THE UNIVERSITY OF CHICAGO

AN INTERIOR PENALTY FINITE ELEMENT METHOD WITH DISCONTINUOUS ELEMENTS

A DISSERTATION SUBMITTED TO
THE FACULTY OF THE DIVISION OF THE PHYSICAL SCIENCES
IN CANDIDACY FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

DEPARTMENT OF MATHEMATICS

BY

DOUGLAS N. ARNOLD

CHICAGO, ILLINOIS

JUNE, 1979

ACKNOWLEDGEMENTS

I wish primarily to thank Professor Jim Douglas, Jr. who gave generously of his time and performed the many duties of advisor and teacher with exceptional skill. I am also grateful to numerous professors and colleagues who read the manuscript and offered valuable suggestions.

TABLE OF CONTENTS

Chapter

ABSTRACT

A nonconforming finite element procedure for the solution of second order, nonlinear parabolic boundary value problems is formulated and analyzed. The finite element space consists of discontinuous piecewise polynomial functions over quite general meshes, with inter-element continuity being enforced approximately by means of penalties. Optimal order error estimates in energy and L^2 norms are stated in terms of locally expressed quantities. They are proved first for a model problem and then in general.

CHAPTER I

INTRODUCTION

In this paper we define a semidiscrete finite element procedure for the numerical solution of a second order parabolic initial-boundary value problem. The method is nonconforming in that the piecewise polynomial trial functions are, in general, discontinuous and therefore do not lie in the Sobolev space H^1 associated with variational form of the differential problem. Approximate continuity is imposed by including in the form which defines the method penalty terms which are weighted L^2 inner products of the jumps in the function values across element edges. In the case of Dirichlet boundary conditions, the penalty terms on the boundary of the domain penalize the deviation of the approximate solution from the specified value of the true solution, exactly as in a well-known method of Nitsche [10].

The primary motivation for the interior penalty method is the enhanced flexibility afforded by discontinuous elements. This allows meshes which are more general in their construction and degree of nonuniformity than is permitted by more conventional finite element methods. Moreover, the local nature of the trial space and the capability to regulate the degree of smoothness of the approximate solution by local variation of the penalty weighting function should enable closer approximation of solutions which vary in character from one part of the domain to another and should allow the incorporation of partial knowledge of the solution into the scheme. An important particular class of difficult equations is that of parabolic equations with dominant transport terms for which the solution varies rapidly on a small moving part of the domain.

The inclusion of penalty terms in the variational form defining a finite element method is not new. The method of Nitsche referred to above and the penalty method of Babuška [2] both employ this technique in order to impose essential boundary conditions weakly. Zienkiewicz [14] discussed the use of penalties in the formulation of nonconforming methods for fourth order problems for which the trial functions, though continuous, are not contained in H^2. Babuška and Zlámal [3] have presented a scheme which does just that, using interior penalties analogous to the boundary penalties of Babuška's method to solve the biharmonic equation. More recently Douglas and Dupont [8] have analyzed a method analogous to ours which uses interior penalties to enforce behavior between C^0 and C^1 on conforming elements for linear elliptic and parabolic problems. Numerical experiments with that method have clearly demonstrated the value of penalties for solving certain problems which have proved intractable to more conventional methods (see e.g, [7]). Closest to the present method are an interior penalty method which Wheeler [13] has presented and analyzed for second order linear elliptic equations, and a similar procedure due to Baker [4] for the biharmonic equation.

Since we wish to allow meshes which are relatively course in some parts of the domain and fine in others, we do not assume quasi-uniformity, and we have stated the error estimates in a manner which relates the size of each finite element to the smoothness of the solution on that element. Thus, if h_T is the diameter of the element T, we have bounded the discretization error by quantities of the form

$$\left(\sum_T h_T^{2j} \|w\|_{H^k(T)}^2\right)^{1/2}$$

rather than the more usual

$$(\max_T h_T)^j \|w\|_{H^k(\cup\{T\})} .$$

(In fact we shall even allow j and k to depend on T.) This provides motivation and some justification for schemes incorporating adaptive mesh refinement, in which a new mesh is selected from time to time using partial knowledge of the solution to equalize $h_T^j \|w\|_{H^k(T)}$. We also feel that finite element methods based on discontinuous elements will prove more amenable to such adaptive schemes than do conforming methods.

The problem considered for most of the paper is the initial-boundary value problem

(1.1a) $w_t(x,t) - \nabla\cdot[a(x,t,w(x,t))\nabla w(x,t) + b(x,t,w(x,t))]$

$= f(x,t,w(x,t)), (x,t) \in \partial\Omega\times I,$

(1.1b) $w(x,t) = g(x,t),\ (x,t) \in \partial\Omega\times I,$

(1.1c) $w(x,0) = w_0(x),\ x \in \Omega.$

Here Ω is a bounded domain in the plane with Lipschitz boundary, $I = [0,t^*] \subset \mathbb{R}$, $a \in C_b^2(\bar\Omega\times I\times \mathbb{R})$, $b \in C_b^1(\bar\Omega\times I\times \mathbb{R})\times C_b^1(\bar\Omega\times I\times \mathbb{R})$, $f \in C_b^1(\bar\Omega\times I\times \mathbb{R})$. ($C_b^n$ is the space of functions with continuous, bounded partial derivatives of order up to n.) It is assumed that $\underline{a} \le a(x,t,\rho) \le \bar{a}$ where $\underline{a}$ and $\bar{a}$ are positive constants. Also, we assume that w and w_t are in $C(I;C^1(\bar\Omega))$ and that $w \in L^\infty(I;H^2(\Omega))$. (For the definition of this latter space see subsection 2.1.)

In the next section we collect essential notations and set out the general framework in which the investigation will preceed. Before considering the general case, however, we present in section 3 the method and the energy estimates for a model problem which exhibits the essential features of the analysis, unencumbered by technical detail. The formulation and analysis in the general case are given in the following four sections, with the stronger results which can be obtained in less generality being presented in remarks. In section 9 several extensions and generalizations are considered, including in subsection 9.3 a method designed to facilitate mesh adaptation. A final section collects various observations concerning the penalty function.

CHAPTER II

PRELIMINARIES

2.1. Function spaces.

We shall use the usual L^2-based Sobolev spaces $H^k(S)$ with norm $\|\cdot\|_{k,S}$ and seminorm $|\cdot|_{k,S}$ and the L^∞-based Sobolev spaces with norm $\|\cdot\|_{W^k_\infty(S)}$. $H^1_0(S)$ denotes the subspace of $H^1(S)$ consisting of functions which vanish on ∂S.

If $S \subseteq \mathbb{R}^2$, $(\cdot,\cdot)_S$ [respectively $\langle\cdot\ \cdot\rangle_S$] will denote the inner product in $L^2(S)$ where S is measured by the Lebesgue [respectively, the one dimensional Hausdorff] measure.

By default, $(\cdot,\cdot)=(\cdot,\cdot)_\Omega$, $\langle\cdot,\cdot\rangle=\langle\cdot,\cdot\rangle_{\partial\Omega}$, $H^k=H^k(\Omega)$, and $\|\cdot\| = \|\cdot\|_{0,\Omega} = \|\cdot\|_{L^2(\Omega)}$.

If K is an interval, X is one of the function spaces introduced above, and φ is a function on $\Omega\times K$ then $\|\varphi\|_{L^p(K;X)}$ denotes the norm in $L^p(K)$ of the function $t\to\|\varphi(\cdot,t)\|_X$. $L^p(X)$ is short for $L^p(I,X)$.

2.2 The Mesh.

We wish to consider both triangular and rectangular finite elements and to allow curved boundary elements. Hence we shall state our definitions with sufficient generality to cover all cases.

By a <u>mesh</u> on Ω we mean a finite set $\mathcal{T}$ of closed subsets of $\bar{\Omega}$ such that:

(i) Each element of $\mathcal{T}$ is the closure of its nonempty interior.

(ii) $\cup\mathcal{T} = \bar{\Omega}$.

(iii) Distinct elements of $\mathcal{T}$ have disjoint interiors.

For $T \in \mathcal{T}$ let $h_T = \text{diam}(T)$, the diameter of T. The mesh $\mathcal{T}$ is said to satisfy a <u>shape constraint</u> with shape constant K if for each

$T \in \mathcal{T}$ there exists a homeomorphism Ψ of T onto a closed disc such that K is a Lipschitz constant for both Ψ and Ψ^{-1}. Note that this notion is independent of the concept of quasi-uniformity. ($\mathcal{T}$ is said to be <u>quasi-uniform</u> with quasi-uniformity constant C if $h_{T_1}/h_{T_2} \leq C$ for $T_1, T_2 \in \mathcal{T}$).

<u>Remark 2.1</u>. A family of meshes satisfying a shape constraint with common constant K is a regular family [5,pg. 124], that is each mesh element T contains a disc of diameter $\rho \cdot h_T$ where $\rho > 0$ depends only on K (in fact one can take $\rho = 1/K^2$). (This condition assures nondegeneracy of the elements.) If we restrict our attention to meshes of convex sets, then the opposite implication holds. For given a compact convex set T containing a disc of diameter $\rho \cdot \mathrm{diam}(T)$, there is a bi-Lipschitz homeomorphism of T onto a disc for which the Lipschitz constants in both directions can be bounded in terms of ρ alone. See [6,pg 55] for the simple construction. The shape constraint is employed to insure in addition to regularity that element boundaries are sufficiently well behaved. The maps Ψ of the definition will be used only analytically and will not enter into the formulation of the procedure. In particular, we are <u>not</u> using the disc as a reference element.

<u>Remark 2.2</u>. If the elements of $\mathcal{T}$ are triangles, then $\mathcal{T}$ satisfies a shape constraint with K depending only on the minimum angle of the elements

of T. Again, if T consists of rectangles then there is a shape constant for T depending only on the maximum ratio between the side lengths of a rectangle in T.

A triangular mesh is a mesh each element of which is the intersection of a triangle with $\bar{\Omega}$. A rectangular mesh is one whose elements are the intersections with $\bar{\Omega}$ of rectangles having sides parallel to the coordinate axes. All meshes appearing below are implicitly assumed to be either triangular or rectangular.

Let T be a mesh and define

$$E_0 = \{T_1 \cap T_2 \mid T_1, T_2 \in T \text{ are distinct, } T_1 \cap T_2 \text{ contains at least two points}\},$$

$$E_\partial = \{T \cap \partial\Omega \mid T \in T,\ T \cap \partial\Omega \text{ contains at least two points}\},$$

$$E = E_0 \cup E_\partial,$$

$$E_T = \{e \in E \mid e \subset T\}, \qquad T \in T.$$

Let ℓ_e be the length of e for $e \in E$.

Next we introduce a property relating adjacent mesh elements. In finite element theory one often works only with edge-to-edge meshes; that is, meshes for which distinct intersecting elements meet in either a common vertex or a common edge. One of the advantages of the present method is that the much weaker condition of gradedness suffices.

Note that, if T is an edge-to-edge mesh of either triangles or rectangles, then E is simply the set of all edges of elements of T. From Remark 2.1 it then follows that

$$K^2 \ell_e \geq h_T, \qquad T \in T,\ e \in E_T.$$

For general (not necessarily edge-to-edge) meshes, the elements of E_T are segments of the edges of T, and will be called the edge-segments of T. T is said to be graded with grade constant K' if

$$K'\ell_e \geq h_T, \qquad T \in T,\ e \in E_T.$$

While it is not true that a triangle need have only three edge-segments, nor a rectangle four, it is easy to see that the cardinality of E_T is at most 3 K' [respectively 4 K'] for all triangles [respectively rectangles] T $\in$ T.

For e $\in E_0$, we select one of the two unit normals to e and denote it n_e or simply n. If e $\in E_\partial$ (in which case n_e may be non-constant on e), we choose n_e to point exterior to Ω.

Since our finite element space will consist of discontinuous elements, it will not lie in $H^1(\Omega)$ but rather in the piecewise Sobolev space defined by

$$H^\ell(T) = \{\varphi \in L^2(\Omega) \,\big|\, \varphi|_T \in H^\ell(T) \text{ for all } T \in T\}.$$

Differential operators will be understood to act on such spaces piecewise and not in the sense of distributions. Thus, for example, if $\varphi \in H^1(T)$, we view $\nabla\varphi$ as a function in $L^2(\Omega)\times L^2(\Omega)$.

For $\varphi \in H^1(T)$, we define the jump and average of φ -- denoted $[\varphi]$ and $\{\varphi\}$, respectively-- as functions on $\cup E$ as follows. For each e $\in E$, $[\varphi]$, $\{\varphi\} \in L^2(e)$. If e $\in E_0$, then e = $T_1 \cap T_2$ with n_e exterior to T_1 for some pair of elements (T_1,T_2). Set

$$[\varphi] = (\varphi|_{T_1})|_e - (\varphi|_{T_2})|_e$$

$$\{\varphi\} = [(\varphi|_{T_1})|_e + (\varphi|_{T_2})|_e]/2.$$

If e $\in E_\partial$, then $[\varphi] = \{\varphi\} = \varphi|_e$. The restrictions to e are taken in the sense of traces.

In this notation we can state the basic integration by parts formula

$$(2.1)\qquad (\nabla\cdot\varphi,\Psi) = -(\varphi,\nabla\Psi) + \langle\varphi\cdot n,\Psi\rangle + \sum_{e\in E_0} (\langle\{\varphi\}\cdot n,[\Psi]\rangle_e + \langle[\varphi]\cdot n,\{\varphi\}\rangle_e),$$

valid for $\varphi \in H^1(T)\times H^1(T)$ and $\Psi \in H^1(T)$.

We shall also require an inequality of the Poincare-Friedrichs type valid for $\varphi \in H^1(T)$. This result in turn requires a geometric lemma, the proof of which appears in the appendix.

LEMMA 2.3. *Let T be a mesh on a polygonal domain with T consisting entirely of triangles [respectively rectangles]. Then there exists a constant C depending only on the domain and the shape and grade constants for T such that for every line L in $\mathbb{R}^2$ [respectively every line L parallel to the axis],*

$$\sum\{\ell_e \mid e \in E, e\cap L \neq \emptyset\} \leq C.$$

THEOREM 2.4. *Let T be a triangular or rectangular mesh on Ω. Then there exists a constant C depending only on Ω and the shape and grade constants for T such that*

$$(2.2)\qquad \|\varphi\|^2 \leq C\ (\|\nabla\varphi\|^2 + \sum_{e\in E} \ell_e^{-1}\|[\varphi]\|_{0,e}^2) \text{ for } \varphi \in H^1(T).$$

Proof. The proof will be given only in the triangular case, the rectangular case being simpler. We begin by reducing to the situation where Ω is polygonal and all the elements of T are triangles.

Note that, if $T_1 \in \mathcal{T}$ and $e_1 \in \mathcal{E}_\partial \cap \mathcal{E}_T$, then

$$\|\varphi\|^2_{0,T_1} \le C(\|\nabla\varphi\|^2_{0,T_1} + \ell^{-1}_{e_1}\|\varphi\|^2_{0,e_1});$$

in fact, the right hand side even bounds $\ell_e^{-2}\|\varphi\|^2_{0,T_1}$. Moreover, if T_2 is a triangle all of whose edge-segments lie in $\mathcal{E}_0$ and which shares an edge-segment e_2 with T_1 (see Figure 1), then

$$\ell^{-1}_{e_2} \|\varphi|_{T_2}\|^2_{0,e_2} \le 2(\ell^{-1}_{e_2} \|\varphi|_{T_1}\|^2_{0,e_2} + \ell^{-1}_{e_2} \|[\varphi]\|^2_{0,e_2})$$

$$\le C(\|\nabla\varphi\|^2_{0,T_1} + \ell^{-1}_{e_1} \|\varphi\|^2_{0,e_1} + \ell^{-1}_{e_2} \|[\varphi]\|^2_{0,e_2}).$$

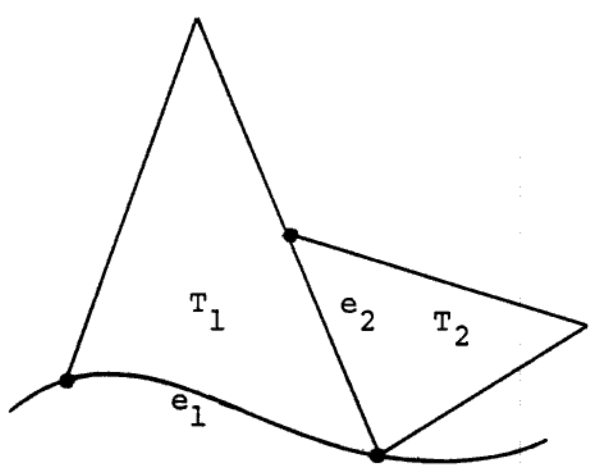

Figure 1. Triangles meeting $\partial\Omega$.

These claims are easily proved on an isoceles right triangle and preserved under a Lipschitz homeomorphism. (Each of our elements is by assumption Lipschitz equivalent to a disc, and hence to an isoceles right triangle.) Consequently, if we prove the assertion for the restriction of the triangulation to the domain obtained by removing from Ω all elements with boundary edges, then the resulting right hand side will be dominated by the right hand side of (2.2). Hence, changing notation, we assume that T consists of triangles.

Let $S = I_1 \times I_2$ be a square of side length 2 diam(Ω) containing $\bar{\Omega}$ in its interior. Extend φ to S by zero. For any $y \in I_2$ such that $I_1 \times \{y\}$ contains no element $e \in E$, define $J_e\varphi(y)$ as follows:

if $I_1 \times \{y\} \cap e = \emptyset$, then $J_e\varphi(y) = 0$;

if $I_1 \times \{y\} \cap e = \{x,y\}$, then $J_e\varphi(y) = \varphi(x+,y) - \varphi(x-,y)$.

Let π_e be the projection of e onto the y-axis and p_e the length of that projection. By the lemma

$$\sum\{p_e | e \in E,\ I_1 \times \{y\} \cap e \neq \emptyset\} \leq M,$$

where M depends only on Ω and the shape and grade constants. Hence, squaring the inequality

$$|\varphi(x,y)| \leq \int_{I_1} |\varphi_x(t,y)|dt + \sum_{e \in E} |J_e\varphi(y)|$$

which holds for almost all $(x,y) \in S$, we see that

$$|\varphi(x,y)|^2 \leq 2[\int_{I_1} |\varphi_x(t,y)|dt]^2 + 2[\sum_{e \in E} |J_e\varphi(y)|]^2$$

$$\leq 4 \text{ diam}(\Omega) \int_{I_1} |\varphi_x(t,y)|^2 dt + 2M \sum_{e \in E} p_e^{-1} |J_e\varphi(y)|^2.$$

We have used the Cauchy-Schwartz inequality in the form

$$\sum \lambda_i^2 \le (\sum \alpha_i)\ (\sum \alpha_i^{-1} \lambda_i^2), \quad \alpha_i, \lambda_i \ge 0.$$

Now

$$\int_{I_2} |J_e \varphi(y)|^2 dy = \int_{\pi_e} |J_e \varphi(y)|^2 dy = \frac{p_e}{\ell_e} \|[\varphi]\|_{0,e}^2 .$$

Therefore integration of the last inequality over $(x,y) \in S$ yields

$$\|\varphi\|^2 \le 8[\mathrm{diam}(\Omega)]^2 \|\varphi_x\|^2 + 4\ M\ \mathrm{diam}(\Omega) \sum_{e \in E} \ell_e^{-1} \|[\varphi]\|_{0,e}^2 .$$

For the remainder of the paper we concentrate -- for simplicity of notation -- on one fixed triangular or rectangular mesh $\mathcal{T}$ on Ω with shape constant K and grade constant K'. Let $h = \max h_T$. We shall prove estimates below in which certain errors are bounded by expressions involving constant multiples of powers of h_T and h. Of course, what we have in mind is a family of meshes for which the corresponding values of h approach zero. Thus, for our results to be meaningful it is essential that the constants appearing depend on $\mathcal{T}$ only through K and K'. In short we shall prove asymptotic estimates valid for any uniformly shape constrained family of uniformly graded meshes.

2.3. The Finite Element Space.

Let r be a fixed positive integer. For each $T \in \mathcal{T}$, let $\mathcal{Q}_r(T)$ be the set of functions which are restrictions to T of polynomials of separate degree at most r in each variable, and let $\mathcal{P}_r(T)$ be the subset consisting of restrictions of polynomials of total degree at most r. Set $\mathcal{M}_r(T) = \mathcal{P}_r(T)$ if $\mathcal{T}$ is triangular and $\mathcal{M}_r(T) = \mathcal{Q}_r(T)$ if T is rectangular.

Now T contains a disc of radius h_T/K^2 (see remark 2.1), from which it follows that there exists a constant C depending only on r and K such that

$$\|\varphi\|_{1,T} \le Ch_T^{-1}\|\varphi\|_{0,T},$$

$$\|\varphi\|_{0,\partial T}^2 \le Ch_T^{-1}\|\varphi\|_{0,T}^2,$$

$$\left\|\frac{\partial\varphi}{\partial n}\right\|_{0,\partial T}^2 \le Ch_T^{-1}\|\varphi\|_{1,T}^2,$$

$$\|\varphi\|_{W_\infty^1(T)} \le Ch_T^{-1}\|\varphi\|_{1,T},$$

for all $\varphi \in M_r(T)$. These relations are referred to as local inverse inequalities.

If φ is continuous on the triangle T, define $I_T\varphi$ to be the unique function in $M_r(T)$ interpolating φ at the $(r+1)(r+2)/2$ points of T with barycentric coordinates in $\{0,1/r,2/r,\ldots,1\}$ (see Figure 2). Similarly, for rectangular T define $I_T\varphi$ to be the unique element of $M_r(T)$ interpolating φ at $(r+1)^2$ evenly spaced grid points on T.

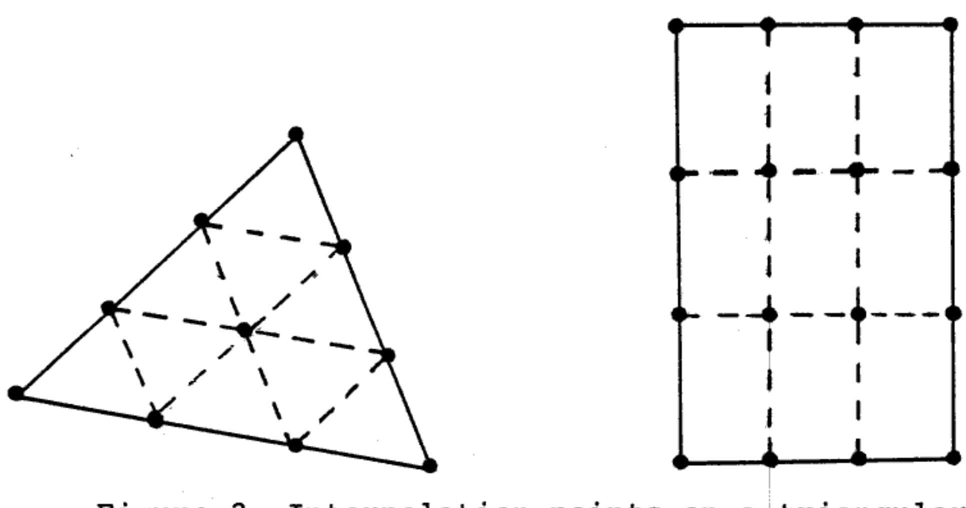

Figure 2. Interpolation points on a triangular and a rectangular finite element. r = 3.

In both cases, there exists a constant C = C(r,K) such that

(2.3) $\|\varphi - I_T\varphi\|_{i,T} \leq Ch_T^{j-1}\ \|\varphi\|_{j,T},\ 0 \leq i \leq j \leq r+1, \quad j \geq 2.$

If $T \in \mathcal{T}$ has a curvilinear edge, we may choose the interpolation points off the interior edges in any fashion which preserves (2.3).

The finite element space we shall employ is

$$M = \prod_{T\in\mathcal{T}} M_r(T).$$

If $\chi \in M$ and $T \in \mathcal{T}$, we denote the T component of χ by $\chi|_T$, i.e., $\chi = (\chi|_T)_{T\in\mathcal{T}}$. χ will be viewed as a function in $L^2(\Omega)$, its value being defined unambiguously on $\Omega\setminus\cup E_0$.

Define I: $H^2(\mathcal{T}) \to M$ by

$$(I\varphi)|_T = I_T(\varphi|_T),\ T \in \mathcal{T}.$$

Then, if we temporarily set $\Pi = I$,

(A1) Π: $H^2(\mathcal{T}) \to M$ is linear.

(A2) $\|\varphi - \Pi\varphi\|_{i,T} \leq C\ (r,K)\ h_T^{j-i}\ \|\varphi\|_{j,T},\ 0 \leq i \leq j \leq r+1, \quad j \geq 2.$

If $\mathcal{T}$ is an edge-to-edge mesh, then in addition

(A3) $\Pi\varphi$ is continuous if φ is continuous.

(A3) does not hold for general triangular or rectangular meshes. This leads to a further definition. We have seen that in all cases there is an approximation operator $\Pi = I$. If there exists Π satisfying also (A3), we shall then say that M approximates smoothly. (More precisely, this could be viewed as a property of a family of triples $(\mathcal{T},M,\Pi)$ but we shall avoid this formality.)

Let Π satisfying (A1) and (A2)(and (A3) if M approximates smoothly) be chosen and fixed. Now for $e \in E_T$ and $\varphi \in H^2(T)$ the following trace inequalities hold ([1, Theorem 3.10]).

$$(2.4)\qquad \|\varphi\|_{0,e}^2 \leq C\,(\ell_e^{-1}\|\varphi\|_{0,T}^2 + \ell_e|\varphi|_{1,T}^2),\quad \varphi \in H^1(T),$$

$$(2.5)\qquad \|\tfrac{\partial\varphi}{\partial n}\|_{0,e}^2 \leq C\,(\ell_e^{-1}\,|\varphi|_{1,T}^2 + \ell_e\,|\varphi|_{2,T}^2),\quad \varphi \in H^2(T).$$

Hence, for all $\varphi \in H^j(T)$,

$$(2.6)\qquad \|\varphi - \Pi\varphi\|_{1,T}^2 + \sum_{e\in E_T} [\ell_e^{-1}\|\varphi - \Pi\varphi\|_{0,e}^2 + \ell_e\|\tfrac{\partial}{\partial n}(\varphi-\Pi\varphi)\|_{0,e}^2]$$

$$\leq Ch_T^{2(j-1)}\,\|\varphi\|_{j,T}^2,\qquad 2 \leq j \leq r+1$$

where C depends only on r, K and K'.

2.4. Forms, Norms and Penalties.

In order to derive a weak formulation of (1.1) we note that (2.1) implies

$$-(\nabla\cdot[a(w)\nabla\varphi],\Psi) = (a(w)\nabla\varphi,\nabla\Psi) - \sum_{e\in E_0}\langle a(w)\tfrac{\partial\varphi}{\partial n},[\Psi]\rangle - \langle a(g)\tfrac{\partial\varphi}{\partial n},\Psi\rangle$$

for $\varphi \in H^2(\Omega)$ and $\Psi \in H^1(T)$. If we symmetrize the form appearing on the right hand side and replace the unknown solution w by a function ρ, we arrive at the form

$$A(\rho;\varphi,\Psi) = (a(\rho)\nabla\varphi,\nabla\Psi) - \sum_{e\in E_0}[\langle a(\{\rho\})\{\tfrac{\partial\varphi}{\partial n}\},[\Psi]\rangle_e$$

$$+\langle a(\{\rho\})[\varphi],\{\tfrac{\partial\Psi}{\partial n}\}\rangle_e] - [\langle a(g)\tfrac{\partial\varphi}{\partial n},\Psi\rangle + \langle a(g)\varphi,\tfrac{\partial\Psi}{\partial n}\rangle]$$

defined for $\rho,\varphi,\Psi \in H^2(T)$.

Penalties will be introduced via the form

$$J(\varphi,\Psi)=\sum_{e\in E} \ell_e^{-1}\langle \sigma[\varphi],[\Psi]\rangle_e, \qquad \varphi,\Psi \in H^1(T),$$

where $\sigma: \bigcup E\times J \to [\gamma_0, \infty) \subset (0,\infty)$ is a measurable function, differentiable in t when viewed as a function into $L^\infty(\bigcup E)$. J depends on t through σ. Note that the definitions of A and J are independent of the choice of the interior normals n_e.

We also set

$$B(\rho;\cdot,\cdot) = A(\rho;\cdot,\cdot) + J(\cdot,\cdot),$$

$$B(\cdot,\cdot) = B(w;\cdot,\cdot).$$

It follows from (2.1) that the solution w satisfies

(2.7) $$(w_t,\chi)+B(w,\chi)+(b(w),\nabla\chi)-\sum_{e\in E_0}\langle b(w)\cdot n,[\chi]\rangle_e-\langle b(g)\cdot n,\chi\rangle$$

$$= (f(w),\chi)-\langle a(g)g,\frac{\partial\chi}{\partial n}\rangle+\sum_{e\in E_\partial}\ell_e^{-1}\langle\sigma g,\chi\rangle$$

for all $\chi \in H^2(T)$. Here and elswhere we suppress some of the arguments of the coefficients in the notation.

On the space $H^\ell(T)$ we place the obvious norm:

$$\|\varphi\|_{\ell,T} = (\sum_{T\in T}\|\varphi\|_{\ell,T}^2)^{1/2}.$$

The following norm, which incorporates a measure of discontinuity into the $H^1(T)$ norm is naturally associated with the form A. Define

$$|||\varphi|||^2 = \|\varphi\|_{1,T}^2 + \sum_{e\in E}(\ell_e^{-1}\|[\varphi]\|_{0,e}^2 + \ell_e\|\{\tfrac{\partial\varphi}{\partial n}\}\|_{0,e}^2),$$

for $\varphi\in H^2(T)$. We have immediately the inequality

(2.8) $$|A(\rho;\varphi,\Psi)| \le \bar{a}\ |||\varphi|||\ |||\Psi|||, \qquad \rho,\varphi,\Psi \in H^2(T).$$

The following lemma shows that restricted to M $|||\cdot|||$ is equivalent to a simpler norm.

LEMMA 2.5. There exists a constant $C = C(K)$ such that

$$|||\varphi|||^2 \leq C\,(\|\varphi\|^2_{1,T} + \sum_{e\in E} \ell_e^{-1} \|[\varphi]\|^2_{0,e}), \qquad \varphi \in M.$$

The proof follows directly from the inverse inequality

$$(2.9) \qquad \sum_{e\in E} \ell_e \, \|\{\tfrac{\partial\varphi}{\partial n}\}\|^2_{0,e} \leq C\|\varphi\|^2_{1,T}, \quad \varphi \in M.$$

From (2.6) it follows that for integers $j(T)$,

$$(2.10) \qquad |||\varphi - \Pi\varphi||| \leq C(r,K,K')\,(\sum_{T\in T} h_T^{2[j(T)-1]} \|\varphi\|^2_{j(T),T})^{1/2},$$

$$2 \leq j(T) \leq r+1, \qquad T \in T.$$

In the course of the analysis we shall impose restrictions on the penalty function σ. These restrictions will refer to various quantities which are collected here for reference.

γ_0 = a positive lower bound for σ;

$\gamma_1 = \sup\,\{\sigma(x,t) \mid x \in \cup E,\ t \in I\}$;

$\gamma_2 = \sup\,\{\sigma(x,t) \mid x \in \partial\Omega,\ t \in I\}$;

$\gamma_3 = \sup\,\{|\sigma_t(x,t)| \mid x \in \cup E,\ t \in I\}$;

$\gamma_4 = \sup \{|\sigma^{-1}(x,t)\,\sigma_t(x,t)| \;\big|\; x \in \cup E,\ t \in I\};$

$\gamma_5 = \sup \{\sigma(x,0) \mid x \in \cup E$

$\gamma_6 = \sup \{\sigma(x,0) \mid x \in \partial\Omega\}.$

When the statement of a result refers to some γ_i, it is tacitly assumed that γ_i exists and is finite.

2.5. Constants

As has already been seen, the letter C will denote a generic constant, changing from appearance to appearance. Occasionally its dependence will be presented explicitly. Other times the dependence will be indicated implicitly; e.g. $C(r,K,K')$. However, dependence on r, K, K', $\bar{a}$, $\underline{a}$, Ω, t^*, $\|a\|_{W^2_\infty}$, $\|Lb\|_{W^1_\infty \times W^1_\infty}$, $\|Lf\|_{W^1_\infty}$, $\|w\|_{L^\infty(W^1_\infty)}$ and $\|w_t\|_{L^\infty(W^1_\infty)}$ will not necessarily be noted. Similar remarks apply to ε, which will be used to denote a generic small positive constant.

CHAPTER III

A MODEL PROBLEM

In this section we consider an interior penalty method for the heat equation. The results we obtain here are special cases of sharper ones which will be proven in the following sections, and a number of dispensable assumptions are made in the interest of simplicity.

Let w be a smooth function satisfying the heat equation

$$w_t - \Delta w = 0 \qquad \text{on } \Omega\times I,$$

$$w_t = 0 \qquad \text{on } \partial\Omega\times I,$$

$$w(\cdot,0) = w_0 \qquad \text{on } \Omega.$$

We assume that Ω is polygonal and that the mesh T consists entirely of triangles or rectangles and is edge-to-edge. T need not be quasi-uniform, but the estimates will not be stated in a manner which reflects the advantage of local refinement.

We consider only a constant penalty function

$$\sigma(x,t) = \gamma_0, \qquad x \in \Omega,\ t \in I.$$

Thus,

$$A(\varphi,\Psi) = (\nabla\varphi,\nabla\Psi) - \sum_{e\in E}\left[\left\langle\{\frac{\partial\varphi}{\partial n}\},[\Psi]\right\rangle_e + \left\langle[\varphi],\ \{\frac{\partial\Psi}{\partial n}\}\right\rangle_e\right],$$

and

$$J(\varphi,\Psi) = \gamma_0 \sum_{e\in E} \ell_e^{-1}\langle[\varphi],[\Psi]\rangle_e.$$

Clearly,

$$|A(\varphi,\Psi)| \leq |||\varphi||| \quad |||\Psi|||,$$

$$|J(\varphi,\Psi)| \leq \gamma_0 \quad |||\varphi||| \quad |||\Psi|||,$$

for $\varphi,\Psi \in H^2(T)$.

Let C_1 and C_2 be the constants appearing in (2.9) and (2.2), respectively. then

$$A(\varphi,\varphi) \geq \|\nabla\varphi\|^2 - 2(\sum \ell_e \|\{\tfrac{\partial\varphi}{\partial n}\}\|_{0,e}^2)^{1/2}(\sum \ell_e^{-1}\|[\varphi]\|_{0,e}^2)^{1/2}$$

$$\geq \tfrac{1}{2}\|\nabla\varphi\|^2 - 2C_1\sum \ell_e^{-1}\|[\varphi]\|_{0,e}^2$$

$$\geq \tfrac{1}{4}\|\nabla\varphi\|^2 + \frac{1}{4C_1}\sum \ell_e \|\{\tfrac{\partial}{\partial n}\}\|_{0,e}^2 - 2C_1\sum \ell_e^{-1}\|[\varphi]\|_{0,e}^2$$

$$\geq \tfrac{1}{8}\|\nabla\varphi\|^2 + \frac{1}{8C_2}\|\varphi\|^2 + \frac{1}{4C_1}\sum \ell_e \|\{\tfrac{\partial\varphi}{\partial n}\}\|_{0,e}^2 - (2C_1 + \tfrac{1}{8})\sum \ell_e^{-1}\|[\varphi]\|_{0,e}^2.$$

for all $\varphi \in M$. Now assume that $\gamma_0 \geq 4C_1 + \frac{1}{2}$.

Then

(3.1) $B(\varphi,\varphi) \geq \varepsilon_1 |||\varphi|||^2 + \frac{1}{2}J(\varphi,\varphi)$, $\varphi \in M$,

where $\varepsilon_1 = \min(\frac{1}{8}, \frac{1}{4C_1}, \frac{1}{8C_2}) > 0$.

Now w satisfies

(3.2) $(w_t,\chi) + B(w,\chi) = 0$, $\chi \in H^2(T)$.

The interior penalty finite element approximation to w is defined by analogy as the unique function $W\colon\ I \to M$ such that

(3.3a) $(W_t,\chi) + B(W,\chi) = 0$, $\chi \in M$,

(3.3b) $W(0) = Iw_0$.

Let $\{\chi_i\}_{i=1}^m$ be a basis for M. Define $m \times m$ matrices α and β and an m-vector ω_0 by

$$\alpha_{ij} = (\chi_j,\chi_i),$$

$$\beta_{ij} = B(\chi_j,\chi_i),$$

$$Iw_0 = \sum_{j=1}^{m} \omega_j^0 \chi_j.$$

Then, writing $W(t) = \Sigma\ \omega_j(t)\chi_j$, (3.3) can be regarded as an ordinary differential initial value problem for the unknown vector :

$$\alpha\omega'(t) + \beta\omega(t) = 0, \quad t \in I,$$
$$\omega(0) = \omega_0.$$

Since the matrix is nonsingular (in fact positive definite), a unique solution exists.

Note that, if we choose $\{\chi_i\}$ in the obvious manner, by selecting a canonical basis for each $M_r(T)$ and extending all the resulting function to by zero, then both the matrices α and β are sparse. Moreover, to evaluate a nonzero entry of requires only an integration over a single element and some one-dimensional quadrature over its edges. Note also that the size of these matrices is considerably larger than for conforming finite element methods of the same degree over the mesh.

We now analyze the proposed procedure by the method of energy estimates. Let $\zeta = W-w$. Then from (3.2) and (3.3),

$$(\zeta_t,\chi) + B(\zeta,\chi) = 0, \ \chi \in M. \tag{3.4}$$

Decompose ζ as $\mu-\nu$ where $\mu = Iw-w$, $\nu = Iw-W$. Note that $[\mu] \equiv 0$ on $\cup E\times I$; thus,

$$(\nu_t,\chi) + B(\nu,\chi) = (\mu_t,\chi) + A(\mu,\chi), \quad \chi \in M. \tag{3.5}$$

Since $\nu(t) \in M$ we can set $\chi=\nu(t)$, obtaining

$$\frac{1}{2}\frac{d}{dt}\|\nu\|^2 + B(\nu,\nu) = (\mu_t,\nu) + A(\mu,\nu)$$
$$\le \frac{1}{2}\|\mu_t\|^2 + \frac{1}{2}\|\nu\|^2 + C|||\mu|||^2 + \frac{\varepsilon_1}{2}|||\nu|||^2.$$

Therefore, we can apply the coercivity result (3.1) to get

$$\frac{d}{dt}\|\nu\|^2 + \varepsilon_1|||\nu|||^2 + J(\nu,\nu) \le \|\nu\|^2 + \|\mu_t\|^2 + C|||\mu|||^2.$$

Since $\nu(0) = 0$, Gronwall's lemma implies that

$$(3.6)\quad \|\nu\|^2_{L^\infty(L^2)} + \int_I |||\nu|||^2\, dt + \int_I J(\nu,\nu)\,dt \le C(\|\mu_t\|^2_{L^2(L^2)} + \int_I |||\mu|||^2 dt,$$

Since $\zeta = \mu - \nu$ and $[\mu] = 0$,

$$\|\zeta\|^2_{L^\infty(L^2)} + \int_I |||\zeta|||^2 dt + \int_I J(\zeta,\zeta)\,dt$$

$$\le C(\|\mu\|^2_{L^\infty(L^2)} + \int_I |||\mu|||^2 dt + \|\mu_t\|^2_{L^2(L^2)}).$$

Thus, as is typical in finite element theory, error bounds for the finite element approximation to the true solution reduce to error bounds for a much simpler sort of approximation, in this case the piecewise polynomial interpolant. These latter bounds have already been noted in (2.3) and (2.10) and hence we have obtained the following theorem.

THEOREM 3.1. _The error ζ in the interior penalty finite element method for the heat equation satisfies the inequality_

$$\|\zeta\|_{L^\infty(L^2)} + (\int_I |||\zeta|||^2 dt)^{1/2} + (\int_I J(\zeta,\zeta)\,dt)^{1/2}$$

$$\le Ch^r(\|w\|_{L^\infty(H^r)} + \|w\|_{L^2(H^{r+1})} + \|w_t\|_{L^2(H^r)}).$$

The constant C _depends only on_ r, K, t^* _and_ Ω.

Remark 3.2. This theorem does not supply an optimal order estimate on $\|\zeta\|_{L^\infty(L^2)}$. In section 6 we use the technique of comparing W to an interior penalty elliptic projection to derive an $O(h^{r+1})$ bound on ζ in $L^\infty(L^2)$.

The choice $\chi = \nu_t$ in (3.5) leads to the second energy estimate. Then,

$$\|\nu_t\|^2 + \frac{1}{2}\frac{d}{dt}B(\nu,\nu) \le \frac{1}{2}\|\mu_t\|^2 + \frac{1}{2}\|\nu_t\|^2 + A(\mu,\nu_t),$$

so

$$(3.7)\quad \int_0^t \|\nu_t\|^2 dt + B(\nu(t),\nu(t)) \le \|\mu_t\|^2_{L^2(L^2)} + 2\left|\int_0^t A(\mu,\nu_t)dt\right|.$$

The final term may be integrated by parts in time. Hence,

$$2\left|\int_0^t A(\mu,\nu_t)dt\right| \le 2|A(\mu(t),\nu(t))| + 2\left|\int_0^t A(\mu_t,\nu)dt\right|$$
$$\le \frac{\varepsilon_1}{2}|||\nu(t)|||^2 + \int_I |||\nu|||^2 dt + C(\sup_I |||\mu|||^2 + \int_I |||\mu_t|||^2 dt).$$

Moreover, by (3.6), $\int_I |||\nu|||^2 dt$ can be absorbed into the last term. If (3.1) is applied to (3.7), it follows that

$$\int_0^t \|\nu_t\|^2 dt + |||\nu(t)||| + J(\nu(t),\nu(t))$$
$$\le C(\sup_I |||\mu|||^2 + \int_I |||\mu_t|||^2 dt).$$

Since this result obviously remains true if we replace ν by μ, it holds also for ζ, giving the following theorem.

THEOREM 3.3. <u>There exists a constant</u> C <u>depending only on</u> r, K, t* <u>and</u> Ω such that

$$\|\zeta_t\|_{L^2(L^2)} + \sup_I |||\zeta||| + \sup_I [J(\zeta,\zeta)]^{1/2}$$
$$\le Ch^r(\|w\|_{L^\infty(H^{r+1})} + \|w_t\|_{L^2(H^{r+1})}).$$

CHAPTER IV

DEFINITION OF THE METHOD

In light of (2.7) we define the approximate solution $W: J \to M$ by the equations

$$(4.1) \qquad (W_t,\chi) + B(W;W,\chi) + (b(W),\nabla\chi) - \sum_{e\in E_0} \langle b(\{W\})\cdot n, [\chi] \rangle_e$$

$$= (f(W),\chi) - \langle a(g)g, \frac{\partial\chi}{\partial n}\rangle + \sum_{e\in E_\partial} \ell_e^{-1} \langle \sigma g,\chi\rangle_e + \langle b(g)\cdot n,\chi\rangle,$$

$$\chi \in M, \quad t \in T.$$

Upon choice of a basis for M, (4.1) may be viewed as a system of ordinary differential equations in the unknown coefficients of W. Once an initial condition is imposed, it follows that W is determined uniquely and moreover is computable from f and g. Assume that the initial value $W(0) \in M$ satisfies

$$(4.2) \qquad \| W(0) - w_0\|^2 \le Ch^2 \sum_{T\in T} h_T^{2[j(T)-1]} \|w_0\|^2_{j(T),T},$$

$$2 \le j(T) \le r+1.$$

(That is, (4.2) is supposed to hold for all integer valued functions $T \mapsto j(T) \in [2, r+1]$.) Acceptable choices of $W(0)$ are, for example, the interpolant Iw_0 of w_0, the L^2 projection of w_0 into M, or the elliptic projection of w_0 defined by the linear system

$$B(W(0),\chi) = B(w_0,\chi), \qquad \chi \in M.$$

In the first two cases,

$$\| W(0) - w_0 \|^2 \le C\sum h_T^{2j(T)} \|w_0\|^2_{j(T),T}, \quad 1 \le j(T) \le r+1,$$

while the error estimate (4.2) will be shown for the elliptic projection in Theorem 6.2 below.

Let $\zeta = W-w$. In the following three sections we shall derive estimates on ζ which are strengthened and generalized versions of the estimates stated in Theorem 3.1, Remark 3.2, and Theorem 3.3, respectively.

CHAPTER V

THE FIRST ENERGY ESTIMATE

The object of this section is the demonstration of the following theorem.

THEOREM 5.1. *Assume that* (4.2) *holds for a selection of* $j(T)$, $T \in \mathcal{T}$, *satisfying* $2 \le j(T) \le r+1$. *Then there exists a constant* C *depending on* γ_1 *such that the error satisfies the inequality*

$$(5.1) \qquad \|\zeta\|^2_{L^\infty(L^2)} + \int_I [|||\zeta|||^2 + J(\zeta,\zeta)]dt$$

$$\le C \sum_{T\in\mathcal{T}} h_T^{2[j(T)-1]} \left(\|w_t\|^2_{L^1(H^{j(T)-1}(T))} + h^2\|w_0\|^2_{j(T),T}\right).$$

We begin by proving a coercivity result for the form B.

THEOREM 5.2. *There exists a positive constant* ε *such that if* γ_0 *is sufficiently large, then*

$$B(\rho;\varphi,\varphi) \ge \varepsilon |||\varphi|||^2 + \frac{1}{2} J(\varphi,\varphi)$$

for all $\varphi \in M$ *and* $\rho \in H^2(\mathcal{T})$.

Proof. $A(\rho;\varphi,\varphi) \ge \underline{a}\,\|\nabla\varphi\|^2 - 2\,\bar{a} \sum_{e\in E} \|\{\frac{\partial\varphi}{\partial n}\}\|_{0,e}\,\|[\varphi]\|_{0,e}$

$$\ge \underline{a}\,\|\nabla\varphi\|^2 - \delta \sum_{e\in E} \ell_e \|\{\frac{\partial\varphi}{\partial n}\}\|^2_{0,e} - C\delta^{-1}\sum_{e\in E} \ell_e^{-1}\|[\varphi]\|^2_{0,e}$$

where $\delta>0$ is arbitrary. Using (2.9) we see that by taking δ small enough we get

$$A(\rho;\varphi,\varphi) \geq \frac{1}{2}\,\underline{a}\|\nabla\varphi\|^2 - \delta'\|\varphi\|^2 - C\sum_{e\in E}\ell_e^{-1}\|[\varphi]\|_{0,e}^2,$$

where $\delta' > 0$ can be taken as small as desired (with C correspondingly large). The theorem therefore follows from Theorem 2.4 and Lemma 2.5.

Hereafter, it is assumed that γ_0 is sufficiently large in the sense of Theorem 5.2. Now ,

(5.2) $$(\zeta_t,\chi) + B(W;\zeta,\chi) = -(b(W) - b(w),\nabla\chi) + \sum_{e\in E_0}\langle[b(\{W\}) - b(w)]\cdot n,[\chi]\rangle_e + (f(W) - f(w),\chi) + A(w;w,\chi) - A(W;w,\chi), \qquad \chi\in M,$$

as one verifies by subtracting (2.7) from (4.1). Next set $\mu=\Pi w-w$ and $\nu=\Pi w-W$, and substitute

(5.3) $$\mu - \nu = \zeta$$

into (5.2), to obtain

(5.4) $$(\nu_t,\chi) + B(W;\nu,\chi) = (\mu_t,\chi) + B(W;\mu,\chi) + (b(W) - b(w),\nabla\chi) - \sum_{e\in E_0}\langle[b(\{W\}) - b(w)]\cdot n,\chi\rangle_e - (f(W) - f(w),\chi) + [A(W;w,\chi) - A(w;w,\chi)], \qquad \chi\in M.$$

The following lemma, will prove useful here and in the sequel.

LEMMA 5.3. <u>Let</u> α,β_1,β_2, <u>and</u> γ <u>be real-valued functions on</u> $\bar{\Omega}\times\mathbb{R}$ <u>each of which satisfies a Lipschitz condition with respect to its second argument uniformly over</u> $\bar{\Omega}$, <u>with Lipschitz constant</u> M. <u>Let</u> $\varphi\in C^1(T)$ <u>for each</u> $T\in\mathcal{T}$ <u>and set</u> $\|\varphi\|_{W^1_\infty(\mathcal{T})} = \sup_T \|\varphi\|_{W^1_\infty(T)}$. <u>Then there is a constant</u> $C = C(M,\|\varphi\|_{W^1_\infty(\mathcal{T})})$ <u>such that, for all</u> $\rho_1,\rho_2,\Psi\in H^2(\mathcal{T})$, (5.5) $$|([\alpha(\rho_1) - \alpha(\rho_2)]\nabla\varphi,\nabla\Psi)| + |(\beta(\rho_1) - \beta(\rho_2),\nabla\Psi)|$$

$$+ \quad |(\gamma(\rho_1) - \gamma(\rho_2),\ \Psi)| \le C\|\rho_1 - \rho_2\| \ \ \|\Psi\|_{1,T};$$

(5.6) $$\sum_{e\in E_0} (|<[\alpha(\{\rho_1\}) - \alpha(\{\rho_2\})]\ \{\frac{\partial\varphi}{\partial n}\},[\Psi]>_e|$$

$$+ \ |<[\beta(\{\rho_1\}) - \beta(\{\rho_2\}]\cdot n,[\Psi]>_e|)$$

$$\le C[\|\rho_1 - \rho_2\| + (\sum_{T\in T} h_T^2\|\rho_1-\rho_2\|_{1,T}^2)^{1/2}](\sum_{e\in E_0} \ell_e^{-1}\|[\Psi]\|_{0,e}^2)^{1/2};$$

(5.7) $$\sum_{e\in E_0} |<[\alpha(\{\rho_1\}) - \ (\{\rho_2\})]\ \ \{\frac{\partial\Psi}{\partial n}\},[\varphi]>_e|$$

$$\le \ C \sup_{e\in E_0} (\ell_e^{-1}\|[\varphi]\|_{L^\infty(e)})[\|\rho_1-\rho_2\| + (\sum_{T\in T} h_T^2\|\rho_1-\rho_2\|_{1,T}^2)^{1/2}]$$

$$\cdot (\sum_{e\in E_0} \ell_e\|\{\frac{\partial\Psi}{\partial n}\}\|_{0,e}^2)^{1/2}.$$

<u>Proof</u>. The inequality (5.5) is clear with $C = M(\|\varphi\|_{W_\infty^1(T)}+2)$.

For (5.6), note that

(5.8) $$|<[\alpha(\{\rho_1\}) - \alpha(\{\rho_2\})]\ \{\frac{\partial\varphi}{\partial n}\},\ [\Psi]>_e|$$

$$+ \ |<[\beta(\{\rho_1\}) - (\{\rho_2\})]\cdot n,[\Psi]>_e|$$

$$\le M(\|\varphi\|_{W_\infty^1(T)} + 1)\ \ \ell_e^{1/2}\|\{\rho_1-\rho_2\}\|_{0,e}\ \ell_e^{-1/2}\|[\Psi]\|_{0,e}.$$

Now, by (2.4),

$$\sum_{e\in E} \ell_e\|\{\rho_1-\rho_2\}\|_{0,e}^2 \le \sum_{T\in T}\ \ \sum_{e\in E_T} \|(\rho_1-\rho_2)|_T\|_{0,e}^2$$

$$\le C\sum_{T\in T} \|\rho_1-\rho_2\|_{0,T}^2 + h_T^2\|\rho_1-\rho_2\|_{1,T}^2].$$

Now, by (2.4),

$$\sum_{e\in E} \ell_e \| \{\rho_1-\rho_2\} \|_{0,e}^2 \le \sum_{T\in T} \sum_{e\in E_T} \| (\rho_1-\rho_2)|_T \|_{0,e}^2 \le C \sum_{T\in T} [\| \rho_1-\rho_2 \|_{0,T}^2$$

$$+ h_T^2 \| \rho_1-\rho_2 \|_{1,T}^2].$$

Thus (5.6) results from summing (5.8) over $e \in E_0$; (5.7) is obtained similarly.

It follows from the lemma and the fact that $[w] \equiv 0$ on $\cup E_0$ that the last four of the six terms on the right hand side of (5.4) are bounded by

$$C \, (\| \zeta \|^2 + \sum_{T\in T} h_T^2 \, \| \zeta \|_{1,T}^2) + \frac{\varepsilon}{2} \, |||\chi|||^2$$

where ε is the value furnished by Theorem 5.2. We next apply the triangle inequality and an inverse inequality to see that

$$(5.9) \qquad \sum_{T\in T} h_T^2 \, \| \zeta \|_{1,T}^2 \le 2 (\sum_{T\in T} h_T^2 \, \| \nu \|_{1,T}^2 + \sum_{T\in T} h_T^2 \| \mu \|_{1,T}^2)$$

$$\le C \, (\| \nu \|^2 + \sum_{T\in T} h_T^2 \, \| \mu \|_{1,T}^2).$$

Thus,

$$(\nu_t,\chi) + B(W;\nu,\chi) \le (\mu_t,\chi) + B(W;\mu,\chi)$$

$$+ C \, (\| \mu \|^2 + \| \nu \|^2 + h_T^2 \| \mu \|_{1,T}^2) + \frac{\varepsilon}{2} \, |||\chi|||^2$$

for all $\chi \in M$ and $t \in I$.

We now set $\chi = \nu(t) \in M$ and apply Theorem 5.2:

$$(5.10) \qquad \frac{d}{dt} \| \nu \|^2 + \varepsilon \, |||\nu|||^2 + J(\nu,\nu)$$

$$\le \quad 2(\mu_t,\nu) + 2B(W;\mu,\nu) + C \, (\| \mu \|^2 + \| \nu \|^2 + \sum h_T^2 \| \mu \|_{1,T}^2).$$

Dominating $B(W;\mu,\nu)$ by $C(\gamma_1)\ |||\mu|||^2 + \frac{\varepsilon}{2}\ |||\nu|||^2$ and integrating (5.10) over $t \in [0,t_0] \subseteq I$, we get

$$\|\nu(t_0)\|^2 + \int_0^{t_0} [\tfrac{\varepsilon}{2}|||\nu|||^2 + J(\nu,\nu)]dt$$

$$\leq \|\nu(0)\|^2 + \tfrac{1}{2}\|\nu\|^2_{L^\infty(L^2)} + C(\|\mu_t\|^2_{L^1(L^2)} + \int_0^{t_0} \|\nu\|^2 dt + \int_I |||\mu|||^2 dt).$$

As this holds for all $t_0 \in I$,

(5.11) $$\|\nu\|^2_{L^\infty(L^2)} + \int_I (|||\nu|||^2 + J(\nu,\nu))\,dt \leq C(\|\nu(0)\|^2 + \|\mu_t\|^2_{L^1(L^2)} + \int_I |||\mu|||^2 dt).$$

By (4.2),

$$\|\nu(0)\|^2 \leq 2(\|\mu(0)\|^2 + \|\zeta(0)\|^2) \leq Ch^2 \sum_{T\in\mathcal{T}} h_T^{2[j(T)-1]} \|w_0\|^2_{J(T),T}.$$

Since $\mu_t = \Pi w_t - w_t$, the property (A2) of the operator Π implies that

$$\|\mu_t\|^2 \leq C\sum h_T^{2[j(T)-1]} \|w_t\|^2_{j(T)-1,T}.$$

Finally, by (2.10),

$$|||\mu|||^2 \leq C\sum h_T^{2[j(T)-1]} \|w\|^2_{j(T),T}.$$

Hence (5.11) implies the assertion of Theorem 5.1 with ζ replaced by ν. Since the assertion is clear when ζ is replaced by μ, Theorem 5.1 is proven.

Remark 5.4. If M approximates smoothly we can strengthen the conclusion of the theorem. Instead of assuming that σ remains bounded on $\mathcal{U}E\times I$ we need only assume that σ stays bounded on $\partial\Omega\times I$ with no restriction made on the behavior of σ on interior edges except that $\sigma \geq \gamma_0$. The constant C in the statement of the theorem will depend on γ_2 rather than the larger quantity γ_1. In fact, the only points in the proof where γ_1 entered was in the bound $B(W;\mu,\nu) \leq C(\gamma_1) \,|||\mu|||^2 + \frac{\varepsilon}{2}\,|||\nu|||^2$ and in bounding $\int_I J(\mu,\mu)dt$ by the right hand side of (5.1) at the very last step. However, if M approximates smoothly, then μ is continuous and the interior penalty terms of $B(W;\mu,\nu)$ and $J(\mu,\mu)$ vanish; thus, only the values of σ on $\partial\Omega$ enter.

CHAPTER VI

AN OPTIMAL ORDER ESTIMATE IN $L^\infty(L^2)$

The bound for $\|\zeta\|_{L^\infty(L^2)}$ provided by Theorem 5.1 is not of optimal order in h. To achieve an optimal order bound we use the technique introduced by Wheeler [12] of comparing the approximate solution to an elliptic projection of the true solution. This approach could also be used to produce optimal order estimates of $|\|\zeta\||$ and $J(\zeta,\zeta)$, but these results would not be as satisfactory as those of the last section for two reasons. First, in this section we shall have to impose mild restrictions on the growth of the penalty function σ as a function of time; second, the bounds derived through the projection are not expressed entirely locally and hence are weaker in the case of a family of meshes which is not quasi-uniform. In this section we also assume that 0-regularity holds for the Laplace operator on Ω, i.e., $U \longmapsto \|\nabla U\|$ is supposed to be a norm on $H^2 \cap H_0^1$ equivalent to the H^2 norm. This is the case if, for instance, Ω is of class C^2 [1, Theorem 9.8] or if Ω is convex [9].

LEMMA 6.1. *Let* $t \in I$ *be fixed and suppose that* $\Phi \in H^2(T)$ *satisfies*

$$B(\Phi,\chi) = F(\chi), \qquad \chi \in M,$$

where $F\colon H^2(T) \to \mathbb{R}$ *is a linear map. Let* M_1 *and* M_2 *be constants for which*

$$|F(\rho)| \le M_1 |\|\rho\||; \qquad \rho \in H^2(T),$$

and

$$|F(\Psi)| \le M_2 \|\Psi\|_{2,\Omega} \qquad \Psi \in H^2 \cap H_0^1.$$

Then, $$\|\Phi\| \le C(|\|\Phi\|| + M_1)h + M_2,$$

where C *depends on* γ_1.

Proof. Define $\Psi \in H^2 \cap H_0^1$ by the relation

$$-\nabla\cdot(a(w)\nabla\Psi) = \Phi.$$

Then,

$$(6.1) \qquad \|\Psi\|_{2,\Omega} \leq C\|\Phi\|,$$

the constant depending on $\underline{a}$, $\|a(t,\cdot,\cdot)\|_{W^1_\infty}$ and $\|w(t,\cdot)\|_{W^1_\infty}$. Now, by (2.10) and (6.1),

$$\|\Phi\|^2 = (\Phi,-\nabla\cdot(a(w)\nabla\Psi)) = (a(w)\nabla\Phi,\nabla\Psi) - \sum_{e\in E} \left\langle a(w)[\Phi], \frac{\partial\Psi}{\partial n}\right\rangle_e$$

$$= B(\Phi,\Psi) = B(\Phi,\Psi-\Pi\Psi) - F(\Psi-\Pi\Psi) + F(\Psi)$$

$$\leq |B(\Phi,\Psi-\Pi\Psi)| + M_1 \;|||\Psi-\Pi\Psi||| + M_2 \;\|\Psi\|_{2,\Omega}$$

$$\leq C(\gamma_1)\;|||\Phi|||\;\;|||\Psi - \Pi\Psi||| + M_1|||\Psi-\Pi\Psi||| + M_2\|\Psi\|_{2,\Omega}$$

$$\leq [C(\gamma_1)(|||\Phi||| + M_1)h + M_2]\;\|\Phi\|.$$

The next theorem introduces the elliptic projection and contains the analysis of the interior penalty method for an associated elliptic problem. It generalizes to our situation Theorem 1 of [13].

THEOREM 6.2. There exists a unique function $Z: J \to M$ satisfying

$$B(Z,\chi) = B(w,\chi), \qquad \chi \in M.$$

The error $\eta = Z-w$ satisfies at each $t \in I$

$$(6.2) \qquad |||\eta|||^2 + J(\eta,\eta) \leq C\sum_{T\in T} h_T^{2[j(T)-1]}\;\|w\|^2_{j(T),T},$$

$$(6.3) \qquad \|\eta\|^2 \leq Ch^2\sum_{T\in T} h_T^{2[j(T)-1]}\;\|w\|^2_{j(T),T},$$

$$(6.4) \qquad |||\eta_t|||^2 \leq C[\sum_{T\in T} h_T^{2[j(T)-1]}\;(\|w\|^2_{j(T),T} + \|w_t\|^2_{j(T),T})],$$

$$(6.5) \qquad \|\eta_t\|^2 \leq Ch^2[\sum_{T\in T} h_T^{2[j(T)-1]}\;(\|w\|^2_{j(T),T} + \|w_t\|^2_{j(T),T})]$$

for $2 \leq j(T) \leq r+1$. The constants depend on γ_1 and γ_3.

Proof. The uniqueness of Z and therefore its existence follow from the positivity of the form B, which was established in Theorem 5.2. Moreover, since

$$B(\eta,\chi) = 0, \qquad \chi \in M,$$

(6.3) is a consequence of (6.2) and Lemma 6.1. To prove (6.2) we apply Theorem 5.2 to $\theta = Z - \Pi w$. It follows that

$$(6.6) \qquad |||\theta|||^2 + J(\theta,\theta) \leq CB(\theta,\theta) = CB(w-\Pi w,\theta) \leq C(\gamma_1)\,|||w-\Pi w|||\ \ |||\theta|||.$$

Thus, by (2.10),

$$(6.7) \qquad |||\eta||| \leq |||\theta||| + |||w-\Pi w||| \leq C|||w-\Pi w||| \leq C\Big(\sum_{T\in\mathcal{T}} h_T^{2[j(T)-1]}\,\|w\|^2_{j(T),T}\Big)^{1/2}.$$

Since $J(\eta,\eta) \leq \gamma_1\ |||\eta|||_1^2$, (6.2) follows from (6.6) and (6.7).

To estimate η_t, differentiate the defining equation for Z to obtain

$$B(\eta_t,\chi) + B'(\eta,\chi) = 0, \qquad \chi \in M,$$

where

$$B'(\varphi,\Psi) = \Big(\frac{d}{dt}\,a(w)\nabla\varphi,\nabla\Psi\Big) - \sum_{e\in E}\Big[\Big\langle\frac{d}{dt}\,a(w)\,[\varphi],\Big\{\frac{\partial\Psi}{\partial n}\Big\}\Big\rangle_e$$

$$+\Big\langle\frac{d}{dt}\,a(w)\Big\{\frac{\partial\varphi}{\partial n}\Big\},\ [\Psi]\Big\rangle_e\Big] + \sum_{e\in E}\ell_e^{-1}\,\langle\sigma_t[\varphi],[\Psi]\rangle_e.$$

Note that

$$|B'(\eta,\rho)| \leq C(\gamma_3)\ |||\eta|||\ \ \ |||\rho|||\,, \qquad \rho \in H^2(\mathcal{T}).$$

Moreover, for $\Psi \in H^2 \cap H_0^1$ an integration by parts using (2.1) shows that

$$|B'(\eta,\Psi)| = |(-\eta,\nabla\cdot[\frac{d}{dt}\,a(w)\nabla\Psi])| \leq C\|\eta\|\ \ \|\Psi\|_{2,\Omega}.$$

Thus, Lemma 6.1 applies, and

$$\|\eta_t\| \leq C[(|||\eta_t||| + |||\eta|||)h + \|\eta\|].$$

Consequently, (6.2)-(6.4) imply (6.5), and it remains only to demonstrate (6.4).

Recall that

$$(6.8) \qquad |||\frac{\partial}{\partial t}(w-\Pi w)|||^2 \le C \sum_{T\in\mathcal{T}} h_T^{2[j(T)-1]} \; \|w_t\|^2_{j(T),T},$$

$$2 \le j(T) \le r+1.$$

Also, by Theorem 5.2,

$$(6.9) \qquad |||\theta_t|||^2 + J(\theta_t,\theta_t) \le CB(\theta_t,\theta_t)$$

$$= C[B(\frac{\partial}{\partial t}(w-\Pi w),\theta_t) - B'(\eta,\theta_t)]$$

$$\le C(\gamma_1)(|||\frac{\partial}{\partial t}(w-\Pi w)||| + |||\eta|||)\; |||\theta_t||| + \sum_{e\in E} \ell_e^{-1} |\langle \sigma_t[\eta],[\theta_t]\rangle_e|$$

$$\le \frac{1}{2}(|||\theta_t|||^2 + \sum_{e\in E} \ell_e^{-1} \|\sigma^{1/2}[\theta_t]\|^2_{0,e})$$

$$+ C(|||\frac{\partial}{\partial t}(w-\Pi w)|||^2 + |||\eta|||^2 + \sum_{e\in E} \ell_e^{-1} \|\sigma_t\sigma^{-1/2}[\eta]\|^2_{0,e}).$$

Therefore,

$$(6.10) \qquad |||\eta_t|||^2 \le 2(|||\theta_t|||^2 + |||\frac{\partial}{\partial t}(w-\Pi w)|||^2)$$

$$\le C[|||\frac{\partial}{\partial t}(w-\Pi w)|||^2 + |||\eta|||^2 + \gamma_3^2\gamma_0^{-2} J(\eta,\eta)],$$

and (6.4) follows from (6.2) and (6.8).

In analogy with (5.3) we shall use the decomposition $\eta-\xi=\zeta$ where $\xi = Z - W \in M$. Substituting this into (5.2) leads to the relation

$$(6.11) \quad (\eta_t,\chi) + B(W;\eta,\chi) = (\xi_t,\chi) + B(W;\xi,\chi) - (b(W) - b(w),\nabla\chi)$$

$$+ \sum_{e\in E_0} \langle [b(\{W\} - b(w)]\cdot n, [\chi]\rangle_e + (f(W) - f(w),\chi) + A(w;w,\chi) - A(W;w,\chi)$$

Now,

$$B(W;\eta,\chi) - A(w;w,\chi) + A(W;w,\chi)$$

$$= B(W;\eta,\chi) - B(w;w,\chi) + B(W;w,\chi)$$

$$= B(W;Z,\chi) - B(w;Z,\chi)$$

$$= A(W;Z,\chi) - A(w;Z,\chi).$$

Thus,

$$(6.12)\quad (\xi_t,\chi) + B(W;\xi,\chi) = (\eta_t,\chi) + [A(W;Z,\chi)-A(w;Z,\chi)]$$
$$+ (b(W)-b(w),\nabla\chi) - \sum_{e\in E_0} \langle [b(\{W\})-b(w)]\cdot n,[\chi]\rangle_e$$
$$- (f(W)-f(w),\chi).$$

The last four of the five terms on the right hand side of (6.12) can be estimated by Lemma 5.3, and

$$(\xi_t,\chi) + B(W;\xi,\chi) \le (\eta_t,\chi) + C(\|\eta\|^2+\|\xi\|^2+\sum_{T\in\mathcal{T}} h_T^2\|\eta\|_{1,T}^2)$$
$$+ \frac{\varepsilon}{4}|||\chi|||^2,$$

where C depends on $\|Z\|_{W^1_\infty(\mathcal{T})}$ and $\sup_{e\in E_0} \ell_e^{-1}\|[Z]\|_{L^\infty(e)}$ and ε is derived from Theorem 5.2. Note also that the triangle inequality and an inverse inequality have been used in the same manner as in (5.9). If the choice $\chi=\xi$ is made and the argument by which (5.11) was derived from (5.10) is adapted, then the inequality

$$\|\xi\|^2_{L^\infty(L^2)} + \int_I [|||\xi|||^2 + J(\xi,\xi)]dt$$
$$\le C(\|\xi(0)\|^2 + \|\eta_t\|^2_{L^1(L^2)} + \|\eta\|^2_{L^2(L^2)} + \sum h_T^2\|\eta\|^2_{L^2(H^1(T))}).$$
$$\le Ch^2\sum h_T^{2[j(T)-1]}(\|w_0\|^2_{j(T),T} + \|w\|^2_{L^2(H^{j(T)}(T))} + \|w_t\|^2_{L^1(H^{j(T)}(T))}).$$

results from (4.2) and Theorem 6.2. This estimate together with (6.3) implies the following theorem.

THEOREM 6.3. *There exists a constant* C *depending on* γ_1, γ_3, $\| Z \|_{L^\infty(W^1_\infty(T))}$ *and* $\sup_{e \in E_0} \ell_e^{-1} \| [Z] \|_{L^\infty(e)}$ *such that the error* ζ *satisfies the inequality*

$$\| \zeta \|_{L^\infty(L^2)} \le Ch[\sum_{T \in \mathcal{T}} h_T^{2[j(T)-1]} (\| w_0 \|^2_{j(T),T} + \| w \|^2_{L^\infty(H^{j(T)}(T))} + \| w_t \|^2_{L^1(H^{j(T)}(T))})]$$

for $2 \le j(T) \le r+1$.

Below we shall remark on cases for which bounds for $\| Z \|_{L^\infty(W^1_\infty(T))}$ and $\sup_{e \in E} \ell_e^{-1} \| [Z] \|_{L^\infty(e)}$ are readily available. We have also had to assume that σ and σ_t are bounded. If M approximates smoothly, we can considerably weaken this constraint. Instead of the dependence on γ_1 and γ_3 in Theorem 6.3, dependence on γ_2 and γ_4 suffices. The necessary alterations to the argument will be outlined briefly. First a slightly modified form of Lemma 6.1 is needed.

LEMMA 6.4. *Let* $t \in I$ *be fixed and suppose that* $\Phi \in H^2(T)$ *satisfies*

$$B(\Phi,\chi) = F(\chi), \qquad \chi \in M,$$

where $F\colon H^2(T) \to \mathbb{R}$ *is a linear map. Let* M_1 *and* M_2 *be constants for which*

$$|F(\rho)| \le M_1 |||\rho|||, \qquad \rho \in H^2(T) \cap C^0(\Omega),$$

and

$$|F(\Psi)| \le M_2 \|\Psi\|_{2,\Omega}, \qquad \Psi \in H^2 \cap H^1_0.$$

Then, $\|\Phi\| \le C(|||\Phi||| + M_1)h + M_2$, *where* C *depends on* γ_2.

Note that the test function ρ now varies only over the continuous subspace of $H^2(T)$. The proof is essentially the same as for Lemma 6.1. Since the function Ψ introduced in that proof is continuous, the hypothesis of smooth approximation implies that $\rho = \Psi - \Pi\Psi$ is continuous. So, the weakened hypothesis suffices, and, as the interior penalty terms of $B(\Phi, \Psi-\Pi\Psi)$ vanish, the improved dependence results.

To adapt the proof of Theorem 6.2, note that similarly γ_1 may be replaced by γ_2 in (6.6). Since $[\eta] = [\theta]$ on $\mathcal{E}_0$, (6.2) then follows from (6.6) and (6.7) with the improved constant. Then, (6.3) follows from Lemma 6.4.

From the definition of B'

$$|B'(\eta,\rho)| \leq C\,(\sup_{\partial\Omega\times I} |\sigma_t|)\ |||\eta|||\ |||\rho|||,\quad \rho \in H^2(\mathcal{T})\cap C^0(\Omega).$$

Moreover, $\gamma_2\cdot\gamma_4$ is a bound for $|\sigma_t|$ on $\partial\Omega\times I$. Thus, Lemma 6.4 implies (6.5) with the constant depending only on γ_2 and γ_4 if such a constant suffices in (6.4). This is possible because the constant $\gamma_3^2\cdot\gamma_0^{-2}$ which was introduced in passing from (6.9) to (6.10) can equally well be replaced by γ_4^2. Consequently, the improved version of Theorem 6.2 follows, and Theorem 6.3 results as before.

We conclude this section by giving two cases in which the dependence of the constant in Theorem 6.3 on $|||Z|||_{W^1_\infty(\mathcal{T})}$ and $\sup_{e\in\mathcal{E}_0} \ell_e^{-1}\|[Z]\|_{L^\infty(e)}$ can be suppressed. This dependence was introduced in bounding $|A(W;Z,\chi) - A(w;Z,\chi)|$ by Lemma 5.3. Hence, if the coefficient a is independent of w, then the constant can be taken independent of these quantities. In particular, this is the case if the differential equation is linear or even semilinear.

Also, in the case of a quasi-uniform family of edge-to-edge meshes the dependence of the constant C of Theorem 6.3 on Z reduces to dependence on the solution.

THEOREM 6.5. <u>Suppose that</u> $\mathcal{I}w$ <u>is continuous</u>. <u>Let</u> $M = \sup_{T\in\mathcal{T}} h/h_T$. <u>Then there exists a constant</u> $C = C(M, \|w\|_{2,\Omega})$ <u>such that</u>

$$\|Z\|_{W^1_\infty(\mathcal{T})} + \sup_{e\in\mathcal{E}_0} \ell_e^{-1} \|[Z]\|_{L^\infty(e)} \leq C, \quad t \in I.$$

<u>Proof</u>. Set $\Theta = Z - \mathcal{I}w$. From an inverse inequality, (6.2), and (2.3), we obtain

$$\|\theta\|_{W^1_\infty(\mathcal{T})} = \sup_{T\in\mathcal{T}} \|\theta\|_{W^1_\infty(T)} \leq C \sup_{T\in\mathcal{T}} h_T^{-1}\|\theta\|_{1,T}$$

$$\leq C\, M\, h^{-1}\|\theta\|_{1,\mathcal{T}} \leq C\, M\, h^{-1}(\|\eta\|_{1,\mathcal{T}} + \|w - \mathcal{I}_w\|_{1,\mathcal{T}})$$

$$\leq C\, M\, \|w\|_{2,\Omega}.$$

Since $\|\mathcal{I}w\|_{W^1_\infty(\Omega)} \leq C\|w\|_{W^1_\infty(\mathcal{T})}$, $\|Z\|_{W^1_\infty(\mathcal{T})} \leq C$.

Finally, for $e \in \mathcal{E}_0$ a one-dimensional inverse inequality and (6.2) imply that

$$\ell_e^{-1} \|[Z]\|_{L^\infty(e)} = \ell_e^{-1} \|[\theta]\|_{L^\infty(e)}$$

$$\leq C\ell_e^{-3/2}\|[\theta]\|_{0,e} = C\,\ell_e^{-3/2}\|[\eta]\|_{0,e} \leq C\|w\|_{2,\Omega}.$$

CHAPTER VII

THE SECOND ENERGY ESTIMATE

In this section we prove an $O(h^r)$ estimate on $\|\zeta_t\|_{L^2(L^2)}$ + sup $|||\zeta|||$. In particular an $O(h^r)$ estimate on $|||\zeta(0)|||$ must hold. Thus, assume that $W(0)$ is chosen to satisfy

$$(7.1) \quad |||W(0)-w_0||| \leq C \sum_{T\in\mathcal{T}} h_T^{2[j(T)-1]} \|w_0\|^2_{j(T),T}, \quad 2\leq j(T)\leq r+1,$$

in addition to (4.2). The elliptic projection of w_0 is a satisfactory choice (by Theorem 6.2), as is the interpolant $\mathcal{I}w_0$ (by (2.10)).

The L^2 projection of w_0 into $\mathcal{M}$ may also be selected to initialize the procedure. Indeed Lemma 2.5 and an inverse assumption imply that, for any $W(0) \in \mathcal{M}$,

$$|||W(0)- \mathcal{I}w_0|||^2 \leq C \sum_{T\in\mathcal{T}} h_T^{-2}\|W(0)-\mathcal{I}w_0\|^2_{0,T}$$

$$\leq C \sum_{T\in\mathcal{T}} h_T^{-2} (\|W(0)-w_0\|^2_{0,T} + \|w_0 - \mathcal{I}w_0\|^2_{0,T}).$$

Now if $W(0)$ is the L^2 projection of w_0, then $W(0)|_T$ is the L^2 projection of $w_0|_T$ into $\mathcal{M}_r(T)$ for each $T \in \mathcal{T}$. Consequently, $\|W(0)-w_0\|_{0,T} \leq Ch_T^{j(T)}\|w_0\|_{j(T),T}$, so $|||W(0)-\mathcal{I}w_0|||^2$

$\leq C\sum h_T^{2[j(T)-1]} \|w_0\|^2_{j(T),T}$, and (7.1) is satisfied.

Because of the nonlinearity and time-dependence of our differential operator it does not commute with time differentiation, a fact which accounts for a multitude of terms appearing in the analysis below. We begin by considering some of these terms separately. Fix a function $j: \mathcal{T} \to \mathbb{Z}$ such that $2 \le j(T) \le r+1$ for all T, and let

$$H_j = \sum_{T \in \mathcal{T}} h_T^{2[j(T)-1]} \left(\|w_t\|^2_{L^2(H^{j(T)}(T))} + \|w\|^2_{L^2(H^{j(T)}(T))} + \|w_0\|^2_{j(T),T} \right).$$

LEMMA 7.1. <u>Given</u> $\varepsilon > 0$, <u>there exist positive constants</u> $C=C(\varepsilon,\gamma_1)$ <u>and</u> $h^* = h^*(\varepsilon)$ <u>such that for</u> $h \le h^*$ <u>and</u> $t \in I$,

$$\int_0^t \{ [A(W;w;\zeta_t)-A(w;w,\zeta_t)] + (b(W)-b(w),\nabla\zeta_t) \tag{7.2}$$

$$+ \sum_{e \in E_0} \langle [b(\{W\})-b(w)]\cdot n, [\zeta_t] \rangle_e - (f(W)-f(w),\zeta_t)\}dt$$

$$\le \varepsilon |||\zeta(t)|||^2 + \varepsilon \int_0^t \|\zeta_t\|^2 dt + CH_j .$$

<u>Proof</u>. We consider only $\int_0^t [A(W;w,\zeta_t)-A(w;w,\zeta_t)]dt$, since the lower order terms can be treated similarly. From integration by parts it follows that

$$\int_0^t [A(W;w,\zeta_t)-A(w;w,\zeta_t)]dt = J_1 + J_2 + J_3 + J_4,$$

where

$$J_1 = [A(W;w,\zeta) - A(w;w,\zeta)]\Big|_0^t,$$

$$J_2 = -\int_0^t [A(W;w_t,\zeta) - A(w;w_t,\zeta)]dt,$$

$$J_3 = -\int_0^t ([\frac{d}{dt} a(W) - \frac{d}{dt} a(w)]\nabla w, \nabla\zeta)\,dt,$$

$$J_4 = -\int_0^t \sum_{e\in E_0} < [\frac{d}{dt} a(W) - \frac{d}{dt} a(w)]\{\frac{\partial w}{\partial n}\}, [\zeta] >_e dt.$$

Now,

$$|A(W;w,\zeta) - A(w;w,\zeta)| \le C\{[\|\zeta\| + (\sum h_T^2\|\zeta\|_{1,T}^2)^{1/2}]\, |||\zeta|||$$

by Lemma 5.3. Thus, if h is sufficiently small, then

$$|J_1| \le \varepsilon(|||\zeta(t)|||^2 + |||\zeta(0)|||^2 + C\|\zeta\|^2_{L^\infty(L^2)}$$

which is bounded by the right hand side of (7.2) by (7.1) and Theorem 5.1.

Next, the same lemma shows that

$$|J_2| \le C[\|\zeta\|_{L^2(L^2)} + (\sum h_T^2\|\zeta\|^2_{L^2(H^1(T))})^{1/2}](\int_0^t |||\zeta|||^2 dt)^{1/2}$$

$$\le C\int_0^t |||\zeta|||^2 dt \le CH_j.$$

To estimate J_3 we use the decomposition

$$\frac{d}{dt} a(W) - \frac{d}{dt} a(w) = [a_t(W) - a_t(w)] + a_w(W)\zeta_t + [a_w(W) - a_w(w)]w_t.$$

Applying (5.5), we obtain the inequality

$$|J_3| \le \frac{\varepsilon}{2}\int_0^t \|\zeta_t\|^2 dt + C\int_0^t |||\zeta|||^2 dt.$$

Similarly,

$$|J_4| \le C[\|\zeta\|^2_{L^2(L^2)} + (\sum h_T^2 \|\zeta\|^2_{L^2(H^1(T))})^{1/2} + (\int_0^t \|\zeta_t\|^2 dt)^{1/2}$$

$$+ (\int_0^t \sum h_T^2 \|\zeta_t\|^2_{1,T} dt)^{1/2}](\int_0^t |||\zeta|||^2 dt)^{1/2}.$$

Now apply an inverse inequality as was done in (5.9), but with ζ, ν, and μ replaced by their time derivatives. Hence,

$$\int_0^t \sum h_T^2 \|\zeta_t\|^2_{1,T} dt \le C\int_0^t (\|\zeta_t\|^2 + \|\mu_t\|^2 + \sum h_T^2 \|\mu_t\|^2_{1,T}) dt \tag{7.3}$$

$$\le C\int_0^t (\|\zeta_t\|^2 + \sum h_T^{2[j(T)-1]} \|w_t\|^2_{L^2(H^{j(T)-1}(T))}) dt.$$

Therefore,

$$|J_4| \le \frac{\varepsilon}{2}\int_0^t \|\zeta_t\|^2 dt + C[\int_0^t |||\zeta|||^2 dt$$

$$+ \sum h_T^{2[j(T)-1]} \|w_t\|^2_{L^2(H^{j(T)-1}(T))}]$$

which is again of the desired form. This completes the proof of the lemma.

LEMMA 7.2. Given $\varepsilon > 0$, there exists $C = C(\gamma_1, \gamma_3 \|\zeta\|_{L^\infty(W^1_\infty(T))})$ such that

$$(7.4)\int_0^t \left|\frac{d}{dt} B(W;\zeta,\zeta) - 2B(W;\zeta,\zeta_t)\right| dt \le \varepsilon \int_0^t \|\zeta_t\|^2 dt + CH_j .$$

Proof. The integrand on the left hand side of (7.4) can be rewritten as

$$(7.5) \qquad \Big| \left(\frac{d}{dt} a(W)\nabla\zeta, \nabla\zeta\right) - \sum_{e\in E_0} \left\langle \frac{d}{dt} a(\{W\}) \left\{\frac{\partial\zeta}{\partial n}\right\}, [\zeta]\right\rangle_e$$
$$-2\left\langle \frac{d}{dt} a(g)\frac{\partial\zeta}{\partial n}, \zeta\right\rangle + \sum_{e\in E} \ell_e^{-1} \left\langle \sigma_t[\zeta], [\zeta]\right\rangle_e \Big| .$$

We bound these terms individually. First,

$$\left|\left(\frac{d}{dt} a(W)\nabla\zeta, \nabla\zeta\right)\right| \le |(a_w(W)\zeta_t\nabla\zeta, \nabla\zeta)| + |(a_w(W)w_t\nabla\zeta, \nabla\zeta)| + |(a_t(W)\nabla\zeta, \nabla\zeta)|$$

$$\le \frac{\varepsilon}{2}\|\zeta_t\|^2 + C\left(\|\zeta\|_{L^\infty(W^1_\infty(T))}\right)\|\zeta\|^2_{1,T},$$

the time integral of which is bounded by the right hand side of (7.4) by Theorem 5.1.

Similarly,

$$\sum_{e\in E_0} \left|\left\langle \frac{d}{dt} a(\{W\}) \left\{\frac{\partial\zeta}{\partial n}\right\}, [\zeta]\right\rangle_e\right| \le \sum \left|\left\langle a_w(\{W\})(\{\zeta_t\}\left\{\frac{\partial\zeta}{\partial n}\right\}, [\zeta]\right\rangle_e\right|$$
$$+\sum\left|\left\langle a_w(\{W\})w_t\left\{\frac{\partial\zeta}{\partial n}\right\}, [\zeta]\right\rangle_e\right| + \sum \left|\left\langle a_t(\{W\}) \left\{\frac{\partial\zeta}{\partial n}\right\}, [\zeta]\right\rangle_e\right| .$$

The latter two terms are clearly bounded by $C|||\zeta|||^2$, while by (5.6) and (7.3),

$$\sum\left|\left\langle a_w(\{W\}) \{\zeta_t\}\left\{\frac{\partial\zeta}{\partial n}\right\}, [\zeta]\right\rangle_e\right| \le C(\|\zeta\|_{L^\infty(W^1_\infty(T))})(\|\zeta_t\|^2 + h_T^2\|\zeta_t\|^2)^{1/2}|||\zeta|||$$
$$\le \frac{\varepsilon}{2}\|\zeta_t\|^2 + C(\|\zeta\|_{L^\infty(W^1_\infty(T))})(|||\zeta|||^2 + \sum h_T^{2[j(T)-1]}\|W_t\|^2_{j(T),T}) .$$

The third term of (7.5) presents no difficulty, and

$$\left|\sum \ell_e^{-1} \langle \sigma_t[\zeta],\ [\zeta] \rangle_e\right| < \gamma_3 \cdot \gamma_0 \sum \ell_e^{-1} \left| \langle \sigma[\zeta]\ ,[\zeta] \rangle_e \right|.$$

Thus the proof is completed by an appeal to Theorem 5.1.

THEOREM 7.3 <u>There exists</u> $C = C(\gamma_1,\gamma_3,\gamma_5,\|\zeta\|_{L^\infty(W^1_\infty(T))})$ <u>so that for</u> h <u>sufficiently small and any selection of integers</u> $j(T) \in [2, r+1]$, <u>the error</u> ζ <u>satisfies the inequality</u>

$$\|\zeta_t\|^2_{L^2(L^2)} + \sup_I |||\zeta|||^2 + \sup_I J(\zeta,\zeta)$$

$$\leq C \sum h_T^{2[j(T)-1]} \left(\|w_t\|^2_{L^2(H^{j(T)-1}(T))} + \|w\|^2_{L^2(H^{j(T)}(T))} + \|w_0\|^2_{j(T),T}\right).$$

<u>Proof</u>. Setting $\chi = \zeta_t - \mu_t$ in (5.2) yields the inequality

$$(7.6)\qquad \|\zeta_t\|^2 + B(W;\zeta,\zeta_t) \leq (\zeta_t,\mu_t) + B(W;\zeta,\mu_t) + F(\zeta_t) - F(\mu_t),$$

where by definition

$$F(\varphi) = -\ (b(W) - b(w), \nabla\varphi) + \sum_{e\in E_0} \langle [b(\{W\}) - b(w)]\cdot n, [\varphi] \rangle_e$$

$$+\quad (f(W) - f(w),\ \varphi) + A(w;w,\ \varphi) - A(W;w,\ \varphi).$$

In view of Lemma 5.3,

$$|B(W;\zeta,\mu_t)| + |F(\mu_t)| \leq C\ |||\zeta|||\ \ |||\mu_t|||.$$

Consequently,

$$\int_I (|B(W;\zeta,\mu_t)| + |F(\mu_t)|)dt \leq CH_j.$$

Also, $|(\zeta_t,\mu_t)| \leq \frac{1}{2}\|\zeta_t\|^2 + CH_j$. Applying these estimates and the lemmas of this section to (7.6) yields the relation

$$\int_0^t [\|\zeta_t\|^2 + \frac{1}{2}\frac{d}{dt} B(W;\zeta,\zeta)]dt \leq \frac{1}{2}\int_0^t \|\zeta_t\|^2 dt + \frac{\varepsilon}{4} |||\zeta(t)|||^2 + CH_j,\ t \in I,$$

where $\varepsilon > 0$ is the value furnished by Theorem 5.2. Application of that theorem and (7.1) completes the demonstration of Theorem 7.3.

Remark 7.4. The dependence of the constant on γ_1 in Theorem 7.3 results from the frequent reliance on the first energy estimate. Hence, we can replace γ_1 by γ_2 in the conclusion if M approximates smoothly (see Remark 5.4). Moreover, in that case the dependence of γ_3 can be reduced to dependence on γ_4. Finally, the dependence on γ_5 arose only at the last step in bounding $J(\zeta(0), \zeta(0))$. Hence, if $W(0)$ is taken to be a continuous function (which is, in principle, possible if M approximates smoothly), then dependence on γ_4 suffices. Thus, if M approximates smoothly, the second energy estimate is valid independent of a bound on the interior penalty function or its time derivative, and depends only on their ratio.

An inspection of the proof reveals that the dependence of the constant on $\|\zeta\|_{L^\infty(W^1_\infty(T))}$ is unnecessary if the coefficient $a(x, t, w)$ is independent of w.

CHAPTER VIII

GENERALIZATIONS AND EXTENSIONS

8.1. A More General Equation.

We sketch briefly the modifications to the method and the proof of Theorem 5.1 necessary to obtain the first energy estimate for the more general differential equation

$$c(x,t,w)\frac{\partial w}{\partial t}-\sum_{p,q=1}^{2}\frac{\partial}{\partial x_p}a_{pq}(x,t,w)\frac{\partial w}{\partial x_q}=f(x,t,w,\nabla w)$$

Here, $f\in C_b^1(\bar{\Omega}\times I\times \mathbb{R}^3)$, apq, $c\in C_b^1(\bar{\Omega}\times I\times \mathbb{R})$, and the matrix (a_{pq}) is symmetric. Moreover, we assume the existence of positive constants $\underline{a}$ and $\underline{c}$ such that

$$\underline{a}\,|\,\omega\,|^2\leq\sum a_{pq}(x,\,t,\,\rho)\;\omega_p\,\omega_q,$$
$$\underline{c}\leq C(x,\,t,\,\rho)$$

for all $\omega\in R^2$ and $(x,\,t,\,\rho)\in\bar{\Omega}\times I\times\mathbb{R}$.

Let $n_e=(n_e^1\,,\,n_e^2)$ and define the conormal $n_e^*=n_e^*\,(x,\,t,\,\rho)$ $=(a_{11}n_e^1+a_{21}n_e^2,\;a_{12}n_e^1+a_{22}n_e^2)$. Redefine the form A to be

$$A(\rho;\varphi,\Psi)=\sum_{p,q}(a_{pq}(\rho)\frac{\partial\varphi}{\partial x_p},\frac{\partial\Psi}{\partial x_q})-\sum_{e\in E}[<\{\frac{\partial\varphi}{\partial n_e^*}\},[\Psi]>_e+<[\varphi],\{\frac{\partial\Psi}{\partial n_e^*}\}>_e].$$

where the final argument of n_e^* is understood to be ρ if $e\in E_0$ and g if $e\in E_\partial$. Again set B = A+J. Then Theorem 5.2 follows as before.

Define W: I $\longrightarrow$ M by

$$(c(W)W_t,\chi)+B(W;W,\chi)=(f(W,\nabla W),\chi)-<g,\frac{\partial\chi}{\partial n^*}>-\sum_{e\in E_\partial}\ell_e^{-1}<\sigma g,\chi>_e$$

for all $\chi\in M$, as the same equations hold with w replacing W.

The relation below is analagous to (5.4):

$$(c(W)\nu_t,\chi) + B(W;\nu,\chi) = (c(W)\mu_t,\chi) + B(W;\mu,\chi)$$

$$- (f(W),\nabla W) - f(w,\nabla w),\chi) + [A(W;w,\chi) - A(w;w,\chi)] + ([c(W)-c(w)]w_t,\chi).$$

As before, the last two of the five terms on the right hand side are bounded by

$$C\,(\|\mu\|^2 + \|\nu\|^2 + \sum_{T\in\mathcal{T}} h_T^2\,\|\mu\|^2_{1,T}) + \frac{\varepsilon}{2}\,|||\chi|||^2.$$

Moreover,

$$|(f(W,\nabla W)-f(w,\nabla w),\chi)| \le |(f(\bar{W},\nabla W)-f(w,\nabla W),\chi)| + |(f(w,\nabla W)-f(w,\nabla w),\chi)|$$

$$\le C\,(\|\zeta\| + \|\nabla\zeta\|)\,\|\chi\| \le \frac{\varepsilon}{4}\,|||\nu|||^2 + \frac{\varepsilon}{4}\,|||\mu||| + C\|\chi\|^2.$$

Thus, setting $\chi=\nu$, we obtain

$$(8.1)\qquad (c(W)\nu_t,\nu) + \varepsilon|||\nu|||^2 + \frac{1}{2}J(\nu,\nu) \le \|\nu\|\,\|\mu_t\| + \frac{3\varepsilon}{4}|||\nu|||^2$$

$$+ C(|||\mu|||^2 + \|\nu\|^2).$$

Now,

$$(c(W)\nu_t,\nu) = \frac{d}{dt}\int_\Omega\int_0^\nu \rho c(\mu-\rho)\,d\rho\,dx - \int_\Omega\int_0^\nu \rho\frac{d}{dt}c(\mu-\rho)\,d\rho\,dx.$$

Since $\displaystyle\int_\Omega\int_0^\nu \rho c(\mu-\rho)\,d\rho\,dx \ge \frac{1}{2}\,\underline{c}\|\nu\|^2$

and

$$\left|\int_\Omega\int_0^\nu \rho\,\frac{d}{dt}\,c(\mu-\rho)\,d\rho\,dx\right| \le C\|\nu\|^2,$$

integration of (8.1) over $(0,t)\subseteq I$ achieves Theorem 5.1.

8.2. Neumann Boundary Conditions.

We now indicate the alterations to be made to the interior penalty method and the foregoing analysis if the Dirichlet boundary condition (1.1b) is replaced by the natural Neumann condition

$$\text{(8.2)} \qquad a(x,t,w(x,t))\ \frac{\partial w}{\partial n}(x,t) + b(x,t,w(x,t))\cdot n = g(x,t),\ (x,t)\in\partial\Omega\times I.$$

In this case the definitions of the forms A, J, and B, and the associated norm are altered as follows:

$$A(\rho;\varphi,\Psi) = (a(\rho)\nabla\varphi,\nabla\Psi) - \sum_{e\in E_0} [\langle a(\{\rho\})\{\tfrac{\partial\varphi}{\partial n}\},[\Psi]\rangle_e + \langle a(\{\rho\})[\varphi],\{\tfrac{\partial\Psi}{\partial n}\}\rangle_e,$$

$$J(\varphi,\Psi) = \sum_{e\in E_0} \ell_e^{-1}\langle\sigma[\varphi]\ ,\ [\Psi]\rangle_e,$$

$$B(\rho;\varphi,\Psi) = A(\rho;\varphi,\Psi) + J(\varphi,\Psi)$$

$$B(\varphi,\Psi) = B(w;\varphi,\Psi),$$

$$|||\varphi|||^2 = \|\varphi\|_{1,T}^2 + \sum_{e\in E_0} \ell_e\ \|\{\tfrac{\partial\varphi}{\partial n}\}\|_{0,e}^2 + \sum_{e\in E_0} \ell_e^{-1}\|[\varphi]\|_{0,e}^2,$$

for $\rho,\varphi,\Psi \in H^2(T)$. Note that in the case of a natural boundry condition the penalty function σ is defined only on the interior edges.

The boundedness inequality (2.8) remains valid. However, the coercivity result stated in Theorem 5.2 does not, as can be seen by taking φ to be constant. Nevertheless, the proof of that theorem can be employed to show that

$$A(\rho;\varphi,\varphi) \geq \frac{1}{2}\ \underline{a}\|\nabla\varphi\|^2 - \delta'\|\varphi\|^2 - c\sum_{e\in E_0} \ell_e^{-1}\|[\varphi]\|_{0,e}^2,$$

where $\delta' > 0$ can be taken arbitrarily small if C is correspondingly large. Hence we get the following Gårding inequality.

THEOREM 8.1. <u>There exists</u> $\varepsilon > 0$ <u>such that for sufficiently large</u> γ_0,

$$B(\rho;\varphi,\varphi) + \|\varphi\|^2 \geq \varepsilon \, |||\varphi|||^2 + \frac{1}{2} J(\varphi,\varphi)$$

<u>for all</u> $\varphi \in M$ <u>and</u> $\rho \in H^2(T)$.

The finite element solution $W: I \to M$ is defined by the equations

$$(W_t,\chi) + B(W;W,\chi) + (b(W),\nabla\chi) - \sum_{e\in E_0} \langle b(\{W\})\cdot n,[\chi]\rangle_e = (f(W),\chi) - \langle g,\chi\rangle$$

for all $\chi \in M$. These equations are also satisfied when W is replaced by the exact solution, w. Equation (5.2) for the error $\zeta=W-w$ then holds. Decomposing ζ as $\mu-\nu$ we obtain (5.4) exactly as in the case of Dirichlet boundary conditions. One can then extract (5.10) as before, noting only that we must add $\|\nu\|^2$ to both sides. The proof of the first energy estimate can then be completed without difficulty, giving the statement of Theorem 5.1 unaltered except that γ_1 and the sum on the left hand side are taken with reference to E_0 rather than all of E. If M approximates smoothly, then the constant in Theorem 5.1 is entirely independent of $\sigma \geq \gamma_0$.

In order to adapt the results of section 6, the elliptic projection Z must be defined in such a manner that it does not hinge on the strict coercivity of the form B. Let $Z: I\to M$ be defined by the equations

$$B(Z,\chi) + (Z,\chi) = B(w,\chi) + (w,\chi), \qquad \chi \in M, \quad t \in I.$$

The following lemma is the appropriate variant of Lemma 6.1.

LEMMA 8.2. *Let* $t \in I$ *be fixed and suppose that* $\Phi \in H^2(T)$ *satisfies the equations*

$$B(\Phi,\chi) + (\Phi,\chi) = F(\chi), \qquad \chi \in M,$$

where $F\colon H^2(T) \to \mathbb{R}$ *is a linear map. Let* M_1 *and* M_2 *be constants for which*

$$|F(\rho)| \le M_1 |||\rho||| \quad \text{if } \rho \in H^2(T),$$

and

$$|F(\Psi)| \le M_2 \|\Psi\|_{2,\Omega} \quad \text{if } \Psi \in H^2 \text{ and } \frac{\partial\Psi}{\partial n} = 0 \text{ on } \partial\Omega.$$

Then, $\|\Phi\| \le C(|||\Phi||| + M_1)h + M_2$, *where* C *depends on*

$$\sup\{\sigma(x,t) \mid (x,t) \in \cup E_0 \times I\}.$$

The proof is a simple modification of that of Lemma 6.1. Naturally, the function Ψ in the proof is defined by the boundary value problem $-\nabla\cdot(a(w)\nabla\Psi) + \Psi = \Phi$, $\frac{\partial\Psi}{\partial n} = 0$ on $\partial\Omega$. Next, one derives without difficulty the Neumann version of Theorem 6.2 in which the form B is replaced by $B(\cdot,\cdot) + (\cdot,\cdot)$ and γ_1 and γ_3 are defined with reference only to $\sigma|\cup E_0 \times J$. This accomplished, the proof of Theorem 6.3 adapts simply to the present case. Indeed, (6.11) is correct as given. Since

$$B(W;\eta,\chi) - A(w;w,\chi) + A(W;w,\chi) = A(W;Z,\chi) - A(w;Z,\chi) - (\eta,\chi),$$

instead of (6.12) we obtain that

$$(\xi_t,\chi) + B(W;\xi,\chi) + (\xi,\chi) = (\eta_t,\chi) + [A(W;Z,\chi) - A[w;Z,\chi)]$$
$$-(\zeta,\chi) + (b(W)-b(w),\nabla\chi) - \sum_{e\in E} \langle [b(\{W\}) - b(w)]\cdot n, [\chi] \rangle_e$$
$$-(f(W) - f(w),\chi).$$

The additional term (ξ,χ) on the left hand side allows us to use the weakened coercivity result of Theorem 8.1, while the term (ζ,χ) on the left hand side causes no additional trouble.

The proof of the second energy estimate is based on (5.2), which holds in the Neumann case. Thus one has little difficulty adapting the proof. The details, which are of the same sort considered above, will be omitted.

8.3. A Multipenalty Method.

If the mesh T is to be changed from time to time as the character of the solution w changes, it is necessary to interpolate the approximate solution from one mesh to another, which inevitably introduces interpolation errors. Let us sketch briefly and heuristically how interior penalties can be used to minimize such errors.

Let

$$J_1(\varphi,\Psi) = \sum_{e\in E_0} \ell_e \langle \sigma_1 \, [\tfrac{\partial\varphi}{\partial n}], \, [\tfrac{\partial\Psi}{\partial n}] \rangle_e,$$

where $\sigma_1 \in L^\infty(\cup E_0 \times I)$ is a non-negative function, and set $B_1=B+J_1$. Define $W^1: I \to M$ by the equations derived from (4.1) by replacing B with B_1. Then it is easy to show, as is indicated below, that

$J(W^1, W^1) + J_1(W^1, W^1)$ is bounded by the right hand side of (5.1). If σ and σ_1 are large on some edge $e \in E_0$, this estimate tells us that the discontinuities of W^1 and its normal derivative across e_0 are small and decrease with h. Suppose now that r=1, so that W^1 is a piecewise linear or piecewise bilinear function. Such a function is determined by its values and those of its normal derivative along a line segment. Hence, when σ and σ_1 are large on e_0, W^1 is essentially the same function on both sides of e. Therefore, if we interpolate W^1 into the mesh derived from T by removing the edge e, the error should be small.

Conversely, to introduce a new edge into the mesh, we can begin with the penalties at that edge large and reduce them to pass smoothly from the old mesh to the refinement.

For $r \geq 1$, the same heuristic considerations apply if we use the form $B_r = B + J_1 + J_2 + \ldots + J_r$, where

$$J_k(\varphi,\Psi) = \sum_{e \in E_0} \ell_e^{2k-1} < \sigma_k [\frac{\partial^k \varphi}{\partial n^k}], \ [\frac{\partial^k \Psi}{\partial n^k}] >_e .$$

Now, for all $\varphi \in H^k(T)$ and $e \in E_T$,

$$\| \frac{\partial^k \varphi}{\partial n^k} \|_{0,e}^2 \leq C(\ell_e^{-1} \ |\varphi|_{k,T}^2 + \ell_e |\varphi|_{k+1,T}^2).$$

It follows that, if $\varphi \in H^j(T)$ with $k+1 \leq j \leq r+1$, then

$$\ell_e^{2k-1} \| \frac{\partial^k}{\partial n^k} (\varphi - I\varphi) \|_{0,e}^2 \leq C(\ell_e^{2k-2} \ |\varphi - I\varphi|_{k,T}^2 + \ell_e^{2k} |\varphi - I\varphi|_{k+1,T}^2).$$

Thus, for $\varphi \in H^{r+1}(T)$,

$$(8.3) \quad \sum_{e \in E} \ell_e^{2k-1} \| [\frac{\partial^k}{\partial n^k} (\varphi - I\varphi)] \|_{0,e}^2 \leq \sum_{T \in T} h_T^{2[j(T)-1]} \|\varphi\|_{j(T),T}^2 ,$$

$$k+1 \leq j(T) \leq r+1.$$

In the multipenalty method we define W^k via the form B_k and let $\zeta_k = W^k - w$.

THEOREM 8.3. <u>There exists a constant depending on</u> γ_1 <u>and</u> $\sup\{|\sigma_i(x,t)| \,\big|\, x \in \cup\mathcal{E}_0,\ t \in I,\ 1\le i\le k\}$ <u>such that</u>

$$\|\zeta_k\|^2_{L^\infty(L^2)} + \int_I [|||\zeta_k|||^2 + J(\zeta_k,\zeta_k) + J_1(\zeta_k,\zeta_k) + \ldots + J_k(\zeta_k,\zeta_k)]\,dt$$

$$\le C \sum_{T\in\mathcal{T}} h_T^{2[j(T)-1]} \left(\|w_t\|^2_{L^1(H^{j(T)-1}(T))} + \|w\|^2_{L^2(H^{j(T)}(T))}\right.$$

$$\left. + h^2\|w_0\|^2_{j(T),T}\right) \text{ \underline{for} } k+1\le J(T)\le r+1.$$

The proof of Theorem 5.1, almost unchanged, gives Theorem 8.3. Since $J_i(w,\chi) = 0$ for all $i \ge 1$ and $\chi \in M$, the error equation (5.2) holds with B_k and ζ_k replacing B and ζ_k replacing B and ζ. Moreover, it is clear from Theorem 5.2 that

$$B_k(\rho;\varphi,\varphi) \ge \varepsilon\ |||\varphi|||^2 + \frac{1}{2} J_1(\varphi,\varphi) + J_1(\varphi,\varphi) + \ldots + J_k(\varphi,\varphi).$$

Thus, the claimed bounds reduce to bounds on μ, which hold by (8.3).

In a similar manner analogues of the $L^\infty(L^2)$ and second energy estimates can be shown for the multipenalty method.

We note that the form J_1 is exactly the one used by Douglas and Dupont [8] in creating their conforming interior penalty method mentioned in the introduction.

8.4. Elements of Varying Degree.

In the context of discontinuous finite elements it is an easy matter to let the degree r vary from element to element. This fact was exploited by Percell and Wheeler [11] in their local residual finite element procedure, and they proposed the strategy of using polynomials of low degree subordinate to a fine mesh in regions where the solution is relatively rough and higher degree polynomials subordinate to a coarse mesh in regions of smoothness of the solution.

There is no difficulty in adapting our analysis to allow for this possibility. Given an integer-valued function $T \in \mathcal{T} \mapsto r(T) \geq 1$, set

$$\mathcal{M} = \{\chi \in L^2(\Omega) \mid \chi|_T \in \mathcal{M}_{r(T)}(T), \quad T \in \mathcal{T}\}.$$

The usual range $2 \leq j(T) \leq r+1$ should then be replaced with $2 \leq j(T) \leq r(T)+1$. All the results previously stated remain valid. Note however that with r variable $\mathcal{M}$ will not generally approximate smoothly even if the mesh is edge-to-edge.

8.5. The Interior Penalty Method in Three Dimensions.

We have thus far restricted our attention to two-dimensional domains because of the greater complexity of the geometry of three dimensions. In particular, general tetrahedral meshes are of a complexity sufficient to make their implementation impractical in most cases. This is certainly not true of rectangular meshes in $\mathbb{R}^3$, which are little more complicated than their two-dimensional counterparts. Because the interior penalty method does not require an edge-to-edge mesh (in three dimensions better termed a "face-to-face mesh"), it offers the possibility of using a refined rectangular mesh only in that part of the domain where it is needed, a consideration of particular importance in three dimensional problems. Therefore, we feel it is worth observing that the previous analysis can easily be applied on domains of three (or n) dimensions.

Let $\mathcal{T}$ be a rectangular mesh of a domain $\Omega \subset \mathbb{R}^3$. Thus each element $T \in \mathcal{T}$ is a rectangular parallelpiped of the form $T = I_1 \times I_2 \times I_3$ with I_j a closed interval. In practice some allowance for boundary elements must also be made. $\mathcal{T}$ is again subject to a shape constraint by assuming that each $T \in \mathcal{T}$ is Lipschitz to a closed ball in $\mathbb{R}^3$ with the Lipschitz constant of the homeomorphism and its inverse bounded by a fixed constant K. Equivalently, we assume the existence of a fixed constant bounding the ratios of the faces of $T \in \mathcal{T}$.

The appropriate definitions of $\mathcal{E}_0$ and $\mathcal{E}_\partial$ are:

$$\mathcal{E}_0 = \{T_1 \cap T_2 \mid T_1, T_2 \in \mathcal{T} \text{ are distinct, } T_1 \cap T_2 \text{ is not contained in a line}\},$$

$$\mathcal{E}_\partial = \{T_1 \cap \partial\Omega \mid T_1 \cap \partial\Omega \text{ is not contained in a line}\}.$$

For $e \in \mathcal{E} = \mathcal{E}_0 \cup \mathcal{E}_\partial$ let $\ell_e = \sqrt{a_e}$, the square root of the area of e. As before, $\mathcal{T}$ is graded with a grade constant K' if $\ell_e \geq K h_T$ for each $e \in \mathcal{E}$ and $T \in \mathcal{T}$ such that $e \subset T$. Thus, as in the two-dimensional case, ℓ_e is taken to be representative of the diameters of the nearby finite elements. Note that the trace inequalities (2.4) and (2.5) remain valid.

The analogue of Lemma 2.3 is the assertion that

$$\sum \{\ell_e |\ e \in E,\ e \cap L \neq \emptyset\} \leq C\ (K, K', \Omega)$$

for each line L parallel to the x-axis in $\mathbb{R}^3$. Its demonstration, as in the rectangular case in $\mathbb{R}^2$, is trivial.

Almost without exception the remainder of the definitions and assertions adapt to several dimensions without difficulty. One result depending on the dimension is the inverse inequality relating $W^1_\infty(T)$ and $H^1(T)$ norms. In n dimensions.

$$\|\varphi\|_{W^1_\infty(T)} \leq Ch^{-n/2}\ \|\varphi\|_{H^1(T)}$$

for all polynomials φ on T, where C depends on the shape of T and the degree of φ. This affects the proof of Theorem 6.5 and the regularity required on w for that result. Similarly, the one-dimensional inverse inequality used in that proof must be replaced by the appropriate two-dimensional inequality.

CHAPTER IX

THE PENALTY FUNCTION

In subsection 8.3 we suggested an application of the interior penalty method to mesh refinement for which it is clearly valuable to be able to choose the penalty functions with some degree of flexibility. In addition, one of the initial motivations of this study was the possibility of using interior penalties to adjust the smoothness of the approximation to the behavior of the solution. For these reasons we have avoided placing undue restrictions on the penalty function σ, even when this would have simplified the analysis. Let us recall what restrictions have been made.

These have been formulated in terms of the boundedness of the quantities γ_i, listed at the end of subsection 2.4. First, we have assumed throughout that γ_0, a lower bound for σ, is sufficiently large. This is necessary for the coercivity result of Theorem 5.2 (or even for a Gårding inequality as in Theorem 8.1) and is entirely to be expected. In case M approximates smoothly the only additional assumption necessary is that γ_2, the supremum on the boundary of the penalty function, remain bounded. This is a very mild restriction since our main interest is in adjusting the interior penalties, and it is not unreasonable to fix $\sigma|_{\partial\Omega\times I}$ at some sufficiently large constant value. However, in the general case our estimates also depended on γ_1, the least upper bound for

σ on all edge-segments. This limitation is not unexpected, for -- reasoning heuristically-- if the constant C in (5.1) were to remain bounded as we let σ tend to infinity, then $W = W_\sigma$ would tend to a continuous optimal order approximation of w in the subspace M. But if M does not approximate smoothly, then there exist functions w for which no continuous optimal order approximation can be found in M.

This distinction of cases is reflected in the $L^\infty(L^2)$ and second energy estimates as well. These require bounds on σ and $|\sigma_t|$ (γ_1 and γ_3, respectively) in general but only on $\sigma|_{\partial\Omega}$ and a growth condition in case M approximates smoothly. Note that in the latter case σ may be arbitrarily large on E_0, so the growth condition, that $\gamma_4 = \sup|\sigma^{-1}\sigma_t|$ be bounded, is much weaker thatn the assumption that $|\sigma_t|$ be bounded.

Since a change in σ causes a change in W, it is reasonable that the estimate of $\|\zeta_t\|_{L^2(L^2)}$ provided by Theorem 7.3 depends on σ_t. In Theorem 6.3, however, this dependence was introduced by the method of proof which bounded $\|\zeta\|_{L^\infty(L^2)}$ in terms of the time derivative of the error in the elliptic projection. The author does not know whether another proof might be found which avoids this difficulty.

In Theorem 7.3 it was also assumed that $\gamma_5 = \sup|\sigma(x,0)|$ (or γ_6 if M approximates smoothly and W(0) is continuous) be bounded. This is easily seen to be necessary.

Finally, let us note that σ need not be furnished as an explicit function of x and t. For example, in the favorable case of smooth approximation the basic energy estimate remains valid if $\sigma|_{\cup E_0 \times I}$ is <u>any</u> function which is bounded below by γ_0. In particular, σ can depend on the approximate solution at an earlier time. The very interesting question of effective choice of the penalty function is presently under investigation.

APPENDIX

We now give the deferred proof of the assertion of Lemma 2.3 that

$$\text{(A.1)} \qquad \sum \{\ell_e \mid e \in \mathcal{E},\ e \cap L \neq \emptyset\} \leq C(K, K', \Omega)$$

for each line L in $\mathbb{R}^2$. Here Ω is a polygonal domain, $\mathcal{E}$ is the set of edge-segments of a mesh $\mathcal{T}$ on Ω which consists entirely of triangles or rectangles, K is a shape constant for $\mathcal{T}$, and K' is a grade constant. Actually, in the rectangular case, we restricted L to be a line parallel to the x-axis since this sufficed for our purposes and rendered the proof trivial. However, the full claim in the rectangular case can be derived from that in the triangular by adding a single diagonal to each rectangle, a construction which does not decrease the sum in (9.1), leaves K' unchanged, and at most doubles K.

To prove the lemma in case $\mathcal{T}$ is a triangulation (i. e., a mesh consisting of triangles), we shall use as a measure of the shape constraint on $\mathcal{T}$ an <u>angle bound</u> for the triangulation; that is, a positive lower bound on all angles of triangles in $\mathcal{T}$ (see Remark 2.2). We first prove the assertion for edge-to-edge triangulations.

LEMMA. Let $\mathcal{E}$ be the set of edges of an edge-to-edge triangulation with angle bound θ. Then,

$$\sum \{\ell_e \mid e \in \mathcal{E},\ e \cap L \neq \emptyset\} \leq C(\theta,\Omega)$$

for all lines L in $\mathbb{R}^2$.

Proof. First we claim that we can assume without loss of generality that L contains no vertex of the triangulation. For, if L^+ and L^- are sufficiently near parallel translations of L in opposite directions, then neither contains a vertex, and every edge intersected by L is also intersected by either L^+ or L^- with the exception of any edges contained in L. Therefore,

$$\sum \{\ell_e \mid e \cap L \neq \emptyset\} \leq \sum\{\ell_e \mid e \cap L^+ \neq \emptyset\} + \sum\{\ell_e \mid e \cap L^- \neq \emptyset\} + \mathrm{diam}(\Omega),$$

and so it suffices to bound the latter two sums.

Let N be the least number of convex subsets of $\overline{\Omega}$ which cover $\overline{\Omega}$. Since $\mathcal{T}$ is a finite convex cover, $N < \infty$. The number of connected components of $L \cap \overline{\Omega}$ is at most N, so it suffices to prove the existence of $C = C(\theta,\Omega)$ such that

$$\sum \{\ell_e \mid e \cap I \neq \emptyset\} \leq C$$

for each connected component I of $L \cap \overline{\Omega}$.

The following notations are illustrated in Figure 3. Let $T_1, T_2, \ldots, T_m$ be the triangles intersecting I taken in order (beginning at either end of I). Let $e_0, e_1, \ldots, e_m$ be the edges intersecting I, again ordered, so $e_{i-1} \cup e_i \subset T_i$, $i = 1, \ldots, m$. Let H^+ and H^- be the two components of $\mathbb{R}^2 \setminus L$, H^+ being the half plane which contains only one vertex of T_1.

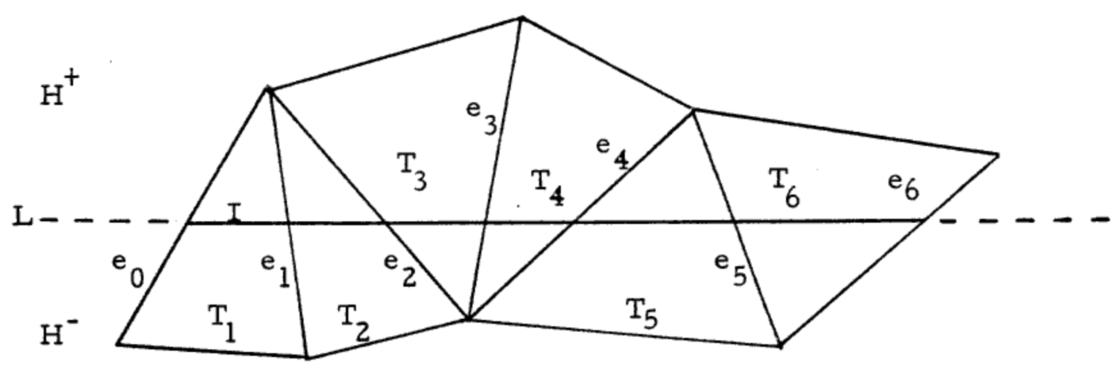

Figure 3. $m = 6$, $k = 3$, $i(1) = 2$, $i(2) = 4$, $i(3) = 5$.

We shall say that the triangle T_i points up if T_i has a single vertex in H^+; otherwise, T_i points down. By construction, T_1 points up. For certain integers i, $1 \leq i \leq m-1$, T_i and T_{i+1} point differently. Let $i(1) < i(2) < \cdots < i(k)$ be all such integers. Also, set $i(0) = 0$, $i(k+1) = m$.

Our next claim is that there exists a constant M_1 depending only on θ such that $i(j+1) - i(j) \leq M_1$ for $j = 0, 1, \ldots, k$. Indeed, the triangles $T_{i(j)+1}, T_{i(j)+2}, \ldots, T_{i(j+1)}$ all point in the same direction, say up, and all share a common vertex, namely the unique vertex of each in H^+.

Now, the sum of the angles of these triangles at this vertex is the angle formed by $e_{i(j)}$ and $e_{i(j+1)}$. Since both these edges meet I, this sum is less than π. Thus, $(i_{(j+1)} - i_{(j)})\theta < \pi$, and we can take $M_1 = \pi/\theta$. As a consequence, there exists a constant M_2 depending only on θ such that

$$\text{(A.2)} \qquad \max\{\ell_{e_i} \mid i(j) \le i \le i(j+1)\} \le M_2 \min\{\ell_{e_i} \mid i(j) \le i \le i(j+1)\}, \quad j = 0,1,\dots,k.$$

Now, if $k = 0$, then $m = i(1) \le M_1 + i(0) = M_1$, so that

$$\sum_{i=0}^{m} \ell_{e_i} \le (M_1 + 1)\operatorname{diam}(\Omega) ,$$

and the lemma is proven. Hence, we assume $k > 0$.

Let $p_j = e_{i(j)} \cap L$, $j = 0,1,\dots,k+1$, and set $v_j = e_{i(j-1)} \cap e_{i(j)}$, $j = 1,2,\dots,k+1$ (see Figure 4).

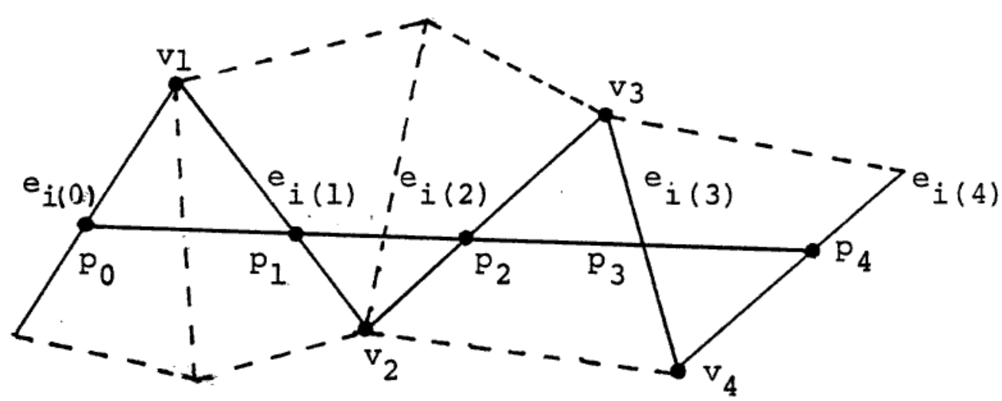

Figure 4. The points p_i and v_i.

Now the angle formed by $e_{i(0)}$ and $e_{i(1)}$ measures at least θ, so $|p_1 - p_0| \ge |v_1 - p_1| \sin\theta$. Similarly, $|p_2 - p_1| \ge |p_1 - v_2| \sin\theta$.

Adding, we get $|p_2 - p_0| \geq \ell_{e_{i(1)}} \sin\theta$. By (A. 2),

$$\sum_{i=i(0)}^{i(2)} \ell_{e_i} \leq (2M_1 + 1)M_2 |p_2 - p_0| / \sin\theta .$$

Similarly,

$$\sum_{i=i(j+1)}^{i(j+1)} \ell_{e_i} \leq M_3 |p_{j+1} - p_{j-1}| \quad , \qquad j = 1, 2, \ldots, k ,$$

where $M_3 = (2M_1 + 1)M_2/\sin\theta$. Summing, we find that

$$\sum_{i=0}^{m} \ell_{e_i} \leq \sum_{j=1}^{k} \sum_{i=i(j-1)}^{i(j+1)} \ell_{e_i} \leq M_3 \sum_{j=1}^{k} |p_{j+1} - p_{j-1}|$$

$$\leq 2M_3 \sum_{j=0}^{k} |p_{j+1} - p_j| \leq 2M_3 \mathrm{diam}(\Omega) .$$

This completes the proof of the lemma.

To prove Lemma 2.3 for general triangulations, we show that an arbitrary triangulation can be refined to an edge-to-edge triangulation without loss of control of the grade constant or angle bound. Given a triangulation $\mathcal{U}$ of Ω, let $\mathcal{E}(\mathcal{U})$ be its set of edge-segments. A second triangulation $\mathcal{V}$ of Ω is a refinement of $\mathcal{U}$ if $\mathcal{E}(\mathcal{U}) \subseteq \mathcal{E}(\mathcal{V})$.

LEMMA. A triangulation $\mathcal{T}$ with grade constant K and angle bound θ has an edge-to-edge refinement $\mathcal{U}$ such that K is a grade constant for $\mathcal{U}$ and $\arcsin(K^{2-k} \sin\theta)$ is an angle bound for $\mathcal{U}$.

<u>Proof.</u> If $\mathcal{T}$ is not already edge-to-edge, let $T \in \mathcal{T}$ be a triangle and E an edge of T such that some vertex of the triangulation lies interior to E. Let $v_0, v_1, \ldots, v_n$, $n > 1$, be all the vertices on E in order so that v_0 and v_n are the endpoints of E. Denote by e_i the edge-segment from v_{i-1} to v_i. Then

$$\operatorname{diam}(T) \geq \sum_{i=1}^{n} \ell_{e_i} \geq \sum_{i=1}^{n} \operatorname{diam}(T)/K \;;$$

so, $n \leq K$.

Let v be the vertex of T opposite E and consider the refinement $\mathcal{T}_1$ of $\mathcal{T}$ obtained by dividing T along the line segment from v to v_1 (see Figure 5). Note that K is a grade constant for $\mathcal{T}_1$, since the new edge has length

$$|v - v_1| \geq \min(|v - v_0|, |v - v_n|) \geq \operatorname{diam}(T)/K.$$

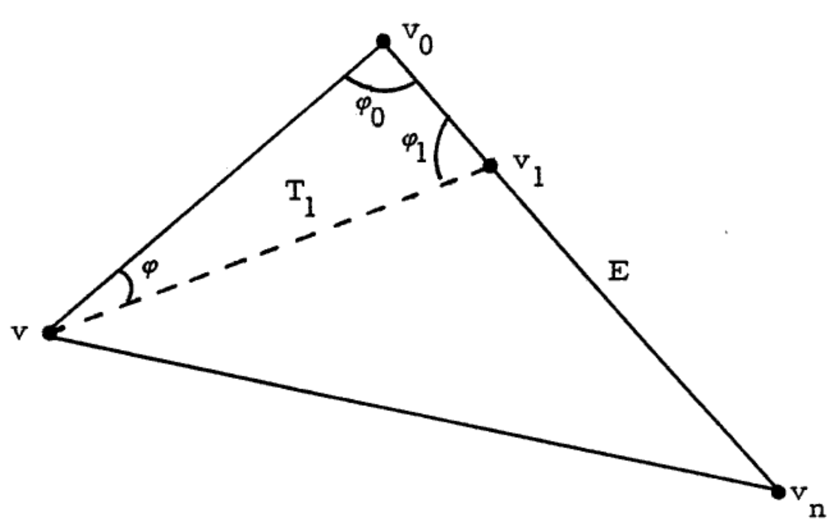

Figure 5. Division of the triangle T.

Moreover, we shall show that $\arcsin(K^{-1} \sin\theta)$ is an angle bound for $\mathcal{T}_1$. Consider the triangle T_1 with vertices v, v_0, and v_1. We must show that $K^{-1}\sin\theta$ is a lower bound for the sines of the angles of T_1. Let φ, φ_0, and φ_1 be the angles at v, v_0, and v_1 respectively. By hypothesis $\sin\varphi_0 \geq \sin\theta$. Using the law of sines we obtain that

$$\sin\varphi_1 = \frac{|v - v_0|}{|v - v_1|} \sin\varphi_0 \geq K^{-1} \sin\theta,$$

$$\sin\varphi = \frac{|v_0 - v_1|}{|v - v_1|} \sin\varphi_0 \geq K^{-1} \sin\theta,$$

as claimed.

We apply the same arguments to the sequence of refinements $\mathcal{T}_2, \mathcal{T}_3, \ldots, \mathcal{T}_{n-1}$, where $\mathcal{T}_i$ is obtained from $\mathcal{T}_{i-1}$ by joining v to v_i. It follows that K is a grade constant for each $\mathcal{T}_i$ while $\arcsin(K^{-i}\sin\theta)$ is an angle bound. In particular, $\arcsin(K^{1-K}\sin\theta)$ is an angle bound for $\mathcal{T}_{n-1}$.

It is also possible that there lie vertices of the original triangulation on other edges of the triangle T. Consider the refinement to $\mathcal{T}_{n-1}$ obtained by joining these vertices to the opposite vertex of the subtriangle of T in which they lie (e. g., vertices on the edge from v to v_0 are joined to v_1). The same argument shows that K and $\arcsin(K^{2-2k}\sin\theta)$, respectively, are grade constant and angle bound for this refined partition.

Finally, we note that the same construction may be applied to all the triangles of T, resulting in the desired edge-to-edge triangulation $\mathcal{U}$.

REFERENCES

[1] S. AGMON, Lectures on Elliptic Boundary Value Problems, Van Nostrand, Princeton, 1965.

[2] I. BABUŠKA, The finite element method with penalty, Math. Comp., 27 (1973), pp. 221-228.

[3] I. BABUŠKA AND M. ZLÁMAL, Nonconforming elements in the finite element method with penalty, SIAM J. Numer. Anal., 10 (1973), pp. 863-875.

[4] G. BAKER, Finite element methods for elliptic equations using nonconforming elements, Math. Comp. 31 (1977), pp. 45-59.

[5] P. CIARLET, The Finite Element Method for Elliptic Problems, North Holland, Amsterdam, 1978.

[6] A. DOLD, Lectures on Algebraic Topology, Springer-Verlag, Berlin, 1972.

[7] J. DOUGLAS, JR., The numerical simulation of miscible displacement in porous media, Proceedings of the Second International Conference on Computational Methods in Nonlinear Mechanics, U. of Texas, Austin, March 26-29, 1979.

[8] J. DOUGLAS, JR. AND T. DUPONT, Interior penalty procedures for elliptic and parabolic Galerkin methods, Lecture Notes in Physics, No. 58, Springer-Verlag, Berlin, 1976.

[9] J. KADLEC, On the regularity of the solution of the Poisson problem on a domain with boundary locally similar to the boundary of a convex set, Czech. Math. J., 14 (1964), pp.386-393.

[10] J.A. NITSCHE, Über ein Variationsprinzip zur Lösung von Dirichlet-Problemen bei Verwendung von Teilräumen, die keinen Randbedingungen unterworfen sind, Abh. Math. Sem. Univ. Hamburg, 36 (1971), pp. 9-15.

[11] P. PERCELL AND M. WHEELER, A local residual finite element procedure for elliptic equations, SIAM J. Numer. Anal., 15 (1978), pp. 705-714.

[12] M. WHEELER, A priori L_2 error estimates for Galerkin approximations to parabolic partial differential equations, SIAM J. Numer. Anal., 10 (1973), pp. 723-759.

[13] M. WHEELER, An elliptic collocation-finite element method with interior penalties,SIAM J. Numer. Anal., 15 (1978), pp. 152-161.

[14] O. ZIENKIEWICZ, Constrained variational principles and penalty function methods in finite element analysis, Conference on the Numerical Solution of Differential, Equations, Lecture Notes in Mathematics, No. 363, Springer, New York, 1973.