deVeubeke [27]. The method requires the solution of a global problem, essentially to satisfy continuity requirements. Unfortunately, the computational effort required in the solution of any global problem is comparable with that of obtaining the finite element approximation itself, in which case it would be simpler to resolve the original problem using a finer discretization. It ought to be unnecessary to carry out any further global computation since there is already global information in the finite element approximation itself to enable a sensible choice of q to be made giving a realistic bound on the error.

Another difficulty is that the equality constraint on q

$$\nabla \cdot q + f = 0 \quad \text{in } \Omega \tag{2.17}$$

must be satisfied exactly. This rules out any possibility of using a simple function q, unless f is itself simple. One would expect it to be sufficient to satisfy the condition approximately.

The constraint may be relaxed by making use of a device used by Babuska and Rheinboldt [12]. Firstly, we define a new bilinear form $\tilde{B}(\cdot, \cdot)$ by

$$\tilde{B}(u, v) = \int_\Omega \nabla u \cdot \nabla v \, \mathrm{d}x + \int_\Omega \lambda u v \, \mathrm{d}x \tag{2.18}$$

where $\lambda > 0$ is a constant specified later. The following problem is a perturbed version of (2.3):

$$y \in V : \tilde{B}(y, w) = L(w) - B(u_X, w) \quad \forall\, w \in V. \tag{2.19}$$

The solution may be characterized as the solution of the primal variational problem

$$y \in V : \tilde{J}(y) \leqslant \tilde{J}(w) \quad \forall\, y \in V \tag{2.20}$$

where

$$\tilde{J}(w) = \frac{1}{2}\, \tilde{B}(w, w) - L(w) + B(u_X, w). \tag{2.21}$$

The following theorem gives the complementary principle associated with the perturbed primal problem.

THEOREM 2.1. Let $\tilde{\mathcal{G}}(p)$ be the quadratic functional on $H(\mathbf{div}, \Omega)$ given by

$$\tilde{\mathcal{G}}(p) = \int_\Omega |p - \nabla u_X|^2 \, \mathrm{d}x + \int_\Omega \frac{1}{\lambda} (f + \nabla \cdot p - c u_X)^2 \, \mathrm{d}x. \tag{2.22}$$

Then the following bound holds:

$$\widetilde{B}(y,y) = \widetilde{\mathcal{G}}\left[\nabla(u_X + y)\right] \leqslant \widetilde{\mathcal{G}}(\boldsymbol{p}) \quad \forall \boldsymbol{p} \in H(\mathbf{div}, \Omega). \tag{2.23}$$

PROOF. At a stationary point of the variational problem (2.23)

$$\nabla \cdot [\nabla(u_X + y)] + f - cu_X = \lambda y. \tag{2.24}$$

Let $\boldsymbol{p} = \nabla(y + u_X)$. Since both u_X and y belong to V

$$\nabla \cdot \boldsymbol{p} = cu_X - f + \lambda y \in L^2(\Omega). \tag{2.25}$$

Consequently, $\boldsymbol{p} \in H(\mathbf{div}, \Omega)$. A direct calculation using (2.24) shows that

$$\widetilde{\mathcal{G}}(\boldsymbol{p}) = \int_\Omega |\nabla y|^2 \,\mathrm{d}\boldsymbol{x} + \lambda \int_\Omega y^2 \,\mathrm{d}\boldsymbol{x} = \widetilde{B}(y,y). \tag{2.26}$$

Now, let $\eta \in [0,1]$ and $\boldsymbol{q}, \boldsymbol{r} \in H(\mathbf{div}, \Omega)$. It is easily shown that

$$\widetilde{\mathcal{G}}[(1-\eta)\boldsymbol{r} + \eta\boldsymbol{q}] \leqslant (1-\eta)\widetilde{\mathcal{G}}(\boldsymbol{r}) + \eta\widetilde{\mathcal{G}}(\boldsymbol{q}) \tag{2.27}$$

so $\widetilde{\mathcal{G}}$ is convex. Furthermore

$$\begin{aligned}\frac{1}{2}\frac{\mathrm{d}}{\mathrm{d}\eta}\{\widetilde{\mathcal{G}}[(1-\eta)\boldsymbol{p} + \eta\boldsymbol{q}]\}|_{\eta=0} &= \int_\Omega (\boldsymbol{q}-\boldsymbol{p}) \cdot \nabla y \,\mathrm{d}\boldsymbol{x} + \int_\Omega y\nabla \cdot (\boldsymbol{q}-\boldsymbol{p}) \,\mathrm{d}\boldsymbol{x} \\ &= \int_\Omega \nabla \cdot [y(\boldsymbol{q}-\boldsymbol{p})] \,\mathrm{d}\boldsymbol{x} \\ &= \int_{\partial\Omega} y(\boldsymbol{q}-\boldsymbol{p}) \cdot \boldsymbol{n} \,\mathrm{d}s = 0\end{aligned} \tag{2.28}$$

where we have used (2.24) and recalled that $y \in V$. Thus, $\widetilde{\mathcal{G}}$ is stationary at $\boldsymbol{p}$ and the result follows. □

The result in Theorem 2.1 shows that the functional $\widetilde{\mathcal{G}}(\boldsymbol{p})$ delivers an upper bound on y measured in the perturbed energy norm $|||y|||_* = \sqrt{\widetilde{B}(y,y)}$. In essence the result is similar to (2.16). However, there is an important difference. If (2.23) is used to find an upper bound on $|||y|||_*$, then there is no equality constraint to satisfy. This makes (2.23) a more amenable result. However, the bounds are on $|||y|||_*$ instead of $|||e|||$. The fact that y is the solution of a perturbed version of (2.3) characterizing e means there is a relationship between the functional $\widetilde{\mathcal{G}}(\boldsymbol{p})$ and $|||e|||$. The following result quantifies this statement.

THEOREM 2.2. Suppose that there exist positive constants C, μ such that

$$\|e\|_{L^2(\Omega)} \leqslant Ch^\mu |||e||| \tag{2.29}$$

where C and μ do not depend on h or u. Let $\lambda = Mh^{-\mu}$ where $M > 0$ is a constant. Then for any $\boldsymbol{p} \in H(\mathbf{div}, \Omega)$ and h sufficiently small

$$|||e|||^2 \leqslant \{1 + \mathrm{O}(h^\mu)\}\widetilde{\mathcal{G}}(\boldsymbol{p}) \tag{2.30}$$

where the constant in the $\mathrm{O}(h^\mu)$ term is independent of u and h.

REMARK. If Ω is a convex domain then the Aubin–Nitsche Trick shows that the assumption (2.29) holds with $\mu = 1$. If the domain Ω is not convex, then the assumption holds with $\mu > 0$ depending on the maximum interior angle.

PROOF. From (2.3) and (2.19) we have

$$B(e,w) = \widetilde{B}(y,w) \quad \forall w \in V. \tag{2.31}$$

Since $e \in V$ and $y \in V$

$$|||e|||^2 = \tilde{B}(e,y) = B(e,y) + ((\lambda - c)e,y) = \tilde{B}(y,y) + ((\lambda - c)e,y). \tag{2.32}$$

For h sufficiently small, $0 \leqslant c \leqslant \lambda$ and so

$$\begin{aligned} |((\lambda - c)e,y)| &\leqslant 2\lambda \|e\|_{L_2(\Omega)} \|y\|_{L_2(\Omega)} \\ &\leqslant 2\lambda^{1/2} \|e\|_{L_2(\Omega)} |||y|||. \\ &\leqslant 2C\lambda^{1/2} h^{\mu} |||e||| \, |||y|||. \end{aligned} \tag{2.33}$$

where (2.29) was used. Replacing λ with $Mh^{-\mu}$ and using the elementary inequality $2ab \leqslant a^2 + b^2$, we obtain

$$|((\lambda - c)e,y)| \leqslant CM^{1/2} h^{\mu/2} (|||e|||^2 + |||y|||^2_*). \tag{2.34}$$

Rearranging (2.33) gives for sufficiently small h

$$|||e|||^2 \leqslant \{1 + O(h^{\mu})\} |||y|||^2_* \tag{2.35}$$

and the result follows on applying Theorem 2.1. □

Theorem 2.2 shows that the functional associated with the perturbed primal problem can be used to obtain approximate upper bounds on $|||e|||$. The introduction of the perturbed variational formulation has resulted in the equality constraint being removed at the expense of introducing a second term into $\tilde{\mathcal{G}}(\boldsymbol{p})$.

Theorem 2.2 is a tool that may be used in the analysis of the various heuristically proposed error estimators. If such an estimator can be shown to be related to a particular choice of $\boldsymbol{p}$ in (2.30), then Theorem 2.2 immediately shows that the resulting estimator will be an asymptotic upper bound on the error. There are many heuristically proposed estimators to be found in the literature, yet it is found that many of them may be profitably viewed as corresponding to a particular choice of $\boldsymbol{p}$ in Theorem 2.2. In addition to the heuristically based estimators, some rigorously analyzed estimators are also found to fit in this scheme.

2.3. Recovery operators

In this section we define and analyze a class of schemes that make use of the finite element approximation u_X to find a suitable approximation to ∇u. Later, we analyze the properties of the resulting a posteriori error estimators.

In particular, we shall identify a set of conditions sufficient for the operators G_X guaranteeing that $G_X(\Pi_p u)$ is a good approximation to the true gradient.

Consistency condition

Naturally, if the error estimator is to be asymptotically exact, the recovery scheme must give an approximation consistent with the true gradient in certain circumstances.

(R1) If u belongs to the finite element subspace of order $p+1$ then

$$G_X(\Pi_p u) \equiv \nabla u \tag{2.36}$$

where Π_p is the interpolation operator of degree p.

The consistency condition does not determine G_X uniquely and provides a convenient and simple criterion to work with.

Localization condition

An important practical requirement is that G_X should be inexpensive to compute. Specifically, it must be possible to compute G_X without recourse to global computations: otherwise it would be simpler to resolve the original finite element problem on a finer mesh. Particularly convenient schemes are those where the recovered gradient at a point $\boldsymbol{x}^*$ is a linear combination of values of the gradient of the finite element approximation sampled in a neighbourhood of the point $\boldsymbol{x}^*$. Let $\tilde{K}$ denote the patch consisting

of the element K and its neighbouring elements

$$\tilde{K} = \operatorname{int}\left\{\bigcup J : \bar{J} \cap \bar{K} \neq \emptyset\right\}. \tag{2.37}$$

The localization condition is:

(R2) If $x^* \in K$ then the value of the recovered gradient $G_X[v](x^*)$ depends only on values of ∇v sampled on the patch $\tilde{K}$.

Boundedness and linearity conditions

Ideally, G_X should be a simple function that may be evaluated and integrated easily. If G_X is similar to functions belonging to the space used to construct the finite element subspace then existing routines from the finite element code may be used to manipulate G_X. These considerations lead to

(R3) $G_X : X \to X \times X$ is a linear operator and there exists a constant C (independent of h) such that

$$\|G_X[v]\|_{L_\infty(K)} \leqslant C\,|v|_{W^{1,\infty}(\tilde{K})} \quad \forall K \in \mathcal{P} \quad \forall v \in X \tag{2.38}$$

where X consists of finite elements of order p.

2.3.1. Approximation properties of G_X

The conditions (R1)–(R3) imply the operator G_X possesses various approximation properties. In particular, when u is smooth, $G_X(\Pi_p u)$ is a good approximation to ∇u.

LEMMA 2.3. Suppose that G_X satisfies (R1)–(R3) and that $u \in H^{p+2}(\tilde{K})$. Then

$$\|\nabla u - G_X(\Pi_p u)\|_{L_2(K)} \leqslant Ch^{p+1}\,|u|_{H^{p+2}(\tilde{K})} \tag{2.39}$$

where $C > 0$ is independent of h and u.

PROOF. Let

$$F[u](x) = [\nabla u - G_X(\Pi_p u)](x) \quad x \in K. \tag{2.40}$$

Suppose $u \in H^{p+2}(\tilde{K})$. By (R1) and the linearity of G_X

$$\begin{aligned}\|F[u]\|_{L_\infty(K)} &= \|F[u - \Pi_{p+1}u]\|_{L_\infty(K)} \\ &\leqslant |u - \Pi_{p+1}u|_{W^{1,\infty}(\tilde{K})} + \|G_X(\Pi_p(u - \Pi_{p+1}u))\|_{L_\infty(K)}.\end{aligned} \tag{2.41}$$

With the aid of (R3)

$$\|G_X[\Pi_p(u - \Pi_{p+1}u)]\|_{L_\infty(K)} \leqslant C\,|\Pi_p(u - \Pi_{p+1}u)|_{W^{1,\infty}(\tilde{K})}. \tag{2.42}$$

The mesh is quasi-uniform on the patch $\tilde{K}$ so

$$|\Pi_p(u - \Pi_{p+1}u)|_{W^{1,\infty}(\tilde{K})} \leqslant Ch^{-1}\,\|\Pi_p(u - \Pi_{p+1}u)\|_{L_\infty(\tilde{K})} \tag{2.43}$$

and

$$\|\Pi_p(u - \Pi_{p+1}u)\|_{L_\infty(\tilde{K})} \leqslant C\,\|u - \Pi_{p+1}u\|_{L_\infty(\tilde{K})}. \tag{2.44}$$

Hence,

$$\|F[u]\|_{L_\infty(K)} \leqslant |u - \Pi_{p+1}u|_{W^{1,\infty}(\tilde{K})} + Ch^{-1}\,\|u - \Pi_{p+1}u\|_{L_\infty(\tilde{K})}. \tag{2.45}$$

Theorem 1.1 then shows that

$$\|F[u]\|_{L_\infty(K)} \leqslant Ch^{p}\,|u|_{H^{p+2}(\tilde{K})}. \tag{2.46}$$

Finally, noting that

$$\|F[u]\|_{L_2(K)} \leqslant Ch\,\|F[u]\|_{L_\infty(K)} \leqslant Ch^{p+1}\,|u|_{H^{p+2}(\tilde{K})} \tag{2.47}$$

gives the desired result. □

This local result can be used to obtain a global estimate:

LEMMA 2.4. Suppose that G_X satisfies (R1)–(R3) and that $u \in H^{p+2}(\Omega)$. Then

$$\|\nabla u - G_X(\Pi_p u)\|_{L_2(\Omega)} \leqslant Ch^{p+1} |u|_{H^{p+2}(\Omega)} \tag{2.48}$$

where $C > 0$ is independent of h and u.

PROOF. Sum over the elements using Lemma 2.3. □

2.4. The superconvergence property

It has been shown that if a recovery operator G_X is found satisfying the conditions (R1)–(R3), then applying the operator to $\Pi_p u$ furnishes us with good approximations to the derivatives of u. Consider now the effect of applying G_X to the finite element approximation itself. In some circumstances, for instance if the *superconvergence phenomenon* is present, then applying G_X to the finite element approximation itself gives equally good approximations to the derivatives.

The superconvergence property can be appreciated by recalling the standard a priori error estimate

$$|||u - u_X||| \leqslant Ch^p |u|_{H^{p+1}(\Omega)}. \tag{2.49}$$

The estimate (2.49) is optimal in the sense that the exponent of h is the largest possible. In fact, for the h-version finite element method one has that

$$|||e||| \geqslant C(u)h^p \tag{2.50}$$

for some positive constant $C(u)$ depending only on u. The constant $C(u)$ vanishes only in trivial cases.

Superconvergence is present if, under appropriate regularity conditions on the partition and the true solution, an estimate of the form

$$|u_X - \Pi_p u|_{H^1(\Omega)} \leqslant C(u)h^{p+1} \tag{2.51}$$

holds. Comparing (2.50) and (2.51) shows that ∇u_X is a better approximation to $\nabla \Pi_p u$ than it is to ∇u. This will be referred to as the superconvergence property.

(SC) There exists a constant C independent of h such that

$$|u_X - \Pi_p u|_{H^1(\Omega)} \leqslant C(u)h^{p+1}. \tag{2.52}$$

The precise assumptions used to obtain such estimates differ according to the type of finite element approximation scheme being used. **It should be stressed that superconvergence will only occur in very special circumstances.** A survey of superconvergence results is contained in [42].

LEMMA 2.5. Suppose $u \in H^{p+2}(\Omega)$, G_X satisfies (R1)–(R3) and (SC) holds. Then

$$\|\nabla u - G_X(u_X)\|_{L_2(\Omega)} \leqslant C(u)h^{p+1} \tag{2.53}$$

holds where $C > 0$ is independent of h and u.

PROOF. By the Triangle Inequality and the linearity of G_X

$$\|\nabla u - G_X(u_X)\|_{L_2(K)} \leqslant |\nabla u - G_X(\Pi_p u)|_{L_2(K)} + |G_X(\Pi_p u - u_X)|_{L_2(K)}. \tag{2.54}$$

The boundedness property (R3) of G_X implies

$$\|G_X(\Pi_p u - u_X)\|_{L_2(K)} \leqslant C\,|\Pi_p u - u_X|_{H^1(\widetilde{K})} \tag{2.55}$$

where the Inverse Property [25, p. 142] has been applied separately on each of the elements in $\widetilde{K}$. Summing over the elements and using Lemma 2.3 and (SC) gives the result as claimed. □

2.4.1. A posteriori error estimators

Consider the class of error estimators obtained by using $G_X(u_X)$ instead of ∇u in the expression for

the error. That is, the estimator on element K is η_K where

$$\eta_K = \|G_X(u_X) - \nabla u_X\|_{L_2(\Omega)}. \tag{2.56}$$

The global estimator is obtained by summing contributions from the elements.

THEOREM 2.6. Let η be the a posteriori error estimator defined above. Assume that (SC), (R1)–(R3) and (2.50) hold. Then

$$\lim_{h\to 0} \frac{\eta}{|||e|||} = 1. \tag{2.57}$$

PROOF. By the Triangle Inequality, (2.29) and the foregoing results

$$\begin{aligned} |\eta - |||e||| | &\leqslant \|G_X(u_X) - \nabla u_X - \nabla e\|_{L_2(\Omega)} + C\,\|e\|_{L_2(\Omega)} \\ &= \|G_X(u_X) - \nabla u\|_{L_2(\Omega)} + Ch^{\mu}|||e||| \\ &\leqslant C(u)h^{p+1} + Ch^{\mu}|||e||| \end{aligned} \tag{2.58}$$

since $\|e\|_{L_2(\Omega)} \leqslant Ch^{\mu}|||e|||$ for some positive constant μ depending on the smoothness of the solution and the domain Ω. The result follows from (2.50). □

Theorem 2.6 reduces the problem of finding a posteriori error estimators to using the existing superconvergence results to define an appropriate recovery operator G_X. Consequently, whenever we have superconvergence results for a particular finite element scheme, it is then possible to define an a posteriori error estimator that is asymptotically exact.

2.5. Examples of recovery based estimators

The theory will be illustrated by deriving error estimators for some particular types of finite element approximation scheme. This will show how an existing estimator fits into the framework; how another popular estimator can be viewed as a simplified recovery based estimator; and, how new estimators can be easily derived.

2.5.1. The Babuska and Rheinboldt estimator

Consider piecewise linear approximation in one dimension. There are many types of a posteriori error estimator available for this case. The purpose here is not to obtain new results, but to show how an existing estimator fits within the framework.

A superconvergence result is known for this situation (cf. [63]): if $u \in H^3(\Omega)$ and the mesh is quasi-uniform then

$$|u_X - \Pi_p u|_{H^1(\Omega)} \leqslant Ch^2\,|u|_{H^3(\Omega)} \tag{2.59}$$

where Π_p interpolates at the endpoints of the elements. The recovery operator is constructed using the process shown in Fig. 1. The operator is linear and based on values of the direct approximation to the gradient sampled on the set $\widetilde{K}$ as shown in Fig. 1. It is easily verified that for any piecewise quadratic function v one has $G_X(\Pi_p v) \equiv v'$. Moreover,

$$|G_X(v)|_{L_\infty(K)} \leqslant 3\,|v|_{W^{1,\infty}(\widetilde{K})}. \tag{2.60}$$

Therefore, (SC) and (R1)–(R3) are valid. The estimator on element K is

$$\eta_K = \|G_X(u_X) - u_X'\|_{L_2(K)}. \tag{2.61}$$

The estimator is precisely the estimator originally proposed and analyzed by Babuska and Rheinboldt [12](Definition 6.3). Previously, the estimator was obtained by an argument based on locally projecting the error onto a quadratic function that vanishes at the nodes. For further details see [12] where numerical examples illustrating the effectiveness of this estimator will also be found.

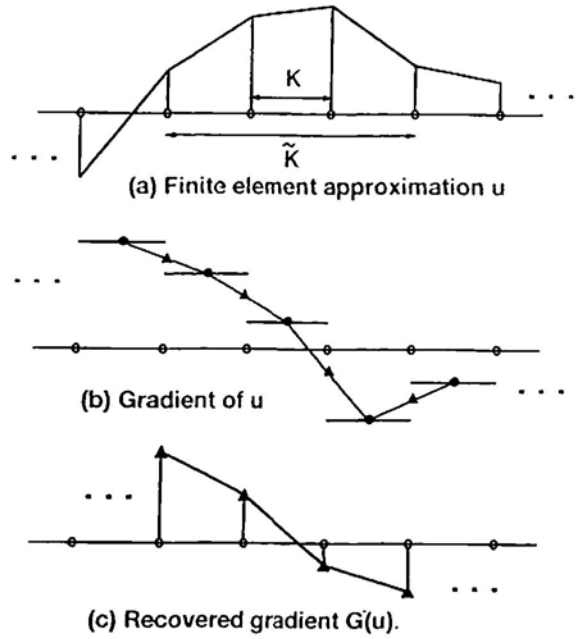


Fig. 1. Construction of recovery operator $G_X = G$ from piecewise linear approximation in one dimension.

2.5.2. An estimator for quadratic approximation

Consider approximation using piecewise quadratic functions in one dimension. The superconvergence property holds in this situation (see [44]). A recovery operator G_X may be defined by exploiting the result: if the true solution is cubic then the true gradient u' and the gradient of the quadratic interpolant $\Pi_p u$, coincide at the nodes used in the 2-point Gauss–Legendre quadrature rule on the element. Therefore, on the element $[x_i, x_{i+1}]$ we sample the gradient at the points

$$x_i^{\pm} = \frac{1}{2}\left\{x_i + x_{i+1} \pm \frac{1}{\sqrt{3}}(x_{i+1} - x_i)\right\}. \tag{2.62}$$

The next step is to define the recovery operator G_X. This is done by first recovering the gradient at the nodes and the centroid of each element. There are many possible ways to carry out this process (many of which fall within the framework). We shall use a *cubic* interpolation process to interpolate the gradient sampled at the Gauss–Legendre points. A procedure based on quadratic interpolation would meet the recovery criteria (R1)–(R3) but would give an unsymmetrical scheme.

The recovery operator G_X is as follows:

- the value of $G_X[v]$ at the centroid of element $[x_i, x_{i+1}]$ is taken to be the value of the of the cubic polynomial interpolating to v' at the points
$$\{x_{i-1}^+, x_i^-, x_i^+, x_{i+1}^-\} \tag{2.63}$$
- the value $G_X[v]$ at a node x_i is the value of the cubic interpolating to v' at the points
$$\{x_{i-1}^-, x_{i-1}^+, x_i^-, x_i^+\} \tag{2.64}$$
- the function $G_X[v]$ is taken to be the X-interpolant of the recovered values at the centroids and nodes.

It is easily verified using elementary manipulations that conditions (R1)–(R3) are satisfied. Theorem 2.6 then shows the estimator is asymptotically exact. One could obtain an explicit expression for the estimator in terms of the values of the finite element approximation at the Gauss–Legendre points. However, this is unnecessary since the recovery process combined with a quadrature rule provides a simple method of implementation. Examples showing the performance of the estimator will be found in [4].

2.5.3. The Kelly, Gago, Zienkiewicz and Babuska estimator

Consider the finite element approximation of Poisson's equations using piecewise bilinear approximation in two dimensions. For the sake of simplicity assume each of the elements K is a square with sides of length h parallel to the axes.

Results from Zlamal [63,64] show that the superconvergence property (SC) holds for this situation

$$|u_X - \Pi_p u|_{H^1(\Omega)} \leqslant Ch^2 |u|_{H^3(\Omega)} \tag{2.65}$$

where Π_p is the bilinear interpolant at the vertices of the mesh. The recovery operator G_X is piecewise bilinear in each component. The values at the vertex x are obtained by a simple averaging of the gradient sampled at the centroids of the elements having a vertex at x (see Fig. 2). If (x, y) is a boundary vertex, then $G_X[v](x, y)$ is the value at (x, y) of the bilinear function which interpolates to ∇v at the centroids of the elements that are nearest to the point (x, y). The operator G_X is linear and bounded since

$$|G_X[v]|_{L_\infty(K)} \leqslant 4 |v|_{W^{1,\infty}(\tilde{K})} \quad \forall v \in X \tag{2.66}$$

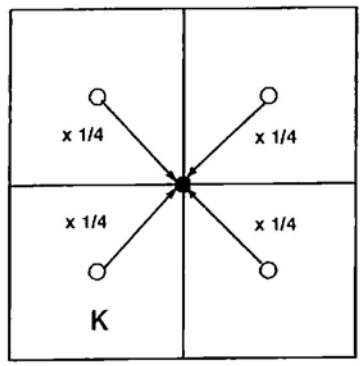
and a straightforward manipulation reveals

$$G_X[\Pi_p v] \equiv \nabla v \tag{2.67}$$

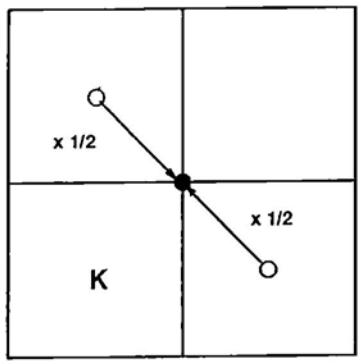
whenever v is piecewise biquadratic. Consequently, the recovery operator satisfies conditions (R1)–(R3) and (SC). The estimator η_K on element K is

$$\eta_K = \|G_X(u_X) - \nabla u_X\|_{L_2(K)}. \tag{2.68}$$

This estimator will be asymptotically exact thanks to Theorem 2.6.



(a) Scheme for new estimator.



(b) Scheme for Kelly et al. estimator.

Fig. 2. Construction of recovered gradient at vertex of element K. The value at • is a linear combination of the values at ○ using the weights indicated.

It is interesting to compare the new estimator with the estimator $\hat{\eta}_K$ proposed by Kelly et al. [40]:

$$\hat{\eta}_K^2 = \frac{h}{24} \int_{\partial K} \left[\frac{\partial u_X}{\partial n} \right]^2 \mathrm{d}s, \tag{2.69}$$

and where

$$\left[\frac{\partial u_X}{\partial n} \right] = \frac{\partial u_X}{\partial n_K}\Big|_K + \frac{\partial u_X}{\partial n_J}\Big|_J \tag{2.70}$$

is the discontinuity in the finite element approximation to the gradient across the edge between neighbouring elements K and J. Using the midpoint rule for integration along each side of the element, the estimator (2.69) may be rewritten as

$$\hat{\eta}_K^2 = \frac{h^2}{24} \sum_{\gamma \subset \partial K} \left[\frac{\partial u_X}{\partial n} \right]^2 \tag{2.71}$$

where the discontinuities are evaluated at the midpoints of the sides. Zienkiewicz et al. [59] state that *the derivation of (2.71) is complex and subject to many heuristic arguments.*

Kelly et al. [40] note that the estimator bears out practical experience that the accuracy of the approximation is related to the discontinuity of the finite element approximation to the gradient on the interelement boundaries. The recovery based estimator (2.68), like (2.71), is found after a lengthy but otherwise straightforward manipulation, to depend on the discontinuities in the finite element approximation to the gradient. The dependence is more intricate than in (2.71) involving, in addition, differences in tangential components at the centroids. The estimator (2.69) uses gradients sampled from the element K and elements sharing a common edge, while the estimator (2.71) also involves elements sharing a common node as shown in Fig. 2.

The estimator (2.68) may be simplified by avoiding terms arising from elements sharing only a common node. This may be achieved by taking the values of the recovered gradient at the vertices to be (see Fig. 2)

$$G_X[v](x, y) = \frac{1}{2} \left[\nabla v|_{x-1/2h, y+1/2h} + \nabla v|_{x+1/2h, y-1/2h} \right]. \tag{2.72}$$

The new recovery operator satisfies (R1)–(R3) and gives rise to an estimator *identical* to $\hat{\eta}_K$. The estimator derived by Kelly et al. therefore may be viewed using the above framework. This approach makes the derivation of (2.71) straightforward.

It might seem that the estimator η_K is too complicated to be of practical use. However, it is in many ways much simpler than the Kelly estimator. For instance, with the Kelly estimator it is not obvious how one should define the value of the jump along the exterior boundary $\partial\Omega$. This difficulty does not arise with the recovery based estimator. The recovery based approach even provides the answer: the value of should be taken to be the jump on the opposite side of the element.

2.5.4. *Zienkiewicz–Zhu patch recovery technique*

An alternative type of recovery estimator was introduced by Zienkiewicz and Zhu [61,62]. Let Ψ denote the set of vertices in the finite element partition. The recovery operator G_X is defined by first identifying the patch $\tilde{\Omega}_m$ of elements having a vertex at $x_m \in \Psi$. That is

$$\tilde{\Omega}_m = \{ K \in \mathcal{P} : K \subset \text{ supp } \theta_m \} \tag{2.73}$$

where θ_m is the pyramid function associated with the node x_m. An intermediate recovered gradient $G_{X,m}$ is then constructed for each patch and the final recovered gradient is obtained by averaging

$$G_X[u_X](x) = \frac{1}{|\Psi|} \sum_{m \in \Psi} G_{X,m}[u_X](x). \tag{2.74}$$

The intermediate functions $G_{X,m}$ are constructed using values of the gradient of the finite element approximation sampled on the patch $\tilde{\Omega}_m$. Let $Z(m)$ denote the set of points at which the gradient is

to be sampled. For instance, working with quadrilateral elements one would use the Gauss–Legendre quadrature points (see Fig. 3). For triangular elements, the set $Z(m)$ consists of the points shown in Fig. 3. Further examples will be found in [61,62].

The function $G_{X,m}$ is obtained by calculating a least-squares fit to the gradient sampled at the points $Z(m)$. The function $G_{X,m}$ is assumed to be of the form $X \times X$ where X is the finite element subspace. That is

$$G_{X,m}(x) = \sum_n \alpha_n \phi_n(x) \tag{2.75}$$

where $\phi_n(x)$ form a basis for the finite element subspace X and α_n are constant vectors chosen to minimize the expression

$$\sum_{z \in Z} \{G_{X,m}(z) - \nabla u_X(z)\}^2. \tag{2.76}$$

Of course, in the summation (2.75), only degrees of freedom associated with elements in the patch $\tilde{\Omega}_m$ need be considered. This means that the actual value of the final recovered gradient G_X on an element K will depend only on values sampled from the patch $\tilde{K}$ of elements neighbouring K. Therefore, condition (R2) will be satisfied. Equally well, the recovery operator is linear and bounded so that condition (R3) is also valid. The superconvergence condition (SC) and condition (R1) can be satisfied by selecting the sampling points $Z(m)$ to consist of the points at which superconvergence occurs. If this is the case, then the estimator will be asymptotically exact according to Theorem 2.6. This is confirmed in numerical examples [61,62]. However, it is often the case in practical computations that the superconvergence property is not satisfied, or, the sampling points may be chosen differently. The estimator might be expected to degrade in such circumstances. In fact, it is found that the estimator is astonishingly robust, continuing to perform satisfactorily even in quite extreme situations [15,16]. While Theorem 2.2 suggests that the estimators might tend to bound the error asymptotically, the reason behind the robustness of the estimators remains an open question.

2.6. Summary

The error estimation techniques often used in the engineering community and sometimes referred to as *averaging based error estimators* have been considered. The development can be thought of as consisting of two main parts. Firstly, the derivation of Theorem 2.2. As a special case, this shows that (for h sufficiently small) the averaging based estimators should tend to overestimate the true error.

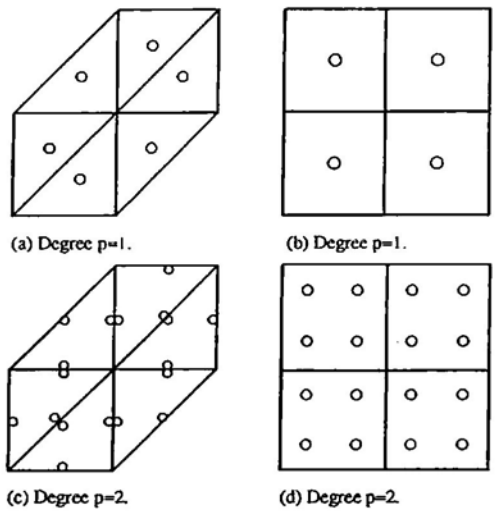


Fig. 3. Sampling set $Z(m)$ for patch recovery technique.

It is only if one wishes to obtain two-sided estimates or asymptotic exactness that it then becomes necessary to make the additional and rather stringent assumption concerning the superconvergence property. Under this condition, it was shown that a class of estimators based on the use of recovery operators G_X will be asymptotically exact. The recovery operators are closely related to the averaging based schemes and in some cases are identical. In this respect, the theory developed for the recovery operators G_X and the associated error estimators can be regarded as providing some indication of the behaviour of the averaging based schemes and how they might be modified to enhance their performance.

3. Explicit a posteriori estimators

3.1. Introduction

Consider the model problem in Section 1.5. Suppose that the finite element approximation u_X has been computed. The basic issue in a posteriori error estimation is embodied in the question: *how can the discretization error e be estimated?* In order to provide an answer one may make use of

- the Galerkin approximation u_X itself
- the data f and g
- equation characterizing the true error:

$$B(e,v) = B(u,v) - B(u_X,v) = L(v) - B(u_X,v) \quad \forall v \in V \tag{3.1}$$

- the Galerkin orthogonality property:

$$B(e,v_X) = 0 \quad \forall v_X \in X. \tag{3.2}$$

The following section illustrates how these may be used to derive a simple a posteriori error estimate.

3.2. A simple a posteriori error estimate

The first step is to decompose Eq. (3.1) for the error into local contributions from each element:

$$\begin{aligned} B(e,v) &= L(v) - B(u_X,v) \\ &= \sum_{K\in\mathcal{P}} \left\{ \int_K fv \,\mathrm{d}x + \int_{\partial K\cap\Gamma_N} gv \,\mathrm{d}x - \int_K (\nabla u_X \cdot \nabla v + cu_X v) \,\mathrm{d}x \right\} \end{aligned} \tag{3.3}$$

for all $v \in V$. Integrating by parts over each element gives

$$B(e,v) = \sum_{K\in\mathcal{P}} \left\{ \int_K rv \,\mathrm{d}x + \int_{\partial K\cap\Gamma_N} Rv \,\mathrm{d}s - \int_{\partial K\setminus\Gamma_N} \frac{\partial u_X}{\partial n_K} v \,\mathrm{d}s \right\} \tag{3.4}$$

where r is the *interior residual*

$$r = f + \Delta u_X - cu_X \quad \text{in } K \tag{3.5}$$

and R is the *boundary residual*

$$R = g - \frac{\partial u_X}{\partial n_K} \quad \text{on } \partial K \cap \Gamma_N \tag{3.6}$$

where n_K is the unit outward normal vector to ∂K. Each of these quantities is well defined thanks to the smoothness of the data and the regularity of the approximation u_X on each element. The contribution from the final term in (3.4) can be rewritten by observing that the (trace of the) function v is continuous along an edge shared by two elements giving

$$B(e,v) = \sum_{K\in\mathcal{P}} \left\{ \int_K rv \,\mathrm{d}x + \int_{\partial K\cap\Gamma_N} Rv \,\mathrm{d}s \right\} - \sum_{\gamma\in\partial\mathcal{P}\setminus\partial\Omega} \int_\gamma \left[\frac{\partial u_X}{\partial n} \right] v \,\mathrm{d}s \tag{3.7}$$

where the final summation is over all the inter-element edges γ on the interior of the mesh. The quantity

$$\left[\frac{\partial u_X}{\partial n}\right] = n_K \cdot (\nabla u_X)_K + n_J \cdot (\nabla u_X)_J \tag{3.8}$$

defined on the edge γ separating elements K and J represents the jump discontinuity in the approximation to the flux. The identity (3.8) can be written more compactly by extending the definition of the boundary residual to incorporate the jump discontinuity in the flux. Therefore, on interior edges the definition (3.6) is augmented by

$$R = -\frac{1}{2}\left[\frac{\partial u_X}{\partial n}\right] \tag{3.9}$$

so that (3.8) then becomes

$$B(e, v) = \sum_{K\in\mathcal{P}} \int_K rv \,\mathrm{d}x + \sum_{\gamma\in\partial\mathcal{P}} \int_\gamma Rv \,\mathrm{d}s \quad \forall v \in V \tag{3.10}$$

where the final summation is over all the edges in the partition $\mathcal{P}$.

The orthogonality property (3.2) may now be used as follows. For given $v \in V$, let $\Pi_X v$ be the interpolant to v from the subspace X as in Theorem 1.1. Thanks to (3.2) and the identity (3.10) there holds

$$0 = \sum_{K\in\mathcal{P}} \int_K r\Pi_X v \,\mathrm{d}x + \sum_{\gamma\in\partial\mathcal{P}} \int_\gamma R\Pi_X v \,\mathrm{d}s. \tag{3.11}$$

Combining this with identity (3.10) gives

$$B(e, v) = \sum_{K\in\mathcal{P}} \int_K r(v - \Pi_X v) \,\mathrm{d}x + \sum_{\gamma\in\partial\mathcal{P}} \int_\gamma R(v - \Pi_X v) \,\mathrm{d}s \quad \forall v \in V. \tag{3.12}$$

The identity (3.12) plays an important role, indirectly or directly, throughout a posteriori error analysis. It may be used to derive the a posteriori error estimate as follows. Applying the Cauchy–Schwarz Inequality gives

$$B(e, v) \leqslant \sum_{K\in\mathcal{P}} \|r\|_{L_2(K)} \|v - \Pi_X v\|_{L_2(K)} + \sum_{\gamma\in\partial\mathcal{P}} \|R\|_{L_2(\gamma)} \|v - \Pi_X v\|_{L_2(\gamma)}. \tag{3.13}$$

Let $\tilde{K}$ denote the subdomain consisting of elements sharing a common edge with element K

$$\tilde{K} = \operatorname{int}\left\{\bigcup J \in \mathcal{P} : \bar{J} \cap \bar{K} \neq \emptyset\right\} \tag{3.14}$$

It may then be shown [26] that there exist a constant C which is independent of v and h_K such that

$$\|v - \Pi_X v\|_{L_2(K)} \leqslant C h_K |v|_{H^1(\tilde{K})} \tag{3.15}$$

and

$$\|v - \Pi_X v\|_{L_2(\partial K)} \leqslant C h_K^{1/2} |v|_{H^1(\tilde{K})} \tag{3.16}$$

where h_K is the diameter of the element K. Inserting these estimates in inequality (3.13) and applying the Cauchy–Schwarz Inequality leads to

$$B(e, v) \leqslant C |v|_{H^1(\Omega)} \left\{\sum_{K\in\mathcal{P}} h_K^2 \|r\|^2_{L_2(K)} + \sum_{\gamma\in\partial\mathcal{P}} h_K \|R\|^2_{L_2(\gamma)}\right\}^{1/2}. \tag{3.17}$$

Finally, noting that $|v|_{H^1(\Omega)} \leqslant |||v|||$ and substituting e in place of v gives the a posteriori error estimate

$$|||e|||^2 \leqslant C \left\{\sum_{K\in\mathcal{P}} h_K^2 \|r\|^2_{L_2(K)} + \sum_{\gamma\in\partial\mathcal{P}} h_K \|R\|^2_{L_2(\gamma)}\right\} \tag{3.18}$$

where $|||\cdot|||$ denotes the energy norm for the model problem $|||v|||^2 = B(v, v)$. Apart from the constant C, all of the quantities on the right-hand can be computed explicitly from the data and the finite element

approximation. Typically, the terms on the right-hand side are regrouped as follows:

$$|||e|||^2 \leqslant C \sum_{K \in \mathcal{P}} \left\{ h_K^2 \, \|r\|_{L_2(K)}^2 + \frac{1}{2} h_K \, \|R\|_{L_2(\partial K)}^2 \right\}. \tag{3.19}$$

The purpose in doing so is that defining the local error indicator by η_K on element K by

$$\eta_K^2 = h_K^2 \, \|r\|_{L_2(K)}^2 + \frac{1}{2} h_K \, \|R\|_{L_2(\partial K)}^2 \tag{3.20}$$

allows one to identify contributions from each of the elements. It is then assumed that each of these quantities is a measure of the *local discretization error* over each element. In this way one can use η_K as a basis for guiding local mesh refinements.

Error estimators of this form were originally derived by Babuska and Rheinboldt [14] in one dimension; Babuska and Miller [10] for bilinear approximation in two dimensions; and Kelly et al. [40]. Estimators that may be computed explicitly from the solution and the data are often referred to as *explicit estimators*.

3.3. A simple error estimator in the L_2 norm

The estimator derived above gives information about the error measured in the energy (or any equivalent) norm. The duality argument due to Aubin and Nitsche [25] plays an important role in the derivation of a priori error estimates in norms other than the energy like norms. It may also be used in the context of a posteriori error estimators.

To apply the technique consider the adjoint of the original model problem:

$$\Phi_F \in V : B(v, \Phi_F) = (F, v) \quad \forall v \in V \tag{3.21}$$

where $F \in L_2(\Omega)$ is given data. It is assumed that this problem is *regular* in the sense that the solution Φ_F has the extra regularity $\Phi_F \in H^2(\Omega) \cap V$ and the solution operator from $L_2(\Omega)$ to $H^2(\Omega)$ is continuous

$$\|\Phi_F\|_{H^2(\Omega)} \leqslant C \, \|F\|_{L_2(\Omega)}. \tag{3.22}$$

The specific choice of data F equal to the error function e then gives

$$\|e\|_{L_2(\Omega)}^2 = B(e, \Phi_e). \tag{3.23}$$

The residual equation (3.12) plays the same role as in the derivation of the energy norm estimator

$$B(e, \Phi_e) = \sum_{K \in \mathcal{P}} \int_K r(\Phi_e - \Pi_X \Phi_e) \, dx + \sum_{\gamma \in \partial \mathcal{P}} \int_\gamma R(\Phi_e - \Pi_X \Phi_e) \, ds \tag{3.24}$$

and, as before, the Cauchy–Schwarz Inequality gives

$$B(e, v) \leqslant \sum_{K \in \mathcal{P}} \|r\|_{L_2(K)} \|v - \Pi_X v\|_{L_2(K)} + \sum_{\gamma \in \partial \mathcal{P}} \|R\|_{L_2(\gamma)} \|v - \Pi_X v\|_{L_2(\gamma)}. \tag{3.25}$$

Slightly different approximation theoretic results are required: there exists a constant C which is independent of v and h_K such that [26]

$$\|v - \Pi_X v\|_{L_2(K)} \leqslant C h_K^2 \, |v|_{H^2(\tilde{K})} \tag{3.26}$$

and

$$\|v - \Pi_X v\|_{L_2(\partial K)} \leqslant C h_K^{3/2} \, |v|_{H^2(\tilde{K})}. \tag{3.27}$$

Substituting these estimates and applying the Cauchy–Schwarz Inequality leads to

$$\|e\|_{L_2(\Omega)}^2 \leqslant C \, |\Phi_e|_{H^2(\Omega)} \left\{ \sum_{K \in \mathcal{P}} h_K^4 \, \|r\|_{L_2(K)}^2 + \sum_{\gamma \in \partial \mathcal{P}} h_K^3 \, \|R\|_{L_2(\gamma)}^2 \right\}^{1/2} \tag{3.28}$$

and so, with the aid of the regularity assumption (after bounding $|\Phi_e|_{H^2(\Omega)}$ in terms of $\|e\|_{L_2(\Omega)}$) there

follows

$$\|e\|^2_{L_2(\Omega)} \leqslant C \left\{ \sum_{K\in\mathcal{P}} h_K^4 \|r\|^2_{L_2(K)} + \sum_{\gamma\in\partial\mathcal{P}} h_K^3 \|R\|^2_{L_2(\gamma)} \right\}. \tag{3.29}$$

A rearrangement of the terms on the right-hand side gives an estimator similar to the one derived before, apart from a higher-order scaling

$$\|e\|^2_{L_2(\Omega)} \leqslant C \sum_{K\in\mathcal{P}} \left\{ h_K^4 \|r\|^2_{L_2(K)} + \frac{1}{2} h_K^3 \|R\|^2_{L_2(\partial K)} \right\}. \tag{3.30}$$

THEOREM 3.1. Suppose that the domain Ω is convex and that $\Gamma_D = \partial\Omega$. Let $\eta_{L_2(K)}$ denote the local error indicator

$$\eta^2_{L_2(K)} = h_K^4 \|r\|^2_{L_2(K)} + \frac{1}{2} h_K^3 \|R\|^2_{L_2(\partial K)} \tag{3.31}$$

where r and R are the interior and boundary residuals. Then there exists a constant C depending only on the shape regularity of the elements such that

$$\|e\|^2_{L_2(\Omega)} \leqslant C \sum_{K\in\mathcal{P}} \eta^2_{L_2(K)}. \tag{3.32}$$

PROOF. It may be shown that the adjoint problem satisfies the regularity assumptions when the domain Ω is convex. The proof then follows immediately from the above steps. □

3.4. Equivalence of estimator

The estimator implied by Eq. (3.19) provides an upper bound on the discretization error up to an unknown constant. If the estimator is to be used as the basis of an adaptive refinement algorithm, then it is desirable that it be an *equivalent* measure of the error. That is to say, there should be a constant C which does not depend on the solution u, the data f and g or the mesh parameter h, such that

$$\sum_{K\in\mathcal{P}} \eta_K^2 \leqslant C |||e|||^2. \tag{3.33}$$

Should such an estimate fail to be valid, one cannot expect that the resulting adaptive procedure will be effective.

The problem of obtaining two sided bounds for explicit estimators such as the one derived in Section 3.2 was addressed by Verfürth [57] who devised the following technique.

3.4.1. Bubble functions

Let $\widehat{K}$ denote a reference element. The *interior bubble function* $\widehat{\psi} : \widehat{K} \mapsto \mathbb{R}$ is the lowest-order polynomial which vanishes on the boundary $\partial\widehat{K}$ and is non-zero on the interior of $\widehat{K}$. An *edge bubble function* is the lowest order polynomial which vanishes on all but one edge and is non-zero on the interior.

Triangular elements

In the case of triangular elements

$$\widehat{K} = \{(\hat{x}, \hat{y}) : 0 \leqslant \hat{x} \leqslant 1; \quad 0 \leqslant \hat{y} \leqslant 1 - \hat{x}\}. \tag{3.34}$$

The functions $\widehat{\lambda}_1$, $\widehat{\lambda}_2$ and $\widehat{\lambda}_3$ denote the barycentric (area) coordinates on $\widehat{K}$. The interior bubble function $\widehat{\psi}$ is defined by

$$\widehat{\psi} = 27\widehat{\lambda}_1\widehat{\lambda}_2\widehat{\lambda}_3 \tag{3.35}$$

and the first edge bubble function is

$$\widehat{\chi} = 4\widehat{\lambda}_2\widehat{\lambda}_3. \tag{3.36}$$

Quadrilateral elements

In the case of quadrilateral elements

$$\widehat{K} = \{(\hat{x}, \hat{y}) : -1 \leqslant \hat{x} \leqslant 1; \quad -1 \leqslant \hat{y} \leqslant 1\}. \tag{3.37}$$

The interior bubble function $\widehat{\psi}$ is defined by

$$\widehat{\psi} = (1 - \hat{x}^2)(1 - \hat{y}^2) \tag{3.38}$$

and the first edge bubble functions is

$$\widehat{\chi} = \frac{1}{2}(1 - \hat{x}^2)(1 - \hat{y}). \tag{3.39}$$

LEMMA 3.2. Let $\widehat{P} \subset H^1(\widehat{K})$ be a finite dimensional space of functions defined on the reference element. Then there exists a constant $\widehat{C}$ such that for all $\hat{v} \in \widehat{P}$

$$\widehat{C}^{-1} \|\hat{v}\|^2_{L_2(\widehat{K})} \leqslant \int_{\widehat{K}} \widehat{\psi}\hat{v}^2 \, \mathrm{d}\hat{\mathbf{x}} \leqslant \widehat{C} \|\hat{v}\|^2_{L_2(\widehat{K})} \tag{3.40}$$

and

$$\widehat{C}^{-1} \|\hat{v}\|_{L_2(\widehat{K})} \leqslant \|\widehat{\psi}\hat{v}\|_{H^1(\widehat{K})} \leqslant \widehat{C} \|\hat{v}\|_{L_2(\widehat{K})} \tag{3.41}$$

where the constant $\widehat{C}$ is independent of $\hat{v}$.

PROOF. It is easily seen that the mappings

$$\hat{v} \mapsto \left\{ \int_{\widehat{K}} \widehat{\psi}\hat{v}^2 \, \mathrm{d}\hat{\mathbf{x}} \right\}^{1/2}$$

and

$$\hat{v} \mapsto \|\widehat{\psi}\hat{v}\|_{H^1(\widehat{K})}$$

both define norms on the finite dimensional space $\widehat{P}$. The results then follow immediately from the fact that all norms on a finite dimensional space are equivalent. □

Let $K \in \mathcal{P}$ be any element and $F_K : \widehat{K} \mapsto K$ be an invertible mapping. The associated bubble functions on element K are then defined by

$$\psi_K = \widehat{\psi} \circ F_K^{-1}; \qquad \chi_\gamma = \widehat{\chi} \circ F_K^{-1}. \tag{3.42}$$

It is assumed that the partition $\mathcal{P}$ is non-degenerate. Therefore, there exists a positive number h_K and constants C_1, C_2 and C_3 such that the following properties hold for each of the local mappings F_K:

$$\|J_K\| \leqslant C_1 h_K; \qquad \|J_K^{-1}\| \leqslant C_2 h_K^{-1}; \qquad C_3^{-1} h_K^2 \leqslant |\det J_K| \leqslant C_3 h_K^2 \tag{3.43}$$

where J_K is the Jacobian of the transformation F_K. The set P is defined by

$$P = \{\hat{v} \circ F_K^{-1} : \hat{v} \in \widehat{P}\} \tag{3.44}$$

where $\widehat{P}$ is a finite dimensional subspace consisting functions of defined on $\widehat{K}$.

THEOREM 3.3. There exists a constant C such that for all $v \in P$

$$C^{-1} \|v\|^2_{L_2(K)} \leqslant \int_K \psi v^2 \, \mathrm{d}\mathbf{x} \leqslant C \|v\|^2_{L_2(K)} \tag{3.45}$$

and

$$C^{-1} \|v\|_{L_2(K)} \leqslant \|\psi v\|_{L_2(K)} + h_K |\psi v|_{H^1(K)} \leqslant C \|v\|_{L_2(K)} \tag{3.46}$$

where the constant C is independent of v and h_K.

PROOF. The result follows by mapping to the reference domain $\widehat{K}$, applying the previous lemma and a standard scaling argument. □

LEMMA 3.4. Let $\widehat{\gamma} \subset \partial\widehat{K}$ be an edge and $\widehat{\chi}$ the corresponding edge bubble function. Let $\widehat{P}$ be a finite dimensional space of functions defined on $\widehat{\gamma}$. Then there exists a constant $\widehat{C}$ such that

$$\widehat{C}^{-1} \|\widehat{v}\|^2_{L_2(\widehat{\gamma})} \leqslant \int_{\widehat{\gamma}} \widehat{\chi}\widehat{v}^2 \,\mathrm{d}\widehat{s} \leqslant \widehat{C} \|\widehat{v}\|^2_{L_2(\widehat{\gamma})}. \tag{3.47}$$

Suppose that $\widehat{v} \in \widehat{P}$ is extended to $\widehat{K}$ according to the rule

$$\widehat{\mathcal{E}}\widehat{v}(\widehat{x},\widehat{y}) = \widehat{\chi}(\widehat{x},\widehat{y})\widehat{v}(\widehat{x}) \tag{3.48}$$

there exists a constant $\widehat{C}$ such that

$$\|\widehat{\chi}\widehat{\mathcal{E}}\widehat{v}\|_{H^1(\widehat{K})} \leqslant \widehat{C} \|\widehat{v}\|_{L_2(\widehat{\gamma})} \tag{3.49}$$

In each case the constants $\widehat{C}$ are independent of $\widehat{v}$.

PROOF. The proof is based on the equivalence of norms on a finite dimensional space similar to before. □

A scaling argument can be used to translate these results to a general element K.

THEOREM 3.5. Let $\gamma \subset \partial K$ be an edge and χ_γ the corresponding edge bubble function. Let P be the finite dimensional space of functions defined on γ obtained by mapping $\widehat{P} \subset H^1(\widehat{K})$. Then there exists a constant C such that

$$C^{-1} \|v\|^2_{L_2(\gamma)} \leqslant \int_\gamma \chi_\gamma v^2 \,\mathrm{d}s \leqslant C \|v\|^2_{L_2(\gamma)}. \tag{3.50}$$

Moreover, there exists an extension of v to K (again denoted by v) such that

$$h_K^{-1/2} \|\chi_\gamma v\|_{L_2(K)} + h_K^{1/2} |\chi_\gamma v|_{H^1(K)} \leqslant C \|v\|_{L_2(\gamma)} \tag{3.51}$$

In each case the constants C are independent of v.

3.4.2. Bounds on the residuals

The proof of equivalence makes use of the residual equation (3.10) in conjunction with special choices of the function v.

The first task is to bound the term $\|r\|_{L_2(K)}$. Let $\bar{r}_K$ be a polynomial approximation to the interior residual r on element K. For instance, $\bar{r}_K$ might be the $L_2(K)$ projection onto piecewise constants. Applying Theorem 3.3 gives

$$\|\bar{r}_K\|^2_{L_2(K)} \leqslant C \int_K \psi_K \bar{r}_K^2 \,\mathrm{d}\mathbf{x}. \tag{3.52}$$

The function $v = \bar{r}_K \psi_K$ vanishes on the boundary of element K. It may therefore be extended to the whole of the domain Ω giving a function v belonging to the space V. The residual equation (3.10) then implies

$$B(e, \bar{r}_K \psi_K) = \int_K \psi_K r \bar{r}_K \,\mathrm{d}\mathbf{x} \tag{3.53}$$

where the first term contains contributions from element K only. Therefore

$$\int_K \psi_K \bar{r}_K^2 \,\mathrm{d}\mathbf{x} = \int_K \psi_K \bar{r}_K (\bar{r}_K - r) \,\mathrm{d}\mathbf{x} + B(e, \bar{r}_K \psi_K) \tag{3.54}$$

The first term on the right-hand side may be bounded as follows: by Cauchy–Schwarz

$$\int_K \psi_K \bar{r}_K (\bar{r}_K - r)\,dx \leqslant \|\psi_K \bar{r}_K\|_{L_2(K)} \|\bar{r}_K - r\|_{L_2(K)} \tag{3.55}$$

and then by Theorem 3.3 (Part 2)

$$\|\psi_K \bar{r}_K\|_{L_2(K)} \leqslant C \|\bar{r}_K\|_{L_2(K)}. \tag{3.56}$$

The second term is dealt with similarly

$$B(e, \bar{r}_K \psi_K) \leqslant |||e|||_K \|\psi_K \bar{r}_K\|_{H^1(K)} \tag{3.57}$$

and by Theorem 3.3 (Part 2)

$$\|\psi_K \bar{r}_K\|_{H^1(K)} \leqslant C h_K^{-1} \|\bar{r}_K\|_{L_2(K)}. \tag{3.58}$$

Therefore,

$$\|\bar{r}_K\|_{L_2(K)} \leqslant C\{h_K^{-1} |||e|||_K + \|\bar{r}_K - r\|_{L_2(K)}\} \tag{3.59}$$

and equally well, by the Triangle Inequality

$$\|r\|_{L_2(K)} \leqslant C\{h_K^{-1} |||e|||_K + \|\bar{r}_K - r\|_{L_2(K)}\}. \tag{3.60}$$

It remains to bound the terms $\|R\|_{L_2(\gamma)}$. The argument proceeds in an analogous fashion. Let $\bar{R}_\gamma$ be a polynomial approximation to the boundary residual or jumps. Theorem 3.5 gives

$$\|\bar{R}_\gamma\|^2_{L_2(\gamma)} \leqslant C \int_\gamma \chi_\gamma \bar{R}_\gamma^2\,ds. \tag{3.61}$$

Let $\tilde{\gamma}$ denote the subdomain of Ω consisting of the union of the side γ and the pair of elements (K and J say) sharing the common side γ. The function $v = \bar{R}_\gamma \chi_\gamma$ vanishes on the boundary of the subdomain $\tilde{\gamma}$ and is continuous. Extending to the whole of the domain Ω gives a function v from the space V. The residual equation (3.10) with this choice of v yields

$$B(e, \bar{R}_\gamma \chi_\gamma) = \int_{\tilde{\gamma}} \chi_\gamma r \bar{R}_\gamma\,dx + \int_\gamma \chi_\gamma R \bar{R}_\gamma\,ds. \tag{3.62}$$

Therefore,

$$\int_\gamma \chi_\gamma \bar{R}_\gamma^2\,ds = \int_\gamma \chi_\gamma \bar{R}_\gamma (\bar{R}_\gamma - R)\,ds + B(e, \chi_\gamma \bar{R}_\gamma) - \int_{\tilde{\gamma}} \chi_\gamma r \bar{R}_\gamma\,ds. \tag{3.63}$$

Each of these terms may be dealt with using Theorem 3.5 and the Cauchy–Schwarz Inequality as follows:

$$\int_\gamma \chi_\gamma \bar{R}_\gamma (\bar{R}_\gamma - R)\,ds \leqslant \|\chi_\gamma \bar{R}_\gamma\|_{L_2(\gamma)} \|\chi_\gamma (\bar{R}_\gamma - R)\|_{L_2(\gamma)} \leqslant C \|\bar{R}_\gamma\|_{L_2(\gamma)} \|\bar{R}_\gamma - R\|_{L_2(\gamma)} \tag{3.64}$$

$$B(e, \chi_\gamma \bar{R}_\gamma) \leqslant C |||e|||_{\tilde{\gamma}} \|\chi_\gamma \bar{R}_\gamma\|_{L_2(\gamma)} \leqslant C h_K^{-1/2} |||e|||_{\tilde{\gamma}} \|\bar{R}_\gamma\|_{L_2(\gamma)} \tag{3.65}$$

and

$$\int_{\tilde{\gamma}} \chi_\gamma r \bar{R}_\gamma\,ds \leqslant \|r\|_{L_2(\tilde{\gamma})} \|\chi_\gamma \bar{R}_\gamma\|_{L_2(\gamma)} \leqslant C h_K^{1/2} \|r\|_{L_2(\tilde{\gamma})} \|\bar{R}_\gamma\|_{L_2(\gamma)}. \tag{3.66}$$

This may be used to estimate $\|R\|_{L_2(\gamma)}$ after using the Triangle Inequality and (3.60) giving

$$\|R\|_{L_2(\gamma)} \leqslant C\{h_K^{-1/2} |||e|||_{\tilde{\gamma}} + h_K^{1/2} \|\bar{r}_K - r\|_{L_2(K)} + \|\bar{R}_\gamma - R\|_{L_2(\gamma)}\}. \tag{3.67}$$

LEMMA 3.6. Let r and R denote the interior and boundary residuals associated with the finite element approximation constructed from the subspace X. Suppose that $\bar{r}_K$ and $\bar{R}_\gamma$ are finite dimensional approximations to the residuals on an element K and an edge $\gamma \subset \partial K$. The approximations need not be globally continuous. Then there exists a constant C depending only on the shape regularity of the elements such that

$$\|r\|_{L_2(K)} \leqslant C\{h_K^{-1} |||e|||_K + \|\bar{r}_K - r\|_{L_2(K)}\} \tag{3.68}$$

and

$$\|R\|_{L_2(\gamma)} \leqslant C\{h_K^{-1/2}|||e|||_{\tilde{\gamma}} + h_K^{1/2}\|\bar{r}_K - r\|_{L_2(K)} + \|\bar{R}_\gamma - R\|_{L_2(\gamma)}\}. \tag{3.69}$$

PROOF. Follows from the previous arguments. □

3.4.3. Proof of equivalence

Lemma 3.6 is used to analyze the error estimator as follows. Choose the approximation $\bar{r}_K$ to be $L_2(K)$-projection of the residual r onto the degree p polynomials which are restrictions of elements of X to the element K. Similarly, the approximation $\bar{R}_\gamma$ is the $L_2(\gamma)$-projection of the boundary residual R onto degree p polynomials (this time in one variable) which are the restrictions of elements of X to the single edge γ. Lemma 3.6 then applies giving

$$\|r\|_{L_2(K)} \leqslant C\{h_K^{-1}|||e|||_K + \|\bar{r}_K - r\|_{L_2(K)}\} \tag{3.70}$$

and

$$\|R\|_{L_2(\gamma)} \leqslant C\{h_K^{-1/2}|||e|||_{\tilde{\gamma}} + h_K^{1/2}\|\bar{r}_K - r\|_{L_2(K)} + \|\bar{R}_\gamma - R\|_{L_2(\gamma)}\}. \tag{3.71}$$

For interior edges, the boundary residual R is simply the jump in the normal derivative which is itself a polynomial of degree p in one variable. Thus, one has $\bar{R}_\gamma \equiv R$ on interior edges. Conversely, on the exterior boundaries, one has $\bar{R}_\gamma - R \equiv g - \Pi_p g$, where $\Pi_p g$ is the polynomial approximation to g on the edge γ. Noting that the data c for the original differential equation was constant leads to $\bar{r}_K - r \equiv f - \Pi_p f$ on each element K. Therefore,

$$\|r\|_{L_2(K)} \leqslant C\{h_K^{-1}|||e|||_K + \|f - \Pi_p f\|_{L_2(K)}\} \tag{3.72}$$

and

$$\|R\|_{L_2(\gamma)} \leqslant C\{h_K^{-1/2}|||e|||_{\tilde{\gamma}} + h_K^{1/2}\|f - \Pi_p f\|_{L_2(K)} + \|g - \Pi_p g\|_{L_2(\gamma\cap\Gamma_N)}\}. \tag{3.73}$$

The local error indicator η_K in (3.20) associated with the element K can then be bounded by

$$\eta_K^2 \leqslant C\left\{|||e|||_{\tilde{K}}^2 + h_K^2\|f - \Pi_p f\|_{L_2(K)}^2 + \sum_{\gamma\subset\partial K\cap\Gamma_N} h_K\|g - \Pi_p g\|_{L_2(\gamma)}^2\right\} \tag{3.74}$$

where the constant C depends only on the shape regularity of the element. The estimate shows that the error indicator is local in a certain sense, since the terms on the right-hand bound involve only contributions from the actual element and its immediate neighbours. Summarizing the results so far

THEOREM 3.7. Let η_K denote the local error indicator

$$\eta_K^2 = h_K^2\|r\|_{L_2(K)}^2 + \frac{1}{2}h_K\|R\|_{L_2(\partial K)}^2 \tag{3.75}$$

where r and R are the interior and boundary residuals. Then there exists a constant C depending only on the shape regularity of the elements such that

$$\frac{1}{C}|||e|||^2 \leqslant \sum_{K\in\mathcal{P}}\eta_K^2 \leqslant C\left\{|||e|||^2 + \sum_{K\in\mathcal{P}} h_K^2\|f - \Pi_p f\|_{L_2(K)}^2 + \sum_{\gamma\subset\Gamma_N} h_K\|g - \Pi_p g\|_{L_2(\gamma)}^2\right\}. \tag{3.76}$$

Moreover, the local bound

$$\eta_K^2 \leqslant C\left\{|||e|||_{\tilde{K}}^2 + h_K^2\|f - \Pi_p f\|_{L_2(K)}^2 + \sum_{\gamma\subset\partial K\cap\Gamma_N} h_K\|g - \Pi_p g\|_{L_2(\gamma)}^2\right\} \tag{3.77}$$

is valid.

PROOF. Follows from previous arguments. □

Typically, the terms involving the differences $f - \Pi_p f$ and $g - \Pi_p g$ will be small compared to the other terms. In this sense, the estimator obtained by summing the local indicators provides an equivalent measure of the actual error in the energy norm. The local bound (3.77) is of importance for the design of adaptive algorithms showing that the estimator gives some indication of the error distribution and not simply a global bound.

3.5. The effect of numerical quadrature

In practice, the error indicator η_K will not be computed exactly since the integrals of the residuals will be performed using numerical quadrature. That is to say, the actual error indicator $\tilde{\eta}_K$ will be given by

$$\tilde{\eta}_K^2 = h_K^2 \|\bar{r}_K\|_{L_2(K)}^2 + \frac{1}{2} h_K \|\overline{R}_\gamma\|_{L_2(\partial K)}^2 \tag{3.78}$$

where $\bar{r}_K$ and $\overline{R}_\gamma$ are once again finite dimensional approximations of the actual data (but not necessarily the same choices as previously). In fact, if the data is continuous then the numerical quadrature would correspond to taking $\bar{r}_K$ to be the polynomial $I_p r$ which interpolates the residual r at the quadrature points. Equally well, the approximation $\overline{R}_\gamma$ would be the polynomial $I_p R$ (this time in one variable) which interpolates to R at the quadrature points on the boundary. This estimator can be analyzed as follows. Firstly, applying the Triangle Inequality gives

$$\tilde{\eta}_K^2 \leqslant \eta_K^2 + h_K^2 \|f - I_p f\|_{L_2(K)}^2 + \sum_{\gamma \subset \partial K \cap \Gamma_N} h_K \|g - I_p g\|_{L_2(\gamma)}^2 . \tag{3.79}$$

With the aid of Lemma 3.6 and proceeding much the same as before one obtains

$$\tilde{\eta}_K^2 \leqslant C \left\{ |||e|||_{\tilde{K}}^2 + h_K^2 \|f - \Pi_p f\|_{L_2(K)}^2 + \sum_{\gamma \subset \partial K \cap \Gamma_N} h_K \|g - \Pi_p g\|_{L_2(\gamma)}^2 \right\} . \tag{3.80}$$

Combining these results gives a result derived by Verfürth [56] (but see also [10] and Section 4.2.4):

THEOREM 3.8. Let $\tilde{\eta}_K$ denote the local error indicator

$$\tilde{\eta}_K^2 = h_K^2 \|\bar{r}_K\|_{L_2(K)}^2 + \frac{1}{2} h_K \|\overline{R}_\gamma\|_{L_2(\partial K)}^2 \tag{3.81}$$

where r and R are the interior and boundary residuals. Then there exists a constant C depending only on the shape regularity of the elements such that

$$|||e|||^2 \leqslant C \left\{ \sum_{K \in \mathcal{P}} \tilde{\eta}_K^2 + \sum_{K \in \mathcal{P}} h_K^2 \|f - \Pi_p f\|_{L_2(K)}^2 + \sum_{\gamma \subset \Gamma_N} h_K \|g - \Pi_p g\|_{L_2(\gamma)}^2 \right\} . \tag{3.82}$$

Moreover, the local bound

$$\tilde{\eta}_K^2 \leqslant C \left\{ |||e|||_{\tilde{K}}^2 + h_K^2 \|f - \Pi_p f\|_{L_2(K)}^2 + \sum_{\gamma \subset \partial K \cap \Gamma_N} h_K \|g - \Pi_p g\|_{L_2(\gamma)}^2 \right\} \tag{3.83}$$

is valid.

If the data f and g is smooth then the extra terms appearing in the bounds can be neglected showing that the estimator is in this sense equivalent to the actual error.

3.6. Error estimators for $W_p^1(\Omega)$ and $L_p(\Omega)$

3.6.1. Estimates in $W_p^1(\Omega)$, $1 < p < \infty$

The basic argument used to derive the simple error estimator in the energy norm can be generalized to obtain estimates in the norm on the space $W_p^1(\Omega)$. Let q denote the conjugate exponent

$$\frac{1}{p} + \frac{1}{q} = 1. \tag{3.84}$$

The residual equation (3.12) gives (with Π_X as in Theorem 1.1)

$$|B(e,v)| = \sum_{K\in\mathcal{P}} \int_K r(v - \Pi_X v)\,\mathrm{d}x + \sum_{\gamma\in\partial\mathcal{P}} \int_\gamma R(v - \Pi_X v)\,\mathrm{d}s \tag{3.85}$$

and applying Hölder's Inequality gives

$$|B(e,v)| \leqslant \sum_{K\in\mathcal{P}} \|r\|_{L_p(K)} \|v - \Pi_X v\|_{L_q(K)} + \sum_{\gamma\in\partial\mathcal{P}} \|R\|_{L_p(\gamma)} \|v - \Pi_X v\|_{L_q(\gamma)}. \tag{3.86}$$

Slightly different approximation theoretic results are needed: there exists a constant C which is independent of v and h_K such that [26]

$$\|v - \Pi_X v\|_{L_q(K)} \leqslant C h_K |v|_{W^1_q(\tilde{K})} \tag{3.87}$$

and

$$\|v - \Pi_X v\|_{L_q(\partial K)} \leqslant C h_K^{1-1/q} |v|_{W^1_q(\tilde{K})}. \tag{3.88}$$

Substituting these estimates and applying Hölder's Inequality leads to

$$|B(e,v)| \leqslant C\,|v|_{W^1_q(\Omega)} \left\{ \sum_{K\in\mathcal{P}} h_K^p \|r\|^p_{L_p(K)} + \sum_{\gamma\in\partial\mathcal{P}} h_K \|R\|^p_{L_p(\gamma)} \right\}^{1/p}. \tag{3.89}$$

As usual, the terms may be rearranged to identify contributions from each element

$$B(e,v) \leqslant C\,\|v\|_{W^1_q(\Omega)} \sum_{K\in\mathcal{P}} \left\{ h_K^p \|r\|^p_{L_p(K)} + \frac{1}{2} h_K \|R\|^p_{L_p(\partial K)} \right\}^2. \tag{3.90}$$

THEOREM 3.9. Let $\eta_{W^1_p(K)}$ denote the local error indicator

$$\eta^p_{W^1_p(K)} = h_K^p \|r\|^p_{L_p(K)} + \frac{1}{2} h_K \|R\|^p_{L_p(\partial K)} \tag{3.91}$$

where r and R are the interior and boundary residuals. Then there exists a constant C depending only on the shape regularity of the elements such that

$$\|e\|_{W^1_p(\Omega)} \leqslant C \left\{ \sum_{K\in\mathcal{P}} \eta^p_{W^1_p(K)} \right\}^{1/p}. \tag{3.92}$$

PROOF. The result follows on observing that

$$\|e\|_{W^1_p(\Omega)} \leqslant C \sup_{v\in W^1_q(\Omega)} \frac{B(e,v)}{\|v\|_{W^1_q(\Omega)}} \tag{3.93}$$

and then using the above arguments. □

The effects of using numerical quadrature to approximate the integrals in the estimator can easily be incorporated.

3.6.2. Estimates in $L_p(\Omega)$, $1 < p < \infty$

One might suspect that the same basic argument used to obtain estimates for the error in the L_2 norm can be extended to the L_p case. Once again, the Aubin–Nitsche Trick [25] is the essential idea. Consider the adjoint of the original model problem:

$$\Phi_F \in V : B(v, \Phi_F) = (F, v) \quad \forall v \in V \tag{3.94}$$

where $F \in L_q(\Omega)$ is given data. It is assumed that this problem is *regular* in the sense that the solution Φ_F

has the extra regularity $\Phi_F \in W_q^2(\Omega) \cap V$ and the solution operator from $L_q(\Omega)$ to $W_q^2(\Omega)$ is continuous

$$\|\Phi_F\|_{W_q^2(\Omega)} \leqslant C \|F\|_{L_q(\Omega)}. \tag{3.95}$$

The specific choice of v equal to the error function e then gives

$$(e, F) = B(e, \Phi_F). \tag{3.96}$$

The residual equation (3.12) is used as before,

$$B(e, \Phi_F) = \sum_{K \in \mathcal{P}} \int_K r(\Phi_F - \Pi_X \Phi_F)\, \mathrm{d}x + \sum_{\gamma \in \partial \mathcal{P}} \int_\gamma R(\Phi_F - \Pi_X \Phi_F)\, \mathrm{d}s \tag{3.97}$$

and applying Hölder's Inequality gives

$$B(e, F) \leqslant \sum_{K \in \mathcal{P}} \|r\|_{L_p(K)} \|v - \Pi_X v\|_{L_q(K)} + \sum_{\gamma \in \partial \mathcal{P}} \|R\|_{L_p(\gamma)} \|v - \Pi_X v\|_{L_q(\gamma)}. \tag{3.98}$$

The appropriate approximation theoretic results are that there exists a constant C which is independent of v and h_K such that [26]

$$\|v - \Pi_X v\|_{L_q(K)} \leqslant C h_K^2 |v|_{W_q^2(\widetilde{K})} \tag{3.99}$$

and

$$\|v - \Pi_X v\|_{L_q(\partial K)} \leqslant C h_K^{2-1/q} |v|_{W_q^2(\widetilde{K})}. \tag{3.100}$$

With the aid of Hölder's Inequality

$$(e, F) \leqslant C |\Phi_F|_{W_q^2(\Omega)} \left\{ \sum_{K \in \mathcal{P}} h_K^{2p} \|r\|_{L_p(K)}^p + \sum_{\gamma \in \partial \mathcal{P}} h_K^{1+p} \|R\|_{L_p(\gamma)}^p \right\}^{1/p} \tag{3.101}$$

and, thanks to the elliptic regularity assumption there follows:

$$(e, F) \leqslant C \|F\|_{L_q(\Omega)} \left\{ \sum_{K \in \mathcal{P}} h_K^{2p} \|r\|_{L_p(K)}^p + \sum_{\gamma \in \partial \mathcal{P}} h_K^{1+p} \|R\|_{L_p(\gamma)}^p \right\}. \tag{3.102}$$

A rearrangement gives an estimator in the familiar form apart from a different scaling

$$(e, F) \leqslant C \|F\|_{L_q(\Omega)} \sum_{K \in \mathcal{P}} \left\{ h_K^{2p} \|r\|_{L_p(K)}^p + \frac{1}{2} h_K^{1+p} \|R\|_{L_p(\partial K)}^p \right\}. \tag{3.103}$$

THEOREM 3.10. Suppose that the domain Ω is convex. Let $\eta_{L_p(K)}$ denote the local error indicator

$$\eta_{L_p(K)}^p = h_K^{2p} \|r\|_{L_p(K)}^p + \frac{1}{2} h_K^{1+p} \|R\|_{L_p(\partial K)}^p \tag{3.104}$$

where r and R are the interior and boundary residuals. Then there exists a constant C depending only on the shape regularity of the elements such that

$$\|e\|_{L_p(\Omega)} \leqslant C \left\{ \sum_{K \in \mathcal{P}} \eta_{L_p(K)}^p \right\}^{1/p}. \tag{3.105}$$

PROOF. The adjoint problem satisfies the regularity assumptions when the domain Ω is convex. The proof follows using the identity

$$\|e\|_{L_p(\Omega)} = \sup_{F \in L_q(\Omega)} \frac{(e, F)}{\|F\|_{L_q(\Omega)}} \tag{3.106}$$

and the previous arguments. □

4. Implicit a posteriori estimators

4.1. Introduction

Sections 2 and 3 deal with estimators that can be computed directly from the finite element approximation and the data for the original problem. The methods described in the present section require the solution of a local boundary value problem approximating the residual equation satisfied by the error itself. The local error estimator is the norm of the solution of the problem. Such schemes give rise to *implicit error estimators*. The boundary value problems are local in the sense that they are posed either over a single element or a small subdomain of elements.

One might wonder whether it is worthwhile considering implicit estimators at all, since explicit schemes are apparently much simpler. There are good reasons for using implicit schemes. Firstly, the explicit estimators lead to local error indicators containing generic constants. These constants are, in general, unknown. In practice, one can try to find a suitable bounds on the values of the constants. However, the values of constants would be dictated by the worst case scenario and therefore would usually give pessimistic estimators. Secondly, there are two types of residual present in the explicit estimators corresponding the interior and the boundary. The correct relative weighting to attach to each type of residual is far from obvious. In addition, it is conceivable that there are cancellations between the two types of residual that is lost when one deals with each separately.

Implicit estimators, and the element residual method in particular, avoid such issues by solving a boundary value problem with the residuals as data. In this way the generic constants are avoided and the correct balance between the two types of residual is catered for by the solution process itself. The drawback is that one is obliged to solve an auxiliary problem requiring an appropriate approximation scheme: as we shall see, this creates its own difficulties.

The ideas will be illustrated by considering the model problem in Section 1.5. The basic idea is to approximate the residual problem characterizing the true error:

$$B(e,v) = B(u,v) - B(u_X,v) = l(v) - B(u_X,v) \quad \forall v \in V. \tag{4.1}$$

4.2. The subdomain residual method

A method for a posteriori error estimation based on solving local residual problems with *homogeneous essential boundary data* over small patches or subdomains of the domain Ω was devised by Babuska and Rheinboldt [13]. The approach is particularly noteworthy because it provides a generally applicable method of a posteriori error estimation with firm theoretical foundations. Here, it will be assumed that the partition $\mathcal{P}$ is locally quasi-uniform, although Babuska and Rheinboldt [13] dealt with more general classes of partitions (see also [10]).

4.2.1. Formulation of subdomain residual problem

Let Ψ denote the set of element vertices in the partition $\mathcal{P}$ and $\{\theta_n\}_{n\in\Psi}$ denote the first-order Lagrange basis functions based at the element vertices. These functions are characterized by the conditions

$$\theta_n(\boldsymbol{x}_m) = \delta_{nm} \tag{4.2}$$

where $\boldsymbol{x}_m$ is any vertex in Ψ and

$$\sum_{n\in\Psi} \theta_n(\boldsymbol{x}) \equiv 1, \quad \boldsymbol{x} \in \overline{\Omega}. \tag{4.3}$$

The support $\tilde{\Omega}_n$ of the nodal function θ_n consists of the patch of elements containing the vertex $\boldsymbol{x}_n$.

The method is formulated starting with Eq. (4.1) characterizing the true error $e \in V$. The usual considerations apply: while, in principle, one could approximate the solution of this equation using a refinement of the finite element subspace X, it would be simpler to compute a new finite element approximation directly. The underlying idea is to replace the single global problem characterizing the error by a sequence of independent problems posed on small subdomains of the partition $\mathcal{P}$. The nodal basis functions may be utilized for this purpose. With the aid of property (4.3)

$$B(e,v) = B\left(e, v\sum_{n\in\Psi}\theta_n\right) = \sum_{n\in\Psi} B(e, v\theta_n) = \sum_{n\in\Psi} L(v\theta_n) - B(u_X, v\theta_n) \tag{4.4}$$

where $v \in V$. Each of the functions $\theta_n v$ belongs to the space $H_0^1(\tilde{\Omega}_n)$. Define the local bilinear form $B_n : H_0^1(\tilde{\Omega}_n) \times H_0^1(\tilde{\Omega}_n) \mapsto \mathbb{R}$ by

$$B_n(u,v) = \int_{\tilde{\Omega}_n} (\nabla u \cdot \nabla v + cuv)\, \mathrm{d}x \tag{4.5}$$

and the local load functional $L_n : H_0^1(\tilde{\Omega}_n) \mapsto \mathbb{R}$ by

$$L_n(v) = \int_{\tilde{\Omega}_n} fv\, \mathrm{d}x + \int_{\partial\tilde{\Omega}_n \cap \Gamma_N} gv\, \mathrm{d}s. \tag{4.6}$$

The *subdomain residual problem* is to find $\phi_n \in H_0^1(\tilde{\Omega}_n)$ such that

$$B_n(\phi_n, v) = L_n(v) - B_n(u_X, v) \quad \forall v \in H_0^1(\tilde{\Omega}_n). \tag{4.7}$$

Motivated by Eq. (4.4), the error estimator η_n associated with the subdomain $\tilde{\Omega}_n$ is taken to be

$$\eta_n = |||\phi_n|||_{\tilde{\Omega}_n} \tag{4.8}$$

and the global error estimator η is obtained by summing

$$\eta = \left\{\sum_{n\in\Psi} \eta_n^2\right\}^{1/2}. \tag{4.9}$$

In practice, the subdomain residual problems are approximated using a finite dimensional subspace of $H_0^1(\tilde{\Omega}_n)$. We shall return to this point in Section 4.2.4.

4.2.2. *Nomenclature and assumptions*

Suppose that the basis function θ_n is non-zero on an element K; then

$$|\theta_n(x)| \leqslant 1, \quad x \in K \tag{4.10}$$

and, moreover, there exists a constant C depending only on the non-degeneracy of the element such that

$$|\nabla\theta_n(x)| \leqslant Ch_K^{-1}. \quad x \in K. \tag{4.11}$$

It is possible to partition the set of vertices Ψ into the union of disjoint subsets $\Psi_1, \Psi_2, \ldots$ such that any pair of nodal basis functions in the same subset Ψ_r have non-overlapping supports. Specifically, the condition that will be required is

$$\forall\theta_n, \theta_m \in \Psi_r : m \neq n \Rightarrow \quad \mathrm{supp}^0\theta_n \cap \mathrm{supp}^0\theta_m \text{ is empty} \tag{4.12}$$

where $\mathrm{supp}^0\theta_n$ denotes the interior of the support of θ_n. It is easy to see that this process is always possible since one can simply choose each of the sets Ψ_r to consist of a single basis function. Later, it will be found advantageous if this is accomplished using as few subsets as possible. The smallest number of subsets will be denoted by ρ and referred to as the *overlap index* for the partition.

Let K be any element in the partition $\mathcal{P}$. The set of basis functions which are non-zero on this element is denoted by $\sigma(K)$. The maximum cardinality of any of these sets is denoted by τ and referred to as the *intersection index*. If the partition is locally quasi-uniform then the intersection index will be equal to the maximum number of vertices in any element. However, if the mesh is not proper then the intersection index may increase without bound as the refinement progresses. Such a situation is disallowed by requiring the intersection index τ and the overlap index ρ are uniformly bounded for the family of partitions.

The subdomain consisting of the elements which neighbour element K is denoted by

$$\tilde{K} = \operatorname{int}\left\{\bigcup \bar{J} : \bar{J} \cap \bar{K} \text{ is non-empty}\right\}. \tag{4.13}$$

Since the elements are non-degenerate the maximum number of elements contained in any of these subdomains is bounded by a multiple of τ. Local approximation properties will be needed. In particular, there exists [26] a constant C such that for any $v \in H^1(\Omega)$ and any element K there holds

$$h_K^{-1} \|v - \Pi_X v\|_{L_2(K)} + |v - \Pi_X v|_{L_2(K)} \leqslant C |v|_{H^1(\tilde{K})} \tag{4.14}$$

where Π_X is as in Theorem 1.1.

The assumptions are true regardless of the polynomial degree actually used to construct the finite element subspace, and merely impose restrictions on the mesh topology and mesh geometry. The conditions are always satisfied should the partition be locally quasi-uniform.

4.2.3. Equivalence of estimator

LEMMA 4.1. Suppose that the above conditions hold. Then there exists a constant C depending only on the non-degeneracy of the elements such that for any $v \in H^1(\Omega)$

$$\sum_{n \in \Psi} \|\theta_n(v - \Pi_X v)\|^2_{H^1(\Omega)} \leqslant C\tau^2 |v|^2_{H^1(\Omega)}. \tag{4.15}$$

PROOF. Let K be any element and $\theta_n \in \Psi$ be a nodal basis function which does not vanish on K. Then,

$$\|\theta_n(v - \Pi_X v)\|^2_{H^1(K)} \leqslant \|\theta_n\|^2_{L_\infty(K)} |v - \Pi_X v|^2_{H^1(K)} + |\theta_n|^2_{W^{1,\infty}(K)} \|v - \Pi_X v\|^2_{L_2(K)}.$$

Thanks to properties (4.10)–(4.11)

$$\|\theta_n(v - \Pi_X v)\|^2_{H^1(K)} \leqslant C\{|v - \Pi_X v|^2_{H^1(K)} + h_K^{-2} \|v - \Pi_X v\|^2_{L_2(K)}\}$$

and using the approximation property gives

$$\|\theta_n(v - \Pi_X v)\|^2_{H^1(K)} \leqslant C |v|^2_{H^1(\tilde{K})}.$$

Summing this inequality over all vertices gives the result

$$\begin{aligned}
\sum_{n \in \Psi} \|\theta_n(v - \Pi_X v)\|^2_{H^1(\Omega)} &= \sum_{K \in \mathcal{P}} \sum_{n \in \sigma(K)} \|\theta_n(v - \Pi_X v)\|^2_{H^1(K)} \\
&\leqslant C \sum_{K \in \mathcal{P}} \sum_{n \in \sigma(K)} |v|^2_{H^1(\tilde{K})} \\
&= C\tau \sum_{K \in \mathcal{P}} |v|^2_{H^1(\tilde{K})} \\
&\leqslant C\tau^2 |v|^2_{H^1(\Omega)}
\end{aligned}$$

as claimed. □

THEOREM 4.2. Let η_n denote the local error estimator obtained using the subdomain residual method. Then there exists a constant C depending only on the non-degeneracy of the elements such that

$$\frac{\eta}{\sqrt{\rho}} \leqslant |||e||| \leqslant C\tau\eta \tag{4.16}$$

where τ is the overlap index and ρ is the intersection index.

PROOF. Using the Galerkin orthogonality

$$B(e, v) = B(e, v - \Pi_X v)$$

and then by property (4.3)

$$B(e,v) = \sum_{n\in\Psi} B(e, \theta_n(v - \Pi_X v))$$

and noting that $\theta_n(v - \Pi_X v) \in H_0^1(\tilde{\Omega}_n)$

$$B(e,v) = \sum_{n\in\Psi} B(\phi_n, \theta_n(v - \Pi_X v)).$$

Hence

$$|B(e,v)| \leqslant \sum_{n\in\Psi} \eta_n |||\theta_n(v - \Pi_X v)|||.$$

Applying the Cauchy–Schwarz Inequality and Lemma 4.1

$$|B(e,v)| \leqslant C\eta \left\{ \sum_{n\in\Psi} \|\theta_n(v - \Pi_X v)\|^2_{H^1(\Omega)} \right\}^{1/2} \leqslant C\eta\tau |v|_{H^1(\Omega)} \leqslant C\eta\tau |||v|||$$

from which the result follows immediately

$$|||e||| \leqslant C\tau\eta.$$

Conversely, consider

$$\begin{aligned}
\eta^2 &= \sum_{n\in\Psi} \eta_n^2 = \sum_{n\in\Psi} B_n(\phi_n, \phi_n) \\
&= \sum_{n\in\Psi} B_n(e, \phi_n) = \sum_{n\in\Psi} B(e, \phi_n) \\
&= B(e, \sum_{n\in\Psi} \phi_n) \\
&\leqslant |||e||| \, |||\sum_{n\in\Psi} \phi_n|||.
\end{aligned}$$

By partitioning the set Ψ of vertices as described above

$$|||\sum_{n\in\Psi} \phi_n|||^2 = |||\sum_{\Psi_i} \sum_{n\in\Psi_i} \phi_n|||^2 \leqslant \rho \sum_{\Psi_i} |||\sum_{n\in\Psi_i} \phi_n|||^2 = \rho \sum_{\Psi_i} \sum_{n\in\Psi_i} |||\phi_n|||^2$$

where property (4.12) has been used. Consequently,

$$|||\sum_{n\in\Psi} \phi_n|||^2 \leqslant \rho\eta^2$$

and the result follows immediately. □

4.2.4. *Treatment of residual problems*

The subdomain residual method described above requires the *exact* solution of the local problems

$$\phi_n \in H_0^1(\tilde{\Omega}_n) : B_n(\phi_n, v_n) = L(v_n) - B_n(u_X, v_n) \quad \forall v_n \in H_0^1(\tilde{\Omega}_n). \tag{4.17}$$

In practice, the method is seldom used, partly because it is inconvenient and relatively expensive to develop approximations over the patches $\tilde{\Omega}_n$. Babuska and Miller [10] circumvented this difficulty by obtaining an equivalent measure for the local error estimator $|||\phi_n|||$ as follows. Integrating the right-hand side of (4.17) by parts gives

$$L(v_n) - B_n(u_X, v_n) = \sum_{K\in\tilde{\Omega}_n} \left\{ \int_K r v_n \, dx + \sum_{\gamma\subset\tilde{\Omega}_n} \int_\gamma R v_n \, ds \right\} \tag{4.18}$$

where r and R are the usual interior and boundary residuals. A routine application of the Cauchy–Schwarz Inequality leads to the bound

$$\eta_n \leqslant C \left\{ \sum_{K \subset \widetilde{\Omega}_n} h_K^2 \|r\|_{L^2(K)}^2 + \sum_{\gamma \subset \widetilde{\Omega}_n} h_K \|R\|_{L^2(\gamma)}^2 \right\}^{1/2}. \tag{4.19}$$

Summing this estimate over all vertices in the partition and regrouping the terms leads to the familiar *explicit error estimator*

$$|||e||| \leqslant C \sum_{K \in \mathcal{P}} \eta_K^2 \tag{4.20}$$

where

$$\eta_K^2 = \|r\|_{L^2(K)}^2 + \sum_{\gamma \subset \partial K} h_K \|R\|_{L^2(\gamma)}^2. \tag{4.21}$$

Babuska and Miller also obtain the following result:

$$\sum_{K \subset \widetilde{\Omega}_n} h_K \|r\|_{L^2(K)} + \sum_{\gamma \subset \widetilde{\Omega}_n} h_K^{1/2} \|R\|_{L^2(\gamma)} \leqslant C(\eta_n + \epsilon) \tag{4.22}$$

where

$$\epsilon = \sum_{K \subset \widetilde{\Omega}_n} h_K^2 \|r\|_{L^2(K)}^2 + \sum_{\gamma \subset \widetilde{\Omega}_n} h_K^{1/2} \|R - \overline{R}_\gamma\|_{L^2(\gamma)}. \tag{4.23}$$

This result is essentially identical to those derived in Section 3.4 when considering explicit error estimators but predates the discovery of those estimates.

4.3. The element residual method

4.3.1. Formulation of local residual problem

In principle, one could approximate the problem (4.1) and obtain an approximation to the actual error function. The optimality of the Galerkin method and the associated orthogonality property of the error, means that one must use a larger subspace than the original finite element subspace X if one is to obtain a non-zero approximation to the error. The cost of solving the problem would be comparable with simply resolving the original problem using the finer discretization. One can attempt to reduce the cost of solving this global problem by replacing it by a number of independent local problems posed over each element in the domain. The local problems could then be approximated relatively inexpensively and even solved in parallel.

Let the error on an element K be denoted by $e = u - u_X$. Thanks to the smoothness of the finite element approximation on the element interiors, one finds that the error satisfies the differential equation

$$-\Delta e + ce = f + \Delta u_X - c u_X \quad \text{in } K. \tag{4.24}$$

The major difficulty is to supplement the equation with appropriate boundary conditions. There are various cases to consider. First, suppose that the element K intersects a portion of the boundary of the domain Ω where an essential boundary condition is imposed. The appropriate boundary condition for the local error residual problem is clearly

$$e = 0 \quad \text{on } \partial K \cap \Gamma_D. \tag{4.25}$$

Here, it has been assumed that the finite element approximation has been constructed so that the essential boundary conditions are satisfied exactly (although this is not essential). Next, suppose that the element intersects a portion of the boundary $\partial\Omega$ where a natural boundary condition is imposed. The local error residual problem is subjected to a natural boundary condition

$$\frac{\partial e}{\partial n_K} = g - \frac{\partial u_X}{\partial n_K} \quad \text{on } \partial K \cap \Gamma_N. \tag{4.26}$$

So far, appropriate boundary conditions have been clear. It remains to consider the case when the element boundary lies on the interior of the domain. The first decision is whether to impose an essential or a natural boundary condition. The *element residual method* is based on using a natural boundary condition. Ideally, one would like to impose the condition

$$\frac{\partial e}{\partial n_K} = \frac{\partial u}{\partial n_K} - \left.\frac{\partial u_X}{\partial n_K}\right|_K \tag{4.27}$$

on the edge separating elements K and J, where $u_X|_K$ denotes the restriction of the finite element approximation to an element K. Unfortunately, the true flux appearing on the right-hand side is unknown in general. However, one may replace the true flux by an approximation obtained from the finite element approximation itself

$$\frac{\partial u}{\partial n_K} \approx \left\langle \frac{\partial u_X}{\partial n_K} \right\rangle \tag{4.28}$$

where we define

$$\left\langle \frac{\partial u_X}{\partial n_K} \right\rangle = \frac{1}{2} n_K \cdot \{(\nabla u_X)_K + (\nabla u_X)_J\}. \tag{4.29}$$

The motivation for this choice is founded on the hope that by averaging the discontinuous finite element approximation to the normal flux one obtains a reasonable approximation to the true flux.

The error residual problem is formulated as a variational equation as follows. On each element K the actual error satisfies the boundary value problem

$$B_K(e, v) = F_K(v) - B_K(u_X, v) + \int_{\partial K} \frac{\partial u}{\partial n_K} v \, \mathrm{d}s \quad \forall v \in V_K. \tag{4.30}$$

Here,

$$V_K = \{v \in H^1(K) : v = 0 \text{ on } \partial K \cap \Gamma_D\} \tag{4.31}$$

and $B_K : V_K \times V_K \mapsto \mathbb{R}$ is the local bilinear form

$$B_K(u, v) = \int_K (\nabla u \cdot \nabla v + cuv) \, \mathrm{d}x \tag{4.32}$$

and $F_K : V_K \mapsto \mathbb{R}$ is the local load functional

$$F_K(v) = \int_K fv \, \mathrm{d}x. \tag{4.33}$$

The *error residual problem* is the weak form of the problem specified by the conditions (4.24)-(4.28): find $\phi_K \in V_K$ such that

$$B_K(\phi_K, v) = F_K(v) - B_K(u_X, v) + \int_{\partial K} \left\langle \frac{\partial u_X}{\partial n_K} \right\rangle v \, \mathrm{d}s \quad \forall v \in V_K. \tag{4.34}$$

Here, the definition of the average has been extended to include the portion Γ_N of the exterior boundary

$$\left\langle \frac{\partial u_X}{\partial n_K} \right\rangle = \begin{cases} \frac{1}{2} n_K \cdot \{(\nabla u_X)_K + (\nabla u_X)_J\}, & \text{on } \overline{K} \cap \overline{J} \\ g, & \text{on } \overline{K} \cap \Gamma_N. \end{cases} \tag{4.35}$$

The question naturally arises as to whether there exists solutions of the residual problem (4.34). In general, the answer is negative! The difficulty arises from the non-trivial kernel of associated with the bilinear form $B_K(\cdot, \cdot)$. Unless the data satisfies appropriate compatibility conditions, the problem will fail to possess solutions. Therefore, extra assumptions must be made to ensure well-posedness. One alternative is to work on a subspace of V_K on which the bilinear form will be coercive. This approach, although apparently crude, leads to useful error estimators. An alternative approach is to try to choose the boundary data more carefully so that the underlying problem is naturally well-posed so that there is no need to resort to altering the space V_K. This approach, known as the equilibrated residual method, will be discussed in the next section.

4.3.2. Subspaces for the element residual method

The error residual problem (4.34) is an infinite dimensional problem. The element residual method is obtained by constructing particular finite dimensional subspaces Y_K of the local spaces V_K. Here, the discussion is restricted to first-order finite element approximation; the general case will be dealt with later.

Quadrilateral elements

Consider the case of finite element approximation using piecewise bilinear functions on quadrilateral elements. The local approximation space for the error residual problem is constructed using the basis functions specified on the reference element

$$\widehat{K} = \{(\hat{x}, \hat{y}) : -1 \leqslant \hat{x} \leqslant 1; \quad -1 \leqslant \hat{y} \leqslant 1\}. \tag{4.36}$$

The edge bubble functions are given by

$$\begin{aligned} \hat{\chi}_1 &= \frac{1}{2}(1-\hat{x}^2)(1-\hat{y}) \\ \hat{\chi}_2 &= \frac{1}{2}(1-\hat{y}^2)(1+\hat{x}) \\ \hat{\chi}_3 &= \frac{1}{2}(1-\hat{x}^2)(1+\hat{y}) \\ \hat{\chi}_4 &= \frac{1}{2}(1-\hat{y}^2)(1-\hat{x}) \end{aligned} \tag{4.37}$$

and the interior bubble function $\hat{\psi}$ is given by

$$\hat{\psi} = (1-\hat{x}^2)(1-\hat{y}^2). \tag{4.38}$$

The space $\widehat{Y}$ is defined by

$$\widehat{Y} = \operatorname{span}\{\hat{\chi}_1, \hat{\chi}_2, \hat{\chi}_3, \hat{\chi}_4, \hat{\psi}\} \tag{4.39}$$

and the error residual problem is approximated using the subspace Y_K obtained by mapping the *bubble space* $\widehat{Y}$ to the element K.

Triangular elements

The paper by Bank and Weiser [23] on the element residual method considered the case of finite element approximation using piecewise affine functions on linear triangular elements. The local approximation space for the error residual problem is constructed using the basis functions specified on the reference element

$$\widehat{K} = \{(\hat{x}, \hat{y}) : 0 \leqslant \hat{x} \leqslant 1; \quad 0 \leqslant \hat{y} \leqslant 1 - \hat{x}\}. \tag{4.40}$$

The functions $\hat{\lambda}_1$, $\hat{\lambda}_2$ and $\hat{\lambda}_3$ denote the barycentric (area) coordinates on $\widehat{K}$ and the edge bubble functions are given by

$$\hat{\chi}_1 = 4\hat{\lambda}_2\hat{\lambda}_3; \qquad \hat{\chi}_2 = 4\hat{\lambda}_1\hat{\lambda}_3; \qquad \hat{\chi}_3 = 4\hat{\lambda}_1\hat{\lambda}_2. \tag{4.41}$$

Let $\widehat{Y}$ denote the space

$$\widehat{Y} = \operatorname{span}\{\hat{\chi}_1, \hat{\chi}_2, \hat{\chi}_3\}. \tag{4.42}$$

The error residual problem is approximated using the subspace Y_K obtained by mapping the bubble space $\widehat{Y}$ to the element K. The bubble space suggested by Bank and Weiser contains no interior bubble function. Verfürth [56] has suggested an alternative bubble space based on including the extra interior bubble function.

In each of these cases, the basic pattern for approximating the local problem (4.34) is to increase the order of the space used to construct the original finite element approximation and then factor out the

functions which are non-zero at the vertices of the element. This may be achieved by subtracting the interpolant in the original finite element space.

The discussion has been restricted to first order finite element approximation. For higher-order approximation, some care has to be exercised when selecting the subspaces with which to approximate the residual problem (see [3] and Section 4.5.3).

4.3.3. The classical element residual method

The *classical error residual method* [22,23,29,48] for a posteriori error estimation is to solve the local error residual problems

$$\phi_K \in Y_K: \quad B_K(\phi_K, v) = F_K(v) - B_K(u_X, v) + \int_{\partial K} \left\langle \frac{\partial u_X}{\partial n_K} \right\rangle v \, ds \quad \forall v \in Y_K \tag{4.43}$$

thereby obtaining a function ϕ_K. The local error estimator η_K on element K is defined by

$$\eta_K = |||\phi_K|||_K \tag{4.44}$$

and the global error estimator is obtained by summing the local contributions

$$\eta = \left\{ \sum_{K \in \mathcal{P}} \eta_K^2 \right\}^{1/2}. \tag{4.45}$$

As remarked earlier, such estimators will be referred to as *implicit estimators*.

4.4. Equivalence of estimator

4.4.1. Relationship with explicit error estimators

The implicit estimators can be related to the explicit estimators of Section 3. By applying Greens identity to the right-hand side of the error residual equation (4.43) one obtains

$$B_K(\phi_K, v) = \int_K rv \, dx + \int_{\partial K} Rv \, ds \quad \forall v \in Y_K \tag{4.46}$$

where r and R are the precisely the interior and boundary residuals appearing in the explicit error estimators. Choosing the function v to be the estimated error function ϕ_K leads to

$$\eta_K^2 = \int_K r\phi_K \, dx + \int_{\partial K} R\phi_K \, ds. \tag{4.47}$$

The function ϕ_K belongs to a finite dimensional space. Moreover, if the gradient of ϕ_K vanishes then ϕ_K itself must be the zero function. Therefore, there exists a constant C which depends only on the shape regularity of the element K but not on its size such that

$$\|\phi_K\|_{L_2(K)} \leqslant C h_K \, |\phi_K|_{H^1(K)} \tag{4.48}$$

and for any edge $\gamma \subset \partial K$

$$\|\phi_K\|_{L_2(\gamma)} \leqslant C h_K^{1/2} \, |\phi_K|_{H^1(K)}. \tag{4.49}$$

With the aid of these results and the Cauchy–Schwarz Inequality

$$\eta_K^2 \leqslant C \left\{ h_K \, \|r\|_{L_2(K)} + \sum_{\gamma \subset \partial K} h_K^{1/2} \, \|R\|_{L_2(\gamma)} \right\} |\phi_K|_{H^1(K)} \tag{4.50}$$

from which one can deduce that

$$\eta_K \leqslant C \left\{ h_K^2 \, \|r\|_{L_2(K)}^2 + \frac{1}{2} \sum_{\gamma \subset \partial K} h_K \, \|R\|_{L_2(\gamma)}^2 \right\}^{1/2} \tag{4.51}$$

where term on the right-hand side is the explicit error estimator discussed earlier.

4.4.2. Further bounds

Suppose that the space Y_K used to approximate the error residual problem is equipped with an interior bubble function. This assumption rules out the space chosen by Bank and Weiser [23] for piecewise affine approximation on linear triangles, but is satisfied if the space is chosen as suggested by Verfürth [56]. Let $\psi_K \in Y_K$ denote the interior bubble constructed in Theorem 3.3. The following result complements Lemma 3.6:

LEMMA 4.3. Let r and R denote the interior and boundary residuals associated with the finite element approximation constructed from the subspace X. Suppose that $\overline{r}_K$ and $\overline{R}_\gamma$ are finite dimensional approximations to the residuals on an element K and on an edge $\gamma \subset \partial K$, and let η_K denote the element residual error estimator. Then there exists a constant C depending only on the shape regularity of the elements such that

$$\|r\|_{L_2(K)} \leqslant C\{h_K^{-1}\eta_K + \|\overline{r}_K - r\|_{L_2(K)}\} \tag{4.52}$$

and

$$\|R\|_{L_2(\gamma)} \leqslant C\{h_K^{1/2}\eta_K + h_K^{1/2}\|\overline{r}_K - r\|_{L_2(K)} + \|\overline{R}_\gamma - R\|_{L_2(\gamma)}\}. \tag{4.53}$$

PROOF. The first result is shown by following the arguments from (3.52) to (3.60), except that the identity (obtained by choosing $v = \overline{r}_K\psi_K$ in (4.43))

$$B_K(\phi_K, \overline{r}_K\psi_K) = \int_K \psi_K r\overline{r}_K \,\mathrm{d}x$$

is used in place of (3.53). The second result is obtained by following similar arguments to those from (3.61) to (3.67) and using the identity (which follows from (4.43) with the choice $v = \overline{R}_\gamma\chi_\gamma$)

$$B_K(\phi_K, \overline{R}_\gamma\chi_\gamma) = \int_K r\overline{R}_\gamma\chi_\gamma \,\mathrm{d}x + \int_\gamma \chi_\gamma R\overline{R}_\gamma \,\mathrm{d}s$$

instead of (3.62). □

4.4.3. Proof of equivalence

Thanks to the relationship between the explicit and implicit error estimators, one immediately obtains the following result as a corollary of Theorem 3.7 (cf. [56]):

THEOREM 4.4. Let η_K denote the local error estimator obtained using the element residual method. Then there exists a constant C depending only on the shape regularity of the elements such that

$$|||e|||^2 \leqslant C\sum_{K\in\mathcal{P}}\left\{\eta_K^2 + h_K^2\|f - \Pi_p f\|_{L_2(K)}^2 + \sum_{\gamma\subset\partial K\cap\Gamma_N} h_K\|g - \Pi_p g\|_{L_2(\gamma)}^2\right\}. \tag{4.54}$$

Moreover, the local bound

$$\eta_K^2 \leqslant C\left\{|||e|||_{\tilde{K}}^2 + h_K^2\|f - \Pi_p f\|_{L_2(K)}^2 + \sum_{\gamma\subset\partial K\cap\Gamma_N} h_K\|g - \Pi_p g\|_{L_2(\gamma)}^2\right\} \tag{4.55}$$

holds for all elements $K \in \mathcal{P}$.

As noted elsewhere, the extra terms depend on the smoothness of the data and will often be negligible in comparison with the estimator and the actual error. In this sense, the element residual method gives an equivalent measure of the discretization error in the energy norm.

4.4.4. The effect of numerical quadrature

In practice, just as for explicit error estimators, the computations of the the integrals appearing in the element residual equation will be performed using numerical quadrature. Therefore, the computed

error estimator $\tilde{\eta}_K$ will be obtained by solving the perturbed problem

$$\tilde{\phi}_K \in Y_K: \quad B_K(\tilde{\phi}_K, v) = \tilde{L}_K(v) - B_K(u_X, v) + \int_{\partial K} \left\langle \frac{\partial u_X}{\partial n_K} \right\rangle v \, ds \quad \forall v \in Y_K \tag{4.56}$$

and where $\tilde{L}_K : Y_K \mapsto \mathbb{R}$ is the functional

$$\tilde{L}_K(v) = \int_K (I_p f) v \, dx + \int_{\partial K \cap \Gamma_N} (I_p g) v \, ds \tag{4.57}$$

where $I_p f$ and $I_p g$ are the polynomials which interpolate to the data at the quadrature points (assuming the data is sufficiently smooth). This estimator can be analyzed by noting that

$$B_K(\phi_K, v) = B_K(\tilde{\phi}_K, v) + \int_K (f - I_p f) v \, dx + \int_{\partial K \cap \Gamma_N} (g - I_p g) v \, ds \tag{4.58}$$

from which one obtains

$$\left| |||\phi_K|||_K - |||\tilde{\phi}_K|||_K \right| \leqslant C \left\{ h_K \|f - I_p f\|_{L_2(K)} + \sum_{\gamma \subset \partial K \cap \Gamma_N} h_K^{1/2} \|g - I_p g\|_{L_2(\gamma)} \right\}. \tag{4.59}$$

This estimate used in conjunction with the previous result gives:

THEOREM 4.5. Let $\tilde{\eta}_K$ denote the error estimator obtained from the element residual method in the presence of numerical quadrature. Then there exists a constant C depending only on the shape regularity of the elements such that

$$|||e|||^2 \leqslant C \left\{ \sum_{K \in \mathcal{P}} \tilde{\eta}_K^2 + \sum_{K \in \mathcal{P}} h_K^2 \|f - I_p f\|_{L_2(K)}^2 + \sum_{\gamma \subset \Gamma_N} h_K \|g - I_p g\|_{L_2(\gamma)}^2 \right\}. \tag{4.60}$$

Moreover, the local bound

$$\tilde{\eta}_K^2 \leqslant C \left\{ |||e|||_{\tilde{K}}^2 + h_K^2 \|f - I_p f\|_{L_2(K)}^2 + \sum_{\gamma \subset \partial K \cap \Gamma_N} h_K \|g - I_p g\|_{L_2(\gamma)}^2 \right\} \tag{4.61}$$

is valid.

4.5. Performance of estimators

The robustness and quality of estimators along with the identification of limits of their performance is of vital importance. Studies of this type have been undertaken by Babuska et al. [15,16]. Here, the influence of the subspace used to solve the element residual problem is studied. In practice, the only feasible approach is to construct an approximate solution to the local problem. The choice of subspace used for first-order finite elements is reasonably well established and understood: see for example [23] and Section 4.3.1. For higher-order finite element approximation, the situation is less clear and some unpleasant surprises are lurking.

Although the classical element residual method is popular, relatively little is known about its performance. For instance, it was only comparatively recently proved by Duran and Rodriguez [31] that the classical element residual error estimator converges to the actual error in the energy norm for regular solutions on parallel meshes of linear triangular elements.

Throughout, it will be assumed that the mesh and the true solution are as regular as necessary for the analysis. This is an unrealistic assumption from the practical viewpoint, but serves to isolate the effects associated with the approximate solution of the local residual problem from effects due to singularities, non-smooth domain, irregular meshes and so on. A simple test of the performance of an error estimator is consistency (or *asymptotic exactness*). Roughly speaking, an a posteriori error estimator is said to be asymptotically exact if the ratio of the estimated error to the actual error tends to unity as the mesh size tends to zero. Whilst asymptotic exactness should not be over-emphasized, it is useful to understand

how the estimator behaves in the most favourable case when the mesh is regular and the underlying problem smooth.

To begin with, we analyze the case when the local residual problem is solved exactly. As already noted, this is not a viable proposition, but will bring into sharp relief the effects not associated with the approximate solution of the local problem. The selection of suitable approximate subspaces is then discussed. It is shown that *the estimators arising from solving the local problems approximately perform better than if the problems are solved exactly*.

4.5.1. Exact solution of element residual problem

In this section, we review the analysis in [2] for the classical element residual technique applied to pth order finite element approximation on quadrilateral elements. It is assumed that the local residual problems determining the error estimator are solved 'exactly'. Of course, it has already been stated that there will not, in general, be an exact solution of the residual problem. Here, the term 'exact' is understood to mean that the space $H^1(K)/\mathbb{R}$ is used in place of V_K. The analysis will show that in the case of *odd* order approximation of regular solutions on meshes consisting of quasi-uniform *square* elements, the classical element residual estimator is asymptotically exact in the energy norm.

The conditions under which the result is demonstrated at first appear unnecessarily strict. It is surprising that the result is the best possible. A counterexample will be given of even order approximation on square elements of a problem with a smooth (polynomial) solution where the element residual scheme asymptotically tends to overestimate the true error by a factor of $\sqrt{2(p+1)}$. A second example (again with polynomial solution) on a uniform partition gives an error estimate which tends to overestimate by a factor of at least $1 + C\cos^2\theta$ where θ is the angle between the normals to the element edges. Asymptotic exactness is therefore seen to be a rather fragile property.

The proof makes use of superconvergence results due to Lesaint and Zlamal [44]. However, while the superconvergence results continue to hold for approximations of all orders and for partitions containing elements other than squares, by themselves they are insufficient to guarantee the effectiveness of the estimator. A difference between the behaviour of error estimators for odd versus even order approximation has already been observed by Babuska and Yu [19]. The results will shed light on the source of this difference.

Assumptions on the mesh

It will be assumed that the domain Ω may be obtained by an *affine* invertible mapping of a reference grid. The reference grid is supposed to consist of squares with sides parallel to the coordinate axes. The images of each of the squares under the mapping generate a partitioning $\mathcal{P}$ of the domain Ω into the union of non-overlapping quadrilaterals. The assumptions are extremely stringent, essentially restricting one to square elements. However, it will be found even these are insufficient to guarantee the asymptotic exactness of the a posteriori error estimators.

These conditions satisfy the assumptions (2.6)–(2.7) and (4.5) in [44], under which they were able to show that if the true solution u of the model problem is sufficiently regular, then the finite element approximation u_X exhibits superconvergence. The following special case of their result will be useful:

THEOREM 4.6. Suppose that the finite element partitions satisfy the above conditions and that all integrals are evaluated exactly. If the true solution u belongs to $H^{p+2}(\Omega)$, then there exists a positive constant C such that

$$|||\Pi_X u - u_X||| \leqslant Ch^{p+1}\,\|u\|_{H^{p+2}(\Omega)} \tag{4.62}$$

where $\Pi_X u \in X$ is the interpolant to u at the Gauss–Lobatto nodes.

PROOF. For any $v \in X$ it follows from the orthogonality of the error in the Galerkin approximation that $B(\Pi_X u - u_X, v) = B(\Pi_X u - u, v)$. The result then follows on recalling the following result (Eq. (4.9) in [44]):

$$|B(\Pi_X u - u, v)| \leqslant Ch^{p+1} \|u\|_{H^{p+2}(\Omega)} |v|_{H^1(\Omega)} \tag{4.63}$$

and choosing $v = \Pi_X u - u_X$. □

Accuracy of averaged flux

The accuracy of the flux approximation on the element boundaries is critical to the performance of the error estimator η. Therefore, let γ be an interelement edge. The key result concerning the accuracy of the averaged flux is

THEOREM 4.7. Let $p \in \mathbb{N}$ be odd and $v \in H^{p+2}(\Omega)$. Suppose that the partition consists of square subdomains. Then there exists a constant C, depending only on p, such that

$$\left\| \frac{\partial v}{\partial n} - \left\langle \frac{\partial \Pi_X v}{\partial n} \right\rangle \right\|_{L_2(\gamma)} \leqslant Ch^{p+1/2} |v|_{H^{p+2}(\tilde{\gamma})} \tag{4.64}$$

where $\tilde{\gamma}$ is the subdomain consisting of the pair of elements sharing the edge γ.

PROOF. See [2]. □

The hypotheses that the polynomial degree be odd and the elements be square might seem to be overly strong. However, counterexamples [2] show that the result is sharp.

4.5.2. Performance of the error estimator

This section contains three principal results. The element residual error estimator is shown to be asymptotically exact for odd degree elements on meshes of squares provided the true solution is sufficiently smooth. It is shown that this result is sharp. Counterexamples are given showing that the estimator is not asymptotically exact for even order elements on square elements and that the estimator is not asymptotically exact for odd order elements when the elements are not approximately square.

Asymptotic exactness on square elements of odd degree

Suppose that the solution u of the model problem belongs to the space $H^{p+2}(\Omega)$. Let ψ_K be the solution of the local problem

Find $\psi_K \in H^1(K)/\mathbb{R}$ such that for all $v \in H^1(K)/\mathbb{R}$

$$B_K(\psi_K, v) = F_K(v) - B_K(\Pi_X u, v) + \oint_{\partial K} \left\langle \frac{\partial \Pi_X u}{\partial n_K} \right\rangle v \, ds . \tag{4.65}$$

For any v belonging to $H^1(K)/\mathbb{R}$, it then follows that

$$B_K(\psi_K - u + \Pi_X u, v) = \oint_{\partial K} \left(\left\langle \frac{\partial \Pi_X u}{\partial n_K} \right\rangle - \frac{\partial u}{\partial n_K} \right) v \, ds . \tag{4.66}$$

By the Cauchy–Schwarz Inequality:

$$\left| \oint_{\partial K} \left(\left\langle \frac{\partial \Pi_X u}{\partial n_K} \right\rangle - \frac{\partial u}{\partial n_K} \right) v \, ds \right| \leqslant \left\| \frac{\partial u}{\partial n} - \left\langle \frac{\partial \Pi_X u}{\partial n} \right\rangle \right\|_{L_2(\partial K)} \|v\|_{L_2(\partial K)} . \tag{4.67}$$

Let $\tilde{K}$ denote be the subdomain consisting of the element K and its neighbours. Using Theorem 4.7 yields

$$\left\| \frac{\partial u}{\partial n} - \left\langle \frac{\partial \Pi_X u}{\partial n} \right\rangle \right\|^2_{L_2(\partial K)} = \sum_{J \in \tilde{K}} \left\| \frac{\partial u}{\partial n} - \left\langle \frac{\partial \Pi_X u}{\partial n} \right\rangle \right\|^2_{L_2(\Gamma_{KJ})} \leqslant Ch^{2p+1} |u|^2_{H^{p+2}(\tilde{K})} . \tag{4.68}$$

Collecting these results gives for any $v \in H^1(K)/\mathbb{R}$:

$$|B_K(\psi_K - (u - \Pi_X u), v)| \leqslant Ch^{p+1/2} |u|_{H^{p+2}(\tilde{K})} \|v\|_{L_2(\partial K)} . \tag{4.69}$$

Hence,

$$|||\psi_K - (u - \Pi_X u)|||_K \leqslant C h^{p+1} |u|_{H^{p+2}(\tilde{K})}. \tag{4.70}$$

Moreover, for any v belonging to X_K:

$$B_K(\phi_K - \psi_K, v) = B_K(u_X - \Pi_X u, v) + \oint_{\partial K} \left\langle \frac{\partial}{\partial n_K} (\Pi_X u - u_X) \right\rangle v \, ds \tag{4.71}$$

and using the Cauchy–Schwarz Inequality and the Trace Theorem leads to the estimate

$$|||\phi_K - \psi_K|||_K \leqslant |||u_X - \Pi_X u|||_K + C|||u_X - \Pi_X u|||_{\tilde{K}}. \tag{4.72}$$

Finally,

$$|||e - \phi_K|||_K \leqslant |||(u - \Pi_X u) - \psi_K|||_K + |||\psi_K - \phi_K|||_K + |||\Pi_X u - u_X|||_K \tag{4.73}$$

and collecting the foregoing results, there follows:

$$\sum_{K \in \mathcal{P}} |||e - \phi_K|||_K^2 \leqslant C \left\{ |||\Pi_X u - u_X|||^2 + h^{2p+2} |u|^2_{H^{p+2}(\Omega)} \right\}. \tag{4.74}$$

The optimal rate of convergence in the energy norm that may be achieved using degree p elements is $O(h^p)$. Only in trivial cases can this rate be exceeded. If there exists a constant $C(u)$ for which the error in the energy is bounded below by $C(u)h^p$ then the error is said to be *properly* $O(h^p)$. The main result may now be stated:

THEOREM 4.8. Let $p \in \mathbf{N}$ be odd and $u \in H^{p+2}(\Omega)$. Suppose the error e measured in the energy norm is properly $O(h^p)$ and that all integrals are evaluated exactly. If the partition consists of square elements, then the error estimator η is asymptotically exact, i.e.

$$\lim_{h \to 0} \frac{\eta}{|||e|||} = 1. \tag{4.75}$$

PROOF. Follows immediately from Eq. (4.74) by using Theorem 4.6 and the assumption on e. □

Non-asymptotic exactness for elements of even order

The estimator η is not asymptotically exact when the approximation is of even order as the following counterexample shows. Consider the problem:

$$\begin{aligned} -\Delta u &= -p(p+1)x^{p-1} \quad \text{in } \Omega \\ u &= 0 \quad \text{on } \Gamma_D \\ \partial u/\partial n &= 0 \quad \text{on } \Gamma_N \end{aligned} \tag{4.76}$$

with $\Omega = (0,1) \times (0,1)$ and Γ_D being the vertical boundaries of Ω. The partition is formed by subdividing Ω into uniform squares of size h. The true solution is

$$u(x,y) = x(x^p - 1) \tag{4.77}$$

and the true error on element K may be computed explicitly [2]

$$|||e|||_K^2 = \frac{2h}{2p+1} \left(\frac{h}{2}\right)^{2p+1} \left(\frac{p+1}{k_p}\right)^2 \tag{4.78}$$

where k_p is the coefficient of the leading term in the Legendre polynomial. For the local error estimator η_K, it suffices to consider only those elements K lying on the interior of the partition, since the combined effect of elements on the boundary of the domain Ω becomes negligible as the partition is refined (provided the true solution is smooth). The estimator when the polynomial degree is even is

$$\eta_K^2 = h\left(\frac{h}{2}\right)^{2p+1}\left(\frac{p+1}{k_p}\right)^2\left\{\frac{2}{2p+1}+2\right\}. \tag{4.79}$$

Consequently, when p is even:

$$\lim_{h\to 0}\frac{\eta}{|||e|||} = \sqrt{2(p+1)}. \tag{4.80}$$

Non-asymptotic exactness on non-square elements

The estimator is not asymptotically exact when the subdomains are not square even if the approximation is of odd degree. Suppose that the domain Ω is a rhombus with angle θ. Details will be given for the case of first-order approximation of the problem:

$$-\Delta u = f \quad \text{in } \Omega \qquad u = 0 \quad \text{on } \Gamma_D \qquad \partial u/\partial n = g \quad \text{on } \Gamma_N. \tag{4.81}$$

The data f and g are chosen so that the true solution is $u(x,y) = y^2$. The domain Ω is partitioned into a mesh of $N \times N$ uniform rhombuses. The finite element approximation of this problem coincides with the interpolant of the true solution. The true error is given by

$$|||e|||_K^2 = \frac{8h}{3}\left(\frac{h}{2}\sin\theta\right)^3. \tag{4.82}$$

The solution of the error residual problem is difficult to compute exactly for this example. However, bounds can be established using variational analysis (see [2]):

$$|||\phi|||_K^2 \geqslant h\left(\frac{h}{2}\sin\theta\right)^3\left\{\frac{8}{3}+\frac{16}{15}\cos^2\theta\right\}. \tag{4.83}$$

Hence, using the expression for the true error reveals

$$\frac{|||\phi|||_K^2}{|||e|||_K^2} \geqslant 1+\frac{2}{5}\cos^2\theta. \tag{4.84}$$

Letting n_1 and n_2 denote unit outward normals on adjacent edges of any subdomain in the partition, we obtain

$$\lim_{h\to 0}\frac{\eta^2}{|||e|||^2} \geqslant 1+\frac{2}{5}|n_1\cdot n_2|^2. \tag{4.85}$$

Therefore the estimator cannot be asymptotically exact unless the normal vectors are orthogonal so that the mesh must consist of squares.

4.5.3. *Analysis and selection of approximate subspaces*

The case of finite element approximation on quadrilateral elements using piecewise polynomials of degree p was studied in the previous section. Following the analysis in [3], the effect of solving the local residual problems *approximately* is studied. At first sight, one might expect that the effect of solving approximately would only exacerbate the already tenuous situation regarding the performance of the estimators. It is perhaps surprising that this need not be the case.

Parallel meshes

Let $\mathcal{P}$ be a regular partitioning of the domain Ω and $F_K : \widehat{K} \mapsto K$ be an invertible, bilinear transformation of the reference element onto an element K. More specifically, we shall consider the class of *parallel meshes*. That is, each element K is a parallelogram with sides of length h_K, k_K making angles α_K and β_K with the coordinate axes (see Fig. 4). It is assumed that there exist positive constants C, θ such that for all $K \in \mathcal{P}$

$$\frac{1}{C} \leqslant \frac{h_K}{k_K} \leqslant C: \qquad \theta \leqslant |\beta_K - \alpha_K| \leqslant \pi - \theta \tag{4.86}$$

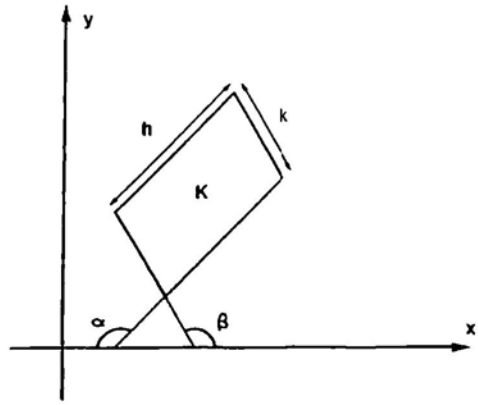


Fig. 4. Notation for parallelogram element.

and for any pair of neighbouring elements K and K' there holds

$$|\alpha_K - \alpha_{K'}| \leqslant Ch_K; \qquad |\beta_K - \beta_{K'}| \leqslant Ch_K. \tag{4.87}$$

If h is the maximum element size, then the constants C and θ should be independent of h. Strictly speaking, parallel meshes are mild distortions of the partitions described above.

Illustration of influence of approximate subspaces

The a posteriori error estimate on element K is obtained by solving a local residual problem of the form

Find $\phi_K \in Y_K$ such that for all $v \in Y_K$

$$B_K(\phi_K, v) = \int_K fv \, dx - B_K(u_X, v) + \int_{\partial K} \left\langle \frac{\partial u_X}{\partial n_K} \right\rangle v \, ds \tag{4.88}$$

where Y_K is a finite dimensional subspace of $H^1(K)$. In the previous section, it was assumed that the subspace Y_K was chosen to be $H^1(K)$. The estimator resulting from solving exactly were rather discouraging, asserting that the estimator gives a consistent error estimator only under highly restrictive conditions. The assumption that the error residual problem be solved exactly is now relaxed. At first sight, one might expect this to lead to an even less encouraging result. However, we shall see that this need not be the case, provided that the subspace is selected with some care.

The choice of subspace for the error residual problem has a significant effect on the performance of the error estimator, as the following simple illustration shows. Consider the problem

Find u such that

$$-\Delta u = f \quad \text{in } \Omega = (0,1) \times (0,1). \tag{4.89}$$

The boundary conditions and data f are chosen so that the true solution is given by

$$u(x, y) = x(x^p - 1) \tag{4.90}$$

for $p \in \mathbf{N}$. The problem is solved using elements of degree p on the meshes consisting of lines parallel to the x and y axes. The finite element approximation is identical to the interpolant. The error residual problem will be solved using various choices of subspace:

Full space

$$\hat{Y} = \hat{Q}(p+1) \backslash \mathbb{R}$$

Uniform

$$\hat{Y} = \text{span}\{\{1, \hat{x}, \ldots, \hat{x}^p\} W_{p+1}(\hat{y}), W_{p+1}(\hat{x}) \{1, \hat{y}, \ldots, \hat{y}^p\}, W_{p+1}(\hat{x}) W_{p+1}(\hat{y})\} \tag{4.91}$$

where

$$W_{p+1}(s) = \Pi_{j=0}^{p}(s - s_j)$$

and $s_j = -1 + 2j/p$, $j = 0, \ldots, p$.

Legendre

$$\hat{Y} = \text{span}\{\{1, \hat{x}, \ldots, \hat{x}^p\} P_{p+1}(\hat{y}), P_{p+1}(\hat{x}) \{1, \hat{y}, \ldots, \hat{y}^p\}, P_{p+1}(\hat{x}) P_{p+1}(\hat{y})\}$$

where P_{p+1} is the degree $p+1$ Legendre polynomial.

Lobatto

$$\hat{Y} = \text{span}\{\{1, \hat{x}, \ldots, \hat{x}^p\} L_{p+1}(\hat{y}), L_{p+1}(\hat{x}) \{1, \hat{y}, \ldots, \hat{y}^p\}, L_{p+1}(\hat{x}) L_{p+1}(\hat{y})\}$$

where $L_{p+1}(s) = (1 - s^2) P'_{p+1}(s)$.

The subspace Y_K used for the solution of the residual problem is then

$$Y_K = \{\hat{v} \circ F_K^{-1} : \hat{v} \in \hat{Y}\}. \tag{4.92}$$

In each case, the error residual problem is solved and the error estimator computed. Table 1 shows the results obtained for uniform mesh spacing ($h_K = k_K = h$ for all K). The performance of the error estimator is seen to be quite sensitive to the choice of subspace used to solve the residual problem. In some cases, the estimator is a gross over-estimate while in others estimates the error as begin zero. Moreover, increasing the dimension of the subspace does not improve the performance of the error estimator: the full space is consistent only when the polynomial degree is odd (as found in the previous section). The Lobatto basis is the only choice that is consistent in all cases.

Suppose that the mesh spacing is non-uniform. The analysis in [3] shows that the estimator based on the Lobatto subspace is consistent when the degree $p > 1$ and inconsistent when $p = 1$. The results of using estimators based on the other spaces are inconsistent for all values of p.

It is worth observing that these results have been obtained for a very simple model problem. The purpose is to illustrate that even in such a simplified setting the error estimator can give misleading results inconsistent with the actual error unless the subspace used to approximate the residual problem is chosen carefully. Even so, one finds that the estimator can still be inconsistent if the mesh is only mildly non-uniform. In practical computations the true solution may be singular and the mesh highly irregular. It is only to be expected that further problems in the performance will arise.

It can be shown [3] that the behaviour illustrated in the examples is true generally whenever the mesh is parallel and the true solution smooth. Equally important is to understand the mechanism leading to these somewhat surprising phenomena. Henceforth, it will be assumed that the Lobatto subspace is used.

Accuracy of boundary flux functionals

The analysis in the previous section showed that the boundary flux approximation plays a key role in the performance of the error estimator. Suppose elements K and K' share a common edge γ. The performance of the error estimator depends crucially on how well the the boundary flux functional $I(\cdot)$

Table 1
Performance of error estimators on uniform mesh

Degree	True error	Square of error estimator			
p	$\|\|\|e\|\|\|_K^2$	Full	Uniform	Legendre	Lobatto
1	$\frac{1}{3}h^4$	$\frac{1}{3}h^4$	$\frac{1}{3}h^4$	$\frac{1}{3}h^4$	$\frac{1}{3}h^4$
2	$\frac{1}{20}h^6$	$\frac{3}{10}h^6$	$\frac{1}{20}h^6$	0	$\frac{1}{20}h^6$
3	$\frac{1}{175}h^8$	$\frac{1}{175}h^8$	$\frac{3565126227}{6400003750000}h^8$	$\frac{1}{250}h^8$	$\frac{1}{175}h^8$
4	$\frac{1}{1764}h^{10}$	$\frac{5}{882}h^{10}$	$\frac{64}{132741}h^{10}$	0	$\frac{1}{1764}h^{10}$

defined for fixed $v \in Y_K$ by

$$\Gamma(u) = \int_\gamma \left\langle \frac{\partial u}{\partial n_K} \right\rangle v \, ds \tag{4.93}$$

is approximated by $\Gamma(u_X)$ where u_X is the finite element approximation to u. The accuracy of the boundary flux functionals plays a key role in the proof of:

THEOREM 4.9. Suppose that the mesh is parallel and the true solution u belongs to $H^{p+2}(\Omega)$. If the error is properly $O(h^p)$ and $p > 1$, then the error estimator obtained using the Lobatto basis is asymptotically exact:

$$\lim_{h \to 0} \frac{\eta}{|||e|||} = 1. \tag{4.94}$$

PROOF. See [3]. □

4.5.4. Conclusions

The error residual method for a posteriori error estimation is based on solving local residual problems for the error. It has been seen that the estimator is quite sensitive to the choice of subspace used to solve the problem. In particular, using a full space of polynomials is inappropriate since the resulting estimator may give results completely inconsistent with the actual error. By considering certain subspaces of polynomials the estimator can be improved. Theorem 4.9 shows that if the mesh is parallel and the true solution smooth, then the resulting estimator is consistent with the actual error for degree $p > 1$ finite element approximation. However, the estimator is still inconsistent on non-uniform meshes when $p = 1$.

The results can be explained as follows. The error in the finite element approximation can be thought of as consisting of a number of components. The true solution of the residual problem can also be regarded as comprising of a number of components, some of which correspond to actual modes in the true error and other *spurious modes* that arise from the formulation of the problem using inexact boundary data. When a full space is used to approximate the problem, all of these components are present in the approximation. The resulting estimator is inconsistent owing to the contributions from the spurious modes. However, when a subspace Y_K is used to approximate the residual problem, then the components orthogonal to Y_K are not present in the approximation. If the subspace can be chosen so that it is precisely the spurious modes that are orthogonal to the space, then the consistency of the estimator will be recovered.

It is not surprising then, that the estimators are so sensitive to the choice of subspace. The results obtained using the subspaces in the illustration can be interpreted in the light of this explanation as follows. First, the subspace based on Uniform Nodes (4.91) is inappropriate since it still contains some spurious modes when the degree p exceeds two, leading to over-estimates of the error. On the opposite extreme, the subspace based on Legendre Nodes is inappropriate since it is not only orthogonal to the spurious modes but also to modes that represent the actual error, leading to gross under-estimates of the error. The subspace based on Lobatto Nodes is orthogonal to spurious modes but at the same time provides sufficient resolution of the true modes (whenever $p > 1$). An alternative possibility is to construct the boundary data for the underlying problem differently, for instance, using the equilibration procedures discussed in the next chapter.

5. The equilibrated residual method

5.1. Introduction

The advantages of using implicit error estimators have been outlined in Section 4. The element residual method requires the solution of local boundary value problem. However, in many ways the basic for-

mulation of the element residual problem is somewhat unsatisfactory. For instance, the local Neumann problems that must be approximated may not have a solution. This 'difficulty' is avoided by solving using a quotient space with the null space factored out. In Section 4, the estimators were shown to perform extremely well in certain circumstances. However, the investigations revealed other deficiencies in the basic formulation of the problem. In particular, the construction of the boundary data for the local problem is somewhat ad hoc.

An alternative criterion for choosing boundary data is to require that the local problem is well posed. This requires that the boundary data be in equilibrium with the interior residual. The *equilibration condition* was first used in a posteriori error analysis by Ladeveze and Leguillon [43], who wished to solve a local dual problem for the error. The element residual method with equilibrated data was first used by Bank and Weiser [23] who, on the basis of numerical experience, conjectured that the resulting estimator gives an upper bound for the error. The method was analyzed and generalized by Ainsworth and Oden [5] which the current exposition follows closely. One by-product of [5] is a proof that Bank and Weiser's conjecture is correct.

Consider the usual model problem described in Section 1.5. Suppose that $X \subset V$ is a finite element subspace constructed on a non-degenerate partition $\mathcal{P}$. The finite element approximation of this problem is to find $u_X \in X$ such that

$$B(u_X, v_X) = L(v_X) \quad \forall v_X \in X. \tag{5.1}$$

The error $e = u - u_X$ belongs to the space V and satisfies

$$B(e, v) = B(u, v) - B(u_X, v) = L(v) - B(u_X, v) \quad \forall v \in V. \tag{5.2}$$

5.2. A posteriori error analysis

5.2.1. Mesh dependent forms and spaces

It will be convenient to reduce the global spaces and forms into sums of contributions from each of the elements in the partition $\mathcal{P}$. With this in mind, define the broken Sobolev spaces for $m \in \mathbb{Z}^+$

$$H^m(\mathcal{P}) = \{v \in L_2(\Omega) : v|_K \in H^m(K) \ \forall K \in \mathcal{P}\}. \tag{5.3}$$

Here, and in what follows, v_K denotes the restriction of v to a single element K. The associated mesh dependent norm is

$$\|v\|_{m,\mathcal{P}} = \left\{ \sum_{K \in \mathcal{P}} \|v_K\|^2_{m,K} \right\}^{1/2}. \tag{5.4}$$

For each element $K \in \mathcal{P}$, let

$$V_K = \{v \in H^1(K) : v = 0 \text{ on } \Gamma_D \cap \partial K\} \tag{5.5}$$

and introduce the bilinear form $B_K : V_K \times V_K \to \mathbb{R}$

$$B_K(u, v) = \int_K (\nabla u \cdot \nabla v + cuv) \, d\mathbf{x}. \tag{5.6}$$

Similarly, $F_K : V_K \to \mathbb{R}$ is defined by

$$F_K(v) = \int_K fv \, d\mathbf{x}. \tag{5.7}$$

Hence for $v, w \in V$

$$B(v, w) = \sum_{K \in \mathcal{P}} B_K(v_K, w_K) \tag{5.8}$$

and

$$L(v) = \sum_{K \in \mathcal{P}} F_K(v_K) + \int_{\Gamma_N} g(s)v(s) \, ds. \tag{5.9}$$

The inner products on the Sobolev spaces are decomposed as sums of contributions from each element in the partition in an analogous fashion. The broken version of the space V is defined by

$$V(\mathcal{P}) = \{v \in L_2(\Omega) : v_K \in V_K \ \forall K \in \mathcal{P}\} . \tag{5.10}$$

Later, we shall wish to consider the space of continuous linear functionals τ on $V(\mathcal{P})$ which vanish on the subspace V. Therefore, let $H(\mathbf{div}, \Omega)$ denote the space

$$H(\mathbf{div}, \Omega) = \{\boldsymbol{A} \in [L_2(\Omega)]^2 : \mathbf{div}\, \boldsymbol{A} \in L_2(\Omega)\} \tag{5.11}$$

equipped with norm

$$\|\boldsymbol{A}\|_{H(\mathbf{div},\Omega)} = \{\|\boldsymbol{A}\|^2_{0,\Omega} + \|\mathbf{div}\, \boldsymbol{A}\|^2_{0,\Omega}\}^{1/2}. \tag{5.12}$$

Let $\mathcal{M}$ denote the subspace

$$\mathcal{M} = \left\{\boldsymbol{A} \in H(\mathbf{div}, \Omega) : \oint_{\partial\Omega} v\, \boldsymbol{n} \cdot \boldsymbol{A}\, ds = 0 \ \forall v \in V\right\}. \tag{5.13}$$

The following result is originally due to Raviart and Thomas [51]:

THEOREM 5.1. A continuous linear functional τ on the space $V(\mathcal{P})$ vanishes on the subspace V if and only if there exists $\boldsymbol{A} \in \mathcal{M}$ such that

$$\tau[v] = \sum_{K\in\mathcal{P}} \oint_{\partial K} v_K\, \boldsymbol{n}_K \cdot \boldsymbol{A}\, ds \tag{5.14}$$

where $\boldsymbol{n}_K$ denotes the unit outward normal on the boundary of K.

PROOF. Suppose $\tau \in V(\mathcal{P})'$ vanishes on V. By Riesz' Theorem, any continuous functional on the space $H^1(K)$ may be written in the form

$$v \to \int_K \left\{\sum_{j=1}^{2} A_j \frac{\partial v}{\partial x_j} + a_0 v\right\} d\boldsymbol{x} \tag{5.15}$$

where A_j and $a_0 \in L_2(K)$. Therefore, summing over the elements shows that for any $v \in V(\mathcal{P})$, τ may be written in the form

$$\tau[v] = \sum_{K\in\mathcal{P}} \int_K \left\{\sum_{j=1}^{2} A_j \frac{\partial v}{\partial x_j} + a_0 v\right\} d\boldsymbol{x} \tag{5.16}$$

where A_j and a_0 now denote elements of the global space $L_2(\Omega)$. Owing to the hypothesis on τ it follows that for any $v \in V$

$$0 = \int_\Omega \left\{\sum_{j=1}^{2} A_j \frac{\partial v}{\partial x_j} + a_0 v\right\} d\boldsymbol{x}. \tag{5.17}$$

Hence, in the sense of distributions

$$-\frac{\partial A_j}{\partial x_j} + a_0 = 0 \tag{5.18}$$

and consequently (A_1, A_2) belongs to the space $H(\mathbf{div}, \Omega)$. With the aid of (5.16), (5.18) and Green's Identity

$$\tau[v] = \sum_{K\in\mathcal{P}} \int_K \sum_{j=1}^{2} \left\{A_j \frac{\partial v}{\partial x_j} + \frac{\partial A_j}{\partial x_j} v\right\} d\boldsymbol{x} = \sum_{K\in\mathcal{P}} \oint_{\partial K} v_K\, \boldsymbol{n}_K \cdot \boldsymbol{A}\, ds. \tag{5.19}$$

Finally, $\boldsymbol{A} \in \mathcal{M}$ since for any $v \in V$

$$0 = \tau[v] = \sum_{K\in\mathcal{P}} \oint_{\partial K} v_K\, n_K \cdot \boldsymbol{A}\, ds = \oint_{\partial\Omega} v\, n \cdot \boldsymbol{A}\, ds\,. \tag{5.20}$$

The converse is shown using similar arguments. □

The import of this result is that one may identify τ with an element $\boldsymbol{A}$ of the space $\mathcal{M}$ and vice versa. In view of Theorem 5.1, we shall abuse the nomenclature slightly and refer to an element of $\mathcal{M}$ as being a linear functional.

5.2.2. Preliminaries

The residual equation (5.2) characterizing the error may be stated equivalently as a minimization problem

$$\min J(v) : v \in V \tag{5.21}$$

where $J : V \to \mathbb{R}$ is the quadratic functional

$$J(v) = \frac{1}{2} B(v,v) - L(v) + B(u_X, v). \tag{5.22}$$

The error e in the finite element approximation is the minimizer of J

$$-\frac{1}{2} B(e,e) = J(e) \leqslant J(v) \quad \forall v \in V \tag{5.23}$$

where (5.2) has been used.

In principle, one could approximate the problem (5.23) directly and thereby obtain a computable bound on the error. However, the cost associated with solving a global problem renders this approach impractical. However, an alternative approach is to attempt to reduce the single global problem (5.23) into a sequence of independent problems posed locally over each element. The advantage would be that each of these smaller problems might then be approximated comparatively inexpensively and even in parallel.

The theme throughout the rest of this section is to recast the global statement (5.23) as a sequence of independent local problems posed on each element $K \in \mathcal{P}$. However, rather than simply decompose the problem indiscriminately and risk losing the valuable bound on the actual discretization error, we shall proceed in a way whereby the relationship is preserved.

The approximation to the true flux on the inter-element boundaries played an important role when we considered the classical element residual method. There a simple averaging of the finite element approximations to the flux from the neighbouring elements on an edge was used. Approximations to the true flux will again be of importance, but at this stage we shall proceed more generally as follows:

Approximation of fluxes on element boundaries

Let $\partial\mathcal{P}$ denote the element edges in the partition. Suppose that we order the elements in some way. For instance, one could order elements according to their global numbers in the finite element code. Let $\sigma_K : \partial K \to \{+1, -1\}$ be the piecewise constant function on the edges of element K defined by

$$\sigma_K(s) = \begin{cases} +1 & s \in \overline{K} \cap \overline{J},\ K > J \\ -1 & s \in \overline{K} \cap \overline{J},\ K < J \\ +1 & s \in \partial\Omega \end{cases}. \tag{5.24}$$

Notice that if elements K and J share a common edge then

$$\sigma_K(s) = -\sigma_J(s) \quad s \in \overline{K} \cap \overline{J}. \tag{5.25}$$

With each element interface γ we associate a smooth function $g_\gamma : \gamma \to \mathbb{R}$. If γ lies on the portion of the boundary Γ_N on which a Neumann boundary condition is prescribed then we shall always choose g_γ to be the Neumann data g on the edge. It will be unnecessary to define g_γ should γ lie on the Dirichlet boundary Γ_D. Specific constructions for the functions g_γ on the interior interfaces will be discussed later.

The approximation to the flux is then defined by

$$g_K = \sigma_K g_\gamma \quad \text{on } \gamma \subset \partial K. \tag{5.26}$$

There is no danger of confusion with the Neumann data g using this notation since the functions g_K agree with the boundary data on Γ_N. The notation $[\cdot]$ is used to denote differences in values of functions v from the broken space $V(\mathcal{P})$ across element boundaries

$$[v] = \begin{cases} \sigma_K(v_K - v_J) & \text{on } \gamma = \overline{K} \cap \overline{J} \\ v & \text{on } \gamma \subset \partial\Omega \end{cases} \tag{5.27}$$

and $\boldsymbol{n}$ is defined by

$$\boldsymbol{n} = \sigma_K \boldsymbol{n}_K \quad \text{on } \gamma \subset \partial K. \tag{5.28}$$

These definitions are unambiguous since on any edge $\overline{K} \cap \overline{J}$ one has

$$\sigma_K(v_K - v_J) = \sigma_J(v_J - v_K) \tag{5.29}$$

and

$$\sigma_K \boldsymbol{n}_K = \sigma_J \boldsymbol{n}_J. \tag{5.30}$$

The following identity valid for $v \in V(\mathcal{P})$ is readily obtained:

$$\sum_{K\in\mathcal{P}} \oint_{\partial K} g_K v \, ds = \sum_{\gamma\in\partial\mathcal{P}} \int_\gamma g_\gamma [v] \, ds. \tag{5.31}$$

5.2.3. Localization

The process of decomposing the global problem (5.23) into smaller, local problems posed over the elements can now be discussed precisely. Two basic steps are involved in breaking up the problem:

- Decomposition of the quadratic functional J into separate contributions from each element.
- Localization of the global space V. The essence is to decompose the space V of globally smooth functions into functions belonging to the broken space $V(\mathcal{P})$. These functions need not be continuous across the interelement boundaries and allow one to deal with a series of problems, on each of the elements independently, thereby substantially reducing the complexity of the problem.

The chief difficulty arises in the second step. If one were to simply substitute the space V with the broken space $V(\mathcal{P})$ then the key relationship (5.23) with the true error would be irretrievably lost.

We begin by extending the functional given by

$$V \ni w \to L(w) - B(u_X, w) \tag{5.32}$$

to the broken space $V(\mathcal{P})$. For any $w \in V(\mathcal{P})$ define the linear functional $\mathcal{R} : V(\mathcal{P}) \to \mathbb{R}$ by

$$\mathcal{R}(w) = \sum_{K\in\mathcal{P}} \left\{ F_K(w) - B_K(u_X, w) + \oint_{\partial K} g_K w_K \, ds \right\} - \sum_{\gamma\in\partial\mathcal{P}} \int_\gamma g_\gamma [w] \, ds. \tag{5.33}$$

Notice that, thanks to (5.31), whenever $w \in V$

$$\mathcal{R}(w) = L(w) - B(u_X, w). \tag{5.34}$$

so that $\mathcal{R}$ is an extension of the functional (5.32) to the whole of $V(\mathcal{P})$.

LEMMA 5.2. Under the above notations and conventions, there exists $\mu_* \in \mathcal{M}$ such that for all $w \in V(\mathcal{P})$

$$\mu_*(w) = \sum_{\gamma\in\partial\mathcal{P}} \int_\gamma g_\gamma [w] \, ds. \tag{5.35}$$

PROOF. The right-hand side of Eq. (5.35) is continuous on $V(\mathcal{P})$ and vanishes on V. Applying Theorem 5.1, the result follows immediately. □

Applying Lemma 5.2 yields the identity

$$\mathcal{R}(w) = \sum_{K\in\mathcal{P}} \left\{ F_K(w) - B_K(u_X, w) + \oint_{\partial K} g_K w_K \, ds \right\} - \mu_*(w) \tag{5.36}$$

valid for all $w \in V(\mathcal{P})$.

5.2.4. Variational analysis

The requirement that v belong to the space V in (5.23) may be regarded as a constraint on the interelement continuity of the function v. Recall that the aim is to relax this constraint and yet retain the bound on the error. Therefore, we follow the standard approach of introducing a Lagrange multiplier to enforce the constraint indirectly. Let the Lagrangian functional $\mathcal{L} : V(\mathcal{P}) \times \mathcal{M} \to \mathbb{R}$ be defined by

$$\mathcal{L}(w, \mu) = \frac{1}{2} B(w, w) - \mathcal{R}(w) - \mu(w) \tag{5.37}$$

and note that

$$\sup_{\mu\in\mathcal{M}} \mathcal{L}(w, \mu) = \begin{cases} \frac{1}{2} B(w, w) - \mathcal{R}(w) & \text{if } w \in V \\ +\infty & \text{otherwise} \end{cases} \tag{5.38}$$

Moreover, for $w \in V$, (5.34) and (5.2) reveal

$$\begin{aligned} \frac{1}{2} B(w, w) - \mathcal{R}(w) &= \frac{1}{2} \{B(w - e, w - e) - B(e, e)\} \\ &\geqslant -\frac{1}{2} B(e, e) = -\frac{1}{2} |||e|||^2. \end{aligned} \tag{5.39}$$

Therefore,

$$-\frac{1}{2} |||e|||^2 = \inf_{w\in V(\mathcal{P})} \sup_{\mu\in\mathcal{M}} \mathcal{L}(w, \mu) = \sup_{\mu\in\mathcal{M}} \inf_{w\in V(\mathcal{P})} \mathcal{L}(w, \mu). \tag{5.40}$$

The interchange in the order of the inf–sup is justified here since a saddle point is obtained when the multiplier μ is the true interelement flux. This choice is a valid multiplier as can be seen by applying Theorem 5.1. Eq. (5.40) immediately gives the bound

$$-\frac{1}{2} |||e|||^2 \geqslant \inf_{w\in V(\mathcal{P})} \mathcal{L}(w, \mu) \tag{5.41}$$

which is valid *for any* $\mu \in \mathcal{M}$. Moreover,

$$\mathcal{L}(w, \mu) = \sum_{K\in\mathcal{P}} \inf_{w_K\in V_K} \left\{ \frac{1}{2} B(w_K, w_K) - F_K(w) + B_K(u_X, w) - \oint_{\partial K} g_K w_K \, ds \right\} + \mu_*(w) - \mu(w) \tag{5.42}$$

and so

$$-\frac{1}{2} |||e|||^2 \geqslant \sum_{K\in\mathcal{P}} \inf_{w_K\in V_K} J_K(w) + \mu_*(w) - \mu(w) \tag{5.43}$$

where

$$J_K(w) = \frac{1}{2} B_K(w, w) - F_K(w) + B_K(u_X, w) - \oint_{\partial K} g_K w_K \, ds. \tag{5.44}$$

Studying the estimate (5.43) one sees that the space V has been replaced by the broken space $V(\mathcal{P})$ while preserving the bound on the error. Thus, the continuity requirements on the choice of admissable

functions imposed by the space V have been relaxed. However, the statement (5.44) has not yet been decomposed into *independent* contributions from each element. The contribution from Lagrange multiplier term $\mu_*(w) - \mu(w)$ preserves the bound on the error. In fact, examining Lemma 5.2 reveals that the interelement continuity requirement is being imposed indirectly through the Lagrange multiplier acting on the jumps $[w]$ on the admissable functions. A key point now emerges: this coupling between elements can be removed by choosing the Lagrange multiplier μ to be equal to μ_*, while simultaneously retaining the bound on the error.

Summarising, we have shown:

THEOREM 5.3. Let $J_K : V_K \to \mathbb{R}$ be the quadratic functional

$$J_K(w) = \frac{1}{2} B_K(w, w) - F_K(w) + B_K(u_X, w) - \oint_{\partial K} g_K w_K \, ds. \tag{5.45}$$

Then,

$$|||e|||^2 \leqslant -2 \sum_{K \in \mathcal{P}} \inf_{w_K \in V_K} J_K(w_K). \tag{5.46}$$

5.3. The equilibration principle

To begin with, assume that the coefficient c appearing in the differential operator is strictly greater than zero. Later, this assumption will be removed. Theorem 5.3 leads us to consider a sequence of local minimization problems posed on each element K

$$\inf_{v \in V_K} J_K(v) \tag{5.47}$$

where J_K is the quadratic functional defined in equation (5.44). For the present purposes it will be convenient to reorganize the terms appearing in the functional J_K. Applying Green's identity gives

$$F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds = \int_K rv \, dx + \int_{\partial K} R_* v \, ds \tag{5.48}$$

where r is the interior residual

$$r = f + \Delta u_X - c u_X \tag{5.49}$$

and R_* is a modified form of the boundary residual

$$R_* = g_K - n_K \cdot \nabla u_X|_K. \tag{5.50}$$

THEOREM 5.4. Let

$$W_K = \{\boldsymbol{p} \in H(\mathbf{div}, K) : n_K \cdot \boldsymbol{p} = R_* \text{ on } \partial K\} \tag{5.51}$$

and define $G_K : W_K \to \mathbb{R}$ by

$$G_K(\boldsymbol{p}) = -\frac{1}{2} \int_K \boldsymbol{p} \cdot \boldsymbol{p} \, dx - \frac{1}{2c} \int_K (\nabla \cdot \boldsymbol{p} + r)^2 \, dx. \tag{5.52}$$

Let ϕ_K be the solution of the problem

$$B_K(\phi_K, v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds \quad \forall v \in V_K. \tag{5.53}$$

Then $\nabla \phi_K \in W_K$ and

$$\inf_{v \in V_K} J_K(v) = J_K(\phi_K) = -\frac{1}{2} |||\phi_K|||_K^2 = G_K(\nabla \phi_K) = \sup_{p \in W_K} G_K(\boldsymbol{p}). \tag{5.54}$$

PROOF. The existence and uniqueness of ϕ_K is a consequence of the Lax–Milgram Theorem. Problem (5.53) means that in the sense of distributions

$$-\mathbf{div}(\nabla\phi_K) + c\phi_K = r \tag{5.55}$$

and hence, since r is smooth on the interior of element K and ϕ_K belongs to $L_2(K)$, one sees that $\nabla\phi_K \in H(\mathbf{div}, K)$. Furthermore, again in a distributional sense, we have

$$\boldsymbol{n} \cdot \nabla\phi_K = R_e \tag{5.56}$$

and hence $\nabla\phi_K$ belongs to W_K.

Eq. (5.53) is the Euler equation for the functional J_K and consequently ϕ_K is a minimizer of J_K, and

$$J_K(\phi_K) = -\frac{1}{2}\,|||\phi_K|||^2. \tag{5.57}$$

Furthermore,

$$G_K(\nabla\phi_K) = -\frac{1}{2}\int_K |\nabla\phi_K|^2\,\mathrm{d}\boldsymbol{x} - \frac{1}{2}\int_K \frac{1}{c}\,(cv_K^*)^2\,\mathrm{d}\boldsymbol{x} = -\frac{1}{2}\,|||\phi_K|||^2. \tag{5.58}$$

Finally, since G_K is strictly concave and quadratic, it suffices to show G_K is stationary when $\nabla\phi_K$. Let

$$\boldsymbol{q} \in \{\boldsymbol{p} \in H(\mathbf{div}, K) : \boldsymbol{n} \cdot \boldsymbol{p} = 0\} \tag{5.59}$$

and let $\lambda \in \mathbb{R}$. Then $\nabla\phi_K + \lambda\boldsymbol{q} \in W_K$ and

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}\lambda} G_K\left(\nabla\phi_K + \lambda\boldsymbol{q}\right)|_{\lambda=0} &= -\int_K \boldsymbol{q}\cdot\nabla\phi_K\,\mathrm{d}\boldsymbol{x} - \int_K \frac{1}{c}(\mathbf{div}\,\boldsymbol{q})(\mathbf{div}\,\nabla\phi_K + r_K)\,\mathrm{d}\boldsymbol{x}\\ &= -\int_K \boldsymbol{q}\cdot\nabla\phi_K\,\mathrm{d}\boldsymbol{x} - \int_K (\mathbf{div}\,\boldsymbol{q})\phi_K\,\mathrm{d}\boldsymbol{x}\\ &= -\oint_{\partial K} \phi_K(\boldsymbol{n}_K\cdot\boldsymbol{q})\,\mathrm{d}s\end{aligned} \tag{5.60}$$

and this vanishes owing to (5.59). □

Using Theorems 5.3 and 5.4 one obtains

COROLLARY 5.5. Let ϕ_K and G_K be as above. Then

$$|||e|||^2 \leqslant -2\sum_{K\in\mathcal{P}} G_K(\boldsymbol{p}_K) \tag{5.61}$$

for any choices of $\boldsymbol{p} \in \Pi_{K\in\mathcal{P}} W_K$.

The implication of this result on the computation of error bounds is that by simply constructing elements of the sets W_K, one can obtain rigorous upper bounds on the error. Of course, care must be exercised in choosing $\boldsymbol{p}_K$ if one is to obtain realistic bounds.

Suppose $\boldsymbol{p} \in W_K$ then and define δ_K to be

$$\delta_K = \int_K (\nabla\cdot\boldsymbol{p} + r)\,\mathrm{d}\boldsymbol{x} \tag{5.62}$$

then

$$\begin{aligned}\delta_K &= \oint_{\partial K} \boldsymbol{n}_K\cdot\boldsymbol{p}\,\mathrm{d}s + \int_K r\,\mathrm{d}\boldsymbol{x}\\ &= \oint_{\partial K} R_e\,\mathrm{d}s + \int_K r\,\mathrm{d}\boldsymbol{x}\end{aligned} \tag{5.63}$$

or alternatively

$$\delta_K = F_K(1) - B_K(u_X, 1) + \oint_{\partial K} g_K\,\mathrm{d}s. \tag{5.64}$$

Observe that δ_K is independent of the particular choice of $\boldsymbol{p}$. Now, with the second term of the functional G_K in mind, we note that by the Cauchy–Schwarz Inequality

$$\delta_K^2 \leqslant \operatorname{meas}(K) \int_K (\nabla \cdot \boldsymbol{p} + r)^2 \, \mathrm{d}\boldsymbol{x} \tag{5.65}$$

and hence for *any* $\boldsymbol{p} \in W_K$,

$$\int_K \frac{1}{c} (\nabla \cdot \boldsymbol{p} + r)^2 \, \mathrm{d}\boldsymbol{x} \geqslant \delta_K^2 / c \operatorname{meas}(K). \tag{5.66}$$

This estimate shows that as we refine the partition so that meas(K) tends to zero then the bound proclaimed by Corollary 5.5 will become trivial, unless the quantity δ_K vanishes too. The value of δ_K is not something which may be altered by our choice of $\boldsymbol{p}$, since it is a generic property arising from the definition of the space W_K itself. Essentially, δ_k represents a third type of residual in addition to the interior residual r and the boundary residual $R_.$.

The means by which we can control the quantity δ_K is provided by the functions g_γ that also determine the boundary conditions on the local error residual problem (5.53). Consequently, a natural strategy is to attempt select these functions in such a way that the quantity δ_K vanishes. This procedure is referred to as *equilibration*.

So far, we have assumed that the parameter c in the differential operator was strictly positive. Much of the foregoing analysis cannot be directly applied when c vanishes. A natural question to ask is whether the equilibration criterion is still relevant. Inserting the function $v \equiv \lambda$, where λ is an arbitrary real number, into the original quadratic functional J_K gives

$$J_K(\lambda) = F_K(\lambda) - B_K(u_X, \lambda) + \oint_{\partial K} g_K \lambda \, \mathrm{d}s = \lambda \delta_K. \tag{5.67}$$

Consequently, unless δ_K vanishes, the value of the functional J_K can be made as negative as we wish by choosing λ appropriately. Therefore, if Theorem 5.3 is to yield useful information, it necessary that the data be equilibrated. If the data is equilibrated then Theorem 5.4 may extended to include the case $c = 0$:

THEOREM 5.6. Suppose that $c = 0$ and let

$$W_K = \{\boldsymbol{p} \in H(\mathbf{div}, K) : \mathbf{div}\, \boldsymbol{p} = r \text{ on } K \text{ and } \boldsymbol{n}_K \cdot \boldsymbol{p} = R_. \text{ on } \partial K\} \tag{5.68}$$

and define $G_K : W_K \to \mathbb{R}$ by

$$G_K(\boldsymbol{p}) = -\frac{1}{2} \int_K \boldsymbol{p} \cdot \boldsymbol{p} \, \mathrm{d}\boldsymbol{x}. \tag{5.69}$$

Let ϕ_K be the solution of the problem

$$B_K(\phi_K, v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, \mathrm{d}s \quad \forall v \in V_K. \tag{5.70}$$

Then $\nabla \phi_K \in W_K$ and

$$\inf_{v \in V_K} J_K(v) = J_K(\phi_K) = -\frac{1}{2} |||\phi_K|||_K^2 = G_K(\nabla \phi_K) = \sup_{\boldsymbol{p} \in W_K} G_K(\boldsymbol{p}). \tag{5.71}$$

The proof is virtually identical to Theorem 5.4. The requirement that the data be equilibrated is necessary if the set W_K is to be non-empty.

When $c = 0$ the local minimization problem is the variational form of the pure Neumann problem

$$-\Delta \phi_K = r \quad \text{in } K \tag{5.72}$$

subject to the boundary conditions $\partial \phi_K / \partial n = R_.$ on the whole of ∂K. The equilibration principle has a natural physical interpretation that the interior load r is in equilibrium with the boundary forces $R_.$.

5.4. Construction of equilibrated fluxes

The discussion above indicates the advantages of constructing approximations to the interelement fluxes in such a way that the data for each of the element residual problems is equilibrated:

$$0 = F_K(1) - B_K(u_X, 1) + \int_{\partial K} g_K \, ds. \tag{5.73}$$

The equilibration condition cannot be satisfied for each element independently of the remaining elements since the flux approximations g_K and g_J on neighbouring elements K and J are related by the condition

$$g_K = \sigma_K g_\gamma = -\sigma_J g_\gamma = -g_J. \tag{5.74}$$

In effect, while the localization arguments from earlier have decoupled the element residual problems, coupling still persists implicitly through the boundary conditions if we demand that the data be equilibrated.

There are now two issues to be resolved:

- is it possible to satisfy the equilibration conditions at all?
- if so, can this be achieved by performing independent local computations?

It is essential to reduce the computation of equilibrated fluxes to independent local computations so that the resulting error estimator is economical to compute. Perhaps surprisingly, we shall find that the answer to both of these questions is affirmative. Moreover, the fluxes are not uniquely determined and there is more than one procedure whereby equilibrated fluxes may be computed locally.

The basic idea used to obtain local problems is to introduce a partition of unity as follows. Let Ψ denote the set of element vertices in the partition $\mathcal{P}$ and for $n \in \Psi$ let θ_n denote the first order Lagrange basis function based at the vertex x_n. That is, θ_n is piecewise linear (or bilinear) on the partition and satisfies

$$\theta_n(x_m) = \delta_{mn} \tag{5.75}$$

where δ_{mn} is the Kronecker symbol. For each element $K \in \mathcal{P}$, let $\Psi(K) \subset \Psi$ consist of the vertices of element K. Notice that

$$\sum_{n \in \Psi(K)} \theta_n(x) \equiv 1. \quad x \in \overline{K}. \tag{5.76}$$

Finally, let $\tilde{\Omega}_n$ denote the elements having a vertex at x_n.

We describe two basic procedures for obtaining equilibrated data: Ladeveze's Method [43] and Flux Splitting [5.6].

5.4.1. Ladeveze's method

Ladeveze and Leguillon [43] (see also [23]) construct equilibrated fluxes by defining the flux approximation on the edge γ between elements K and J to be

$$g_K = \left\langle \frac{\partial u_X}{\partial n_K} \right\rangle + \sigma_K \beta_\gamma \tag{5.77}$$

where

$$\left\langle \frac{\partial u_X}{\partial n_K} \right\rangle = \begin{cases} \frac{1}{2} n_K \cdot \{\nabla u_X|_K + \nabla u_X|_J\}, & \text{on } \overline{K} \cap \overline{J} \\ g, & \text{on } \overline{K} \cap \Gamma_N \end{cases} \tag{5.78}$$

and $\beta_\gamma : \gamma \to \mathbb{R}$ is a linear function on each edge and is identically zero on Γ_N. It is easily seen that this choice satisfies our earlier conditions on g_K. Substituting the definition (5.77) into the equilibration condition (5.73) and rewriting gives

$$-\sum_{\gamma \subset \partial K} \sigma_K \int_\gamma \beta_\gamma \, ds = \delta_K(1) \tag{5.79}$$

where

$$\delta_K(v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} \left\langle \frac{\partial u_X}{\partial n_K} \right\rangle v \, ds. \tag{5.80}$$

The equilibration condition (5.79) is localized by inserting the sum (5.76) in place of unity giving the following condition for each element K

$$-\sum_{n\in\Psi(K)}\sum_{\gamma\subset\partial K}\sigma_K\int_\gamma\theta_n\beta_\gamma\,ds=\sum_{n\in\Psi(K)}\delta_K(\theta_n). \tag{5.81}$$

Let ψ_n, $n\in\Psi$ be piecewise linear functions defined on the edges. The functions are constructed so that on any edge γ with an endpoint at x_m there holds

$$\int_\gamma\psi_n(s)\theta_m(s)\,ds=\delta_{mn} \tag{5.82}$$

where δ_{nm} is again the Kronecker symbol. A straightforward computation shows that on an edge γ with endpoints x_m and x_n, one has

$$\psi_n(s)=\frac{2}{|\gamma|}(2\phi_n(s)-\phi_m(s)) \tag{5.83}$$

where $|\gamma|$ is the length of γ. On this edge the piecewise linear function β_γ may be written in terms of the functions ψ_m and ψ_n

$$\beta_\gamma(s)=\beta_\gamma^m\psi_m(s)+\beta_\gamma^n\psi_n(s) \tag{5.84}$$

where β_γ^m and β_γ^n are constants to be determined. Inserting this definition into condition (5.81) and simplifying gives the condition that for each element K

$$-\sum_{n\in\Psi(K)}\sum_{\gamma\subset\partial K}\sigma_K\beta_\gamma^n=\sum_{n\in\Psi(K)}\delta_K(\theta_n). \tag{5.85}$$

A sufficient condition for (5.85) to hold is

$$-\sum_{\gamma\subset\partial K}\sigma_K\beta_\gamma^n=\delta_K(\theta_n)\quad\forall n\in\Psi(K) \tag{5.86}$$

which may in turn be reformulated: for each $n\in\Psi$

$$-\sum_{\gamma\subset\partial K}\sigma_K\beta_\gamma^n=\delta_K(\theta_n)\quad\forall K\in\tilde{\Omega}_n. \tag{5.87}$$

The difference is that before we sought to satisfy the condition for each element by solving over the nodes of the element. Now, we seek to satisfy the condition for each node by solving over the elements containing the node. Later, we shall show that there exist constants β_γ^n such that the conditions (5.87) are satisfied. Once these constants have been ascertained, one then reconstructs the flux approximation from the definitions (5.77) and (5.83).

5.4.2. *Flux splitting*

An alternative approach to constructing equilibrated fluxes was developed in [5.6]. The basic approach stems from the construction of the flux approximation employed in the classical element residual method. There, a simple averaging of the finite element approximations to the flux from the neighbouring elements on an edge was used. A natural alternative is to use a weighted average of the fluxes from the neighbouring elements. To this end, introduce smooth functions $\alpha_\gamma:\gamma\to\mathbb{R}$ on the interior edges. The functions α_γ are identically zero on exterior edges. Suppose that elements K and J share a common edge γ, then the approximation to the flux on the edge is taken to be the weighted average

$$g_K=\sigma_K\left\langle\frac{\partial u_X}{\partial n}\right\rangle_\cdot \tag{5.88}$$

where

$$\left\langle\frac{\partial u_X}{\partial n}\right\rangle_\cdot=\left(\frac{1}{2}+\sigma_K\alpha_\gamma\right)n\cdot\nabla u_X|_K+\left(\frac{1}{2}+\sigma_J\alpha_\gamma\right)n\cdot\nabla u_X|_J. \tag{5.89}$$

Notice that the scheme is consistent in the sense that if the approximation from each of the neighbouring elements is the same then the approximation g_K will agree with this value. This expression may be rewritten as

$$\left\langle \frac{\partial u_X}{\partial n} \right\rangle_{\gamma} = \left\langle \frac{\partial u_X}{\partial n} \right\rangle + \alpha_\gamma \left[\frac{\partial u_X}{\partial n} \right] \tag{5.90}$$

where

$$\left[\frac{\partial u_X}{\partial n} \right] = n_K \cdot \nabla u_X|_K + n_J \cdot \nabla u_X|_J \tag{5.91}$$

is the jump in the approximation to the flux across the edge γ. Inserting (5.91) into the equilibration condition and simplifying leads to the condition on each element K

$$-\sum_{\gamma \subset \partial K} \sigma_K \int_\gamma \alpha_\gamma \left[\frac{\partial u_X}{\partial n} \right] \mathrm{d}s = \delta_K(1). \tag{5.92}$$

As before, this is still a global problem thanks to the coupling across edges expressed through the functions α_γ. The functions α_γ are chosen to be piecewise linear on the edges

$$\alpha_\gamma(s) = \alpha_\gamma^l \theta_l(s) + \alpha_\gamma^r \theta_r(s) \tag{5.93}$$

where α_γ^l are constants to be determined. In an attempt to decouple the problem, we replace unity on the right-hand side by the sum (5.76) and insert the definition (5.93) to obtain the condition for each element K

$$-\sum_{\gamma \subset \partial K} \sigma_K \alpha_\gamma^l \int_\gamma \theta_l(s) \left[\frac{\partial u_X}{\partial n} \right] \mathrm{d}s - \sum_{\gamma \subset \partial K} \sigma_K \alpha_\gamma^r \int_\gamma \theta_r(s) \left[\frac{\partial u_X}{\partial n} \right] \mathrm{d}s = \sum_{n \in \Psi(K)} \delta_K(\theta_n). \tag{5.94}$$

One way to obtain a solution of this system is to satisfy the stronger conditions: for each vertex $x_n \in \Psi$

$$-\sum_{\gamma \subset \partial K} \sigma_K \alpha_\gamma^n \int_\gamma \theta_n(s) \left[\frac{\partial u_X}{\partial n} \right] \mathrm{d}s = \delta_K(\theta_n) \quad \forall K \in \tilde{\Omega}_n. \tag{5.95}$$

If conditions (5.95) hold then by summing over all vertices $n \in \Psi(K)$ one obtains (5.94). The systems (5.95) consist of independent problems to be solved for α_γ^n over the support of the function θ_n. As before, it is not immediately obvious that there is a solution of the problems. However, if one can find constants α_γ^n satisfying these conditions then the flux approximation is obtained from definitions (5.93) and (5.88).

5.4.3. Solvability of local equilibration systems

Both Ladeveze's Method and Flux Splitting reduce the single coupled global equilibration condition into a sequence of independent equilibration problems. Each such problem is localized over the patch $\tilde{\Omega}_n$ of elements surrounding each of the nodes x_n in the partition. Both systems of equations are of the form: for each vertex $x_n \in \Psi$

$$-\sum_{\gamma \subset \partial K} \sigma_K \mu_\gamma^n = \delta_K(\theta_n) \quad \forall K \subset \tilde{\Omega}_n \tag{5.96}$$

where, in the case of Ladeveze's method

$$\mu_\gamma^n = \beta_\gamma^n \tag{5.97}$$

and in the case of Flux Splitting

$$\mu_\gamma^n = \alpha_\gamma^n \int_\gamma \theta_n(s) \left[\frac{\partial u_X}{\partial n} \right] \mathrm{d}s. \tag{5.98}$$

This correspondence means that we may deal with the solvability of the local equilibration conditions for both methods at the same time. To simplify the notation we shall omit the superscript n.

Proper partitions in the plane

Suppose that the partition $\mathcal{P}$ is proper. That is, each element edge is either a subset of the exterior boundary $\partial\Omega$ or a complete edge of another element in the partition. The domain Ω is supposed to be

two dimensional. There are now two possibilities to be considered depending on whether the vertex x_n lies on the boundary of Ω or the interior.

Interior Vertex. If x_n is an interior vertex then the subdomain $\tilde{\Omega}_n$ is of the form shown in Fig. 5. The system (5.96) in this case is

$$\begin{aligned} -\mu_1 + \mu_2 &= \delta_1(\theta_n) \\ -\mu_2 + \mu_3 &= \delta_2(\theta_n) \\ &\vdots \\ -\mu_N + \mu_1 &= \delta_N(\theta_n). \end{aligned} \tag{5.99}$$

This represents a system of N linear equations in N unknowns. However, the equations are linearly dependent as may be seen by summing all equations. The left-hand side then vanishes. Fortunately, summing the right-hand sides gives

$$\begin{aligned} \sum_{K=1}^{N} \delta_K(\theta_n) &= \sum_{K=1}^{N} F_K(\theta_n) - B_K(u_X, \theta_n) + \oint_{\partial K} \left\langle \frac{\partial u_X}{\partial n_K} \right\rangle \theta_n(s)\, ds \\ &= F(\theta_n) - B(u_X, \theta_n) \end{aligned} \tag{5.100}$$

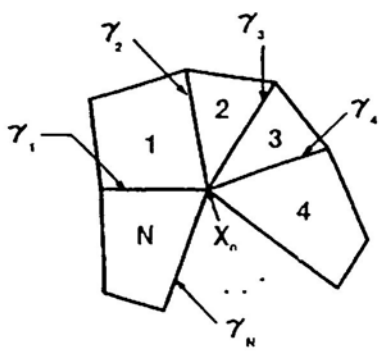
where the final step follows since the integrals over the boundary cancel pairwise across on each edge, and because we have summed over all elements on which θ_n is non-zero. Noting that x_n is an interior vertex it follows from the definition of the finite element approximation u_X itself that

$$B(u_X, \theta_n) = L(\theta_n) = F(\theta_n) \tag{5.101}$$

and so the sum of the right-hand sides also vanishes. Consequently, the system (5.99) has solutions determined up to an arbitrary constant.

Boundary vertex. Suppose that the vertex x_n lies on the boundary Γ_N as in Fig. 6. The values of the constants μ on the edges γ_1 and γ_{N+1} are required to vanish in both Ladeveze's and the Flux Splitting methods. This is a natural condition since the value of the flux on these boundaries is known exactly and it is undesirable to introduce any perturbations. Incorporating these constraints, the system (5.96) has the form

$$\begin{aligned} \mu_2 &= \delta_1(\theta_n) \\ -\mu_2 + \mu_3 &= \delta_2(\theta_n) \\ &\vdots \\ -\mu_N &= \delta_N(\theta_n). \end{aligned} \tag{5.102}$$



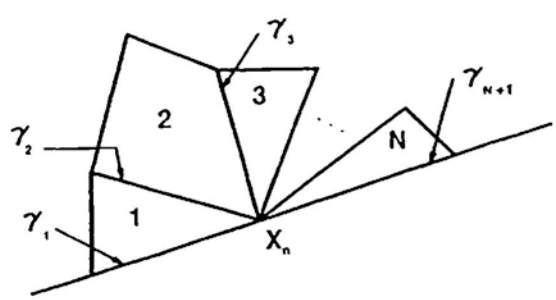


Fig. 5. Patch $\tilde{\Omega}_n$ associated with interior vertex x_n.

Fig. 6. Patch $\tilde{\Omega}_n$ associated with boundary vertex x_n.

This time there are N equations in only $N-1$ unknowns. The first $N-1$ conditions in (5.102) uniquely determine the unknowns. Summing the first $N-1$ equations reveals that

$$\mu_N = -\delta_N(\theta_n) + \sum_{K=1}^{N} \delta_K(\theta_n) \tag{5.103}$$

and then a calculation similar to the case of an interior vertex shows that the second term on the right-hand side vanishes. Consequently, the Nth equation in the system (5.102) is automatically satisfied. The system therefore has a unique solution when the vertex lies on the boundary.

There is an additional possibility that may arise in the Flux Splitting algorithm. It may happen that the approximation to the normal fluxes between two elements is continuous. This imposes the extra constraint that the constant μ_γ^n must vanish. If there is only one edge on which to enforce the constraint then there is no problem since the solution of the original problem was determined only up to an arbitrary constant. However, for more than one edge the conditions cannot be satisfied in general. In practice, these situations occur due to symmetry which often allows the system to be solved anyway.

Irregular partitions in the plane

Several finite element codes now incorporate local refinements in which irregular partitions are created (see Fig. 7). Previously, these cases have not be singled for attention because little is changed from the analysis of the proper partitions. It is worth pointing out the differences in the equilibration procedures for such meshes. The discussion is based on Ainsworth and Oden [5].

The first difference comes in the classification of the vertices of the partition. The open nodes in Fig. 7 show where the linear degrees of freedom must be constrained so that a conforming approximation is obtained. Such vertices must now be excluded from the set Ψ of *regular* vertices in the partition. The Lagrange basis function associated with the vertex x_n shown in Fig. 8 is still defined to be the piecewise bilinear function satisfying the conditions

$$\theta_n(x_m) = \delta_{mn} \quad m \in \Psi. \tag{5.104}$$

However, a little reflection reveals that θ_n is also non-zero on elements not containing the vertex x_n. In particular, the support of θ_n consists of elements 2, ..., 7. The previous definition is modified to become

$$\tilde{\Omega}_n = \{K \in \mathcal{P} : K \subset \operatorname{supp} \theta_n\}. \tag{5.105}$$

That is, $\tilde{\Omega}_n$ consists of the elements on which θ_n is non-zero. The definition of the set $\Psi(K)$ is also modified to

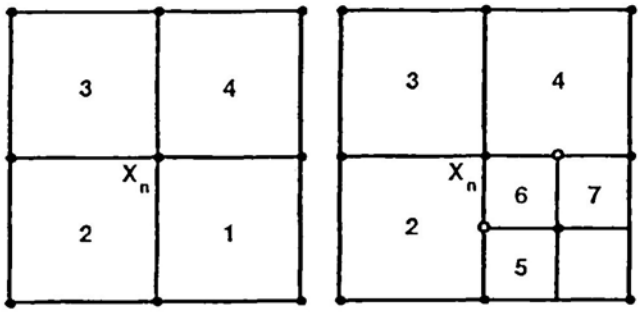


Fig. 7. An irregular partition created by local refinement.

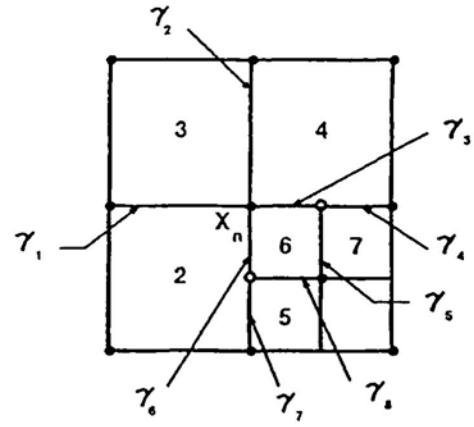


Fig. 8. Notations for equilibration problem on patch $\tilde{\Omega}_n$.

$$\Psi(K) = \{n \in \Psi : K \subset \operatorname{supp} \theta_n\}. \tag{5.106}$$

Each of these definitions generalize the original conditions. It is worth noting that the key identity (5.76) used in the localization process is preserved

$$\sum_{n \in \Psi(K)} \theta_n(x) = 1, \quad x \in \overline{K}. \tag{5.107}$$

Both Ladeveze's method and the Flux Splitting algorithm are derived following the same steps as before. This leads to a series of local systems of equations that are expressed exactly as in (5.96): for each vertex $x_n \in \Psi$

$$-\sum_{\gamma \subset \partial K} \sigma_K \mu_\gamma^n = \delta_K(\theta_n) \quad \forall K \subset \tilde{\Omega}_n. \tag{5.108}$$

However, the modified notations mean that the issue of solvability has to be considered afresh. For example, expanding the conditions for the vertex x_n using the notations shown in Fig. 8 gives (omitting superscripts)

$$\begin{aligned}
-\mu_1 - \mu_6 - \mu_7 &= \delta_2(\theta_n) \\
\mu_1 - \mu_2 &= \delta_3(\theta_n) \\
\mu_2 - \mu_3 - \mu_4 &= \delta_4(\theta_n) \\
\mu_7 - \mu_8 &= \delta_5(\theta_n) \\
\mu_3 - \mu_5 + \mu_6 + \mu_8 &= \delta_6(\theta_n) \\
\mu_4 + \mu_5 &= \delta_7(\theta_n).
\end{aligned} \tag{5.109}$$

The system is singular as may be seen by summing the left-hand sides. As before the sum of the right-hand sides reduces to $L(\theta_n) - B(u_X, \theta_n)$ which vanishes since (x_n is a regular node) θ_n belongs to the finite element subspace. It suffices to show that this is the only linear dependency within the system. For proper partitions this was a trivial fact.

Consider then a general regular vertex x_n. The system (5.108) may be expanded as a matrix equation

$$M\mu = \delta \tag{5.110}$$

where $\delta \in \mathbb{R}^E$ with E the number of elements contained in the patch $\tilde{\Omega}_n$. We examine the null space $\operatorname{Ker} M^t$, where M^t is the transpose of M. Suppose that $\xi \in \operatorname{Ker} M^t$. Then $M^t\xi = 0$. First, notice that each column of M (and therefore each row of M^t) has precisely two non-zero entries corresponding a single edge γ in the patch $\tilde{\Omega}_n$. Moreover, if the edge γ separates elements L and R, with $L > R$ say, then these entries are: $+1$, in the row corresponding to element L; and -1 in the row corresponding to element R. In turn this means that $\xi_L - \xi_R = 0$ since $M^t\xi = 0$. Thus, for any pair of neighbouring elements in the patch we have that the corresponding components of ξ must be identical. However, for any pair of elements K and J in the patch, starting from K we can always find a path leading to element J by passing across element edges. As we pass across an edge the value of the component of ξ in the new element is the same as the value in the initial element K. Hence, $\xi_K = \xi_J$ for all pair of elements in the patch. We have shown that if $\xi \in \operatorname{Ker} M^t$ then ξ must have all its components equal, i.e.

$$\operatorname{Ker} M^t = \operatorname{span}\{\lambda\} \tag{5.111}$$

where $\lambda = (1, \ldots, 1)^t$. Therefore, a necessary and sufficient condition for the systems (5.108) to have a solution is that the sum of the components on the right-hand sides be zero. This has already been seen to be true. The solution exists and is determined up to an arbitrary constant (as for proper partitions).

The systems (5.108) have a sufficiently simple structure on proper partitions for it to be possible to write down solutions explicitly. However, on irregular partitions this is not so straightforward. It is undesirable to attempt to enumerate every possible mesh configuration. A simple general procedure for constructing solutions was presented in [5]. A typical system is of the form

$$M\boldsymbol{\mu} = \boldsymbol{\delta}. \tag{5.112}$$

Suppose that we can find a solution of the related system

$$MM^{\mathrm{t}}\boldsymbol{\xi} = \boldsymbol{\delta} \tag{5.113}$$

then we simply take $\boldsymbol{\mu} = M^{\mathrm{t}}\boldsymbol{\xi}$. The matrix MM^{t} is symmetric, positive and indefinite. Moreover

$$\mathrm{Ker}\,(MM^{\mathrm{t}}) = \mathrm{Ker}\,M^{\mathrm{t}} = \mathrm{span}\{\boldsymbol{\lambda}\} \tag{5.114}$$

with $\boldsymbol{\lambda}$ as before. Therefore, the system (5.113) has solutions. The key observation is that in addition, the matrix MM^{t} has a particularly simple structure. Assume K and J are elements in the patch $\tilde{\Omega}_n$. Earlier arguments show that the components of M are given by

$$M_{K,\gamma} = \begin{cases} 1, & \text{if } \gamma = \overline{K} \cap \overline{J} \text{ with } K > J \\ -1, & \text{if } \gamma = \overline{K} \cap \overline{J} \text{ with } K < J \\ 0, & \text{otherwise} \end{cases}. \tag{5.115}$$

Consider the diagonal element of MM^{t} corresponding to element K:

$$[MM^{\mathrm{t}}]_{K,K} = \sum_{\gamma} M^2_{K,\gamma} \tag{5.116}$$

where the summation is over all edges in the patch. Therefore,

$$[MM^{\mathrm{t}}]_{K,K} = \text{number of elements in patch adjacent to element } K. \tag{5.117}$$

A similar argument can be used to obtain the off-diagonal elements ($K \neq J$):

$$[MM^{\mathrm{t}}]_{KJ} = \begin{cases} -1, & \text{if } K \text{ and } J \text{ share an edge} \\ 0, & \text{otherwise} \end{cases}. \tag{5.118}$$

The net result is that the *topology matrix* MM^{t} is readily constructed directly from the topological information in the patch and has integer entries. The conjugate gradient algorithm is suitable to solve (5.113) since the matrix vector products may be efficiently implemented and the method generally requires at most three iterations (even for irregular meshes in three dimensions). Having obtained $\boldsymbol{\xi}$, the action of M^{t} is performed using the information in (5.115). The advantage of this approach is that it copes with all mesh topologies in a straightforward and numerically stable manner. Further details and operation counts will be found in [5].

Once a solution of the systems (5.108) has been computed, the fluxes are constructed using (5.83) for Ladeveze's method or (5.93) for Flux Splitting. There is the difficulty with Ladeveze's method that the definition (5.82) characterizing the function ψ_n has to be altered to accommodate the various topologies.

Partitions in three dimensions

Equilibration on (proper and irregular) partitions in three dimensions is essentially the same as in two dimensions with element faces taking over the role of the element edges. The solvability of the systems can be dealt with by precisely the same argument used for irregular partitions in two dimensions. Furthermore, the system can be solved efficiently even on irregular meshes by following the approach based on the topology matrix discussed above.

The minor difference with the two-dimensional algorithm is that the definition of the functions ψ_n used in Ladeveze's method must be altered. The chief advantage of the Flux Splitting approach is that it can cope with the myriad of possible mesh configurations when using irregular meshes in three dimensions. The drawback is the careful treatment sometimes needed for continuous fluxes.

5.4.4. Higher-order equilibration

Let X be the finite element subspace and for each element let X_K consist of the restrictions of functions from X to the element K. It has been demonstrated that flux functions g_γ on the edges may

be constructed so that on each element K in the partition

$$0 = F_K(v) - B_K(u_X, v) + \int_{\partial K} g_K v \, ds \quad \text{for all } v \in X_K. \tag{5.119}$$

The equilibration condition follows immediately from this fact. So far all the arguments have been for first-order approximation. The question arises of whether (5.119) can be satisfied when X_K is constructed using higher-order basis functions. The following simple inductive argument will show how property (5.119) can be satisfied when X_K is constructed using higher-order basis functions.

Suppose that the finite element subspace is based on polynomials of degree $p = 2$. Associated with each edge γ in the partition is a test function $\theta_\gamma \in X$ supported on the two elements sharing the edge. Often θ_γ is referred to as the edge bubble function. Let

$$\tilde{\Omega}_\gamma = \{K \in \mathcal{P} : K \subset \operatorname{supp} \theta_\gamma\} \tag{5.120}$$

so that $\tilde{\Omega}_\gamma$ consists of the elements of which γ is an edge.

Let g'_γ be edge functions constructed so that condition (5.119) holds whenever v is a first-order polynomial basis function. New edge functions approximating the boundary flux are defined by

$$g_\gamma = g'_\gamma + \mu_\gamma \psi_\gamma \tag{5.121}$$

where $\psi_\gamma : \gamma \to \mathbb{R}$ is the quadratic function uniquely defined by the conditions

$$\int_\gamma \psi_\gamma(s)\theta_n(s) \, ds = 0 \quad \text{for all } n \in \Psi \tag{5.122}$$

and

$$\int_\gamma \psi_\gamma^2(s) \, ds = 1. \tag{5.123}$$

The constant μ_γ will be chosen to satisfy the higher-order equilibration condition. The boundary function is then given by $g_K = \sigma_K g_\gamma$. The equilibration condition

$$0 = F_K(v) - B_K(u_X, v) + \int_{\partial K} g_K v \, ds \tag{5.124}$$

can easily be seen to hold whenever $v \in X_K$ is a first order basis function. Moreover, the condition holds when $v \in X_K$ is a second-order interior basis function supported on the single element K, from the definition of the finite element approximation itself. Thus, it suffices to deal with the case of v being an edge bubble function. Inserting the expression (5.121) into the condition (5.119) with $v = \theta_\gamma$ leads to

$$-\sigma_K \mu_\gamma = \delta'_K(\theta_\gamma) \quad \forall K \in \tilde{\Omega}_\gamma \tag{5.125}$$

where

$$\delta'_K(v) = F_K(v) - B_K(u_X, v) + \int_{\partial K} g'_K v \, ds. \tag{5.126}$$

This is analogous to the condition (5.87) but has a simpler form owing to θ_γ being supported on only one edge. Letting R and L denote the pair of elements sharing edge γ with $R > L$ gives the conditions

$$\begin{aligned} -\mu_\gamma &= \delta'_R(\theta_\gamma) \\ \mu_\gamma &= \delta'_L(\theta_\gamma). \end{aligned} \tag{5.127}$$

The system has a solution since

$$\delta'_R(\theta_\gamma) + \delta'_R(\theta_\gamma) = L(\theta_\gamma) - B(u_X, \theta_\gamma) = 0 \tag{5.128}$$

thanks to the definition of the finite element approximation. The construction is repeated for all edges allowing boundary fluxes to be found such that condition (5.119) is satisfied.

The process described above shows how boundary fluxes equilibrated up to first order may be extended to second order. The same arguments may be used inductively to extend the equilibration to the full order p of the finite element approximation. Of course, one cannot then continue to higher-orders since the argument relies on properties of the finite element approximation.

The results obtained in this section are summarized in:

THEOREM 5.7. Let $u_X \in X$ be a Galerkin finite element approximation:

$$B(u_X, v) = L(v) \quad \forall v \in X. \tag{5.129}$$

Then, there exist smooth (polynomial) functions defined on the interelement edges such that the equilibration condition is satisfied

$$0 = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds \quad \forall v \in X_K \tag{5.130}$$

where

$$F_K(v) = \int_K f v \, dx \tag{5.131}$$

and $g_K + g_J = 0$ on $\overline{K} \cap \overline{J}$. Moreover, the functions may be computed locally using the algorithms described above.

5.5. A posteriori error estimators

Two key results have been obtained: Theorem 5.3, is the basic result showing the possibility of obtaining computable upper bounds on the energy norm of error in the finite element approximation; and Theorem 5.7, showing that it is possible to compute approximations to the boundary flux so that the equilibration condition is satisfied. The task is to use these results to derive a posteriori error estimators. There are two basic approaches: solve the *primal* problem or solve the *dual* problem.

5.5.1. Primal method

The primal method consists of solving the *primal* problem (5.53):

$$B_K(\phi_K, v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds \quad \forall v \in V_K \tag{5.132}$$

where V_K is the space

$$V_K = \{v \in H^1(K) : \gamma v = 0 \text{ on } \Gamma_D \cap \partial K\}. \tag{5.133}$$

A solution exists provided that the equilibration condition is satisfied. According to Theorems 5.3 and 5.4 (or Theorem 5.6 in case $c = 0$) one obtains

$$|||e|||^2 \leqslant \sum_{K \in \mathcal{P}} |||\phi_K|||_K^2. \tag{5.134}$$

Thus, an appropriate choice for the error estimator η_K is to take

$$\eta_K = |||\phi_K|||_K. \tag{5.135}$$

This method is not a viable algorithm since the space V_K is infinite dimensional.

Relationship with classical element residual method

The problem (5.132) is reminiscent of the classical element residual method discussed in the previous section. In fact, suppose that the finite element approximation u_X is based on first order basis functions and that we simply take the boundary flux functions to be

$$g_Y = \left\langle \frac{\partial u_X}{\partial n} \right\rangle. \tag{5.136}$$

The residual problem (5.141) may not have a solution with this choice of data. However, if we first replace the space V_K by the finite dimensional subspace

$$V_K \approx Y_K \tag{5.137}$$

where Y_K is one of the subspaces defined in Section 4.3.1, then problem (5.141) now has a solution. The method is precisely the classical element residual method. The approximation provided by the subspace Y_K means that one will not have a guaranteed upper bound on the error.

5.5.2. *Duality method*

An alternative approach is to solve the *dual* problem suggested by Theorems 5.4 and 5.6: find $\boldsymbol{p} \in W_K$ such that

$$G_K(\boldsymbol{p}) \geqslant G_K(\boldsymbol{q}) \quad \forall \boldsymbol{q} \in W_K \tag{5.138}$$

where W_K is defined in Theorems 5.4 and 5.6. The equilibration principle assures us that a solution $\boldsymbol{p}$ exists. An error estimator may be defined by

$$\eta_K = -2G_K(\boldsymbol{p}) \tag{5.139}$$

and Theorem 5.3 then gives the upper bound

$$|||e|||^2 \leqslant \sum_{K \in \mathcal{P}} |||\phi_K|||_K^2. \tag{5.140}$$

The infinite dimensional problem (5.138) has to be approximated by a suitable finite dimensional approximation. Ladeveze and Leguillon [43] construct a finite dimensional subspace of W_K using a finite element discretization on a partitioning of the element K into three or four subelements. Many other constructions suggest themselves. The basic approach to a posteriori error estimation based on constructing a dual variational principle seems to be due to de Veubeke [27].

5.6. *The equilibrated residual method*

The infinite dimensional problem to be approximated is: find $\phi_K \in V_K$ such that

$$B_K(\phi_K, v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds \quad \forall v \in V_K \tag{5.141}$$

where the finite element approximation u_X is based on polynomials of degree p. Suppose that the boundary fluxes have been chosen so that the equilibration condition

$$0 = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds \quad \forall v \in X_K \tag{5.142}$$

is satisfied. We construct a family of subspaces $Y_K^{(q)}$, $q \in \mathbb{N}$ as follows. If the element K is the image of the reference element $\widehat{K}$ under an invertible mapping F_K then

$$Y_K^{(q)} = \{\widehat{v} \circ F_K^{-1} : \widehat{v} \in \widehat{R}(q)\} \tag{5.143}$$

where $\widehat{R}(q)$ is the space

$$\widehat{R}(q) = \begin{cases} \widehat{Q}(q) & \text{if } \widehat{K} \text{ is a square} \\ \widehat{P}(q) & \text{if } \widehat{K} \text{ is a triangle} \end{cases} \tag{5.144}$$

In particular, note that the local finite element space $X_K = Y_K^{(p)}$. A sequence of error estimators for $q = p, p+1, \ldots$ may be defined by: find $\phi_K^{(q)} \in Y_K^{(q)}$ such that

$$B_K(\phi_K^{(q)}, v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds \quad \forall v \in Y_K^{(q)} \tag{5.145}$$

with the error estimator given by

$$|||\phi_K^{(q)}|||_K = \sqrt{B_K(\phi_K^{(q)}, \phi_K^{(q)})}. \tag{5.146}$$

By increasing the polynomial degree q, the accuracy of the approximation to the infinite dimensional problem (5.141) is improved. In effect, one is using the p version finite element method on the single element K; cf. Babuska et al. [17,18].

It is impractical to calculate the error estimators using the full space $Y_K^{(q)}$ for larger values of q owing to the expense of the computation. The alternative is to solve on a subspace of the full space $Y_K^{(q)}$. For instance, one might use the same functions used for the classical element residual method. However, the equilibration property may be exploited by using a more appropriate subspace $\mathcal{B}_K^{(q)}$

$$\mathcal{B}_K^{(q)} = \{v \in Y_K^{(q)} : B(v, w) = 0 \; \forall w \in Y_K^{(q-1)}\}. \tag{5.147}$$

Clearly, these basis functions will be problem dependent and are discussed more fully later. Importantly, the dimension of the space $\mathcal{B}_K^{(q)}$ is significantly less than the full space $Y_K^{(q)}$

$$\dim Y_K^{(q)} = (q+1)^2 \quad \text{and} \quad \dim \mathcal{B}_K^{(q)} = 2q + 1. \tag{5.148}$$

The size of the error residual problem using the space $\mathcal{B}_K^{(q)}$ is relatively modest but the danger is that accuracy may have been sacrificed. The principal property of the subspaces is that the solution of the error residual problem is *identical* with the solution obtained using the full space:

THEOREM 5.8. Suppose that the boundary fluxes have been chosen so that the equilibration condition is satisfied. For $q = p+1, p+2, \ldots$ let $\phi_K^{(q)} \in Y_K^{(q)}$ be such that

$$B_K(\phi_K^{(q)}, v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds \quad \forall v \in Y_K^{(q)} \tag{5.149}$$

and let $\tilde{\phi}_K^{(q)} \in \mathcal{B}_K^{(q)}$ be such that

$$B_K(\tilde{\phi}_K^{(q)}, v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds \quad \forall v \in \mathcal{B}_K^{(q)}. \tag{5.150}$$

Then

$$\phi_K^{(q)} = \sum_{j=p+1}^{q} \tilde{\phi}_K^{(j)} \tag{5.151}$$

and

$$|||\phi_K^{(q)}|||_K^2 = \sum_{j=p+1}^{q} |||\tilde{\phi}_K^{(j)}|||_K^2 . \tag{5.152}$$

PROOF. The definition of the spaces $\mathcal{B}_K^{(k)}$ implies that for any $v_j \in \mathcal{B}_K^{(j)}$ and $w_k \in \mathcal{B}_K^{(k)}$

$$B(v_j, w_k) = 0 \quad \text{if } j \neq k \tag{5.153}$$

and for $q > p$

$$Y_K^{(q)} = X_K \cup \bigcup_{j=p+1}^{q} \mathcal{B}_K^{(j)}. \tag{5.154}$$

Therefore, we may write

$$\phi_K^{(q)} = w_X + \sum_{j=p+1}^{q} w_j \tag{5.155}$$

where $w_X \in X_K$ and $w_j \in \mathcal{B}_K^{(j)}$. Let $D_K(v)$ denote the linear functional appearing in the error residual problem

$$D_K(v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds. \tag{5.156}$$

Then, for any $v_X \in X_K$, the equilibration property reveals that

$$0 = D_K(v_X) = B(\phi_K^{(q)}, v_X) = \sum_{j=p+1}^{q} B(w_j, v_X) + B(w_X, v_X) = B(w_X, v_X) \tag{5.157}$$

and hence, $w_X = 0$. Therefore, for any $v_k \in \mathcal{B}_K^{(k)}$

$$B(\tilde{\phi}_K^{(k)}, v_k) = D_K(v_k) = B(\phi_K^{(q)}, v_k) = \sum_{j=p+1}^{q} B(w_j, v_k) = B(w_k, v_k) \tag{5.158}$$

and hence $w_k = \tilde{\phi}_K^{(k)}$. Consequently,

$$\phi_K^{(q)} = \sum_{j=p+1}^{q} \tilde{\phi}_K^{(j)}. \tag{5.159}$$

Property (5.153) immediately shows that

$$B(\phi_K^{(q)}, \phi_K^{(q)}) = \sum_{j=p+1}^{q} B(\tilde{\phi}_K^{(j)}, \tilde{\phi}_K^{(j)}). \quad \square \tag{5.160}$$

The result shows not only that one may use the reduced spaces $\mathcal{B}_K^{(k)}$ to construct the solution on the full space, but also reveals that the functions $\tilde{\phi}_K^{(k)}$, $k = p+1, p+2, \ldots$ may be computed independently and then summed. In each case, the resulting error estimator is identical. The result is of considerable practical importance with regard to computing the estimators economically.

A posteriori error estimates for the residual problem

One can actually assess the accuracy of the approximation $\phi_K^{(q)} \approx \phi_K$ to the solution of the infinite dimensional problem. If the p-version finite element method is used to approximate a solution having a singularity on a corner of the domain then the rate of convergence is (see [18,17])

$$|||\phi_K - \phi_K^{(q)}|||_K \leqslant C N_q^{-\alpha} \tag{5.161}$$

where $N_q = (q+1)^2$ and C, α are positive constants depending only on ϕ_K. A simple computation reveals that

$$|||\phi_K - \phi_K^{(q)}|||_K^2 = |||\phi_K|||_K^2 - |||\phi_K^{(q)}|||_K^2. \tag{5.162}$$

Therefore, we make the assumption that

$$|||\phi_K|||_K^2 - |||\phi_K^{(q)}|||_K^2 = C N_q^{-2\alpha}. \tag{5.163}$$

Suppose that we compute $|||\tilde{\phi}_K^{(p+1)}|||_K$ and $|||\tilde{\phi}_K^{(p+2)}|||_K$. The value of $|||\tilde{\phi}_K^{(p)}|||_K$ is known to be zero owing to the equilibration condition. Theorem 5.8 may be invoked to obtain the values of $|||\phi^{(q)}|||_K$, $q = p$, $p+1$ and $p+2$. Three equations may be obtained from Eq. (5.163) by choosing q to be p, $p+1$ and $p+2$. Eliminating the constants C and α between these equations leaves the following equation for the 'unknown' $|||\phi_K|||_K$:

$$\frac{|||\phi_K|||_K^2 - |||\phi_K^{(p+2)}|||_K^2}{|||\phi_K|||_K^2 - |||\phi_K^{(p+1)}|||_K^2} = \left(\frac{|||\phi_K|||_K^2 - |||\phi_K^{(p+1)}|||_K^2}{|||\phi_K|||_K^2 - |||\phi_K^{(p)}|||_K^2} \right)^{\frac{\log(N_{p+2}/N_{p+1})}{\log(N_{p+1}/N_p)}}. \tag{5.164}$$

This equation may be solved to obtain an approximation to the value of $|||\phi_K|||_K$. This type of extrapolation technique has been used to obtain global a posteriori error estimates for p-version finite element computations [54].

Summary

The purpose of this section has been to show how, in principle, one may resolve the infinite dimensional local error residual problem (5.53) exploiting the equilibration property. A key role is played by the finite dimensional subspaces $\mathcal{B}_K^{(q)}$. In particular, one may control the accuracy of the approximation of the infinite dimensional problem, and even make use of an extrapolation procedure to obtain enhanced approximations to the energy of the true solution. Summarizing, we have

THEOREM 5.9. Suppose that the boundary fluxes have been chosen so that the equilibration condition is satisfied. For each element $K \in \mathcal{P}$, let $\tilde{\phi}_K^{(q)} \in \mathcal{B}_K^{(q)}$, $q = p+1, p+2, \ldots$, be defined by

$$B_K(\tilde{\phi}_K^{(q)}, v) = F_K(v) - B_K(u_X, v) + \oint_{\partial K} g_K v \, ds \quad \forall v \in \mathcal{B}_K^{(q)}. \tag{5.165}$$

Then

$$\sum_{k=p+1}^{q} |||\tilde{\phi}_K^{(k)}|||_K^2 \to |||\phi_K|||_K^2 \quad \text{as } q \to \infty \tag{5.166}$$

and the error in the finite element approximation is bounded by

$$|||e|||^2 \leqslant \sum_{K \in \mathcal{P}} |||\phi_K|||_K^2. \tag{5.167}$$

PROOF. The limit (5.166) follows from the convergence estimates for the p-version finite element method and Theorem 5.8. The bound (5.167) follows from Theorems 5.3, 5.4 and 5.6. □

5.7. Treatment of the local spaces $\mathcal{B}^{(q)}$

5.7.1. Construction

Let $\hat{K}$ be a reference element and define the space

$$\hat{\mathcal{B}}^{(q)} = \{v \in \hat{R}(q) : \hat{B}(v, w) = 0 \ \forall w \in \hat{R}(q-1)\} \tag{5.168}$$

where $\hat{R}(q)$ is either $\hat{Q}(q)$ or $\hat{P}(q)$ depending on whether $\hat{K}$ is a square or triangle. A basis for $\hat{\mathcal{B}}^{(q)}$ is easily constructed. Suppose that the basis functions for $\hat{R}(q)$ are ordered so that functions belonging to the subspace $\hat{R}(q-1)$ appear first. Let Ψ_q be the vector whose components are the basis functions, so that

$$\Psi_q = [\Psi_L, \Psi_H] \tag{5.169}$$

where Ψ_L (respectively, Ψ_H) is formed using the basis functions from $\hat{R}(q-1)$ (respectively, $\hat{R}(q) \backslash \hat{R}(q-1)$). This partitioning induces a block structure on the element stiffness matrix

$$\begin{bmatrix} \hat{A}_{LL} & \hat{A}_{HL} \\ \hat{A}_{LH} & \hat{A}_{HH} \end{bmatrix}. \tag{5.170}$$

A basis for the space $\hat{\mathcal{B}}^{(q)}$ is obtained by letting

$$\Psi'_H = \Psi_H - \hat{A}_{HL}\hat{A}_{LL}^{-1}\Psi_L \tag{5.171}$$

and taking the basis functions to be the components of Ψ'_H. It is readily verified that these functions form a basis for the space $\hat{\mathcal{B}}^{(q)}$. Figs. 9–11 show typical basis functions for the spaces $\hat{\mathcal{B}}^{(2)}$, $\hat{\mathcal{B}}^{(3)}$ and $\hat{\mathcal{B}}^{(4)}$ when $\hat{K}$ is the square reference element and the bilinear form corresponds to the Laplace operator. The

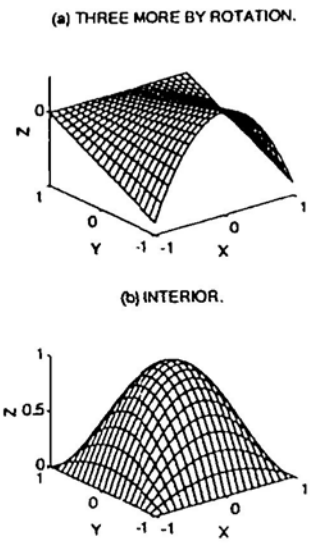


Fig. 9. Typical basis functions in the space $\widehat{\mathcal{B}}^{(2)}$.

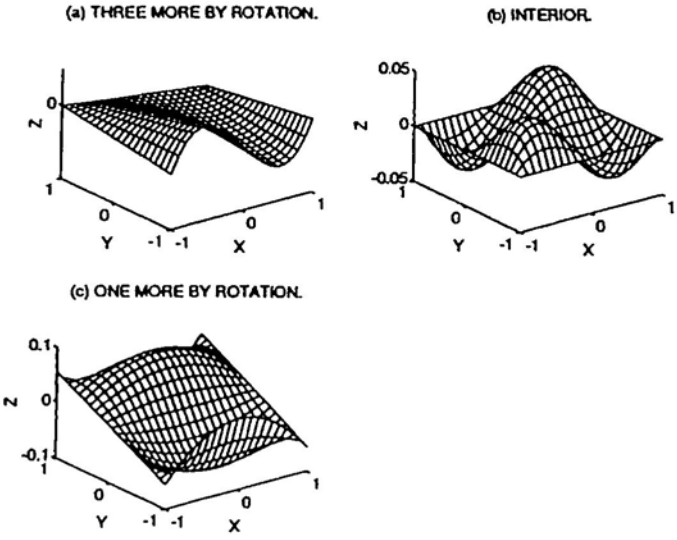


Fig. 10. Typical basis functions in the space $\widehat{\mathcal{B}}^{(3)}$.

matrix $\widehat{A}_{LL}$ is singular unless the constant mode is factored out by prescribing a Dirichlet condition at a single (arbitrary) point.

A basis for the spaces $\mathcal{B}_K^{(q)}$ can be found in the same manner.

5.7.2. *Approximate subspaces*

An obvious drawback in constructing the spaces $\mathcal{B}_K^{(q)}$ is that the computation has to be repeated for each element. An alternative is to define approximate subspaces

$$\tilde{\mathcal{B}}_K^{(q)} = \{\hat{v} \circ F_K^{-1} : \hat{v} \in \widehat{\mathcal{B}}^{(q)}\} \tag{5.172}$$

where $F_K : \hat{K} \to K$ is the usual mapping of the reference element. The space $\widehat{\mathcal{B}}^{(q)}$ is then constructed once and for all on the appropriate reference element and the basis functions 'hard-wired' into the finite

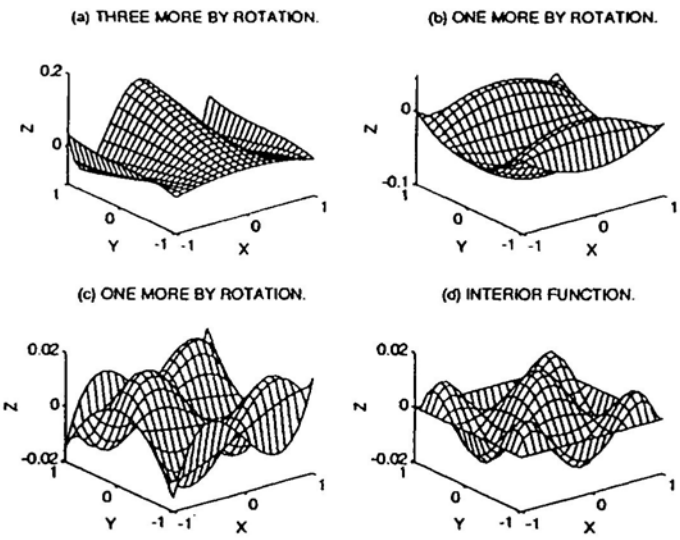


Fig. 11. Typical basis functions in the space $\widehat{\mathcal{B}}^{(4)}$.

element code. The bilinear form $\widehat{B}(\cdot,\cdot)$ is taken to be the operator for the class of problems under consideration with frozen coefficients. If the elements are not too distorted then one might expect that as the partition is refined, the spaces $\widetilde{\mathcal{B}}_K^{(q)}$ produce results approaching those obtained if the true spaces $\mathcal{B}_K^{(q)}$ were used. This statement will be quantified after the next section.

5.7.3. Framework for analysis of approximate subspaces

The analysis of the effect of choosing a subspace with which to approximate the full space may be dealt with under a general framework. The problem on the full space is of the form: find $\phi \in Y$ such that

$$B(\phi, v) = D(v) \quad \forall v \in Y \tag{5.173}$$

where the data $D(\cdot)$ satisfies the equilibration condition

$$D(v) = 0 \quad \forall v \in X \subset Y. \tag{5.174}$$

Typically, we choose an approximate subspace Y' to be the space Y with elements of the subspace X excluded. Therefore, let $I\!I : Y \to X$ be bounded, linear and surjective. The approximate subspace has the form

$$Y' = \{v - I\!I v : v \in Y\}. \tag{5.175}$$

The approximation to the problem on the full space is to find: $\phi' \in Y'$

$$B(\phi', v) = D(v) \quad \forall v \in Y'. \tag{5.176}$$

The issue is to assess the accuracy of the approximation $|||\phi||| \approx |||\phi'|||$. The accuracy is related to the angle between the subspaces Y' and X expressed through a *Strengthened Cauchy–Schwarz Inequality*.

THEOREM 5.10. Suppose that the Strengthened Cauchy–Schwarz Inequality holds

$$|B(v, w)| \leqslant \mu |||v||| \, |||w||| \quad \forall v \in Y', w \in X \tag{5.177}$$

where $0 \leqslant \mu < 1$. *Then*

$$|||\phi'||| \leqslant |||\phi||| \leqslant \frac{1}{\sqrt{1-\mu^2}} |||\phi'|||. \tag{5.178}$$

PROOF. Let $v \in Y$ be arbitrary. Then

$$\begin{aligned} |||v - \Pi v|||^2 &= |||v|||^2 - |||\Pi v|||^2 + 2B(\Pi v, \Pi v - v) \\ &\leqslant |||v|||^2 - |||\Pi v|||^2 + 2\mu |||\Pi v||| |||v - \Pi v||| \\ &\leqslant |||v|||^2 + \mu^2 |||v - \Pi v|||^2 \end{aligned} \tag{5.179}$$

where the inequality $2\mu ab \leqslant a^2 + \mu^2 b^2$ has been used. Hence,

$$|||v - \Pi v||| \leqslant \frac{1}{\sqrt{1-\mu^2}} |||v|||. \tag{5.180}$$

Note that

$$|||\phi||| = \sup_{v \in Y} \frac{|B(\phi, v)|}{|||v|||} \tag{5.181}$$

and so by the equilibration condition and the definitions

$$\begin{aligned} |||\phi||| &= \sup_{v \in Y} \frac{|B(\phi, v - \Pi v)|}{|||v|||} \\ &= \sup_{v \in Y} \frac{|B(\phi', v - \Pi v)|}{|||v|||} \\ &\leqslant |||\phi'||| \sup_{v \in Y} \frac{|||v - \Pi v|||}{|||v|||} \\ &\leqslant \frac{1}{\sqrt{1-\mu^2}} |||\phi'|||. \end{aligned} \tag{5.182}$$

Finally, since $Y' \subset Y$

$$|||\phi'|||^2 = D(\phi') = B(\phi, \phi') \leqslant |||\phi||| \, |||\phi'||| \tag{5.183}$$

and the result is proved. □

The subspaces $\mathcal{B}_K^{(q)}$ defined earlier are constructed so that the Strengthened Cauchy–Schwarz Inequality is valid with the constant $\mu = 0$. Theorem 5.10 confirms the earlier finding that the solution on the subspace coincides with the solution on the full space. The effects of using the approximate subspaces $\widetilde{\mathcal{B}}_K^{(q)}$ can be analyzed by estimating the size of the constant μ. Intuitively, if the elements are not too distorted then μ will be small meaning that the subspaces provided a satisfactory approximation.

5.7.4. *Alternative choices of subspace*

Oden et al. [47] propose an alternative set of subspaces for dealing with the error residual problem which have been popular. The subspaces are based on a 'p-interpolation' operator $\Pi_p^* : \widehat{Q}(p+1) \to \widehat{Q}(p)$ defined by

$$\Pi_p^* u = u_L + u_E + u_I \tag{5.184}$$

where u_L is the standard bilinear interpolant to u at the vertices of the square reference element $\widehat{K}$. To describe the functions u_E and u_I, we first introduce the space $\mathbb{P}_0^{(p)}(I)$ of polynomials of degree p on the interval I that vanish at the endpoints. The restriction of the function u_E to each edge $\widehat{\gamma}$ of the reference element $\widehat{K}$ belongs to the space $\mathbb{P}_0^{(p)}(\widehat{\gamma})$ and satisfies

$$\int_{\widehat{\gamma}} (u_E - (u - u_L))' v' \, ds = 0 \quad \forall v \in \mathbb{P}_0^{(p)}(\widehat{\gamma}) \tag{5.185}$$

where the prime denotes the derivative with respect to arc length. The interior function $u_I \in \widehat{Q}(p)$ vanishes on the boundary $\partial\widehat{K}$ and satisfies

$$\int_{\widehat{K}} \nabla(u_I - (u - u_L - u_E)) \cdot \nabla v \, dx = 0 \quad \forall v \in \widehat{Q}(p) \cap H_0^1(\widehat{K}). \tag{5.186}$$

The operator Π_p^* is used to construct an approximate subspace for solving the error residual problems

$$\widehat{\mathcal{M}}^{(p+1)} = \{v \in \widehat{Q}(p+1) : \Pi_p^* v \equiv 0\}. \tag{5.187}$$

We shall analyze the effectiveness of this choice using the framework of the previous section. First, we find a simpler description for the space.

The interior function u_I vanishes along each of the boundaries $\widehat{\gamma}$ and so Eq. (5.185) may be rewritten as

$$\int_{\widehat{\gamma}} (\Pi_p^* u)' v' \, ds = \int_{\widehat{\gamma}} u' v' \, ds \quad \forall v \in \mathbb{P}_0^{(p)}(\widehat{\gamma}). \tag{5.188}$$

Similarly, Eq. (5.186) is equivalent to

$$\int_{\widehat{K}} \nabla(\Pi_p^* u) \cdot \nabla v \, dx = \int_{\widehat{K}} \nabla u \cdot \nabla v \, dx \quad \forall v \in \widehat{Q}(p) \cap H_0^1(\widehat{K}). \tag{5.189}$$

The operator Π_p^* has a far simpler interpretation:

LEMMA 5.11. Let $\Pi_p : \widehat{Q}(p+1) \to \widehat{Q}(p)$ denote the interpolation operator that interpolates at the Gauss–Lobatto points. That is,

$$(\Pi_p u)(\zeta_j, \zeta_k) = u(\zeta_j, \zeta_k) \quad \forall j, k = 0, \ldots, p \tag{5.190}$$

where ζ_j are the zeros of the polynomial $(1 - s^2)L_p'(s)$ with L_p the pth degree Legendre polynomial. Then

$$\Pi_p^* u \equiv \Pi_p u. \tag{5.191}$$

PROOF. Let $\delta \in \widehat{Q}(p)$ be the difference

$$\delta = \Pi_p^* u - \Pi_p u. \tag{5.192}$$

Each of the operators Π_p^* and Π_p interpolate at the vertices of $\widehat{K}$ and hence δ also vanishes at the vertices. Let $\widehat{\gamma}$ be any edge of $\widehat{K}$. Then by (5.188)

$$\int_{\widehat{\gamma}} \delta' v' \, ds = \int_{\widehat{\gamma}} (\Pi_p^* u - \Pi_p u)' v' \, ds \int_{\widehat{\gamma}} (u - \Pi_p u)' v' \, ds \tag{5.193}$$

but on the edge $u - \Pi_p u \propto (1 - s^2)L_p'(s)$ and since L_p satisfies Legendre's differential equation

$$\int_{\widehat{\gamma}} \delta' v'(s) \, ds \propto \int_{\widehat{\gamma}} L_p(s) v'(s) \, ds = 0 \tag{5.194}$$

using the orthogonality property of Legendre polynomials. Hence, δ vanishes on the boundary $\partial\widehat{K}$. Finally, by (5.189), for any $v \in \widehat{Q}(p) \cap H_0^1(\widehat{K})$

$$\int_{\widehat{K}} \nabla\delta \cdot \nabla v \, dx = \int_{\widehat{K}} \nabla(\Pi_p^* u - \Pi_p u) \cdot \nabla v \, dx = \int_{\widehat{K}} \nabla(u - \Pi_p u) \cdot \nabla v \, dx. \tag{5.195}$$

However, $u - \Pi_p u \propto (1 - x^2)L_p'(x)(1 - y^2)L_p'(y)$ and so similarly to above one finds

$$\int_{\widehat{K}} \nabla\delta \cdot \nabla v \, dx = 0 \quad \forall v \in \widehat{Q}(p) \cap H_0^1(\widehat{K}). \tag{5.196}$$

Hence, since δ vanishes on the boundary it follows that δ is identically zero. □

Table 2
Values of strengthened Cauchy–Schwarz constant μ for Oden et al. basis function

Degree p	μ	$\frac{1}{\sqrt{1-\mu^2}}$
2	0.6741	1.359
3	0.8898	2.191
4	0.9229	2.595
5	0.9387	2.901
6	0.9539	3.333
7	0.9559	3.406
8	0.9636	3.740
9	0.9662	3.878
10	0.9705	4.149

An immediate consequence is that the space $\widehat{\mathcal{M}}^{(p+1)}$ may be rewritten more simply in the form

$$\widehat{\mathcal{M}}^{(p+1)} = \text{span}\left\{\chi_j(x)\chi_p(y),\ \chi_p(x)\chi_j(y),\ \chi_p(x)\chi_p(y) \quad j = 1, \ldots, p-1\right\} \tag{5.197}$$

where $\chi_j(s) = (1-s^2)L_j'(s)$. The constant μ in the strengthened Cauchy–Schwarz Inequality may be computed by solving an eigenvalue problem. Table 2 contains the values obtained for various polynomial degrees for the Laplace operator. One sees that the value of the all important quotient $1/\sqrt{1-\mu^2}$ grows relatively slowly. The spaces $\mathcal{M}_K^{(p+1)}$ used on an actual element are obtained through the usual mapping principle and may result in much larger Cauchy–Schwarz constants if the elements are distorted or if the operator is not the Laplacian.

6. Applications

6.1. Stokes and Oseen's equations

There are a number of specific issues that must be resolved when developing a posteriori error estimates for the Stokes problem. Firstly, the Stokes problem involves an incompressibility constraint and one must decide how to take proper account of the condition. In addition, the Oseen approximation of the incompressible Navier–Stokes equations contains a non-self-adjoint operator in the momentum equations. This means that there is no natural energy norm in which to measure the error.

Explicit a posteriori error estimates for the Stokes' problem have been derived by Baranger and El Amri [24] and Verfurth [56]. Generalizations of the classical element residual method have been developed by Bank and Welfert [21] and Verfurth [56]. The estimator is obtained by solving a local Stokes problem on each element, yielding a pair of functions whose norm is then used as an estimate of the true discretization error. One might have expected to be faced with an element residual problem requiring the solution of a local *Stokes* problem. However, there are drawbacks with this approach. For instance, it has been shown in earlier sections that the local residual problem has to be approximated using an appropriate subspace. However, when dealing with the Stokes problem, the subspaces used to approximate the pressure and velocity components should also be constructed so that the *inf–sup* stability condition is satisfied. Consequently, one faces additional, rather awkward difficulties in designing stable schemes with which to approximate the local problem.

The present discussion follows [7]. The error estimator is based on solving local residual problems. Firstly, the basic question of the norm in which the error will be estimated is considered. One outcome is that, perhaps surprisingly, it is unnecessary to solve a local Stokes problem in order to obtain an a posteriori error estimate. The significance of this conclusion in the design of a general a posteriori error estimation procedure for incompressible fluid flow is vital. In particular, the approximation of the local residual problems can, to a large extent, be developed independently of the type of element used to approximate the original fluid flow problem since there is no *inf–sup* condition to be satisfied.

The analysis is valid for essentially any conforming discretization scheme for the Stokes problem. The approach reveals an appropriate equilibration principle for the determination of the boundary data and the error estimator provides an upper bound on the true error in an energy like norm.

6.1.1. Model problem

Introduce function spaces V and W as follows:

$$V = H_0^1(\Omega) \times H_0^1(\Omega) \quad \text{and} \quad W = L_2(\Omega). \tag{6.1}$$

Let $B : V \times V \to \mathbb{R}$ and $b : V \times W \to \mathbb{R}$ be the bilinear forms

$$b(\boldsymbol{v}, q) = -\int_\Omega q \, \mathbf{div}\, \boldsymbol{v} \, \mathrm{d}\boldsymbol{x} \tag{6.2}$$

and

$$B(\boldsymbol{v}, \boldsymbol{w}) = \int_\Omega \{\nu \nabla \boldsymbol{v} \cdot \nabla \boldsymbol{w} + \boldsymbol{w} \cdot (\boldsymbol{U} \cdot \nabla)\boldsymbol{v}\} \, \mathrm{d}\boldsymbol{x} \tag{6.3}$$

where $\nu > 0$ is the viscosity parameter and $\boldsymbol{U}$ is a smooth solenoidal vector field on Ω (i.e. $\mathbf{div}\, \boldsymbol{U} = 0$). For given data $\boldsymbol{f} \in L_2(\Omega) \times L_2(\Omega)$ we seek the solution of the problem:

Find $(\boldsymbol{u}, p) \in V \times W$ such that for all $(\boldsymbol{v}, q) \in V \times W$

$$B(\boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, p) + b(\boldsymbol{u}, q) = F(\boldsymbol{v}) \tag{6.4}$$

where $F : V \to \mathbb{R}$ is the linear functional

$$F(\boldsymbol{v}) = \int_\Omega \boldsymbol{f} \cdot \boldsymbol{v} \, \mathrm{d}\boldsymbol{x}. \tag{6.5}$$

Eq. (6.4) may be written equivalently as a pair of equations by choosing $\boldsymbol{v} = \boldsymbol{O}$ and $q = 0$ in turn. For ease of exposition we consider only homogeneous Dirichlet boundary conditions. More general conditions may be dealt with in an analogous fashion. In order to describe sufficient conditions for the existence of a solution to (6.4) we introduce inner products $a(\cdot, \cdot)$ and $c(\cdot, \cdot)$ on V and W, respectively:

$$a(\boldsymbol{v}, \boldsymbol{w}) = \int_\Omega \nu \nabla \boldsymbol{v} \cdot \nabla \boldsymbol{w} \, \mathrm{d}\boldsymbol{x} \tag{6.6}$$

and

$$c(p, q) = \int_\Omega p \, q \, \mathrm{d}\boldsymbol{x}. \tag{6.7}$$

These inner products induce norms on V and W denoted by $\|\cdot\|_a$ and $\|\cdot\|_c$, respectively. The following facts [36] concerning B and b will be useful:

- there exists a positive constant C_B such that

$$|B(\boldsymbol{v}, \boldsymbol{w})| \leqslant C_B \|\boldsymbol{v}\|_a \, \|\boldsymbol{w}\|_a \quad \forall \boldsymbol{v}, \boldsymbol{w} \in V \tag{6.8}$$

-

$$|b(\boldsymbol{v}, q)| \leqslant \nu^{-1/2} \|\boldsymbol{v}\|_a \, \|q\|_c \quad \forall (\boldsymbol{v}, q) \in V \times W \tag{6.9}$$

- b satisfies an *inf–sup* condition: i.e. there exists a positive constant α_b such that

$$\sup_{\boldsymbol{v} \in V} \frac{|b(\boldsymbol{v}, q)|}{\|\boldsymbol{v}\|_a} \geqslant \alpha_b \|q\|_c \quad \forall q \in W \tag{6.10}$$

- B is coercive: i.e. owing to the vector field $\boldsymbol{U}$ being solenoidal there holds

$$B(\boldsymbol{v}, \boldsymbol{v}) = \|\boldsymbol{v}\|_a^2 \quad \forall \boldsymbol{v} \in V. \tag{6.11}$$

Under these conditions it follows [36] that there is a unique solution to (6.4).

6.1.2. Norm on $V \times W$

The usual choice of norm on the product space $V \times W$ is

$$(\boldsymbol{v}, q) \mapsto \left\{\|\boldsymbol{v}\|_a^2 + \|q\|_c^2\right\}^{1/2}. \tag{6.12}$$

It will be convenient to establish an equivalent norm for the space $V \times W$. Let $(e, E) \in V \times W$ be *arbitrary*. The pair $(\phi, \psi) \in V \times W$ is defined to be the Ritz projection of the residuals. That is,

$$a(\phi, v) + c(\psi, q) = B(e, v) + b(v, E) + b(e, q) \tag{6.13}$$

for all $(v, q) \in V \times W$. The existence and uniqueness of the pair (ϕ, ψ) follows from the continuity of the forms B and b. Therefore, we may define

$$|||(e, E)||| = \{\|\phi\|_a^2 + \|q\|_c^2\}^{1/2}. \tag{6.14}$$

The following result confirms that this quantity is a norm on $V \times W$ equivalent to the usual norm:

THEOREM 6.1. Under the foregoing assumptions and definitions, there exist positive constants k_1 and k_2 such that

$$k_1 |||(e, E)|||^2 \leqslant \|e\|_a^2 + \|E\|_c^2 \leqslant k_2 |||(e, E)|||^2 \tag{6.15}$$

where k_1 depends only on C_B and ν; and, k_2 depends only on C_B and α_b.

PROOF. Right-hand inequality. Making use of (6.10), (6.13), (6.8) and the Cauchy–Schwarz Inequality yields:

$$\|E\|_c \leqslant \frac{1}{\alpha_b} \{\|\phi\|_a + C_B \|e\|_a\}. \tag{6.16}$$

Using (6.11), (6.13) (with $q = -E$ and $v = e$) and the Cauchy–Schwarz Inequality:

$$\|e\|_a^2 \leqslant \|\phi\|_a \ \|e\|_a + \|\psi\|_c \ \|E\|_c. \tag{6.17}$$

From (6.16) and (6.17) one finds

$$\frac{1}{2} \|e\|_a^2 \leqslant \frac{1}{2} \left(\|\phi\|_a + \frac{C_B}{\alpha_b} \|\psi\|_c \right)^2 + \frac{1}{\alpha_b} \|\phi\|_a \ \|\psi\|_c. \tag{6.18}$$

Combining (6.18) with (6.16); once again using (6.18) gives the result with k_2 a constant depending on α_b and C_B.

Left-hand inequality. Using (6.13), (6.8), (6.9) and the Cauchy–Schwarz Inequality gives

$$\|\phi\|_a \leqslant C_B \|e\|_a + \nu^{-1/2} \|E\|_c. \tag{6.19}$$

Using (6.13) and (6.9) gives

$$\|\psi\|_c^2 = b(e, \psi) \leqslant \nu^{-1/2} \|e\|_a \ \|\psi\|_c. \tag{6.20}$$

Combining (6.19) and (6.20) yields the estimate claimed where k_1 depends only on C_B and ν. □

6.1.3. Discretization

Let $\mathcal{P}$ be a locally quasi-uniform partition of the domain Ω. Suppose that each element K is the image of an appropriate reference element $\widehat{K}$ under the usual type of transformation F_K. The basis functions on the element K are of the form

$$X_K = \text{Span}\{\widehat{v} \circ F_K^{-1} : \widehat{v} \in \mathbb{P}(p_X)\}^2 \tag{6.21}$$

and

$$M_K = \text{Span}\{\widehat{q} \circ F_K^{-1} : \widehat{q} \in \mathbb{P}'(p_M)\} \tag{6.22}$$

where $p_X \in \mathbb{N}$ and $p_M \in \mathbb{Z}^+$ and $\mathbb{P}$, $\mathbb{P}'$ are appropriate polynomial spaces on the reference element. The global finite element subspaces X and M are constructed in the usual manner so that the inclusion $X \times M \subset V \times W$ holds. The finite element approximation to (6.4) is then:

Find $(u_X, p_X) \in X \times M$ such that for all $(v_X, q) \in X \times M$

$$B(u_X, v_X) + b(v_X, p_X) + b(u_X, q) = F(v_X). \tag{6.23}$$

A few remarks concerning the construction of the finite element subspace $X \times M$ are in order. It will have been noted that there was no requirement for a discrete *inf–sup* condition to hold. The stability

of the discretization scheme does not affect the a posteriori error analysis since only stability of the underlying continuous problem is used. Of course, the indiscriminate use of unstable discretizations is not recommended.

6.1.4. A posteriori error analysis

The argument closely parallels the discussion of the element residual method with equilibration from Section 5 with which we assume the reader is familiar. Throughout we shall use the same notations and conventions.

Mesh dependent forms and spaces

The local velocity space on each subdomain $K \in \mathcal{P}$ is

$$\boldsymbol{V}_K = \{\boldsymbol{v} \in H^1(K) \times H^1(K) : \boldsymbol{v} = \boldsymbol{O} \text{ on } \partial\Omega \cap \partial K\} \tag{6.24}$$

and the local pressure space is

$$W_K = L_2(K). \tag{6.25}$$

The bilinear forms $B_K : \boldsymbol{V}_K \times \boldsymbol{V}_K \to \mathbb{R}$ and $b_K : \boldsymbol{V}_K \times W_K \to \mathbb{R}$ are defined as follows:

$$b_K(\boldsymbol{v}, q) = -\int_K q \, \mathbf{div}\, \boldsymbol{v} \, \mathrm{d}\boldsymbol{x} \tag{6.26}$$

and

$$B_K(\boldsymbol{v}, \boldsymbol{w}) = \int_K \{\nu \nabla \boldsymbol{v} : \nabla \boldsymbol{w} + \boldsymbol{w} \cdot (\boldsymbol{U} \cdot \nabla)\boldsymbol{v}\} \, \mathrm{d}\boldsymbol{x}. \tag{6.27}$$

Similarly, $F_K : \boldsymbol{V}_K \to \mathbb{R}$ is defined by

$$F_K(\boldsymbol{v}) = \int_K \boldsymbol{f} \cdot \boldsymbol{v} \, \mathrm{d}\boldsymbol{x}. \tag{6.28}$$

Hence, for $\boldsymbol{v}, \boldsymbol{w} \in \boldsymbol{V}$ and $q \in W$

$$b(\boldsymbol{v}, q) = \sum_{K \in \mathcal{P}} b_K(\boldsymbol{v}_K, q_K) \tag{6.29}$$

$$B(\boldsymbol{v}, \boldsymbol{w}) = \sum_{K \in \mathcal{P}} B_K(\boldsymbol{v}_K, \boldsymbol{w}_K) \tag{6.30}$$

and

$$F(\boldsymbol{v}) = \sum_{K \in \mathcal{P}} F_K(\boldsymbol{v}_K). \tag{6.31}$$

The broken space $\boldsymbol{V}(\mathcal{P}) \times W(\mathcal{P})$ is defined by

$$\boldsymbol{V}(\mathcal{P}) \times W(\mathcal{P}) = \{(\boldsymbol{v}, q) \in [L_2(\Omega)]^3 : (\boldsymbol{v}, q)|_K \in \boldsymbol{V}_K \times W_K \quad \forall K \in \mathcal{P}\}. \tag{6.32}$$

Examining the previous notations reveals that $W(\mathcal{P}) = W$. As before, we consider the space of continuous linear functionals τ on $\boldsymbol{V}(\mathcal{P}) \times W(\mathcal{P})$ that vanish on the subspace $\boldsymbol{V} \times W$. Let $\mathbb{H}(\mathbf{div}, \Omega)$ be the space

$$\mathbb{H}(\mathbf{div}, \Omega) = \{\boldsymbol{A} \in L_2(\Omega)^{2\times 2} : \mathbf{div}\, \boldsymbol{A} \in L_2(\Omega)^2\} \tag{6.33}$$

equipped with norm

$$\|\boldsymbol{A}\|_{\mathbb{H}(\mathbf{div}, \Omega)} = \{\|\boldsymbol{A}\|^2_{L_2(\Omega)} + \|\mathbf{div}\, \boldsymbol{A}\|^2_{L_2(\Omega)}\}^{1/2}. \tag{6.34}$$

The following result generalizes Theorem 5.1:

THEOREM 6.2. A continuous linear functional τ on the space $\boldsymbol{V}(\mathcal{P}) \times W(\mathcal{P})$ vanishes on the subspace $\boldsymbol{V} \times W$ if and only if there exists $\boldsymbol{A} \in \mathbb{H}(\mathbf{div}, \Omega)$ such that

$$\tau[(\boldsymbol{v}, q)] = \sum_{K \in \mathcal{P}} \oint_{\partial K} \boldsymbol{n}_K \cdot \boldsymbol{A} \cdot \boldsymbol{v}_K \, ds \tag{6.35}$$

where $\boldsymbol{n}_K$ denotes the unit outward normal on the boundary of K.

PROOF. Essentially identical with the proof of Theorem 5.1. □

Thanks to this result we may refer to the functional τ as belonging to the space $\mathbb{H}(\mathbf{div}, \Omega)$.

Error analysis

Let $(\boldsymbol{u}_X, p_X) \in \boldsymbol{X} \times M$ be the finite element approximation to $(\boldsymbol{u}, p) \in \boldsymbol{V} \times W$. In view of the inclusion $\boldsymbol{X} \times M \subset \boldsymbol{V} \times W$, the discretization error $(\boldsymbol{e}, E)$

$$\boldsymbol{e} = \boldsymbol{u} - \boldsymbol{u}_X; \qquad E = p - p_X \tag{6.36}$$

belongs to the space $\boldsymbol{V} \times W$. Define a pair $(\boldsymbol{\phi}, \psi) \in \boldsymbol{V} \times W$ such that:

$$a(\boldsymbol{\phi}, \boldsymbol{v}) + c(\psi, q) = B(\boldsymbol{e}, \boldsymbol{v}) + b(\boldsymbol{v}, E) + b(\boldsymbol{e}, q) \tag{6.37}$$

for all $(\boldsymbol{v}, q) \in \boldsymbol{V} \times W$. Theorem 6.1 reveals that the norm of the discretization error is given by

$$|||(\boldsymbol{e}, E)|||^2 = \|\boldsymbol{\phi}\|_a^2 + \|\psi\|_c^2 . \tag{6.38}$$

The problem is therefore to estimate $\|\boldsymbol{\phi}\|_a$ and $\|\psi\|_c$ numerically. As before, we reduce the single global problem (6.37) into a sequence of independent problems posed locally over each element.

Inter-element boundary flux

The stress tensor $\boldsymbol{\rho}(\boldsymbol{v}, q)$ is defined by

$$\rho_{ij}(\boldsymbol{v}, q) = \nu \frac{\partial v_i}{\partial x_j} - q\delta_{ij} \tag{6.39}$$

where δ_{ij} is the Kronecker symbol. The interelement fluxes played a vital role in Section 5. The normal flux on the boundary of element K is given by $(n_K)_{ij}\rho_j$. Let σ_K be as defined in (5.24) and $\boldsymbol{g}_\gamma : \gamma \to \mathbb{R}^2$ be smooth functions on the edges γ on the interior of the domain. The approximation to the flux on the boundary of element K is given by $\boldsymbol{g}_K$ where

$$\boldsymbol{g}_K = \sigma_K \boldsymbol{g}_\gamma \quad \text{on } \gamma \subset \partial K. \tag{6.40}$$

As before, the notation $[\cdot]$ will be used to denote differences in quantities across element boundaries as before. The following identity valid for $\boldsymbol{v} \in \boldsymbol{V}(\mathcal{P})$ is analogous to Eq. (5.31):

$$\sum_{K \in \mathcal{P}} \oint_{\partial K} \boldsymbol{g}_K \cdot \boldsymbol{v} \, ds = \sum_{\gamma \in \partial \mathcal{P}} \int_\gamma \boldsymbol{g}_\gamma \cdot [\boldsymbol{v}] \, ds. \tag{6.41}$$

Localization

The next step is to decompose the global problem (6.37) into local problems posed over the elements. Firstly, the unknowns $(\boldsymbol{u}, p)$ in (6.37) are replaced by appealing to (6.4):

$$\begin{aligned} a(\boldsymbol{\phi}, \boldsymbol{w}) + c(\psi, q) &= B(\boldsymbol{e}, \boldsymbol{w}) + b(\boldsymbol{w}, E) + b(\boldsymbol{e}, q) \\ &= \sum_{K \in \mathcal{P}} \{F_K(\boldsymbol{w}) - B_K(\boldsymbol{u}_X, \boldsymbol{w}) - b_K(\boldsymbol{w}, p_X) - b_K(\boldsymbol{u}_X, q)\} . \end{aligned} \tag{6.42}$$

The global space $\boldsymbol{V} \times W$ is decomposed into functions that are smooth on each of the elements but not necessarily continuous between elements. The functional given by (6.42) is then extended to the broken space $\boldsymbol{V}(\mathcal{P}) \times W(\mathcal{P})$. For any $(\boldsymbol{w}, q) \in \boldsymbol{V}(\mathcal{P}) \times W(\mathcal{P})$ define the linear functional $\mathcal{R} : \boldsymbol{V}(\mathcal{P}) \times W(\mathcal{P}) \to \mathbb{R}$ by

$$\mathcal{R}[(w,q)] = \sum_{K\in\mathcal{P}} \left\{ F_K(w) - B_K(u_X, w) - b_K(w, p_X) - b_K(u_X, q) + \oint_{\partial K} g_K \cdot w_K \, ds \right\} - \sum_{\gamma\in\partial\mathcal{P}} \int_\gamma g_\gamma \cdot [w] \, ds \tag{6.43}$$

so that whenever $(w,q) \in V \times W$

$$\mathcal{R}[(w,q)] = a(\phi, w) + c(\psi, q). \tag{6.44}$$

LEMMA 6.3. Under the above notations and conventions, there exists $\mu_* \in \mathbb{H}(\mathbf{div}, \Omega)$ *such that for all* $(w,q) \in V(\mathcal{P}) \times W(\mathcal{P})$

$$\mu_*[(w,q)] = \sum_{\gamma\in\partial\mathcal{P}} \int_\gamma g_\gamma \cdot [w] \, ds. \tag{6.45}$$

PROOF. The right-hand side of Eq. (6.45) vanishes on $V \times W$. The result then follows immediately from Theorem 6.2. □

Applying Lemma 6.3 yields:

$$\mathcal{R}[(w,q)] = \sum_{K\in\mathcal{P}} \left\{ F_K(w) - B_K(u_X, w) - b_K(w, p_X) - b_K(u_X, q) + \oint_{\partial K} g_K \cdot w_K \, ds \right\} - \mu_*[(w,q)] \tag{6.46}$$

for all $(w,q) \in V(\mathcal{P}) \times W(\mathcal{P})$.

Variational analysis

Introduce the Lagrangian functional $\mathcal{L} : V(\mathcal{P}) \times W(\mathcal{P}) \times \mathbb{H}(\mathbf{div}, \Omega) \to \mathbb{R}$ given by

$$\mathcal{L}[(w,q), \mu] = \frac{1}{2}\{a(w,w) + c(q,q)\} - \mathcal{R}[(w,q)] - \mu[(w,q)] \tag{6.47}$$

so that

$$\sup_{\mu\in\mathbb{H}(\mathbf{div},\Omega)} \mathcal{L}[(w,q), \mu] = \begin{cases} \frac{1}{2}\{a(w,w) + c(q,q)\} - \mathcal{R}[(w,q)] & \text{if } (w,q) \in V \times W \\ +\infty & \text{otherwise} \end{cases} \tag{6.48}$$

and analogously to (5.39),

$$\frac{1}{2}\{a(w,w) + c(q,q)\} - \mathcal{R}[(w,q)] \geqslant -\frac{1}{2} |||(e, E)|||^2 \tag{6.49}$$

for any $(w,q) \in V \times W$. Therefore,

$$\begin{aligned} -\frac{1}{2} |||(e,E)|||^2 &= \inf_{(w,q)\in V(\mathcal{P})\times W(\mathcal{P})} \sup_{\mu\in\mathbb{H}(\mathbf{div},\Omega)} \mathcal{L}[(w,q),\mu] \\ &= \sup_{\mu\in\mathbb{H}(\mathbf{div},\Omega)} \inf_{(w,q)\in V(\mathcal{P})\times W(\mathcal{P})} \mathcal{L}[(w,q),\mu] \\ &\geqslant \inf_{(w,q)\in V(\mathcal{P})\times W(\mathcal{P})} \mathcal{L}[(w,q),\mu_*] \\ &= \sum_{K\in\mathcal{P}} \inf_{w_K\in V_K} \left\{ \frac{1}{2} a(w_K, w_K) - F_K(w) + B_K(u_X, w) + b_K(w, p_X) \right. \\ &\qquad \left. - \oint_{\partial K} g_K \cdot w_K \, ds - \frac{1}{2} \|\mathbf{div}\, u_X\|_{c,K}^2 \right\} \end{aligned} \tag{6.50}$$

where the infimum over the space $W(\mathcal{P})$ has been computed explicitly. As usual, the order of the inf–sup may be changed since a saddle point is obtained when the multiplier μ is the true interelement flux. This choice is a valid multiplier as can be seen by applying Theorem 6.2. Summarising, we have shown:

THEOREM 6.4. Let $J_K : V_K \to \mathbb{R}$ be the quadratic functional

$$J_K(w_K) = \frac{1}{2}\, a(w_K, w_K) - F_K(w) + B_K(u_X, w) + b_K(w, p_X) - \oint_{\partial K} g_K \cdot w_K \, ds. \tag{6.51}$$

Then,

$$|||(e, E)|||^2 \leqslant \sum_{K \in \mathcal{P}} \left\{ -2 \inf_{w_K \in V_K} J_K(w_K) + \|\mathbf{div}\, u_X\|^2_{c,K} \right\}. \tag{6.52}$$

6.1.5. Analysis of local error residual problems

The analysis has led to problems on each subdomain of the form

$$\inf_{w_K \in V_K} J_K(w_K). \tag{6.53}$$

Suppose for a moment that a minimum exists, then the minimizing element is characterized by finding $\phi_K \in V_K$ such that

$$a(\phi_K, v) = F_K(v) - B_K(u_X, v) - b_K(v, p_X) + \oint_{\partial K} g_K \cdot v \, ds \quad \forall v \in V_K. \tag{6.54}$$

This problem decouples into a pair of Poisson type problem with Neumann data. The result of the foregoing analysis has been that one can obtain a local a posteriori error estimator for the Stokes problem by solving auxiliary Neumann type problems for the residual in the momentum equations. The contribution from the incompressibility constraint may be calculated explicitly. This has a considerable impact in the computation of the error estimator since one need not solve a local Stokes type problem as, for example, is the case with [21] and [56]. The approach suggested above could be used in the context of those papers, yielding significantly simpler local problems.

The necessary and sufficient conditions for the existence of a minimum are that the data satisfy the following compatibility or *equilibration* condition:

$$0 = F_K(\theta) - B_K(u_X, \theta) - b_K(\theta, p_X) + \oint_{\partial K} g_K \cdot \theta \, ds \tag{6.55}$$

for all $\theta \in \mathrm{Ker}[a, V_K]$ where

$$\mathrm{Ker}[a, V_K] = \{ \theta \in V_K : a_K(w, \theta) = 0 \; \forall w \in V_K \}. \tag{6.56}$$

If the subdomain K lies on the boundary $\partial\Omega$ then the local problem (6.54) will be subject to a homogeneous Dirichlet condition on a portion of their boundaries and thus will be automatically well posed. However, elements away from the boundary are subject to pure Neumann conditions and the null space will consist of the rigid body motions

$$\mathrm{Ker}[a, V_K] = \mathrm{Span}\,\{\theta_1, \theta_2\} \tag{6.57}$$

where

$$\theta_1 = \begin{bmatrix} 1 \\ 0 \end{bmatrix}, \qquad \theta_2 = \begin{bmatrix} 0 \\ 1 \end{bmatrix}. \tag{6.58}$$

The equilibration theory and procedures described in Section 5 can be applied to each of the equations in the system (6.55) in turn, giving boundary data satisfying condition (6.55). It is worth noting that there is only an equilibration requirement for the momentum equations and not for the incompressibility constraint.

In summary, the numerical procedure is to first calculate equilibrated boundary data that ensures the local problems (6.54) are well posed. These problems are then solved numerically using the approximate subspaces discussed in Section 5 giving an approximate solution ϕ_K. The process then yields an a posteriori error estimate η_K on the subdomain K

$$\eta_K = \{\|\phi_K\|_{a,K}^2 + \|\mathbf{div}\, u_X\|_{c,K}^2\}^{1/2}. \tag{6.59}$$

A global error estimate may be obtained by summing the local estimates. Theorem 6.4 guarantees that the estimate bounds the true error $|||(e, E)|||$ from above, provided that the local approximate subspace provides sufficient resolution.

6.1.6. *Summary and conclusions*

An important point of the analysis is that one does not have to solve a local Stokes problem, it is sufficient to solve a pair of independent local Poisson problems. This means that one is solving a system of two equations (since the residual corresponding to the incompressibility condition can be treated directly) rather than the system of three coupled equations needed for other techniques. Importantly, when one comes to construct the basis functions used in approximating the local problems, there is no issue of stability (inf–sup) conditions. These conditions can be quite problematic if one is trying to solve a local Stokes problem using an appropriate space, requiring a careful stability analysis [21]. This issue does not arise with the approach presented here. These features make the computation of the estimators less expensive, and more easily applicable to general finite element schemes for Stokes' type problems.

Although the analysis suggests that the boundary data for the local residual type problems should be chosen to satisfy an equilibration condition, the above comments are equally valid whether one is equilibrating the boundary fluxes or not. Of course, one loses the upper bound property if the equilibration condition is not satisfied, but this may not be of primary importance in some applications.

One can question the usefulness of an upper bound in the unorthodox energy norm $|||\cdot|||$ albeit equivalent with the H^1 type norms. The analysis could be used to obtain an estimator in H^1 norm. The energy of the actual solution can be estimated in the same $|||\cdot|||$, being computed at the same time as the error estimator by modifying the right-hand sides used in the error estimation process (after omitting the terms $B_K(u_X, v)$ and $b(v, p_X)$ in Eq. (6.54)). The process yields a sufficiently good estimate for practical purposes and may be used to scale the error estimator giving an estimate of relative error. It is therefore possible to perform rigorous and quantitative error control for Stokes' and Oseen type problems.

Summarizing, it has been shown that the equilibration principle carries over from the scalar case along with the basic steps in the analysis. In addition, the procedure for the treatment of side conditions, such as the incompressibility constraints, has been outlined.

6.2. *Incompressible Navier–Stokes equations*

Following Oden et al. [49] the analysis for the Stokes and Oseen problem will be extended to the incompressible Navier–Stokes equations with small data. Let $D : V \times V \times V \to \mathbb{R}$ be the trilinear form

$$D(u, v, w) = \int_\Omega u \cdot \nabla v \cdot w \, \mathrm{d}x. \tag{6.60}$$

The form D is continuous in the sense that there exists a constant C_B such that

$$D(u, v, w) \leqslant C_D |u|_{H^1(\Omega)} |v|_{H^1(\Omega)} |w|_{H^1(\Omega)}. \tag{6.61}$$

We shall assume that C_D is the best possible constant such that (6.61) holds. The incompressible Navier–Stokes problem is to

Find $(u, p) \in V \times W$ such that for all $(v, q) \in V \times W$

$$a(u, v) + b(v, p) + D(u, u, v) + b(u, q) = F(v) \tag{6.62}$$

where $F : V \to \mathbb{R}$ is the linear functional

$$F(v) = \int_\Omega f \cdot v \, \mathrm{d}x. \tag{6.63}$$

The problem (6.63) is known [36] to possess a unique solution whenever the data is sufficiently small. In particular, if

$$|F(v)| \leqslant \theta \frac{\nu^2}{C_D} |v|_{H^1(\Omega)} \quad \forall v \in V \tag{6.64}$$

for some fixed number $\theta \in [0, 1)$ then there is a unique solution $u \in V$ satisfying

$$|u|_{H^1(\Omega)} \leqslant \theta \frac{\nu}{C_D}. \tag{6.65}$$

The finite element approximation to (6.62) is then:

Find $(u_X, p_X) \in X \times M$ such that for all $(v_X, q) \in X \times M$

$$a(u_X, v) + b(v, p_X) + D(u_X, u_X, v) + b(u_X, q) = F(v). \tag{6.66}$$

The finite element subspaces X and M are constructed as described previously. It will be assumed that the finite element approximation u_X converges to the velocity u as the partition is refined.

6.2.1. A posteriori error analysis

The basic idea behind the extension of the analysis for the Stokes and Oseen type problem to the Navier–Stokes equations was suggested by Wu [58]. Let (e, E) be the error in the finite element approximation and define a pair $(\phi, \psi) \in V \times W$ to be the Ritz projection of the *modified residuals*

$$a(\phi, v) + c(\psi, q) = a(e, v) + b(v, E) + b(e, q) + \delta(u, u_X, v) \tag{6.67}$$

for all $(v, q) \in V \times W$, where

$$\delta(u, u_X, v) = D(u, u, v) - D(u_X, u_X, v). \tag{6.68}$$

The data on the right-hand side of (6.67) defines a continuous linear functional and so the pair (ϕ, ψ) exists and is unique so that we may define

$$|||(e, E)||| = \{\|\phi\|_a^2 + \|q\|_c^2\}^{1/2}. \tag{6.69}$$

The idea echoes the basic step (6.13) used before. However, there is a significant difference: before the pair (e, E) was arbitrary but now it is essential that (e, E) be the error in the finite element approximation. Furthermore, owing to the presence of the non-linear term $\delta(\cdot, \cdot, \cdot)$ one cannot directly apply Theorem 6.1. However, suppose for a moment that

$$k_1 |||(e, E)|||^2 \leqslant \|e\|_a^2 + \|E\|_c^2 \leqslant k_2 |||(e, E)|||^2 \tag{6.70}$$

for non-negative quantities k_1 and k_2. Let $\phi_K \in V_K$ be such that

$$a(\phi_K, v) = F_K(v) - a_K(u_X, v) - b_K(v, p_X) - D_K(u_X, u_X, v) + \oint_{\partial K} g_K \cdot v \, ds \tag{6.71}$$

for all $v \in V_K$ where the boundary data has been chosen so that the equilibration condition

$$0 = F_K(\theta) - B_K(u_X, \theta) - b_K(\theta, p_X) - D_K(u_X, u_X, \theta) + \oint_{\partial K} g_K \cdot \theta \, ds \tag{6.72}$$

is satisfied for all $\theta \in \mathrm{Ker}[a, V_K]$. The local a posteriori error estimate on element K is taken to be

$$\eta_K = \{\|\phi_K\|_{a,K}^2 + \|\mathbf{div}\, u_X\|_{c,K}^2\}^{1/2}. \tag{6.73}$$

An argument identical to the linear case reveals that

$$|||(e, E)|||^2 \leqslant \sum_{K \in \mathcal{P}} \eta_K^2 = \eta^2. \tag{6.74}$$

Consequently, thanks to the equivalence (6.70), η provides an error estimator for the incompressible Navier–Stokes problem with small data. It remains to prove the equivalence (6.70).

PROOF. First, we obtain bounds for the form $\delta(\cdot,\cdot,\cdot)$. Suppose $v \in V$ then

$$\begin{aligned}
\delta(u, u_X, v) &= D(u, u, v) - D(u_X, u_X, v) \\
&= D(u, e, v) + D(e, u_X, v) \\
&\leqslant C_D\{|u|_{H^1(\Omega)}|e|_{H^1(\Omega)}|v|_{H^1(\Omega)} + |e|_{H^1(\Omega)}|u_X|_{H^1(\Omega)}|v|_{H^1(\Omega)}\} \\
&\leqslant C_D\{2|u|_{H^1(\Omega)}|e|_{H^1(\Omega)}|v|_{H^1(\Omega)} + |e|^2_{H^1(\Omega)}|v|_{H^1(\Omega)}\} \\
&\leqslant C_D\left\{2\theta\frac{\nu}{C_D}|e|_{H^1(\Omega)}|v|_{H^1(\Omega)} + |e|^2_{H^1(\Omega)}|v|_{H^1(\Omega)}\right\} \\
&= \left\{2\theta + \frac{C_D}{\nu}|e|_{H^1(\Omega)}\right\}\|e\|_a\|v\|_a
\end{aligned} \tag{6.75}$$

where (6.61) and (6.65) have been used. If $v = e$ then a sharper bound may be obtained by first noting [36, Eq. (6), p. 285]

$$D(u, e, e) = 0 \tag{6.76}$$

and then following similar steps to before giving

$$\delta(u, u_X, e) \leqslant \left\{\theta + \frac{C_D}{\nu}|e|_{H^1(\Omega)}\right\}\|e\|_a^2. \tag{6.77}$$

Left-hand inequality. Using (6.9), (6.67) and (6.75) gives

$$\|\phi\|_a \leqslant \nu^{-1/2}\|E\|_c + \left\{3 + \frac{C_D}{\nu}|e|_{H^1(\Omega)}\right\}\|e\|_a. \tag{6.78}$$

Using (6.9) and (6.67) gives

$$\|\psi\|_c \leqslant \nu^{-1/2}\|e\|_a \tag{6.79}$$

and hence

$$k_1\left\{1 + O(|e|_{H^1(\Omega)})\right\}|||(e, E)|||^2 \leqslant \|e\|_a^2 + \|E\|_c^2 \tag{6.80}$$

where k_1 depends on C_D and ν.

Right-hand inequality. Using (6.10), (6.67) and (6.75) gives

$$\begin{aligned}
\alpha_b\|E\|_c &\leqslant \sup_{v\in V}\frac{|b(v, q)|}{\|v\|_a} \\
&\leqslant \sup_{v\in V}\frac{1}{\|v\|_a}|a(\phi, v) - a(e, v) - \delta(u, u_X, v)| \\
&\leqslant \|\phi\|_a + \|e\|_a\left\{3 + \frac{C_D}{\nu}|e|_{H^1(\Omega)}\right\}.
\end{aligned} \tag{6.81}$$

Equally well, using (6.9), (6.67) and (6.77) gives

$$\begin{aligned}
\|e\|_a^2 &= a(\phi, e) - b(e, E) - \delta(u, u_X, e) \\
&= a(\phi, e) - c(\psi, E) - \delta(u, u_X, e) \\
&\leqslant \|\phi\|_a\|e\|_a + \|\psi\|_c\|E\|_c + \left\{\theta + \frac{C_D}{\nu}|e|_{H^1(\Omega)}\right\}\|e\|_a^2.
\end{aligned} \tag{6.82}$$

Since $|e|_{H^1(\Omega)} \to 0$, we may choose $\epsilon > 0$ sufficiently small that

$$\theta + \frac{\epsilon^2}{2} + \frac{C_D}{\nu}|e|_{H^1(\Omega)} = \theta' < 1. \tag{6.83}$$

Hence, from (6.82) and (6.81) we obtain

$$\|e\|_a^2 \leqslant \frac{1}{\alpha_b}\|\phi\|_a\|\psi\|_c + \|e\|_a\left\{\|\phi\|_a + \frac{1}{\alpha_b}\left(3 + \frac{C_D}{\nu}|e|_{H^1(\Omega)}\right)\right\} + \|e\|_a^2\left\{\theta + \frac{C_D}{\nu}|e|_{H^1(\Omega)}\right\}$$
$$\leqslant \frac{1}{\alpha_b}\|\phi\|_a\|\psi\|_c + \theta'\|e\|_a^2 + \frac{1}{2\epsilon^2}\left\{\|\phi\|_a + \frac{1}{\alpha_b}\left(3 + \frac{C_D}{\nu}|e|_{H^1(\Omega)}\right)\right\}^2. \tag{6.84}$$

Hence, for $|e|_{H^1(\Omega)}$ sufficiently small

$$\|e\|_a^2 \leqslant C(\alpha_b, \nu, C_D, |e|_{H^1(\Omega)}, \theta)\{\|\phi\|_a^2 + \|\psi\|_c^2\} \tag{6.85}$$

and using (6.77) gives

$$\|\psi\|_c^2 \leqslant C(\alpha_b, \nu, C_D, |e|_{H^1(\Omega)}, \theta)\{\|\phi\|_a^2 + \|\psi\|_c^2\} \tag{6.86}$$

and the result follows. □

6.3. *Variational inequalities*

A class of variational inequalities describing the flow of an ideal fluid through an unsaturated porous medium [41] will be considered. The strong form or *linear complementarity problem* governing this situation is to find u such that on the domain Ω there holds

$$-\Delta u \geqslant f; \quad u \geqslant \psi; \quad (\Delta u + f)(u - \psi) \equiv 0 \tag{6.87}$$

subject to Dirichlet conditions on the boundary $\partial\Omega$. Evidently, the chief feature is the presence of the inequality conditions. Previously, error estimators have been obtained by solving local problems analogous to the original global problem. The type of local problem for the error estimator that might be derived from the system (6.87) is unclear. Naturally, one may expect to obtain the same type of complementarity condition as (6.87) on the interior of the elements but appropriate boundary conditions to impose on the boundaries of the elements is not obvious. However, by following the basic idea used in Section 5 the correct formulations of the local problems will emerge. The approach is based on [8].

6.3.1. *Model problem*

Let $\Omega \subset \mathbb{R}^2$ be an open bounded domain with smooth boundary $\partial\Omega$. For sufficiently smooth data f, ψ and u_0 consider the variational inequality:

Find $u \in \mathcal{K}$ such that

$$B(u, v - u) \geqslant F(v - u) \quad \forall v \in \mathcal{K} \tag{6.88}$$

where $\mathcal{K}$ is the convex set

$$\mathcal{K} = \{v \in H^1(\Omega) : v \geqslant \psi \text{ on } \Omega \text{ and } v = u_0 \text{ on } \partial\Omega\} \tag{6.89}$$

and $B : \mathcal{K} \times \mathcal{K} \to \mathbb{R}$ and $F : \mathcal{K} \to \mathbb{R}$ are the usual bilinear and linear forms

$$B(u, v) = \int_\Omega \nabla u \cdot \nabla v \, dx; \qquad F(v) = \int_\Omega f v \, dx. \tag{6.90}$$

Let $\mathcal{P}$ be a regular partition of Ω, X be a finite element space on the partition and let $\mathcal{K}_X = \mathcal{K} \cap X$. The discretised version of (6.88) is:

Find $u_X \in \mathcal{K}_X$ such that

$$B(u_X, v_X - u_X) \geqslant F(v_X - u_X) \quad \forall v_X \in \mathcal{K}_X. \tag{6.91}$$

It is known [41] that if ψ, f and u_0 are sufficiently smooth and $\psi \leqslant u_0$ on $\partial\Omega$ then there exists a unique solution of both the continuous and discrete problems (6.88) and (6.91).

6.3.2. *A posteriori error analysis*

Let $\mathcal{W}$ denote the convex set

$$\mathcal{W} = \{w : w = v - u_X, \text{ for some } v \in \mathcal{K}\} \tag{6.92}$$

or equally well

$$\mathcal{W} = \{w \in H^1(\Omega) : w + u_X \geqslant \psi \text{ in } \Omega \text{ and } w + u_X = 0 \text{ on } \partial\Omega\}. \tag{6.93}$$

From the variational inequality (6.88) it follows that the error e is characterised as the solution of the problem:

Find $e \in \mathcal{W}$ such that

$$B(e, w - e) \geqslant F(w - e) - B(u_X, w - e) \quad \forall w \in \mathcal{W}. \tag{6.94}$$

The existence of a unique solution of (6.94) follows immediately from the existence and uniqueness of the solution u of the original problem (6.88). The relation (6.94) is the analogue of the residual equation from which error estimators were obtained for linear problems. An alternative form equivalent to (6.94) is:

Find $e \in \mathcal{W}$ such that

$$J(e) \leqslant J(w) \quad \forall w \in \mathcal{W} \tag{6.95}$$

where

$$J(w) = \frac{1}{2} B(w, w) - F(w) + B(u_X, w). \tag{6.96}$$

Now, since $\mathcal{K}_X \subset \mathcal{K}$, there follows from (6.88)

$$\begin{aligned} |||e|||^2 &= B(e, e) \\ &= B(u, e) - B(u_X, e) \\ &\leqslant F(e) - B(u_X, e) \\ &= -J(e) + \frac{1}{2} |||e|||^2 \end{aligned} \tag{6.97}$$

and hence from (6.95) we obtain

$$\inf_{w \in \mathcal{W}} J(w) = J(e) \leqslant -\frac{1}{2} |||e|||^2. \tag{6.98}$$

6.3.3. Localization

The notations and conventions used in Section 5 will be used. In addition, the space $\mathcal{W}(\mathcal{P})$ is defined by

$$\mathcal{W}(\mathcal{P}) = \{w \in H^1(\mathcal{P}) : w + u_X \geqslant \psi \text{ in } \Omega \text{ and } w + u_X = 0 \text{ on } \partial\Omega\}. \tag{6.99}$$

The estimate (6.98) is analogous to the basic result (5.23) used in Section 5 to derive the a posteriori error estimates. The steps leading from (5.23) to Theorem 5.3 may be repeated giving:

THEOREM 6.5. Let $J_K : \mathcal{W}_K \to \mathbb{R}$ be the quadratic functional

$$J_K(w) = \frac{1}{2} B_K(w, w) - F_K(w) + B_K(u_X, w) - \oint_{\partial K} g_K w \, ds. \tag{6.100}$$

Then

$$|||e|||^2 \leqslant -2 \inf_{w \in \mathcal{W}(\mathcal{P})} \sum_{K \in \mathcal{P}} J_K(w). \tag{6.101}$$

As anticipated, Theorem 6.5 shows that in order to obtain a bound on the discretisation error one need only construct elements of the space $\mathcal{W}(\mathcal{P})$. The set $\mathcal{W}(\mathcal{P})$ does not impose any interelement restrictions on the choice. Consequently, the statement (6.101) reduces to a series of local problems of the form:

Find $\phi_K \in \mathcal{W}_K$ such that

$$J_K(\phi_K) \leqslant J_K(w_K) \quad \forall w_K \in \mathcal{W}_K \tag{6.102}$$

where, with a slight abuse of notation, we define

$$\mathcal{W}_K = \{w_K \in H^1(K) : w_K + u_X \geqslant \psi \text{ on } K \text{ and } w_K + u_X = 0 \text{ on } \partial K \cap \partial\Omega\}. \tag{6.103}$$

The main advantage associated with dealing with local problems is that the computational cost is negligible in comparison with the expense entailed in obtaining u_X. This contrasts with the related ideas presented in [37] where it is necessary to solve a problem of comparable complexity to (6.91) to obtain an a posteriori error estimate.

6.3.4. Analysis of local problems

Consider the local problem (6.102) on an element K in the interior of the domain Ω. The strong form of the problem consists of a linear complementarity condition on the interior of the element

$$-\Delta\phi \geqslant f + \Delta u_X; \qquad \phi \geqslant \psi - u_X; \qquad (\Delta\phi + f + \Delta u_X)(\phi - \psi + u_X) \equiv 0 \tag{6.104}$$

supplemented with a *linear complementarity condition on the boundary* ∂K

$$\frac{\partial\phi}{\partial n} \geqslant g_K - \frac{\partial u_X}{\partial n}; \qquad \phi \geqslant \psi - u_X; \qquad \left(\frac{\partial\phi}{\partial n} \geqslant g_K - \frac{\partial u_X}{\partial n}\right)(\phi - \psi + u_X) \equiv 0 \quad \text{on } \partial K. \tag{6.105}$$

Therefore, we conclude that the appropriate local error residual problem to be solved for the error consists of the weak form of the problem specified by conditions (6.104) and (6.105).

The local problem (6.102) is automatically well posed if the element K lies on the boundary of the domain Ω. However, if the element lies on the interior of the domain, then it is subject the linear complementarity conditions (6.105) on the whole of the boundary ∂K and it may be that the problem possesses no solution. A routine application of arguments found in [37] shows that the local problem has a unique solution if and only if the condition

$$-F_K(1) + B_K(u_X, 1) - \oint_{\partial K} g_K \, ds > 0 \tag{6.106}$$

is satisfied. If the inequality is not strict then the solution is unique up to the addition of a positive constant.

The condition (6.106) plays the role of the equilibration condition in Section 5 and provides a criterion for selecting the boundary data g_K. The discussion in Section 5.4 may be extended to the case of an inequality equilibration condition and used to deduce that it is indeed possible to construct boundary data so that condition (6.106) is satisfied.

The approximate solution of the local variational problems (6.102) is complicated by the unilateral condition in the definition (6.103). In particular, one cannot easily use a p-version finite element method to approximate the local problem. An alternative is to subdivide the element K into a small number of subelements and compute a local h-version finite element approximation of the local problem. Numerical examples based on this method will be found in [8].

Acknowledgement

The support of a portion of the work of M.A. by a TICAM Research Fellowship and of J.T.O. by a contract from the Office of Naval Research is gratefully acknowledged.

References

[1] R.A. Adams, Sobolev Spaces, Pure and Applied Mathematics Series, Vol. 65 (Academic Press, 1978).
[2] M. Ainsworth, The performance of Bank–Weiser's error estimator for quadrilateral finite elements, Numer. Methods PDE 10 (1994) 609–623.
[3] M. Ainsworth, The influence and selection of subspaces for a posteriori error estimators, Numer. Math. (to appear).

[4] M. Ainsworth and A.W. Craig, A posteriori error estimators in the finite element method, Numer. Math. 60 (1991) 429–463.
[5] M. Ainsworth and J.T. Oden, A posteriori error estimators for second order elliptic systems Part 2. An optimal order process for calculating self equilibrating fluxes. Comput. Math. Appl. 26 (1993) 75–87.
[6] M. Ainsworth and J.T. Oden, A unified approach to a posteriori error estimation based on element residual methods. Numer. Math. 65 (1993) 23–50.
[7] M. Ainsworth and J.T. Oden, A posteriori error estimates for Stokes' and Oseen's equations. SIAM J. Numer. Anal. (to appear).
[8] M. Ainsworth, J.T. Oden and C.Y. Lee, Local a posteriori error estimators for variational inequalities. Numer. Methods PDE 9 (1993) 23–33.
[9] J.P. Aubin and H.G. Burchard, Some aspects of the method of the hypercircle applied to elliptic variational problems, in: B.Hubbard, ed., Numerical Solution of Partial Differential Equations—II SYNSPADE (Academic Press, 1970).
[10] I. Babuska and A.D. Miller, A feedback finite element method with a posteriori error estimation Part 1, Comput. Methods Appl. Mech. Engrg. 61 (1987) 1–40.
[11] I. Babuska and W.C. Rheinboldt, Error estimates for adaptive finite element computations. SIAM J. Numer. Anal. 18 (1978) 736–754.
[12] I. Babuska and W.C. Rheinboldt, A posteriori error analysis of finite element solutions for one dimensional problems. SIAM J. Numer. Anal. 18 (1981) 565–589.
[13] I. Babuska and W.C. Rheinboldt, A posteriori error estimates for the finite element method, Int. J. Numer. Methods Engrg. 12 (1978) 1597–1615.
[14] I. Babuska and W.C. Rheinboldt, Analysis of optimal finite element meshes in R^1, Math. Comput. 33 (1979) 435–463.
[15] I. Babuska, T. Strouboulis, C.S. Upadhyay and S.K. Gangaraj, A model study of the quality of a posteriori estimators for linear elliptic problems error estimation in the interior of patchwise uniform grids of triangles, Comput. Methods Appl. Mech. Engrg. 114 (1994) 307–378.
[16] I. Babuska, T. Strouboulis, C.S. Upadhyay, S.K. Gangaraj and K. Copps, Validation of a posteriori error estimators by a numerical approach, Int. J. Numer. Methods Engrg. 37 (1994) 1073–1123.
[17] I. Babuska and M. Suri, The optimal convergence rate of the p version of the finite element method, SIAM J. Numer. Anal. 24 (1987) 750–776.
[18] I. Babuska, B.A. Szabo and I.N. Katz, The p version of the finite element method, SIAM J. Numer. Anal. 18 (1981) 512–545.
[19] I. Babuska and D. Yu, Asymptotically exact a posteriori error estimator for biquadratic elements, Technical Note Inst. Phys. Sci. and Tech BN-1050, University of Maryland, 1986.
[20] I. Babuska, O.C. Zienkiewicz, J. Gago and E.A. Oliveira, Accuracy Estimates and Adaptive Refinements in Finite Element Computations (Wiley, 1986).
[21] R. Bank and B.D. Welfert, A posteriori error estimates for the Stokes problem, SIAM J. Numer. Anal. 28 (1991) 591–623.
[22] R.E. Bank, Analysis of a local a posteriori error estimate for elliptic equations, in: I. Babuska et al., ed., Accuracy Estimates and Adaptive Refinements in Finite Element Computations (Wiley, New York, 1986) 119–128.
[23] R.E. Bank and A. Weiser, Some a posteriori error estimators for elliptic partial differential equations, Math. Comput. 44 (1985) 283–301.
[24] J. Baranger and H. ElAmri, Estimateurs a posteriori d'erreur pour le calcul adaptatif d'ecoulements quasi-Newtoniens, RAIRO Anal. Numér. 25 (1991) 31–48.
[25] P.G. Ciarlet, The Finite Element Method for Elliptic Problems (North-Holland, 1978).
[26] P. Clément, Approximation by finite element functions using local regularization, RAIRO Anal. Numér. 2 (1975) 77–84.
[27] B. Fraeijs de Veubeke, Displacement and equilibrium models in the finite element method, in: Zienkiewicz and Holister, eds., Stress Analysis (Wiley, London, 1965).
[28] L. Demkowicz, Ph. Devloo and J.T. Oden, On an h-type mesh refinement strategy based on minimization of interpolation errors, Comput. Methods Appl. Mech. Engrg. 53 (1985) 67–89.
[29] L. Demkowicz, J.T. Oden and T. Strouboulis, Adaptive finite elements for flow problems with moving boundaries. Part 1: Variational principles and a posteriori error estimates, Comput. Methods Appl. Mech. Engrg. 46 (1984) 217–251.
[30] L. Demkowicz, J.T. Oden and T. Strouboulis, An adaptive p-version finite element method for transient flow problems with moving boundaries, in: R.H. Gallagher et al., ed., Finite Elements in Fluids VI (John Wiley, 1985) 291–305.
[31] R. Duran and R. Rodriguez, On the asymptotic exactness of Bank–Weiser's estimator, Numer. Math. 62 (1992) 297–304.
[32] K. Eriksson and C. Johnson, Error-estimates and automatic time step control for nonlinear parabolic problems. Part 1, SIAM J. Numer. Anal. 24(1) (1987) 12–23.
[33] K. Eriksson and C. Johnson, Adaptive finite-element methods for parabolic problems. Part 1. A linear-model problem, SIAM J. Numer. Anal. 28(1) (1991) 43–77.
[34] K. Eriksson and C. Johnson, Adaptive finite-element methods for parabolic problems. Part 2. Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$, SIAM J. Numer. Anal. 32(3)(1995).
[35] R.E. Ewing, A posteriori error estimation, Comput. Methods Appl. Mech. Engrg. 82(1–3) (1990) 59–72.
[36] V. Girault and P.A. Raviart, Finite Element Methods for Navier Stokes Equations, Springer Series in Computational Mathematics, Vol. 5 (Springer-Verlag, 1986).
[37] I. Hlavacek, J. Haslinger, J. Necas and J. Lovisek, Solution of Variational Inequalities in Mechanics, Applied Mathematical Sciences, Vol. 66 (Springer-Verlag, 1980).
[38] C. Johnson and P. Hansbo, Adaptive finite element methods in computational mechanics, Comput. Methods Appl. Mech. Engrg. 101(1–3) (1992) 143–181.

[39] D.W. Kelly, The self-equilibration of residuals and complementary a posteriori error estimates in the finite element method, Int. J. Numer. Methods Engrg. 20 (1984) 1491–1506.
[40] D.W. Kelly, J.R. Gago, O.C. Zienkiewicz and I. Babuska, A posteriori error analysis and adaptive processes in the finite element method. Part I—Error analysis, Int. J. Numer. Methods Engrg. 19 (1983) 1593–1619.
[41] D. Kinderlehrer and G. Stampacchia, An Introduction to Variational Inequalities and Their Applications (Academic Press, 1980).
[42] M. Krizek and P. Neitaanmaki, On superconvergence techniques, Acta Applic. Math. 9 (1987) 175–198.
[43] P. Ladeveze and D. Leguillon, Error estimate procedure in the finite element method and applications, SIAM J. Numer. Anal. 20 (1983) 485–509.
[44] P. Lesaint and M. Zlamal, Superconvergence of the gradient of finite element solutions, RAIRO Anal. Numér. 13 (1979) 139–166.
[45] A. K. Noor and I. Babuska, Quality assessment and control of finite element solutions, Finite Elem. Des. 3 (1987) 1–26.
[46] J.T. Oden and L. Demkowicz, Advances in adaptive improvements: A survey of adaptive finite element methods in computational mechanics, in: Accuracy Estimates and Adaptive Refinements in Finite Element Computations (ASME, 1987) 1–43.
[47] J.T. Oden, L. Demkowicz, W. Rachowicz and T.A. Westermann, Toward a universal h-p adaptive finite element strategy. Part 2 A posteriori error estimation, Comput. Methods Appl. Mech. Engrg. 77 (1989) 113–180.
[48] J.T. Oden, L. Demkowicz, T. Strouboulis and Ph. Devloo, Adaptive methods for problems in solid and fluid mechanics, in: I. Babuska et al., ed., Accuracy Estimates and Adaptive Refinements in Finite Element Computations (Wiley, New York, 1986) 249–280.
[49] J.T. Oden, W. Wu and M. Ainsworth, A posteriori error estimators for the Navier–Stokes problem, Comput. Methods Appl. Mech. Engrg. 111 (1994) 185–202.
[50] J. Peraire, M. Vahdati, K. Morgan and O.C. Zienkiewicz, Adaptive remeshing for compressible flow computations, J. Comput. Phys. 72 (1987) 449–466.
[51] P.A. Raviart and J.M. Thomas, Primal hybrid finite element methods for 2nd order elliptic equations, Math. Comput. 31 (1977) 391–413.
[52] L.R. Scott and S. Zhang, Finite element interpolation of non-smooth functions satisfying boundary conditions, Math. Comput. 54 (1992) 483–493.
[53] E.M. Stein, Singular Integrals and Differentiability Properties of Functions (Princeton University Press, 1970).
[54] B. A. Szabo, Estimation and control of error based on p convergence, in: I. Babuska et al., ed., Accuracy Estimates and Adaptive Refinements in Finite Element Computations (Wiley, New York, 1986) 61–70.
[55] B.A. Szabo, Mesh design for the p version of the finite element, Comput. Methods Appl. Mech. Engrg. 55 (1986) 181–197.
[56] R. Verfurth, A posteriori error estimators for the Stokes equations, Numer. Math. 55 (1989) 309–325.
[57] R. Verfürth, A posteriori error estimation and adaptive mesh refinement techniques, J. Comput. Appl. Math. 50 (1994) 67–83.
[58] W. Wu, h-p Adaptive methods for incompressible viscous flow problems, Ph.D. Thesis, University of Texas, 1993.
[59] O.C. Zienkiewicz, J.P. Gago and D.W. Kelly, The hierarchical concept in the finite element method, Comput. Struct. 16 (1983) 53–65.
[60] O.C. Zienkiewicz and J.Z. Zhu, A simple error estimator and adaptive procedure for practical engineering analysis, Int. J. Numer. Methods Engrg. 24 (1987) 337–357.
[61] O.C. Zienkiewicz and J.Z. Zhu, The superconvergent patch recovery and a posteriori error estimates. Part 1: The recovery technique, Int. J. Numer. Methods Engrg. 33 (1992) 1331–1364.
[62] O.C. Zienkiewicz and J.Z. Zhu, The superconvergent patch recovery and a posteriori error estimates. Part 2: Error estimates and adaptivity, Int. J. Numer. Methods Engrg. 33 (1992) 1365–1382.
[63] M. Zlamal, Some superconvergence results in the finite element method, in: A. Dold and B. Eckmann, eds., Mathematical Aspects of Finite Element Methods, Number 606 in Springer Lecture Notes in Mathematics, 1975.
[64] M. Zlamal, Superconvergence and reduced integration in the finite element method, Math. Comput. 32 (1978) 663–685.