THEORETICAL AND APPLIED MECHANICS
25, pp. 21-31, 1999, UDK 539.3

Least-squares mixed finite elements for the linear elasticity problem

Boško Jovanović and Ivan Šestak

Submitted 14 April 1997

Abstract

We expose a theoretical analysis of a least-squares mixed finite elements method for the linear elasticity problem in two- and three-dimensional domains. The coerciveness of the weak form of the problem is proved. It is shown that the finite element approximation yields a symmetric positive definite linear system with condition number $O(h^{-2})$. The error estimate is obtained.

1 Introduction

Many classical boundary value problems for second order linear elliptic equations can be transformed into the first order systems of equations. Such formulation of the problem is called mixed formulation [5].

Unknown functions in mixed boundary value problems often have appropriate physical or mechanical interpretation (temperature and flux, displacements and stress etc.). In such a manner, solution of mixed problem gives more complex information about the considered object or process. For these reasons, numerical methods for solving mixed boundary value problems are of great importance.

Ritz-Galerkin formulation of the finite element method for mixed boundary value problems requires relatively weak smoothness of input data. In such a manner, it requires relatively weak assumptions for the smoothness of finite element approximations. On the other hand, the finite dimensional spaces used in the Ritz-Galerkin method must satisfy so called Ladyzhenskaya–Babuška-Brezzi (LBB) condition (see [10], [2], [3]) which substantially restrict the choice of feasible finite element spaces. In the Ritz-Galerkin method the solution of the problem is the saddle point of a quadratic functional, while the corresponding linear system is indefinite. This may cause difficulties in the solution algorithm.

Contrary to the Ritz-Galerkin method, the least-squares method for the mixed boundary value problems is not the subject to LBB condition, while the corresponding linear system is definite. (As a rule, its condition number is $O(h^{-2})$). Of course, the price for that is slightly stronger smoothness required for the input data.

The least-squares mixed finite element method for the second-order elliptic equation is considered in [11]. In [4] and [8] for the Stokes problem a mixed method which does not require the LBB condition is considered.

2 Problem Formulation

Let be Ω a bounded domain in $\mathbb{R}^N$, $N = 2$ or 3. The linear elasticity problem is given by (see [9])

$$
\begin{aligned}
div\,\sigma + f &= 0 \qquad \text{in } \Omega,\\
\sigma - \mathbb{C}\,\varepsilon(u) &= 0 \qquad \text{in } \Omega,\\
2\varepsilon(u) &= \nabla u + (\nabla u)^t \qquad \text{in } \Omega, \qquad (1)\\
u &= 0 \qquad \text{on } \Gamma_1,\\
\sigma \cdot n &= 0 \qquad \text{on } \Gamma_2,
\end{aligned}
$$

where $u : \Omega \to \mathbb{R}^N$ is the displacement vector, $\sigma : \Omega \to \mathbb{R}^{N\times N}$ is the stress tensor, $\varepsilon(u)$ is the strain tensor, $f \in (L^2(\Omega))^N$ is the given body load, $\partial\Omega = \bar{\Gamma}_1 \cup \bar{\Gamma}_2$, $\Gamma_1 \cap \Gamma_2 = \emptyset$, $\Gamma_1 \neq 0$, while $(\cdot)^t$ denotes the transposition. For the simplicity, in the following we assume that Ω is a polygonal (polyhedral) domain in $\mathbb{R}^2$ ($\mathbb{R}^3$). Further, we define

$$(\sigma \cdot n)_i = \sigma_{ij} n_j,$$

$$(div\ \sigma)_i = \partial\sigma_{ij}/\partial x_j,$$

$$(\mathbb{C}\varepsilon(u))_{ij} = C_{ijkl}\varepsilon_{kl}(u),$$

where n is the unit outward normal to $\partial\Omega$, $i, j, k, l = 1, ..., N$, and summation is implied for repeated indices. The elasticity tensor $\mathbb{C} = \{C_{ijkl}\}$ $(C_{ijkl} \in L^\infty(\Omega),\ i, j, k, l = 1, ..., N)$ is completely symmetric and elliptic, i.e.,

$$C_{ijkl} = C_{jikl} = C_{ijlk} = C_{klij}, \tag{2}$$

and

$$\exists m > 0 : C_{ijkl}\varepsilon_{ij}\varepsilon_{kl} \geq m\varepsilon_{ij}\varepsilon_{ij}, \quad \forall\varepsilon \in \mathbb{R}^{N\times N}. \tag{3}$$

By $(\cdot,\cdot)_{s,\Omega}$, $|\cdot|_{s,\Omega}$ and $\|\cdot\|_{s,\Omega}$ we denote respectively the inner product, seminorm and norm of the Sobolev space $H^s(\Omega)$ (see [1]), while by $(\cdot,\cdot)_{s,\Omega;\alpha}$, $|\cdot|_{s,\Omega;\alpha}$ and $\|\cdot\|_{s,\Omega;\alpha}$ we denote the inner product, seminorm and norm of the space $(H^s(\Omega))^\alpha$ where $\alpha > 1$ is integer. Further, we introduce the following functional spaces with corresponding norms and inner products:

$$L = (L^2(\Omega))^N,$$

$$V = \left\{v \in (H^1(\Omega))^N : v = 0 \text{ on } \Gamma_1\right\},$$

$$\|v\|_V = \|v\|_{1,\Omega;N} = \left(\sum_{i=1}^N \|v_i\|_{1,\Omega}^2\right)^{1/2},$$

$$T = \left\{\tau \in (L^2(\Omega))^{N\times N} : \tau_{ij} = \tau_{ji},\quad i, j = 1, ..., N\right\},$$

$$\|\tau\|_T = \|\tau\|_{0,\Omega;N\times N} = \left(\sum_{i,j=1}^N \|\tau_{ij}\|_{0,\Omega}^2\right)^{1/2},$$

$$H = H(div, \Omega) = \{\tau \in T : div\, \tau \in L,\ \tau \cdot n = 0 \text{ on } \Gamma_2\},$$

$$\|\tau\|_H = \left(\|\tau\|_T^2 + \|div\, \tau\|_L^2\right)^{1/2}.$$

In the sequel, by C and C_i we shall denote positive generic constants, which can take different values in different formulae.

The least-squares functional for the problem (1) is

$$J(v,\tau) = (div\, \tau + f,\ div\, \tau + f)_L + (\tau - \mathbb{C}\varepsilon(v),\ \tau - \mathbb{C}\varepsilon(v))_T, \quad (4)$$

where $(\cdot,\cdot)_L$ and $(\cdot,\cdot)_T$ are inner products in L and T.

The least-squares minimization problem is given by

$$\text{find } (u,\sigma) \in V \times H \text{ such that}$$

$$J(u,\sigma) = \inf_{(v,\tau)\in V\times H} J(v,\tau).$$

Taking variations of the functional (4) with respect to v and τ we obtain the following weak statement of the problem

$$\begin{gathered}\text{find } (u,\sigma) \in V \times H \text{ such that} \\ a((u,\sigma),(v,\tau)) = l(v,\tau), \text{ for all } (v,\tau) \in V \times H,\end{gathered} \quad (5)$$

where

$$a((u,\sigma),(v,\tau)) = (div\, \sigma,\ div\, \tau)_L + (\sigma - \mathbb{C}_\varepsilon(u), \tau - \mathbb{C}_\varepsilon(v))_T, \quad (6)$$

and

$$l(v,\tau) = -(f, div\, \tau)_L. \quad (7)$$

The following theorem assert the coerciveness of the bilinear form (6).

Theorem 1. *There exists a constant* $C > 0$ *such that for all* $(v,\tau) \in V \times H$

$$a((v,\tau),(v,\tau)) \geq C\left(\|v\|_V^2 + \|\tau\|_H^2\right). \quad (8)$$

Proof. From (6)

$$
\begin{aligned}
a\left((v,\tau),(v,\tau)\right) = \; & \left(div\ \tau,\ div\ \tau\right)_L + (\tau,\tau)_T \\
& -2\left(\tau, \mathbb{C}\varepsilon(v)\right)_T + \left(\mathbb{C}\varepsilon(v), \mathbb{C}\varepsilon(v)\right)_T \\
& \pm 2\beta\left(\tau,\varepsilon(v)\right)_T \pm \beta^2 (v,v)_L \\
& \pm \left((\beta\mathbf{I}-\mathbb{C})\,\varepsilon(v), (\beta\mathbf{I}-\mathbb{C})\,\varepsilon(v)\right)_T,
\end{aligned} \tag{9}
$$

where $\beta > 0$ is a constant and $\mathbf{I}$ is the unit tensor of the same order as $\mathbb{C}$. Using the identity (Green's formula)

$$
\left(\tau,\varepsilon(v)\right)_T = -\left(div\ \tau, v\right)_L
$$

from (9) we get

$$
\begin{aligned}
a\left((v,\tau),(v,\tau)\right) = \; & \left(div\ \tau + \beta v,\ div\ \tau + \beta v\right)_L \\
& + \left(\tau + (\beta\mathbf{I}-\mathbb{C})\,\varepsilon(v), \tau + (\beta\mathbf{I}-\mathbb{C})\,\varepsilon(v)\right)_T \\
& -\beta^2 (v,v)_L + \left(\mathbb{C}\varepsilon(v), \mathbb{C}\varepsilon(v)\right)_T \\
& -\left((\beta\mathbf{I}-\mathbb{C})\,\varepsilon(v), (\beta\mathbf{I}-\mathbb{C})\,\varepsilon(v)\right)_T \\
\geq \; & -\beta^2 (v,v)_L - \beta^2\left(\varepsilon(v),\varepsilon(v)\right)_T + 2\beta\left(\varepsilon(v), \mathbb{C}\varepsilon(v)\right)_T .
\end{aligned} \tag{10}
$$

Due to the compact imbedding $H^1(\Omega) \subset L^2(\Omega)$, Korn's inequality (9) and relation (3), from (10) follows

$$
a\left((v,\tau),(v,\tau)\right) \geq \beta\left[2mC_2 - \beta\left(C_1 + C_2\right)\right] \|v\|_V^2 .
$$

Let $\beta = mC_2/\left(C_1 + C_2\right)$. Then

$$
\beta\left[2mC_2 - \beta\left(C_1 + C_2\right)\right] = \frac{\left(mC_2\right)^2}{C_1 + C_2} = C_3 > 0,
$$

and finally

$$
a\left((v,\tau),(v,\tau)\right) \geq C_3 \|v\|_V^2 . \tag{11}
$$

It is obvious from definition (6) that

$$a\left(\left(v,\tau\right),\left(v,\tau\right)\right)\geq\left\|\tau-\mathbb{C}\varepsilon\left(v\right)\right\|_{T}^{2}, \tag{12}$$

$$a\left(\left(v,\tau\right),\left(v,\tau\right)\right)\geq\left\|div\ \tau\right\|_{L}^{2}. \tag{13}$$

Using (12) and (11) we get

$$\left\|\tau\right\|_{T}^{2}=\left\|\left(\tau-\mathbb{C}\varepsilon\left(v\right)\right)+\mathbb{C}\varepsilon\left(v\right)\right\|_{T}^{2}\leq C_{4}a\left(\left(v,\tau\right),\left(v,\tau\right)\right). \tag{14}$$

Combining (11), (13) and (14) we obtain (8).□

Theorem 2. *Let $f\in L$. Then*

$$\left\|u\right\|_{V}+\left\|\sigma\right\|_{H}\leq C\left\|f\right\|_{L}. \tag{15}$$

Proof. The proof of the theorem follows easily from Theorem 1.□

3 Finite Element Approximation

Let $\mathcal{T}_h$ be a partition of the domain Ω into finite elements: $\Omega=\cup_{K\in\mathcal{T}_h}K$ and $h=max\left\{diam\left(K\right):K\in\mathcal{T}_h\right\}$. Let V_h and H_h be finite-dimensional subspaces of V and H, which have the following approximation properties:

$$\inf_{v_h\in V_h}\left\|v-v_h\right\|_{V}\leq Ch^{k}\left\|v\right\|_{k+1,\Omega;N}, \tag{16}$$

and

$$\inf_{\tau_h\in H_h}\left\|\tau-\tau_h\right\|_{H}\leq Ch^{l}\left\|\tau\right\|_{l+1,\Omega;N\times N}, \tag{17}$$

where $k>0$ and $l>0$ are integers. Standard choices for V_h and H_h are spaces with piecewise polynomial bases, i.e.,

$$V_h=\left\{v_h\in V:v_{h,i}\in C^{0}\left(\bar{\Omega}\right),v_{h,i|K}\in\mathcal{P}_k\left(K\right),\right.$$
$$\left.\forall K\in\mathcal{T}_h;\ v_{h,i}=0\text{ on }\Gamma_1;\ i=1,...,N\right\},$$
$$H_h=\left\{\tau_h\in H:\tau_{h,ij|K}\in\mathcal{P}_l\left(K\right),\ \forall K\in\mathcal{T}_h;\ i,j=1,...,N\right\}.$$

Here $\mathcal{P}_s(K)$ is the space of polynomials of degree s on K. Another possible choice for H_h are the Raviart-Thomas spaces [12]. In the following we will assume that the domain Ω is covered by finite elements exactly, and that the integration is exact.

The finite element approximation of the problem (5) is

$$\begin{gathered}\text{find } (u_h, \sigma_h) \in V_h \times H_h \text{ such that} \\ a((u_h, \sigma_h), (v_h, \tau_h)) = l(v_h, \tau_h), \text{ for all } (v_h, \tau_h) \in V_h \times H_h.\end{gathered} \tag{18}$$

From Theorem 1 it follows that the problem (18) has a unique solution. Moreover, the error has the orthogonality property

$$\begin{gathered}a((u,\sigma) - (u_h, \sigma_h), (v_h, \tau_h)) = a((u - u_h, \sigma - \sigma_h), (v_h, \tau_h)) = 0 \\ \text{for all } (v_h, \tau_h) \in V_h \times H_h.\end{gathered} \tag{19}$$

Let us estimate the condition number of the linear system, which we obtain from (18). Suppose that the finite element partition is regular [6] and that the "inverse assumption" is satisfied, i.e.

$$\begin{gathered}\text{there exists a constant } \delta > 0 \text{ such that} \\ \delta h \le diam(K) \le h, \quad \text{for all} \quad K \in \mathcal{T}_h.\end{gathered} \tag{20}$$

If we use isoparametric elements, then the elements should not be too much distorted (see [7]).

Let $\varphi_1, ..., \varphi_L$ and $\psi_1, ..., \psi_M$ be sets of basis functions in V_h and H_h. We suppose that there exist positive constants A_1, A_2, B_1 and B_2 such that for all real vectors $(\alpha_1, ..., \alpha_L)$ and $(\beta_1, ..., \beta_M)$

$$A_1 h^N \sum_{i=1}^{L} \alpha_i^2 \le \left(\sum_{i=1}^{L} \alpha_i \varphi_i, \sum_{i=1}^{L} \alpha_i \varphi_i \right)_{0,\Omega;N} = \left\| \sum_{i=1}^{L} \alpha_i \varphi_i \right\|_L^2 \le A_2 h^N \sum_{i=1}^{L} \alpha_i^2, \tag{21}$$

$$B_1 h^N \sum_{j=1}^{L} \beta_j^2 \le \left(\sum_{j=1}^{M} \beta_j \psi_j, \sum_{j=1}^{M} \beta_j \psi_j \right)_{0,\Omega;N\times N} = \left\| \sum_{j=1}^{M} \beta_j \psi_j \right\|_T^2 \le B_2 h^N \sum_{j=1}^{M} \beta_j^2. \tag{22}$$

Notice that these inequalities are fulfilled for all well-known finite element spaces if (20) is satisfied.

Theorem 3. *Under the above assumptions the condition number of the linear system from (18) is* $O\left(h^{-2}\right)$.

Proof. From coerciveness and boundedness of the bilinear form follows

$$C_1\left(\|v_h\|_V^2+\|\tau_h\|_H^2\right)\le a\left(\left(v_h,\tau_h\right),\left(v_h,\tau_h\right)\right)\le C_2\left(\|v_h\|_V^2+\|\tau_h\|_H^2\right). \tag{23}$$

Using (20) and the inverse estimate [6]

$$|v_{h,i}|_{1,K}\le Ch^{-1}\|v_{h,i}\|_{0,K}\,,\quad K\in\mathcal{T}_h,$$

we obtain

$$\|v_h\|_L\le\|v_h\|_V\le Ch^{-1}\|v_h\|_L\,,$$

$$\|\tau_h\|_T\le\|\tau_h\|_H\le Ch^{-1}\|\tau_h\|_T\,.$$

Substituting in (23),

$$C_1\left(\|v_h\|_L^2+\|\tau_h\|_T^2\right)\le a\left(\left(v_h,\tau_h\right),\left(v_h,\tau_h\right)\right)\le C_3h^{-2}\left(\|v_h\|_L^2+\|\tau_h\|_T^2\right). \tag{24}$$

Since the system is symmetric, setting $v_h=\sum_{i=1}^{L}\alpha_i\varphi_i$, $\tau_h=\sum_{j=1}^{M}\beta_j\psi_j$ and using relations (21) and (22), the desired result follows from (24).□

4 Error Estimate

Let $u_I\in V_h$ and $\sigma_I\in H_h$ be the standard finite element interpolants of u and σ. Then we have [6]

$$\|u-u_I\|_V\le Ch^k\|u\|_{k+1,\Omega;N}\,, \tag{25}$$

$$\|\sigma-\sigma_I\|_H\le Ch^l\|\sigma\|_{l+1,\Omega;N\times N}\,. \tag{26}$$

From the coerciveness and boundedness of bilinear form and the orthogonality condition (19) follows

$$C\left(\|u-u_h\|_V^2+\|\sigma-\sigma_h\|_H^2\right)\leq((u-u_h,\sigma-\sigma_h),(u-u_h,\sigma-\sigma_h))=$$

$$=a\left((u-u_h,\sigma-\sigma_h),(u-u_I,\sigma-\sigma_I)\right)+$$

$$a\left((u-u_h,\sigma-\sigma_h),(u_I-u_h,\sigma_I-\sigma_h)\right)=$$

$$=a\left((u-u_h,\sigma-\sigma_h),(u-u_I,\sigma-\sigma_I)\right)\leq$$

$$\leq C_1\left(\|u-u_h\|_V^2+\|\sigma-\sigma_h\|_H^2\right)^{1/2}\left(\|u-u_I\|_V^2+\|\sigma-\sigma_I\|_H^2\right)^{1/2}.$$

From here, using (25) and (26) we immediately obtain

$$\|u-u_h\|_V+\|\sigma-\sigma_h\|_H\leq C\left(h^k\|u\|_{k+1,\Omega;N}+h^l\|\sigma\|_{l+1,\Omega;N\times N}\right)$$

$$\leq Ch^s\left(\|u\|_{k+1,\Omega;N}+\|\sigma\|_{l+1,\Omega;N\times N}\right),$$

where $s=\min\{k,l\}$. In such a manner, one obtains the optimal error estimate when $k=l=s$.

In such a way, we proved the following assertion.

Theorem 4. *For $k=l$ the error estimate*

$$\|u-u_h\|_V+\|\sigma-\sigma_h\|_H\leq Ch^k\left(\|u\|_{k+1,\Omega;N}+\|\sigma\|_{l+1,\Omega;N\times N}\right)\tag{27}$$

holds.

References

[1] R. A. Adams, Sobolev spaces, Academic Press, New York, 1975.

[2] I. Babuška, The finite element method with Lagrange multipliers, Numer. Math., 20, (1973), 179-192.

[3] F. Brezzi, On the existence, uniqueness and approximation of saddle point problems arising from Lagrange multipliers, RAIRO Ser. Anal. Numer., 8, (1974), 129-151.

[4] F. Brezzi and J. Jr. Douglas, Stabilized mixed method for the Stokes problem, Numer. Math., 53, (1988), 225-235.

[5] F. Brezzi and M. Fortin, Mixed and hybrid finite element methods, Springer-Verlag, New York, 1991.

[6] P. G. Ciarlet, The finite element method for elliptic problems, North Holland, Amsterdam, 1978.

[7] P. G. Ciarlet and P. A. Raviart, Interpolation theory over curved elements, with application to finite element methods, Comput. Methods Appl. Mech. Engrg., 1, (1972), 217-249.

[8] J. Jr. Douglas and J. P. Wang, An absolutely stabilized finite element method for the Stokes problem, Math. Comput., 52, (1989), 495-508.

[9] G. Duvaut and J. L. Lions, Les inéquations en méchanique et en physique, Dunod, Paris, 1972.

[10] O. A. Ladyzhenskaya, The mathematical theory of viscous incompressible flows, Gordon and Breach, London, 1969.

[11] A. I. Pehlivanov, G. F. Carey and R. D. Lazarov, Least-squares mixed finite elements for second-order elliptic problems, SIAM J. Numer. Anal., 31, (1994), 1368-1377.

[12] P. A. Raviart and J.M. Thomas, A mixed finite element method for 2nd order elliptic problems, Lect. Notes in Math., Springer-Verlag, 606, (1977), 292-315.

Acknowledgment - This research was supported by MST of Republic of Serbia, grant number 04M01.

Boško Jovanović
University of Belgrade
Faculty of Mathematics
Studentski trg 16, POB 550
11001 Belgrade
Yugoslavia

Ivan Šestak
Technical Faculty
19210 Bor
Yugoslavia

Metod najmanjih ostataka za mešovite konačne elemente u linearnoj teoriji elastičnosti

U radu je analiziran metod najmanjih ostataka za formulaciju jednačina konačnih elemenata pri rešavanju problema linearne teorije elastičnosti sa dvo i trodimenzijskim domenom. Dokazana je koercivnost slabe formulacije problema. Pokazano je da aproksimacija konačnim elementima dovodi do simetričnog pozitivno-definitnog linearnog sistema čija je uslovljenost reda $O(h^{-2})$. Takodje, data je i ocena greške.