ANALYSIS OF A MODIFIED FIRST-ORDER SYSTEM LEAST SQUARES METHOD FOR LINEAR ELASTICITY WITH IMPROVED MOMENTUM BALANCE

GERHARD STARKE *, ALEXANDER SCHWARZ , AND JÖRG SCHRÖDER†

Abstract. A modified first-order system least squares formulation for linear elasticity, obtained by adding the antisymmetric displacement gradient in the test space, is analyzed. This approach leads to surprisingly small momentum balance error compared to standard least squares approaches. It is shown that the modified least squares formulation is well-posed and its performance is illustrated by adaptive finite element computation based on using a closely related least squares functional as a posteriori error estimator. The results of our numerical computations show that, for the modified least squares approach, the momentum balance error converges at a much faster rate than the overall error. We prove that this is due to a strong connection of the stress approximation to that obtained from a mixed formulation based on the Hellinger-Reissner principle (with exact local momentum balance). The practical significance is that our proposed approach is almost momentum-conservative at a smaller number of degrees of freedom than mixed approximations with exact local momentum balance.

Key words. First-order system least squares, linear elasticity, incompressible, momentum balance, a posteriori error estimator, Raviart-Thomas elements.

AMS subject classifications. 65M60, 65M15

1. Introduction. In this paper, a modified first-order system least squares formulation for linear elasticity with improved momentum balance is analyzed. Its construction is obtained from the least squares approach in [11] by adding the antisymmetric displacement gradient in the test space. This formulation was introduced in [15] from an engineering perspective and shown to give much better results in bending dominated situations. The resulting variational formulation is based on a nonsymmetric bilinear form. The ellipticity properties of this bilinear form are shown to be essentially the same as for the symmetric bilinear form associated with the “true” least squares formulation from [11].

The numerical results in [15] show a significantly improved momentum balance for our nonsymmetric formulation compared to the standard least squares method. This seems to give an explanation for the improved results, especially for bending dominated problems. In this paper, we are able to show that the reduction rate for the momentum balance error is indeed of higher order than the overall approximation error. The improvement of the reduction rate depends on the regularity of the problem. A major step in our analysis is the derivation of a close connection between the stress approximation of our method and that obtained by a mixed formulation using the Arnold-Winther element ([3], generalized to three dimensions in [1]). Such a connection was established in [7] for the least squares formulation of second-order elliptic boundary value problem and its related saddle point and primal formulations.

Least squares finite element methods became very popular in recent years in many different application areas, see [5]. An important property of mixed formulations based on least squares functionals is the fact that its local evaluation provides a local a posteriori error estimate (see e.g. [4] or, in the context of linear elasticity

*Institut für Angewandte Mathematik, Leibniz Universität Hannover, Welfengarten 1, 30167 Hannover, Germany (gcs@ifam.uni-hannover.de).
†Institut für Mechanik, Fakultät für Ingenieurwissenschaften, Universität Duisburg-Essen, Universitätsstr. 15, 45117 Essen, Germany ({alexander.schwarz,j.schroeder}@uni-due.de).

1
Most importantly, such approaches simultaneously construct approximations to all variables of interest (like displacements and stresses) in appropriate spaces. Alternatively, mixed methods based on saddle point formulations may be used for this purpose. For this approach, however, suitable combinations of finite element spaces are restricted by the inf-sup conditions. Families of finite element spaces for such mixed formulations of saddle point type may be found e.g. in [12]. Those families of finite element spaces are based on Raviart-Thomas elements (as a generalization of PEERS, see [2]) or Brezzi-Douglas-Marini elements (as a generalization of BDMS, see [16]). A posteriori error estimators were constructed in [13] for these saddle point approaches based on the Helmholtz decomposition. Of course, the Arnold-Winther elements from [3] which are used in this contribution as a theoretical tool could in principle also be chosen. However, these finite element spaces involve an even higher number of unknowns, in general.

The outline of this paper is as follows. In the following section our modified least squares formulation for the linear elasticity model is introduced. Section 3 contains the ellipticity result for our new variational approach. We proceed with the finite element formulation and the results of our numerical computations on uniformly and adaptively refined triangulations in section 4. Finally, the strong connection between our new formulation and the mixed formulation of saddle point type based on the Arnold-Winther elements as well as the standard displacement formulation is established in section 5.

2. The Modified First-Order System Least Squares Formulation. We start from the equations of linear elasticity in the form

\[\begin{align*}
\text{div } \sigma &= 0, \\
\sigma - C\varepsilon(u) &= 0, \quad (2.1)
\end{align*}\]

where \(\varepsilon(u) = (\nabla u + \nabla u^T)/2\) denotes the linear strain tensor and \(C\) describes the linear material law given by

\[C\varepsilon(u) = \frac{E}{1 + \nu} \left( \varepsilon(u) + \frac{\nu}{1 - 2\nu} (\text{tr } \varepsilon(u)) I \right)\]

with the elasticity modulus \(E\) and Poisson ratio \(\nu\). The boundary of \(\Omega \subset \mathbb{R}^d\) is divided into two parts as \(\partial \Omega = \Gamma_D \cup \Gamma_N\). We consider boundary conditions of the form \(u = 0\) on \(\Gamma_D\), \(\sigma \cdot n = g\) on \(\Gamma_N\) and define

\[\begin{align*}
H^1_D(\Omega) &= \{ q \in H^1(\Omega) : q = 0 \text{ on } \Gamma_D \} \\
H_N(\text{div}, \Omega) &= \{ w \in H(\text{div}, \Omega) : w \cdot n = 0 \text{ on } \Gamma_N \}.
\end{align*}\]

Throughout this paper, the norm on \(L^2(\Omega)\) (or \(L^2(\Omega)^d, L^2(\Omega)^{d \times d}\), respectively) will be abbreviated by \(\| \cdot \|\) and the corresponding inner product by \((\cdot, \cdot)\). We assume that \(\Gamma_D\) is a subset of \(\partial \Omega\) with positive measure (length if \(d = 2\), area in the three-dimensional case) such that Korn’s inequality is valid in the form

\[\|v\|^2 + \|\nabla v\|^2 \leq C_K \|C^{1/2}\varepsilon(v)\|^2 \quad \text{for all } v \in H^1_D(\Omega)^d \quad (2.2)\]

with a constant \(C_K > 0\) (cf. [6, Section VI.3]). Note that \(C_K\) is independent of \(\nu\) due to the fact that

\[\|C^{1/2}\varepsilon(v)\|^2 \geq \frac{E}{1 + \nu} \|\varepsilon(u)\|^2 + \frac{\nu}{1 - 2\nu} \|\text{tr } \varepsilon(u)\|^2 \geq \frac{E}{1 + \nu} \|\varepsilon(u)\|^2.\]
We will frequently use the decomposition of an arbitrary matrix-valued function \( \tau \in L^2(\Omega)^{d \times d} \) into its symmetric and antisymmetric part,

\[
\tau = \text{sy} \tau + \text{as} \tau \quad \text{with} \quad \text{sy} \tau = \frac{\tau + \tau^T}{2}, \quad \text{as} \tau = \frac{\tau - \tau^T}{2}.
\]

Obviously, \((\text{sy} \tau, \text{as} \tau) = 0\) which implies

\[
\|\tau\|^2 = \|\text{sy} \tau\|^2 + \|\text{as} \tau\|^2.
\]

If \(\sigma^N \in H(\text{div}, \Omega)^d\) is chosen such that \(\sigma^N \cdot n = g\) on \(\Gamma_N\), then the solution of (2.1) may be obtained from minimizing the least squares functional

\[
G(\sigma, u) = \|\text{div} \sigma\|^2 + \|C^{-1/2} \sigma - C^{1/2} \varepsilon(u)\|^2
\]

(2.3) among all \(\sigma \in \sigma^N + H_N(\text{div}, \Omega)^d\) and \(u \in H_D^1(\Omega)^d\). This is exactly the least squares functional studied in [9]. One may also extend this functional by the redundant equation

\[
(\text{as} \tau, \text{as} \tau) = 0
\]

(cf. [10]) which leads to

\[
F(\sigma, u) = \|\text{div} \sigma\|^2 + \|C^{-1/2} \sigma - C^{1/2} \varepsilon(u)\|^2 + \|\text{as} (C^{-1/2} \sigma)\|^2.
\]

The minimization of (2.4) is equivalent to the variational problem of finding \(\sigma \in \sigma^N + H_N(\text{div}, \Omega)^d\) and \(u \in H_D^1(\Omega)^d\) such that

\[
(d \sigma, d \tau) + (\text{C}^{-1} \sigma - \varepsilon(u), \tau) + (\text{as} (\text{C}^{-1} \sigma), \text{as} \tau) = 0,
\]

\[
(\text{as} \sigma, \text{as} \varepsilon(u), \varepsilon(v)) = 0
\]

(2.5)

holds for all \(\tau \in H_N(\text{div}, \Omega)^d\) and \(v \in H_D^1(\Omega)^d\). The additional term associated with the antisymmetric stress part is not necessary in the above least squares formulation but will be required in the following modification. Replacing \(\varepsilon(v)\) by \(\nabla v\) in the displacement variation of (2.5) leads to our modified least squares method. It consists in finding \(\sigma \in \sigma^N + H_N(\text{div}, \Omega)^d\) and \(u \in H_D^1(\Omega)^d\) such that

\[
(d \sigma, d \tau) + (\text{C}^{-1} \sigma - \varepsilon(u), \tau) + (\text{as} (\text{C}^{-1} \sigma), \text{as} \tau) = 0,
\]

\[
(\text{as} \sigma, \text{as} \varepsilon(u), \nabla v) = 0
\]

(2.6)

holds for all \(\tau \in H_N(\text{div}, \Omega)^d\) and \(v \in H_D^1(\Omega)^d\).

3. Ellipticity. We consider the nonsymmetric bilinear form

\[
B(\sigma, u; \tau, v) = (d \sigma, d \tau) + (\text{C}^{-1} \sigma - \varepsilon(u), \tau - C \nabla v) + (\text{as} (\text{C}^{-1} \sigma), \text{as} \tau)
\]

(3.1)

associated with the modified least squares formulation (2.6). We will show that this bilinear form is coercive and continuous with respect to the product space \(H_N(\text{div}, \Omega)^d \times H_D^1(\Omega)^d\) equipped with the norms

\[
\|\tau\|_{\text{div}, C^{-1}} = \left(\|d \tau\|^2 + \|\text{C}^{-1/2} \tau\|^2\right)^{1/2},
\]

\[
\|v\|_{1, C} = \left(\|\text{C}^{1/2} \nabla v\|^2 + \|v\|^2\right)^{1/2}.
\]

(3.2)
For these estimates we rely on the assumption that $E$ is on the order of one. Note that this can always be achieved by a suitable rescaling of the variables. Coercivity and continuity of the bilinear form is then achieved uniformly in the incompressible limit, i.e. as $\nu \to 1/2$. This is the statement of the following theorem.

**Theorem 3.1.** There exist positive constants $\alpha$ and $\beta$ which are independent of $\nu$ such that

$$B(\sigma, u; \nu, \nu) \geq \alpha \left( \|\sigma\|_{\text{div}, C^{-1}}^2 + \|\nu\|_{1,C}^2 \right)$$  \hspace{1cm} (3.3)

holds for all $(\sigma, u) \in H_N(\text{div}, \Omega)^d \times H^1_B(\Omega)^d$ and

$$B(\sigma, u; \nu, \nu) \leq \beta \left( \|\sigma\|_{\text{div}, C^{-1}}^2 + \|\nu\|_{1,C}^2 \right)^{1/2} \left( \|\sigma\|_{\text{div}, C^{-1}}^2 + \|\nu\|_{1,C}^2 \right)^{1/2}$$  \hspace{1cm} (3.4)

holds for all $(\sigma, u) \in H_N(\text{div}, \Omega)^d \times H^1_B(\Omega)^d$.

**Proof.** The proof of (3.4) is rather straightforward. The Cauchy-Schwarz inequality immediately implies

$$B(\sigma, u; \nu, \nu) \leq \left( \|\text{div} \sigma\|^2 + \|C^{-1/2} \sigma - C^{1/2} \nu\|^2 + \|\text{as}(C^{-1/2})\|^2 \right)^{1/2}$$

$$\left( \|\text{div} \nu\|^2 + \|C^{-1/2} \nu - C^{1/2} \text{div} \nu\|^2 + \|\text{as}(C^{-1/2})\|^2 \right)^{1/2}.$$  \hspace{1cm} (3.5)

Using the orthogonal splitting of an arbitrary matrix into its symmetric and antisymmetric part, we obtain

$$\|C^{1/2} \text{div} \nu\|^2 = (C \text{ div} \nu, \text{ div} \nu) = (\text{sy}(C \text{ div} \nu), \nu) + (\text{as}(C \text{ div} \nu), \text{ div} \nu)$$

$$= (C \nu, \nu) + (\text{as} \text{ div} \nu, \text{ div} \nu)$$

$$= \|C^{1/2} \nu\|^2 + \|C^{1/2} \text{ as} \text{ div} \nu\|^2,$$

which holds due to

$$\text{sy}(C \text{ div} \nu) = \frac{E}{1+\nu} \text{sy} \text{ div} \nu + \frac{\nu}{1-2\nu}(\text{tr} \text{ div} \nu) I$$

$$= \frac{E}{1+\nu} \nu + \frac{\nu}{1-2\nu}(\text{tr} \nu) I = \nu \text{ div} \nu,$$

as $(C \text{ div} \nu) = \frac{E}{1+\nu} \text{ as} \text{ div} \nu = C \text{ as} \text{ div} \nu$.

In an analogous way, one obtains

$$\|C^{-1/2} \nu\|^2 = \|\text{as}(C^{-1/2} \nu)\|^2 + \|\text{sy}(C^{-1/2} \nu)\|^2.$$  \hspace{1cm} (3.6)

These identities imply

$$\|C^{1/2} \nu\| \leq \|C^{1/2} \text{div} \nu\|,$$

$$\|\text{as}(C^{-1/2})\| \leq \|C^{-1/2} \nu\|,$$

respectively. Using (3.6), (3.5) can be further estimated as

$$B(\sigma, u; \nu, \nu) \leq \left( \|\text{div} \sigma\|^2 + 2\|C^{-1/2} \sigma\|^2 + 2\|C^{1/2} \text{div} \nu\|^2 + \|\text{as}(C^{-1/2})\|^2 \right)^{1/2}$$

$$\left( \|\text{div} \nu\|^2 + 2\|C^{-1/2} \nu\|^2 + 2\|C^{1/2} \text{div} \nu\|^2 + \|\text{as}(C^{-1/2})\|^2 \right)^{1/2}.  \hspace{1cm} (3.7)$$
This means that (3.4) holds with $\beta = 3$.

Naturally, the coercivity estimate (3.3) is harder to establish. We start from
\[
B(\tau, v; \tau, v) = \|\text{div} \tau\|^2 + (C^{-1} \tau - \varepsilon(v), \tau - C \nabla v) + \|\text{as}(C^{-1/2} \tau)\|^2
\]
and note that all three terms on the right hand side are nonnegative. This is trivial for the first and third term. The second term constitutes a quadratic functional which clearly attains its minimum for $(\tau^*, v^*) \in H_N(\text{div}, \Omega)^d \times H^1_D(\Omega)^d$ satisfying
\[
2(C^{-1} \tau^*, \tau^*) - (\varepsilon(v^*) + \nabla v^*, \tau^*) = 0 \quad \text{for all } \tau \in H_N(\text{div}, \Omega)^d,
\]
\[
2(C \varepsilon(v^*), \varepsilon(v)) - (\tau^*, \varepsilon(v) + \nabla v) = 0 \quad \text{for all } v \in H^1_D(\Omega)^d.
\]
In particular,
\[
2(C^{-1} \tau^*, \tau^*) - (\varepsilon(v^*) + \nabla v^*, \tau^*) = 0,
\]
\[
2(C \varepsilon(v^*), \varepsilon(v^*)) - (\tau^*, \varepsilon(v^*) + \nabla v^*) = 0
\]
which implies
\[
(C^{-1} \tau^* - \varepsilon(v^*), \tau^* - C \nabla v^*) = (C^{-1} \tau^*, \tau^*) - (\tau^*, \nabla v^* + \varepsilon(v^*)) + (C \varepsilon(v^*), v^*) = 0.
\]
With a constant $C_1 \geq 1$ which is still free to be chosen appropriately, we therefore have
\[
B(\tau; v; \tau, v) \geq \frac{1}{C_1} \left(\|\text{div} \tau\|^2 + (C^{-1} \tau - \varepsilon(v), \tau - C \nabla v) + C_1 \|\text{as}(C^{-1/2} \tau)\|^2\right)
\]
\[
= \frac{1}{C_1} \left(\|\text{div} \tau\|^2 + (C^{-1} \tau - \varepsilon(v), \tau - C \varepsilon(v)) - (C^{-1} \tau - \varepsilon(v), \nabla v) + C_1 \|\text{as}(C^{-1/2} \tau)\|^2\right)
\]
\[
= \frac{1}{C_1} \left(\|\text{div} \tau\|^2 + ||C^{-1/2} \tau - C^{1/2} \varepsilon(v)||^2 - (\text{as}(C^{-1/2} \tau), \text{as}(C^{1/2} \nabla v)) + C_1 ||\text{as}(C^{-1/2} \tau)||^2\right).
\]
We will rely on the coercivity proof for the symmetric bilinear form from [9, Theorem 2.1] which states that, with a positive constant $C_0$ (independently of $\nu$),
\[
\mathcal{G}(\tau, v) \geq C_0 \left(\|\text{div} \tau\|^2 + ||C^{-1/2} \tau||^2 + ||C^{1/2} \varepsilon(v)||^2\right)
\]
holds for all $(\tau, v) \in H_N(\text{div}, \Omega)^d \times H^1_D(\Omega)^d$. From Korn’s inequality (2.2) we deduce
\[
\|\text{as}(C^{1/2} \nabla v)\|^2 = (\text{as}(C \nabla v), \nabla v) = \frac{E}{1 + \nu} \|\nabla v\|^2 \leq \frac{E}{1 + \nu} \|\nabla v\|^2
\]
\[
\leq \frac{C_K E}{1 + \nu} \|\varepsilon(v)\|^2 \leq \frac{C_K E}{1 + \nu} \left(\|\varepsilon(v)\|^2 + \frac{\nu}{1 - 2\nu} \|\text{div} v\|^2\right)
\]
\[
= C_K (\varepsilon(v), \varepsilon(v)) = C_K \|C^{1/2} \varepsilon(v)\|^2.
\]
Combining (3.10) with (3.9) leads to
\[
\|\text{as}(C^{1/2} \nabla v)\|^2 \leq \frac{C_K}{C_0} \mathcal{G}(\tau, v).
\]
Therefore, \((3.8)\) can be further bounded from below as follows:

\[
\mathcal{B}(\tau, v; \tau, v) \geq \frac{1}{C_1} \left( \|\text{div}\tau\|^2 + \|C^{-1/2}\tau - C^{1/2}\varepsilon(v)\|^2 + C_1 \|\text{as}\ C^{-1/2}\tau\|^2 \\
- \left(\text{as}\ (C^{-1/2}\tau), \text{as}\ (C^{1/2}\nabla v))\right) \right)
\]

\[
\geq \frac{1}{C_1} \left( \|\text{div}\ \tau\|^2 + \|C^{-1/2}\tau - C^{1/2}\varepsilon(v)\|^2 + C_1 \|\text{as}\ (C^{-1/2}\tau)\|^2 \\
- C_1 \|\text{as}\ (C^{-1/2}\tau)\|^2 - \frac{1}{4C_1} \|\text{as}\ (C^{1/2}\nabla v)\|^2 \right)
\]

\[
= \frac{1}{C_1} \left( \mathcal{G}(\tau, v) - \frac{1}{4C_1} \|\text{as}\ (C^{1/2}\nabla v)\|^2 \right)
\]

\[
\geq \frac{1}{C_1} \left( 1 - \frac{C_K}{4C_0C_1} \right) \mathcal{G}(\tau, v).
\]

If we choose \(C_1 = \max\{1, C_K/(2C_0)\}\), we have \(C_K/(4C_0C_1) \leq 1/2\) and therefore

\[
\mathcal{B}(\tau, v; \tau, v) \geq \frac{1}{2C_1} \mathcal{G}(\tau, v). \tag{3.13}
\]

Combined with (3.9) and (2.2), this leads to

\[
\mathcal{B}(\tau, v; \tau, v) \geq \frac{C_0}{2C_1} \left( \|\text{div}\ \tau\|^2 + \|C^{-1/2}\tau\|^2 + \frac{1}{C_K} \left( \|\tau\|^2 + \|C^{1/2}\nabla v\|^2 \right) \right) \tag{3.14}
\]

which finishes the proof. \(\square\)

The constants \(\alpha\) and \(\beta\) in Theorem 1 are independent of \(\nu\) assuming that \(E\) is on the order of 1. This means that our coercivity and continuity estimates hold uniformly in the incompressible limit as \(\nu \to 1/2\). Coercivity and continuity could also be established with respect to unscaled norms

\[
\|\cdot\|_{\text{div}} := \|\cdot\|_{\text{div}, 1} = \left( \|\text{div}\ (\cdot)\|^2 + \|\cdot\|^2 \right)^{1/2},
\]

\[
\|\cdot\|_1 := \|\cdot\|_{1, 1} = \left( \|
abla\ (\cdot)\|^2 + \|\cdot\|^2 \right)^{1/2}.
\]

However, these estimates would not be uniform in the incompressible limit.

The fact that (3.3) and (3.4) hold ensures that our modified least squares formulation is well-posed, i.e. the variational problem (2.6) has a unique solution which depends on a continuous way on the boundary data.

4. **Finite Element Approximation.** For the finite element approximation of the variational problem (2.6), we choose finite-dimensional subspaces \(\Sigma_h \subset \Sigma := H_N(\text{div}, \Omega)^d\) and \(V_h \subset V := H_D^1(\Omega)^d\). The discrete variational problem consists in finding \(\sigma_h = \sigma^* + \Sigma_h\) and \(u_h \in V_h\) such that

\[
(\text{div} \sigma_h, \text{div} \tau_h) + (C^{-1}\sigma_h - \varepsilon(u_h), \tau_h - C
abla v_h) + (\text{as}\ C^{-1}\sigma_h, \text{as} \tau_h) = 0 \tag{4.1}
\]

holds for all \(\tau_h \in \Sigma_h\) and \(v_h \in V_h\). Combined with (2.6), this implies that

\[
(\text{div} (\sigma - \sigma_h), \text{div} \tau_h) + (C^{-1}(\sigma - \sigma_h) - \varepsilon(u - u_h), \tau_h - C \nabla v_h)
+ (\text{as}\ C^{-1}(\sigma - \sigma_h), \text{as}\ \tau_h) = 0 \tag{4.2}
\]

is satisfied for all \(\tau_h \in \Sigma_h\) and \(v_h \in V_h\).
In our computational results, next-to-lowest order Raviart-Thomas elements \( (RT_1) \) for the stress space \( \Sigma_h \) are combined with conforming quadratic elements \( (P_2) \) for the displacement space \( V_h \).

**Example 1.** We consider Cook’s membrane problem with corners \((0, 0), (48, 44), (48, 60) \) and \((0, 44)\) which is clamped at the left boundary segment (see figure 4.1). The rest of the boundary is subjected to surface traction forces which are zero on the upper and lower boundary segments and pointing upwards on the right boundary segment. We consider an almost incompressible material with Poisson ratio \( \nu = 0.49 \).

<table>
<thead>
<tr>
<th>( l )</th>
<th># elem.</th>
<th>( \dim \Sigma_h )</th>
<th>( \dim V_h )</th>
<th>( \mathcal{F}(\sigma_h, u_h) )</th>
<th>( | \text{div} \sigma_h |^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>40</td>
<td>72</td>
<td>12.729</td>
<td>2.640</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>144</td>
<td>304</td>
<td>5.402</td>
<td>2.125</td>
</tr>
<tr>
<td>2</td>
<td>128</td>
<td>544</td>
<td>1248</td>
<td>2.642</td>
<td>1.240</td>
</tr>
<tr>
<td>3</td>
<td>512</td>
<td>2112</td>
<td>5056</td>
<td>1.206</td>
<td>0.566</td>
</tr>
<tr>
<td>4</td>
<td>2048</td>
<td>8320</td>
<td>20352</td>
<td>0.617</td>
<td>0.387</td>
</tr>
<tr>
<td>5</td>
<td>8192</td>
<td>33024</td>
<td>81664</td>
<td>0.421</td>
<td>0.334</td>
</tr>
</tbody>
</table>

**Table 4.1**
Symmetric least squares formulation (2.5): \( RT_1 \times P_2 \), uniform refinement

<table>
<thead>
<tr>
<th>( l )</th>
<th># elem.</th>
<th>( \dim \Sigma_h )</th>
<th>( \dim V_h )</th>
<th>( \mathcal{F}(\sigma_h, u_h) )</th>
<th>( | \text{div} \sigma_h |^2 )</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>40</td>
<td>72</td>
<td>19.145</td>
<td>4.527</td>
</tr>
<tr>
<td>1</td>
<td>32</td>
<td>144</td>
<td>304</td>
<td>9.421</td>
<td>2.336</td>
</tr>
<tr>
<td>2</td>
<td>128</td>
<td>544</td>
<td>1248</td>
<td>4.468</td>
<td>0.901</td>
</tr>
<tr>
<td>3</td>
<td>512</td>
<td>2112</td>
<td>5056</td>
<td>1.631</td>
<td>0.499</td>
</tr>
<tr>
<td>4</td>
<td>2048</td>
<td>8320</td>
<td>20352</td>
<td>0.753</td>
<td>0.374</td>
</tr>
<tr>
<td>5</td>
<td>8192</td>
<td>33024</td>
<td>81664</td>
<td>0.527</td>
<td>0.289</td>
</tr>
</tbody>
</table>

**Table 4.2**
Nonsymmetric least squares formulation (2.6): \( RT_1 \times P_2 \), uniform refinement

A comparison of the results in tables 4.1 and 4.2 shows that, on the same triangulations resulting from uniform refinement, the nonsymmetric formulation produces a faster decrease of the momentum balance error. Note that the overall least squares functional must be smaller for the symmetric formulation (on identical triangulations) since it is actually minimized. However, the higher reduction rate of the momentum balance error is not very pronounced in this example.

We also compare these two formulations on a sequence of adaptively refined triangulations based on the a posteriori error estimator provided by the local evaluation of the functional \( \mathcal{F}(\sigma_h, u_h) \). A fixed proportion of elements with largest local contribution to the functional is refined in our adaptive approach (about 15% of elements in our computations).

The results in tables 4.3 and 4.4 show that on adaptively refined triangulations, the difference between the two formulations is much more striking. The reduction rate of the momentum balance term in the functional is significantly faster in the nonsymmetric formulation. Surprisingly, even the least squares functional on the finest refinement level is smaller for the nonsymmetric formulation. This means that these refined triangulations are better suited to our problem.

The adaptively refined triangulations as well as the deformed domain are shown in figure 4.1. The most striking difference is that the nonsymmetric formulation (on
<table>
<thead>
<tr>
<th>$l$</th>
<th># elem.</th>
<th>dim $\Sigma_h$</th>
<th>dim $V_h$</th>
<th>$\mathcal{F}(\sigma_h, u_h)$</th>
<th>$|\text{div} \sigma_h|^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>40</td>
<td>72</td>
<td>12.729</td>
<td>2.040</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>90</td>
<td>176</td>
<td>5.476</td>
<td>2.048</td>
</tr>
<tr>
<td>2</td>
<td>43</td>
<td>194</td>
<td>408</td>
<td>2.584</td>
<td>1.257</td>
</tr>
<tr>
<td>3</td>
<td>90</td>
<td>396</td>
<td>864</td>
<td>1.627</td>
<td>0.975</td>
</tr>
<tr>
<td>4</td>
<td>184</td>
<td>796</td>
<td>1780</td>
<td>1.061</td>
<td>0.718</td>
</tr>
<tr>
<td>5</td>
<td>372</td>
<td>1572</td>
<td>3636</td>
<td>0.748</td>
<td>0.543</td>
</tr>
<tr>
<td>6</td>
<td>732</td>
<td>3044</td>
<td>7204</td>
<td>0.519</td>
<td>0.394</td>
</tr>
<tr>
<td>7</td>
<td>1359</td>
<td>5586</td>
<td>13440</td>
<td>0.413</td>
<td>0.323</td>
</tr>
<tr>
<td>8</td>
<td>2440</td>
<td>9960</td>
<td>24200</td>
<td>0.341</td>
<td>0.253</td>
</tr>
<tr>
<td>9</td>
<td>4351</td>
<td>17634</td>
<td>43280</td>
<td>0.267</td>
<td>0.162</td>
</tr>
<tr>
<td>10</td>
<td>7568</td>
<td>30600</td>
<td>75352</td>
<td>0.185</td>
<td>0.082</td>
</tr>
</tbody>
</table>

Table 4.3: Symmetric least squares formulation (2.5): $RT_1 \times P_2$, adaptive refinement

<table>
<thead>
<tr>
<th>$l$</th>
<th># elem.</th>
<th>dim $\Sigma_h$</th>
<th>dim $V_h$</th>
<th>$\mathcal{F}(\sigma_h, u_h)$</th>
<th>$|\text{div} \sigma_h|^2$</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>8</td>
<td>40</td>
<td>72</td>
<td>19.145</td>
<td>4.527</td>
</tr>
<tr>
<td>1</td>
<td>19</td>
<td>90</td>
<td>176</td>
<td>10.773</td>
<td>3.361</td>
</tr>
<tr>
<td>2</td>
<td>41</td>
<td>186</td>
<td>388</td>
<td>4.595</td>
<td>1.874</td>
</tr>
<tr>
<td>3</td>
<td>87</td>
<td>382</td>
<td>836</td>
<td>2.558</td>
<td>1.010</td>
</tr>
<tr>
<td>4</td>
<td>190</td>
<td>812</td>
<td>1848</td>
<td>1.277</td>
<td>0.548</td>
</tr>
<tr>
<td>5</td>
<td>370</td>
<td>1560</td>
<td>3620</td>
<td>0.796</td>
<td>0.379</td>
</tr>
<tr>
<td>6</td>
<td>710</td>
<td>2948</td>
<td>6992</td>
<td>0.592</td>
<td>0.269</td>
</tr>
<tr>
<td>7</td>
<td>1346</td>
<td>5528</td>
<td>13316</td>
<td>0.476</td>
<td>0.151</td>
</tr>
<tr>
<td>8</td>
<td>2464</td>
<td>10044</td>
<td>24452</td>
<td>0.360</td>
<td>0.065</td>
</tr>
<tr>
<td>9</td>
<td>4415</td>
<td>17902</td>
<td>43908</td>
<td>0.230</td>
<td>0.020</td>
</tr>
<tr>
<td>10</td>
<td>7774</td>
<td>31408</td>
<td>77428</td>
<td>0.129</td>
<td>0.005</td>
</tr>
</tbody>
</table>

Table 4.4: Nonsymmetric least squares formulation (2.6): $RT_1 \times P_2$, adaptive refinement

the right) puts much more emphasis at refining near the upper left corner.

**Example 2.** We consider the more regular Cook's membrane problem with corners
$(0,0), (48,14), (48,30)$ and $(0,44)$ where the singularities are much less severe. This is
a good illustration of our analysis in the next section which will rely on the regularity
of the underlying boundary value problem. It will indicate that the decrease of the
momentum balance error relative to the overall error will be more pronounced the
more regular the problem is. We repeat our computations for this more regular
modification of Cook's membrane on a sequence of uniformly refined triangulations.

The results in table 4.6 for the nonsymmetric formulation clearly show a faster
reduction of the momentum balance error which the numbers in table 4.5 for the
symmetric formulation do not.

**5. A Strong Connection To Mixed and Displacement Formulations.** In
[7], Brandts, Chen and Yang have established a strong connection for second-order
elliptic problems between the least squares formulation and corresponding mixed and
standard Galerkin approaches. Such a strong connection also holds for the linear elas-
ticity problem between our modified variational formulation (2.6) and an appropriate
mixed and the standard displacement-based approach. This result is the basis for
our proof of the improved momentum balance observed in the computational results shown in Section 4.

In our analysis, we will use the finite element spaces introduced by Arnold and
Winther in [3] which are defined as follows for $k \geq 1$:

$$\Sigma_{h}^{AW} = \{ \tau \in \Sigma : \tau = 0 , \quad \tau|_{T} \in \mathcal{P}_{k+2}^{2} , \quad \text{div} \tau|_{T} \in \mathcal{P}_{k}^{1} \}, \quad (5.1)$$

where $\mathcal{P}_{k}$ denotes the space of polynomials of degree $k$ (in two dimensions). Important for our consideration is the fact that $\text{div} \Sigma_{h}^{AW}$ coincides with the space of piecewise polynomials of degree $k$ (without any inter-element continuity conditions). We restrict our analysis to two dimensional problems.

For some fixed integer $k \geq 1$, let $\Sigma_{h}^{AW} \subseteq \Sigma$ denote the Arnold-Winther space of degree $k$ introduced above. Moreover, let $V_{h} \subseteq V$ be the standard $H^{1}(\Omega)$-conforming finite element space of piecewise polynomials of degree $k + 1$. Finally, denote by $X_{h} \subseteq L^{2}(\Omega)^{2}$ the piecewise polynomial spaces of degree $k$ without any continuity requirements.

Let us denote the finite element approximation based on the variational formulation (4.1) with $\Sigma_{h} = \Sigma_{h}^{AW}$ by $(\sigma_{h}^{AW}, u_{h}^{AW})$. Note that $(\sigma_{h}^{AW}, u_{h}^{AW}) \in (\sigma^{N} + \Sigma_{h}^{AW}) \times V_{h}$ satisfies $\sigma_{h} = 0$ and therefore also minimizes the least squares functional $G(\sigma_{h}, u_{h})$ with respect to $\Sigma_{h}^{AW} \times V_{h}$. The first step of our analysis consists in establishing a close connection of $(\sigma_{h}^{AW}, u_{h}^{AW})$ to the following mixed and standard displacement approaches.

The mixed formulation that we consider is based on Hellinger-Reissner principle and uses the finite element space by Arnold and Winther [3]. The mixed finite element approximation $(\sigma_{h}^{m}, \xi_{h}^{m}) \in (\sigma^{N} + \Sigma_{h}^{AW}) \times X_{h}$ with these spaces satisfy

$$\begin{align*}
(C^{-1}(\sigma_{h}^{m}, \tau_{h}) + (u - \xi_{h}^{m}, \text{div} \tau_{h}) &= 0 \quad (5.2)) \quad \text{for all } \tau_{h} \in \Sigma_{h}^{AW} \quad \text{and}

\text{(div} \sigma_{h}^{m}, \eta_{h}) &= 0 \quad \text{for all } \eta_{h} \in X_{h} \quad . \quad (5.3)
\end{align*}$$

For the combination $\Sigma_{h}^{AW} \times X_{h}$ introduced above (Arnold-Winther space combined with piecewise polynomials of degree $k$) the first constraint in (5.3) implies that $\text{div} \sigma_{h}^{m} = 0$ holds exactly in $L^{2}(\Omega)$. Finally, the standard displacement formulation $u_{h}^{d} \in V_{h}$ satisfies

$$\begin{align*}
(C\varepsilon(u - u_{h}^{d}), \varepsilon(v_{h})) &= 0 \quad (5.4) \quad \text{for all } v_{h} \in V_{h}.
\end{align*}$$

We are now ready to prove the generalization of the Brandts-Chen-Yang formula (cf. [7, Lemma 3.2]) to our modified least squares formulation (2.6).

**Lemma 5.1.** For some integer $k \geq 1$, let $\Sigma_{h}^{AW}$ denote the Arnold-Winther space of degree $k$ and let $V_{h}$ denote the standard continuous elements of degree $k + 1$. For the finite element approximation $(\sigma_{h}^{AW}, u_{h}^{AW}) \in (\sigma^{N} + \Sigma_{h}^{AW}) \times V_{h}$ based on (4.1), the mixed approximation $\sigma_{h}^{m} \in \sigma^{N} + \Sigma_{h}^{AW}$ based on (5.2) and (5.3), and the displacement approximation $u_{h}^{d} \in V_{h}$ based on (5.4), the following relation holds:

$$\begin{align*}
E(\sigma_{h}^{AW} - \sigma_{h}^{m}, u_{h}^{AW} - u_{h}^{d}, \sigma_{h}^{AW} - \sigma_{h}^{m}, u_{h}^{AW} - u_{h}^{d})
&= (C^{-1}(\sigma_{h}^{m}, \sigma_{h}^{AW} - \sigma_{h}^{m}) - (\varepsilon(u - u_{h}^{d}), \sigma_{h}^{AW} - \sigma_{h}^{m})). \quad (5.5)
\end{align*}$$
Proof. The definition of the bilinear form (3.1) leads to
\[
\mathcal{B}(\sigma_h^{AW} - \sigma_h^m, u_h^{AW} - u_h^d, \sigma_h^{AW} - \sigma_h^m, u_h^{AW} - u_h^d) \\
= \|\text{div} (\sigma_h^{AW} - \sigma_h^m)\|^2 \\
+ (C^{-1}(\sigma_h^{AW} - \sigma_h^m) - \varepsilon(u_h^{AW} - u_h^d), \sigma_h^{AW} - \sigma_h^m - C \nabla (u_h^{AW} - u_h^d)) \\
+ (\text{as}(\sigma_h^{AW} - \sigma_h^m), (\sigma_h^{AW} - \sigma_h^m)) \\
= (\text{div} (\sigma_h^{AW} - \sigma_h^m)) \\
+ (C^{-1}(\sigma_h^{AW} - \sigma_h^m) - \varepsilon(u_h^{AW} - u_h^d), \sigma_h^{AW} - \sigma_h^m - C \nabla (u_h^{AW} - u_h^d)) \\
+ (\text{as}(\sigma_h^{AW} - \sigma_h^m), (\sigma_h^{AW} - \sigma_h^m)) \\
= (C^{-1}(\sigma - \sigma^m_h) - \varepsilon(u_h^{AW} - u_h^d), \sigma_h^{AW} - \sigma_h^m - C \nabla (u_h^{AW} - u_h^d)) \\
+ (\text{as}(\sigma - \sigma^m_h), (\sigma_h^{AW} - \sigma_h^m)) \\
= (C^{-1}(\sigma - \sigma^m_h), \sigma_h^{AW} - \sigma_h^m - \text{div} (u_h^{AW} - u_h^d)) \\
- (\varepsilon(u_h^{AW} - u_h^d), \sigma_h^{AW} - \sigma_h^m) \\
+ (\text{as}(\sigma - \sigma^m_h), (\sigma_h^{AW} - \sigma_h^m)),
\]
where we used (4.2) with \(\Sigma_h = \Sigma_h^{AW}\) in the last equality (note that \(\sigma_h^{AW} - \sigma_h^m \in \Sigma_h^{AW}\) and \(u_h^{AW} - u_h^d \in V_h\)). The second and fourth terms in (5.6) vanish due to (5.3) and (5.4), respectively. The last term vanishes since symmetry is imposed in the stress space \(\Sigma_h^{AW}\). This leads us to the desired generalization of the Brandts-Chen-Yang formula (5.5).\[\Box\]

In the sequel, we will use discrete inverses of the divergence operator with respect to the spaces \(\Sigma_h^{RT} \subseteq \Sigma\) and \(\Sigma_h^{AW} \subseteq \Sigma\) (cf. [14, Sect. 2]). To this end, for any \(\xi_h \in X_h\), we may define \(\mathbb{E}_h^{AW} \in \Sigma_h^{AW}\) to be the solution of the saddle point problem
\[
(C^{-1}\mathbb{E}_h^{AW}, \tau_h) - (\xi_h, \text{div} \tau_h) = 0 \quad \text{for all } \tau_h \in \Sigma_h^{AW}, \\
(\text{div} \mathbb{E}_h^{AW}, \eta_h) = (\xi_h, \eta_h) \quad \text{for all } \eta_h \in X_h. 
\]
(5.7)

Similarly, we may define \(\mathbb{E}_h^{RT} \in \Sigma_h^{RT}\) to be the solution of
\[
(C^{-1}\mathbb{E}_h^{RT}, \tau_h) + (\text{as}(C^{-1}\mathbb{E}_h^{RT}), as \tau_h) - (\xi_h, \text{div} \tau_h) = 0 \quad \text{for all } \tau_h \in \Sigma_h^{RT}, \\
(\text{div} \mathbb{E}_h^{RT}, \eta_h) = (\xi_h, \eta_h) \quad \text{for all } \eta_h \in X_h. 
\]
(5.8)

If we define subspaces
\[
\Sigma_h^{RT/\text{AW}, 0} = \{\tau_h \in \Sigma_h^{RT/\text{AW}} : \text{div} \tau_h = 0\} 
\]
(5.9)
and the inner product \((\cdot, \cdot)_{C^{-1}, as}\) on \(\Sigma\) given by
\[
(\sigma, \tau)_{C^{-1}, as} = (C^{-1} \sigma, \tau) + (\text{as}(C^{-1} \sigma), as \tau),
\]
(5.10)
we obtain \((\mathbb{E}_h^{AW}, \tau_h)_{C^{-1}, as} = 0 \text{ for all } \tau_h \in \Sigma_h^{AW}, 0\) and, similarly, \((\mathbb{E}_h^{RT}, \tau_h)_{C^{-1}, as} = 0 \text{ for all } \tau_h \in \Sigma_h^{RT}, 0\). Furthermore, we define orthogonal subspaces
\[
\Sigma_h^{RT/\text{AW}, \perp} = \{\tau_h \in \Sigma_h^{RT/\text{AW}} : (\tau_h, \omega_h)_{C^{-1}, as} = 0 \text{ for all } \omega_h \in \Sigma_h^{RT/\text{AW}, 0}\}.
\]
(5.11)

The construction in (5.7) and (5.8) provides discrete inverses of the divergence operator mapping \(\Sigma_h^{RT/\text{AW}, \perp}\), respectively, onto \(X_h\). We therefore use the notation
\[
\mathbb{E}_h^{RT} = \text{div}^{-1}_R \xi_h \text{ and } \mathbb{E}_h^{AW} = \text{div}^{-1}_A \xi_h. 
\]
For any \( \xi \in L^2(\Omega)^2 \), we may proceed in any similar way as in (5.7) and (5.8) with respect to the entire Sobolev space \( \Sigma = H_N(\text{div}, \Omega)^2 \). To this end, let \( \Xi \in \Sigma \) be the solution of

\[
(\mathcal{L}^{-1} \Xi, \tau) + (\text{as}(\mathcal{L}^{-1} \Xi), \text{as} \tau) - (\xi, \text{div} \tau) = 0 \quad \text{for all } \tau \in \Sigma,
\]

\[
(\text{div} \Xi, \eta) = (\xi, \eta) \quad \text{for all } \eta \in L^2(\Omega)^2.
\]

This defines the inverse of the divergence operator mapping the orthogonal subspace \( \Sigma^\perp \) of \( \Sigma \) with respect to \( \Sigma_0 = \{ \tau \in \Sigma : \text{div} \tau = 0 \} \) onto \( L^2(\Omega)^2 \), \( \Xi = \text{div}^{-1} \xi \).

From now on, we will use the notation \( \iota_h \lesssim \kappa_h \) to denote that \( \iota_h \leq C \kappa_h \) holds with some constant \( C \) independently of \( h \). Some fundamental properties of the operators introduced in (5.7) and (5.8) are summarized in the following lemma. The statements of Lemma 5.2 are well-known and, in the case of the Arnold-Winther spaces, also used in [14]. We state them explicitly here for the sake of better readability.

**Lemma 5.2.** Under the assumptions that the problem (5.13) satisfies the regularity assumption

\[
\| \Xi \|_{1, \Omega} \lesssim \| \xi \|_1,
\]

the following holds for the discrete inverses of the divergence operator defined in (5.7) and (5.8):

\[
\| \text{div}^{-1} \xi_h - \text{div}^{-1} \text{RT}_h \xi_h \| \lesssim h \| \xi_h \|, \quad \| \text{div}^{-1} \text{RT}_h \xi_h \| \lesssim \| \xi_h \| \tag{5.16}
\]

for all \( \xi_h \in X_h \) and

\[
\| \text{div}^{-1} \xi_h - \text{div}^{-1} \text{AW}_h \xi_h \| \lesssim h \| \xi_h \|, \quad \| \text{div}^{-1} \text{AW}_h \xi_h \| \lesssim \| \xi_h \| \tag{5.17}
\]

for all \( \xi_h \in X_h \).

**Proof.** The regularity assumption (5.15) implies

\[
\| \Xi \|_{1, \Omega} \lesssim \| \text{div} \Xi \|.
\]

The standard approximation properties of Raviart-Thomas spaces (see [8, Prop. 3.9]) lead to

\[
\| \Xi - \Xi_h^{\text{RT}} \| \lesssim h \| \Xi \|_{1, \Omega}
\]

which, combined with (5.18), proves (5.16). Finally, the first approximation property stated in [3, Theorem 6.1] implies

\[
\| \Xi - \Xi_h^{\text{AW}} \| \lesssim h \| \Xi \|_{1, \Omega}
\]

which, combined with (5.18), leads to (5.17). \( \square \)

The next step consists in showing that the momentum balance error of the formulation (4.1) using Arnold-Winther elements converges at a faster rate than the error norm measured by the least squares functional.
Theorem 5.3. Let $\sigma^A_h, \sigma^m_h, u^A_h, u^d_h$ be as in Lemma 5.1 and assume that (5.15) is satisfied. Furthermore, assume that the boundary value problem associated with (2.1) is $H^2$-regular in the sense that, for any $g \in H^{1/2}(\Gamma_N)^2$,

$$u \in H^2(\Omega)^2 \text{ with } \|u\|_{2,\Omega} \lesssim \|g\|_{1/2,\Gamma_N}$$

(5.19)

holds. Then,

$$\|\text{div } \sigma^A_h\| \lesssim h \left(\|C^{-1/2}(\sigma - \sigma^m_h)\|^2 + \|C^{1/2}(\varepsilon(u) - \varepsilon(u^d_h))\|^2\right)^{1/2}$$

(5.20)

and

$$\|\text{div } \sigma^A_h\| \lesssim h \mathcal{F}(\sigma^A_h, u^A_h)^{1/2}.$$  

(5.21)

Proof. From (5.5) and (3.3), we obtain (note that $\text{div } \sigma^m_h = 0$)

$$\|\text{div } \sigma^A_h\|^2 = \|\text{div } (\sigma^A_h - \sigma^m_h)\|^2 \lesssim (C^{-1}(\sigma - \sigma^m_h), \sigma^A_h - \sigma^m_h) - (\varepsilon(u - u^d_h), \sigma^A_h - \sigma^m_h).$$

(5.22)

For the first term in (5.22), (5.2) leads to

$$(C^{-1}(\sigma - \sigma^m_h), \sigma^A_h - \sigma^m_h) = (u - \xi^\text{h}_m, \text{div } (\sigma^A_h - \sigma^m_h))$$

$$= (u - \xi^\text{h}_m, \text{div } \sigma^A_h) = (u - \xi^\text{h}_m, \text{div } \Xi^\text{AW}_h)$$

(5.23)

$$= (C^{-1}(\sigma - \sigma^m_h), \Xi^\text{AW}_h),$$

where $\Xi^\text{AW}_h \in \Sigma^\text{AW}_h$ is the solution of (5.7) with $\xi^\text{h} = \text{div } \sigma^A_h$ (note that $\text{div } \Xi^\text{AW}_h = \text{div } \sigma^A_h$ holds). We also have

$$(C^{-1}(\sigma - \sigma^m_h), \Xi) = (C^{-1}(\sigma - \sigma^m_h), \Xi) + (C^{-1}(\sigma - \sigma^m_h), \Xi) = 0,$$

(5.24)

if $\Xi \in \Sigma$ is the solution of (5.13) with $\xi = \text{div } \sigma^A_h$ (note that $\text{div } (\sigma - \sigma^m_h) = 0$ and as $\sigma - \sigma^m_h$ = 0). The combination of (5.23) and (5.24) leads to

$$(C^{-1}(\sigma - \sigma^m_h), \sigma^A_h - \sigma^m_h) = (C^{-1}(\sigma - \sigma^m_h), \Xi^\text{AW}_h - \Xi) \leq \|C^{-1/2}(\sigma - \sigma^m_h)\| \|\Xi^\text{AW}_h - \Xi\|$$

(5.25)

$$\lesssim h \|\text{div } \sigma^A_h\|,$$

where $\|C^{-1/2}(\Xi - \Xi^\text{AW}_h)\| \lesssim h\|\text{div } \sigma^A_h\|$ is used which holds due to Lemma 5.2.

For the second term in (5.22), integration by parts gives

$$- (\varepsilon(u - u^d_h), \sigma^A_h - \sigma^m_h) = (u - u^d_h, \text{div } (\sigma^A_h - \sigma^m_h)) = (u - u^d_h, \text{div } \sigma^A_h),$$

where as $(\sigma^A_h - \sigma^m_h) = 0$ was used. This leads to

$$- (\varepsilon(u - u^d_h), \sigma^A_h - \sigma^m_h) \leq \|u - u^d_h\| \|\text{div } \sigma^A_h\|$$

(5.26)

$$\lesssim h \|C^{1/2} \varepsilon(u - u^d_h)\| \|\text{div } \sigma^A_h\|,$$

where the estimate $\|u - u^d_h\| \lesssim h \|C^{1/2} \varepsilon(u - u^d_h)\|$ was used which holds due to the classical Aubin-Nitsche duality argument and (2.2) under our assumption of $H^2$ regularity for $u$. The combination of (5.25) and (5.26) proves (5.20).
The estimate (5.21) is obtained from (5.20) in the following way: The relation (5.2) implies
\[ \|C^{-1/2}(\sigma - \sigma_h^m)\|^2 \leq \|C^{-1/2}(\sigma - \tau_h)\|^2 \]
for all \( \tau_h \in \Sigma_{h,AW} \) and therefore
\[
\|C^{-1/2}(\sigma - \sigma_h^m)\|^2 \leq \|C^{-1/2}(\sigma - (\sigma_h^{AW} - \text{div}_h^{AW} \text{div} \sigma_h^{AW}))\|^2 \\
\leq 2 \left( \|C^{-1/2}(\sigma - \sigma_h^{AW})\|^2 + \|\text{div}_h^{AW} \text{div} \sigma_h^{AW}\|^2 \right) \\
\leq 2 \left( \|C^{-1/2}(\sigma - \sigma_h^{AW})\|^2 + \|\text{div} \sigma_h^{AW}\|^2 \right) \\
= 2 \left( \|C^{-1/2}(\sigma - \sigma_h^{AW})\|^2 + \|\text{div} (\sigma - \sigma_h^{AW})\|^2 \right),
\]
(5.27)
where Lemma 5.2 is used again. Similarly, the relation (5.4) leads to
\[ \|C^{1/2}(\varepsilon(u) - \varepsilon(u_h^d))\|^2 \leq \|C^{1/2}(\varepsilon(u) - \varepsilon(u_h^{AW}))\|^2 \]
for all \( v_h \in V_h \) which implies
\[ \|C^{1/2}(\varepsilon(u) - \varepsilon(u_h^d))\|^2 \leq \|C^{1/2}(\varepsilon(u) - \varepsilon(u_h^{AW}))\|^2. \]
(5.28)
From (5.27), (5.28), together with (3.9),
\[ \|C^{-1/2}(\sigma - \sigma_h^m)\|^2 + \|C^{1/2}(\varepsilon(u) - \varepsilon(u_h^d))\|^2 \leq \mathcal{G}(\sigma_h^{AW}, u_h^{AW}) = \mathcal{F}(\sigma_h^{AW}, u_h^{AW}) \]
(5.29)
is obtained which finishes the proof of (5.21).

**Theorem 5.4.** Let \( (\sigma_h^{RT}, u_h^{RT}) \in (\sigma^N + \Sigma_h^{RT}) \times V_h \) and \( (\sigma_h^{AW}, u_h^{AW}) \in (\sigma^N + \Sigma_h^{AW}) \times V_h \) be the approximations of our modified least squares method (4.1) in the respective finite element spaces. Then, under the assumption that (5.15) holds,
\[ \|\text{div} \sigma_h^{RT}\| \leq \|\text{div} \sigma_h^{AW}\| + h \mathcal{F}(\sigma_h^{RT}, u_h^{RT})^{1/2} \]
(5.30)

**Proof.** We start with the observation that \( \text{div} \sigma_h^{AW} \) as well as \( \text{div} \sigma_h^{RT} \) are contained in \( X_h \). Let
\[
\text{div}_h^{AW} \text{div} \sigma_h^{AW} = P_h^{AW} \sigma_h^{AW}, \\
\text{div}_h^{RT} \text{div} \sigma_h^{RT} = P_h^{RT} \sigma_h^{RT}
\]
(5.31)
be the orthogonal projections with respect to \( \langle \cdot, \cdot \rangle_c^{-1,as} \) onto \( \Sigma_h^{AW,\perp} \) and \( \Sigma_h^{RT,\perp} \), respectively. Plugging the special test functions \( \tau_h = \text{div}_h^{AW} \xi_h \in \Sigma_h^{AW,\perp} \) with \( \xi_h \in X_h \) into (4.1) leads to
\[ \left( \text{div} \sigma_h^{AW}, \xi_h \right) + \left( \sigma_h^{AW}, \text{div}_h^{AW} \xi_h \right)_{C^{-1,as}} - \left( \varepsilon(u_h^{AW}), \text{div}_h^{AW} \xi_h \right) = 0, \]
\[ \left( \sigma_h^{AW} - \mathcal{C} \varepsilon(u_h^{AW}), \nabla v_h \right) = 0 \]
(5.32)
for all \( \xi_h \in X_h \) and \( v_h \in V_h \). Similarly, setting \( \tau_h = \text{div}_h^{RT} \xi_h \in \Sigma_h^{RT,\perp} \) in (4.1), we obtain
\[ \left( \text{div} \sigma_h^{RT}, \xi_h \right) + \left( \sigma_h^{RT}, \text{div}_h^{RT} \xi_h \right)_{C^{-1,as}} - \left( \varepsilon(u_h^{RT}), \text{div}_h^{RT} \xi_h \right) = 0, \]
\[ \left( \sigma_h^{RT} - \mathcal{C} \varepsilon(u_h^{RT}), \nabla v_h \right) = 0 \]
(5.33)
for all $\xi_h \in X_h$ and $v_h \in V_h$. Combining (5.32) and (5.33) leads to

$$
(\text{div} (\sigma^\text{AW}_h - \sigma^\text{RT}_h), \xi_h) + (\sigma^\text{AW}_h - \sigma^\text{RT}_h, \text{div}^{-1}_\text{AW} \xi_h) + (\varepsilon(\sigma^\text{AW}_h - \sigma^\text{RT}_h), \text{div}^{-1}_\text{AW} \xi_h) = 0
$$

or, equivalently,

$$
(\text{div} (\sigma^\text{AW}_h - \sigma^\text{RT}_h), \xi_h) + (C^{-1}(P_{\text{AW}}(\sigma^\text{AW}_h - \sigma^\text{RT}_h)) - \nabla(u^\text{AW}_h - u^\text{RT}_h), \text{div}^{-1}_\text{AW} \xi_h) = 0
$$

for all $\xi_h \in X_h$ and $v_h \in V_h$. In terms of the differences $\eta_h = \text{div}\sigma^\text{AW}_h - \text{div}\sigma^\text{RT}_h \in X_h$ and $e_h = u^\text{AW}_h - u^\text{RT}_h \in V_h$, this may be written as

$$
(\eta_h, \xi_h) + (C^{-1}(\text{div}^{-1}_\text{AW} \eta_h), \text{div}^{-1}_\text{AW} \xi_h) + (e_h, \xi_h) = 0. \tag{5.34}
$$

Setting $\xi_h = \eta_h \in X_h$ and $v_h = e_h \in V_h$ in (5.34) leads to

$$
(\eta_h, \eta_h) + (C^{-1}(\text{div}^{-1}_\text{AW} \eta_h), \text{div}^{-1}_\text{AW} \eta_h) + (e_h, \eta_h) = 0.
$$

which may be combined into one equation as

$$
\|\eta_h\|^2 + \|C^{-1/2}(\text{div}^{-1}_\text{AW} \eta_h)\|^2 + 2(e_h, \eta_h) + \|\varepsilon(e_h)\|^2 = 0 \tag{5.35}
$$

For the left-hand side in (5.35),

$$
(e_h, \eta_h) = (e_h, \text{div}^{-1}_\text{AW} \eta_h) = -\text{div}^{-1}_\text{AW} e_h = -\varepsilon(e_h, \text{div}^{-1}_\text{AW} \eta_h)
$$

implies

$$
\|\eta_h\|^2 + \|C^{-1/2}(\text{div}^{-1}_\text{AW} \eta_h)\|^2 + 2(e_h, \eta_h) + \|\varepsilon(e_h)\|^2 = 0
$$

with the functional $\mathcal{G}$ defined in (2.3). For the right-hand side in (5.35), we obtain

$$
(C^{-1}\sigma^\text{RT}_h - \varepsilon(u^\text{RT}_h) + \|\text{div}^{-1}_\text{AW} \eta_h\|^2) \leq (C^{-1}\sigma^\text{RT}_h - \varepsilon(u^\text{RT}_h) + \|\text{div}^{-1}_\text{AW} \eta_h\|^2) \leq h \mathcal{F}(\sigma^\text{RT}_h, u^\text{RT}_h)^{1/2} \|\eta_h\|.
$$
where the estimate
\[
\|\text{div}^{-1}_{RT} \eta_h - \text{div}^{-1}_{AW} \eta_h\| \leq \|\text{div}^{-1} \eta_h - \text{div}^{-1}_{AW} \eta_h\| + \|\text{div}^{-1} \eta_h - \text{div}^{-1}_{RT} \eta_h\| \lesssim h \|\eta_h\|
\]
is used which follows again from Lemma 5.2.

Therefore, (5.35) implies
\[
\mathcal{G}(\text{div}^{-1}_{AW} \eta_h, e_h) \lesssim h \mathcal{F}(\sigma_h^{RT}, u_h^{RT})^{1/2} \|\eta_h\|,
\]
which, combined with (3.9), leads to
\[
\|\eta_h\| \lesssim h \mathcal{F}(\sigma_h^{RT}, u_h^{RT})^{1/2}
\]
and from there to (5.30).

The implication of Theorems 5.3 and 5.4 is that the momentum balance error of \(\sigma_h^{RT}\) is of higher order compared to the error measured in the least squares functional,
\[
\|\text{div} \sigma_h^{RT}\| \lesssim h \left(\mathcal{F}(\sigma_h^{AW}, u_h^{AW})^{1/2} + \mathcal{F}(\sigma_h^{RT}, u_h^{RT})^{1/2}\right).
\]

This property relies on the close connection of our modified least squares formulation using Raviart-Thomas elements to the classical mixed formulation using Arnold-Winther elements. The regularity assumptions (5.15) and (5.19) are satisfied if all interior angles at points where traction boundary conditions meet displacement conditions are sufficiently small. If the problem is not \(H^2\) regular in this sense, then, an estimate similar to (5.37) holds with \(h\) replaced by \(h^\rho\) for some \(\rho < 1\). This can be shown in the usual way using interpolation arguments in scales of Sobolev spaces. The best location for the observation of (5.37) in our numerical results is in Table 4.6 where a more regular Cook's membrane problem is treated. While the functional is reduced at a rate somewhat slower than proportional to \(h\), the square of the momentum balance error almost behaves proportionally to \(h^2\).

We end this section with a comparison of our approach to mixed methods of saddle point type in terms of the overall dimension of the systems. We concentrate on the case \(k = 1\) also used in our computations and denote by \(N_P\), \(N_E\), and \(N_T\) the number of points, edges and triangles, respectively, in our triangulation. Then, the stress space using \(RT_1\) elements involves \(4N_E + 4N_T\) unknowns. For the approach studied in our paper this needs to be augmented with the \(2N_P + 2N_E\) degrees of freedom associated with the conforming quadratic elements. In the \(PEERS_1\) formulation of [12] with comparable approximation properties the \(RT_1\) elements are combined with discontinuous linear elements for the displacements (6\(N_T\) unknowns) and discontinuous quadratic elements for the antisymmetry (6\(N_T\) unknowns). Using the fact that for sufficiently fine triangulations \(N_E \approx 3N_P\) and \(N_T \approx 2N_P\) holds, we have approximately 28\(N_P\) degrees of freedom in our formulation compared to 44\(N_P\) in the \(PEERS_1\) approach.

We thank the anonymous referees for their careful reading and valuable suggestions.

**REFERENCES**


