MATHEMATICS OF COMPUTATION
VOLUME 41, NUMBER 164
OCTOBER 1983, PAGES 399-423

Analysis of Some Mixed Finite Element Methods for Plane Elasticity Equations

By J. Pitkäranta and R. Stenberg

Abstract. We analyze some mixed finite element methods, based on rectangular elements, for solving the two-dimensional elasticity equations. We prove error estimates for a method proposed by Taylor and Zienkiewicz and for some new variants of the known equilibrium methods. A numerical example is given demonstrating the performance of the various algorithms considered.

1. Introduction. In the numerical solution of problems of continuum mechanics, the stresses are normally of primary interest in the elastic region. It is therefore natural to design the numerical algorithms so that the stresses can be obtained directly without first computing the displacements. Such methods can be derived from the dual variational formulation of the elasticity problem. The corresponding finite element algorithms are usually formulated as mixed methods where both the displacements and the stresses are first approximated, and the displacements are then eliminated from the discrete equations. In many cases the elimination can be rather effectively done using penalty/perturbation techniques or their iterative variants; cf. [3], [11], [12].

The best known finite element methods of the above type are the so-called equilibrium methods, first proposed by Fraejis de Veubeke [17] (cf. also [14], [16], [18]) and analyzed theoretically by Johnson and Mercier [9] (cf. also [8]). In these methods, one uses specific composite elements which allow the equilibrium condition between the stresses and the volume load to be satisfied exactly in the case where the volume load is zero.

The main drawback of the equilibrium methods proposed so far is the relatively high number of free parameters as compared with displacement methods of the same order of accuracy. For example, if the composite quadrilateral element of [17], [9] is used on a regular rectangular grid, one has eight degrees of freedom per each interior node of the grid (after the local condensation of three extra degrees of freedom per node, cf. [9]) and the convergence rate $O(h^2)$ for the stresses in L_2 [9]. On the other hand, using the displacement method with reduced biquadratic elements (cf. [6]), one has the same convergence rate with six parameters per node, so the displacement method seems superior.

It is clear from the above example that the mixed or equilibrium methods should be further developed if they are desired to be competitive with displacement

Received August 18, 1982.
1980 *Mathematics Subject Classification.* Primary 65N30.

methods. This is the motivation of the present paper. In particular, we try to find mixed or equilibrium methods which are simpler than those considered in [9] and still preserve the quadratic convergence rate of the stresses in L_2. We analyze in detail two candidates for such methods. In the first method, called Method I below, the stresses are approximated by continuous piecewise bilinear functions on a rectangular grid, and the displacements are taken to be piecewise constant on the same grid. This method (which probably is the simplest possible mixed method one can think of) was proposed recently by Taylor and Zienkiewicz [15]. The convergence rate of this method, however, does not seem to be quadratic. We are able to prove, under various restrictive assumptions, that the stresses converge with the rate $O(h^{3/2})$ if the exact solution is sufficiently smooth. That this result is actually optimal is confirmed numerically.

As another alternative, called Method II below, we consider a class of algorithms based on the composite quadrilateral element of [17], [9]. We show that many of the degrees of freedom can be eliminated without affecting the convergence rate. In particular, we derive a method which contains only the average of four free parameters per node and still gives quadratic convergence rate for stresses in L_2.

We consider only the case of a uniform rectangular mesh on a rectangular domain in this paper. The assumption on mesh uniformly seems essential for Method I, but for Method II the results can very likely be extended to more general quadrilateral meshes.

The plan of the paper is as follows. In Section 2 we state the problem and its finite element discretization in a general form. In Sections 3 and 4 we analyze the two methods, and in Section 5 we present some results of numerical computations with both methods.

2. Notation and Preliminaries. Let us recall the basic problem of linear elasticity in two dimensions (plane stress or plane strain): given $f = (f_1, f_2)$ find a symmetric stress tensor $\sigma = \{\sigma_{ij}\}$, $i, j = 1, 2$ and a displacement (u_1, u_2) satisfying

$$\begin{aligned} \varepsilon(u) &= \lambda \operatorname{tr}(\sigma)\delta + \mu\sigma \quad \text{in } \Omega, \\ \operatorname{div}\sigma + f &= 0 \qquad\qquad \text{in } \Omega, \end{aligned} \tag{2.1}$$

subject to the boundary conditions

$$\begin{aligned} u &= 0 \qquad \text{on } \Gamma_1, \\ \sigma \cdot n &= 0 \qquad \text{on } \Gamma_2. \end{aligned}$$

Here

$$\varepsilon(u) = \{\varepsilon_{ij}(u)\}, \qquad \varepsilon_{ij} = \frac{1}{2}\left(\frac{\partial u_i}{\partial x_j} + \frac{\partial u_j}{\partial x_i}\right), \qquad i, j = 1, 2,$$

is the deformation tensor,

$$\begin{aligned} &\operatorname{tr}(\sigma) = \sigma_{11} + \sigma_{22}, \\ &\operatorname{div}\sigma = \left\{\frac{\partial \sigma_{11}}{\partial x_1} + \frac{\partial \sigma_{12}}{\partial x_2}, \frac{\partial \sigma_{21}}{\partial x_1} + \frac{\partial \sigma_{22}}{\partial x_2}\right\}, \\ &\sigma \cdot n = (\sigma_{n,1}, \sigma_{n,2}) = (n_1\sigma_{11} + n_2\sigma_{12}, n_1\sigma_{21} + n_2\sigma_{22}), \\ &\delta = \{\delta_{ij}\}, \qquad \delta_{ij} = \begin{cases} 1 & \text{if } i = j, \\ 0 & \text{if } i \neq j, \end{cases} \quad i, j = 1, 2, \end{aligned}$$

λ and μ are constants satisfying $\mu > 0$ and $2\lambda + \mu > 0$, Ω is a bounded region in the plane with boundary $\Gamma = \Gamma_1 \cup \Gamma_2$, and $n = (n_1, n_2)$ denotes the unit outward normal vector on Γ. We have assumed homogeneous boundary conditions for simplicity.

We assume below that Ω is a rectangle: $\Omega = \{x = (x_1, x_2) \in \mathbf{R}^2,\ 0 < x_1 < a_1,\ 0 < x_2 < a_2\}$. Each side of Ω is assumed to be fully contained in either Γ_1 or Γ_2, i.e., the boundary conditions can change only at the vertices of Ω. We also assume that Γ_1 is nonempty.

Introducing the spaces

$$V = V(\Omega) = [L_2(\Omega)]^2,$$
$$Y = Y(\Omega) = \{\tau = (\tau_{ij}) : \tau_{ij} \in L_2(\Omega), \tau_{ij} = \tau_{ji}, i, j = 1, 2\},$$
$$H = H(\Omega) = \{\tau \in Y(\Omega) : \operatorname{div} \tau \in V(\Omega), \tau \cdot n = 0 \text{ on } \Gamma_2\},$$

the elasticity problem can be given the following variational formulation: Find $(\sigma, u) \in H \times V$ such that

$$\begin{cases} a(\sigma, \tau) + (u, \operatorname{div} \tau) = 0, & \tau \in H, \\ (\operatorname{div} \sigma, v) + (f, v) = 0, & v \in V, \end{cases} \tag{2.2}$$

where $(\cdot, \cdot)$ denotes the scalar product in V and

$$a(\sigma, \tau) = \int_\Omega (\lambda \operatorname{tr}(\sigma)\operatorname{tr}(\tau) + \mu \sigma \cdot \tau)\, dx,$$

where

$$\sigma \cdot \tau = \sum_{i,j=1}^{2} \sigma_{ij}\tau_{ij}.$$

Since Γ_1 is nonempty, (2.2) has a unique solution; cf. [7].

Below we consider finite element methods of the form: Find $(\sigma_h, u_h) \in H_h \times V_h$ such that

$$a(\sigma_h, \tau) + (u_h, \operatorname{div} \tau) = 0, \qquad \tau \in H_h, \tag{2.3a}$$

$$(\operatorname{div} \sigma_h, v) + (f, v) = 0, \qquad v \in V_h, \tag{2.3b}$$

where H_h and V_h are finite element subspaces of H and V, respectively. In practice, Eqs. (2.3) are often solved by introducing a small perturbation parameter $\varepsilon > 0$ and replacing (2.3b) by

$$-\varepsilon(u_h, v) + (\operatorname{div} \sigma_h, v) + (f, v) = 0, \qquad v \in V_h. \tag{2.3b'}$$

If π_h denotes the orthgonal projection of $[L_2(\Omega)]^2$ onto V_h, (2.3b') may be written as

$$u_h = \frac{1}{\varepsilon}\pi_h(\operatorname{div} \sigma_h + f).$$

Upon substituting this into (2.3a) one obtains

$$a(\sigma_h, \tau) + \frac{1}{\varepsilon}(\pi_h(\operatorname{div} \sigma_h + f), \pi_h \operatorname{div} \tau) = 0, \qquad \tau \in H_h. \tag{2.4}$$

This is a modified penalty method for approximately solving (2.1). The operator π_h corresponds frequently to the use of "selective reduced integration", i.e., some low-order quadrature rule for computing the integral in the penalty term; cf. [3], [12].

Below we associate the spaces V_h and H_h to a uniform rectangular partitioning of Ω defined by

$$\mathcal{C}^h = \{K_{ij},\ i = 1,\dots,m_1,\ j = 1,\dots,m_2\}, \tag{2.5}$$

where

$$K_{ij} = \{x \in \mathbf{R}^2 \colon (i-1)h_1 < x_1 < ih_1,\ (j-1)h_1 < x_2 < jh_1\}.$$

Here $m_i h_i = a_i$, $i = 1, 2$, and h_1 and h_2 are associated to the mesh parameter $h \equiv h_1$ in such a way that h_1/h_2 is bounded from both above and below by a positive constant independent of h.

The set of nodal points in the grid induced by the partitioning $\mathcal{C}^h$ is denoted by $\mathfrak{N}$:

$$\mathfrak{N} = \{P_{ij} = (ih_1, jh_2),\ i = 0,\dots,m_1,\ j = 0,\dots,m_2\}.$$

In all of the methods considered in this paper, the functions in V_h are fully discontinuous along the mesh lines, so that each $v \in V_h$ can be expressed in the form

$$v(x) = \sum_{i=1}^{m_1} \sum_{j=1}^{m_2} \sum_{k=1}^{\nu} \alpha_{ijk}\phi_{ijk}(x), \qquad \alpha_{ijk} \in \mathbf{R},$$

where ϕ_{ijk} vanishes outside $K_{ij} \in \mathcal{C}^h$. Similarly, the space H_h is chosen to consist of functions of the form

$$\tau(x) = \sum_{P \in \mathfrak{N}} \sum_{k=1}^{\nu_P} \alpha_P^k \xi_P^k(x), \qquad \alpha_P^k \in \mathbf{R},$$

where the basis function $\xi_P^k \in H_h$ vanishes in any $K \in \mathcal{C}^h$ that does not have a vertex at P. For this type of choice of the subspaces, the projection π_h in (2.4) can be computed locally in each rectangle $K \in \mathcal{C}^h$. It is also obvious that in (2.4) there is coupling only between the nodes that are vertices of the same rectangle $K \in \mathcal{C}^h$.

We shall denote by $|\cdot|_{s,p,T}$ and $\|\cdot\|_{s,p,T}$, respectively, the seminorm and the norm of the Sobolev space $[W^{s,p}(T)]^\alpha$, s and α integers, $1 \leq p \leq \infty$. For nonintegral s, $s \geq 0$, $\|\cdot\|_{s,p,T}$ is defined as usual by interpolation. If $p = 2$, we set $\|\cdot\|_{s,2,T} = \|\cdot\|_{s,T}$. The subscript T is dropped if $T = \Omega$.

In the analysis below we need frequently the partial integration formula

$$\int_T \operatorname{div}\tau \cdot v\, dx = \int_{\partial T} (\tau \cdot n) \cdot v\, ds - \int_T \tau \cdot \varepsilon(v)\, dx, \tag{2.6}$$

where n is the unit outward normal vector to ∂T. (2.6) is valid if τ is a symmetric tensor such that $\tau_{ij} \in L_2(T)$ and $\operatorname{div}\tau \in [L_2(T)]^2$, and $v \in [H^1(T)]^2$. We also recall that since $\mu > 0$ and $2\lambda + \mu > 0$, we have (cf. [7])

$$a(\tau, \tau) \geq C\|\tau\|_0^2, \qquad \tau \in Y. \tag{2.7}$$

Here and below, C denotes a positive constant which may depend on Ω and on the parameters λ and μ in (2.1) but is independent of other parameters unless indicated explicitly. We shall finally denote by $P_k(T)$, $T \subset \mathbf{R}^2$, the set of polynomials in two variables of degree $\leq k$ defined on T.

3. Analysis of Method I. We consider here the method proposed in [15]. Let S_h be the set of continuous piecewise bilinear functions on $\mathcal{C}_h$ and let Q_h be the set of functions that are piecewise constant on $\mathcal{C}^h$. Then the method of [15] is equivalent to choosing the subspace H_h and V_h in (2.2) as

$$H_h = H(\Omega) \cap [S_h]^4, \qquad V_h = [Q_h]^2.$$

Obviously any $\tau \in H_h$ is determined uniquely by the values of $\tau_{ij}(P)$, $P \in \mathfrak{N}$, $P \notin \bar{\Gamma}_2$. If $P \in \bar{\Gamma}_2$, then $(\tau \cdot n)(P) = 0$, $\tau \in H_h$.

Let (σ, u) be the solution of (2.2), let $\tilde{\sigma} \in H_h$ be the interpolant of σ defined by

$$\tilde{\sigma}(P) = \sigma(P), \qquad P \in \mathfrak{N},$$

(assuming that σ is continuous) and let $\tilde{u}$ be the L_2-projection of u into V_h. We need the following three lemmas.

LEMMA 3.1. *Assume that Γ_1 contains two adjacent sides of Ω. Then there is a constant C such that for all $v \in V_h$*

$$\sup_{\tau \in H_h} \frac{(\operatorname{div} \tau, v)}{\|\tau\|_0} \geqslant C\|v\|_0. \quad \square$$

LEMMA 3.2. *For all $v \in V_h$,*

$$|(\operatorname{div}(\sigma - \tilde{\sigma}), v)| \leqslant Ch^k |\sigma|_{k+1}\|v\|_0, \qquad k = 2, 3. \quad \square$$

LEMMA 3.3. *For all $\tau \in H_h$,*

$$|(u - \tilde{u}, \operatorname{div} \tau)| \leqslant C\big(h^2 |u|_3 + h^{3/2} |u|_{2,\infty}\big)\|\tau\|_0. \quad \square$$

Remark. Lemma 3.1 states a weak Babuška-Brezzi-type stability estimate (cf. [1], [5]) for the method (2.3). Without the additional assumption on the boundary condition, we have only been able to prove that

$$\sup_{\tau \in H_h} \frac{(\operatorname{div} \tau, v)}{\|\tau\|_0} \geqslant Ch\|v\|_0, \qquad v \in V_h. \quad \square$$

Before proving the lemmas, let us show that they imply the following error estimate.

THEOREM 3.1. *Assume that Γ_1 contains two adjacent sides of Ω. Let (σ, u) be the solution of* (2.2) *and (σ_h, u_h) the solution of* (2.3), *where the subspaces H_h and V_h are defined as above. Further let $\tilde{u} \in V_h$ be the L_2-projection of u. Then we have the error estimate*

$$\|\sigma - \sigma_h\|_0 + \|u_h - \tilde{u}\|_0 \leqslant Ch^{3/2}\|u\|_{7/2}. \quad \square$$

PROOF. By (2.2) and (2.3) we have the identity

$$\mathfrak{B}(\sigma_h - \tilde{\sigma}, u_h - \tilde{u}; \tau, v) = \mathfrak{B}(\sigma - \tilde{\sigma}, u - \tilde{u}; \tau, v), \qquad (\tau, v) \in H_h \times V_h, \tag{3.1}$$

where

$$\mathfrak{B}(\sigma, u; \tau, v) = a(\sigma, \tau) + (u, \operatorname{div} \tau) - (v, \operatorname{div} \sigma). \tag{3.2}$$

By a standard argument (cf. [1], [5]) (2.7) and Lemma 3.1 imply the existence of $(\tau, v) \in H_h \times V_h$ satisfying

$$\|\tau\|_0 + \|v\|_0 \leqslant C, \tag{3.3a}$$

$$\mathfrak{B}(\sigma_h - \tilde{\sigma}, u_h - \tilde{u}; \tau, v) \geqslant \|\sigma_h - \tilde{\sigma}\|_0 + \|u_h - \tilde{u}\|_0. \tag{3.3b}$$

Combining (3.3b) with (3.1), we see that

$$\|\sigma_h - \tilde{\sigma}\|_0 + \|u_h - \tilde{u}\|_0 \leqslant |a(\sigma - \tilde{\sigma}, \tau)| + |(u - \tilde{u}, \operatorname{div}\tau)| + |(v, \operatorname{div}(\sigma - \tilde{\sigma}))|. \tag{3.4}$$

Using (3.3a) and standard approximation theory, the first term on the right side of (3.4) is estimated as

$$|a(\sigma - \tilde{\sigma}, \tau)| \leqslant C\|\sigma - \tilde{\sigma}\|_0 \|\tau\|_0 \leqslant C_1 h^2 |\sigma|_2 \leqslant C_2 h^2 \|u\|_3.$$

In the second term we apply Lemma 3.3, (3.3a) and Sobolev imbedding to obtain

$$|(u - \tilde{u}, \operatorname{div}\tau)| \leqslant Ch^{3/2}(|u|_3 + |u|_{2,\infty})\|\tau\|_0 \leqslant C_1 h^{3/2}\|u\|_{7/2}.$$

Finally, interpolating in Lemma 3.2 and using (3.3a), we have

$$|(\operatorname{div}\sigma - \tilde{\sigma}, v)| \leqslant Ch^{3/2}\|\sigma\|_{5/2} \leqslant C_1 h^{3/2}\|u\|_{7/2}.$$

Upon combining these inequalities with (3.4), using the triangle inequality and recalling the estimate for $\|\sigma - \tilde{\sigma}\|_0$, the asserted estimate follows. □

Remark. Using the triangle inequality and the standard estimate $\|u - \tilde{u}\|_0 \leqslant Ch|u|_1$, it follows from Theorem 3.1 that $\|u - u_h\|_0 \leqslant Ch\|u\|_{7/2}$. □

Remark. Without the extra assumption on the boundary condition we can only prove the estimate

$$\|\sigma - \sigma_h\|_0 + \|u - u_h\|_0 \leqslant Ch^{1/2}\|u\|_{7/2}$$

(see the remark following Lemma 3.3). □

Proof of Lemma 3.1. Assume first that $m_1 \leqslant 2$ in (2.5). Then, since $m_2 = a_2(h_2/h_1)m_1$, m_2 is bounded by a constant independent of h. Thus, also $\dim(H_h)$ and $\dim(V_h)$ are bounded by constants independent of h. By the equivalence of norms in a finite-dimensional space, the assertion then follows if one can show that

$$(\operatorname{div}\tau, v) = 0 \quad \forall v \in H_h \Rightarrow v = 0. \tag{3.5}$$

To see that this is valid, let $K \in \mathcal{C}^h$ be such that K has two sides on Γ_1, and choose $\tau \in H_h$ so that τ is nonzero only in K. It is easy to see that one can have $(\operatorname{div}\tau, v) = 0$ for all such τ ony if $v = 0$ on K. Repeating this argument, (3.5) follows easily. We omit the details.

The case $m_2 \leqslant 2$ can be handled as above, so let us assume that $m_1, m_2 \geqslant 3$. Let $r_i \geqslant 0$ be the largest integer such that $3 \leqslant m_i - 3r_i \leqslant 5$, $i = 1, 2$, and let $\tilde{\mathcal{C}}^h$ be a coarser subdivision of Ω into rectangles $\tilde{K}_{\nu\mu}$, $\nu = 1, \ldots, r_1 + 1$, $\mu = 1, \ldots, r_2 + 1$, where

$$\tilde{K}_{\nu\mu} = \{x \in \mathbf{R}^2 : d_{1,\nu-1} < x_1 < d_{1\nu},\ d_{2,\mu-1} < x_2 < d_{2\mu}\},$$

with

$$d_{i\nu} = \begin{cases} 3\nu h_i, & 0 \leqslant \nu \leqslant r_i, \\ a_i, & \nu = r_i + 1, \end{cases} \quad i = 1, 2.$$

Thus, each $\tilde{K} \in \tilde{\mathcal{C}}^h$ consists of $l_1 \times l_2$ rectangles of $\mathcal{C}^h$ with $3 \leqslant l_1, l_2 \leqslant 5$.

For each $\tilde{K} \in \tilde{\mathcal{C}}^h$, we define the following finite-dimensional spaces

$$U = U(\tilde{K}) = \{\tau_{|\tilde{K}} : \tau \in H_h, \tau = 0 \text{ on } \Omega \setminus \tilde{K}\},$$

$$N = \left\{v \in V_{h|\tilde{K}} : \int_{\tilde{K}} \operatorname{div} \tau \cdot v \, dx = 0, \forall \tau \in U\right\},$$

$$W = \left\{w \in V_{h|\tilde{K}} : \int_{\tilde{K}} w \cdot v \, dx = 0, \forall v \in N\right\}.$$

Since $\tilde{K}$ consists of at most 5×5 rectangles of $\mathcal{C}^h$, we have $\dim(U) \leqslant 27$ and $\dim(W) \leqslant 50$. Therefore, using a scaling argument and the equivalence of norms in a finite-dimensional space, we conclude the existence of a positive constant C independent of $\tilde{K}$ such that

$$\sup_{\tau \in U(\tilde{K})} \frac{\int_{\tilde{K}} \operatorname{div} \tau \cdot v \, dx}{\|\tau\|_{0,\tilde{K}}} \geqslant Ch^{-1}\|v\|_{0,\tilde{K}} \quad \forall v \in W(\tilde{K}). \tag{3.6}$$

Now let $v \in V_h$ be given, and write

$$v = v_0 + v_1, \quad (v_0)|_{\tilde{K}} \in W(\tilde{K}), \quad (v_1)|_{\tilde{K}} \in N(\tilde{K}), \quad \tilde{K} \in \tilde{\mathcal{C}}^h. \tag{3.7}$$

By (3.6), there exists for any $\tilde{K} \in \tilde{\mathcal{C}}^h$ a function $\tau_{\tilde{K}} \in U(\tilde{K})$ such that, for some constant C,

$$\|\tau_{\tilde{K}}\|_{0,\tilde{K}} \leqslant Ch^{-1}\|v_0\|_{0,\tilde{K}}, \qquad \int_{\tilde{K}} \operatorname{div} \tau_{\tilde{K}} \cdot v_0 \, dx \geqslant h^{-2}\|v_0\|^2_{0,\tilde{K}}.$$

Let τ_0 be defined on Ω so that

$$\tau_0(x) = \tau_{\tilde{K}}(x), \qquad x \in \tilde{K} \in \tilde{\mathcal{C}}^h.$$

Then $\tau_0 \in H_h$ (since $\tau_{\tilde{K}} = 0$ on $\partial\tilde{K}$). Moreover, $\|\tau_0\|_0 \leqslant Ch^{-1}\|v_0\|_0$, and

$$(\operatorname{div} \tau_0, v) = \sum_{\tilde{K} \in \tilde{\mathcal{C}}^h} \int_{\tilde{K}} \operatorname{div} \tau_0 \cdot v \, dx = \sum_{\tilde{K} \in \tilde{\mathcal{C}}^h} \int_{\tilde{K}} \operatorname{div} \tau_0 \cdot v_0 \, dx \geqslant h^{-2}\|v_0\|_0^2.$$

Let us assume for a while that there also exists $\tau_1 \in H_h$ satisfying

$$\|\tau_1\|_0 \leqslant C\|v_1\|_0, \tag{3.8a}$$

$$(\operatorname{div} \tau_1, v_1) \geqslant \|v_1\|_0^2. \tag{3.8b}$$

Then, setting $\tau = \tau_0 + \gamma\tau_1$, where $\gamma \in (0, 1]$ will be chosen below, we have $\tau \in H_h$ and

$$\|\tau\|_0 \leqslant C\big(h^{-2}\|v_0\|_0^2 + \|v_1\|_0^2\big)^{1/2}. \tag{3.9}$$

Moreover, using the inverse inequality

$$\|\operatorname{div} \tau_1\|_0 \leqslant Ch^{-1}\|\tau_1\|_0$$

together with the above inequalities, we have

$$(\operatorname{div} \tau, v) \geqslant h^{-2}\|v_0\|_0^2 + \gamma\|v_1\|_0^2 + \gamma(\operatorname{div} \tau_1, v_0) \geqslant h^{-2}(1 - C\gamma)\|v_0\|_0^2 + \tfrac{\gamma}{2}\|v_1\|_0^2.$$

Thus, choosing $\gamma = \min\{1, 1/2C\}$, we have

$$(\operatorname{div} \tau, v) \geqslant C\big(h^{-2}\|v_0\|_0^2 + \|v_1\|_0^2\big).$$

Together with (3.9) this proves the assertion of the lemma, so the remaining task is to construct $\tau_1 \in H_h$ which satisfies (3.8).

We begin the construction of τ_1 by characterizing the spaces $N(\tilde{K})$. First, if $\tilde{K}$ is a macroelement consisting of $2 \times 3 = 6$ rectangles of $\mathcal{C}^h$, one has $\dim(U(\tilde{K})) = 6$ and $\dim(V_{H_{\tilde{K}}}) = 12$. Hence, one expects $N(\tilde{K})$ to be six-dimensional in this case, and this can indeed be verified by a straightforward computation. Let ν and μ be such that $\tilde{K}$ contains the rectangles $K_{\nu+i,\mu+j} \in \mathcal{C}^h$ for $i = 1,2$ and $j = 1,2,3$. Then we may choose for the basis of $N(\tilde{K})$ the set $\{\xi_1,\dots,\xi_6\}$, where the functions ξ_k are defined on the subrectangles $K_{\nu+i,\mu+j} \subset \tilde{K}$ as

$$
\begin{aligned}
(3.10)\qquad & \xi_1(x) = (1,0), \quad \xi_2(x) = (0,1), \quad \xi_3(x) = (i-1, 1-j), \\
& \xi_4(x) = \big((-1)^{i+j}, 0\big), \quad \xi_5(x) = \big(0, (-1)^{i+j}\big), \\
& \xi_6(x) = \big((1-j)(-1)^{i+j}, (i-1)(-1)^{i+j}\big), \\
& \qquad\qquad x \in K_{\nu+i,\mu+j}, \quad i, j \geq 1, \; K_{\nu+i,\mu+j} \subset \tilde{K}.
\end{aligned}
$$

We omit the details of showing that if $\tilde{K}$ is any macroelement consisting of $l_1 \times l_2$ rectangles of $\mathcal{C}^h$, the space $N(\tilde{K})$ is always six-dimensional and is spanned by the functions ξ_k defined by (3.10), provided that $\min\{l_1, l_2\} \geq 2$ and $\max\{l_1, l_2\} \geq 3$. (This can be shown by splitting $\tilde{K}$ into smaller subrectangles and using a construction similar to that given below.) In particular, this is the case for any $\tilde{K} \in \tilde{\mathcal{C}}^h$. The values of the basis functions $\xi_k = (\xi_{k1}, \xi_{k2})$ on the subrectangles of $\tilde{K}$ are shown in Figure 1 in the case where $\tilde{K}$ consists of 3×3 subrectangles.

k	1			2			3		
ξ_{k1}	1	1	1	0	0	0	2	2	2
	1	1	1	0	0	0	1	1	1
	1	1	1	0	0	0	0	0	0
ξ_{k2}	0	0	0	1	1	1	0	-1	-2
	0	0	0	1	1	1	0	-1	-2
	0	0	0	1	1	1	0	-1	-2

k	4			5			6		
ξ_{k1}	1	-1	1	0	0	0	0	1	-2
	-1	1	-1	0	0	0	0	-1	2
	1	-1	1	0	0	0	0	1	-2
ξ_{k2}	0	0	0	1	-1	1	2	-2	2
	0	0	0	-1	1	-1	-1	1	-1
	0	0	0	1	-1	1	0	0	0

FIGURE 1

The basis of $N(\tilde{K})$

Note that the first three functions represent the physical degrees of freedom in $N(\tilde{K})$, i.e., the rigid translations along the coordinate axes ($k = 1, 2$) and rotation ($k = 3$). The remaining degrees of freedom represent purely numerical "zero energy modes" (cf. [10] for similar modes in other mixed methods).

It follows from the above considerations that the function v_1 in (3.7) can be written as

$$v_1 = \sum_{\nu=1}^{r_1+1} \sum_{\mu=1}^{r_2+1} \sum_{k=1}^{6} \alpha_{\nu\mu k} \xi_{\nu\mu k}, \qquad \alpha_{\nu\mu k} \in \mathbf{R},$$

where $\xi_{\nu\mu k} = \xi_k(\tilde{K}_{\nu\mu})$, $\tilde{K}_{\nu\mu} \in \tilde{\mathcal{C}}^h$. Consider now a given $\tilde{K}_{\nu\mu} \in \tilde{\mathcal{C}}^h$, and let A, B, C, $D \in \mathfrak{N}$ be nodes of the finite element grid located on the sides of $\tilde{K}_{\nu\mu}$ as in Figure 2.

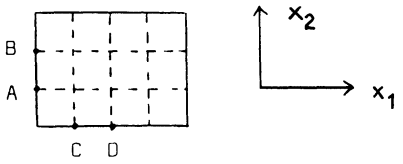

FIGURE 2

Let $\tau \in H_h$ be such that $\tau_{12} = 0$ and τ_{11} and τ_{22} vanish at all nodes except at A, B, C and D. Then if none of these four nodes is on the boundary Γ, we find that

$$\begin{aligned}(3.11)\qquad (\operatorname{div}\tau, v_1) = {} & h_2(\tau_{11}(A) + \tau_{11}(B))(\alpha_{\nu-1,\mu 1} - \alpha_{\nu\mu 1}) \\ & + \tfrac{1}{2}h_2(\tau_{11}(A) + 3\tau_{11}(B))(\alpha_{\nu-1,\mu 3} - \alpha_{\nu\mu 3}) \\ & + h_2(\tau_{22}(A) - \tau_{22}(B))(\alpha_{\nu-1,\mu 5} + \alpha_{\nu\mu 5}) \\ & + \tfrac{1}{2}h_2(\tau_{22}(A) - 3\tau_{22}(B))(\alpha_{\nu-1,\mu 6} + \alpha_{\nu\mu 6}) \\ & + h_1(\tau_{22}(C) + \tau_{22}(D))(\alpha_{\nu,\mu-1,2} - \alpha_{\nu\mu 2}) \\ & + \tfrac{1}{2}h_1(\tau_{22}(C) + 3\tau_{22}(D))(\alpha_{\nu\mu 3} - \alpha_{\nu,\mu-1,3}) \\ & + h_1(-\tau_{11}(C) + \tau_{11}(D))(\alpha_{\nu,\mu-1,4} + \alpha_{\nu\mu 4}) \\ & + \tfrac{1}{2}h_1(\tau_{11}(C) - 3\tau_{11}(D))(\alpha_{\nu,\mu-1,6} + \alpha_{\nu\mu 6}).\end{aligned}$$

By our assumption on the boundary conditions, we may assume that Γ_1 contains the sides of Ω at $x_1 = 0$ and $x_2 = 0$. Then the values of τ_{ij} at the nodes A, B, C, D are free parameters in H_h, and hence (3.11) holds even if some of the nodes are on the boundary if we set $\alpha_{0\mu k} = \alpha_{\nu 0 k} = 0$.

Let us now choose $\tilde{\tau} \in H_h$ in such a way that if A, B, C, D are nodes located on the boundary of some $\tilde{K}_{\nu\mu} \in \tilde{\mathcal{C}}^h$ as in Figure 2, then

$$\begin{aligned}\tilde{\tau}_{11}(A) + \tilde{\tau}_{11}(B) &= h^{-1}(\alpha_{\nu-1,\mu 1} - \alpha_{\nu\mu 1}), \\ \tilde{\tau}_{11}(A) + 3\tilde{\tau}_{11}(B) &= h^{-1}(\alpha_{\nu-1,\mu 3} - \alpha_{\nu\mu 3}), \\ \tilde{\tau}_{22}(A) - \tilde{\tau}_{22}(B) &= h^{-1}(\alpha_{\nu-1,\mu 5} + \alpha_{\nu\mu 5}), \\ \tilde{\tau}_{22}(A) - 3\tilde{\tau}_{22}(B) &= h^{-1}(\alpha_{\nu-1,\mu 6} + \alpha_{\nu\mu 6}), \\ \tilde{\tau}_{22}(C) + \tilde{\tau}_{22}(D) &= h^{-1}(\alpha_{\nu,\mu-1,2} - \alpha_{\nu\mu 2}),\end{aligned}$$

$$\tilde{\tau}_{22}(C) + 3\tilde{\tau}_{22}(D) = h^{-1}(\alpha_{\nu\mu 3} - \alpha_{\nu,\mu-1,3}),$$
$$-\tilde{\tau}_{11}(C) + \tilde{\tau}_{11}(D) = h^{-1}(\alpha_{\nu,\mu-1,4} + \alpha_{\nu\mu 4}),$$
$$\tilde{\tau}_{11}(C) - 3\tilde{\tau}_{11}(D) = h^{-1}(\alpha_{\nu,\mu-1,6} + \alpha_{\nu\mu 6}).$$

At the remaining nodes we set $\tilde{\tau} = 0$. Then $\tilde{\tau}$ is uniquely determined and $\|\tilde{\tau}\|_0 \leq C\|v_1\|_h$, where

$$\|v_1\|_h^2 = \sum_{\nu=1}^{r_1+1}\sum_{\mu=1}^{r_2+1}\Bigg[\sum_{k=2,3}(\alpha_{\nu\mu k} - \alpha_{\nu,\mu-1,k})^2 + \sum_{k=4,6}(\alpha_{\nu\mu k} + \alpha_{\nu,\mu-1,k})^2 + \sum_{k=1,3}(\alpha_{\nu\mu k} - \alpha_{\nu-1,\mu k})^2 + \sum_{k=5,6}(\alpha_{\nu\mu k} + \alpha_{\nu-1,\mu k})^2\Bigg],$$
$$(\alpha_{0\mu k} = \alpha_{\nu 0 k} = 0).$$

Moreover, it follows from (3.11) that $(\operatorname{div}\tilde{\tau}, v_1) \geq C\|v_1\|_h^2$. Using the inequality

$$\alpha_1^2 + \sum_{i=2}^{n}(\alpha_i \pm \alpha_{i-1})^2 \geq n^{-2}\sum_{i=1}^{n}\alpha_i^2$$

and noting that $r_i \leq Ch^{-1}$, $i = 1, 2$, we see that $\|v_1\|_h \geq C\|v\|_0$. Combining this with the above inequalities we see that $\tau_1 = (\|v_1\|_0/C\|v_1\|_h)\tilde{\tau}$ satisfies (3.8) for C sufficiently large, and so the proof is complete. □

Proof of Lemma 3.2. The proof is based on the following result which can be verified by direct computation; cf. also [10].

LEMMA 3.4. *If* $\sigma \in [P_2(K)]^4$, $K \in \mathcal{C}^h$ *and if* $\tilde{\sigma}$ *is the bilinear interpolant of* σ *on* K, *then*

$$\int_K \operatorname{div}(\sigma - \tilde{\sigma})\,dx = 0. \quad \square$$

Using Lemma 3.4, it follows from the Bramble-Hilbert lemma [4] that, with $\tilde{\sigma}$ as in Lemma 3.2,

$$\left|\int_K \operatorname{div}(\sigma - \tilde{\sigma})v\,dx\right| \leq Ch^k|\sigma|_{k+1,K}\|v\|_{0,K}, \qquad k = 2, 3,\ v \in V_h,\ K \in \mathcal{C}^h.$$

By summing over $K \in \mathcal{C}^h$, the asserted estimate follows. □

Proof of Lemma 3.3. We use the following easy-to-prove result. Note that this is only valid for a uniform mesh.

LEMMA 3.5. *Let* $\tau \in H_h$ *be such that* τ *vanishes at all nodes except at a given node* $P \in \mathcal{N}$. *Denote by* T *the support of* τ, *let* $u \in [L_2(T)]^2$ *and let* $\tilde{u}$ *be the* L_2*-projection of* u *into the subspace* $V_{h|T}$. *Then*

$$\int_T (u - \tilde{u}) \cdot \operatorname{div}\tau\,dx = 0,$$

if either (a) P *is an interior point of* Ω *and* $u \in [P_2(T)]^2$ *or* (b) P *is a boundary point but not a vertex of* Ω *and* $u \in [P_1(T)]^2$. □

Now let $\tilde{u}$ be as in Lemma 3.3, let $P \in \mathfrak{N}$ be a node interior to Ω and let τ satisfy the assumptions of Lemma 3.5. Then Lemma 3.5 and the Bramble-Hilbert lemma [4] imply that

$$\left|\int_T (u - \tilde{u}) \operatorname{div} \tau \, dx\right| \leq Ch^2 |u|_{3,T} \|\tau\|_{0,T}. \tag{3.12}$$

If P is a boundary node but not a vertex of Ω, we have by the same argument

$$\left|\int_T (u - \tilde{u}) \cdot \operatorname{div} \tau \, dt\right| \leq Ch^2 |u|_{2,\infty,T} \|\tau\|_{0,T}. \tag{3.13}$$

Finally if P is a vertex of Ω, we have three possibilities:

(i) The neighborhood of P is contained in Γ_2. In this case the boundary conditions imply that $\tau(P) = 0$ for all $\tau \in H_h$.

(ii) The neighborhood of P is contained in Γ_1. In this case we have $u = 0$ on the two sides meeting at P, so by Taylor expansion

$$|u(x)| \leq Ch^2 |u|_{2,\infty,T}, \qquad x \in T.$$

Using this we have

$$\left|\int_T (u - \tilde{u}) \cdot \operatorname{div} \tau \, dx\right| \leq C\|u\|_{0,\infty,T} \|\tau\|_{1,1,T} \leq C_1 h^2 |u|_{2,\infty,T} \|\tau\|_{0,T},$$

so (3.13) is valid also in this case.

(iii) The boundary condition changes at P. Assume, for example, that $P = (0,0)$ and that the sides at $x_i = 0$ are contained in Γ_i, $i = 1,2$. If $u \in [W^{2,\infty}(\Omega)]^2$, it follows from (2.1) and from the boundary conditions that $u(P) = 0$, $\partial u_i(P)/\partial x_2 = 0$, $i = 1,2$ and $\partial u_2(P)/\partial x_1 = 0$. Therefore, u admits the Taylor expansion

$$u(x) = (0, \alpha x_2) + v(x), \qquad x \in T,$$

where $\alpha = \partial u_2(P)/\partial x_2$ and $\|v\|_{0,\infty,T} \leq Ch^2 |u|_{2,\infty,T}$. On the other hand, $\tau_{12}(P) = \tau_{22}(P) = 0$ by the boundary conditions, so that $\operatorname{div} \tau = (\partial \tau_{11}/\partial x_1, 0)$. Combining these observations it is easy to see that (3.13) is again valid.

Now let $\tau \in H_h$ be arbitrary and let τ_P, $P \in \mathfrak{N}$, be such that $\tau_P(P) = \tau(P)$ and $\tau_P(P') = 0$ for $P' \in \mathfrak{N}$, $P' \neq P$. Denote the support of τ_P by T_P. Then, by (3.12) and (3.13),

$$\begin{aligned}
|(u - \tilde{u}, \operatorname{div} \tau)| &= \left|\sum_{P \in \mathfrak{N}} (u - \tilde{u}, \operatorname{div} \tau_P)\right| \\
&\leq Ch^2 \sum_{\substack{P \in \mathfrak{N} \\ P \notin \Gamma}} |u|_{3,T_P} \|\tau_P\|_0 + Ch^2 \sum_{\substack{P \in \mathfrak{N} \\ P \in \Gamma}} |u|_{2,\infty,T_P} \|\tau_P\|_0 \\
&\leq C_1 h^2 \left\{ \sum_{\substack{P \in \mathfrak{N} \\ P \notin \Gamma}} |u|^2_{3,T_P} \right\}^{1/2} \|\tau\|_0 + C_1 h^2 \left\{ \sum_{\substack{P \in \mathfrak{N} \\ P \in \Gamma}} |u|^2_{2,\infty,T_P} \right\}^{1/2} \|\tau\|_0 \\
&\leq C_2 \left(h^2 |u|_3 + h^{3/2} |u|_{2,\infty}\right) \|\tau\|_0,
\end{aligned}$$

which proves the assertion. □

4. Method II. In this section we consider a class of methods based on (2.3), where the subspace H_h is defined in terms of a composite rectangular element described in [9]. Let $K \in \mathcal{C}^h$, and let K be subdivided into four triangles T_i as in Figure 3.

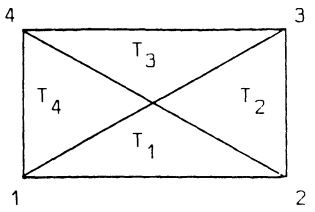

FIGURE 3
Subdivision of $K \in \mathcal{C}^h$

We introduce the spaces

$$R_K = \{\tau \in Y(K): \tau \in [P_1(T)]^4, \tau \cdot n = 0 \text{ on the diagonal 1–3 and } \tau = 0 \text{ on } K \setminus T\},$$

$$R'_K = \{\tau \in Y(K): \tau \in [P_1(T')]^4, \tau \cdot n' = 0 \text{ on the diagonal 2–4 and } \tau = 0 \text{ on } K \setminus T'\},$$

$$S_K = Y(K) \cap [P_1(K)]^4 \quad \text{and} \quad G_K = S_K \oplus R_K \oplus R'_K,$$

where n and n' are normals to the diagonals 1–3 and 2–4, respectively, $T = T_1 \cup T_2$ and $T' = T_1 \cup T_4$.

Setting $V_K = [P_1(K)]^2$ and $H_K = G_K$, $K \in \mathcal{C}^h$ and defining the spaces V_h and H_h in (2.3) as

$$V_h = \{v \in [L_2(\Omega)]^2: v_{|K} \in V_K, K \in \mathcal{C}^h\}, \tag{4.1a}$$

$$H_h = \{\tau \in H(\Omega): \tau_{|K} \in H_K, K \in \mathcal{C}^h\}, \tag{4.1b}$$

we obtain the method analyzed in [9]. Here we have (see below)

$$\dim(H_h) = 11 m_1 m_2 + O(h^{-1}),$$

i.e., there are about 11 degrees of freedom per node in H_h. In what follows we consider methods where V_h and H_h are chosen as proper subspaces of those defined by (4.1).

The first modification we consider is as follows. Let V_k be the space of rigid displacements on K, i.e.,

$$V_K = \{v(x) = (\alpha_1 + \alpha_3 x_2, \alpha_2 - \alpha_3 x_1), x \in K, \alpha_i \in \mathbf{R}\}.$$

Further, let

$$N_K = \left\{v \in [P_1(K)]^2: \int_K v \cdot w \, dx = 0 \ \forall v \in V_K\right\}$$

and

$$H_K = \left\{\tau \in G_K: \int_K \operatorname{div} \tau \cdot v \, dx = 0 \ \forall v \in N_K\right\}, \tag{4.2}$$

where G_K is as above. With this new definition of V_K and H_K, define now V_h and H_h by (4.1).

To see how H_K in (4.2) is constructed, we recall from [9] that any $\tau \in G_K$ is defined uniquely by the following 19 degrees of freedom:

(i) the values of $\tau \cdot n$ at two points on each side of K,

(ii) $\int_K \tau_{ij}\,dx$, $i, j = 1, 2$.

Let $Q = (q_1, q_2)$ be the midpoint of K. Then if $\tau \in G_K$, we have $\tau \in H_K$ if and only if

$$\int_K \operatorname{div} \tau \cdot v_{ij}\,dx = 0, \qquad i, j = 1, 2, \tag{4.3}$$

where $v_{ij} \in N_K$ are defined by

$$\begin{aligned} v_{11}(x) &= (x_1 - q_1, 0), \\ v_{12}(x) &= v_{21}(x) = \tfrac{1}{2}(x_2 - q_2, x_1 - q_1), \\ v_{22}(x) &= (0, x_2 - q_2), \qquad x \in K. \end{aligned}$$

Applying (2.6) and noting that $\tau \cdot \varepsilon(v_{ij}) = \tau_{ij}$, we see that (4.3) is equivalent to

$$\int_K \tau_{ij}\,dx = \int_{\partial K} (\tau \cdot n) \cdot v_{ij}\,ds, \qquad i, j = 1, 2. \tag{4.4}$$

Thus, H_K is constructed by eliminating from G_K the inner degrees of freedom (ii) using (4.4).

If H_K is defined by (4.2), the inclusion $H_K \supset S_K$ obviously remains valid, since $\int_K \operatorname{div} \tau \cdot v\,dx = 0$ if $v \in N_K$ and $\tau \in [P_1(K)]^4$. Therefore the space H_K has the same approximation properties as G_K [9]: If $\sigma \in Y(K) \cap [H^2(K)]^4$ and $\tilde{\sigma} \in H_K$ is the interpolant of σ, i.e., $\tilde{\sigma} \cdot n = \sigma \cdot n$ at two points on each side of K, then

$$\|\sigma - \tilde{\sigma}\|_{0,K} \leqslant Ch^2 |\sigma|_{2,K}. \tag{4.5}$$

In the sequel we choose for the degrees of freedom of H_K the limiting values of $(\tau \cdot n)(x)$ on each side of K as x approaches a vertex of K. Thus, there are eight degrees of freedom in H_h associated to each interior node. In order to further reduce the space H_h, we need some notation. For $\tau \in H_h$ and P an interior node, let $A_P(\tau)$ be a 2×4 matrix with coefficients

$$a_{ij} = \lim_{\substack{P' \to P \\ P' \in l_j}} (\tau \cdot n_j)_i, \qquad i = 1, 2, \quad j = 1, \ldots, 4,$$

where l_j, $j = 1, \ldots, 4$, denote the semi-infinite mesh lines starting at P, numbered as in Figure 4.

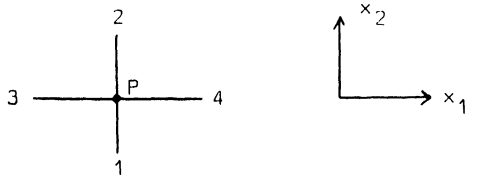

FIGURE 4

Local numbering of mesh lines

If P is located on Γ, we think the mesh to be extended outside Ω and set $a_{ij} = 0$, $i = 1, 2$, if $l_j \cap \Omega = \varnothing$.

Let H^P denote the subspace of H_h consisting of those functions that vanish at all nodes except at P. If P is an interior node, then $\dim(H^P) = 8$, and we may choose

for the basis of H^P the set $\{\tau_i^P, i = 1,\dots,8\}$ such that

$$A(\tau_1^P) = \begin{pmatrix} 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix}, \quad A(\tau_2^P) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 1 \end{pmatrix},$$
$$A(\tau_3^P) = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 1 & 0 & 0 \end{pmatrix}, \quad A(\tau_4^P) = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 0 & 0 & 0 \end{pmatrix},$$
$$A(\tau_5^P) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{pmatrix}, \quad A(\tau_6^P) = \begin{pmatrix} 0 & 0 & -1 & -1 \\ 1 & 1 & 0 & 0 \end{pmatrix},$$
$$A(\tau_7^P) = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 1 & -1 & 0 & 0 \end{pmatrix}, \quad A(\tau_8^P) = \begin{pmatrix} 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 \end{pmatrix}.$$

If P is a boundary node, we define τ_i^P on Ω as above.

With the above notation, the space H_h can be defined as

$$H_h = \left\{ \tau = \sum_{P \in \mathfrak{N}} \sum_{i=1}^{8} \alpha_i^P \tau_i^P,\ \alpha_i^P \in \mathbf{R},\ \tau \cdot n = 0 \text{ on } \Gamma_2 \right\}. \tag{4.6}$$

We note that if $\sigma \in H(\Omega)$ is continuous and $\tilde{\sigma} \in H_h$ is the interpolant of σ defined by

$$A_P(\tilde{\sigma}) = A_P(\sigma), \qquad P \in \mathfrak{N}, \tag{4.7}$$

then

$$\tilde{\sigma} = \sum_{P \in \mathfrak{N}} \left[\sigma_{11}(P)\tau_1^P + \sigma_{22}(P)\tau_2^P + \sigma_{12}^P(\tau_3^P) \right]. \tag{4.8}$$

Thus, only the functions τ_i^P, $i = 1, 2, 3$, $P \in \mathfrak{N}$, are required in the interpolation. In particular, if $\sigma \in H(\Omega) \cap [H^2(\Omega)]^4$ and if $\tilde{\sigma}$ is defined by (4.8), then (4.5) holds and so, by summing over $K \in \mathcal{C}^h$ in (4.5),

$$\|\sigma - \tilde{\sigma}\|_0 \leqslant Ch^2 |\sigma|_2. \tag{4.9}$$

From now on we denote by $H_{h,1}$ the subspace defined by (4.6) and by $H_{h,0}$ the space

$$H_{h,0} = \left\{ \tau = \sum_{P \in \mathfrak{N}} \sum_{i=1}^{3} \alpha_i^P \tau_i^P,\ \alpha_i^P \in \mathbf{R},\ \tau \cdot n = 0 \text{ on } \Gamma_2 \right\}. \tag{4.10}$$

We shall see below that with $H_h = H_{h,1}$ and V_h as above, we have the quasioptimal estimate

$$\|\sigma - \sigma_h\|_0 \leqslant Ch^2 |\sigma|_2, \tag{4.11}$$

where σ and σ_h are solutions to (2.2) and (2.3), respectively. Our aim is now to construct smaller subspaces H_h, with $H_{h,0} \subset H_h \subset H_{h,1}$, so that (4.11) still holds.

Let us first formulate a stability criterion for the method (2.3), assuming that V_h is defined by (4.1a) and $H_{h,0} \subset H_h \subset H_{h,1}$. We follow here the lines of Babuška [1] and Brezzi [5] (cf. also [2], [10]). Let $\tau \in H(K)$, $K \in \mathcal{C}^h$, and let $v \in V^h$. Then, by (2.6) and since $\varepsilon(v) = 0$ on K,

$$\int_K \operatorname{div} \tau \cdot v \, dx = \int_{\partial K} (\tau \cdot n) \cdot v \, ds.$$

Summing this over all $K \in \mathcal{C}^h$, we obtain

$$(4.12) \qquad (\operatorname{div}\tau, v) = \int_{\Gamma_h} (\tau \cdot n) \cdot (v^- - v^+)\, ds + \int_{\Gamma_1} (\tau \cdot n) \cdot v\, ds, \qquad \tau \in H(\Omega), v \in V_h,$$

where Γ_h denotes the union of mesh lines in the interior of Ω, n is a normal to Γ_h or an exterior normal to Γ and

$$v^{\pm}(x) = \lim_{\varepsilon \to 0^{\pm}} v(x + \varepsilon n), \qquad x \in \Gamma_h.$$

By (4.12),

$$(4.13) \qquad |(\operatorname{div}\tau, v)| \leqslant |\tau|_{0,h} |v|_{1,h}, \qquad \tau \in H(\Omega), v \in V_h,$$

where

$$|\tau|_{0,h}^2 = h \int_{\Gamma_h \cup \Gamma_1} |\tau \cdot n|^2\, ds,$$

$$|v|_{1,h}^2 = h^{-1} \int_{\Gamma_h} |v^+ - v^-|^2\, ds + h^{-1} \int_{\Gamma_1} |v|^2\, ds.$$

By the following lemma, the seminorms $|\cdot|_{0,h}$ and $|\cdot|_{1,h}$ are norms on H_h and V_h, respectively.

LEMMA 4.1. *There is a constant C such that for all $\tau \in H_h$ and $v \in V_h$*

$$C^{-1}\|\tau\|_0 \leqslant |\tau|_{0,h} \leqslant C\|\tau\|_0,$$

and

$$|v|_{1,h} \geqslant \begin{cases} C^{-1}\|v\|_0, \\ C^{-1}(1 + |\log h|)^{-1/2}\|v\|_{0,\infty}. \end{cases} \quad \square$$

Proof. The equivalence of $|\cdot|_{0,h}$ and $\|\cdot\|_0$ on H_h follows easily from local scaling arguments (cf. also [2]), so let us only prove the lower bounds for $|\cdot|_{1,h}$. Let $\tilde{\mathcal{C}}^h$ be another rectangular subdivision of Ω such that the interior nodes in $\tilde{\mathcal{C}}^h$ are at the midpoints of each $K \in \mathcal{C}^h$. For $v \in V_h$ given, let $\tilde{v}$ be a smoothening of v defined in terms of the bicubic Bogner-Fox-Schmidt element [6] as follows: For any $\tilde{K} \in \tilde{\mathcal{C}}^h$, $\tilde{v}_{|\tilde{K}}$ is a bicubic polynomial, and if $\tilde{P}$ is a vertex of $\tilde{K} \in \tilde{\mathcal{C}}^h$, then

$$\tilde{v}(P) = v(P), \qquad \frac{\partial}{\partial x_i}\tilde{v}(P) = \frac{\partial}{\partial x_i} v(P), \qquad i = 1, 2,$$

$$\frac{\partial^2}{\partial x_1 \partial x_2}\tilde{v}(P) = \frac{\partial^2}{\partial x_1 \partial x_2} v(P).$$

As is well known (cf. [6]), $\tilde{v}$ is uniquely determined and $\tilde{v} \in H^1(\Omega)$. Now if $\tilde{K} \in \tilde{\mathcal{C}}^h$, it is easy to see that we can have

$$\int_{\Gamma_h \cap \tilde{K}} |v^+ - v^-|^2\, ds = 0$$

if and only if $\|\varepsilon(\tilde{v})\|_{0,\tilde{K}} = 0$, i.e., if and only if $v_{|\tilde{K}} = \tilde{v}_{|\tilde{K}}$ is a rigid displacement on $\tilde{K}$. Therefore, by a scaling argument and by the equivalence of norms in a finite-dimensional space, there is a positive constant C independent of $\tilde{K}$ such that

$$h^{-1}\int_{\Gamma_h \cap \tilde{K}} |v^+ - v^-|^2 \, ds \geqslant C\|\varepsilon(\tilde{v})\|^2_{0,\tilde{K}}.$$

Summing this over all $\tilde{K} \in \tilde{\mathcal{C}}^h$, we obtain

$$h^{-1}\int_{\Gamma_h} |v^+ - v^-|^2 \, ds \geqslant C\|\varepsilon(\tilde{v})\|^2_0.$$

Noting also that, by the construction of $\tilde{v}$, $\|v\|^2_{0,\Gamma_1} \geqslant C\|\tilde{v}\|^2_{0,\Gamma_1}$, we obtain the inequality

$$|v|^2_{1,h} \geqslant C\big(\|\varepsilon(\tilde{v})\|^2_0 + \|\tilde{v}\|^2_{0,\Gamma_1}\big).$$

Further, since Γ_1 contains at least one side of Ω, we have by Korn's inequality (cf. [7], [13])

$$\|\varepsilon(\tilde{v})\|^2_0 + \|\tilde{v}\|^2_{0,\Gamma_1} \geqslant C\|\tilde{v}\|^2_1.$$

We finally note that, by the construction of $\tilde{v}$, $\|\tilde{v}\|_0 \geqslant C\|v\|_0$. Upon combining the last three inequalities, the first part of the assertion follows.

For the second part of the assertion, we need the additional estimates

$$\|v\|_{0,\infty} \leqslant C\|\tilde{v}\|_{0,\infty} \leqslant C_1(1 + |\log h|)^{1/2}\|\tilde{v}\|_1. \tag{4.14}$$

Combining (4.14) with the above inequalities, the proof is completed, so it remains to prove (4.14).

The first inequality in (4.14) follows easily from the construction of $\tilde{v}$, so let us only prove the second part of (4.14). Let w be an extension of $\tilde{v}$ to $\mathbf{R}^2$ such that w vanishes outside some disc S of finite radius and satisfies

$$\|w\|_{H^1(\mathbf{R}^2)} \leqslant C\|\tilde{v}\|_1.$$

Let $x \in \mathfrak{N}$, $x \notin \partial\Omega$. Then $\tilde{v}(x)$ may be written, by Green's formula, as

$$\tilde{v}(x) = \frac{1}{2\pi}\int_S \nabla(\log|x-y|) \cdot \nabla w(y)\, dy.$$

Let S_h be a disc centered at x and of radius $\rho = \min\{h_1, h_2\}$. Then $S_h \subset \Omega$, and we have by a scaling argument

$$\left|\int_{S_h} \nabla(\log|x-y|) \cdot \nabla w(y)\, dy\right| \leqslant \int_{S_h} \frac{1}{|x-y|}|\nabla\tilde{v}(y)|dy \leqslant C|\tilde{v}|_{1,S_h}.$$

Since, on the other hand,

$$\left|\int_{S\setminus S_h} |\nabla \log(x-y) \cdot \nabla w(y)\, dy|\right|$$

$$\leqslant \left\{\int_{S\setminus S_h} \frac{1}{|x-y|^2}dy\right\}^{1/2} \|w\|_{H^1(\mathbf{R}^2)} \leqslant C(1 + |\log h|)^{1/2}\|\tilde{v}\|_1,$$

it follows that, for $x \in \mathfrak{N}$, $x \notin \partial\Omega$,

$$|\tilde{v}(x)| \leqslant C(1 + |\log h|)^{1/2}\|\tilde{v}\|_1. \tag{4.15}$$

Finally, if $x \in \bar{\Omega}$ is arbitrary, we find by a scaling argument that $|\tilde{v}(x)| \leqslant |v(y)| + C|\tilde{v}|_{1,K}$, where $K \in \mathcal{C}^h$ is such that $x \in \bar{K}$ and y is any vertex of K. Thus, (4.15) holds for any $x \in \bar{\Omega}$, and so the proof is complete. □

Using the above mesh-dependent norms, we can state a basic stability condition for method (2.3) as: There is a constant C such that

$$\sup_{\tau \in H_h} \frac{(\operatorname{div} \tau, v)}{|\tau|_{0,h}} \geqslant C|v|_{1,h}, \qquad v \in V_h. \tag{4.16}$$

Note that, by Lemma 4.1, (4.16) is a stronger stability inequality than that proved for Method I (Lemma 3.1).

Let us first prove

LEMMA 4.2. *If* $H_h = H_{h,1}$, *then* (4.16) *holds with* $C = 1$. □

Proof. If $H_h = H_{h,1}$, the trace space $\{\tau \cdot n: \tau \in H_h\}$ contains the trace space of V_h on $\Gamma_h \cup \Gamma_1$. Thus if $v \in V_h$ is given, there exists $\tau \in H_h$ such that $\tau \cdot n = v^- - v^+$ on Γ_h and $\tau \cdot n = v$ on Γ_1. Using (4.12) we see that $(\operatorname{div} \tau, v) = |\tau|_{0,h}|v|_{1,h}$, so the assertion follows. □

Remark. If one chooses $H_h = H_{h,0}$, it is easy to see that (4.16) can only be valid if C depends on h. In fact, it is not difficult to verify that, even if the space V_h is reduced to consist only of functions that are piecewise constant, (4.16) does not hold if $H_h = H_{h,0}$. Note that such a method has essentially the same properties as Method I above. □

Before showing examples where the stability assumption (4.16) holds with $H_h \neq H_{h,1}$, let us prove an error estimate for the method (2.3) assuming merely that $H_{h,1} \supset H_h \supset H_{h,0}$ and that (4.16) holds. We need the following nonstandard interpolant in V_h: if $u \in V$, define $\tilde{u} \in V_h$ by requiring

$$\int_K (u - \tilde{u}) \operatorname{div} \tau \, dx = 0, \qquad \tau \in H_K, K \in \mathcal{C}^h, \tag{4.17}$$

where H_K is defined as above.

LEMMA 4.3. *The interpolant* $\tilde{u}$ *is uniquely determined. Moreover, one has the estimates*

$$\|u - \tilde{u}\|_0 \leqslant Ch|u|_1 \quad \text{and} \quad |(u - \tilde{u})(Q_K)| \leqslant Ch^2|u|_{2,\infty,K}, \qquad K \in \mathcal{C}^h,$$

where Q_K *is the midpoint of* K. □

Proof. Let $W_K = \{\operatorname{div} \tau, \tau \in G_K\}$, where G_K is as in (4.2). In [9] it is shown that $W_K = \pi_K([P_1(K)]^2)$, where π_K denotes the orthogonal projection of $[L_2(K)]^2$ into the subspace consisting of functions v such that $v_{|T_i}$ is constant on each of the four subtriangles of K. By the definition of H_K we then have

$$\{\operatorname{div} \tau: \tau \in H_K\} = \pi_K(V_K).$$

Thus, we may set $\operatorname{div} \tau = \pi_K \tilde{u}$ in (4.17) to obtain

$$\int_K \tilde{u} \pi_K \tilde{u} \, dx = \int_K (\pi_K \tilde{u})^2 \, dx = \int_K u \pi_K \tilde{u} \, dx.$$

Using here the easy-to-prove inequality

$$\|\pi_K v\|_{0,K}^2 \geqslant \tfrac{2}{3}\|v\|_{0,K}^2, \qquad v \in V_K,$$

we see that u satisfies $\|\tilde{u}\|_{0,K} \leqslant \frac{3}{2}\|u\|_{0,K}$. So, $\tilde{u}$ is uniquely determined. Moreover, since $\tilde{u} = u$ if u is constant, we conclude from the Bramble-Hilbert lemma [4] that $\|u - \tilde{u}\|_{0,K} \leqslant Ch|u|_{1,K}$. This proves the first inequality in the lemma.

To prove the second inequality, note that if we write

$$\tilde{u}(x) = (\alpha_1, \alpha_2) + \alpha_3(x_2 - q_2, -x_1 + q_1) = \tilde{u}_0(x) + \tilde{u}_1(x), \qquad x \in K,$$

where $Q = (q_1, q_2)$ is the midpiont of K, then by (4.17) $\int_K (u - \tilde{u}_0)\,dx = 0$. Thus, $\tilde{u}_0$ is the L_2-projection of u into $[P_0(K)]^2$. Since $(u - \tilde{u})(Q) = (u - \tilde{u}_0)(Q) = 0$ if $u \in [P_1(K)]^2$, we obtain from the Bramble-Hilbert lemma the estimate

$$|(u - \tilde{u})(Q)| \leqslant Ch^2|u|_{2,\infty,K},$$

and the proof is complete. □

We are now ready to state the main convergence result.

THEOREM 4.1. *Let* (σ, u) *be the solution of* (2.2), *and let* (σ_h, u_h) *be the solution of* (2.3), *where* V_h *is defined by* (4.1a) *and* $H_h \supset H_{h,0}$ *is chosen so the stability condition* (4.16) *holds. Then we have the error estimates*

$$\|\sigma - \sigma_h\|_0 \leqslant Ch^2|\sigma|_2,$$

$$|(u - u_h)(Q_K)| \leqslant Ch^2(1 + |\log h|)^{1/2}|\sigma|_2 + Ch^2|u|_{2,\infty,K}, \qquad K \in \mathcal{C}^h,$$

where Q_K *is the midpoint of* K. □

Proof. Let $\tilde{\sigma} \in H_h$ be defined by (4.7) (or equivalently by (4.8)), and let $\tilde{u} \in V_h$ be defined by (4.17). Then by (2.7) and (4.16) and since $|\tau|_{0,h} \leqslant C\|\tau\|_0$ for $\tau \in H_h$, by Lemma 4.1, we conclude by the standard argument (cf. [1], [5]) that there exists $(\tau, v) \in H_h \times V_h$ satisfying

$$\|\tau\|_0 + |v|_{1,h} \leqslant C, \tag{4.18a}$$

$$\mathcal{B}(\sigma_h - \tilde{\sigma}, u_h - \tilde{u}; \tau, v) \geqslant \|\sigma_h - \tilde{\sigma}\|_0 + |u_h - \tilde{u}|_{1,h}, \tag{4.18b}$$

where $\mathcal{B}$ is as in (3.2). Using (4.17) and (4.18b) in the identity (3.1), we obtain

$$\|\sigma_h - \tilde{\sigma}\|_0 + |u_h - \tilde{u}|_{1,h} \leqslant |(\sigma - \tilde{\sigma}, \tau)| + |(\operatorname{div}(\sigma - \tilde{\sigma}), v)|.$$

Using (4.18a) and (4.9), we have

$$|(\sigma - \tilde{\sigma}, \tau)| \leqslant \|\sigma - \tilde{\sigma}\|_0\|\tau\|_0 \leqslant Ch^2|\sigma|_2.$$

Similarly using (4.13) and (4.18a),

$$|(\operatorname{div}(\sigma - \tilde{\sigma}), v)| \leqslant |\sigma - \tilde{\sigma}|_{0,h}|v|_{1,h} \leqslant C|\sigma - \tilde{\sigma}|_{0,h} \leqslant C_1h^2|\sigma|_2.$$

Here the interpolation estimate in the norm $|\cdot|_{0,h}$ is proved by standard techniques; cf. [2] for details of the argument.

Combining these estimates, using the triangle inequality and recalling (4.9), we obtain

$$\|\sigma - \sigma_h\|_0 + |u_h - \tilde{u}|_{1,h} \leqslant Ch^2|\sigma|_2. \tag{4.19}$$

This proves the asserted estimate for $\|\sigma - \sigma_h\|_0$. Finally, by (4.19), Lemma 4.1 and Lemma 4.3, if Q_K is the midpoint of $K \in \mathcal{C}^h$, then

$$\begin{aligned}|(u - u_h)(Q_K)| &\leq |(u_h - \tilde{u})(Q_K)| + |(u - \tilde{u})(Q_K)| \\ &\leq C(1 + |\log h|)^{1/2} |u_h - \tilde{u}|_{1,h} + Ch^2 |u|_{2,\infty,K} \\ &\leq Ch^2(1 + |\log h|)^{1/2} |\sigma|_2 + Ch^2 |u|_{2,\infty,K},\end{aligned}$$

which completes the proof. □

In the remaining part of this section we consider the practical problem of constructing H_h so that the stability condition (4.16) holds. We first formulate a general criterion which is sufficient (and probably also necessary) for the validity of (4.16). Let $\mathcal{E}^h$ be a collection of "macroelements", i.e., a collection of open rectangles $\tilde{K}$ such that if $K \in \mathcal{C}^h$ and $\tilde{K} \in \mathcal{E}^h$, then either $K \subset \tilde{K}$ or $K \cap \tilde{K} = \varnothing$. We associate to each $\tilde{K} \in \mathcal{E}^h$ the following subspaces:

$$U(\tilde{K}) = \left\{\tau_{|\tilde{K}} : \tau \in H_h, \tau = 0 \text{ in } \Omega \setminus \tilde{K}\right\}, \tag{4.20a}$$

$$N(\tilde{K}) = \left\{v_{|\tilde{K}} : v \in V_h, \int_{\tilde{K}} \operatorname{div} \tau \cdot v \, dx = 0 \ \forall \tau \in U(\tilde{K})\right\}. \tag{4.20b}$$

Let us now assume that there exists an integer $M \geq 1$ independent of h such that $\mathcal{E}^h$ satisfies the following hypotheses:

(i) Each $\tilde{K} \in \mathcal{E}^h$ contains at most M different rectangles $K \in \mathcal{C}^h$.

(ii) If l is a side of $K \in \mathcal{C}^h$ and l is not on Γ, then l is contained in at least one and not more than M different rectangles $\tilde{K} \in \mathcal{E}^h$.

(iii) For each $\tilde{K} \in \mathcal{E}^h$, $N(\tilde{K}) = \{0\}$ if $\partial\tilde{K} \cap \Gamma_1 \neq \varnothing$, otherwise $N(\tilde{K})$ is the space of rigid displacements of $\tilde{K}$, i.e., $\dim(N(\tilde{K})) = 3$.

We can now prove

LEMMA 4.4. *If there is a collection $\mathcal{E}^h$ which satisfies the above assumptions, then* (4.16) *holds.* □

Proof. For $\tilde{K} \in \mathcal{E}^h$ let

$$W(\tilde{K}) = \left\{w_{|\tilde{K}} : w \in V_h, \int_{\tilde{K}} w \cdot v \, dx = 0 \ \forall v \in N(\tilde{K})\right\}.$$

and let

$$|v|^2_{1,h,\tilde{K}} = h^{-1} \int_{\Gamma_h \cap \tilde{K}} |v^+ - v^-|^2 \, ds + h^{-1} \int_{\Gamma_1 \cap \partial\tilde{K}} |v|^2 \, ds, \qquad v \in V_h.$$

It follows from assumption (iii) that $|\cdot|_{1,h,\tilde{K}}$ is a norm on $W(\tilde{K})$. For, if $v \in W(\tilde{K})$ and $|v|_{1,h,\tilde{K}} = 0$, then $v = 0$ if $\tilde{K}$ has a side on Γ_1, and otherwise v is a rigid displacement of $\tilde{K}$. By assumption (iii), this implies that $v \in N(\tilde{K})$, so $v \in W(\tilde{K}) \cap N(\tilde{K})$, which is possible only if $v = 0$.

On the other hand, if we define

$$|v|_{h,\tilde{K}} = \sup_{\tau \in U(\tilde{K})} \frac{\int_{\tilde{K}} \operatorname{div} \tau \cdot v \, dx}{\|\tau\|_{0,\tilde{K}}},$$

then $|\cdot|_{h,\tilde{K}}$ is another norm on $W(\tilde{K})$, by the definition of $W(\tilde{K})$. Using now a scaling argument and the equivalence of norms in a finite-dimensional space, recalling the assumption (i), we conclude the existence of a constant which depends only on M such that, for all $v \in W(\tilde{K})$,

$$|v|_{h,\tilde{K}} \geqslant C|v|_{1,h,\tilde{K}}. \tag{4.21}$$

Since $|v|_{h,\tilde{K}} = |v|_{1,h,\tilde{K}} = 0$ if $v \in N(\tilde{K})$, we further conclude that (4.21) actually holds for all $v \in V_{h|\tilde{K}}$. In view of the definition of $|\cdot|_{h,\tilde{K}}$, this may be interpreted as: For each $v \in V_{h|\tilde{K}}$, there exists $\tau_{\tilde{K}} \in U(\tilde{K})$ such that

$$\|\tau_{\tilde{K}}\|_0 \leqslant C|v|_{1,h,\tilde{K}}, \tag{4.22a}$$

$$\int_{\tilde{K}} \operatorname{div} \tau_{\tilde{K}} \cdot v \, dx \geqslant |v|^2_{1,h,\tilde{K}}, \tag{4.22b}$$

where C depends only on M.

Now let $v \in V_h$ be given, define $\tau_{\tilde{K}} \in H_h$ for each $\tilde{K} \in \mathfrak{S}^h$ so that (4.22) holds and $\tau_{\tilde{K}} = 0$ on $\Omega \setminus \tilde{K}$, and let $\tau \in H_h$ be defined by

$$\tau = \sum_{\tilde{K} \in \mathfrak{S}^h} \tau_{\tilde{K}}.$$

Then, by (4.22a) and by assumptions (i) and (ii),

$$\|\tau\|_0^2 = \sum_{K \in \mathcal{C}^h} \left\| \sum_{\substack{\tilde{K} \in \mathfrak{S}^h \\ \tilde{K} \supset K}} \tau_{\tilde{K}} \right\|^2_{0,K} \leqslant M \sum_{\tilde{K} \in \mathfrak{S}^h} \|\tau_{\tilde{K}}\|^2_{0,\tilde{K}}$$

$$\leqslant CM \sum_{\tilde{K} \in \mathfrak{S}^h} |v|^2_{1,h,\tilde{K}} \leqslant 2CM^2 |v|^2_{1,h}.$$

Here we also used the fact that if l is a side of $K \in \mathcal{C}^h$ on Γ_1, then l is contained in $\partial\tilde{K}$ for at most $2M$ different rectangles $\tilde{K} \in \mathfrak{S}^h$. This is a consequence of assumption (ii). From (ii) it also follows that each side of each $K \in \mathcal{C}^h$ is contained in at least one $\tilde{K} \in \mathfrak{S}^h$ if l is in the interior of Ω or in at least one $\partial\tilde{K}$ if $l \subset \Gamma$. Therefore, and by (4.22b),

$$(\operatorname{div} \tau, v) = \sum_{\tilde{K} \in \mathfrak{S}^h} \int_{\tilde{K}} \operatorname{div} \tau_{\tilde{K}} \cdot v \, dx \geqslant \sum_{\tilde{K} \in \mathfrak{S}^h} |v|^2_{1,h,\tilde{K}} \geqslant |v|^2_{1,h}.$$

Combining these inequalities, (4.16) follows. □

We consider now two examples where the above hypotheses can be verified.

Example 1. Let H_h be defined by

$$H_h = \left\{ \tau = \sum_{\substack{P \in \mathfrak{N} \\ P \notin \Gamma}} \sum_{i=1}^{6} \alpha_i^P \tau_i^P + \sum_{\substack{P \in \mathfrak{N} \\ P \in \Gamma}} \sum_{i=1}^{3} \alpha_i^P \tau_i^P, \ \alpha_i^P \in \mathbf{R}, \ \tau \cdot n = 0 \text{ on } \Gamma_2 \right\},$$

where the functions τ_i^P are as above. □

Assume that $m_1, m_2 \geqslant 3$ in (2.5), and let $\mathfrak{S}^h$ be a collection of rectangles $\tilde{K}$ such that assumption (ii) is satisfied for some finite M and such that each $\tilde{K} \in \mathfrak{S}^h$ is of the form

$$\tilde{K} = \{x \in \Omega : ih_1 < x_1 < (i + \nu_1)h_1, jh_2 < x_2 < (j + \nu_2)h_2\},$$

where $\nu_i = 3$ if $\tilde{K}$ has a side l on Γ parallel with the x_i-axis and otherwise $\nu_1 = \nu_2 = 2$. For such $\mathfrak{S}^h$, assumption (i) holds with $M = 9$, so it remains to verify (iii).

Assume first that $\partial\tilde{K} \cap \Gamma = \varnothing$. Then if $\tau \in H_h$ and $\tau = 0$ on $\tilde{K} \setminus \Omega$, we have (see Figure 5)

$$\tau = \sum_{i=1}^{6} \alpha_i \tau_i^A + \alpha_7\left(\tau_1^B - \tau_4^B\right) + \alpha_8\left(\tau_1^C + \tau_4^C\right) + \alpha_9\left(\tau_2^D - \tau_5^D\right) + \alpha_{10}\left(\tau_2^E + \tau_5^E\right), \qquad \alpha_i \in \mathbf{R}. \tag{4.23}$$

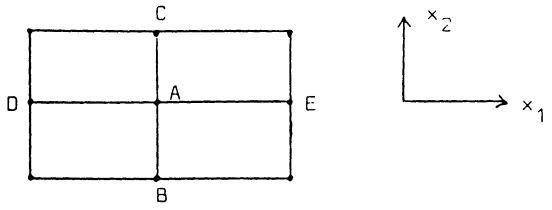

FIGURE 5

Hence if $U(\tilde{K})$ is defined by (4.20a) we have $\dim(U(\tilde{K})) = 10$. If $N(\tilde{K})$ is now defined as in (4.20b), then $v \in N(\tilde{K})$ if and only if

$$\int_K \operatorname{div}\tau \cdot v\, dx = \int_{\Gamma_h \cap K} (\tau \cdot n) \cdot (v^+ - v^-)\, ds = 0$$

for all τ of the form (4.23). It is straightforward to verify that this is possible only if $v^+ = v^-$ on Γ_h, i.e., v is a rigid displacement of $\tilde{K}$.

Assume next that $\tilde{K}$ has a side on Γ_1 which contains the nodes $A, B \in \mathfrak{N}$ (see Figure 6). It follows from the above consideration that if $v \in N(\tilde{K})$, then v can be at most a rigid displacement of K. But by (4.20) and by the definition of H_h we also have

$$\int_{\tilde{K}} \operatorname{div}\tau_i^P \cdot v\, dx = \int_{\partial\tilde{K} \cap \Gamma_1} \left(\tau_i^P \cdot n\right) \cdot v\, ds = 0, \qquad P = A, B, i = 1,2,3. \tag{4.24}$$

It is easy to see that if v is a rigid displacement on $\tilde{K}$, then (4.24) implies that $v = 0$. Thus, assumption (iii) is verified and (4.16) now follows from Lemma 4.4.

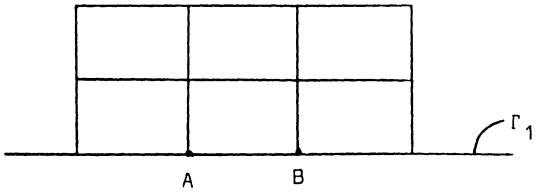

FIGURE 6

Example 2. We assume here that m_1 and m_2 in (2.5) are even numbers, with $m_1, m_2 \geq 4$. Denote by $\tilde{\mathcal{C}}^h$ a coarser rectangular subdivision of Ω such that each $\tilde{K} \in \tilde{\mathcal{C}}^h$ is of the type

$$\tilde{K} = \{x \in \Omega: 2ih_1 < x_1 < 2(i+1)h_1,\ 2jh_2 < x_2 < 2(j+1)h_2\}$$

for some i, j, $0 \leq i \leq m_1/2$, $0 \leq j \leq m_2/2$. Let $\mathfrak{N}_i$, $i = 0,\dots,3$, be subsets of the nodal set $\mathfrak{N}$ such that $\mathfrak{N}_0$ is the set of midpoints of $\tilde{K} \in \tilde{\mathcal{C}}^h$, $\mathfrak{N}_i$, $i = 1, 2$, is the set of midpoints of the sides of $\tilde{K} \in \tilde{\mathcal{C}}^h$ that are parallel with the x_i-axis, and $\mathfrak{N}_3$ is the set of vertices of $\tilde{K} \in \tilde{\mathcal{C}}^h$. Then define H_h as

$$H_h = \Bigg\{ \tau = \sum_{P \in \mathfrak{N}} \sum_{i=1}^{3} \alpha_i^P \tau_i^P + \sum_{P \in \mathfrak{N}_0} \sum_{i=4,6} \alpha_i^P \tau_i^P + \sum_{P \in \mathfrak{N}_1} \alpha_4^P \tau_4^P + \sum_{P \in \mathfrak{N}_2} \alpha_5^P \tau_5^P,\ \alpha_i^P \in \mathbf{R},\ \tau \cdot n = 0 \text{ on } \Gamma_2 \Bigg\}. \quad \square$$

Note that

$$\dim(H_h) = 4m_1 m_2 + O(h^{-1}),$$

so there are only about four degrees of freedom per node in this case.

To verify assumptions (i) through (iii) for the above choice of H_h, let $\mathcal{E}^h$ be a collection of rectangles $\tilde{K}$ such that (ii) is satisfied for some finite M and such that each $\tilde{K} \in \mathcal{E}^h$ is of the type

$$\tilde{K} = \{x \in \Omega\colon 2ih_1 < x_1 < (2i+4)h_1,\ 2jh_2 < x_2 < (2j+4)h_2\},$$

where i, $j \geq 0$, $(2i+4)h_1 \leq a_1$, $(2j+4)h_2 \leq a_2$. Then assumption (i) holds with $M = 16$.

To verify assumption (iii), consider first a given macroelement $\tilde{K}$ which contains four rectangles of $\mathcal{C}^h$. With the notation of Figure 5, if $\tau \in H_h$ vanishes outside $\tilde{K}$, then

$$\tau = \sum_{i=1}^{4} \alpha_i \tau_i^A + \alpha_5 \tau_6^A + \alpha_6(\tau_1^B - \tau_4^B) + \alpha_7(\tau_1^C + \tau_4^C) + \alpha_8(\tau_2^D - \tau_5^D) + \alpha_9(\tau_2^E + \tau_5^E), \qquad \alpha_i \in \mathbf{R}.$$

We omit the details of showing that if $v \in V_h$ and $\int_{\tilde{K}} \operatorname{div} \tau \cdot v\, dx = 0$ for all τ of the above form, then $v_{|\tilde{K}}$ is a rigid displacment.

Consider now a rectangle $\tilde{K} \in \tilde{\mathcal{E}}^h$ consisting of four subrectangles $\tilde{K}_i$ as in Figure 7, and let $U(\tilde{K})$ and $N(\tilde{K})$ be defined by (4.20).

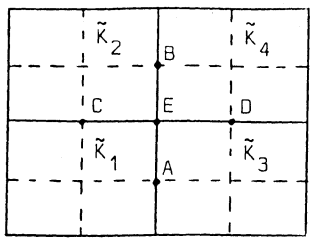

FIGURE 7

We noted above that if $v \in N(\tilde{K})$, then $v_{|\tilde{K}_i}$ is a rigid displacement for $i = 1,\dots,4$. Therefore, if $v \in N(\tilde{K})$, we may write

$$v(x) = \frac{h_2}{h}\alpha_i(1,0) + \frac{h_1}{h}\beta_i(0,1) + \frac{1}{2h}\gamma_i(x_2 - q_{2i}, q_{1i} - x_1), \qquad x \in \tilde{K}_i,\ i = 1,\dots,4, \tag{4.25}$$

where $\alpha_i, \beta_i, \gamma_i \in \mathbf{R}$ and (q_{1i}, q_{2i}) is the midpoint of $\tilde{K}_i$. Now, since $\tau_{i|\tilde{K}}^P \in U(\tilde{K})$ for $i = 1, 2, 3$ and $P = A, B, C, D, E$ (see Figure 7), any $v \in N(\tilde{K})$ satisfies, for example,

$$\int_{\tilde{K}} \operatorname{div} \tau_i^P \cdot v \, dx = 0 \quad \text{for} \begin{cases} P = A, B, & i = 1, 3, \\ P = C, D, & i = 2, 3, \\ P = E, & i = 1. \end{cases}$$

With v given by (4.25), these equations are written equivalently as

$$\begin{cases} \alpha_1 - \alpha_3 = 0 \\ \beta_1 - \beta_3 - \gamma_1 - \gamma_3 = 0 \\ \alpha_2 - \alpha_4 = 0 \\ \beta_2 - \beta_4 - \gamma_2 - \gamma_4 = 0 \\ \beta_1 - \beta_2 = 0 \\ \alpha_1 - \alpha_2 + \gamma_1 + \gamma_2 = 0 \\ \beta_3 - \beta_4 = 0 \\ \alpha_3 - \alpha_4 + \gamma_3 + \gamma_4 = 0 \\ \alpha_1 - \alpha_2 - \alpha_3 - \alpha_4 + \frac{2}{3}(\gamma_1 - \gamma_2 - \gamma_3 + \gamma_4) = 0 \end{cases}$$

From this it is easy to see that the equations are all linearly independent. Thus, if $\partial\tilde{K} \cap \Gamma_1 = \varnothing$, $\dim(N(\tilde{K})) = 12 - 9 = 3$, and so $N(\tilde{K})$ consists of the rigid displacements only.

Finally if $\partial\tilde{K} \cap \Gamma_1 \neq \varnothing$, one can further show that $N(\tilde{K}) = \{0\}$ using the same argument as in Example 1 above. Thus, assumption (iii) is verified, and so (4.16) follows.

5. A Numerial Example. We give here the results of numerical computations using the methods presented in the preceding sections. We consider a simple model problem where Ω is the unit square, $\lambda = -0.3$ and $\mu = 1.3$ in (2.1), and $f = (f_1, f_2)$ is chosen so that the exact displacements under the boundary condition $u = 0$ on Γ are

$$u_1 = 16x_1(1 - x_1)x_2(1 - x_2)e^{(x_1 - x_2)},$$
$$u_2 = \sin \pi x_1 \sin \pi x_2, \qquad (x_1, x_2) \in \Omega.$$

In solving the discrete equations, the following iterative version of the penalty method (2.4) is used (cf. [11, p. 321])

$$a(\sigma_h^k, \tau) + \frac{1}{\varepsilon}(\pi_h \operatorname{div} \sigma_h^k, \pi_h \tau)$$
$$= -\frac{1}{\varepsilon}(\pi_h f, \pi_h \operatorname{div} \tau) - (u_h^{k-1}, \operatorname{div} \tau), \qquad \tau \in H_h, \tag{5.1}$$
$$u_h^k = u_h^{k-1} + \frac{1}{\varepsilon}\pi_h(\operatorname{div} \sigma_h^k + f), \qquad k = 1, 2, \ldots.$$

Using $u_h^0 = 0$ as the starting guess, we see that the first step in (5.1) is equivalent to the penalty method (2.3a, b′).

The main benefit of using (5.1) instead of (2.3a, b′) is that one has more freedom in choosing the parameter ε. For, if the scheme (2.3a, b′) is used, one has to take $\varepsilon = O(h^2)$ in order to obtain the convergence rate $\|\sigma - \sigma_h\|_0 = O(h^2)$ (cf. [10]). This

makes the system (2.4) rather ill-conditioned (the condition number is of the order of h^{-4}). Also, one has to compute $\pi_h f$ with high precision in (2.4) since the error is multiplied by $1/\varepsilon$.

In practice the iteration (5.1) seems to converge quite fast: not more than six iterations were required in the computations. Since only one Cholesky decomposition is required in (5.1), the additional cost due to iteration is relatively small.

Table I shows the computed L_2-errors $\|\sigma - \sigma_h\|_0$ in the above model problem. Method IIA corresponds to the choice $H_h = H_{h,1}$, in IIB H_h is chosen as in Example 1 and in IIC as in Example 2 of Section 4. In parentheses is shown the relative number of operations required in the Cholesky reduction of the matrix in (5.1).

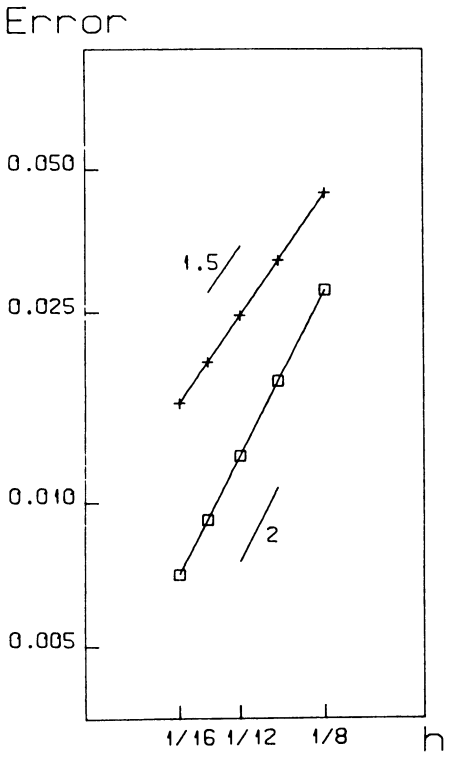

FIGURE 8

L_2-errors $\|\sigma - \sigma_h\|_0$ in a model problem for Method I (+) *and Method* IIC (□). *Two line segments with slopes* 3/2 *and* 2 *are drawn for comparison.*

As shown in Figure 8, the rate of convergence is $O(h^{3/2})$ for Method I and $O(h^2)$ for Method II, as expected theoretically. Notice also that the constant in the error estimate is practically the same in the three variants of Method II. Thus, Method IIC should be preferred as it involves the least amount of computational work. As compared with Method I, Method IIC is superior roughly below the error level $\|\sigma - \sigma_h\|_0 \simeq 0.03$.

TABLE I

L_2-errors $\|\sigma - \sigma_h\|_0$ for various methods in a model problem. Numbers in parentheses indicate the relative cost of Cholesky reduction.

$1/h$	Method I		Method II					
			A		B		C	
4	0.123	(1.0)	0.090	(19)	0.101	(8.0)	0.102	(3.0)
6	0.068	(5.1)	0.042	(97)	0.047	(41)	0.048	(17)
8	0.045	(16)	0.024	(300)	0.026	(130)	0.027	(48)
10	0.032	(39)	0.016	(740)	0.017	(310)	0.018	(120)
12	0.025	(81)	0.011	(1500)	0.012	(650)	0.012	(240)

Institute of Mathematics
Helsinki University of Technology
SF-02150 Espoo 15, Finland

1. I. BABUŠKA, "Error bounds for finite element method," *Numer. Math.*, v. 16, 1971, pp. 322–333.

2. I. BABUŠKA, J OSBORN & J. PITKÄRANTA, "Analysis of mixed methods using mesh dependent norms," *Math. Comp.*, v. 35, 1980, pp. 1039–1062.

3. M. BERCOVIER, "Perturbation of mixed variational problems. Application to mixed finite element methods," *RAIRO Anal. Numér.*, v. 12, 1978, pp. 211–236.

4. J. BRAMBLE & S. HILBERT, "Estimation of linear functionals on Sobolev spaces with application to Fourier transforms and spline interpolation," *SIAM J. Numer. Anal.*, v. 7, 1970, pp. 112–124.

5. F. BREZZI, "On the existence, uniqueness and approximation of saddle-point problems arising from Lagrange multipliers," *RAIRO Ser. Rouge*, v. 8, 1974, pp. 129–151.

6. P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

7. G. DUVAUT & J. L. LIONS, *Inequalities in Mechanics and Physics*, Springer-Verlag, Berlin and New York, 1976.

8. I. HLAVAČEK, "Convergence of an equilibrium finite element method for plane elastostatics," *Apl. Math.*, v. 24, 1979, pp. 427–456.

9. C. JOHNSON & B. MERCIER, "Some equilibrium finite element methods for two-dimensional elasticity problems," *Numer. Math.*, v. 30, 1978, pp. 103–116.

10. C. JOHNSON & J. PITKÄRANTA, "Analysis of some mixed finite element methods related to reduced integration," *Math. Comp*, v. 38, 1982, pp. 375–400.

11. D. G. LUENBERGER, *Introduction to Linear and Nonlinear Programming*, Addison-Wesley, Reading, Mass., 1973.

12. D. MALKUS & T. HUGHES, "Mixed finite element methods—reduced and selective integration techniques: A unification of concepts," *Comp. Methods Appl. Mech. Engrg.*, v. 15, 1978, pp. 63–81.

13. J. NITSCHE, "On Korn's second inequality," *RAIRO Anal. Numér.*, v. 15, 1981, pp. 237–248.

14. G. SANDER, *Application of the Dual Analysis Principle*, Proc. of IUTAM Symp. on High Speed Computing of Elastic Structures, Univ. de Liège, 1971, pp. 167–207.

15. R. TAYLOR & O. C. ZIENKIEWICZ, "Complementary energy with penalty functions in finite element analysis," in *Energy Methods in Finite Element Analysis* (R. Glowinski, E. Y. Rodin and O. C. Zienkiewicz, eds.), Wiley, New York, 1979, pp. 143–174.

16. V. B. WATWOOD, JR. & B. J. HARTZ, "An equilibrium stress field model for finite element solutions of two-dimensional elastostatic problems," *Internat. J. Solids and Structures*, v. 4, 1968, pp. 857–873.

17. B. FRAEJIS DE VEUBEKE, "Displacement and equilibrium models in the finite element method," in *Stress Analysis* (O. C. Zienkiewicz and G. S. Holister, eds.), Wiley, New York, 1965, pp. 145–197.

18. B. FRAEJIS DE VEUBEKE, *Finite Element Method in Aerospace Engineering Problems*, Proc. Internat. Sympos. Computing Methods in Appl. Sci. and Eng., Versailles, 1973, Part 1, pp. 224–258.