

## ON THE STABILITY AND CONVERGENCE OF HIGHER-ORDER MIXED FINITE ELEMENT METHODS FOR SECOND-ORDER ELLIPTIC PROBLEMS

MANIL SURI

**ABSTRACT.** We investigate the use of higher-order mixed methods for second-order elliptic problems by establishing refined stability and convergence estimates which take into account both the mesh size  $h$  and polynomial degree  $p$ . Our estimates yield asymptotic convergence rates for the  $p$ - and  $h-p$ -versions of the finite element method. They also describe more accurately than previously proved estimates the increased rate of convergence expected when the  $h$ -version is used with higher-order polynomials. For our analysis, we choose the Raviart-Thomas and the Brezzi-Douglas-Marini elements and establish optimal rates of convergence in both  $h$  and  $p$  (up to arbitrary  $\varepsilon > 0$ ).

### 1. INTRODUCTION

There have been several variational mixed formulations proposed for the solution of second-order elliptic problems like the Poisson equation. One such formulation involves writing the equation as a first-order system with both the displacement and velocity as unknowns. The Raviart-Thomas (RT) elements introduced in [14] provide a finite element discretization for this mixed variational principle and have been defined for arbitrary polynomial degree  $p$ . These elements, which are particularly useful when the velocity is the main physical quantity of interest, have received much attention in the literature (see, e.g., [12] and the references contained therein). All analysis carried out so far in connection with these elements concentrates on the  $h$ -version of the finite element method, where a fixed low degree  $p$  of elements is used (usually  $p = 1$  or  $2$ ) and accuracy is achieved by decreasing the mesh size  $h$ . Another class of elements for the same problem, the Brezzi-Douglas-Marini (BDM) elements (employing fewer degrees of freedom), was introduced in [9]. Like the RT elements, these, too, have been analyzed in the context of keeping  $p$  fixed and decreasing  $h$ . The error estimates that follow from such analysis usually yield a rate of convergence for the relative error bounded by a term of the form  $Ch^\gamma$ , where  $C$  is a constant independent of  $h$  but not  $p$ , and  $\gamma$  depends upon  $p$  and the smoothness of the solution.

---

Received October 11, 1988.

1980 *Mathematics Subject Classification* (1985 Revision). Primary 65N15, 65N30.

Research partially supported by the Air Force Office of Scientific Research, Air Force Systems Command, USAF, under Grant No. AFOSR-85-0322.

This work was completed while the author was a visitor at Institut National de Recherche en Informatique et en Automatique (INRIA), Domaine de Voluceau, Rocquencourt, France.

In recent times, there has been a large amount of interest shown in the use of the  $h$ -version with higher-order elements ( $p \geq 3$ ) because of the possible advantages of such elements over lower-order elements. For example, in [4], several methods have been tested for the rhombic (Kirchhoff) plate problem, and one of the conclusions reached is that higher-order elements are more efficient and more robust than lower-order elements. In [15], it was shown that in the elasticity problem the locking effect (for  $\nu \approx 0.5$ ) is completely eliminated when  $p \geq 4$ . Other advantages of higher-order elements have been discussed in [2].

Usual estimates of the form  $Ch^\gamma$  do not fully reflect the increase in order of convergence that may be expected when higher-order elements are used. This is because when  $p$  is increased, in addition to the exponent of  $h$  being increased, the constant  $C$ , which depends on  $p$ , also decreases. Consequently, more carefully derived estimates are needed, with the exact dependence of  $C$  on  $p$  being investigated.

The use of higher-order elements and the dependence of  $C$  on  $p$  are also important in the context of the  $p$ - and  $h$ - $p$  versions of the finite element method. In the  $p$ -version, a fixed mesh with constant  $h$  is used and accuracy is increased solely by increasing  $p$ . In the  $h$ - $p$  version, both  $h$  and  $p$  are changed. Basic approximation results for these methods first appeared in 1981 (in [8] and [3], respectively). Since then, they have become quite popular, owing to much higher rates of convergence than that possible with the  $h$ -version. These methods have been implemented for two-dimensional problems in the industrial code PROBE (Noetic Technologies, St. Louis). A survey of their theoretical and computational properties may be found in [1].

From the above discussion, it is clear that several finite element methods that have been analyzed in the context of the  $h$ -version (with estimates of the form  $Ch^\gamma$ ) would profit from further analysis, determining exactly how this behavior changes when  $p$  is increased. In this paper, we are interested in carrying out this analysis for some mixed methods, for which convergence depends upon two factors—the *stability* of the subspaces used and their *approximation* properties. Our goal is to investigate the rectangular RT and BDM elements and specifically answer the following two questions.

First, we determine how the stability constants for these spaces behave when  $p$  is increased. This is necessary to find out whether the  $p$ - and the  $h$ - $p$  versions would be stable if these methods are used.

Second, we establish rates of convergence for these methods which are uniform in both  $h$  and  $p$  (with the constant  $C$  being independent of both  $h$  and  $p$ ). This gives a more complete picture for the convergence of the  $h$ -version with high  $p$  and also establishes rates of convergence for the  $p$ - and  $h$ - $p$  versions.

We mention another reference [11] where the  $p$ -version of a mixed method (for Stokes' flow) has been analyzed. The problem considered there is Stokes' flow, whereas here we consider the Poisson equation. In that paper, it was found that the methods proposed had stability constants which, in general, behaved like  $p^{-\alpha}$  as  $p$  increased (with  $1 \leq \alpha \leq \frac{3}{2}$  for a family of elements analyzed in detail). Consequently, the error estimates that follow for the pressure are

nonoptimal in  $p$ . In contrast, we show that for the RT elements, the stability constant is independent of  $p$  (as well as  $h$ ), while for BDM, the dependence is not worse than  $p^{-\varepsilon}$ ,  $\varepsilon$  arbitrarily small. In §4 we show how this leads to optimal error estimates in both  $h$  and  $p$  (up to arbitrary  $\varepsilon > 0$  for  $p$ ) for both the velocity and the displacement.

## 2. PRELIMINARY RESULTS

Let  $\Omega$  be a bounded convex polygonal domain,  $\Omega \subseteq \mathbb{R}^2$ , with boundary  $\Gamma$ . We consider the model elliptic second-order problem,

$$(2.1) \quad -\Delta u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \Gamma.$$

To formulate our mixed method, we introduce the gradient of  $u$  as a new variable  $\sigma$  to obtain

$$(2.2) \quad -\operatorname{div} \sigma = f, \quad \sigma = \operatorname{grad} u \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \Gamma.$$

An equivalent variational formulation of (2.2) is then obtained by defining the spaces

$$V = L_2(\Omega), \quad S = H(\operatorname{div}, \Omega) = \{\tau \in (L_2(\Omega))^2; \operatorname{div} \tau \in L_2(\Omega)\}$$

and finding  $(u, \sigma) \in V \times S$  satisfying

$$(2.3) \quad (\sigma, \tau)_\Omega + (u, \operatorname{div} \tau)_\Omega = 0 \quad \forall \tau \in S,$$

$$(2.4) \quad (\operatorname{div} \sigma, v)_\Omega + (f, v)_\Omega = 0 \quad \forall v \in V,$$

where  $(\cdot, \cdot)_\Omega$  denotes the usual  $(L_2(\Omega))^n$  ( $n = 1, 2$ ) inner products. The boundary condition is built into equation (2.3). We will use  $\|\cdot\|_V$  and  $\|\cdot\|_S$  to denote the  $L_2(\Omega)$  and  $H(\operatorname{div}, \Omega)$  norms, respectively. Moreover,  $|\cdot|_{r,\Omega}$  and  $\|\cdot\|_{r,\Omega}$  will be used to denote the seminorm and norm on  $(H^r(\Omega))^n$ ,  $n = 1, 2$  for any region  $\Omega$ .

(2.3)–(2.4) may be discretized by choosing a pair of finite-dimensional subspaces  $V_N \subset V$ ,  $S_N \subset S$  and finding  $(u_N, \sigma_N) \in V_N \times S_N$  such that

$$(2.5) \quad (\sigma_N, \tau_N)_\Omega + (u_N, \operatorname{div} \tau_N)_\Omega = 0 \quad \forall \tau_N \in S_N,$$

$$(2.6) \quad (\operatorname{div} \sigma_N, v_N)_\Omega + (f, v_N)_\Omega = 0 \quad \forall v_N \in V_N.$$

(2.5)–(2.6) will only have a solution when certain compatibility conditions, described later, between  $V_N$  and  $S_N$  are satisfied.

We assume that there is a family  $\{V_N \times S_N\}$  of such spaces, with  $N$  being a parameter related to the dimensions of  $V_N$ ,  $S_N$ . The finite element spaces to be considered consist of piecewise polynomial spaces defined on grids on  $\Omega$  with mesh size  $h$ .  $N$  will depend on both  $h$  and the polynomial degree  $p$  used, so that  $N = N(h_N, p_N)$ . In order to increase accuracy, one employs an *extension procedure*, by which pairs of spaces  $(V_N, S_N)$  with increasing dimension  $N$  are selected. In the usual extension procedure, the degree of polynomials is kept fixed while  $h_N$  is decreased. We will be interested in analyzing the combined effect of changing both  $h_N$  and  $p_N$ , either together or separately. We will require the following theorem (see [14]).

**Theorem 2.1.** *Let  $\{V_n, S_n\}$  be a family of spaces such that:*

(1) *For any  $\tau_n \in S_n$ ,*

$$(2.7) \quad (v_n, \operatorname{div} \tau_n)_\Omega = 0 \quad \forall v_n \in V_n \Rightarrow \operatorname{div} \tau_n = 0.$$

(2) *There exists  $\alpha = \alpha(N) > 0$  such that for any  $v_n \in V_n$ ,*

$$(2.8) \quad \sup_{\tau_n \in S_n} \frac{(v_n, \operatorname{div} \tau_n)_\Omega}{\|\tau_n\|_S} \geq \alpha(N) \|v_n\|_V.$$

*Then the problem (2.5)–(2.6) has a unique solution, and there exists a constant  $C > 0$  independent of  $N$  such that*

$$(2.9) \quad \|\sigma_n\|_S + \|u_n\|_V \leq \frac{C}{\alpha(N)} \{\|\sigma\|_S + \|u\|_V\},$$

$$(2.10) \quad \|\sigma - \sigma_n\|_S + \|u - u_n\|_V \leq \frac{C}{\alpha(N)} \left\{ \inf_{\tau \in S_n} \|\sigma - \tau\|_S + \inf_{v \in V_n} \|u - v\|_V \right\}.$$

Let us now define the RT spaces (denoted by  $\{V_n^1, S_n^1\}$ ) and the BDM spaces ( $\{V_n^2, S_n^2\}$ ). Like in [11], our analysis will be restricted to the case of *parallel-ogram* elements. Let  $Q$  denote the standard square,  $[-1, 1] \times [-1, 1]$ . For  $\Omega \subset \mathbb{R}^1$  or  $\mathbb{R}^2$ ,  $P_k(\Omega)$  will denote the set of all polynomials on  $\Omega$  of total degree  $\leq k$ . When  $\Omega \equiv Q$ , we will use  $P_k$  to denote  $P_k(Q)$ . By  $P_{l,m}$ , we will denote the set of polynomials on  $Q$  with degree  $\leq l$  in  $\xi$  and degree  $\leq m$  in  $\eta$ . Then we define

$$(2.11) \quad V_k^1(Q) = P_{k,k},$$

$$(2.12) \quad S_k^1(Q) = P_{k+1,k} \times P_{k,k+1},$$

$$(2.13) \quad V_k^2(Q) = P_{k-1},$$

$$(2.14) \quad S_k^2(Q) = (P_k \times P_k) \oplus \operatorname{span}\{(\xi^{k+1}, -\xi^k \eta)^T, (\xi \eta^k, -\eta^{k+1})^T\}.$$

Note that  $P_k \times P_k \subset S_k^i(Q)$ ,  $i = 1, 2$ .

Now let  $\{T_N\}$  be a quasi-uniform family of meshes on  $\Omega$  consisting of parallelograms  $K$ .  $h_K, \rho_K$  will denote the diameters of  $K$  and of the largest circle that can be inscribed in  $K$ , respectively. Let  $h_N = \max_{K \in T_N} h_K$ . We assume there exist constants  $C_1, C_2$  independent of  $h_N$  such that for all  $K \in T_N$ , for all  $N$ ,

$$(2.15) \quad \frac{h_N}{h_K} \leq C_1, \quad \frac{h_K}{\rho_K} \leq C_2.$$

Further, we assume that each pair  $K_1, K_2 \in T_N$  has either an entire side or a vertex in common, or has empty intersection.

For  $K \in T_N$ , let  $\mathbf{F}_K$  be the affine invertible mapping such that  $K = \mathbf{F}_K(Q)$ ,

$$(2.16) \quad (x, y) = \mathbf{F}_K((\xi, \eta)) = B_K(\xi, \eta)^T + \mathbf{b}_K,$$

where  $B_k$  is a  $2 \times 2$  matrix. With any scalar function  $\tilde{v}$  defined on  $Q$  (or  $\partial Q$ ) we associate the function  $v$  defined on  $K$  (or  $\partial K$ ) by

$$(2.17) \quad v = \tilde{v} \circ \mathbf{F}_K^{-1} \quad (\tilde{v} = v \circ \mathbf{F}_K).$$

For vector-valued functions, the correspondence between  $\tilde{\tau}$  defined on  $Q$  and  $\tau$  defined on  $K$  is given by

$$(2.18) \quad \tau = \frac{1}{J_K} B_K \tilde{\tau} \circ \mathbf{F}_K^{-1} \quad (\tilde{\tau} = J_K B_K^{-1} \tau \circ \mathbf{F}_K),$$

where  $J_K = \det(B_K)$ . The one-to-one correspondences  $\tilde{v} \leftrightarrow v$  and  $\tilde{\tau} \leftrightarrow \tau$  will be understood in the sequel.

The following lemmas follow from Lemmas 2 and 3, respectively, of [14].

**Lemma 2.1.** *For any function  $\tilde{\tau} \in (H^1(Q))^2$ ,*

$$(2.19) \quad (\operatorname{div} \tilde{\tau}, \tilde{\phi})_Q = (\operatorname{div} \tau, \phi)_K \quad \forall \tilde{\phi} \in L_2(Q),$$

$$(2.20) \quad \int_{\partial Q} \tilde{\tau} \cdot \tilde{\nu} \tilde{\phi} d\tilde{s} = \int_{\partial K} \tau \cdot \nu \phi ds \quad \forall \tilde{\phi} \in L_2(\partial Q).$$

**Lemma 2.2.** *For any integer  $l \geq 0$ ,*

$$(2.21) \quad |\tilde{\tau}|_{l,Q} \leq C h_K^l |\tau|_{l,K},$$

$$(2.22) \quad |\tau|_{l,K} \leq C h_K^{-l} |\tilde{\tau}|_{l,Q},$$

where the constant  $C$  depends on  $l$  but is independent of  $\tau, h_K$ .

(We have used condition (2.15) in (2.21)–(2.22).)

With  $K$ , we now associate the spaces ( $i = 1, 2$ )

$$(2.23) \quad S_k^i(K) = \{\tau : K \rightarrow \mathbb{R}^2, \tilde{\tau} \in S_k^i(Q)\},$$

$$(2.24) \quad V_k^i(K) = \{v : K \rightarrow \mathbb{R}^1, \tilde{v} \in V_k^i(Q)\}.$$

Then, we set, for  $i = 1, 2$ ,

$$(2.25) \quad S_N^i = \{\tau \in S, \tau|_K \in S_{p_N}^i(K) \forall K \in T_N\} \subset S,$$

$$(2.26) \quad V_N^i = \{v \in V, v|_K \in V_{p_N}^i(K) \forall K \in T_N\} \subset V.$$

Note that the inclusion  $S_N^i \subset S$  is equivalent to the condition that the normal component of  $\tau$  along any  $\partial K$  must be continuous.

Since the spaces  $V_N^i$  consist of piecewise polynomials on regular quasi-uniform meshes, the following inverse inequality will be true:

**Lemma 2.3.** *There exists an  $\varepsilon_0 > 0$  such that for  $v_N \in V_N^i$  and  $0 \leq \varepsilon < \varepsilon_0$ ,*

$$\|v_N\|_{\varepsilon,\Omega} \leq C h_N^{-\varepsilon} p_N^{2\varepsilon} \|v_N\|_{0,\Omega},$$

where  $C$  is a constant independent of  $h_N, p_N$  and  $v_N$ .

*Proof.* The proof follows easily from the separate inverse inequalities in terms of  $h_N$  (see [10]) and in terms of  $p_N$  (see [8]).  $\square$

It is easy to see that the above spaces satisfy

$$(2.27) \quad \operatorname{div}(S_N^i) \subset V_n^i,$$

so that condition (2.7) of Theorem 2.1 is automatically satisfied. Moreover, it has been shown in [14, 9] that (2.8) is satisfied with  $\alpha(N) > 0$  independent of  $h_N$  (but depending possibly on  $p_N$ ). In order to get our desired convergence rate, we must now estimate  $\alpha(N)$  in terms of both  $h_N$  and  $p_N$  and also estimate the approximation properties of our spaces, to be used in (2.9)–(2.10). Our analysis will be facilitated by families of projections

$$\Pi_N^i : S \rightarrow S_N^i \quad \text{and} \quad P_N^i : V \rightarrow V_N^i$$

defined for  $i = 1$  in [14] and  $i = 2$  in [9] such that the following commutative property holds:

$$(2.28) \quad \operatorname{div} \circ \Pi_N^i = P_N^i \circ \operatorname{div}.$$

We now describe the above projections, which are constructed locally on each  $K \in T_N$ .  $P_N^i$  will simply be the  $L_2$  projection satisfying

$$(2.29) \quad (v - P_N^i v, w_N)_K = 0 \quad \forall w_N \in V_{p_N}^i(K), \quad K \in T_N.$$

The following theorem follows from the approximation theory of the  $h$ - $p$  version.

**Theorem 2.2.** *Let  $P_N^i : V \rightarrow V_N^i$  be defined piecewise over each  $K \in T_N$  by (2.29). Then for any  $v \in H^r(\Omega)$ ,  $r \geq 0$ ,*

$$(2.30) \quad \|v - P_N^i v\|_V \leq C h_N^{\mu_1} p_N^{-r} \|v\|_{r, \Omega},$$

where  $\mu_1 = \min(p_N + 1, r)$  and  $\mu_2 = \min(p_N, r)$  and where  $C$  is a constant independent of  $h_N$ ,  $p_N$  and  $v$ .

*Proof.* Since  $P_N^i$  is the  $L_2$  projection, we know that over each  $K$ ,

$$(2.31) \quad \|v - P_N^i v\|_{0, K} \leq C \inf_{w \in V_{p_N}^i(K)} \|v - w\|_{0, K} \leq C h_N^{\mu_1} p_N^{-r} \|v\|_{r, K}$$

by Lemma 4.5 of [5]. (2.30) follows by squaring and summing (2.31) over all  $K \in T_N$ .  $\square$

**Remark 2.1.** The powers  $\mu_1, \mu_2$  are different in view of the fact that the polynomials used to define  $V_{p_N}^2(Q)$  are of one degree less than those for  $V_{p_N}^1(Q)$ .

Now let  $\tau$  be a function in  $S$ . The projections  $\Pi_N^i \tau$  are defined locally over each  $K$  in terms of a projection  $\Pi_{p_N}^i \tilde{\tau}$  on the standard square  $Q$  (where  $\tilde{\tau}$  satisfies (2.18)). Let for  $k \geq 1$ ,

$$(2.32) \quad M_k^1 = P_{k-1, k} \times P_{k, k-1}, \quad M_k^2 = P_{k-2} \times P_{k-2},$$

where  $M_1^2$  is understood to be empty. Then  $\Pi_{p_N}^i \tau$  is defined by the conditions

$$(2.33) \quad ((\Pi_{p_N}^i \tilde{\tau} - \tilde{\tau}), \tilde{w})_Q = 0 \quad \text{for all } \tilde{w} \in M_{p_N}^i,$$

$$(2.34) \quad \int_l (\Pi_{p_N}^i \tilde{\tau} - \tilde{\tau}) \cdot \tilde{\nu} \tilde{v} \, ds = 0 \quad \text{for all } \tilde{v} \in P_{p_N}(l),$$

where (2.34) holds for any side  $l$  of  $Q$  and  $\tilde{\nu}$  is the outer normal to  $\partial Q$ . The unisolvence of (2.33), (2.34) has been established in [14, 9]. Note that (2.11), (2.13), (2.32) imply that  $\text{grad}(V_k^i(Q)) \subset M_k^i$  and also that  $\tilde{v} \in V_k^i(Q)$  implies  $\tilde{v}|_l \in P_k(l)$ . Hence, for any  $\tilde{v} \in V_{p_N}^i(Q)$ , we obtain by (2.33), (2.34) that

$$(2.35) \quad (\text{div}(\Pi_{p_N}^i \tilde{\tau} - \tilde{\tau}), \tilde{v})_Q = \int_{\partial Q} (\Pi_{p_N}^i \tilde{\tau} - \tilde{\tau}) \cdot \tilde{\nu} \tilde{v} \, ds - (\Pi_{p_N}^i \tilde{\tau} - \tilde{\tau}, \text{grad } \tilde{v})_Q = 0.$$

We now define  $\Pi_N^i \tau$  on  $\Omega$  such that

$$(2.36) \quad (\widetilde{\Pi_N^i \tau})|_K = \Pi_{p_N}^i \tilde{\tau}.$$

Then the following holds.

**Theorem 2.3.** *For  $\tau \in S$ , let  $\Pi_N^i \tau$  be defined by (2.33), (2.34), (2.36). Then  $\Pi_N^i \tau \in S_N^i$  and is uniquely defined. Moreover,*

$$(2.37) \quad \Pi_N^i \tau = \tau \quad \text{for all } \tau \in S_N^i,$$

$$(2.38) \quad (\text{div}(\Pi_N^i \tau - \tau), v)_\Omega = 0 \quad \text{for all } v \in V_N^i.$$

*Proof.* Using (2.35) together with (2.19), and summing over  $K \in T_N$ , gives (2.38). Moreover, by (2.20) and (2.34),  $\Pi_N^i \tau \cdot \nu$  is continuous for any  $l$  in the triangulation, so that  $\Pi_N^i \tau \in S_N^i$ . Finally, (2.37) follows from the unisolvence of (2.33)–(2.34).  $\square$

Note that (2.38) implies (2.28). In the next section, we derive error estimates for  $\|\Pi_N^i \tau - \tau\|_S$  that are uniform in both  $h_N$  and  $p_N$ .

### 3. ERROR ESTIMATES FOR THE PROJECTIONS $\Pi_N^i$

Let  $I = [-1, +1]$ . Then  $\{L_j(\xi)\}$ ,  $j = 0, 1, \dots$ , will denote the Legendre polynomials on  $I$  which are orthogonal in the following sense:

$$(3.1) \quad \int_{-1}^{+1} L_k(\xi) L_j(\xi) \, d\xi = \frac{2}{2j+1} \quad \text{if } j = k, \quad = 0 \text{ otherwise.}$$

For any  $\tau \in L_2(Q)$ , where  $Q$  is the standard square  $I \times I$ , we may expand  $\tau$  as

$$(3.2) \quad \tau = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} a_{ij} L_i(\xi) L_j(\eta).$$

Then we have, using the orthogonality properties of  $\{L_j(\xi)\}$  and their derivatives (see [7]),

$$(3.3) \quad \|\tau\|_{0,Q}^2 = \iint_Q \tau^2 \, d\xi \, d\eta = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \frac{4a_{ij}^2}{(2i+1)(2j+1)},$$

$$\begin{aligned}
\|\tau\|_{r,Q}^2 &\geq \sum_{0 \leq s+t \leq r} \iint_Q (1 - \xi^2)^s (1 - \eta^2)^t \left( \frac{\partial^{s+t} \tau}{\partial \xi^s \partial \eta^t} \right)^2 d\xi d\eta \\
(3.4) \quad &\geq C \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \frac{a_{ij}^2 (1 + i^2 + j^2)^r}{(2i+1)(2j+1)}.
\end{aligned}$$

**3.1. The Raviart-Thomas elements.** Let  $\tilde{\tau} = (\tau_1, \tau_2) \in H(\text{div}, Q)$ . Then, for the RT elements, the projection  $\Pi_k^1$  defined in (2.33)–(2.34) may be written as  $\Pi_k^1 \tilde{\tau} = (\tau_1^k, \tau_2^k) \in S_k^1(Q)$ , where

$$(3.5) \quad \iint_Q (\tau_1^k - \tau_1) \phi(\xi) \zeta(\eta) d\xi d\eta = 0, \quad \phi \in P_{k-1}(I), \zeta \in P_k(I),$$

$$(3.6) \quad \iint_Q (\tau_2^k - \tau_2) \phi(\xi) \zeta(\eta) d\xi d\eta = 0, \quad \phi \in P_k(I), \zeta \in P_{k-1}(I),$$

$$(3.7) \quad \int_{-1}^{+1} (\tau_1^k - \tau_1)(\pm 1, \eta) \zeta(\eta) d\eta = 0, \quad \zeta \in P_k(I),$$

$$(3.8) \quad \int_{-1}^{+1} (\tau_2^k - \tau_2)(\xi, \pm 1) \phi(\xi) d\xi = 0, \quad \phi \in P_k(I).$$

We are interested in estimating  $\|\tilde{\tau} - \Pi_k^1 \tilde{\tau}\|_{0,Q}$ . Since  $\tau_1 \in L_2(Q)$ , we may assume it has the expansion (3.2). Moreover, the polynomial  $\tau_1^k \in P_{k+1,k}$  may be expanded as

$$(3.9) \quad \tau_1^k = \sum_{j=0}^k \sum_{i=0}^{k+1} b_{ij} L_i(\xi) L_j(\eta).$$

Let us calculate the coefficients  $b_{ij}$ . We first use the fact that the Legendre polynomials form an orthogonal basis for  $P_{l,m}$  with respect to the  $L_2(Q)$  inner product. Taking  $\phi(\xi) = L_i(\xi)$ ,  $\zeta(\eta) = L_j(\eta)$  in (3.5) yields

$$(3.10) \quad b_{ij} = a_{ij}, \quad 0 \leq i \leq k-1, \quad 0 \leq j \leq k.$$

To calculate  $b_{kj}$  and  $b_{k+1,j}$ , we use the boundary conditions (3.7). First, on the side  $\xi = -1$ , taking  $\zeta(\eta) = L_l(\eta)$ ,  $0 \leq l \leq k$ , gives

$$\begin{aligned}
&\int_{-1}^{+1} \left( \sum_{j=0}^k \sum_{i=0}^{k+1} b_{ij} L_i(-1) L_j(\eta) \right) L_l(\eta) d\eta \\
&= \int_{-1}^{+1} \left( \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} a_{ij} L_i(-1) L_j(\eta) \right) L_l(\eta) d\eta.
\end{aligned}$$

Since  $L_i(-1) = (-1)^i$ , this yields (using (3.1) and (3.10)),

$$b_{kl} - b_{k+1,l} = \sum_{i=k}^{\infty} a_{il} (-1)^{i-k}, \quad 0 \leq l \leq k.$$



Similarly, the condition at  $\xi = +1$  gives

$$b_{kl} + b_{k+1,l} = \sum_{i=k}^{\infty} a_{il}, \quad 0 \leq l \leq k,$$

so that

$$(3.11) \quad b_{kl} = \sum_{i=0}^{\infty}{}'' a_{k+i,l}, \quad b_{k+1,l} = \sum_{i=0}^{\infty}{}' a_{k+i,l},$$

where  $\sum''$  stands for summation over even integers and  $\sum'$  over odd.

We will now use (3.9)–(3.11) to estimate  $\|\tau_1 - \tau_1^k\|_{0,Q}$ . By (3.3) we have

$$(3.12) \quad \begin{aligned} \|\tau_1 - \tau_1^k\|_{0,Q}^2 &= \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \frac{4(a_{ij} - b_{ij})^2}{(2i+1)(2j+1)} \quad (\text{where } b_{ij} = 0 \text{ for } j > k \text{ or } i > k+1) \\ &= 4 \left( \sum_{j=0}^k + \sum_{j=k+1}^{\infty} \right) \left( \sum_{i=0}^{k-1} + \sum_{i=k}^{k+1} + \sum_{i=k+2}^{\infty} \right) \frac{(a_{ij} - b_{ij})^2}{(2i+1)(2j+1)} \\ &= 4 \left\{ \sum_{j=0}^{\infty} \sum_{i=k}^{k+1} \frac{(a_{ij} - b_{ij})^2}{(2i+1)(2j+1)} \right. \\ &\quad \left. + \left( \sum_{j=k+1}^{\infty} \sum_{i=0}^{k-1} + \sum_{j=0}^k \sum_{i=k+2}^{\infty} + \sum_{j=k+1}^{\infty} \sum_{i=k+2}^{\infty} \right) \frac{a_{ij}^2}{(2i+1)(2j+1)} \right\}. \end{aligned}$$

Now for  $r \geq 0$ ,

$$(3.13) \quad \begin{aligned} &4 \sum_{j=k+1}^{\infty} \sum_{i=k+2}^{\infty} \frac{a_{ij}^2}{(2i+1)(2j+1)} \\ &\leq 4 \sum_{j=k+1}^{\infty} \sum_{i=k+2}^{\infty} \frac{a_{ij}^2}{(2i+1)(2j+1)} \frac{(1+i^2+j^2)^r}{(1+k^2+k^2)^r} \\ &\leq \frac{C}{k^{2r}} \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \frac{a_{ij}^2 (1+i^2+j^2)^r}{(2i+1)(2j+1)} \\ &\leq \frac{C}{k^{2r}} \|\tau_1\|_{r,Q}^2, \end{aligned}$$

using (3.4). Similarly,

$$(3.14) \quad \begin{aligned} 4 \sum_{j=0}^k \sum_{i=k+2}^{\infty} \frac{a_{ij}^2}{(2i+1)(2j+1)} &\leq 4 \sum_{j=0}^k \sum_{i=k+2}^{\infty} \frac{a_{ij}^2}{(2i+1)(2j+1)} \frac{(1+i^2+j^2)^r}{(1+i^2)^r} \\ &\leq \frac{C}{k^{2r}} \|\tau_1\|_{r,Q}^2. \end{aligned}$$

The term

$$4 \sum_{j=k+1}^{\infty} \sum_{i=0}^{k-1} \frac{a_{ij}^2}{(2i+1)(2j+1)}$$

is similarly bounded. We now bound the first term. Let  $i = k$ . Then, using (3.11), we have

$$A = 4 \sum_{j=0}^{\infty} \frac{(a_{kj} - b_{kj})^2}{(2k+1)(2j+1)} \leq 4 \sum_{j=0}^{\infty} \frac{(\sum_{i=2}^{\infty} ''a_{k+i,j})^2}{(2k+1)(2j+1)}.$$

Now for  $r_1 > 1/2$ ,

$$\begin{aligned} \left( \sum_{i=2}^{\infty} ''a_{k+i,j} \right)^2 &\leq \sum_{i=k+2}^{\infty} a_{ij}^2 (1+i^2+j^2)^{r_1} \sum_{i=k+2}^{\infty} (1+i^2+j^2)^{-r_1} \\ &\leq Ck^{-(2r_1-1)} \sum_{i=k+2}^{\infty} a_{ij}^2 (1+i^2+j^2)^{r_1} \\ &\leq Ck^{-(2r_1-1)} \sum_{i=k+2}^{\infty} \frac{a_{ij}^2 (1+i^2+j^2)^{r_1+1/2}}{(2i+1)}, \end{aligned}$$

so that

$$\begin{aligned} (3.15) \quad A &\leq \frac{4}{(2k+1)} \sum_{j=0}^{\infty} \frac{(\sum_{i=2}^{\infty} ''a_{k+i,j})^2}{(2j+1)} \\ &\leq Ck^{-2r_1} \sum_{i=0}^{\infty} \sum_{j=0}^{\infty} \frac{a_{ij}^2 (1+i^2+j^2)^{r_1+1/2}}{(2i+1)(2j+1)} \\ &\leq Ck^{-2r_1} \|\tau_1\|_{r_1+1/2,Q}^2 = Ck^{-2(r-1/2)} \|\tau_1\|_{r,Q}^2, \end{aligned}$$

provided  $r = r_1 + \frac{1}{2} > 1$ . (3.13)–(3.15) show that for  $r > 1$ ,

$$\|\tau_1 - \tau_1^k\|_{0,Q} \leq Ck^{-(r-1/2)} \|\tau_1\|_{r,Q}.$$

A similar argument may be used to bound  $\|\tau_2 - \tau_2^k\|_{0,Q}$ . We have therefore proven

**Lemma 3.1.** *Let  $\tilde{\tau} \in (H^r(Q))^2$ ,  $r > 1$ . Let  $\Pi_k^1 \tilde{\tau} = (\tau_1^k, \tau_2^k) \in S_k^1(Q)$  be defined by (3.5)–(3.8). Then*

$$(3.16) \quad \|\tilde{\tau} - \Pi_k^1 \tilde{\tau}\|_{0,Q} \leq Ck^{-(r-1/2)} \|\tilde{\tau}\|_{r,Q},$$

where  $C$  is a constant independent of  $k$  and  $\tilde{\tau}$  but depending on  $r$ .

In order to prove a corresponding estimate for  $\Pi_N^1$  (in both  $h_N$  and  $p_N$ ), we need the following lemma.

**Lemma 3.2.** *Let  $\tilde{\tau} \in (H^r(Q))^2$  and  $\tau \in (H^r(K))^2$ ,  $r \geq 0$ , be related by (2.18). Then*

$$(3.17) \quad \inf_{\tilde{\omega} \in P_k \times P_k} \|\tilde{\tau} - \tilde{\omega}\|_{r,Q} \leq Ch_K^{\min(k+1, r)} \|\tau\|_{r,K},$$

where  $C$  depends on  $r$  but is independent of  $h_K$ ,  $k$  and  $\tau$ .

*Proof.* The above lemma is simply a vector form of Lemma 4.4 proved in [5]. The proof follows identically, using the scaling result (2.21) and Theorem 3.1.2 of [10].  $\square$

We now prove our main estimates for  $\Pi_N^1$ .

**Theorem 3.1.** *Let  $\tau \in (H^r(\Omega))^2$ ,  $r > 1$ . Let  $\Pi_N^1 : S \rightarrow S_N^1$  be as defined in §2. Then*

$$(3.18) \quad \|\tau - \Pi_N^1 \tau\|_{0,\Omega} \leq C h_N^{\min(p_N+1, r)} p_N^{-(r-1/2)} \|\tau\|_{r,\Omega},$$

where  $C$  is a constant independent of  $h_N$ ,  $p_N$  and  $\tau$  but depends upon  $r$ .

Moreover, if  $\operatorname{div} \tau \in H^r(\Omega)$ , then

$$(3.19) \quad \|\tau - \Pi_N^1 \tau\|_S \leq C h_N^{\mu_1} p_N^{-\beta_1} (\|\tau\|_{r,\Omega} + \|\operatorname{div} \tau\|_{r,\Omega}),$$

where  $\mu_1 = \min(p_N + 1, r)$  and  $\beta_1 = r - \frac{1}{2}$ .

*Proof.* Let  $K \subset T_N$ . Then we have by (2.37), for any  $\tilde{\omega} \in P_k \times P_k$ ,

$$(3.20) \quad \begin{aligned} \|\tilde{\tau} - \Pi_{p_N}^1 \tilde{\tau}\|_{0,Q} &= \|(\tilde{\tau} - \tilde{\omega}) - \Pi_{p_N}^1(\tilde{\tau} - \tilde{\omega})\|_{0,Q} \\ &\leq C p_N^{-(r-1/2)} h_K^{\min(p_N+1, r)} \|\tau\|_{r,K}, \end{aligned}$$

where we have used Lemmas 3.1 and 3.2. Using (3.20) with (2.22) then gives

$$(3.21) \quad \|\tau - \Pi_N^1 \tau\|_{0,K} \leq C p_N^{-(r-1/2)} h_K^{\min(p_N+1, r)} \|\tau\|_{r,K}.$$

Squaring (3.21), summing over all  $K \in T_N$  and noting that  $h_K \leq h_N$  yields (3.18).

To obtain (3.19), we note that by (2.28),

$$(3.22) \quad \|\operatorname{div} \tau - \operatorname{div}(\Pi_N^1 \tau)\|_{0,\Omega} = \|\operatorname{div} \tau - P_N^1(\operatorname{div} \tau)\|_{0,\Omega}.$$

Using Theorem 2.2 gives

$$(3.23) \quad \|\operatorname{div} \tau - \operatorname{div}(\Pi_N^1 \tau)\|_{0,\Omega} \leq C h_N^{\mu_1} p_N^{-r} \|\operatorname{div} \tau\|_{r,\Omega}.$$

Combining (3.23) with (3.18) gives (3.19).  $\square$

**3.2. The Brezzi-Douglas-Marini elements.** Let  $\tilde{\tau} = (\tau_1, \tau_2) \in H(\operatorname{div}, Q)$  be given by

$$(3.24) \quad \tau_n = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} a_{ij}^n L_i(\xi) L_j(\eta), \quad n = 1, 2.$$

Then  $\Pi_k^2 \tilde{\tau} = (\tau_1^k, \tau_2^k) \in S_k^2(Q)$  (defined by (2.14)) may be written in the form

$$(3.25) \quad \begin{aligned} \tau_1^k &= \sum_{0 \leq i+j \leq k} b_{ij}^1 L_i(\xi) L_j(\eta) + b_{k+1,0}^1 L_{k+1}(\xi) L_0(\eta) \\ &\quad - c_k b_{0,k+1}^2 L_1(\xi) L_k(\eta), \end{aligned}$$

$$(3.26) \quad \begin{aligned} \tau_2^k &= \sum_{0 \leq i+j \leq k} b_{ij}^2 L_i(\xi) L_j(\eta) - c_k b_{k+1,0}^1 L_k(\xi) L_1(\eta) \\ &\quad + b_{0,k+1}^2 L_0(\xi) L_{k+1}(\eta), \end{aligned}$$

where

$$c_k = \frac{\text{leading coefficient of } L_{k+1}}{\text{leading coefficient of } L_k} = \frac{2k+1}{k+1}$$

satisfies  $1 \leq c_k < 2$ . Using (2.33)–(2.34), we see that  $(\tau_1^k, \tau_2^k)$  satisfies

$$(3.27) \quad \iint_Q (\tau_i^k - \tau_j^k) \phi(\xi, \eta) d\xi d\eta = 0, \quad \phi \in P_{k-2}(Q),$$

together with conditions (3.7)–(3.8). By the orthogonality of Legendre polynomials, we obtain from (3.27)

$$(3.28) \quad b_{ij}^n = a_{ij}^n, \quad 0 \leq i+j \leq k-2, \quad n = 1, 2.$$

Next, taking  $\zeta(\eta) = L_l(\eta)$ , the conditions (3.7) on the sides  $\xi = \pm 1$  give

$$(3.29) \quad \sum_{i=0}^{k-l} b_{il}^1 L_i(\pm 1) = \sum_{i=0}^{\infty} a_{il}^1 L_i(\pm 1) \quad \text{for } l = 1, 2, \dots, k-1,$$

$$(3.30) \quad \sum_{i=0}^k b_{i0}^1 L_i(\pm 1) + b_{k+1,0}^1 L_{k+1}(\pm 1) = \sum_{i=0}^{\infty} a_{i0}^1 L_i(\pm 1) \quad \text{for } l = 0,$$

$$(3.31) \quad b_{0k}^1 L_0(\pm 1) - c_k b_{0,k+1}^2 L_i(\pm 1) = \sum_{i=0}^{\infty} a_{ik}^1 L_i(\pm 1) \quad \text{for } l = k.$$

Using (3.28) with  $n = 1$ , and the fact that  $L_i(1) = 1$ ,  $L_i(-1) = (-1)^i$ , (3.29)–(3.31) give respectively

$$(3.32) \quad b_{k-l-1,l}^1 = \sum_{i=0}^{\infty} a_{k-l-1+i,l}^1, \quad b_{k-l,l}^1 = \sum_{i=0}^{\infty} a_{k-l+i,l}^1, \\ l = 1, 2, \dots, k-1,$$

$$(3.33) \quad b_{k-1,0}^1 + b_{k+1,0}^1 = \sum_{i=0}^{\infty} a_{k-1+i,0}^1, \quad b_{k0}^1 = \sum_{i=0}^{\infty} a_{k+i,0}^1,$$

$$(3.34) \quad b_{0k}^1 = \sum_{i=0}^{\infty} a_{ik}^1, \quad b_{0,k+1}^2 = -\frac{1}{c_k} \sum_{i=1}^{\infty} a_{ik}^1.$$

Similarly, the conditions on the sides  $\eta = \pm 1$  give

$$(3.35) \quad b_{l,k-l-1}^2 = \sum_{j=0}^{\infty} a_{l,k-l-1+j}^2, \quad b_{l,k-l}^2 = \sum_{j=0}^{\infty} a_{l,k-l+j}^2, \\ l = 1, 2, \dots, k-1,$$

$$(3.36) \quad b_{0,k-1}^2 + b_{0,k+1}^2 = \sum_{j=0}^{\infty} a_{0,k-1+j}^2, \quad b_{0k}^2 = \sum_{j=0}^{\infty} a_{0,k+j}^2,$$

$$(3.37) \quad b_{k0}^2 = \sum_{j=0}^{\infty} a_{kj}^2, \quad b_{k+1,0}^1 = -\frac{1}{c_k} \sum_{j=1}^{\infty} a_{kj}^2.$$

The only unknowns not explicitly solved in the above equations are  $b_{k-1,0}^1$  and  $b_{0,k-1}^2$ . These are given by

$$(3.38) \quad \begin{aligned} b_{k-1,0}^1 &= \sum_{i=0}^{\infty}'' a_{k-1+i,0}^1 + \frac{1}{c_k} \sum_{j=1}^{\infty}' a_{kj}^2, \\ b_{0,k-1}^2 &= \sum_{j=0}^{\infty}'' a_{0,k-1+j}^2 + \frac{1}{c_k} \sum_{i=1}^{\infty}' a_{ik}^1. \end{aligned}$$

We now use (3.28), (3.32)–(3.38) to estimate the error  $\|\tau_1 - \tau_1^k\|_{0,Q}$ . Let  $b_{ij}^1 = 0$  for those not explicitly specified above. Then we have

$$(3.39) \quad \begin{aligned} A &= \|\tau_1 - \tau_1^k\|_{0,Q}^2 = \sum_{j=0}^{\infty} \sum_{i=0}^{\infty} \frac{4(a_{ij}^1 - b_{ij}^1)^2}{(2i+1)(2j+1)} \quad (\text{where } b_{1k}^1 = -c_k b_{0,k+1}^2) \\ &= \sum_{i+j > k+1} \sum \frac{4(a_{ij}^1)^2}{(2i+1)(2j+1)} + \sum_{k-1 \leq i+j \leq k+1} \sum \frac{4(a_{ij}^1 - b_{ij}^1)^2}{(2i+1)(2j+1)} \\ &\leq C \left\{ \sum_{i+j \geq k-1} \sum \frac{(a_{ij}^1)^2}{(2i+1)(2j+1)} + \sum_{k-1 \leq i+j \leq k+1} \sum \frac{(b_{ij}^1)^2}{(2i+1)(2j+1)} \right\}. \end{aligned}$$

Now for  $r \geq 0$ ,

$$(3.40) \quad \begin{aligned} \sum_{i+j \geq k-1} \sum \frac{(a_{ij}^1)^2}{(2i+1)(2j+1)} &\leq C \sum_{i+j \geq k-1} \sum \frac{(a_{ij}^1)^2}{(2i+1)(2j+1)} \frac{(1+i^2+j^2)^r}{k^{2r}} \\ &\leq \frac{C}{k^{2r}} \|\tau_1\|_{r,Q}^2 \leq \frac{C}{k^{2r}} \|\tilde{\tau}\|_{r,Q}^2. \end{aligned}$$

Also, we know from (3.32)–(3.38) that for  $i+j \geq k-1$ ,

$$(3.41) \quad |b_{ij}^1| \leq \sum_{l=i}^{\infty} |a_{lj}^1| + C_{ij} \sum_{l=1}^{\infty}' |a_{kl}^2|,$$

where

$$\begin{aligned} C_{ij} &= 1 \quad \text{for } (i, j) = (k-1, 0) \text{ or } (k+1, 0), \\ &= 0 \quad \text{otherwise.} \end{aligned}$$

Now for  $r_1 > 1/2$  (since  $i+j \geq k-1$ ),

$$(3.42) \quad \begin{aligned} \left( \sum_{l=i}^{\infty} |a_{lj}^1| \right)^2 &\leq \sum_{l=i}^{\infty} (a_{lj}^1)^2 (1+l^2+j^2)^{r_1} \sum_{l=i}^{\infty} (1+l^2+j^2)^{-r_1} \\ &\leq \frac{C}{k^{2r_1-1}} \sum_{l=i}^{\infty} \frac{(a_{lj}^1)^2 (1+l^2+j^2)^{r_1+1/2}}{(2l+1)}. \end{aligned}$$

Similarly,

$$(3.43) \quad \left( \sum_{l=1}^{\infty}' |a_{kl}^2| \right)^2 \leq \frac{C}{k^{2r_1-1}} \sum_{l=0}^{\infty} \frac{(a_{kl}^2)^2 (1+k^2+l^2)^{r_1+1/2}}{(2l+1)}.$$

Hence, we see that by (3.41)–(3.43),

$$\begin{aligned}
 & \sum_{i+j=k-1} \sum \frac{(b_{ij}^1)^2}{(2i+1)(2j+1)} \\
 &= \sum_{j=0}^{k-1} \frac{(b_{k-1-j,j}^1)^2}{(2k-2j-1)(2j+1)} \leq \sum_{j=0}^{k-1} \frac{(b_{k-1-j,j}^1)^2}{(2j+1)} \\
 (3.44) \quad & \leq \frac{C}{k^{2r_1-1}} \left\{ \sum_{j=0}^{\infty} \sum_{l=0}^{\infty} \frac{(a_{lj}^1)^2 (1+l^2+j^2)^{r_1+1/2}}{(2l+1)(2j+1)} \right. \\
 & \quad \left. + \sum_{l=0}^{\infty} \frac{(a_{kl}^2)^2 (1+k^2+l^2)^{r_1+1/2}}{(2l+1)(2k+1)} \right\} \\
 & \leq \frac{C}{k^{2r-2}} \|\tilde{\tau}\|_{r,Q},
 \end{aligned}$$

where  $r = r_1 + \frac{1}{2} > 1$ .

The term

$$\sum_{k \leq i+j \leq k+1} \sum \frac{(b_{ij})^2}{(2i+1)(2j+1)}$$

may be bounded similarly to (3.44). Hence, by (3.39), (3.40), (3.44) we obtain for  $r > 1$ ,

$$(3.45) \quad \|\tau_1 - \tau_1^k\|_{0,Q}^2 \leq \frac{C}{k^{2r-2}} \|\tilde{\tau}\|_{r,Q}.$$

The term  $\|\tau_2 - \tau_2^k\|_{0,Q}^2$  can be treated the same way, leading to the following lemma.

**Lemma 3.3.** *Let  $\tilde{\tau} \in (H^r(Q))^2$ ,  $r > 1$ . Let  $\Pi_k^2 \tilde{\tau} = (\tau_1^k, \tau_2^k) \in S_k^2(Q)$  be defined by (3.27), (3.7)–(3.8). Then*

$$(3.46) \quad \|\tilde{\tau} - \Pi_k^2 \tilde{\tau}\|_{0,Q} \leq C k^{-(r-1)} \|\tilde{\tau}\|_{r,Q},$$

where  $C$  is a constant independent of  $k$ ,  $\tilde{\tau}$ .

Lemma 3.3 then yields the following theorem, which can be proved the same way as Theorem 3.1.

**Theorem 3.2.** *Let  $\tau \in (H^r(\Omega))^2$ ,  $r > 1$ . Let  $\Pi_N^2 : S \rightarrow S_N^2$  be as defined in §2. Then*

$$(3.47) \quad \|\tau - \Pi_N^2 \tau\|_{0,\Omega} \leq C h_N^{\min(p_N+1, r)} p_N^{-(r-1)} \|\tau\|_{r,\Omega},$$

where  $C$  is a constant independent of  $h_N$ ,  $p_N$  and  $\tau$  but depends upon  $r$ .

Moreover, if  $\operatorname{div} \tau \in H^r(\Omega)$ , then

$$(3.48) \quad \|\tau - \Pi_N^2 \tau\|_S \leq C h_N^{\mu_2} p_N^{-\beta_2} (\|\tau\|_{r,\Omega} + \|\operatorname{div} \tau\|_{r,\Omega}),$$

where  $\mu_2 = \min(p_N, r)$  and  $\beta_2 = r - 1$ .

## 4. STABILITY AND CONVERGENCE RESULTS

In this section, we examine the dependence of the stability constant  $\alpha(N) = \alpha(h_N, p_N)$  in (2.8) on  $h_N$  and  $p_N$ , when the RT and the BDM spaces are used. We also use Theorems 2.1, 3.1 and 3.2 to derive error estimates for the mixed method defined by (2.5)–(2.6).

We first examine the question of stability.

**Theorem 4.1.** *The spaces  $\{V_N^i, S_N^i\}$ ,  $i = 1, 2$ , satisfy condition (2.8) of Theorem 2.1 with the stability constant  $\alpha_i(N) = \alpha_i(h_N, p_N)$  being given by*

$$(4.1) \quad \alpha_1(N) \geq C_1, \quad \text{for RT spaces,}$$

$$(4.2) \quad \alpha_2(N) \geq \frac{C_3}{1 + C_2 h_N p_N^\varepsilon} \quad \text{for BDM spaces,}$$

where  $C_i$ ,  $i = 1, 2, 3$ , are constants independent of  $h_N$ ,  $p_N$ , and  $\varepsilon > 0$  may be chosen to be arbitrarily small.

*Proof.* Let  $v_N \in V_N^i$ . In order to establish (2.8), it is sufficient to find  $\tau_N \in S_N^i$  such that

$$(4.3) \quad \operatorname{div} \tau_N = v_N,$$

$$(4.4) \quad \|\tau_N\|_S \leq \frac{1}{\alpha_i(N)} \|v_N\|_V.$$

We first solve the following elliptic problem on  $\Omega$ :

$$\Delta u = v_N \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega.$$

Let  $\tau = \operatorname{grad} u$ . Then we have

$$(4.5) \quad \operatorname{div} \tau = v_N.$$

Moreover, since  $\Omega$  is convex, there exists a  $0 < \varepsilon_0 < \frac{1}{2}$  such that the following shift theorem holds:

$$(4.6) \quad \|\tau\|_{1+\varepsilon, \Omega} \leq \tilde{C} \|v_N\|_{\varepsilon, \Omega} \quad \text{for all } 0 \leq \varepsilon < \varepsilon_0.$$

(Note that  $v_N \in V_N^i \subset H^s(\Omega)$  for any  $0 \leq s < \frac{1}{2}$ .) We now take

$$(4.7) \quad \tau_N = \Pi_N^i \tau.$$

Then, since  $\operatorname{div} \tau \in V_N^i$ , (4.3) follows by (2.38). Moreover,

$$\begin{aligned} \|\tau_N\|_{0, \Omega} &\leq \|\tau\|_{0, \Omega} + \|\tau - \Pi_N^i \tau\|_{0, \Omega} \\ &\leq \|\tau\|_{1, \Omega} + C h_N^{1+\varepsilon} p_N^{-(1+\varepsilon-\gamma_i)} \|\tau\|_{1+\varepsilon, \Omega}, \end{aligned}$$

where  $\gamma_1 = \frac{1}{2}$ ,  $\gamma_2 = 1$  and  $0 < \varepsilon < \varepsilon_0$ , by (3.18), (3.47). This gives, by (4.6),

$$(4.8) \quad \begin{aligned} \|\tau_N\|_{0, \Omega} &\leq \|\tau\|_{1, \Omega} + C h_N^{1+\varepsilon} p_N^{-(1+\varepsilon-\gamma_i)} \|v_N\|_{\varepsilon, \Omega} \\ &\leq C(1 + C_2 h_N^{1+\varepsilon} p_N^{-(1+\varepsilon-\gamma_i)} h_N^{-\varepsilon} p_N^{2\varepsilon}) \|v_N\|_{0, \Omega}, \end{aligned}$$

where we have used Lemma 2.3 (with  $\varepsilon$  small enough). Since  $\|\operatorname{div} \tau_N\|_{0,\Omega} = \|v_N\|_{0,\Omega}$ , (4.8) shows that (4.4) holds with

$$\begin{aligned} \frac{1}{\alpha_1(N)} &\leq C(1 + C_2 h_N p_N^{-(1/2-\varepsilon)}) \leq C_1^{-1} \quad \text{for RT}, \\ \frac{1}{\alpha_2(N)} &\leq C(1 + C_2 h_N p_N^\varepsilon) \quad \text{for BDM}. \end{aligned}$$

The assertions (4.1)–(4.2) follow immediately.  $\square$

We see, therefore, that the RT spaces are stable. For the BDM spaces, (4.2) guarantees stability up to an arbitrarily small power  $p_N^{-\varepsilon}$ . (Obviously, if  $h_N p_N^\varepsilon$  remains bounded, then in (4.2) we obtain  $\alpha_2(N) \geq C$ .) We may now apply Theorem 2.1 and obtain the following rates of convergence, using the approximation estimates in Theorems 2.2, 3.1 and 3.2.

**Corollary 4.1.** *Let  $u$  be the solution of (2.1), with  $\sigma = \operatorname{grad} u$ , so that  $(u, \sigma) \in V \times S$  satisfy (2.3)–(2.4). Let  $(u_N^1, \sigma_N^1) \in V_N^1 \times S_N^1$  be the finite element solutions corresponding to the RT spaces. Then there exists a constant  $C$  independent of  $h_N, p_N, u$  such that for  $r \geq 1$ ,*

$$(4.9) \quad \|\sigma_N^1\|_S + \|u_N^1\|_V \leq C\{\|u\|_0 + \|\sigma\|_0 + \|\operatorname{div} \sigma\|_0\},$$

$$(4.10) \quad \|\sigma - \sigma_N^1\|_S + \|u - u_N^1\|_V \leq C h_N^{\min(p_N+1, r)} p_N^{-(r-1/2)} \{\|u\|_r + \|\sigma\|_r + \|\operatorname{div} \sigma\|_r\}.$$

**Corollary 4.2.** *Let  $(u, \sigma)$  be as in Corollary 4.1 and  $(u_N^2, \sigma_N^2) \in V_N^2 \times S_N^2$  be the finite element solutions corresponding to the BDM spaces. Then there exists a constant  $C$  independent of  $h_N, p_N$  and  $u$  such that for any  $\varepsilon > 0$ , for  $r \geq 1$ ,*

$$(4.11) \quad \|\sigma_N^2\|_S + \|u_N^2\|_V \leq C p_N^\varepsilon \{\|u\|_0 + \|\sigma\|_0 + \|\operatorname{div} \sigma\|_0\},$$

$$(4.12) \quad \begin{aligned} &\|\sigma - \sigma_N^2\|_S + \|u - u_N^2\|_V \\ &\leq C h_N^{\min(p_N, r)} p_N^{-(r-1-\varepsilon)} \{\|u\|_r + \|\sigma\|_r + \|\operatorname{div} \sigma\|_r\}. \end{aligned}$$

The above estimates are optimal in  $h_N$  but not in  $p_N$ . We now show how they can be improved in terms of  $p_N$  to give optimal estimates (up to an arbitrary  $\varepsilon > 0$ ). The argument used was first introduced in [8].

We first require the following interpolation result.

**Lemma 4.1.** *For any  $r \geq 0$ , let  $X^r = H^r(\Omega) \times Y^r$ , where  $Y^r$  denotes the completion of  $(C_0^\infty(\Omega))^2$  functions under the following norm:*

$$\|\sigma\|_{Y^r}^2 = \|\sigma\|_r^2 + \|\operatorname{div} \sigma\|_r^2.$$

*Let for  $q = r_1 + \theta(r_2 - r_1)$ ,  $r_2 \geq r_1 \geq 0$ ,  $0 \leq \theta \leq 1$ ,  $\tilde{X}^q$  denote the interpolation space  $[X^{r_1}, X^{r_2}]_\theta$  using the  $K$ -method of interpolation (see [13]). Then*

$$(4.13) \quad \tilde{X}^q = X^q.$$



*Proof.* We first show that

$$(4.14) \quad \tilde{Y}^q = Y^q,$$

where  $\tilde{Y}^q = [Y^{r_1}, Y^{r_2}]_\theta$ . We note that for  $i = 1, 2$ ,  $Y^{r_i}$  may be defined as

$$\{\sigma \mid \sigma \in (H^{r_i}(\Omega))^2, \partial\sigma \in H^{r_i}(\Omega)\},$$

where  $\partial \equiv \text{div} \in \mathcal{L}((H^r(\Omega))^2, H^{r-1}(\Omega))$  for any  $r \geq 0$ . Moreover, as shown below, there exists an operator  $\mathcal{G}$  which belongs to  $\mathcal{L}(H^{r-1}(\Omega), (H^r(\Omega))^2)$  for all  $r \geq 0$  such that

$$(4.15) \quad \partial\mathcal{G}\chi = \chi \quad \forall \chi \in H^{r-1}(\Omega).$$

Hence, Theorem 14.3 of [13] allows us to interpolate between spaces  $Y^{r_i}$  and obtain (4.14). Then (4.13) follows easily by a standard result on the interpolation of products of spaces (equation (6.42), Chapter 2 of [13]).

To define the operator  $\mathcal{G}$ , we first let  $\tilde{\chi}$  denote an extension of  $\chi \in H^{r-1}(\Omega)$  ( $r \geq 0$ ) to  $\mathbb{R}^2$  such that

$$(4.16) \quad \|\tilde{\chi}\|_{H^{r-1}(\mathbb{R}^2)} \leq C\|\chi\|_{H^{r-1}(\Omega)}.$$

(The details of this extension may be found, for example, in [16].) Next, let  $\tilde{w}$  satisfy  $\Delta\tilde{w} = \tilde{\chi}$  on  $\mathbb{R}^2$  such that the shift theorem holds. Taking

$$\mathcal{G}\chi = \text{grad } \tilde{w}|_\Omega,$$

we see that (4.15) holds, and

$$\|\text{grad } \tilde{w}\|_{(H^r(\mathbb{R}^2))^2} \leq C\|\tilde{\chi}\|_{H^{r-1}(\mathbb{R}^2)} \leq C\|\chi\|_{H^{r-1}(\Omega)}$$

by (4.16) and

$$\|\mathcal{G}\chi\|_{(H^r(\Omega))^2} = \|\text{grad } \tilde{w}|_\Omega\|_{(H^r(\Omega))^2} \leq \|\text{grad } \tilde{w}\|_{(H^r(\mathbb{R}^2))^2}$$

for any  $r \geq 0$ , so that  $\mathcal{G} \in \mathcal{L}((H^r(\Omega))^2, H^{r-1}(\Omega))$  as required.  $\square$

We now prove the following theorem.

**Theorem 4.2.** *Let  $u$  be the solution of (2.1) with  $\sigma = \text{grad } u$ , so that  $(u, \sigma) \in V \times S$  satisfy (2.3)–(2.4). Let  $(u_N^1, \sigma_N^1) \in V_N^1 \times S_N^1$  be the finite element solutions corresponding to the RT spaces. Then, given any  $\varepsilon > 0$ ,  $k \geq 0$ , there exists a constant  $C$  independent of  $h_N$ ,  $p_N$ ,  $u$  but depending upon  $\varepsilon$  and  $k$  such that*

$$(4.17) \quad \|\sigma - \sigma_N^1\|_S + \|u - u_N^1\|_V \leq Ch_N^{\min(p_N+1, k)} p_N^{-(k-\varepsilon)} \{\|u\|_k + \|\sigma\|_k + \|\text{div } \sigma\|_k\}.$$

*Proof.* We first use (4.9) to obtain the estimate

$$(4.18) \quad \|\sigma - \sigma_N^1\|_S + \|u - u_N^1\|_V \leq Ch_N^{\min(p_N+1, 0)} p_N^0 \|(u, \sigma)\|_{X^0},$$

where  $X^r$  is as defined in Lemma 4.1. Next, given  $\varepsilon > 0$ ,  $k \geq 0$ , choose  $r$  in Corollary 4.1 such that  $r > k/2\varepsilon$ . Then (4.10) states that

$$(4.19) \quad \|\sigma - \sigma_N^1\|_S + \|u - u_N^1\|_V \leq Ch_N^{\min(p_N+1, r)} p_N^{-(r-1/2)} \|(u, \sigma)\|_{X^r}.$$

We now interpolate between (4.18) and (4.19) with  $\theta = k/r$  so that  $0 < \theta/2 < \varepsilon$ . This gives

$$\begin{aligned} \|\sigma - \sigma_N^1\|_S + \|u - u_N^1\|_V &\leq Ch_N^{\min(p_N+1, \theta r)} p_N^{-(\theta r - \theta/2)} \|(u, \sigma)\|_{\tilde{X}^{\theta r}}, \\ &\leq Ch_N^{\min(p_N+1, k)} p_N^{-(k-\varepsilon)} \|(u, \sigma)\|_{X^k}, \end{aligned}$$

where we have used Lemma 4.1. This proves the theorem.  $\square$

**Remark 4.1.** Although the constant  $C$  in (4.17) depends on  $\varepsilon$ , Theorem 4.2 asserts that if one chooses any positive  $\varepsilon$ , no matter how small, and fixes it, then one can find a constant  $C$  such that (4.17) holds. Hence the rate of convergence in  $p_N$  is optimal up to any arbitrarily small  $\varepsilon > 0$ .

We can obtain a theorem similar to this for the BDM spaces. The proof is essentially the same, except that  $r$  must now satisfy  $r > k/\varepsilon$ .

**Theorem 4.3.** *Let  $(u, \sigma)$  be as in Theorem 4.2. Let  $(u_N^2, \sigma_N^2) \in V_N^2 \times S_N^2$  be the finite element solutions corresponding to the BDM spaces. Then given any  $\varepsilon > 0$ ,  $k \geq 0$ , there exists a constant  $C$  independent of  $h_N$ ,  $p_N$ ,  $u$  but depending upon  $\varepsilon$  and  $k$  such that*

$$(4.20) \quad \|\sigma - \sigma_N^2\|_S + \|u - u_N^2\|_V \leq Ch_N^{\min(p_N, k)} p_N^{-(k-\varepsilon)} \{\|u\|_k + \|\sigma\|_k + \|\operatorname{div} \sigma\|_k\}.$$

**Remark 4.2.** Theorems 4.2 and 4.3 can now be used to give an estimate for the asymptotic rate of convergence when  $h$  and  $p$  are changed either separately or together. It may be observed, for example, that using the  $h$ -version with  $p_N > 1$  changes not only the exponent of  $h_N$  but can also lead to a substantial decrease in the “constant” which decays asymptotically like  $Cp_N^{-(k-\varepsilon)}$ .

## BIBLIOGRAPHY

1. I. Babuška, *The  $p$ - and  $h-p$  versions of the finite element method. The state of the art*, in Finite Elements Theory and Application (D. L. Dwyer, M. Y. Hussaini and R. G. Voigt, eds.), Springer-Verlag, New York, 1988, pp. 199–239.
2. —, *Are high degree elements preferable? Some aspects of the  $h$  and  $h-p$  versions of the finite element method*, in Numerical Techniques for Engineering Analysis and Design, Vol. I (G. N. Pande and J. Middleton, eds.), Martinus Nijhoff Publishers, 1987.
3. I. Babuška and M. R. Dorr, *Error estimates for the combined  $h$  and  $p$  version of the finite element method*, Numer. Math. **37** (1981), 257–277.
4. I. Babuška and T. Scapolla, *The computational aspects of the  $h$ ,  $p$  and  $h-p$  versions of the finite element method*, in Advances in Computer Methods for PDEs–VI (R. Vichnevetsky and R. S. Stepleman, eds.), IMACS, 1987.
5. I. Babuška and M. Suri, *The  $h-p$  version of the finite element method with quasiuniform meshes*, RAIRO Modél. Math. Anal. Numér. **21** (1987), 199–238.
6. —, *The treatment of nonhomogeneous Dirichlet boundary conditions by the  $p$ -version of the finite element method*, Numer. Math. **55** (1989), 97–121.
7. I. Babuška and B. A. Szabo, *Lecture notes on finite element analysis*, In preparation.
8. I. Babuška, B. A. Szabo and I. N. Katz, *The  $p$ -version of the finite element method*, SIAM J. Numer. Anal. **18** (1981), 515–545.
9. F. Brezzi, J. Douglas and L. D. Marini, *Two families of mixed finite elements for second order elliptic problems*, Numer. Math. **47** (1985), 217–235.
10. P. G. Ciarlet, *The finite element method for elliptic problems*, North-Holland, Amsterdam, 1978.

11. S. Jensen and M. Vogelius, *Divergence stability in connection with the  $p$ -version of the finite element method*, to appear in SIAM J. Numer. Anal., 1990.
12. C. Johnson and V. Thomée, *Error estimates for some mixed finite element methods for parabolic type problems*, RAIRO Anal. Numér. **15** (1981), 41–78.
13. J. L. Lions and E. Magenes, *Non-homogeneous boundary value problems and applications I*, Springer-Verlag, Berlin and New York, 1972.
14. P. A. Raviart and J. M. Thomas, *A mixed finite element method for second order elliptic problems*, Proc. Sympos. Mathematical Aspects of the Finite Element Method (Rome, 1975), Lecture Notes in Math., vol. 606, Springer-Verlag, Berlin, 1977, pp. 292–315.
15. L. R. Scott and M. Vogelius, *Norm estimates for a maximal right inverse of the divergence operator in spaces of piecewise polynomials*, RAIRO Modél. Math. Anal. Numér. **19** (1985), 111–143.
16. E. M. Stein, *Singular integrals and differentiability properties of functions*, Princeton Univ. Press, Princeton, N.J., 1970.

DEPARTMENT OF MATHEMATICS AND STATISTICS, UNIVERSITY OF MARYLAND BALTIMORE COUNTY, BALTIMORE, MARYLAND 21228. E-mail: suri@umbc1.umbc.edu