ENERGY NORM A POSTERIORI ERROR ESTIMATES FOR MIXED FINITE ELEMENT METHODS

CARLO LOVADINA AND ROLF STENBERG

Abstract. The paper deals with the a-posteriori error analysis of mixed finite element methods for second order elliptic equations. It is shown that a reliable and efficient error estimator can be constructed using a postprocessed solution of the method. The analysis is performed in two different ways; under a saturation assumption and using a Helmholtz decomposition for vector fields.

1. Introduction

We consider the mixed finite element approximation of second order elliptic equations with the Poisson problem as a model:

$$-\Delta u = f \quad \text{in} \quad \Omega \subset \mathbb{R}^n, \tag{1.1}$$

$$u = 0 \quad \text{on } \partial\Omega. \tag{1.2}$$

The problem is written as the system

$$\boldsymbol{\sigma} - \nabla u = \mathbf{0}, \tag{1.3}$$

$$\operatorname{div} \boldsymbol{\sigma} + f = 0, \tag{1.4}$$

which is approximated with the

Mixed method. *Find* $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h \subset \boldsymbol{H}(\operatorname{div}:\Omega) \times L^2(\Omega)$ *such that*

$$(\boldsymbol{\sigma}_h, \boldsymbol{\tau}) + (\operatorname{div} \boldsymbol{\tau}, u_h) = 0 \quad \forall \boldsymbol{\tau} \in \boldsymbol{S}_h, \tag{1.5}$$

$$(\operatorname{div} \boldsymbol{\sigma}_h, v) + (f, v) = 0 \quad \forall v \in V_h. \tag{1.6}$$

In the method the polynomial used for approximating the flux $\boldsymbol{\sigma}$ is of higher degree than that used for the displacement u, which is counterintuitive in view of (1.3). As a consequence, the mixed method has to be carefully designed in order to satisfy the Babuška-Brezzi conditions, c.f. e.g. [8]. There are two ways of posing these conditions, both yielding the same a priori estimates. The more common one is to use the $\boldsymbol{H}(\operatorname{div}:\Omega)$ norm for the flux and the $L^2(\Omega)$ norm for the displacement. The other one is to use so called mesh dependent norms [3] which are close to the energy norm of the continuous problem.

The a posteriori error analysis of mixed methods has been performed in [1], [10] and [5]. In [10] the estimate is for the $\boldsymbol{H}(\operatorname{div}:\Omega)$ norm. This is in a way unsatisfactory since the "div" part of the norm is trivially computable and also may dominate the error, see Remark 3.4 below. In [5] an estimate for the L^2-norm

1991 *Mathematics Subject Classification.* 65N30.

Key words and phrases. mixed finite element methods, a-posteriori error estimates, post-processing.

This work has been supported by the European Project HPRN-CT-2002-00284 "New Materials, Adaptive Systems and their Nonlinearities. Modelling, Control and Numerical Simulation".

of the flux is derived but it is, however, not optimal. The reason for this is that the estimator includes the element residual in the constitutive relation (1.3). As the polynomial degree of approximation for the displacement is lower than that for the flux, it is clear that this residual is large.

The purpose of this paper is to point out a simple remedy to this. Since the work of Arnold and Brezzi [2] it is known that the mixed finite element solution can be locally postprocessed in order to obtain an improved displacement. Later other postprocessing has been proposed [6, 9, 7, 17, 16]. On each element the postprocessed displacement is of one degree higher than the flux, which is in accordance with (1.3). Hence, it is natural to use it in the a posteriori estimate. In this paper, we will focus on the postprocessing introduced in [17, 16]. In Section 2 we develop an a-priori error analysis by recognizing that the postprocessed output can be viewed as the direct solution of a suitable modified method. In Section 3 we introduce our estimator based on the postprocessed solution, and we prove its efficiency and reliability.

Throughout the paper we will use standard notations for Sobolev norms and seminorms. Moreover, we will denote with C and C_i $(i = 1, 2, ...)$ generic constants independent of the mesh parameter h, which may take different values in different occurrences.

2. A-priori estimates and postprocessing

In this section we will consider the mixed methods, their postprocessing and error analysis. We will also give the stability and error analysis by treating the method and the postprocessing as one method. This will be useful for the a posteriori analysis.

We will use standard notation used in connection with (mixed) FE methods. By $\mathcal{C}_h$ we denote the finite element regular partitioning and by Γ_h the collection of edges or faces of $\mathcal{C}_h$. The subspaces $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h \subset \boldsymbol{H}(\mathrm{div}:\Omega) \times L^2(\Omega)$ are piecewise polynomial spaces defined on $\mathcal{C}_h$. In this paper we will consider the following families of elements. (The results are, however, easily applicable for other families as well.)

- *RTN elements* – the triangular elements of Raviart-Thomas [15] and their tetrahedral counterparts of Nedelec [14];
- *BDM elements* – the triangular elements of Brezzi-Douglas-Marini [9] and their tetrahedral counterparts of Brezzi-Douglas-Duran-Fortin [7].

Accordingly, given an integer $k \geq 1$, we define:

$$\boldsymbol{S}_h^{RTN} = \{\, \boldsymbol{\tau} \in \boldsymbol{H}(\mathrm{div}:\Omega) \,|\, \boldsymbol{\tau}|_K \in [P_{k-1}(K)]^n \oplus \boldsymbol{x}\tilde{P}_{k-1}(K) \ \forall K \in \mathcal{C}_h \,\} \tag{2.1}$$

$$\boldsymbol{S}_h^{BDM} = \{\, \boldsymbol{\tau} \in \boldsymbol{H}(\mathrm{div}:\Omega) \,|\, \boldsymbol{\tau}|_K \in [P_k(K)]^n \ \forall K \in \mathcal{C}_h \,\} \tag{2.2}$$

$$V_h^{RTN} = V_h^{BDM} = \{\, v \in L^2(\Omega) \mid v|_K \in P_{k-1}(K) \ \forall K \in \mathcal{C}_h \,\}, \tag{2.3}$$

where $\tilde{P}_{k-1}(K)$ denotes the homogeneous polynomials of degree $k-1$. For quadrilateral and hexahedral meshes there exist a wide choice of different alternatives, c.f. [8].

By defining the following bilinear form

$$\mathcal{B}(\boldsymbol{\varphi}, w; \boldsymbol{\tau}, v) = (\boldsymbol{\varphi}, \boldsymbol{\tau}) + (\mathrm{div}\,\boldsymbol{\tau}, w) + (\mathrm{div}\,\boldsymbol{\varphi}, v) \tag{2.4}$$

the mixed method can compactly be defined as:
Find $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h$ such that

$$\mathcal{B}(\boldsymbol{\sigma}_h, u_h; \boldsymbol{\tau}, v) + (f, v) = 0 \qquad \forall (\boldsymbol{\tau}, v) \in \boldsymbol{S}_h \times V_h. \tag{2.5}$$

For the displacement and the flux we will use the following norms:

$$\|v\|_{1,h}^2 = \sum_{K \in \mathcal{C}_h} \|\nabla v\|_{0,K}^2 + \sum_{E \in \Gamma_h} h_E^{-1} \|[\![v]\!]\|_{0,E}^2, \tag{2.6}$$

and

$$\|\boldsymbol{\tau}\|_{0,h}^2 = \|\boldsymbol{\tau}\|_0^2 + \sum_{E \in \Gamma_h} h_E \|\boldsymbol{\tau} \cdot \boldsymbol{n}\|_{0,E}^2, \tag{2.7}$$

where $\boldsymbol{n}$ is the unit normal to $E \in \Gamma_h$ and $[\![v]\!]$ is the jump in v along interior edges/faces and v on edges/faces on $\partial\Omega$. By an element by element partial integration we have

$$|(\operatorname{div} \boldsymbol{\tau}, v)| \leq \|\boldsymbol{\tau}\|_{0,h} \|v\|_{1,h} \qquad \forall (\boldsymbol{\tau}, v) \in \boldsymbol{S}_h \times V_h. \tag{2.8}$$

In the FE subspace the norm for the flux is equivalent to the L^2-norm:

$$C\|\boldsymbol{\tau}\|_{0,h} \leq \|\boldsymbol{\tau}\|_0 \leq \|\boldsymbol{\tau}\|_{0,h} \quad \forall \boldsymbol{\tau} \in \boldsymbol{S}_h. \tag{2.9}$$

Hence, it also holds

$$|(\operatorname{div} \boldsymbol{\tau}, v)| \leq C\|\boldsymbol{\tau}\|_0 \|v\|_{1,h} \qquad \forall (\boldsymbol{\tau}, v) \in \boldsymbol{S}_h \times V_h. \tag{2.10}$$

With this choice of norms the Babuška-Brezzi *stability condition* is the following.

Lemma 2.1. *There is a positive constant C such that*

$$\sup_{\boldsymbol{\tau} \in \boldsymbol{S}_h} \frac{(\operatorname{div} \boldsymbol{\tau}, v)}{\|\boldsymbol{\tau}\|_0} \geq C\|v\|_{1,h} \quad \forall v \in V_h. \tag{2.11}$$

Proof. We first point out that since $V_h^{RTN} = V_h^{BDM}$ and $\boldsymbol{S}_h^{RTN} \subset \boldsymbol{S}_h^{BDM}$ the result for BDM is a consequence of that for RTN. Therefore, we focus on the RTN family, first recalling that the local degrees of freedom for the flux variable are the following:

$$\langle \boldsymbol{\tau} \cdot \boldsymbol{n}, z \rangle_E \qquad \forall z \in P_{k-1}(E), \qquad E \subset \partial K, \tag{2.12}$$

$$(\boldsymbol{\tau}, \boldsymbol{z})_K \qquad \forall \boldsymbol{z} \in [P_{k-2}(K)]^n. \tag{2.13}$$

Above and in the rest of the paper, we use the notation $(\cdot, \cdot)_K$ and $\langle \cdot, \cdot \rangle_E$ for the L^2 inner product on the element K and on the edge/face E, respectively.

Hence, given $v \in V_h$ we can define $\boldsymbol{\tau} \in \boldsymbol{S}_h$ by

$$\langle \boldsymbol{\tau} \cdot \boldsymbol{n}, z \rangle_E = h_E^{-1} \langle [\![v]\!], z \rangle_E \qquad \forall z \in P_{k-1}(E), \qquad E \in \Gamma_h, \tag{2.14}$$

$$(\boldsymbol{\tau}, \boldsymbol{z})_K = -(\nabla v, \boldsymbol{z})_K \qquad \forall \boldsymbol{z} \in [P_{k-2}(K)]^n, \qquad K \in \mathcal{C}_h. \tag{2.15}$$

Noting that $\nabla v_{|K} \in [P_{k-2}(K)]^n$, $[\![v]\!]_{|E} \in P_{k-1}(E)$, from (2.14)–(2.15) we obtain

$$\langle \boldsymbol{\tau} \cdot \boldsymbol{n}, [\![v]\!] \rangle_E = h_E^{-1} \|[\![v]\!]\|_{0,E}^2, \tag{2.16}$$

$$(\boldsymbol{\tau}, \nabla v)_K = -\|\nabla v\|_{0,K}^2. \tag{2.17}$$

It follows that (cf. also (2.6))

$$(2.18)\qquad \begin{aligned}(\operatorname{div}\boldsymbol{\tau}, v) &= -\sum_{K\in\mathcal{C}_h}(\boldsymbol{\tau},\nabla v)_K + \sum_{E\in\Gamma_h}\langle \boldsymbol{\tau}\cdot\boldsymbol{n}, [\![v]\!]\rangle_E \\ &= \sum_{K\in\mathcal{C}_h}\|\nabla v\|_{0,K}^2 + \sum_{E\in\Gamma_h} h_E^{-1}\|[\![v]\!]\|_{0,E}^2 = \|v\|_{1,h}^2\end{aligned}$$

Using scaling arguments (2.14)–(2.15) imply

$$(2.19)\qquad ||\boldsymbol{\tau}||_{0,h} \le C||v||_{1,h}.$$

The assertion now follows from (2.18) and (2.19). □

From this stability estimate, the following full stability result holds.

Lemma 2.2. *There is a positive constant C such that*

$$(2.20)\qquad \sup_{(\boldsymbol{\tau},v)\in\boldsymbol{S}_h\times V_h}\frac{\mathcal{B}(\boldsymbol{\varphi},w;\boldsymbol{\tau},v)}{\|\boldsymbol{\tau}\|_0+\|v\|_{1,h}} \ge C\big(\|\boldsymbol{\varphi}\|_0 + \|w\|_{1,h}\big) \quad \forall(\boldsymbol{\varphi},w)\in\boldsymbol{S}_h\times V_h.$$

In our analysis we will exploit the *interpolation operator* $\boldsymbol{R}_h : \boldsymbol{H}(\operatorname{div}:\Omega)\cap[L^s(\Omega)]^n \to \boldsymbol{S}_h$, with $s>2$, such that

$$(2.21)\qquad (\operatorname{div}(\boldsymbol{\tau}-\boldsymbol{R}_h\boldsymbol{\tau}), v) = 0 \quad \forall v\in V_h,$$

which can be constructed by using the degrees of freedom for $\boldsymbol{S}_h$, cf. [15, 14, 9, 7]. In addition, we will use the *equilibrium property*

$$(2.22)\qquad \operatorname{div}\boldsymbol{S}_h \subset V_h.$$

When denoting by $P_h : L^2(\Omega)\to V_h$ the L^2-projection, this implies that

$$(2.23)\qquad (\operatorname{div}\boldsymbol{\tau}, u - P_h u) = 0 \quad \forall \boldsymbol{\tau}\in\boldsymbol{S}_h.$$

The projection and interpolation operators satisfy the following *commuting property*:

$$(2.24)\qquad \operatorname{div}\boldsymbol{R}_h = P_h \operatorname{div}.$$

Theorem 2.3. *There is a positive constant C such that*

$$(2.25)\qquad \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0 + \|P_h u - u_h\|_{1,h} \le C\|\boldsymbol{\sigma}-\boldsymbol{R}_h\boldsymbol{\sigma}\|_0.$$

Proof. By Lemma 2.2 there is a pair $(\boldsymbol{\tau},v)\in\boldsymbol{S}_h\times V_h$, with $\|\boldsymbol{\tau}\|_0+\|v\|_{1,h}\le C$, such that

$$(2.26)\qquad \|\boldsymbol{\sigma}_h-\boldsymbol{R}_h\boldsymbol{\sigma}\|_0 + \|u_h - P_h u\|_{1,h} \le \mathcal{B}(\boldsymbol{\sigma}_h-\boldsymbol{R}_h\boldsymbol{\sigma}, u_h-P_hu;\boldsymbol{\tau},v).$$

Next, (2.21), (2.23) and (2.24) give

$$(2.27)\qquad \begin{aligned}&\mathcal{B}(\boldsymbol{\sigma}_h-\boldsymbol{R}_h\boldsymbol{\sigma}, u_h-P_hu;\boldsymbol{\tau},v)\\ &\quad= (\boldsymbol{\sigma}_h-\boldsymbol{R}_h\boldsymbol{\sigma},\boldsymbol{\tau}) + (\operatorname{div}\boldsymbol{\tau}, u_h-P_hu) + (\operatorname{div}(\boldsymbol{\sigma}_h-\boldsymbol{R}_h\boldsymbol{\sigma}), v)\\ &\quad= (\boldsymbol{\sigma}-\boldsymbol{R}_h\boldsymbol{\sigma},\boldsymbol{\tau}) \le \|\boldsymbol{\sigma}-\boldsymbol{R}_h\boldsymbol{\sigma}\|_0\|\boldsymbol{\tau}\|_0 \le C\|\boldsymbol{\sigma}-\boldsymbol{R}_h\boldsymbol{\sigma}\|_0.\end{aligned}$$

The assertion then follows from the triangle inequality. □

This gives (assuming full regularity):

$$(2.28)\qquad \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0 + \|P_hu-u_h\|_{1,h} \le Ch^{k+1}|\boldsymbol{\sigma}|_{k+1} \qquad \text{for BDM},$$

$$(2.29)\qquad \|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0 + \|P_hu-u_h\|_{1,h} \le Ch^{k}|\boldsymbol{\sigma}|_{k} \qquad \text{for RTN}.$$

We note that these estimates contain a *superconvergence* result for $\|P_h u - u_h\|_{1,h}$. This, together with the fact that $\boldsymbol{\sigma}_h$ is a good approximation of ∇u, implies that an improved approximation for the displacement can be constructed by local *post-processing*. Below we will consider the method introduced in [17, 16]. The postprocessed displacement is sought in a FE space $V_h^* \supset V_h$. For our choices, the spaces are

$$V_h^{*BDM} = \{\, v \in L^2(\Omega) \mid v|_K \in P_{k+1}(K) \ \forall K \in \mathcal{C}_h \,\}, \tag{2.30}$$

$$V_h^{*RTN} = \{\, v \in L^2(\Omega) \mid v|_K \in P_k(K) \ \forall K \in \mathcal{C}_h \,\}. \tag{2.31}$$

Postprocessing method. *Find* $u_h^* \in V_h^*$ *such that*

$$P_h u_h^* = u_h \tag{2.32}$$

and

$$(\nabla u_h^*, \nabla v)_K = (\boldsymbol{\sigma}_h, \nabla v)_K \quad \forall v \in (I - P_h) V_h^*|_K. \tag{2.33}$$

The error analysis of this postprocessing is done in [17, 16]. Here we proceed in a slightly different way by considering the method and the postprocessing as one method. To this end we define the bilinear form

$$\begin{aligned} \mathcal{B}_h(\boldsymbol{\varphi}, w^*; \boldsymbol{\tau}, v^*) =& (\boldsymbol{\varphi}, \boldsymbol{\tau}) + (\operatorname{div} \boldsymbol{\tau}, w^*) + (\operatorname{div} \boldsymbol{\varphi}, v^*) \\ &+ \sum_{K \in \mathcal{C}_h} (\nabla w^* - \boldsymbol{\varphi}, \nabla (I - P_h) v^*)_K. \end{aligned} \tag{2.34}$$

Then we have the following equivalence to the original problem.

Lemma 2.4. *Let* $(\boldsymbol{\sigma}_h, u_h^*) \in \boldsymbol{S}_h \times V_h^*$ *be the solution to the problem*

$$\mathcal{B}_h(\boldsymbol{\sigma}_h, u_h^*; \boldsymbol{\tau}, v^*) + (P_h f, v^*) = 0 \qquad \forall (\boldsymbol{\tau}, v^*) \in \boldsymbol{S}_h \times V_h^*, \tag{2.35}$$

and set $u_h = P_h u_h^* \in V_h$. *Then* $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h$ *coincides with the solution of* (1.5)–(1.6). *Conversely, let* $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h$ *be the solution of* (1.5)–(1.6), *and let* $u_h^* \in V_h^*$ *be the postprocessed displacement defined by* (2.32)–(2.33). *Then* $(\boldsymbol{\sigma}_h, u_h^*) \in \boldsymbol{S}_h \times V_h^*$ *is the solution to* (2.35).

Proof. Testing by $(\boldsymbol{\tau}, 0) \in \boldsymbol{S}_h \times V_h^*$ in (2.35) gives

$$(\boldsymbol{\sigma}_h, \boldsymbol{\tau}) + (\operatorname{div} \boldsymbol{\tau}, u_h^*) = 0 \quad \forall \boldsymbol{\tau} \in \boldsymbol{S}_h. \tag{2.36}$$

The equilibrium property (2.22) implies

$$(\operatorname{div} \boldsymbol{\tau}, u_h^*) = (\operatorname{div} \boldsymbol{\tau}, u_h). \tag{2.37}$$

Hence, (1.5) is satisfied. Next, for a generic $v^* \in V_h^*$ set $v = P_h v^* \in V_h$ and observe that $V_h = P_h(V_h^*)$. Testing in (2.35) with $(\mathbf{0}, v)$, and using the fact that $(P_h f, v) = (f, v)$, we obtain

$$(\operatorname{div} \boldsymbol{\sigma}_h, v) + (f, v) = 0 \quad \forall v \in V_h, \tag{2.38}$$

i.e. the equation (1.6). Conversely, let $(\boldsymbol{\sigma}_h, u_h) \in \boldsymbol{S}_h \times V_h$ be the solution of (1.5)–(1.6), and let $u_h^* \in V_h^*$ be defined by (2.32)–(2.33). Splitting a generic $v^* \in V_h^*$ as $v^* = P_h v^* + (I - P_h) v^*$ we have

(2.39)
$$\begin{aligned}
\mathcal{B}_h(\boldsymbol{\sigma}_h, u_h^*; \boldsymbol{\tau}, v^*) &= \mathcal{B}_h(\boldsymbol{\sigma}_h, u_h^*; \boldsymbol{\tau}, P_h v^*) + \mathcal{B}_h(\boldsymbol{\sigma}_h, u_h^*; \mathbf{0}, (I-P_h)v^*) \\
&= (\boldsymbol{\sigma}_h, \boldsymbol{\tau}) + (\operatorname{div}\boldsymbol{\tau}, u_h^*) + (\operatorname{div}\boldsymbol{\sigma}_h, P_h v^*) + \sum_{K\in\mathcal{C}_h} (\nabla u_h^* - \boldsymbol{\sigma}_h, \nabla(I-P_h)P_h v^*)_K \\
&\quad + (\operatorname{div}\boldsymbol{\sigma}_h, (I-P_h)v^*) + \sum_{K\in\mathcal{C}_h} (\nabla u_h^* - \boldsymbol{\sigma}_h, \nabla(I-P_h)(I-P_h)v^*)_K \\
&= (\boldsymbol{\sigma}_h, \boldsymbol{\tau}) + (\operatorname{div}\boldsymbol{\tau}, u_h) - (P_h f, P_h v^*) = -(P_h f, v^*) \qquad \forall(\boldsymbol{\tau}, v^*) \in \boldsymbol{S}_h \times V_h^*.
\end{aligned}$$

Therefore, $(\boldsymbol{\sigma}_h, u_h^*) \in \boldsymbol{S}_h \times V_h^*$ solves (2.35). □

Next, we prove the stability. In the proof we will use the following norm equivalence.

Lemma 2.5. *There are positive constants C_1 and C_2, such that*

$$\|w^*\|_{1,h} \le \|P_h w^*\|_{1,h} + \|(I-P_h)w^*\|_{1,h} \le C_2 \|w^*\|_{1,h} \tag{2.40}$$

and

$$C_1\|w^*\|_{1,h} \le \|P_h w^*\|_{1,h} + \Big(\sum_{K\in\mathcal{C}_h} \|\nabla(I-P_h)w^*\|_{0,K}^2\Big)^{1/2} \le C_2\|w^*\|_{1,h}, \tag{2.41}$$

for every $w^* \in V_h^*$.

Proof. We first prove (2.40). The estimate

$$\|w^*\|_{1,h} \le \|P_h w^*\|_{1,h} + \|(I-P_h)w^*\|_{1,h}$$

follows immediately from the triangle inequality. To continue, we notice that

$$\|P_h w^*\|_{1,h} + \|(I-P_h)w^*\|_{1,h} \le 2\|P_h w^*\|_{1,h} + \|w^*\|_{1,h} \tag{2.42}$$

We now fix an interior edge/face E, and we consider the elements K_1 and K_2 such that $E = K_1 \cap K_2$. A scaling argument shows that

$$h_E^{-1}\|[\![P_h w^*]\!]\|_{0,E}^2 + \sum_{i=1}^{2} \|\nabla P_h w^*\|_{0,K_i}^2 \le C\Big(h_E^{-1}\|[\![w^*]\!]\|_{0,E}^2 + \sum_{i=1}^{2}\|\nabla w^*\|_{0,K_i}^2\Big). \tag{2.43}$$

If $E \subset K$ is an edge/face lying in $\partial\Omega$, a similar argument gives

$$h_E^{-1}\|P_h w^*\|_{0,E}^2 + \|\nabla P_h w^*\|_{0,K}^2 \le C\Big(h_E^{-1}\|w^*\|_{0,E}^2 + \|\nabla w^*\|_{0,K}^2\Big). \tag{2.44}$$

The estimate

$$\|P_h w^*\|_{1,h} + \|(I-P_h)w^*\|_{1,h} \le C_2\|w^*\|_{1,h} \tag{2.45}$$

easily follows from (2.42)–(2.44) (cf. also (2.6)). Hence, (2.40) is proved.

To prove (2.41) we first notice that (2.45) implies

$$\|P_h w^*\|_{1,h} + \Big(\sum_{K\in\mathcal{C}_h} \|\nabla(I-P_h)w^*\|_{0,K}^2\Big)^{1/2} \le C_2\|w^*\|_{1,h}. \tag{2.46}$$

Next, scaling arguments lead to

$$h_E^{-1}\|[\![w^*]\!]\|_{0,E}^2 + \sum_{i=1}^{2} \|\nabla w^*\|_{0,K_i}^2 \leq C\Big(h_E^{-1}\|[\![P_h w^*]\!]\|_{0,E}^2 + \sum_{i=1}^{2} \big(\|\nabla P_h w^*\|_{0,K_i}^2 + \|\nabla (I-P_h) w^*\|_{0,K_i}^2\big)\Big), \tag{2.47}$$

for an interior edge/face E, and to

$$h_E^{-1}\|w^*\|_{0,E}^2 + \|\nabla w^*\|_{0,K}^2 \leq C\Big(h_E^{-1}\|P_h w^*\|_{0,E}^2 + \|\nabla P_h w^*\|_{0,K}^2 + \|\nabla (I-P_h) w^*\|_{0,K}^2\Big), \tag{2.48}$$

for a boundary edge/face E. The estimate

$$C_1\|w^*\|_{1,h} \leq \|P_h w^*\|_{1,h} + \Big(\sum_{K\in\mathcal{C}_h} \|\nabla (I-P_h) w^*\|_{0,K}^2\Big)^{1/2}$$

is a consequence of (2.47)–(2.48). The proof is complete. □

Lemma 2.6. *There is a positive constant constant C such that*

$$\sup_{(\boldsymbol{\tau},v^*)\in \boldsymbol{S}_h\times V_h^*} \frac{\mathcal{B}_h(\boldsymbol{\varphi},w^*;\boldsymbol{\tau},v^*)}{\|\boldsymbol{\tau}\|_0+\|v^*\|_{1,h}} \geq C\big(\|\boldsymbol{\varphi}\|_0 + \|w^*\|_{1,h}\big) \quad \forall (\boldsymbol{\varphi},w^*)\in \boldsymbol{S}_h\times V_h^*. \tag{2.49}$$

Proof. Let $(\boldsymbol{\varphi},w^*)\in \boldsymbol{S}_h\times V_h^*$ be arbitrary. By choosing $v^*=v\in V_h$ and using the equilibrium condition (2.22) we then get

$$\begin{aligned}\mathcal{B}_h(\boldsymbol{\varphi},w^*;\boldsymbol{\tau},v) &= (\boldsymbol{\varphi},\boldsymbol{\tau}) + (\operatorname{div}\boldsymbol{\tau},w^*) + (\operatorname{div}\boldsymbol{\varphi},v)\\ &= (\boldsymbol{\varphi},\boldsymbol{\tau}) + (\operatorname{div}\boldsymbol{\tau},P_h w^*) + (\operatorname{div}\boldsymbol{\varphi},v)\\ &= \mathcal{B}(\boldsymbol{\varphi},P_h w^*;\boldsymbol{\tau},v),\end{aligned} \tag{2.50}$$

Hence, the stability of Lemma 2.2 implies that we can choose $(\boldsymbol{\tau},v)$ such that

$$\mathcal{B}_h(\boldsymbol{\varphi},w^*;\boldsymbol{\tau},v) \geq \big(\|\boldsymbol{\varphi}\|_0^2 + \|P_h w^*\|_{1,h}^2\big) \tag{2.51}$$

and

$$\|\boldsymbol{\tau}\|_0 + \|v\|_{1,h} \leq C_1\big(\|\boldsymbol{\varphi}\|_0 + \|P_h w^*\|_{1,h}\big). \tag{2.52}$$

Next, (2.10) and Schwarz inequality give

$$
\begin{aligned}
&\mathcal{B}_h(\boldsymbol{\varphi}, w^*; \mathbf{0}, (I-P_h)w^*) \\
&= (\operatorname{div}\boldsymbol{\varphi}, (I-P_h)w^*) + \sum_{K\in\mathcal{C}_h} (\nabla w^* - \boldsymbol{\varphi}, \nabla (I-P_h)w^*)_K \\
&\geq -C_2\|\boldsymbol{\varphi}\|_0 \|(I-P_h)w^*\|_{1,h} + \sum_{K\in\mathcal{C}_h} (\nabla w^*, \nabla (I-P_h)w^*)_K \\
&= -C_2\|\boldsymbol{\varphi}\|_0 \|(I-P_h)w^*\|_{1,h} + \sum_{K\in\mathcal{C}_h} (\nabla P_h w^*, \nabla (I-P_h)w^*)_K \\
&\qquad + \sum_{K\in\mathcal{C}_h} \|\nabla (I-P_h)w^*\|_{0,K}^2 \\
&\geq -\big(C_2\|\boldsymbol{\varphi}\|_0 + \|P_h w^*\|_{1,h}\big)\|(I-P_h)w^*\|_{1,h} \\
&\qquad + \sum_{K\in\mathcal{C}_h} \|\nabla (I-P_h)w^*\|_{0,K}^2 .
\end{aligned} \tag{2.53}
$$

We now notice that $(I-P_h)w^*$ is L^2–orthogonal to the piecewise constant functions; therefore, a scaling argument shows that

$$
\|(I-P_h)w^*\|_{1,h} \leq C_3 \left(\sum_{K\in\mathcal{C}_h} \|\nabla (I-P_h)w^*\|_{0,K}^2 \right)^{1/2} . \tag{2.54}
$$

For $\alpha > 0$, we obtain from (2.53) and (2.54)

$$
\begin{aligned}
&\mathcal{B}_h(\boldsymbol{\varphi}, w^*; \mathbf{0}, (I-P_h)w^*) \\
&\geq -\frac{1}{2\alpha}\big(C_2\|\boldsymbol{\varphi}\|_0 + \|P_h w^*\|_{1,h}\big)^2 - \frac{\alpha}{2}\|(I-P_h)w^*\|_{1,h}^2 \\
&\qquad + \sum_{K\in\mathcal{C}_h} \|\nabla (I-P_h)w^*\|_{0,K}^2 \\
&\geq -\frac{1}{2\alpha}\big(C_2\|\boldsymbol{\varphi}\|_0 + \|P_h w^*\|_{1,h}\big)^2 \\
&\qquad + \left(1 - \frac{\alpha C_3^2}{2}\right) \sum_{K\in\mathcal{C}_h} \|\nabla (I-P_h)w^*\|_{0,K}^2 .
\end{aligned} \tag{2.55}
$$

Choosing $\alpha > 0$ sufficiently small, we get

$$
\mathcal{B}_h(\boldsymbol{\varphi}, w^*; \mathbf{0}, (I-P_h)w^*) \geq C_4 \left(\sum_{K\in\mathcal{C}_h} \|\nabla (I-P_h)w^*\|_{0,K}^2 - \|\boldsymbol{\varphi}\|_0^2 - \|P_h w^*\|_{1,h}^2 \right) . \tag{2.56}
$$

Combining (2.51) and (2.56), with $\delta > 0$ to be chosen, we have

$$
\begin{aligned}
&\mathcal{B}_h(\boldsymbol{\varphi}, w^*; \boldsymbol{\tau}, v + \delta(I-P_h)w^*) \\
&\qquad \geq (1-\delta C_4)\left(\|\boldsymbol{\varphi}\|_0^2 + \|P_h w^*\|_{1,h}^2\right) + \delta C_4 \sum_{K\in\mathcal{C}_h} \|\nabla (I-P_h)w^*\|_{0,K}^2 .
\end{aligned} \tag{2.57}
$$

Next, by (2.41) we have

$$
\|P_h w^*\|_{1,h}^2 + \delta \sum_{K\in\mathcal{C}_h} \|\nabla (I-P_h)w^*\|_{0,K}^2 \geq C_5 \|w^*\|_{1,h}^2 . \tag{2.58}
$$

From (2.52) and (2.40) we have

$$
\begin{aligned}
&\|\boldsymbol{\tau}\|_0 + \|v + \delta(I - P_h)w^*\|_{1,h} \\
&\le \|\boldsymbol{\tau}\|_0 + \|v\|_{1,h} + \delta\|(I - P_h)w^*\|_{1,h} \\
&\le C_1\big(\|\boldsymbol{\varphi}\|_0 + \|P_h w^*\|_{1,h}\big) + \delta\|(I - P_h)w^*\|_{1,h} \\
&\le C_6\big(\|\boldsymbol{\varphi}\|_0 + \|w^*\|_{1,h}\big).
\end{aligned} \tag{2.59}
$$

Choosing $\delta = 1/(2C_4)$, estimate (2.49) is proved by combining (2.57)–(2.59). □

Theorem 2.7. *The following a priori error estimate holds*

$$
\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h} \le C\big(\|\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}\|_0 + \inf_{v^* \in V_h^*} \|u - v^*\|_{1,h}\big).
$$

Proof. From Lemma 2.6 it follows that there is $(\boldsymbol{\varphi}, w^*) \in \boldsymbol{S}_h \times V_h^*$, with $\|\boldsymbol{\varphi}\|_0 + \|w^*\|_{1,h} \le C$, such that

$$
\big(\|\boldsymbol{\sigma}_h - \boldsymbol{R}_h\boldsymbol{\sigma}\|_0 + \|u_h^* - v^*\|_{1,h}\big) \le \mathcal{B}_h(\boldsymbol{\sigma}_h - \boldsymbol{R}_h\boldsymbol{\sigma}, u_h^* - v^*; \boldsymbol{\varphi}, w^*). \tag{2.60}
$$

Next, from the definition of $\mathcal{B}_h$ and the equations (1.3)–(1.4) it follows that

$$
\mathcal{B}_h(\boldsymbol{\sigma}, u; \boldsymbol{\varphi}, w^*) + (f, w^*) = 0. \tag{2.61}
$$

Hence it holds

$$
\begin{aligned}
&\mathcal{B}_h(\boldsymbol{\sigma}_h - \boldsymbol{R}_h\boldsymbol{\sigma}, u_h^* - v^*; \boldsymbol{\varphi}, w^*) \\
&\quad = \mathcal{B}_h(\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}, u - v^*; \boldsymbol{\varphi}, w^*) + (f - P_h f, w^*).
\end{aligned} \tag{2.62}
$$

Writing out the right hand side we have

$$
\begin{aligned}
&\mathcal{B}_h(\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}, u - v^*; \boldsymbol{\varphi}, w^*) + (f - P_h f, w^*) \\
&= (\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}, \boldsymbol{\varphi}) + (\operatorname{div}\boldsymbol{\varphi}, u - v^*) + (\operatorname{div}(\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}), w^*) \\
&+ \sum_{K \in \mathcal{C}_h} (\nabla(u - v^*) - (\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}), \nabla(I - P_h)w^*)_K + (f - P_h f, w^*).
\end{aligned} \tag{2.63}
$$

The commuting property (2.24) gives

$$
(\operatorname{div}(\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}), w^*) = -(f - P_h f, w^*). \tag{2.64}
$$

Hence, the third and the last term on the right hand side of (2.63) cancel. The other terms are directly estimated

$$
(\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}, \boldsymbol{\varphi}) \le \|\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}\|_0\|\boldsymbol{\varphi}\|_0 \le C\|\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}\|_0, \tag{2.65}
$$

$$
(\operatorname{div}\boldsymbol{\varphi}, u - v^*) \le C\|\boldsymbol{\varphi}\|_0\|u - v^*\|_{1,h} \le C\|u - v^*\|_{1,h} \tag{2.66}
$$

and using (2.41)

$$
\begin{aligned}
&\sum_{K \in \mathcal{C}_h} (\nabla(u - v^*) - (\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}), \nabla(I - P_h)w^*)_K \\
&\qquad \le C\big(\|u - v^*\|_{1,h} + \|\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}\|_0\big)\|w^*\|_{1,h} \\
&\qquad \le C\big(\|u - v^*\|_{1,h} + \|\boldsymbol{\sigma} - \boldsymbol{R}_h\boldsymbol{\sigma}\|_0\big).
\end{aligned} \tag{2.67}
$$

The assertion then follows by collecting the above estimate and using the triangle inequality. □

For our choices of spaces we obtain the estimates (with the assumption of a sufficiently smooth solution).

Corollary 2.8. *There are positive constants* C *such that*

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h} \leq Ch^{k+1}|u|_{k+2} \qquad \text{for BDM}, \tag{2.68}$$

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h} \leq Ch^{k}|u|_{k+1} \qquad \text{for RTN}. \tag{2.69}$$

3. A-posteriori estimates

We define the following local error indicators on the elements

$$\eta_{1,K} = \|\nabla u_h^* - \boldsymbol{\sigma}_h\|_{0,K}, \quad \eta_{2,K} = h_K\|f - P_h f\|_{0,K}, \tag{3.1}$$

and on the edges

$$\eta_E = h_E^{-1/2}\|[\![u_h^*]\!]\|_{0,E}. \tag{3.2}$$

Using these quantities, the global estimator is

$$\eta = \Big(\sum_{K\in\mathcal{C}_h} \big(\eta_{1,K}^2 + \eta_{2,K}^2\big) + \sum_{E\in\Gamma_h} \eta_E^2 \Big)^{1/2}. \tag{3.3}$$

The efficiency of the estimator is given by the following lower bounds, which directly follow from (1.3) using the triangle inequality, and from (3.2) noting that $[\![u]\!] = 0$ on each edge E.

Theorem 3.1. *It holds*

$$\begin{aligned} \eta_{1,K} &\leq \|\nabla(u - u_h^*)\|_{0,K} + \|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_{0,K}, \\ \eta_E &= h_E^{-1/2}\|[\![u - u_h^*]\!]\|_{0,E}. \end{aligned} \tag{3.4}$$

As far as the estimator reliability is concerned, below we will use two different techniques.

3.1. Reliability via a saturation assumption.

The first technique to prove the upper bound is based on the following saturation assumption. We let $\mathcal{C}_{h/2}$ be the mesh obtained from $\mathcal{C}_h$ by refined each element into 2^n $(n = 2, 3)$ elements. For clarity all variables in the spaces defined on $\mathcal{C}_h$ will be equipped with the subscript h whereas $h/2$ will be used for those defined on $\mathcal{C}_{h/2}$. Accordingly, we let $(\boldsymbol{\sigma}_{h/2}, u_{h/2}^*) \in \boldsymbol{S}_{h/2} \times V_{h/2}^*$ be the solution to

$$\mathcal{B}_{h/2}(\boldsymbol{\sigma}_{h/2}, u_{h/2}^*; \boldsymbol{\tau}_{h/2}, v_{h/2}^*) + (P_{h/2}f, v_{h/2}^*) = 0 \quad \forall(\boldsymbol{\tau}_{h/2}, v_{h/2}^*) \in \boldsymbol{S}_{h/2} \times V_{h/2}^*. \tag{3.5}$$

As already done in [5], we make the following assumption for the solutions of (2.35) and (3.5).

Saturation assumption. *There exists a positive constant* $\beta < 1$ *such that*

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{h/2}\|_0 + \|u - u_{h/2}^*\|_{1,h/2} \leq \beta\big(\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h}\big). \tag{3.6}$$

Since it holds

$$\|u - u_h^*\|_{1,h} \leq \|u - u_h^*\|_{1,h/2} \tag{3.7}$$

we also have

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_{h/2}\|_0 + \|u - u_{h/2}^*\|_{1,h/2} \leq \beta\big(\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h/2}\big). \tag{3.8}$$

Using the triangle inequality we then get

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h/2} \leq \frac{1}{1-\beta}\big(\|\boldsymbol{\sigma}_{h/2} - \boldsymbol{\sigma}_h\|_0 + \|u_{h/2}^* - u_h^*\|_{1,h/2}\big). \tag{3.9}$$

By again using (3.7) we obtain

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h} \le \frac{1}{1-\beta}\big(\|\boldsymbol{\sigma}_{h/2} - \boldsymbol{\sigma}_h\|_0 + \|u_{h/2}^* - u_h^*\|_{1,h/2}\big). \tag{3.10}$$

We now prove the following result.

Theorem 3.2. *Suppose that the saturation assumption* (3.6) *holds. Then there exists a positive constant* C *such that*

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h} \le C\eta. \tag{3.11}$$

Proof. By (3.10) it is sufficient to prove the following bound

$$\|\boldsymbol{\sigma}_{h/2} - \boldsymbol{\sigma}_h\|_0 + \|u_{h/2}^* - u_h^*\|_{1,h/2} \le C\eta. \tag{3.12}$$

By Lemma 2.6 applied to the finer mesh $\mathcal{C}_{h/2}$, there is $(\boldsymbol{\tau}_{h/2}, v_{h/2}^*) \in \boldsymbol{S}_{h/2} \times V_{h/2}^*$, with $\|\boldsymbol{\tau}_{h/2}\|_0 + \|v_{h/2}^*\|_{1,h/2} \le C$, such that

$$\begin{aligned}&\big(\|\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}\|_0 + \|u_h^* - u_{h/2}^*\|_{1,h/2}\big)\\ &\qquad \le \mathcal{B}_{h/2}(\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}, u_h^* - u_{h/2}^*; \boldsymbol{\tau}_{h/2}, v_{h/2}^*).\end{aligned} \tag{3.13}$$

Using the fact that

$$(\boldsymbol{\sigma}_{h/2}, \boldsymbol{\tau}_{h/2}) + (\operatorname{div}\boldsymbol{\tau}_{h/2}, u_{h/2}^*) = 0 \tag{3.14}$$

we have

$$\begin{aligned}&\mathcal{B}_{h/2}(\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}, u_h^* - u_{h/2}^*; \boldsymbol{\tau}_{h/2}, v_{h/2}^*)\\ &\quad = (\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}, \boldsymbol{\tau}_{h/2}) + (\operatorname{div}\boldsymbol{\tau}_{h/2}, u_h^* - u_{h/2}^*) + (\operatorname{div}(\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}), v_{h/2}^*)\\ &\qquad + \sum_{K\in\mathcal{C}_{h/2}} (\nabla(u_h^* - u_{h/2}^*) - (\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}), \nabla(I - P_{h/2})v_{h/2}^*)_K\\ &\quad = (\boldsymbol{\sigma}_h, \boldsymbol{\tau}_{h/2}) + (\operatorname{div}\boldsymbol{\tau}_{h/2}, u_h^*) + (\operatorname{div}(\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}), v_{h/2}^*)\\ &\qquad + \sum_{K\in\mathcal{C}_{h/2}} (\nabla u_h^* - \boldsymbol{\sigma}_h, \nabla(I - P_{h/2})v_{h/2}^*)_K,\end{aligned} \tag{3.15}$$

We now notice that it holds (cf. (2.9))

$$C\|\boldsymbol{\tau}_{h/2}\|_{0,h} \le \|\boldsymbol{\tau}_{h/2}\|_0 \le \|\boldsymbol{\tau}_{h/2}\|_{0,h} \quad \forall \boldsymbol{\tau}_{h/2} \in \boldsymbol{S}_{h/2}. \tag{3.16}$$

Therefore, using (3.16) and (3.1)–(3.3), we obtain

$$\begin{aligned}&(\boldsymbol{\sigma}_h, \boldsymbol{\tau}_{h/2}) + (\operatorname{div}\boldsymbol{\tau}_{h/2}, u_h^*)\\ &\quad = \sum_{K\in\mathcal{C}_h} (\boldsymbol{\sigma}_h - \nabla u_h^*, \boldsymbol{\tau}_{h/2})_K + \sum_{E\in\Gamma_h} \langle \boldsymbol{\tau}_{h/2}\cdot\boldsymbol{n}, [\![u_h^*]\!]\rangle_E\\ &\quad \le \sum_{K\in\mathcal{C}_h} \|\boldsymbol{\sigma}_h - \nabla u_h^*\|_{0,K}\|\boldsymbol{\tau}_{h/2}\|_{0,K} + \sum_{E\in\Gamma_h} \|\boldsymbol{\tau}_{h/2}\cdot\boldsymbol{n}\|_{0,E}\|[\![u_h^*]\!]\|_{0,E}\\ &\quad \le \eta\|\boldsymbol{\tau}_{h/2}\|_{0,h} \le \eta C\|\boldsymbol{\tau}_{h/2}\|_0 \le C\eta.\end{aligned} \tag{3.17}$$

Similarly for the last term in (3.15) we get using (2.40)

$$\begin{aligned}\sum_{K\in\mathcal{C}_{h/2}} (\nabla u_h^* - \boldsymbol{\sigma}_h, \nabla(I - P_{h/2})v_{h/2}^*)_K &\le C\eta\|(I - P_{h/2})v_{h/2}^*\|_{1,h/2}\\ &\le C\eta\|v_{h/2}^*\|_{1,h/2} \le C\eta.\end{aligned} \tag{3.18}$$

When estimating the term $(\operatorname{div}(\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}), v^*_{h/2})$ in (3.15) we recall that

$$\operatorname{div}\boldsymbol{\sigma}_h = -P_h f \quad \text{and} \quad \operatorname{div}\boldsymbol{\sigma}_{h/2} = -P_{h/2} f,$$

and that $P_h,\ P_{h/2}$ are L^2-projection operators. Therefore, we have

$$\begin{aligned}(3.19)\qquad (\operatorname{div}(\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}), v^*_{h/2}) &= (P_{h/2}f - P_h f, v^*_{h/2}) \\ &= (P_{h/2}f - f, v^*_{h/2}) + (f - P_h f, v^*_{h/2}) \\ &= (P_{h/2}f - f, v^*_{h/2} - P_{h/2}v^*_{h/2}) + (f - P_h f, v^*_{h/2} - P_h v^*_{h/2}).\end{aligned}$$

Next, we use the following interpolation estimates, which are easily proved by standard scaling arguments (cf. [5, Lemma 3.1]):

$$\|v^*_{h/2} - P_h v^*_{h/2}\|_{0,K} \le C h_K |v^*_{h/2}|_{1,h/2,K}, \quad \forall K \in \mathcal{C}_h,$$

where

$$|v^*_{h/2}|^2_{1,h/2,K} = \sum_{K_i} \|\nabla v^*_{h/2}\|^2_{0,K_i} + \sum_{E_i} h^{-1}_{E_i} \|[\![v^*_{h/2}]\!]\|^2_{0,E_i}.$$

Here $K_i \subset K$ are the elements of $\mathcal{C}_{h/2}$ and E_i are the edges of $\Gamma_{h/2}$ lying in the interior of K. This gives

$$\begin{aligned}(3.20)\quad (f - P_h f, v^*_{h/2} - P_h v^*_{h/2}) &\le C\Big(\sum_{K\in\mathcal{C}_h} h_K^2 \|f - P_h f\|^2_{0,K}\Big)^{1/2} \|v^*_{h/2}\|_{1,h/2} \\ &\le C\Big(\sum_{K\in\mathcal{C}_h} h_K^2 \|f - P_h f\|^2_{0,K}\Big)^{1/2} \le C\eta.\end{aligned}$$

We also have

$$\begin{aligned}(P_{h/2}f - f, v^*_{h/2} - P_{h/2}v^*_{h/2}) &\le \sum_{K\in\mathcal{C}_{h/2}} \|f - P_{h/2}f\|_{0,K} \|v^*_{h/2} - P_{h/2}v^*_{h/2}\|_{0,K} \\ &\le C \sum_{K\in\mathcal{C}_{h/2}} h_K \|f - P_{h/2}f\|_{0,K} \|\nabla v^*_{h/2}\|_{0,K} \\ (3.21)\qquad &\le C\Big(\sum_{K\in\mathcal{C}_{h/2}} h_K^2 \|f - P_{h/2}f\|^2_{0,K}\Big)^{1/2} \|v^*_{h/2}\|_{1,h/2} \\ &\le C\Big(\sum_{K\in\mathcal{C}_{h/2}} h_K^2 \|f - P_{h/2}f\|^2_{0,K}\Big)^{1/2} \\ &\le C\Big(\sum_{K\in\mathcal{C}_{h}} h_K^2 \|f - P_{h/2}f\|^2_{0,K}\Big)^{1/2}.\end{aligned}$$

Since, by the properties of L^2-projection operators, it holds

$$\|f - P_{h/2}f\|_{0,K} \le \|f - P_h f\|_{0,K} \quad \forall K \in \mathcal{C}_h,$$

from (3.21) we obtain

$$(3.22)\qquad (P_{h/2}f - f, v^*_{h/2} - P_{h/2}v^*_{h/2}) \le C\Big(\sum_{K\in\mathcal{C}_h} h_K^2 \|f - P_h f\|^2_{0,K}\Big)^{1/2} \le C\eta.$$

By collecting the estimates (3.17)–(3.20) and (3.22), from (3.15) we get

$$(3.23)\qquad \mathcal{B}_{h/2}(\boldsymbol{\sigma}_h - \boldsymbol{\sigma}_{h/2}, u^*_h - u^*_{h/2}; \boldsymbol{\tau}_{h/2}, v^*_{h/2}) \le C\eta.$$

The assertion now follows from (3.13). □

We have presented the above proof since this is rather general and can be used for other problems as well. In [13] we use it for a plate bending method.

3.2. **Reliability via a Helmholtz decomposition.** Now, let us give another proof of the estimator reliability, not relying on the saturation assumption.

Theorem 3.3. *Suppose that $\Omega \subset \mathbb{R}^2$ is a simply connected domain. Then there exists a positive constant C such that*

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_0 + \|u - u_h^*\|_{1,h} \leq C\eta. \tag{3.24}$$

Proof. We use the techniques of [11] and [10]. We first notice that

$$||\boldsymbol{\sigma} - \boldsymbol{\sigma}_h||_0 = \sup_{\boldsymbol{\varphi} \in \boldsymbol{L}^2(\Omega)} \frac{(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \boldsymbol{\varphi})}{||\boldsymbol{\varphi}||_0}. \tag{3.25}$$

For a generic $\boldsymbol{\varphi} \in \boldsymbol{L}^2(\Omega)$, we consider the $\boldsymbol{L}^2$-orthogonal Helmholtz decomposition (see, e.g. [12]):

$$\boldsymbol{\varphi} = \nabla\psi + \mathbf{curl}\, q, \qquad \psi \in H_0^1(\Omega), \quad q \in H^1(\Omega)/\mathbb{R}, \tag{3.26}$$

with

$$||\boldsymbol{\varphi}||_0 = \Big(||\nabla\psi||_0^2 + ||\mathbf{curl}\, q||_0^2\Big)^{1/2}. \tag{3.27}$$

Therefore, from (3.25)–(3.27) we see that it holds

$$||\boldsymbol{\sigma} - \boldsymbol{\sigma}_h||_0 \leq \sup_{\psi \in H_0^1(\Omega)} \frac{(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \nabla\psi)}{|\psi|_1} + \sup_{q \in H^1(\Omega)/\mathbb{R}} \frac{(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{curl}\, q)}{|q|_1}. \tag{3.28}$$

Given $\psi \in H_0^1(\Omega)$, from (1.4) and (1.6) it follows that

$$\big(\mathrm{div}\,(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h), P_h\psi\big) = 0. \tag{3.29}$$

Hence, we have

$$\begin{aligned}(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \nabla\psi) &= -\,(\mathrm{div}\,(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h), \psi)\\ &= -\,(\mathrm{div}\,(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h), \psi - P_h\psi)\\ &\leq C\Big(\sum_{K \in \mathcal{C}_h} h_K^2 ||\mathrm{div}\,(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h)||_{0,K}^2\Big)^{1/2} |\psi|_1\\ &\leq C\Big(\sum_{K \in \mathcal{C}_h} h_K^2 ||f - P_h f||_{0,K}^2\Big)^{1/2} |\psi|_1.\end{aligned} \tag{3.30}$$

As a consequence, we get (cf. (3.1))

$$\sup_{\psi \in H_0^1(\Omega)} \frac{(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \nabla\psi)}{|\psi|_1} \leq C\Big(\sum_{K \in \mathcal{C}_h} h_K^2 ||f - P_h f||_{0,K}^2\Big)^{1/2} = C\Big(\sum_{K \in \mathcal{C}_h} \eta_{2,K}^2\Big)^{1/2}. \tag{3.31}$$

To continue, let $I_h q$ be the Clément interpolant of q in the space of continuous piecewise linear functions (see [4], for instance) satisfying

$$\|q - I_h q\|_1 + \Big(\sum_{E\in\Gamma_h} h_E^{-1} ||q - I_h q||_{0,E}^2 \Big)^{1/2} \le C|q|_1. \tag{3.32}$$

Noting that $\mathbf{curl}\, I_h q \in \boldsymbol{S}_h$, and $\operatorname{div} \mathbf{curl}\, I_h q = 0$, from (1.3) and (1.5) we get

$$(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{curl}\, I_h q) = 0. \tag{3.33}$$

Therefore, using (3.32), one has

$$\begin{aligned}
(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{curl}\, q) &= \big(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{curl}(q - I_h q)\big) \\
&= \big(\nabla u - \boldsymbol{\sigma}_h, \mathbf{curl}(q - I_h q)\big) = -\big(\boldsymbol{\sigma}_h, \mathbf{curl}(q - I_h q)\big) \\
&= - \sum_{K\in\mathcal{C}_h} \big(\boldsymbol{\sigma}_h - \nabla u_h^*, \mathbf{curl}(q - I_h q)\big)_K + \sum_{K\in\mathcal{C}_h} \big(\nabla u_h^*, \mathbf{curl}(q - I_h q)\big)_K \\
&\le C\Big(\sum_{K\in\mathcal{C}_h} ||\boldsymbol{\sigma}_h - \nabla u_h^*||_{0,K}^2 \Big)^{1/2} |q|_1 + \sum_{K\in\mathcal{C}_h} \big(\nabla u_h^*, \mathbf{curl}(q - I_h q)\big)_K.
\end{aligned} \tag{3.34}$$

Furthermore, an integration by parts and standard arguments and (3.32) give

$$\begin{aligned}
\sum_{K\in\mathcal{C}_h} \big(\nabla u_h^*, \mathbf{curl}(q - I_h q)\big)_K &= - \sum_{K\in\mathcal{C}_h} \langle \nabla u_h^* \cdot \boldsymbol{t}, q - I_h q\rangle_{\partial K} \\
&= - \sum_{E\in\Gamma_h} \langle [\![\nabla u_h^* \cdot \boldsymbol{t}]\!], q - I_h q\rangle_E \\
&\le \Big(\sum_{E\in\Gamma_h} h_E ||[\![\nabla u_h^* \cdot \boldsymbol{t}]\!]||_{0,E}^2 \Big)^{1/2} \Big(\sum_{E\in\Gamma_h} h_E^{-1} ||q - I_h q||_{0,E}^2 \Big)^{1/2} \\
&\le C\Big(\sum_{E\in\Gamma_h} h_E^{-1} ||[\![u_h^*]\!]||_{0,E}^2 \Big)^{1/2} |q|_1.
\end{aligned} \tag{3.35}$$

From (3.34) and (3.35) we obtain (see (3.1) and (3.2))

$$\begin{aligned}
\sup_{q\in H^1(\Omega)/\mathbb{R}} \frac{(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \mathbf{curl}\, q)}{|q|_1} &\le C\Big(\sum_{K\in\mathcal{C}_h} ||\boldsymbol{\sigma}_h - \nabla u_h^*||_{0,K}^2 + \sum_{E\in\Gamma_h} h_E^{-1} ||[\![u_h^*]\!]||_{0,E}^2 \Big)^{1/2} \\
&= C\Big(\sum_{K\in\mathcal{C}_h} \eta_{1,K}^2 + \sum_{E\in\Gamma_h} \eta_E^2 \Big)^{1/2}.
\end{aligned} \tag{3.36}$$

Using (3.31) and (3.36) we deduce

$$||\boldsymbol{\sigma} - \boldsymbol{\sigma}_h||_0 \le C\Big(\sum_{K\in\mathcal{C}_h} (\eta_{1,K}^2 + \eta_{2,K}^2) + \sum_{E\in\Gamma_h} \eta_E^2 \Big)^{1/2}. \tag{3.37}$$

We now estimate the term $||u - u_h^*||_{1,h}$. We first recall that

$$||u - u_h^*||_{1,h} = \Big(\sum_{K\in\mathcal{C}_h} \|\nabla(u - u_h^*)\|_{0,K}^2 + \sum_{E\in\Gamma_h} h_E^{-1} \|[\![u - u_h^*]\!]\|_{0,E}^2 \Big)^{1/2} \tag{3.38}$$

and we notice that (cf. (3.2))

$$\left(\sum_{E\in\Gamma_h} h_E^{-1}\|[\![u-u_h^*]\!]\|_{0,E}^2\right)^{1/2} = \left(\sum_{E\in\Gamma_h} h_E^{-1}\|[\![u_h^*]\!]\|_{0,E}^2\right)^{1/2} = \left(\sum_{E\in\Gamma_h} \eta_E^2\right)^{1/2}. \tag{3.39}$$

We have

$$\begin{aligned}\|\nabla(u-u_h^*)\|_{0,K}^2 &= \big(\nabla u - \nabla u_h^*\,,\nabla(u-u_h^*)\big)_K = \big(\boldsymbol{\sigma} - \nabla u_h^*\,,\nabla(u-u_h^*)\big)_K\\ &= \big(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\nabla(u-u_h^*)\big)_K + \big(\boldsymbol{\sigma}_h - \nabla u_h^*\,,\nabla(u-u_h^*)\big)_K\\ &\le \Big(||\boldsymbol{\sigma}-\boldsymbol{\sigma}_h||_{0,K} + ||\boldsymbol{\sigma}_h - \nabla u_h^*||_{0,K}\Big)||\nabla(u-u_h^*)||_{0,K},\end{aligned} \tag{3.40}$$

by which we obtain

$$\|\nabla(u-u_h^*)\|_{0,K} \le ||\boldsymbol{\sigma}-\boldsymbol{\sigma}_h||_{0,K} + ||\boldsymbol{\sigma}_h - \nabla u_h^*||_{0,K}. \tag{3.41}$$

Hence we infer

$$\left(\sum_{K\in\mathcal{C}_h} ||\nabla(u-u_h^*)||_{0,K}^2\right)^{1/2} \le ||\boldsymbol{\sigma}-\boldsymbol{\sigma}_h||_0 + \left(\sum_{K\in\mathcal{C}_h} ||\boldsymbol{\sigma}_h - \nabla u_h^*||_{0,K}^2\right)^{1/2}. \tag{3.42}$$

Using (3.37) and recalling (3.1), from (3.42) we get

$$\left(\sum_{K\in\mathcal{C}_h} ||\nabla(u-u_h^*)||_{0,K}^2\right)^{1/2} \le C\Big(\sum_{K\in\mathcal{C}_h}(\eta_{1,K}^2+\eta_{2,K}^2) + \sum_{E\in\Gamma_h}\eta_E^2\Big)^{1/2}. \tag{3.43}$$

Therefore, joining (3.39) and (3.43) we obtain

$$||u-u_h^*||_{1,h} \le C\Big(\sum_{K\in\mathcal{C}_h}(\eta_{1,K}^2+\eta_{2,K}^2) + \sum_{E\in\Gamma_h}\eta_E^2\Big)^{1/2}. \tag{3.44}$$

From (3.37) and (3.44) we finally deduce (see (3.3))

$$||\boldsymbol{\sigma}-\boldsymbol{\sigma}_h||_0 + ||u-u_h^*||_{1,h} \le C\Big(\sum_{K\in\mathcal{C}_h}(\eta_{1,K}^2+\eta_{2,K}^2) + \sum_{E\in\Gamma_h}\eta_E^2\Big)^{1/2} = C\eta. \tag{3.45}$$

□

We end the paper by the following

Remark 3.4. *On the estimate in the* $\boldsymbol{H}(\mathrm{div}:\Omega)$*-norm.* In the paper we have repeatedly used the fact that by the equilibrium property (2.22) we have $\mathrm{div}\,(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h) = P_h f - f$ and hence $\|\mathrm{div}\,(\boldsymbol{\sigma}-\boldsymbol{\sigma}_h)\|_0 = \|f - P_h f\|_0$ is a quantity that is directly computable from the data to the problem. For the BDM spaces it furthermore holds that for a general loading and a smooth solution it holds $\|f-P_hf\|_0 = \mathcal{O}(h^k)$, whereas $\|\boldsymbol{\sigma}-\boldsymbol{\sigma}_h\|_0 = \mathcal{O}(h^{k+1})$, and hence this trivial component in the $\boldsymbol{H}(\mathrm{div}:\Omega)$ norm can dominate the whole estimate. □

References

[1] A. Alonso, *Error estimators for a mixed method*, Numer. Math., 74 (1994), pp. 385–395.

[2] D. N. Arnold and F. Brezzi, *Mixed and nonconforming finite element methods: implementation, postprocessing and error estimates*, RAIRO Modél. Math. Anal. Numér., 19 (1985), pp. 7–32.

[3] I. Babuška, J. Osborn, and J. Pitkäranta, *Analysis of mixed methods using mesh dependent norms*, Math. Comp., 35 (1980), pp. 1039–1062.

[4] D. Braess, *Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics*, Cambridge University Press, 2001.

[5] D. Braess and R. Verfürth, *A posteriori error estimators for the Raviart-Thomas element*, SIAM J. Numer. Anal., 33 (1996), pp. 2431–2444.

[6] J. H. Bramble and J. Xu, *A local post-processing technique for improving the accuracy in mixed finite-element approximations*, SIAM J. Numer. Anal., 26 (1989), pp. 1267–1275.

[7] F. Brezzi, J. Douglas, Jr., R. Durán, and M. Fortin, *Mixed finite elements for second order elliptic problems in three variables*, Numer. Math., 51 (1987), pp. 237–250.

[8] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, Springer-Verlag, 1991.

[9] F. Brezzi, J. Douglas Jr., and L. Marini, *Two families of mixed finite elements for second order elliptic problems.*, Numer. Math., 47 (1985), pp. 217–235.

[10] C. Carstensen, *A posteriori error estimate for the mixed finite element method*, Math. Comp., 66 (1997), pp. 465–476.

[11] E. Dari, R. Durán, C. Padra, and V. Vampa, *A posteriori error estimators for nonconforming finite element methods*, RAIRO Modél. Math. Anal. Numér., 30 (1996), pp. 385–400.

[12] V. Girault and P. Raviart, *Finite Element Methods for Navier-Stokes Equations. Theory and Algorithms*, Springer-Verlag, 1986.

[13] C. Lovadina and R. Stenberg, *A posteriori error analysis of the linked interpolation technique for plate bending problems*, SIAM J. Numer. Anal., Accepted for publication.

[14] J.-C. Nédélec, *A new family of mixed finite elements in* R^3, Numer. Math., 50 (1986), pp. 57–81.

[15] P. Raviart and J. Thomas, *A mixed finite element method for second order elliptic problems*, in Mathematical Aspects of the Finite Element Method. Lecture Notes in Math. 606, Springer-Verlag, 1977, pp. 292–315.

[16] R. Stenberg, *Some new families of finite elements for the Stokes equations*, Numer. Math., 56 (1990), pp. 827–838.

[17] R. Stenberg, *Postprocessing schemes for some mixed finite elements*, RAIRO Modél. Math. Anal. Numér., 25 (1991), pp. 151–167.

Dipartimento di Matematica, Università di Pavia and IMATI-CNR, VIa Ferrata 1, Pavia 27100, Italy

E-mail address: carlo.lovadina@unipv.it

Institute of Mathematics, Helsinki University of Technology, P.O. Box 1100, 02015 TKK, Finland

E-mail address: rolf.stenberg@tkk.fi