MATHEMATICS OF COMPUTATION
Volume 74, Number 252, Pages 1653–1677
S 0025-5718(05)01741-2
Article electronically published on March 1, 2005

ERROR ANALYSIS OF VARIABLE DEGREE MIXED METHODS FOR ELLIPTIC PROBLEMS VIA HYBRIDIZATION

BERNARDO COCKBURN AND JAYADEEP GOPALAKRISHNAN

Abstract. A new approach to error analysis of hybridized mixed methods is proposed and applied to study a new hybridized variable degree Raviart-Thomas method for second order elliptic problems. The approach gives error estimates for the Lagrange multipliers without using error estimates for the other variables. Error estimates for the primal and flux variables then follow from those for the Lagrange multipliers. In contrast, traditional error analyses obtain error estimates for the flux and primal variables first and then use it to get error estimates for the Lagrange multipliers. The new approach not only gives new error estimates for the new variable degree Raviart-Thomas method, but also new error estimates for the classical uniform degree method with less stringent regularity requirements than previously known estimates. The error analysis is achieved by using a variational characterization of the Lagrange multipliers wherein the other unknowns do not appear. This approach can be applied to other hybridized mixed methods as well.

1. Introduction

In this paper, we introduce a new approach to obtaining error estimates for hybridized mixed methods. It allows error analysis of the Lagrange multipliers *without* using error estimates for the other variables, as is customarily done. Moreover, it yields previously unknown error estimates with weaker regularity assumptions on the exact solution. The analysis is achieved by adopting a new point of view inspired by a recent variational characterization [7] of the Lagrange multipliers in which the flux and primal variable do not appear.

We illustrate this new approach by applying it to a *new* hybridized variable degree Raviart-Thomas (RT) mixed method for the following second order elliptic boundary value problem:

$$\boldsymbol{q} + \boldsymbol{a}(\boldsymbol{x})\ \mathbf{grad}\, u = 0, \quad \text{on } \Omega, \tag{1.1}$$

$$\operatorname{div} \boldsymbol{q} + d(\boldsymbol{x})\, u = f \quad \text{on } \Omega, \tag{1.2}$$

$$u = g \quad \text{on } \partial\Omega. \tag{1.3}$$

Here $\Omega \subset \mathbb{R}^N$ is a polyhedral domain ($N \geq 2$), $f \in L^2(\Omega)$, $g \in H^{1/2}(\partial\Omega)$, $\boldsymbol{a}(\boldsymbol{x})$ is a symmetric $N \times N$ matrix function with components in $L^\infty(\Omega)$, $\boldsymbol{a}(\boldsymbol{x})$ is uniformly

Received by the editor November 26, 2003 and, in revised form, August 1, 2004.
2000 *Mathematics Subject Classification.* Primary 65N30.
Key words and phrases. Mixed finite elements, hybrid methods, elliptic problems.
The first author is supported in part by the National Science Foundation (Grant DMS-0411254) and by the University of Minnesota Supercomputing Institute.
The second author is supported by the National Science Foundation (Grant DMS-0410030).

©2005 American Mathematical Society

positive definite on Ω, and $d(\boldsymbol{x})$ is a nonnegative function in $L^\infty(\Omega)$. This approach can also be applied to other variable degree methods, like the method proposed by Brezzi, Douglas and Marini [5] or the one introduced by Demkowicz et al. [8] (as discussed in the last section of this paper). We choose to introduce and analyze a variable degree version of the simplicial RT method because, to our knowledge, no such method exists in the literature. A variable degree RT method (in nonhybridized form) employing rectangles has been studied previously by Suri [16].

To place the results of this paper into proper perspective, let us discuss the increasing importance of hybridization to the study of mixed methods. In its inception in 1965 by Fraejis de Veubeke [9], hybridization was thought of merely as a clever implementation technique for mixed methods. However, twenty years later, it was realized that it also leads to better approximations. Indeed, Arnold and Brezzi [2] considered the RT method and showed how to use the Lagrange multiplier unknowns introduced by the hybridization procedure to enhance the accuracy of solutions by means of a local post-processing. Later, Brezzi, Douglas and Marini [4] applied a similar approach to a mixed method now known as the BDM method.

Hybridization has other well-known merits as well. The matrix system for the Lagrange multipliers is significantly smaller than that of the original mixed method. Indeed, when local spaces of polynomials of degree k are used, the number of degrees of freedom within an element increases like $O(k^N)$, whereas the number of degrees of freedom of the Lagrange multipliers on a face increases only like $O(k^{N-1})$. Furthermore, the recent simplifications in [7] that allow easy and direct computation of the matrix equation for the multipliers make hybridization even more appealing.

Another use of hybridization, namely its use as a device for avoiding implementation of transition elements in variable degree methods, was pointed out by Kirby and Dawson [13]. To put this in perspective, recall that in 1985, Brezzi, Douglas and Marini [5] proposed a natural variable degree extension of their original BDM method by introducing suitably defined transition elements. These elements were constructed in order to maintain the inter-element continuity of the normal component of the flux when joining two BDM elements of different degree. Kirby and Dawson [13] showed how the method of [5] can be conveniently implemented without the use of transition elements by hybridizing it.

In this paper, we show that hybridization helps in the construction *and* analysis of variable degree mixed methods. The main difficulty in the construction and analysis of variable degree versions of mixed methods is to ensure that the variation of the polynomial degree does not destroy the delicate stability of the methods, as manifested in the inf-sup condition. We overcome this difficulty by obtaining a variational formulation of the Lagrange multipliers for which stability is obvious: Indeed, on the space of the Lagrange multipliers $M_{\mathcal{T}}$, we will give a bilinear form $a_{\mathcal{T}}(\cdot,\cdot)$ and a linear form $b_{\mathcal{T}}(\cdot)$ (which depends on mesh $\mathcal{T}$) such that the discrete Lagrange multiplier $\lambda_{\mathcal{T}}$ is the only element of the space $M_{\mathcal{T}}$ satisfying

$$a_{\mathcal{T}}(\lambda_{\mathcal{T}},\mu) = b_{\mathcal{T}}(\mu) \qquad \text{for all } \mu \in M_{\mathcal{T}}.$$

This result shapes our approach to error estimation: Because of it, it is reasonable to expect an error estimate in the "energy" norm $a_{\mathcal{T}}(\mu,\mu)^{1/2}$. But it turns out that the "consistency error"

$$\tau(\mu) := a_{\mathcal{T}}(u,\mu) - b_{\mathcal{T}}(\mu),$$

where u is the exact solution of (1.1)–(1.3), is *nonzero*, in general. This and the mesh dependent nature of $a_{\mathcal{T}}(\cdot,\cdot)$ makes the error analysis reminiscent of finite element methods with variational crimes. See, e.g., [3] for a general theory of variational crimes and [14] for the application of such an approach to the analysis of discontinuous Galerkin methods. Previous papers have studied other relationships between mixed and nonconforming methods (see, e.g., [1, 2]).

Where we depart from all the above-mentioned analyses is in showing that the truncation error $\tau(\mu)$ admits a particularly simple bound in terms of projections that have become classical tools in the analysis of mixed methods. In the traditional analysis of mixed methods, one such projection is well known to be an essential tool in proving the inf-sup condition. In contrast, in our approach, we only use it to bound the consistency error $\tau(\mu)$. While the construction of such projections is very well known for uniform degree mixed methods, their generalization to the variable degree case is not entirely simple. Such a generalization was studied in the early work on the variable degree BDM method [5], but results were achieved only under the assumption that the maximum variation of the polynomial degrees of two adjacent elements be one. This restriction was removed by Demkowicz et al. [8] who constructed projections for BDM-type spaces with arbitrarily varying polynomial degrees. We will adapt their ideas to construct an appropriate projection for the variable degree RT method.

The paper is organized as follows. In Section 2, we introduce and discuss the hybridized variable degree method. In particular, we give a variational characterization of the Lagrange multipliers. This is a particular case of an abstract version of the results in [7] and is essential for the analysis of the method. The abstract result is discussed in Appendix A. In Section 3, we state and discuss the error estimates for the Lagrange multipliers as well as how error estimates for the other two variables follow from them. The error estimates are stated in terms of an extension of the well-known Raviart-Thomas projection to the variable degree case. Details of this projection, constructed following [8], are in Appendix B. Section 4 is devoted to obtaining key auxiliary lemmas. Section 5 is devoted to the proofs of the error estimates in L^2-like norms and Section 6 to proving an equivalence between the energy norm and a more transparent norm. We end in Section 7 by briefly showing how to extend this analysis to the variable degree BDM method.

2. The method

In this section, we introduce the hybridized variable degree RT method. We begin by recalling the classical uniform degree RT method and its hybridization. We then present its variable degree version and prove that it is well defined. Finally, we state the result that characterizes the approximate solution and then we briefly discuss the new approach to error analysis.

2.1. The variable degree hybridized RT method.

Given a triangulation of Ω, $\mathcal{T}$, made of simplices, the RT mixed method seeks an approximation $(\boldsymbol{q}_h, u_h)$ in the finite element space $\mathbb{V}_h^k \times \mathbb{W}_h^k$ given by

$$\begin{aligned}
\mathbb{V}_h^k &= \{\boldsymbol{v} \in H(\mathrm{div},\Omega): \quad \boldsymbol{v}|_K \in \boldsymbol{\mathcal{V}}^k(K) \quad \text{for all } \ K \in \mathcal{T}\},\\
\mathbb{W}_h^k &= \{w \in L^2(\Omega): \qquad w|_K \in \mathcal{W}^k(K) \quad \text{for all } \ K \in \mathcal{T}\},
\end{aligned}$$

where the local spaces $\boldsymbol{\mathcal{V}}^k(K)$ and $\mathcal{W}^k(K)$ are given by

$$\boldsymbol{\mathcal{V}}^k(K) = (P^k(K))^N + \boldsymbol{x}\, P^k(K), \tag{2.1}$$

$$\mathcal{W}^k(K) = P^k(K). \tag{2.2}$$

Here, $P^k(K)$ denotes the space of polynomials on K of degree at most k, $k \geq 0$. The classical uniform degree method defines the approximation $(\boldsymbol{q}_h, u_h)$ by requiring that, for all $(\boldsymbol{v}, w) \in \mathbb{V}_h \times \mathbb{W}_h$,

$$\int_\Omega \boldsymbol{c}\, \boldsymbol{q}_h \cdot \boldsymbol{v}\, \mathrm{d}x - \int_\Omega u_h \operatorname{div} \boldsymbol{v}\, \mathrm{d}x = -\int_{\partial\Omega} g\, \boldsymbol{v} \cdot \boldsymbol{n}\, \mathrm{d}s,$$
$$\int_\Omega w \operatorname{div} \boldsymbol{q}_h\, \mathrm{d}x + \int_\Omega d\, u_h\, w\, \mathrm{d}x = \int_\Omega f\, w\, \mathrm{d}x,$$

where $\boldsymbol{c} = \boldsymbol{a}^{-1}$.

Next, let us introduce the hybridized version of the above method. Such a method seeks an approximation $(\boldsymbol{q}_h, u_h, \lambda_h)$ to $(\boldsymbol{q}, u, u|_{\mathcal{E}_i})$, where $\mathcal{E}_i$ denotes the set of interior faces. It is sought in the finite element space $\boldsymbol{V}_h^k \times W_h^k \times M_h^k$ defined by

$$\boldsymbol{V}_h^k = \{\boldsymbol{v} \in (L^2(\Omega))^N : \quad \boldsymbol{v}|_K \in \boldsymbol{\mathcal{V}}^k(K) \quad \text{for all } K \in \mathcal{T}\},$$
$$W_h^k = \{w \in L^2(\Omega) : \quad w|_K \in \mathcal{W}^k(K) \quad \text{for all } K \in \mathcal{T}\},$$
$$M_h^k = \{\mu \in L^2(\mathcal{E}_i) : \quad \mu|_e \in \mathcal{M}^k(e) \quad \text{for all } e \in \mathcal{E}_i\},$$

where the local space $\mathcal{M}^k(e)$ is given by

$$\mathcal{M}^k(e) = P^k(e), \tag{2.3}$$

and is defined by requiring that, for all $(\boldsymbol{v}, w, \mu) \in \boldsymbol{V}_h^k \times W_h^k \times M_h^k$,

$$\int_\Omega \boldsymbol{c}\, \boldsymbol{q}_h \cdot \boldsymbol{v}\, \mathrm{d}x - \sum_{K \in \mathcal{T}} \int_K u_h \operatorname{div} \boldsymbol{v}\, \mathrm{d}x + \sum_{e \in \mathcal{E}_i} \int_e \lambda_h\, [\boldsymbol{v}]\, \mathrm{d}s = -\int_{\partial\Omega} g\, [\boldsymbol{v}]\, \mathrm{d}s,$$
$$\sum_{K \in \mathcal{T}} \int_K w \operatorname{div} \boldsymbol{q}_h\, \mathrm{d}x + \int_\Omega d\, u_h\, w\, \mathrm{d}x = \int_\Omega f\, w\, \mathrm{d}x,$$
$$\sum_{e \in \mathcal{E}_i} \int_e \mu\, [\boldsymbol{q}_h]\, \mathrm{d}s = 0,$$

where $[\boldsymbol{v}] = \boldsymbol{v} \cdot \boldsymbol{n}$ on $\partial\Omega$, and $[\boldsymbol{v}] = \boldsymbol{v}_e^+ \cdot \boldsymbol{n}_e^+ + \boldsymbol{v}_e^- \cdot \boldsymbol{n}_e^-$, for all $e \in \mathcal{E}_i$. Here, $\boldsymbol{n}$ denotes the outward unit normal to Ω, $\boldsymbol{n}_e^+ = -\boldsymbol{n}_e^-$ is an arbitrary unit vector normal to e and $\boldsymbol{v}_e^\pm(\boldsymbol{x}) = \lim_{\epsilon \downarrow 0} \boldsymbol{v}(\boldsymbol{x} - \epsilon\, \boldsymbol{n}_e^\pm)$. It is well known that the above method is well defined and that $(\boldsymbol{q}_h, u_h)$ given by this method *coincides* with the approximate solution given by the original RT method (see, e.g., [6, Section V.1]).

The definition of the hybridized variable degree RT method can be obtained by simply *replacing* the finite dimensional space $\boldsymbol{V}_h^k \times W_h^k \times M_h^k$ by the finite dimensional space $\boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}} \times M_{\mathcal{T}}$ defined by

$$\boldsymbol{V}_{\mathcal{T}} = \{\boldsymbol{v} \in (L^2(\Omega))^N : \quad \boldsymbol{v}|_K \in \boldsymbol{V}(K) \quad \text{for all } K \in \mathcal{T}\}, \tag{2.4}$$

$$W_{\mathcal{T}} = \{w \in L^2(\Omega) : \quad w|_K \in W(K) \quad \text{for all } K \in \mathcal{T}\}, \tag{2.5}$$

$$M_{\mathcal{T}} = \{\mu \in L^2(\mathcal{E}_i) : \quad \mu|_e \in M(e) \quad \text{for all } e \in \mathcal{E}_i\}, \tag{2.6}$$

where

$$\boldsymbol{V}(K) = \boldsymbol{\mathcal{V}}^{k(K)}(K), \quad W(K) = \mathcal{W}^{k(K)}(K), \quad \text{and} \quad M(e) = \mathcal{M}^{k(e)}(e).$$

Here $k(K)$ for $K \in \mathcal{T}$ lies in an arbitrary but fixed set of nonnegative integers and

$$k(e) = \max\{k(K_e^+), k(K_e^-)\}, \tag{2.7}$$

where $K_e^{\pm}$ are the two simplices in $\mathcal{T}$ that share the interior face e. Thus, we consider the approximation $(\boldsymbol{q}_{\mathcal{T}}, u_{\mathcal{T}}, \lambda_{\mathcal{T}})$ in the space $\boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}} \times M_{\mathcal{T}}$ satisfying

$$\int_\Omega \boldsymbol{c}\, \boldsymbol{q}_{\mathcal{T}} \cdot \boldsymbol{v} \,\mathrm{d}x - \sum_{K\in\mathcal{T}} \int_K u_{\mathcal{T}} \operatorname{div} \boldsymbol{v} \,\mathrm{d}x + \sum_{e\in\mathcal{E}_i} \int_e \lambda_{\mathcal{T}}\, [\boldsymbol{v}] \,\mathrm{d}s = -\int_{\partial\Omega} g\, [\boldsymbol{v}] \,\mathrm{d}s, \tag{2.8}$$

$$\sum_{K\in\mathcal{T}} \int_K w \operatorname{div} \boldsymbol{q}_{\mathcal{T}} \,\mathrm{d}x + \int_\Omega d\, u_{\mathcal{T}}\, w \,\mathrm{d}x = \int_\Omega f\, w \,\mathrm{d}x, \tag{2.9}$$

$$\sum_{e\in\mathcal{E}_i} \int_e \mu\, [\boldsymbol{q}_{\mathcal{T}}] \,\mathrm{d}s = 0 \tag{2.10}$$

for all $(\boldsymbol{v}, w, \mu) \in \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}} \times M_{\mathcal{T}}$. These equations yield a well-defined method, as we see in the next result.

Proposition 2.1. *There exists a unique* $(\boldsymbol{q}_{\mathcal{T}}, u_{\mathcal{T}}, \lambda_{\mathcal{T}}) \in \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}} \times M_{\mathcal{T}}$ *satisfying the discrete formulation* (2.8), (2.9) *and* (2.10).

Proof. To prove the existence and uniqueness of $(\boldsymbol{q}_{\mathcal{T}}, u_{\mathcal{T}}, \lambda_{\mathcal{T}})$, it suffices to show that the only solution for the case in which the boundary data g and the right-hand side f are equal to zero is the trivial one. In such a case, if we extend $\lambda_{\mathcal{T}}$ by zero to $\partial\Omega$, the first two equations defining the approximate solution become

$$\sum_{K\in\mathcal{T}} \left(\int_K \boldsymbol{c}\boldsymbol{q}_{\mathcal{T}} \cdot \boldsymbol{v} \,\mathrm{d}x - \int_K u_{\mathcal{T}} \operatorname{div} \boldsymbol{v} \,\mathrm{d}x + \int_{\partial K} \lambda_{\mathcal{T}} \boldsymbol{v} \cdot \boldsymbol{n}_K \,\mathrm{d}s \right) = 0,$$

$$\sum_{K\in\mathcal{T}} \left(\int_K w \operatorname{div} \boldsymbol{q}_{\mathcal{T}} \,\mathrm{d}x + \int_K d\, u_{\mathcal{T}}\, w \,\mathrm{d}x \right) = 0,$$

for all $\boldsymbol{v} \in \boldsymbol{V}_{\mathcal{T}}$ and all $w \in W_{\mathcal{T}}$. Choosing $w = u_{\mathcal{T}}$ and $\boldsymbol{v} = \boldsymbol{q}_{\mathcal{T}}$, adding the above equations, and using (2.10), we obtain

$$\int_\Omega \boldsymbol{c}\boldsymbol{q}_{\mathcal{T}} \cdot \boldsymbol{q}_{\mathcal{T}} \,\mathrm{d}x + \int_\Omega d u_{\mathcal{T}}^2 \,\mathrm{d}x = 0.$$

Thus $\boldsymbol{q}_{\mathcal{T}} \equiv 0$.

Next, we show that $\lambda_{\mathcal{T}} \equiv 0$ as well. Since $\boldsymbol{q}_{\mathcal{T}} = 0$, we have, for all $\boldsymbol{v} \in \boldsymbol{V}(K)$,

$$\begin{aligned} 0 &= -\int_K u_{\mathcal{T}} \operatorname{div} \boldsymbol{v} \,\mathrm{d}x + \int_{\partial K} \lambda_{\mathcal{T}} \boldsymbol{v} \cdot \boldsymbol{n}_K \,\mathrm{d}s \qquad (2.11)\\ &= \int_K \boldsymbol{v} \cdot \mathbf{grad}\, u_{\mathcal{T}} \,\mathrm{d}x + \int_{\partial K} (\lambda_{\mathcal{T}} - u_{\mathcal{T}}) \boldsymbol{v} \cdot \boldsymbol{n}_K \,\mathrm{d}s \\ &= \int_K \boldsymbol{v} \cdot \mathbf{grad}\, u_{\mathcal{T}} \,\mathrm{d}x + \int_{\partial K} (P_{\partial K}\lambda_{\mathcal{T}} - u_{\mathcal{T}})\, \boldsymbol{v} \cdot \boldsymbol{n}_K \,\mathrm{d}s, \qquad (2.12) \end{aligned}$$

where the function $P_{\partial K}\lambda_{\mathcal{T}}$ is defined on each face e of ∂K as the element of $\mathcal{M}^{k(K)}(e)$ satisfying

$$\int_e (P_{\partial K}\lambda_{\mathcal{T}})\, \mu \,\mathrm{d}s = \int_e \lambda_{\mathcal{T}}\, \mu \,\mathrm{d}s, \qquad \text{for all } \mu \in \mathcal{M}^{k(K)}(e). \tag{2.13}$$

Using the properties of the local space $\boldsymbol{V}(K)\times W(K)$ (see, e.g., [6]), we can find a function $\boldsymbol{r}$ in $\boldsymbol{V}(K)$ such that

$$\boldsymbol{r}\cdot\boldsymbol{n}|_e = (P_{\partial K}\lambda_{\mathcal{T}} - u_{\mathcal{T}})\,|_e \in \mathcal{M}^{k(K)}(e) \quad \text{for all } e\in\partial K,$$

and

$$\int_K \boldsymbol{r}\cdot\mathbf{grad}\, w\, dx = 0, \quad \text{for all } w\in W(K).$$

Using this function in equation (2.12), we obtain that

$$(P_{\partial K}\lambda_{\mathcal{T}} - u_{\mathcal{T}})|_{\partial K} = 0 \quad \text{and} \quad \int_K \boldsymbol{v}\cdot\mathbf{grad}\, u_{\mathcal{T}}\, \mathrm{d}x = 0,$$

for all $\boldsymbol{v}\in\boldsymbol{V}(K)$ and for all $K\in\mathcal{T}$. Hence, $u_{\mathcal{T}}$ is constant on K and, as a consequence, $P_{\partial K}\lambda_{\mathcal{T}}$ is constant on ∂K for every $K\in\mathcal{T}$.

Now, since, for every face $e\in\mathcal{E}_i$, we have

$$(P_{\partial K_e^+}\lambda_{\mathcal{T}})|_e \in \mathcal{M}^{k(K_e^+)}(e), \quad (P_{\partial K_e^-}\lambda_{\mathcal{T}})|_e \in \mathcal{M}^{k(K_e^-)}(e) \quad \text{and} \quad \lambda_{\mathcal{T}}|_e \in M(e),$$

and since, by equation (2.7), $k(e) = \max\{k(K_e^+), k(K_e^-)\}$, one of the functions $(P_{\partial K_e^+}\lambda_{\mathcal{T}})|_e$ and $(P_{\partial K_e^-}\lambda_{\mathcal{T}})|_e$ must coincide with $\lambda_{\mathcal{T}}$. Hence $\lambda_{\mathcal{T}}$ must be constant on e as $P_{\partial K_e^\pm}\lambda_{\mathcal{T}}$ is constant on $\partial K_e^\pm$. But if $\lambda_{\mathcal{T}}$ is constant on every edge $e\in\mathcal{E}_i$, then $\lambda_{\mathcal{T}}$ coincides with $P_{\partial K}\lambda_{\mathcal{T}}$ on all $K\in\mathcal{T}$, and $\lambda_{\mathcal{T}}|_{\partial K}$ is constant on ∂K for all $K\in\mathcal{T}$. Since we extended $\lambda_{\mathcal{T}}$ by zero to $\partial\Omega$, we conclude that $\lambda_{\mathcal{T}}\equiv 0$.

This implies that we can now write equation (2.11) as

$$0 = -\int_K u_{\mathcal{T}}\,\mathrm{div}\,\boldsymbol{v}\,\mathrm{d}x.$$

As a consequence, $u_{\mathcal{T}}$ vanishes on all $K\in\mathcal{T}$ because $\mathrm{div}\,\boldsymbol{V}(K) = W(K)$. This completes the proof. □

2.2. **A characterization of the approximate solution.** Next, we give a characterization of the approximate solution which is a straightforward extension of an analogous result in [7] for uniform degree hybridized RT methods. This result is also the main tool for the error analysis.

To state it, we introduce some local mappings. The first mapping lifts functions on faces of the simplices of the triangulation $\mathcal{T}$ to functions on Ω. Let $\mathcal{E}$ be the set of all faces of the triangulation $\mathcal{T}$. Notwithstanding a slight abuse of notation, we denote the set of all square integrable functions on the union of all faces of $\mathcal{E}$ by $L^2(\mathcal{E})$. The lifting associates to each $\mathsf{m}\in L^2(\mathcal{E})$ the pair $(\boldsymbol{\mathfrak{Q}}\mathsf{m}\,,\,\mathfrak{U}\mathsf{m}\,)\in\boldsymbol{V}_{\mathcal{T}}\times W_{\mathcal{T}}$ defined by requiring that

$$\int_\Omega c\,\boldsymbol{\mathfrak{Q}}\mathsf{m}\,\cdot\boldsymbol{v}\,\mathrm{d}x - \sum_{K\in\mathcal{T}}\int_K \mathfrak{U}\mathsf{m}\,\mathrm{div}\,\boldsymbol{v}\,\mathrm{d}x = -\sum_{e\in\mathcal{E}}\int_e \mathsf{m}\,[\boldsymbol{v}]\,\mathrm{d}s, \tag{2.14}$$

$$\sum_{K\in\mathcal{T}}\int_K w\,\mathrm{div}\,\boldsymbol{\mathfrak{Q}}\mathsf{m}\,\mathrm{d}x + \int_\Omega d\,\mathfrak{U}\mathsf{m}\,w\,\mathrm{d}x = 0, \tag{2.15}$$

hold for all $(\boldsymbol{v},w)\in\boldsymbol{V}_{\mathcal{T}}\times W_{\mathcal{T}}$.

The other local mappings are defined similarly: For any $f \in L^2(\Omega)$ the pair $(\mathsf{Q}f\,,\,\mathsf{U}f\,) \in \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}}$ and is defined by requiring that

$$\int_\Omega c\,\mathsf{Q}f\;\cdot \boldsymbol{v}\,\mathrm{d}x - \sum_{K\in\mathcal{T}} \int_K \mathsf{U}f\;\operatorname{div}\boldsymbol{v}\,\mathrm{d}x = 0, \tag{2.16}$$

$$\sum_{K\in\mathcal{T}} \int_K w \operatorname{div}\mathsf{Q}f\;\mathrm{d}x + \int_\Omega d\;\mathsf{U}f\;w\,\mathrm{d}x = \int_\Omega f\,w\,\mathrm{d}x \tag{2.17}$$

hold for all $(\boldsymbol{v}, w) \in \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}}$. For any $\boldsymbol{\alpha} \in (L^2(\Omega))^N$ the pair $(\mathfrak{Q}\boldsymbol{\alpha}\;,\,\mathfrak{U}\boldsymbol{\alpha}\;) \in \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}}$ and is defined by requiring that

$$\int_\Omega c\,\mathfrak{Q}\boldsymbol{\alpha}\;\cdot \boldsymbol{v}\,\mathrm{d}x - \sum_{K\in\mathcal{T}} \int_K \mathfrak{U}\boldsymbol{\alpha}\;\operatorname{div}\boldsymbol{v}\,\mathrm{d}x = \int_\Omega \boldsymbol{\alpha}\cdot\boldsymbol{v}\,\mathrm{d}x, \tag{2.18}$$

$$\sum_{K\in\mathcal{T}} \int_K w \operatorname{div}\mathfrak{Q}\boldsymbol{\alpha}\;\mathrm{d}x + \int_\Omega d\;\mathfrak{U}\boldsymbol{\alpha}\;w\,\mathrm{d}x = 0 \tag{2.19}$$

hold for all $(\boldsymbol{v}, w) \in \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}}$.

The above three pairs of maps are all locally defined, e.g., the value of $(\mathfrak{Q}\mathsf{m}\,,\,\mathfrak{U}\mathsf{m}\,)$ on an element K is completely determined as the solution of the mixed method on the single element K with data determined by the value of m on ∂K. Note that the three local maps are obtained using three different data functionals, but using the same mixed problem. The solvability of the mixed problem on one element, ensured by Proposition 2.1, thus shows that the above maps are well defined. Although all three maps use the same mixed problem, we have chosen to explicitly distinguish each of them so as to delineate the dependence of the final solution on the data components and the Lagrange multipliers (as in Theorem 2.2 below).

In order to emphasize the underlying matrix structure of the problem, we rewrite the hybridized method in terms of the operators $\boldsymbol{A} : \boldsymbol{V}_{\mathcal{T}} \mapsto \boldsymbol{V}_{\mathcal{T}}$, $B : \boldsymbol{V}_{\mathcal{T}} \mapsto W_{\mathcal{T}}$, $C : \boldsymbol{V}_{\mathcal{T}} \mapsto M_{\mathcal{T}}$, and $D : M_{\mathcal{T}} \mapsto M_{\mathcal{T}}$ defined by

$$\begin{aligned} (\boldsymbol{A}\boldsymbol{p}, \boldsymbol{r})_\Omega &= \int_\Omega \boldsymbol{c}\,\boldsymbol{p}\cdot\boldsymbol{r}\;\mathrm{d}x, & (B\boldsymbol{p}, v)_\Omega &= \sum_{K\in\mathcal{T}} \int_K v\;\operatorname{div}\boldsymbol{p}\;\mathrm{d}x, \\ (C\boldsymbol{p}, \mu)_{\mathcal{E}} &= -\sum_{K\in\mathcal{T}} \int_{\partial K} \mu\,\boldsymbol{p}\cdot\boldsymbol{n}_K\;\mathrm{d}s, & (Dw, v)_\Omega &= \int_\Omega d\,wv\;\mathrm{d}x, \end{aligned} \tag{2.20}$$

for all $\boldsymbol{p} \in \boldsymbol{V}_{\mathcal{T}}, \boldsymbol{r} \in \boldsymbol{V}_{\mathcal{T}}$, $w \in W_{\mathcal{T}}, v \in W_{\mathcal{T}}$, and $\mu \in M_{\mathcal{T}}$. Here and elsewhere, for any domain $\mathcal{O}$, $(\cdot,\cdot)_{\mathcal{O}}$ denotes the $L^2(\mathcal{O})$ (or $L^2(\mathcal{O})^N$) inner product. To simplify notation, we do not distinguish between functions μ defined on $\mathcal{E}_i$ and their extension by zero on $\partial\Omega$, so we have written $\int_{\partial K} \mu\,\boldsymbol{p}\cdot\boldsymbol{n}_K\;\mathrm{d}s$ above rather than $\int_{\partial K\setminus\partial\Omega} \mu\,\boldsymbol{p}\cdot\boldsymbol{n}_K\;\mathrm{d}s$. Finally, we denote by $\boldsymbol{\alpha}_g$ the element of $\boldsymbol{V}_{\mathcal{T}}$ defined by

$$(\boldsymbol{\alpha}_g, \boldsymbol{v})_\Omega = -(g, \boldsymbol{v}\cdot\boldsymbol{n})_{\partial\Omega} \quad \text{for all } \boldsymbol{v} \in \boldsymbol{V}_{\mathcal{T}}. \tag{2.21}$$

With the above notation, the equations defining the approximate solution, namely, the equations (2.8)–(2.10), can be rewritten as

$$\begin{pmatrix} \boldsymbol{A} & -B^t & -C^t \\ B & D & 0 \\ C & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{q}_{\mathcal{T}} \\ u_{\mathcal{T}} \\ \lambda_{\mathcal{T}} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_g \\ \mathsf{P}f \\ 0 \end{pmatrix}, \tag{2.22}$$

where the superscript t denotes L^2-adjoints and $\mathsf{P}f$ denotes the $L^2(\Omega)$-orthogonal projection of f onto $W_{\mathcal{T}}$. Solutions of systems of this type can be characterized in terms of the local mappings introduced above, as stated in the following theorem.

Theorem 2.2 (Characterization of approximate solutions)**.** *Suppose* $(\boldsymbol{p}, v, \lambda) \in \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}} \times M_{\mathcal{T}}$ *satisfies*

$$\begin{pmatrix} \boldsymbol{A} & -B^t & -C^t \\ B & D & 0 \\ C & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{p} \\ v \\ \lambda \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha} \\ \beta \\ 0 \end{pmatrix}, \tag{2.23}$$

for some functions $\boldsymbol{\alpha} \in \boldsymbol{V}_{\mathcal{T}}$ *and* $\beta \in T_{\mathcal{T}}$*. Then*

$$\begin{aligned} \boldsymbol{p} &= \boldsymbol{\mathfrak{Q}}\boldsymbol{\alpha} + \mathbf{Q}\beta + \boldsymbol{\mathfrak{Q}}\lambda\,, \\ v &= \mathfrak{U}\boldsymbol{\alpha} + \mathsf{U}\beta + \mathfrak{U}\lambda\,, \end{aligned}$$

and λ *is the unique function in* $M_{\mathcal{T}}$ *that satisfies*

$$a_{\mathcal{T}}(\lambda, \mu) = b_{\mathcal{T}}(\mu), \qquad \text{for all } \mu \in M_{\mathcal{T}}, \tag{2.24}$$

where

$$a_{\mathcal{T}}(\lambda, \mu) = \int_\Omega \boldsymbol{c}\,\boldsymbol{\mathfrak{Q}}\lambda \cdot \boldsymbol{\mathfrak{Q}}\mu \,\mathrm{d}x + \int_\Omega d\ \mathfrak{U}\lambda\ \mathfrak{U}\mu \,\mathrm{d}x, \tag{2.25}$$

$$b_{\mathcal{T}}(\mu) = (\beta,\, \mathfrak{U}\mu\,)_\Omega + (\boldsymbol{\alpha}, \boldsymbol{\mathfrak{Q}}\mu\,)_\Omega, \tag{2.26}$$

for all η *and* $\mu \in M_{\mathcal{T}}$*.*

Proof. We apply Theorem A.1 in Appendix A. All conditions of Theorem A.1 can be easily verified using the well-known properties of the Raviart-Thomas pair of spaces on an element. It only remains to show that (2.24) has a unique solution. But this follows immediately from the uniqueness of solutions of (2.23) established in Proposition 2.1. □

3. The main results

In this section, we state and discuss our main results. We begin by stating an error estimate for the Lagrange multiplier in the "energy" norm

$$\|\mu\|_a := a_{\mathcal{T}}(\mu, \mu)^{1/2}$$

on the space of the Lagrange multipliers. We then state the error estimates for the flux and primal variables ($\boldsymbol{q}_{\mathcal{T}}$ and $u_{\mathcal{T}}$) which follow from it. We also give a result that elucidates the nature of this norm. Finally, we obtain error estimates in weaker L^2-like norms which imply superconvergence estimates for the primal variable. Our analysis critically relies on the linearity of the problem.

3.1. **Error estimates in energy norms.** Our first error estimate bounds the distance between the Lagrange multiplier $\lambda_{\mathcal{T}}$ and the $L^2(\mathcal{E})$-orthogonal projection into $M_{\mathcal{T}}$ of the exact solution u, which we denote by Pu; i.e., $Pv \in M_{\mathcal{T}}$ is defined by

$$\int_e (Pv)\,\eta\, d\gamma = \int_e v\,\eta\, d\gamma, \qquad \text{for all } \eta \in M_{\mathcal{T}} \text{ for all } e \in \mathcal{E}_i.$$

The bound is written in terms of two other projections. One is a projection into the flux space: $\Pi : \boldsymbol{V} \mapsto \boldsymbol{V}_{\mathcal{T}}$, where the domain of definition of Π, namely $\boldsymbol{V}$, is a

subspace of $\boldsymbol{H}(\mathrm{div},\Omega)$. In our analysis we require that Π is such that the following diagram commutes:

$$\begin{array}{ccc} \boldsymbol{V} & \xrightarrow{\ \mathrm{div}\ } & L^2(\Omega) \\ {\scriptstyle\Pi}\big\downarrow & & \big\downarrow{\scriptstyle\mathsf{P}} \\ \boldsymbol{V}_{\mathcal{T}} & \xrightarrow{\ \mathrm{div}_{\mathcal{T}}\ } & W_{\mathcal{T}}, \end{array} \tag{3.1}$$

where $\mathrm{div}_{\mathcal{T}}$ denotes "piecewise divergence", i.e., for $\boldsymbol{v} \in \boldsymbol{V}_{\mathcal{T}}$, $(\mathrm{div}_{\mathcal{T}}\,\boldsymbol{v})|_K = \mathrm{div}(\boldsymbol{v}|_K)$ for all $K \in \mathcal{T}$. Such projections have become a classical tool in the analysis of mixed finite element methods [6]. The most well known example of a projection Π satisfying (3.1) is the Raviart-Thomas interpolation operator Π_R which has been used to analyze the uniform degree case traditionally [6, 15]. Although Π_R continues to satisfy (3.1) even in the variable degree case, an important difference between the uniform degree case and the variable degree case is that while in the former case Π_R satisfies

$$\mathrm{Range}(\Pi_R) \subseteq \boldsymbol{V}_{\mathcal{T}} \cap \boldsymbol{H}(\mathrm{div},\Omega), \tag{3.2}$$

in the variable degree case Π no longer satisfies (3.2). The traditional analysis of mixed methods requires that both (3.1) and (3.2) hold.

It is possible to construct a projection Π that satisfies both (3.1) and (3.2) using the ideas in [8]. We detail this construction in Appendix B. It is this projection that appears in all our error estimates. Note that in contrast to the traditional mixed analysis, it is possible to use our approach to get error estimates even when (3.2) does not hold (see Remark 3.1). However, if both (3.1) and (3.2) hold, then our estimates are simpler and more elegant. This is the reason we have chosen to use the projection constructed in Appendix B in all our error estimates.

The remaining projection that appears in the following theorem, namely $\mathsf{P}u$, is the $L^2(\Omega)$-orthogonal projection into $W_{\mathcal{T}}$ of the exact solution u. Define the norms

$$\|\boldsymbol{r}\|_{\boldsymbol{c}} = \left(\int_\Omega \boldsymbol{c}\,\boldsymbol{r}\cdot\boldsymbol{r}\ \mathrm{d}x\right)^{1/2} \quad \text{and} \quad \|v\|_d = \left(\int_\Omega d\,v^2\ \mathrm{d}x\right)^{1/2}.$$

Note that the former is a weighted L^2-norm, while the latter is, in general, only a seminorm. Recall that $\boldsymbol{q}$ and u denote the exact solution components defined by (1.1)–(1.3); we tacitly assume that $\boldsymbol{q} \in \boldsymbol{V}$. The following theorem holds without any uniformity assumptions on the mesh and contains no unknown constants.

Theorem 3.1. *For any $\mu \in M_{\mathcal{T}}$,*

$$a_{\mathcal{T}}(\lambda_{\mathcal{T}} - Pu, \mu) = \int_\Omega \boldsymbol{c}\,(\Pi\boldsymbol{q} - \boldsymbol{q}) \cdot \boldsymbol{\Omega}\mu\ \mathrm{d}x + \int_\Omega d(u - \mathsf{P}u)\,\mathfrak{U}\mu\ \mathrm{d}x. \tag{3.3}$$

Consequently,

$$|\!|\!|\lambda_{\mathcal{T}} - Pu|\!|\!|_a^2 \ \le\ \|\boldsymbol{q} - \Pi\boldsymbol{q}\|_{\boldsymbol{c}}^2 + \|u - \mathsf{P}u\|_d^2. \tag{3.4}$$

Notice that the above result implies control of the consistency error mentioned in the introduction: Since $\tau(\mu) := a_{\mathcal{T}}(u,\mu) - b_{\mathcal{T}}(\mu) = a_{\mathcal{T}}(Pu,\mu) - a_{\mathcal{T}}(\lambda_{\mathcal{T}},\mu)$, the theorem gives

$$\sup_{\mu\in M_{\mathcal{T}}} \frac{\tau(\mu)}{|\!|\!|\mu|\!|\!|_a} \le \left(\|\boldsymbol{q} - \Pi\boldsymbol{q}\|_{\boldsymbol{c}}^2 + \|u - \mathsf{P}u\|_d^2\right)^{1/2}.$$

The proof of Theorem 3.1 is extremely simple. It follows directly from the characterization of the multipliers given by Theorem 2.2 and the commuting diagram

property. The use of the latter provides a precise description of the structure of the error as shown in the next lemma.

Lemma 3.2 (The error equations)**.** *The discrete error functions defined by*

$$e_{\boldsymbol{q}} = \Pi \boldsymbol{q} - \boldsymbol{q}_{\mathcal{T}}, \quad e_u = \mathsf{P}u - u_{\mathcal{T}} \quad \text{and} \quad e_\lambda = Pu - \lambda_{\mathcal{T}},$$

satisfy

$$\begin{pmatrix} \boldsymbol{A} & -B^t & -C^t \\ B & D & 0 \\ C & 0 & 0 \end{pmatrix} \begin{pmatrix} e_{\boldsymbol{q}} \\ e_u \\ e_\lambda \end{pmatrix} = \begin{pmatrix} \boldsymbol{\delta} \\ \epsilon \\ 0 \end{pmatrix}, \tag{3.5}$$

where

$$\boldsymbol{\delta} = \boldsymbol{\mathcal{P}}(\boldsymbol{c}(\Pi \boldsymbol{q} - \boldsymbol{q})) \quad \text{and} \quad \epsilon = \mathsf{P}(d(\mathsf{P}u - u)), \tag{3.6}$$

with $\boldsymbol{\mathcal{P}}$ denoting the L^2-orthogonal projection into $\boldsymbol{V}_{\mathcal{T}}$.

Proof. We begin by noting that the exact solution $(\boldsymbol{q}, u, u)$ satisfies

$$\begin{aligned}
&\int_\Omega \boldsymbol{c}\, \boldsymbol{q} \cdot \boldsymbol{v} \,\mathrm{d}x - \sum_{K \in \mathcal{T}} \int_K u \operatorname{div} \boldsymbol{v} \,\mathrm{d}x + \sum_{e \in \mathcal{E}_i} \int_e u\, [\boldsymbol{v}] \,\mathrm{d}s = -\int_{\partial\Omega} g\, [\boldsymbol{v}] \,\mathrm{d}s, \\
&\sum_{K \in \mathcal{T}} \int_K w \operatorname{div} \boldsymbol{q}_{\mathcal{T}} \,\mathrm{d}x + \int_\Omega d\, u\, w \,\mathrm{d}x = \int_\Omega f\, w \,\mathrm{d}x, \\
&\sum_{e \in \mathcal{E}_i} \int_e \mu\, [\boldsymbol{q}] \,\mathrm{d}s = 0,
\end{aligned}$$

for all $(\boldsymbol{v}, w, \mu) \in \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}} \times M_{\mathcal{T}}$. By using the commuting diagram property in the second equation and the fact that P is the L^2-projection into $M_{\mathcal{T}}$, we obtain

$$\begin{aligned}
&\int_\Omega \boldsymbol{c}\, \Pi \boldsymbol{q} \cdot \boldsymbol{v} \,\mathrm{d}x - \sum_{K \in \mathcal{T}} \int_K \mathsf{P}u \operatorname{div} \boldsymbol{v} \,\mathrm{d}x + \sum_{e \in \mathcal{E}_i} \int_e Pu\, [\boldsymbol{v}] \,\mathrm{d}s = -\int_{\partial\Omega} g\, [\boldsymbol{v}] \,\mathrm{d}s \\
&\qquad\qquad + \int_\Omega \boldsymbol{c}\, (\Pi \boldsymbol{q} - \boldsymbol{q}) \cdot \boldsymbol{v} \,\mathrm{d}x, \\
&\sum_{K \in \mathcal{T}} \int_K w \operatorname{div} \Pi \boldsymbol{q}_{\mathcal{T}} \,\mathrm{d}x + \int_\Omega d\, \mathsf{P}u\, w \,\mathrm{d}x = \int_\Omega f\, w \,\mathrm{d}x + \int_\Omega d\, (\mathsf{P}u - u)\, w \,\mathrm{d}x, \\
&\sum_{e \in \mathcal{E}_i} \int_e \mu\, [\Pi \boldsymbol{q}] \,\mathrm{d}s = 0,
\end{aligned}$$

which we can rewrite as

$$\begin{pmatrix} \boldsymbol{A} & -B^t & -C^t \\ B & D & 0 \\ C & 0 & 0 \end{pmatrix} \begin{pmatrix} \Pi \boldsymbol{q} \\ \mathsf{P}u \\ Pu \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha}_g + \boldsymbol{\delta} \\ \mathsf{P}f + \epsilon \\ 0 \end{pmatrix}, \tag{3.7}$$

where $\boldsymbol{\alpha}_g$ is as in (2.21). Subtracting (2.22) from the above equation, we have the result. □

Proof of Theorem 3.1. The identity (3.3) immediately follows from a direct application of Theorem 2.2 to the system of equations satisfied by the discrete error functions, as described in Lemma 3.2. The inequality (3.4) can now be obtained from (3.3) after a simple application of the Cauchy-Schwarz inequality. Thus Theorem 3.1 is proved. □

Remark 3.1. If we had chosen to use a projection that satisfies (3.1), but not (3.2), then in the third equation of (3.5), the right-hand side would be nonzero. By showing that this nonzero right-hand side is small and employing a more general form of Theorem 2.2 (see Theorem A.1), it is possible to get an error estimate similar to (3.4).

Error estimates for the the flux and primal variables follow from Theorem 3.1 and the simple continuity properties given in the next lemma. In the following, C denotes a generic constant independent of the element sizes.

Lemma 3.3 (Continuity properties of the local mappings)**.** *For all* $(\boldsymbol{\alpha}, \beta, \lambda) \in \boldsymbol{V}(K) \times W(K) \times M(K)$,

$$\begin{aligned} \|\mathfrak{Q}\boldsymbol{\alpha}\|_{\boldsymbol{c},K} &\le \mathrm{C}\|\boldsymbol{\alpha}\|_{\boldsymbol{c},K}, & \|\mathfrak{U}\boldsymbol{\alpha}\|_{0,K} &\le \mathrm{C}h_K\|\boldsymbol{\alpha}\|_{0,K}, \\ \|\mathsf{Q}\beta\|_{\boldsymbol{c},K} &\le \mathrm{C}h_K\|\beta\|_{0,K}, & \|\mathsf{U}\beta\|_{0,K} &\le \mathrm{C}h_K^2\|\beta\|_{0,K}, \\ \|\mathfrak{Q}\lambda\|_{\boldsymbol{c},K} &\le \mathrm{C}h_K^{-1}\|\lambda\|_K, & \|\mathfrak{U}\lambda\|_{0,K} &\le \mathrm{C}\|\lambda\|_K. \end{aligned}$$

These continuity properties are proved in Section 4.

Corollary 3.4. *The discrete solution components* $\boldsymbol{q}_{\mathcal{T}}$ *and* $u_{\mathcal{T}}$ *satisfy*

$$\begin{aligned} \|\boldsymbol{q} - \boldsymbol{q}_{\mathcal{T}}\|_{\boldsymbol{c}} &\le \mathrm{C}\big(\|\boldsymbol{q} - \Pi\boldsymbol{q}\|_{\boldsymbol{c}} + \|u - \mathsf{P}u\|_d\big), \\ \|u - u_{\mathcal{T}}\|_d &\le \mathrm{C}\big(\|\boldsymbol{q} - \Pi\boldsymbol{q}\|_{\boldsymbol{c}} + \|u - \mathsf{P}u\|_d\big). \end{aligned}$$

Proof. Thanks to Lemma 3.2, a direct application of the characterization of solutions, as given by Theorem 2.2, yields

$$e_{\boldsymbol{q}} = \mathfrak{Q}\boldsymbol{\delta} + \mathsf{Q}\epsilon - \mathfrak{Q}e_\lambda \quad \text{and} \quad e_u = \mathfrak{U}\boldsymbol{\delta} + \mathsf{U}\epsilon - \mathfrak{U}e_\lambda\,.$$

Hence

$$\begin{aligned} \|\boldsymbol{q}_{\mathcal{T}} - \Pi\boldsymbol{q}\|_{\boldsymbol{c}} &\le \|\mathfrak{Q}\boldsymbol{\delta}\|_{\boldsymbol{c}} + \|\mathsf{Q}\epsilon\|_{\boldsymbol{c}} + \|\mathfrak{Q}e_\lambda\|_{\boldsymbol{c}}, \\ \|u_{\mathcal{T}} - \mathsf{P}u\|_d &\le \|\mathfrak{U}\boldsymbol{\delta}\|_d + \|\mathsf{U}\epsilon\|_d + \|\mathfrak{U}e_\lambda\|_d. \end{aligned}$$

By the triangle inequality and the definition of the norm $|\!|\!|\cdot|\!|\!|_a$, this implies that

$$\begin{aligned} \|\boldsymbol{q}_{\mathcal{T}} - \boldsymbol{q}\|_{\boldsymbol{c}} &\le 2\left(\|\boldsymbol{q} - \Pi\boldsymbol{q}\|_{\boldsymbol{c}} + \|u - \mathsf{P}u\|_d\right) + \|\mathfrak{Q}\boldsymbol{\delta}\|_{\boldsymbol{c}} + \|\mathsf{Q}\epsilon\|_{\boldsymbol{c}}, \\ \|u_{\mathcal{T}} - u\|_d &\le 2\left(\|\boldsymbol{q} - \Pi\boldsymbol{q}\|_{\boldsymbol{c}} + \|u - \mathsf{P}u\|_d\right) + \|\mathfrak{U}\boldsymbol{\delta}\|_d + \|\mathsf{U}\epsilon\|_d. \end{aligned}$$

The corollary follows after using the continuity properties of the local mappings, Lemma 3.3, and the definition of $\boldsymbol{\delta}$ and ϵ, (3.6). □

It is now clear that rates of convergence can be obtained for the approximate solution by using approximation properties of the projections Π and P. In particular, we obtain the following result under minimal regularity assumptions. We use $\|\cdot\|_X$ and $|\cdot|_X$ to denote the standard norm and seminorm, respectively, in any Sobolev space X.

Corollary 3.5. *Suppose the exact solution* u *is in* $H^1(\Omega)$, $\boldsymbol{q} = -\boldsymbol{a}\,\mathbf{grad}\,u \in H^1(\Omega)^N$, *and the mesh* $\mathcal{T}$ *is a quasi-uniform mesh of mesh-size* h. *Then,*

$$|\!|\!|\lambda_{\mathcal{T}} - Pu|\!|\!|_a \le \mathrm{C}h(|\boldsymbol{q}|_{H^1(\Omega)} + |u|_{H^1(\Omega)}).$$

In the case of the standard uniform degree Raviart-Thomas method, as far as we know, previously stated estimates for $\lambda_{\mathcal{T}} - Pu$ required at least $H^3(\Omega)$-regularity of u; see, e.g., the estimate for the RT method in [2, Corollary 1.5] or the one for the BDM method in [4, Lemma 4.1]. Such stringent regularity requirements create difficulties when attempting duality arguments and in applications such as analysis of multigrid methods. The estimate of Corollary 3.5 is useful in these situations.

3.2. **The nature of the energy norm.** Next, we elucidate the nature of the energy norm $\|\!|\cdot|\!\|_a$. To avoid assumptions on meshes, we state a local result involving the following analogue of $a_{\mathcal{T}}(\cdot,\cdot)$ on one simplex $K \in \mathcal{T}$:

$$a_K(\eta,\mu) = \int_K \boldsymbol{c}\,\boldsymbol{\Omega}\eta \cdot \boldsymbol{\Omega}\mu \ \mathrm{d}x + \int_K d\ \mathcal{U}\eta\ \mathcal{U}\mu\ \mathrm{d}x.$$

We show that $a_K(\lambda,\lambda)$ admits two-sided bounds involving a transparent norm and seminorm defined by

$$\|\lambda\|_K^2 = h_K\ \|\lambda\|_{0,\partial K}^2 \quad \text{and} \quad \|\!|\lambda|\!\|_K^2 = \frac{1}{h_K}\|\lambda - m_K(\lambda)\|_{0,\partial K}^2,$$

where

$$m_K(\lambda) = \frac{1}{|\partial K|}\int_{\partial K} \lambda\ \mathrm{d}s.$$

As before, the estimates contain constants (e.g., C_1 and C_2 in Theorem 3.6) that in general depend on the ratio $r_K = h_K/\rho_K$ but are otherwise independent of h_K. Let $d_{\max}$ and $d_{\min}$ denote the essential supremum and infimum of $d(\boldsymbol{x})$ on K, respectively. Recall that $P_{\partial K}\lambda$ is as defined in (2.13).

Theorem 3.6. *Let $\lambda \in M_{\mathcal{T}}$. There are positive constants C_1 and C_2 independent of h_K such that for every $K \in \mathcal{T}$,*

$$\mathrm{C}_1\big(\|\!|P_{\partial K}\lambda|\!\|_K^2 + d_{\min}\|P_{\partial K}\lambda\|_K^2\big) \le\ a_K(\lambda,\lambda)\ \le \mathrm{C}_2\big(\|\!|P_{\partial K}\lambda|\!\|_K^2 + d_{\max}\|P_{\partial K}\lambda\|_K^2\big)$$

for all $\lambda \in M_{\mathcal{T}}$.

This result was obtained for the case $d = 0$ and for uniform polynomial degrees in [12, Theorem 2.2]. A detailed proof of this result for $d \ge 0$ is given in Section 6. Note that in the particular case of the uniform degree Raviart-Thomas method, we can remove $P_{\partial K}$ from the estimates of the theorem because in such a case, $P_{\partial K}\mu = \mu$ for all $\mu \in M_{\mathcal{T}}$.

Let us briefly discuss a consequence in the uniform degree case. Suppose $\mathcal{T}$ is a nondegenerate mesh. Define

$$\|\!|\lambda|\!\| = \Big(\sum_{K\in\mathcal{T}} \|\!|\lambda|\!\|_K^2\Big)^{1/2} \quad \text{and} \quad \|\lambda\| = \Big(\sum_{K\in\mathcal{T}} \|\lambda\|_K^2\Big)^{1/2}. \tag{3.8}$$

Since $d(\boldsymbol{x}) \ge 0$, one bound of the theorem implies that

$$\mathrm{C}_1\|\!|\lambda|\!\|^2 \le a_{\mathcal{T}}(\lambda,\lambda),$$

while the other implies that

$$a_{\mathcal{T}}(\lambda,\lambda) \le \mathrm{C}(\|\!|\lambda|\!\|^2 + \|\lambda\|^2) \le \mathrm{C}\|\!|\lambda|\!\|^2,$$

where the last inequality follows from a Poincaré -type inequality; see [12]. Combining the last two inequalities, we find that $\|\!|\lambda|\!\|_a$ is equivalent to $\|\!|\lambda|\!\|$ with mesh independent constants. One consequence of this is that the error estimate of Theorem 3.1 implies that we have control of the error in the more transparent $\|\!|\cdot|\!\|$-norm. Another consequence is in preconditioning the matrix equation defining the Lagrange multipliers of the hybridized mixed method: To precondition $a_{\mathcal{T}}(\cdot,\cdot)$, it suffices to precondition the norm $\|\!|\lambda|\!\|$ on $M_{\mathcal{T}}$, even when $d \ne 0$. One such preconditioner is already known [12].

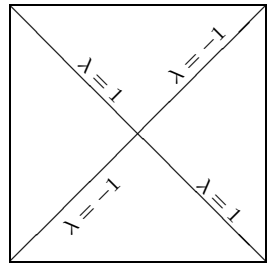

FIGURE 1. An example of a mesh and a multiplier $\lambda \in M_{\mathcal{T}}$ for which $|\lambda|_{\mathcal{U}} = 0$. (Here the domain is the unit square divided into four congruent elements, $d(\boldsymbol{x}) \equiv 0$, $\boldsymbol{c}(\boldsymbol{x})$ is the identity, and $k(K) = 0$ for all the four elements.)

3.3. **Error estimates in L^2-like norms.** The standard technique to obtain error estimates in weaker L^2-like norms from stronger energy norms is the Aubin-Nitsche duality argument. The results we now present are obtained by adapting such a duality argument to the mesh-dependent bilinear form $a_{\mathcal{T}}(\cdot,\cdot)$. First, we present error estimates in the seminorm

$$|\mu|_{\mathcal{U}} \equiv \| \mathcal{U}\mu \|_{0,\Omega}.$$

Then, we show how error estimates in this seminorm naturally yield superconvergence results for the primal variable. Error estimates in the L^2-like norm $\|\cdot\|$ defined by (3.8) also follow as a corollary. Note that $|\cdot|_{\mathcal{U}}$ is only a seminorm in general (see Figure 1). Our results are obtained using the first or both of the following assumptions.

Assumption 3.1.

(1) For every $f \in L^2(\Omega)$, the solution U of the boundary value problem

$$\begin{aligned} -\operatorname{div}(\boldsymbol{a}(\boldsymbol{x})\,\mathbf{grad}\,U) + d(\boldsymbol{x})U &= f \qquad \text{on } \Omega, \\ U &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{3.9}$$

satisfies the regularity property

$$\|U\|_{H^1(\Omega)} + |\boldsymbol{a}\,\mathbf{grad}\,U|_{H^1(\Omega)} \le \mathrm{C}\|f\|_{0,\Omega}.$$

(2) The function $d(\boldsymbol{x})$ is piecewise smooth in the sense that $\|d\|_{W^1_\infty(K)} \le \mathrm{C}$ for all $K \in \mathcal{T}$. Let $d_{1,\infty} = \max_{K\in\mathcal{T}} \|d\|_{W^1_\infty(K)}$.

(3) The mesh $\mathcal{T}$ is quasi-uniform of mesh size h.

Assumption 3.2. (1) The function $\boldsymbol{c}(\boldsymbol{x})$ is piecewise smooth in the sense that $\|\boldsymbol{c}\|_{W^1_\infty(K)} \le \mathrm{C}$ for all $K \in \mathcal{T}$. Let $c_{1,\infty} = \max_{K\in\mathcal{T}} \|\boldsymbol{c}\|_{W^1_\infty(K)}$.

(2) For all $K \in \mathcal{T}$, $k(K) \ge 1$.

We have the following error estimates.

Theorem 3.7. *Let $\boldsymbol{q}$ and u solve* (1.1)–(1.2). *Suppose Assumption* 3.1 *holds. Then,*

$$|\lambda_{\mathcal{T}} - Pu|_{\mathcal{U}} \le \mathrm{C}\,h\Big(\|\Pi\boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + \sigma\,\|\operatorname{div}(\Pi\boldsymbol{q} - \boldsymbol{q})\|_{0,\Omega} + d_{1,\infty}\|\mathsf{P}u - u\|_{0,\Omega}\Big),$$

where the factor σ equals one. If in addition Assumption 3.2 *holds, then $\sigma = 0$.*

A direct consequence of this theorem is the following result which shows that $\|u_{\mathcal{T}} - \mathsf{P}u\|_{0,\Omega}$ is one order higher than the best approximation error $\|u - \mathsf{P}u\|_{0,\Omega}$ under suitable regularity assumptions.

Corollary 3.8 (Superconvergence)**.** *Suppose Assumption* 3.1 *holds. Then*

$$\|u_{\mathcal{T}} - \mathsf{P}u\|_{0,\Omega} \le \mathrm{C}\, h\Big(\|\Pi \boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + \sigma \,\|\operatorname{div}(\Pi \boldsymbol{q} - \boldsymbol{q})\|_{0,\Omega} + d_{1,\infty}\|\mathsf{P}u - u\|_{0,\Omega}\Big),$$

where the factor σ equals one. If in addition Assumption 3.2 *holds, then $\sigma = 0$.*

As a consequence, we also obtain error estimates for the Lagrange multiplier in the L^2-like norm $\|\cdot\|$, as the following corollary shows. While similar results are known in the uniform degree case [2, 4], the estimate is new for the variable degree method.

Corollary 3.9. *The statements of Theorem* 3.7 *continue to hold if we replace $|\cdot|_{\mathcal{U}}$ with $\|\cdot\|$.*

All these results will be proven in Section 5.

4. The continuity properties of the local mappings

In this section, we prove the continuity properties of the local mappings given in Lemma 3.3. We begin by stating and proving an auxiliary result containing key properties of the operators $\boldsymbol{A}, B^t$ and C^t.

Lemma 4.1 (Bounds for $\boldsymbol{A}$, B^t and C^t)**.** *For all $(\boldsymbol{v}, w, \lambda) \in \boldsymbol{V}(K) \times W(K) \times M(K)$ and all $K \in \mathcal{T}$,*

$$\begin{aligned} c\|\boldsymbol{v}\|_{0,K} &\le \quad\quad \|\boldsymbol{A}\boldsymbol{v}\|_{0,K} \le \mathrm{C}\|\boldsymbol{v}\|_{0,K}, \\ c\|w\|_{0,K} &\le h_K\|B^t w\|_{0,K} \le \mathrm{C}\|w\|_{0,K}, \\ & \quad\quad h_K\|C^t\lambda\|_{0,K} \le \mathrm{C}\|\lambda\|_K. \end{aligned}$$

Proof. The estimates for $\boldsymbol{A}$ follow from its definition (2.20), in a straightforward way. Let us prove the estimates for B^t. From the definition of B, we see that $B^t w|_K$ is the only element of $W(K)$ such that $(\boldsymbol{r}, B^t w)_K = (w, \operatorname{div} \boldsymbol{r})_K$ for all $\boldsymbol{r} \in \boldsymbol{V}(K)$. Hence, the upper bound for the operator B^t follows from a simple inverse inequality. The lower bound is a direct consequence of the well-known surjectivity of B as a mapping from $\boldsymbol{V}(K)$ to $W(K)$. Indeed, because of the surjectivity, we have the following inf-sup condition on the reference element $\hat{K}$:

$$\|\hat{w}\|_{0,\hat{K}} \le \hat{\mathrm{C}} \sup_{\boldsymbol{r} \in \boldsymbol{V}(\hat{K})} \frac{(\hat{w}, \operatorname{div} \boldsymbol{r})_{\hat{K}}}{\|\boldsymbol{r}\|_{0,\hat{K}}}, \quad \text{for all } \hat{w} \in W(\hat{K}),$$

where $\boldsymbol{V}(\hat{K}) \times W(\hat{K})$ is the Raviart-Thomas space of index $k(K)$ on $\hat{K}$. This implies the following estimate after mapping back to K:

$$\mathrm{C}\|w\|_{0,K} \le h_K \sup_{\boldsymbol{r} \in \boldsymbol{V}(K)} \frac{(w, \operatorname{div} \boldsymbol{r})_K}{\|\boldsymbol{r}\|_{0,K}} = h_K\|B^t w\|_{0,K}, \quad \text{for all } w \in W(K).$$

This completes the proof of the lower bound for B.

It remains to prove the estimate for C^t. By definition, $C^t\lambda$ is the element in $\boldsymbol{V}_{\mathcal{T}}$ such that

$$\int_\Omega \boldsymbol{r} \cdot C^t\lambda \, dx = (C\boldsymbol{r}, \lambda)_{\mathcal{E}} = -\sum_{K \in \mathcal{T}} \int_{\partial K} \lambda\, \boldsymbol{r} \cdot \boldsymbol{n}_K \, \mathrm{d}s,$$

for all $\boldsymbol{r} \in \boldsymbol{V}_{\mathcal{T}}$. If we take $\boldsymbol{r} = C^t\lambda$ on K and $\boldsymbol{r} = 0$ elsewhere, we get

$$\| C^t\lambda \|_{0,K}^2 = -\int_{\partial K} \lambda\, C^t\lambda\cdot\boldsymbol{n}_K \,\mathrm{d}s \le C \| \lambda \|_{0,\partial K} \| C^t\lambda \|_{0,\partial K},$$

by the Cauchy-Schwarz inequality. Now, a simple inverse inequality gives us that

$$\| C^t\lambda \|_{0,K} \le C\, h_K^{-1/2} \| \lambda \|_{0,\partial K} = C\, h_K^{-1} \| \lambda \|_K,$$

by the definition of $\| \lambda \|_K$. This completes the proof. □

Proof of Lemma 3.3. Let us begin by proving the continuity estimates for the mapping $(\mathfrak{Q}\boldsymbol{\alpha}\,, \mathfrak{U}\boldsymbol{\alpha}\,)$. By definition (see (2.18) and (2.19)) we have

$$\begin{pmatrix} \boldsymbol{A} & -B^t \\ B & D \end{pmatrix} \begin{pmatrix} \mathfrak{Q}\boldsymbol{\alpha} \\ \mathfrak{U}\boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \mathcal{P}\boldsymbol{\alpha} \\ 0 \end{pmatrix},$$

where $\mathcal{P}$ denotes the L^2-orthogonal projection into $\boldsymbol{V}_{\mathcal{T}}$. It follows that

$$(\boldsymbol{A}\mathfrak{Q}\boldsymbol{\alpha}\,, \mathfrak{Q}\boldsymbol{\alpha}\,)_K + (D\,\mathfrak{U}\boldsymbol{\alpha}\,, \mathfrak{U}\boldsymbol{\alpha}\,)_K = (\boldsymbol{\alpha}, \mathfrak{Q}\boldsymbol{\alpha}\,)_K.$$

Since D is positive semidefinite, this implies that $\| \mathfrak{Q}\boldsymbol{\alpha} \|_{\boldsymbol{c},K}^2 \le C\|\boldsymbol{\alpha}\|_{0,K} \| \mathfrak{Q}\boldsymbol{\alpha} \|_{\boldsymbol{c},K}$. Thus $\| \mathfrak{Q}\boldsymbol{\alpha} \|_{\boldsymbol{c},K} \le C \| \boldsymbol{\alpha} \|_{0,K}$.

To obtain the continuity estimate for $\mathfrak{U}$, we proceed as follows:

$$\begin{aligned} \| \mathfrak{U}\boldsymbol{\alpha} \|_{0,K} &\le C\, h_K \| B^t\, \mathfrak{U}\boldsymbol{\alpha} \|_{0,K} && \text{by the invertibility of } B^t\text{, Lemma 4.1,} \\ &= C\, h_K \| \boldsymbol{A}\mathfrak{Q}\boldsymbol{\alpha} - \mathcal{P}\boldsymbol{\alpha} \|_{0,K} && \text{by the definition of } (\mathfrak{Q}\boldsymbol{\alpha}\,, \mathfrak{U}\boldsymbol{\alpha}\,), \\ &\le C\, h_K \| \boldsymbol{\alpha} \|_{0,K}. \end{aligned}$$

Next, let us prove the continuity properties of the mapping $(\mathsf{Q}\beta\,, \mathsf{U}\beta\,)$. By definition (see (2.16) and (2.17)) we have

$$\begin{pmatrix} \boldsymbol{A} & -B^t \\ B & D \end{pmatrix} \begin{pmatrix} \mathsf{Q}\beta \\ \mathsf{U}\beta \end{pmatrix} = \begin{pmatrix} 0 \\ \beta \end{pmatrix}.$$

It follows that

$$(\boldsymbol{A}\mathsf{Q}\beta\,, \mathsf{Q}\beta\,)_K + (D\,\mathsf{U}\beta\,, \mathsf{U}\beta\,)_K = (\beta, \mathsf{U}\beta\,)_K,$$

and consequently

$$\| \mathsf{Q}\beta \|_{\boldsymbol{c},K}^2 \le C \| \beta \|_{0,K} \| \mathsf{U}\beta \|_{0,K}. \tag{4.1}$$

To estimate $\| \mathsf{U}\beta \|_{0,K}$, we proceed as follows:

$$\begin{aligned} \| \mathsf{U}\beta \|_{0,K} &\le C\, h_K \| B^t\, \mathsf{U}\beta \|_{0,K} && \text{by the invertibility of } B^t\text{, Lemma 4.1,} \\ &= C\, h_K \| \boldsymbol{A}\mathsf{Q}\beta \|_{0,K} && \text{by the definition of } (\mathsf{Q}\beta\,, \mathsf{U}\beta\,), \\ &\le C\, h_K \| \beta \|_{0,K}^{1/2} \| \mathsf{U}\beta \|_{0,K}^{1/2} && \text{by (4.1).} \end{aligned}$$

As a consequence,

$$\| \mathsf{U}\beta \|_{0,K} \le C\, h_K^2 \| \beta \|_{0,K} \quad \text{and} \quad \| \mathsf{Q}\beta \|_{\boldsymbol{c},K} \le C\, h_K \| \beta \|_{0,K}.$$

It remains to obtain the estimates for the lifting $(\mathfrak{Q}\lambda\,, \mathfrak{U}\lambda\,)$. Observe that

$$(\mathfrak{Q}\lambda\,, \mathfrak{U}\lambda\,) = (\mathfrak{Q}C^{\,t}\lambda, \mathfrak{U}C^{\,t}\lambda),$$

by the definitions of $(\mathfrak{Q}\lambda\,, \mathfrak{U}\lambda\,)$ and $(\mathfrak{Q}\cdot\,, \mathfrak{U}\cdot\,)$. Hence the required estimates follow from the estimates of $(\mathfrak{Q}\cdot\,, \mathfrak{U}\cdot\,)$ and the continuity of C^t as given in Lemma 4.1. □

5. Proofs of the error estimates in L^2-like norms

We prove Theorem 3.7 in two steps.

5.1. **Proof of Theorem 3.7: Step 1.** The proof of the error estimates in L^2-like norms given in Theorem 3.7 is based on the continuity properties of the local mappings and on a suitable version of the Aubin-Nitsche lemma. We describe the latter next. Consider the dual problem in mixed form

$$\begin{aligned} \boldsymbol{c}\boldsymbol{\psi} + \mathbf{grad}\,\phi &= 0 && \text{on } \Omega,\\ \operatorname{div}\boldsymbol{\psi} + d\phi &= \mathcal{U}e_\lambda && \text{on } \Omega,\\ \phi &= 0 && \text{on } \partial\Omega, \end{aligned}$$

where, as before, $e_\lambda = \lambda_{\mathcal{T}} - Pu$. Let $(\boldsymbol{\psi}_{\mathcal{T}}, \phi_{\mathcal{T}}, \varepsilon_{\mathcal{T}}) \in \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}} \times M_{\mathcal{T}}$ be the hybridized mixed approximation of $(\boldsymbol{\psi}, \phi, \phi|_{\mathcal{E}_i})$. With this notation, we have the following result.

Lemma 5.1 (The Aubin-Nitsche duality argument)**.** *We have*

$$|e_\lambda|_{\mathcal{U}}^2 = \int_\Omega \boldsymbol{c}\,(\Pi\boldsymbol{q} - \boldsymbol{q}) \cdot \boldsymbol{\mathcal{Q}}\varepsilon_{\mathcal{T}}\ \mathrm{d}x + \int_\Omega d(u - \mathsf{P}u)\,\mathcal{U}\varepsilon_{\mathcal{T}}\ \mathrm{d}x.$$

Moreover, the right-hand side is the sum of the terms

$$\begin{aligned} I_1 &= \int_\Omega \boldsymbol{c}\,(\Pi\boldsymbol{q} - \boldsymbol{q}) \cdot \boldsymbol{\psi}\,\mathrm{d}x + \int_\Omega d\,(u - \mathsf{P}u)\phi\,\mathrm{d}x,\\ I_2 &= \int_\Omega \boldsymbol{c}\,(\Pi\boldsymbol{q} - \boldsymbol{q}) \cdot (\boldsymbol{\psi}_{\mathcal{T}} - \boldsymbol{\psi})\,\mathrm{d}x + \int_\Omega d\,(u - \mathsf{P}u)(\phi_{\mathcal{T}} - \phi)\,\mathrm{d}x,\\ I_3 &= -\int_\Omega \boldsymbol{c}\,(\Pi\boldsymbol{q} - \boldsymbol{q}) \cdot \mathbf{Q}\,\mathcal{U}e_\lambda\ \mathrm{d}x - \int_\Omega d\,(u - \mathsf{P}u)\,\mathsf{U}\,\mathcal{U}e_\lambda\ \mathrm{d}x. \end{aligned}$$

Proof. A direct application of the characterization of the approximate solution, Theorem 2.2, to the approximate solution of the dual problem, $(\boldsymbol{\psi}_{\mathcal{T}}, \phi_{\mathcal{T}}, \varepsilon_{\mathcal{T}})$, gives

$$a_{\mathcal{T}}(\mu, \varepsilon_{\mathcal{T}}) = \int_\Omega \mathcal{U}e_\lambda\ \mathcal{U}\mu\ \mathrm{d}x \qquad \text{for all } \mu \in M_{\mathcal{T}}.$$

Hence,

$$|e_\lambda|_{\mathcal{U}}^2 = a_{\mathcal{T}}(e_\lambda, \varepsilon_{\mathcal{T}}) = \int_\Omega \boldsymbol{c}\,(\Pi\boldsymbol{q} - \boldsymbol{q}) \cdot \boldsymbol{\mathcal{Q}}\varepsilon_{\mathcal{T}}\ \mathrm{d}x + \int_\Omega d(u - \mathsf{P}u)\,\mathcal{U}\varepsilon_{\mathcal{T}}\ \mathrm{d}x,$$

by Theorem 3.1. Now, using Theorem 2.2 once more, we find that

$$\begin{aligned} \boldsymbol{\mathcal{Q}}\varepsilon_{\mathcal{T}} &= \boldsymbol{\psi}_{\mathcal{T}} - \mathbf{Q}\,\mathcal{U}e_\lambda = \boldsymbol{\psi} + (\boldsymbol{\psi}_{\mathcal{T}} - \boldsymbol{\psi}) - \mathbf{Q}\,\mathcal{U}e_\lambda\ ,\\ \mathcal{U}\varepsilon_{\mathcal{T}} &= \phi_{\mathcal{T}} - \mathsf{U}\,\mathcal{U}e_\lambda = \phi + (\phi_{\mathcal{T}} - \phi) - \mathsf{U}\,\mathcal{U}e_\lambda\ , \end{aligned}$$

and the result follows. □

5.2. **Proof of Theorem 3.7: Step 2.** It only remains to estimate the terms I_i, $i = 1, 2, 3$. The terms I_2 and I_3 can be immediately seen to be small:

$$\begin{aligned} I_2 &\le \|\Pi\boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}}\|\boldsymbol{\psi} - \boldsymbol{\psi}_{\mathcal{T}}\|_{\boldsymbol{c}} + \|u - \mathsf{P}u\|_d\|\phi - \phi_{\mathcal{T}}\|_d\\ &\le \mathrm{C}h(|\boldsymbol{\psi}|_{H^1(\Omega)} + |\phi|_{H^1(\Omega)})(\|\Pi\boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + \|u - \mathsf{P}u\|_d),\\ &\le \mathrm{C}h|e_\lambda|_{\mathcal{U}}\,(\|\Pi\boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + \|u - \mathsf{P}u\|_d), \end{aligned} \tag{5.1}$$

by Assumption 3.1 and

$$\begin{aligned} I_3 &\le \|\Pi \boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} \|\mathbf{Q}\, \mathcal{U} e_\lambda \,\|_{\boldsymbol{c}} + \|u - \mathsf{P}u\|_d \|\, \mathsf{U}\, \mathcal{U} e_\lambda \,\|_d \\ &\le Ch|e_\lambda|_{\mathcal{U}} \left(\|\Pi \boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + h\|u - \mathsf{P}u\|_d\right), \end{aligned}$$

by the continuity properties of local mappings, Lemma 3.3.

To estimate I_1, we begin by estimating the last term in its definition:

$$\begin{aligned} \int_\Omega d\,(u - \mathsf{P}u)\phi \,\mathrm{d}x &= \int_\Omega (u - \mathsf{P}u)(d\phi - \mathsf{P}(d\phi))\,\mathrm{d}x \le Ch\|u - \mathsf{P}u\|_{0,\Omega} d_{1,\infty} \|\phi\|_{H^1(\Omega)} \\ &\le Chd_{1,\infty}\|u - \mathsf{P}u\|_{0,\Omega}|e_\lambda|_{\mathcal{U}}\,, \end{aligned} \tag{5.2}$$

by Assumption 3.1.

There are two ways to estimate the other term in the definition of I_1. First,

$$\begin{aligned} \int_\Omega \boldsymbol{c}\,(\Pi \boldsymbol{q} - \boldsymbol{q}) \cdot \boldsymbol{\psi}\,\mathrm{d}x &= -\int_\Omega \phi \operatorname{div}(\boldsymbol{q} - \Pi \boldsymbol{q})\,\mathrm{d}x \\ &= -\int_\Omega \phi (I - \mathsf{P}) \operatorname{div} \boldsymbol{q}\,\mathrm{d}x \\ &= -\int_\Omega (I - \mathsf{P})\phi \operatorname{div}(\boldsymbol{q} - \Pi \boldsymbol{q})\,\mathrm{d}x \\ &\le Ch|\phi|_{H^1(\Omega)} \|\operatorname{div}(\boldsymbol{q} - \Pi \boldsymbol{q})\|_{0,\Omega} \\ &\le Ch|e_\lambda|_{\mathcal{U}}\, \|\operatorname{div}(\boldsymbol{q} - \Pi \boldsymbol{q})\|_{0,\Omega}, \end{aligned}$$

by Assumption 3.1. Now, if we assume that Assumption 3.2 *also* holds, then $k(K) \ge 1$, and the same term can be estimated alternately:

$$\begin{aligned} (\boldsymbol{c}\,(\Pi \boldsymbol{q} - \boldsymbol{q}), \boldsymbol{\psi})_K =& (\boldsymbol{c}(\boldsymbol{q} - \Pi \boldsymbol{q}), \boldsymbol{a}\,\mathbf{grad}\,\phi)_K \\ =& ((\boldsymbol{c} - \overline{\boldsymbol{c}})(\boldsymbol{q} - \Pi \boldsymbol{q}), \boldsymbol{a}\,\mathbf{grad}\,\phi)_K \\ &+ (\overline{\boldsymbol{c}}(\boldsymbol{q} - \Pi \boldsymbol{q}), \boldsymbol{a}\,\mathbf{grad}\,\phi - \boldsymbol{\kappa})_K \\ &+ (\overline{\boldsymbol{c}}(\boldsymbol{q} - \Pi \boldsymbol{q}), \boldsymbol{\kappa})_K, \end{aligned}$$

where $\overline{\boldsymbol{c}}$ and $\boldsymbol{\kappa}$ denote the averages of $\boldsymbol{c}(\boldsymbol{x})$ and $\boldsymbol{a}(\boldsymbol{x})\,\mathbf{grad}\,\phi$ on K, respectively. Because $k(K) \ge 1$, by the definition of Π, $(\boldsymbol{q} - \Pi \boldsymbol{q}, \overline{\boldsymbol{c}}^t \boldsymbol{\kappa})_K = 0$, so by Friedrichs inequality,

$$\begin{aligned} \int_K \boldsymbol{c}\,(\Pi \boldsymbol{q} - \boldsymbol{q}) \cdot \boldsymbol{\psi}\,\mathrm{d}x &\le Ch(c_{1,\infty}\|\boldsymbol{a}\,\mathbf{grad}\,\phi\|_{0,\Omega} + |\boldsymbol{a}\,\mathbf{grad}\,\phi|_{H^1(\Omega)})\|\boldsymbol{q} - \Pi \boldsymbol{q}\|_{0,\Omega} \\ &\le Ch|e_\lambda|_{\mathcal{U}}\, \|\boldsymbol{q} - \Pi \boldsymbol{q}\|_{0,\Omega}, \end{aligned}$$

by Assumption 3.1.

Combining the estimates for I_1, I_2, and I_3, we get the estimate of the theorem. This completes the proof.

5.3. **Proof of the superconvergence result of Corollary 3.8.** To prove this result, we begin by noting that

$$\|e_u\|_{0,\Omega} \le Ch\|\, B^t e_u\|_{0,\Omega},$$

by the invertibility of B^t, Lemma 4.1. Now, since, by the error equation, Lemma 3.2,

$$\begin{aligned} B^t e_u &= \boldsymbol{\delta} - \boldsymbol{A} e_{\boldsymbol{q}} - C^t e_\lambda \\ &= \boldsymbol{\delta} - \boldsymbol{A} e_{\boldsymbol{q}} - \boldsymbol{A}\boldsymbol{\Omega} e_\lambda + B^t\, \mathcal{U} e_\lambda\,, \end{aligned}$$

by the definition of $(\boldsymbol{\Omega}\cdot\,,\,\mathcal{U}\cdot\,)$, (2.14)-(2.15), we obtain that

$$\begin{aligned}\|e_u\|_{0,\Omega} &\le \mathrm{C}h\big(\|\boldsymbol{\delta}\|_{0,\Omega} + \|\boldsymbol{A}e_{\boldsymbol{q}}\|_{0,\Omega} + \|\boldsymbol{A}\boldsymbol{\Omega}e_\lambda\,\|_{0,\Omega} + \|B^t\,\mathcal{U}e_\lambda\,\|_{0,\Omega}\big)\\ &\le \mathrm{C}\big(\,h\,\|\boldsymbol{\delta}\|_{0,\Omega} + h\,\|e_{\boldsymbol{q}}\|_{\boldsymbol{c}} + h\,|\!|\!|e_\lambda|\!|\!|_a + \|\,\mathcal{U}e_\lambda\,\|_{0,\Omega}\big),\end{aligned}$$

by the continuity properties of the operators A and B^t, Lemma 4.1, and the definition of the norm $|\!|\!|\cdot|\!|\!|_a$. The estimate now easily follows by using the definition of $\boldsymbol{\delta}$, (3.6), Corollary 3.4, and Theorems 3.1 and 3.7. This completes the proof.

5.4. **Proof of Corollary 3.9.** Although it is possible to prove Corollary 3.9 by means of a duality argument similar to the one used to prove Theorem 3.7, for brevity, we prefer to use Corollary 3.8 together with the approach used in a similar proof in [2]: Since e_λ satisfies

$$\int_{\partial K} e_\lambda\,\boldsymbol{v}\cdot\boldsymbol{n}\,\mathrm{d}s = \int_K \boldsymbol{c}\,(\boldsymbol{q}_{\mathcal{T}} - \boldsymbol{q})\cdot\boldsymbol{v}\,\mathrm{d}x + \int_K e_u\,\mathrm{div}\,\boldsymbol{v}\,\mathrm{d}x,$$

for all $\boldsymbol{v}\in\boldsymbol{V}(K)$, by choosing $\boldsymbol{v}$ such that $\boldsymbol{v}\cdot\boldsymbol{n} = e_\lambda$ and $\|\boldsymbol{v}\|_{0,K}\le \mathrm{C}\|e_\lambda\|_K$, and using an inverse inequality, we have

$$\|e_\lambda\|_K^2 \le (\mathrm{C}h\|\boldsymbol{q}-\boldsymbol{q}_{\mathcal{T}}\|_{0,K} + \|e_u\|_{0,K})\;\|e_\lambda\|_K.$$

Hence the result follows after an application of Corollaries 3.4 and 3.8.

6. Proof of the norm equivalence result

To prove the norm equivalence result of Theorem 3.6, we begin by proving two lemmas. Recall that, by the definition of the operators $\boldsymbol{A}$ and C, $\mathcal{G}\lambda$ is the only element of $\boldsymbol{V}_{\mathcal{T}}$ that satisfies

$$\int_K \boldsymbol{c}\,(\mathcal{G}\lambda\,)\cdot\boldsymbol{q}\,\mathrm{d}x = -\int_{\partial K}\lambda(\boldsymbol{q}\cdot\boldsymbol{n}_K)\,\mathrm{d}s, \qquad \text{for all } \boldsymbol{q}\in\boldsymbol{V}(K) \text{ and } K\in\mathcal{T}. \tag{6.1}$$

As a consequence, we have the following relation between the operator $\mathcal{G}$ and the local liftings $(\boldsymbol{\Omega}\lambda\,,\,\mathcal{U}\lambda\,)$:

$$\begin{pmatrix}\boldsymbol{A} & -B^t\\ B & D\end{pmatrix}\begin{pmatrix}\boldsymbol{\Omega}\lambda\\ \mathcal{U}\lambda\end{pmatrix} = \begin{pmatrix}\boldsymbol{A}\mathcal{G}\lambda\\ 0\end{pmatrix}. \tag{6.2}$$

We also need to introduce an orthogonal projection into the following subspace of divergence-free functions on an element K:

$$\boldsymbol{V}^0(K) \equiv \{\boldsymbol{p}\in\boldsymbol{V}(K) : \mathrm{div}\,\boldsymbol{p} = 0\}.$$

Define an orthogonal projection $\boldsymbol{J}$ into $\boldsymbol{V}^0(K)$ as follows: For any $K\in\mathcal{T}$ and $\boldsymbol{r}\in\boldsymbol{V}_{\mathcal{T}}$, the restriction $\boldsymbol{J}\boldsymbol{r}|_K$ is defined to be the only element in $\boldsymbol{V}^0(K)$ satisfying

$$\int_K \boldsymbol{c}\,(\boldsymbol{J}\boldsymbol{r})\cdot\boldsymbol{p}\,\mathrm{d}x = \int_K \boldsymbol{c}\,\boldsymbol{r}\cdot\boldsymbol{p}\,\mathrm{d}x \qquad \text{for all } \boldsymbol{p}\in\boldsymbol{V}^0(K). \tag{6.3}$$

Note the following properties of $\mathcal{G}$ and $\boldsymbol{J}$.

Lemma 6.1. *For all $(\boldsymbol{r},\lambda)\in\boldsymbol{V}_{\mathcal{T}}\times M_{\mathcal{T}}$ and all $K\in\mathcal{T}$, we have*

$$\begin{aligned}\|P_{\partial K}\lambda\|_K &\le \mathrm{C}h_K\|\mathcal{G}\lambda\,\|_{\boldsymbol{c},K},\\ \|(\boldsymbol{I}-\boldsymbol{J})\boldsymbol{r}\|_{0,K} &\le \mathrm{C}h_K\|B\boldsymbol{r}\|_{0,K}.\end{aligned}$$

Proof. To prove the first inequality, we choose a $\boldsymbol{q}$ in (6.1) that satisfies $\boldsymbol{q}\cdot\boldsymbol{n}_k = P_{\partial K}\lambda$ and $\|\boldsymbol{q}\|_{0,K} \le \mathrm{C}h_K^{1/2}\|P_{\partial K}\lambda\|_{0,\partial K}$. Such a $\boldsymbol{q}$ can easily be seen to exist in $\boldsymbol{V}(K)$. Then,

$$\begin{aligned}\int_{\partial K}\lambda(\boldsymbol{q}\cdot\boldsymbol{n}_K)\,\mathrm{d}s = \int_{\partial K}|P_{\partial K}\lambda|^2\,\mathrm{d}s &= -\int_K \boldsymbol{c}\,(\mathcal{G}\lambda\,)\cdot\boldsymbol{q}\,\mathrm{d}x\\ &\le \|\mathcal{G}\lambda\,\|_{\boldsymbol{c},K}\|\boldsymbol{q}\|_{\boldsymbol{c},K} \le \mathrm{C}h_K^{1/2}\|\mathcal{G}\lambda\,\|_{\boldsymbol{c},K}\|P_{\partial K}\lambda\|_{0,\partial K},\end{aligned}$$

from which the bound follows.

To prove the second inequality, we note that if $B\boldsymbol{r}=0$, then $(\boldsymbol{I}-\boldsymbol{J})\boldsymbol{r} = \boldsymbol{r}-\boldsymbol{J}\boldsymbol{r} = 0$, by definition of B and $\boldsymbol{J}$. The result now follows by a simple scaling argument. This completes the proof. □

The proof of the norm equivalence result proceeds by splitting the liftings of λ to the flux space into a divergence free part and a remainder. The following lemma collects some results pertaining to both the parts.

Lemma 6.2. *For any $\lambda \in M_{\mathcal{T}}$ and $K \in \mathcal{T}$ we have*

$$\boldsymbol{J}\boldsymbol{\Omega}\lambda = \boldsymbol{J}\mathcal{G}\lambda\,, \tag{6.4}$$

$$c\,|\!|\!|P_{\partial K}\lambda|\!|\!|_K \le \|\boldsymbol{J}\boldsymbol{\Omega}\lambda\,\|_{\boldsymbol{c},K} \le \mathrm{C}|\!|\!|P_{\partial K}\lambda|\!|\!|_K, \tag{6.5}$$

$$\|\,(\boldsymbol{I}-\boldsymbol{J})\boldsymbol{\Omega}\lambda\,\|_{\boldsymbol{c},K} \le \mathrm{C}h_K d_{\max}^{1/2}\,\|\,\mathcal{U}\lambda\,\|_{0,K}, \tag{6.6}$$

$$d_{\min}\mathrm{C}\,h_k^2\|(\boldsymbol{I}-\boldsymbol{J})(\boldsymbol{\Omega}\lambda\,-\mathcal{G}\lambda\,)\|_{\boldsymbol{c},K}^2 \le \|\,\mathcal{U}\lambda\,\|_{d,K}^2. \tag{6.7}$$

Proof. To prove (6.4), observe that by the definitions of $\boldsymbol{J}$ and $\boldsymbol{\Omega}$,

$$\int_K \boldsymbol{c}\,(\mathcal{G}\lambda\,)\cdot\boldsymbol{r}\,\mathrm{d}x = \int_K \boldsymbol{c}\,(\boldsymbol{\Omega}\lambda\,)\cdot\boldsymbol{r}\,\mathrm{d}x = -\int_{\partial K}\lambda(\boldsymbol{r}\cdot\boldsymbol{n}_K)\,\mathrm{d}s, \quad \text{for all } \boldsymbol{r}\in\boldsymbol{V}^0(K).$$

Hence $\boldsymbol{J}\boldsymbol{\Omega}\lambda = \boldsymbol{J}\mathcal{G}\lambda$.

To prove (6.5), note that because of (6.4), $\boldsymbol{J}\boldsymbol{\Omega}\lambda$ is nothing but the $\boldsymbol{\Omega}$-lifting of λ when $d\equiv 0$, so $a_K(\lambda,\lambda) = \|\,\boldsymbol{J}\boldsymbol{\Omega}\lambda\,\|_{\boldsymbol{c},K}^2$. By a slight modification of the norm equivalence established for the $d\equiv 0$ (and uniform degree) case in [12, Lemma 2.1 and Theorem 2.2] the inequalities of (6.5) can be established.

To prove (6.6),

$$\begin{aligned}\|(\boldsymbol{I}-\boldsymbol{J})\boldsymbol{\Omega}\lambda\,\|_{\boldsymbol{c},K} &\le \mathrm{C}\,\|(\boldsymbol{I}-\boldsymbol{J})\boldsymbol{\Omega}\lambda\,\|_{0,K} && \text{by (6.4)},\\ &\le \mathrm{C}\,h_K\,\|B\boldsymbol{\Omega}\lambda\,\|_{0,K} && \text{by Lemma 6.1},\\ &= \mathrm{C}\,h_K\,\|D\,\mathcal{U}\lambda\,\|_{0,K} && \text{by the definition of the liftings},\\ &\le \mathrm{C}\,h_K\,d_{\max}^{1/2}\,\|\,\mathcal{U}\lambda\,\|_{0,K}.\end{aligned}$$

To prove (6.7), first note that by Lemma 4.1,

$$\|\,\mathcal{U}\lambda\,\|_{d,K}^2 \ge d_{\min}\|\,\mathcal{U}\lambda\,\|_{0,K}^2 \ge d_{\min}\mathrm{C}\,h_k^2\,\|B^t\,\mathcal{U}\lambda\,\|_{0,K}^2.$$

By (6.2), we also have

$$B^t\,\mathcal{U}\lambda = \boldsymbol{A}(\boldsymbol{\Omega}\lambda\,-\mathcal{G}\lambda\,) = \boldsymbol{A}(\boldsymbol{I}-\boldsymbol{J})(\boldsymbol{\Omega}\lambda\,-\mathcal{G}\lambda\,),$$

where we have also used (6.4). This completes the proof. □

With the above lemmas, we can now prove the norm equivalence result of Theorem 3.6.

Proof of Theorem 3.6. We obtain the upper bound as follows:

$$\begin{aligned} a_K(\lambda,\lambda) &= \|\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|\,\mathcal{U}\lambda\|^2_{d,K} \\ &= \|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|(\boldsymbol{I}-\boldsymbol{J})\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|\,\mathcal{U}\lambda\|^2_{d,K} \quad \text{by orthogonality of } \boldsymbol{J}, \\ &\le \mathrm{C}\,|\!|\!|P_{\partial K}\lambda|\!|\!|^2_K + \mathrm{C}\,d_{\max}\,\|\,\mathcal{U}\lambda\|^2_{0,K} \quad \text{by Lemma 6.2}, \\ &\le \mathrm{C}\left(|\!|\!|P_{\partial K}\lambda|\!|\!|^2_K + d_{\max}\,\|P_{\partial K}\lambda\|^2_K\right), \end{aligned}$$

where the last inequality was obtained by noting that $\mathcal{U}\lambda|_K = \mathcal{U}P_{\partial K}\lambda\,|_K$ and using the continuity property of $\mathcal{U}$ given by Lemma 3.3.

To obtain the lower bound of the theorem, we proceed as follows:

$$\begin{aligned} a_K(\lambda,\lambda) &= \|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|(\boldsymbol{I}-\boldsymbol{J})\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|\,\mathcal{U}\lambda\|^2_{d,K} \\ &\ge \|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|(\boldsymbol{I}-\boldsymbol{J})\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \nu\,\|(\boldsymbol{I}-\boldsymbol{J})(\boldsymbol{\mathfrak{Q}}\lambda - \mathfrak{G}\lambda)\|^2_{\boldsymbol{c},K}, \end{aligned}$$

where $\nu \equiv d_{\min}\mathrm{C}\,h_K^2$ by (6.7) of Lemma 6.2. Note that we can always assume that $\nu \le 1$ for all elements K. Hence

$$\begin{aligned} a_K(\lambda,\lambda) &\ge \|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \nu\left(\|(\boldsymbol{I}-\boldsymbol{J})\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|(\boldsymbol{I}-\boldsymbol{J})(\boldsymbol{\mathfrak{Q}}\lambda - \mathfrak{G}\lambda)\|^2_{\boldsymbol{c},K}\right) \\ &\ge \|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \frac{\nu}{2}\,\|(\boldsymbol{I}-\boldsymbol{J})\mathfrak{G}\lambda\|^2_{\boldsymbol{c},K} \quad \text{by the triangle inequality}, \\ &= (1-\frac{\nu}{2})\|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \frac{\nu}{2}\left(\|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|(\boldsymbol{I}-\boldsymbol{J})\mathfrak{G}\lambda\|^2_{\boldsymbol{c},K}\right) \\ &\ge \frac{1}{2}\|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \frac{\nu}{2}\left(\|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|(\boldsymbol{I}-\boldsymbol{J})\mathfrak{G}\lambda\|^2_{\boldsymbol{c},K}\right), \end{aligned}$$

since $\nu \le 1$. Now, since by Lemma 6.2, $\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda = \boldsymbol{J}\mathfrak{G}\lambda$,

$$\|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \|(\boldsymbol{I}-\boldsymbol{J})\mathfrak{G}\lambda\|^2_{\boldsymbol{c},K} = \|\boldsymbol{J}\mathfrak{G}\lambda\|^2_{\boldsymbol{c},K} + \|(\boldsymbol{I}-\boldsymbol{J})\mathfrak{G}\lambda\|^2_{\boldsymbol{c},K} = \|\mathfrak{G}\lambda\|^2_{\boldsymbol{c},K},$$

by the orthogonality property of $\boldsymbol{J}$. Consequently,

$$a_K(\lambda,\lambda) \ge \frac{1}{2}\|\boldsymbol{J}\boldsymbol{\mathfrak{Q}}\lambda\|^2_{\boldsymbol{c},K} + \frac{\nu}{2}\|\mathfrak{G}\lambda\|^2_{\boldsymbol{c},K} \ge \mathrm{C}\left(|\!|\!|P_{\partial K}\lambda|\!|\!|^2_K + d_{\min}\|P_{\partial K}\lambda\|^2_K\right),$$

by Lemmas 6.2 and 6.1. This completes the proof of Theorem 3.6. □

7. Extension to the variable degree BDM method

We now briefly describe how the analysis extends to the variable degree BDM method [4, 8]. The method (say in three space dimensions) is obtained by setting

$$\boldsymbol{V}(K) = (P^{k(K)}(K))^3, \quad W(K) = P^{k(K)-1}(K), \quad M(e) = \mathcal{M}^{k(e)}(e),$$

in the formulation in Section 2. That the method is well defined can be seen by the same type of arguments as in Proposition 2.1. Moreover, it is easy to see that the entire error analysis goes through and analogues of Theorems 2.2, 3.1, and 3.7 hold. In particular, we have the following analogue of (3.3):

(7.1)

$$a^{\mathrm{BDM}}_{\mathcal{T}}(\lambda^{\mathrm{BDM}}_{\mathcal{T}} - Pu, \mu) = \int_\Omega \boldsymbol{c}\,(\Pi^{\mathrm{BDM}}\boldsymbol{q} - \boldsymbol{q})\cdot\boldsymbol{\mathfrak{Q}}\mu\ \mathrm{d}x + \int_\Omega d(u - \mathsf{P}^{\mathrm{BDM}}u)\,\mathcal{U}\mu\ \mathrm{d}x.$$

Note that when previously introduced notation is superscripted by "BDM", it is to be understood as defined as before after replacing RT spaces by the BDM spaces.

Since $\mathsf{P}^{\mathrm{BDM}}$ projects onto a space that is one polynomial degree less on every mesh element than the corresponding RT space, it may appear at first sight that the last term in (7.1) is only $O(h^k)$ (when $k(K) = k$ and $d(\boldsymbol{x}) > 0$), one order less than the corresponding RT estimate. The following corollary shows otherwise. For simplicity we only give results under both Assumptions 3.1 and 3.2, although similar estimates can be proved when only the former holds.

Corollary 7.1. *Suppose Assumptions* 3.1 *and* 3.2 *hold. Then*

$$\|\!|\lambda_{\mathcal{T}}^{\mathrm{BDM}} - Pu|\!\|_a \le \mathrm{C}\Big(\|\Pi^{\mathrm{BDM}}\boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + d_{1,\infty}h\|u - \mathsf{P}^{\mathrm{BDM}}u\|_{0,\Omega}\Big). \tag{7.2}$$

If in addition the solution U of (3.9) *satisfies $|U|_{H^2(\Omega)} \le \mathrm{C}\|f\|_{0,\Omega}$ and $d_{2,\infty} \equiv \max_{K\in\mathcal{T}} \|d\|_{W^2_\infty(K)} < \infty$, then*

$$\|u_{\mathcal{T}}^{\mathrm{BDM}} - \mathsf{P}^{\mathrm{BDM}}u\|_{0,\Omega} \le \mathrm{C}\,h\Big(\|\Pi^{\mathrm{BDM}}\boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + d_{2,\infty}h\|\mathsf{P}^{\mathrm{BDM}}u - u\|_{0,\Omega}\Big), \tag{7.3}$$

$$\|\lambda_{\mathcal{T}}^{\mathrm{BDM}} - Pu\| \le \mathrm{C}\,h\Big(\|\Pi^{\mathrm{BDM}}\boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + d_{2,\infty}h\|\mathsf{P}^{\mathrm{BDM}}u - u\|_{0,\Omega}\Big). \tag{7.4}$$

Proof. Let $\bar{d}$ denote the piecewise constant function which on every $K \in \mathcal{T}$ equals the average of $d(\boldsymbol{x})$ on K. By (7.1)

$$\begin{aligned}
\|\!|e_\lambda^{\mathrm{BDM}}|\!\|_a^2 &\le \|\boldsymbol{q} - \Pi^{\mathrm{BDM}}\boldsymbol{q}\|_{\boldsymbol{c}}\|\!|e_\lambda^{\mathrm{BDM}}|\!\|_a + \int_\Omega (d - \bar{d})(u - \mathsf{P}^{\mathrm{BDM}}u)\,\mathcal{U}e_\lambda^{\mathrm{BDM}}\,\mathrm{d}x \\
&\le \|\boldsymbol{q} - \Pi^{\mathrm{BDM}}\boldsymbol{q}\|_{\boldsymbol{c}}\|\!|e_\lambda^{\mathrm{BDM}}|\!\|_a + \mathrm{C}hd_{1,\infty}\|u - \mathsf{P}^{\mathrm{BDM}}u\|_{0,\Omega}|e_\lambda^{\mathrm{BDM}}|_{\mathcal{U}} \\
&\le \|\boldsymbol{q} - \Pi^{\mathrm{BDM}}\boldsymbol{q}\|_{\boldsymbol{c}}\|\!|e_\lambda^{\mathrm{BDM}}|\!\|_a \\
&\qquad + \mathrm{C}h^2 d_{1,\infty}\|u - \mathsf{P}^{\mathrm{BDM}}u\|_{0,\Omega}(\|\boldsymbol{q} - \Pi^{\mathrm{BDM}}\boldsymbol{p}\|_{\boldsymbol{c}} + d_{1,\infty}\|u - \mathsf{P}^{\mathrm{BDM}}u\|_{0,\Omega}),
\end{aligned}$$

by the BDM analogue of Theorem 3.7. Hence (7.2) follows after applying the arithmetic geometric mean inequality.

We now prove that

$$|e_\lambda^{\mathrm{BDM}}|_{\mathcal{U}} \le \mathrm{C}\,h(\|\Pi^{\mathrm{BDM}}\boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + d_{2,\infty}h\|\mathsf{P}^{\mathrm{BDM}}u - u\|_{0,\Omega}). \tag{7.5}$$

It is not difficult to see that the remaining two estimates (7.3) and (7.4) follow from (7.5) as in the analysis for the RT method. To prove (7.5), first we repeat the arguments in the proofs of Theorem 3.7 and Corollary 3.8 for the BDM case and get

$$\|u_{\mathcal{T}}^{\mathrm{BDM}} - \mathsf{P}^{\mathrm{BDM}}u\|_{0,\Omega} \le \mathrm{C}\,h(\|\Pi^{\mathrm{BDM}}\boldsymbol{q} - \boldsymbol{q}\|_{\boldsymbol{c}} + d_{1,\infty}\|\mathsf{P}^{\mathrm{BDM}}u - u\|_{0,\Omega}). \tag{7.6}$$

Then, we again follow the lines of the proof of Theorem 3.7 but this time with the following changes: First, in (5.1), we estimate $\|\phi - \phi_{\mathcal{T}}\|_d$ differently: Since

$$\begin{aligned}
\|\phi - \mathsf{P}^{\mathrm{BDM}}\phi\|_{0,\Omega} &\le \mathrm{C}h^2|\phi|_{H^2(\Omega)}, \\
\|\mathsf{P}^{\mathrm{BDM}}\phi - \phi_{\mathcal{T}}\|_{0,\Omega} &\le \mathrm{C}h(h|\boldsymbol{\psi}|_{H^1(\Omega)} + h|\phi|_{H^1(\Omega)}),
\end{aligned}$$

by (7.6). Hence, by triangle inequality $\|\phi - \phi_{\mathcal{T}}\|_{0,\Omega} \le \mathrm{C}h^2(|\boldsymbol{\psi}|_{H^1(\Omega)} + \|\phi\|_{H^2(\Omega)})$. Second, in (5.2), we estimate $d\phi - \mathsf{P}^{\mathrm{BDM}}(d\phi)$ using

$$\|d\phi - \mathsf{P}^{\mathrm{BDM}}(d\phi)\|_{0,\Omega} \le \mathrm{C}h^2 d_{2,\infty}|\phi|_{H^2(\Omega)}.$$

With these revised estimates, completing the proof as before, we obtain (7.5). $\square$

We conclude by emphasizing that the approach we have used here to analyze the variable degree hybridized RT and BDM mixed methods can also be applied to other methods with the same structure. As we have seen, hybridization is useful not only in the construction of mixed methods and their implementation, but it is also a useful theoretical device in their error analysis.

Appendix A. The abstract version of the characterization theorem

In this appendix, we state and prove an abstract version of a result obtained in [7]. Since our intention is to emphasize the generality of the result, we prove it in the following abstract framework:

Assumption A.1. Let $\boldsymbol{V}, W$, and M be arbitrary finite dimensional spaces with inner products $(\cdot,\cdot)_{\boldsymbol{V}}$, $(\cdot,\cdot)_W$, and $(\cdot,\cdot)_M$, respectively. Assume that

(1) $\boldsymbol{A} : \boldsymbol{V} \mapsto \boldsymbol{V}$ is a symmetric and positive definite operator,
(2) $B : \boldsymbol{V} \mapsto W$ is surjective, and
(3) $D : W \mapsto W$ is a symmetric positive semidefinite operator.

We also need to introduce several auxiliary operators. First, we define $(\boldsymbol{\mathfrak{Q}}, \mathfrak{U}) : M \mapsto \boldsymbol{V} \times W$ by

$$\begin{pmatrix} \boldsymbol{A} & -B^t \\ B & D \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathfrak{Q}}\lambda \\ \mathfrak{U}\lambda \end{pmatrix} = \begin{pmatrix} C^t\lambda \\ 0 \end{pmatrix}, \tag{A.1}$$

where C is an operator from $\boldsymbol{V}$ to M. We also define $(\boldsymbol{\mathfrak{Q}}, \mathfrak{U}) : \boldsymbol{V}_{\mathcal{T}} \mapsto \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}}$ by

$$\begin{pmatrix} \boldsymbol{A} & -B^t \\ B & D \end{pmatrix} \begin{pmatrix} \boldsymbol{\mathfrak{Q}}\boldsymbol{\alpha} \\ \mathfrak{U}\boldsymbol{\alpha} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha} \\ 0 \end{pmatrix} \tag{A.2}$$

and $(\mathsf{Q}, \mathsf{U}) : W_{\mathcal{T}} \mapsto \boldsymbol{V}_{\mathcal{T}} \times W_{\mathcal{T}}$ by

$$\begin{pmatrix} \boldsymbol{A} & -B^t \\ B & D \end{pmatrix} \begin{pmatrix} \mathsf{Q}\beta \\ \mathsf{U}\beta \end{pmatrix} = \begin{pmatrix} 0 \\ \beta \end{pmatrix}. \tag{A.3}$$

By virtue of assumption A.1, the above operators are well defined. The next theorem deals with solutions of linear systems of the following form:

$$\begin{pmatrix} \boldsymbol{A} & -B^t & -C^t \\ B & D & 0 \\ C & 0 & 0 \end{pmatrix} \begin{pmatrix} \boldsymbol{p} \\ v \\ \lambda \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha} \\ \beta \\ \gamma \end{pmatrix}. \tag{A.4}$$

Similar equations have been studied earlier [10, 11]. With the above notation, we have the following theorem.

Theorem A.1. *Suppose Assumption* A.1 *holds. Then* (A.4) *holds for some* $(\boldsymbol{p}, v, \lambda) \in \boldsymbol{V} \times W \times M$ *and* $(\boldsymbol{\alpha}, \beta, \gamma) \in \boldsymbol{V} \times W \times M$*, if and only if*

$$\boldsymbol{p} = \boldsymbol{\mathfrak{Q}}\boldsymbol{\alpha} + \mathsf{Q}\beta + \boldsymbol{\mathfrak{Q}}\lambda, \tag{A.5}$$

$$v = \mathfrak{U}\boldsymbol{\alpha} + \mathsf{U}\beta + \mathfrak{U}\lambda, \tag{A.6}$$

and λ *satisfies*

$$a(\lambda, \mu) = b(\mu),$$

for all $\mu \in M$*, where*

$$a(\lambda, \mu) = (\boldsymbol{A}\boldsymbol{\mathfrak{Q}}\lambda, \boldsymbol{\mathfrak{Q}}\mu)_{\boldsymbol{V}} + (D\,\mathfrak{U}\lambda, \mathfrak{U}\mu)_{\boldsymbol{V}},$$
$$b(\mu) = (\boldsymbol{\alpha}, \boldsymbol{\mathfrak{Q}}\mu)_{\boldsymbol{V}} + (\beta, \mathfrak{U}\mu)_W + (\gamma, \mu)_M.$$

Proof. Since we have that

$$\begin{pmatrix} \boldsymbol{A} & -B^t \\ B & D \end{pmatrix} \begin{pmatrix} \boldsymbol{p} \\ v \end{pmatrix} = \begin{pmatrix} \boldsymbol{\alpha} + C^t\lambda \\ \beta \end{pmatrix},$$

by the linearity of the problem and the definition of the auxiliary operators, we immediately get that

$$\boldsymbol{p} = \mathfrak{Q}\boldsymbol{\alpha} + \mathbf{Q}\beta + \mathfrak{Q}\lambda \quad \text{and that} \quad v = \mathfrak{U}\boldsymbol{\alpha} + \mathsf{U}\beta + \mathfrak{U}\lambda\,.$$

It remains to prove the characterization of λ. Inserting the above expression for $\boldsymbol{p}$ in the equation $C\,\boldsymbol{p} = \gamma$, we get

$$C\mathfrak{Q}\lambda = -C\mathfrak{Q}\boldsymbol{\alpha} - C\mathbf{Q}\beta + \gamma,$$

or equivalently,

$$(C\mathfrak{Q}\lambda\,, \mu)_M = -(C\mathfrak{Q}\boldsymbol{\alpha} - C\mathbf{Q}\beta + \gamma, \mu)_M,$$

where μ is an arbitrary element of M. Next, we show that

$$a(\lambda, \mu) = (C\mathfrak{Q}\lambda\,, \mu)_M \quad \text{and that} \quad b(\mu) = (-C\mathfrak{Q}\boldsymbol{\alpha} - C\mathbf{Q}\beta + \gamma, \mu)_M.$$

To do that, we simply use the definitions of the local operators.

Since $(C\mathfrak{Q}\lambda\,, \mu)_M = (\mathfrak{Q}\lambda\,, C^t\mu)_{\boldsymbol{V}}$, if we replace λ by μ in the first equation of (A.1) and then multiply by $\mathfrak{Q}\lambda$, we get

$$(C\mathfrak{Q}\lambda\,, \mu)_M = (\mathfrak{Q}\lambda\,, \boldsymbol{A}\mathfrak{Q}\mu\,)_{\boldsymbol{V}} - (\mathfrak{Q}\lambda\,, B^t\,\mathfrak{U}\mu\,)_{\boldsymbol{V}}.$$

Now, since $(\mathfrak{Q}\lambda\,, B^t\,\mathfrak{U}\mu\,)_{\boldsymbol{V}} = (\,\mathfrak{U}\mu\,, B\mathfrak{Q}\lambda\,)_W$, multiplying the second equation of (A.1) by $\mathfrak{U}\mu$, we obtain

$$-(\,\mathfrak{U}\mu\,, B\mathfrak{Q}\lambda\,)_W = (\,\mathfrak{U}\mu\,, D\,\mathfrak{U}\lambda\,)_W.$$

Hence

$$(C\mathfrak{Q}\lambda\,, \mu)_M = (\mathfrak{Q}\lambda\,, \boldsymbol{A}\mathfrak{Q}\mu\,)_{\boldsymbol{V}} + (\,\mathfrak{U}\mu\,, D\,\mathfrak{U}\lambda\,)_W = a(\lambda, \mu),$$

as claimed.

Using similar algebraic manipulations, we easily get that

$$-(C\mathfrak{Q}\boldsymbol{\alpha}\,, \mu)_M = (\boldsymbol{\alpha}, \mathfrak{Q}\mu\,)_{\boldsymbol{V}} \quad \text{and} \quad -(C\mathbf{Q}\beta\,, \mu)_M = (\beta, \mathfrak{U}\mu\,)_{\boldsymbol{V}}.$$

This completes the proof. □

Appendix B. The interpolation operator Π

Here we give details of the construction of a variable degree projection Π that preserves $\boldsymbol{H}(\mathrm{div}, \Omega)$-continuity. The projection given here is a straightforward modification of the projection introduced in [8]. Although the Raviart-Thomas spaces were not considered in [8], it is easy to adapt their arguments to this case, as we now show.

Let $\hat{K}$ be a fixed reference simplex. To every face e of $\hat{K}$, let us associate a nonnegative integer ℓ_e and let ℓ be an integer such that $\ell_e \le \ell$. Define

$$\widehat{\boldsymbol{V}}_{\ell,\ell_e} = \{\boldsymbol{r} \in \boldsymbol{x}P^\ell(\hat{K}) + P^\ell(\hat{K})^N : \boldsymbol{r}\cdot\boldsymbol{n}_e \in P^{\ell_e}(e) \text{ for all faces } e \text{ of } \hat{K}\}.$$

The projection $\hat{\Pi}_{\ell,\ell_e}\boldsymbol{r}$, for any smooth $\boldsymbol{r}$ is defined as the unique function $\boldsymbol{\pi} \in \widehat{\boldsymbol{V}}_{\ell,\ell_e}$ satisfying

$$\begin{aligned}
(\boldsymbol{\pi}\cdot\boldsymbol{n}_e, \mu)_e &= (\boldsymbol{r}\cdot\boldsymbol{n}_e, \mu)_e && \text{for all } \mu \in P^{\ell_e}(e) \text{ and faces } e,\\
(\mathrm{div}\,\boldsymbol{\pi}, \mathrm{div}\,\boldsymbol{v})_{\hat{K}} &= (\mathrm{div}\,\boldsymbol{r}, \mathrm{div}\,\boldsymbol{v})_{\hat{K}} && \text{for all } \boldsymbol{v} \in \widehat{\boldsymbol{V}}_{\ell,\ell_e} \text{ with } \boldsymbol{v}\cdot\boldsymbol{n}|_{\partial\hat{K}} = 0,\\
(\boldsymbol{\pi}, \boldsymbol{v})_{\hat{K}} &= (\boldsymbol{r}, \boldsymbol{v})_{\hat{K}} && \text{for all } \boldsymbol{v} \in \widehat{\boldsymbol{V}}_{\ell,\ell_e} \text{ with } \boldsymbol{v}\cdot\boldsymbol{n}|_{\partial\hat{K}} = 0 \text{ and } \mathrm{div}\,\boldsymbol{v} = 0.
\end{aligned}$$

These conditions uniquely specify a $\boldsymbol{\pi} \in \widehat{\boldsymbol{V}}_{\ell,\ell_e}$ because if the right-hand sides above are zero, then obviously the only $\boldsymbol{\pi} \in \widehat{\boldsymbol{V}}_{\ell,\ell_e}$ that satisfies the conditions is zero.

The global interpolant Π can now be defined by mapping $\hat{\Pi}_{\ell,\ell_e}$ to K by the Piola map after appropriately setting the degrees ℓ and ℓ_e. First let $\widehat{K}$ be mapped one-to-one onto K by the standard affine mapping $\boldsymbol{x} = D_K\,\widehat{\boldsymbol{x}} + b_K$, and let vector functions be transformed between $\widehat{K}$ and K by

$$\widehat{\boldsymbol{q}}(\widehat{\boldsymbol{x}}) \equiv |\det D_K|\, D_K^{-1}\, \boldsymbol{q}(\boldsymbol{x}).$$

With this notation, we define the global projection $\Pi\boldsymbol{q}$ for a smooth $\boldsymbol{q}$, element by element: On any element K, the restriction $(\Pi\boldsymbol{q})|_K$ is defined by

$$\widehat{\Pi\boldsymbol{q}} = \hat{\Pi}_{\ell,\ell_e}\widehat{\boldsymbol{q}},$$

with ℓ equal to $k(K)$, with $\ell_e = \min(k(K_e^+), k(K_e^-))$ whenever e is an interior face (see (2.7)), and $\ell_e = k(K)$ if e is a face on $\partial\Omega$. It is easy to see that $\Pi\boldsymbol{q} \in \boldsymbol{H}(\mathrm{div},\Omega)$ for any smooth $\boldsymbol{q}$. It is possible to find the domain of continuity of Π by usual arguments, but we omit such details; e.g., Π is a continuous operator on $\boldsymbol{H}(\mathrm{div},\Omega) \cap L^r(\Omega)^2$ for $r > 2$ in two space dimensions.

This projection can be applied in the earlier analysis provided we can prove the critical commutativity property (3.1). Because of the properties of the Piola mapping, we only need to prove that

$$\int_{\hat{K}} p\ \mathrm{div}(\widehat{\Pi\boldsymbol{q}})\ \mathrm{d}x = \int_{\hat{K}} p\ \mathrm{div}\,\widehat{\boldsymbol{q}}\ \mathrm{d}x, \qquad \text{for all } p \in P^{k(K)}(\hat{K}). \tag{B.1}$$

First, observe that this identity holds with $p = 1$, because by the definition of the projection,

$$\int_{\partial\hat{K}} (\widehat{\Pi\boldsymbol{q}} - \widehat{\boldsymbol{q}})\cdot\boldsymbol{n}\ \mathrm{d}s = 0. \tag{B.2}$$

Thus (B.1) holds for lowest order elements. Now consider the case $k(K) > 0$. Set $\ell = k(K)$. Because of (B.2), it suffices to prove (B.1) for all $p \in P^\ell(\hat{K})$ with zero mean. This follows from the definition of the projection, provided $\mathrm{div} : \{\boldsymbol{r} \in \boldsymbol{x}P^\ell(\hat{K}) + P^\ell(\hat{K})^N : \boldsymbol{r}\cdot\boldsymbol{n} = 0 \text{ on } \partial\hat{K}\} \mapsto \{p \in P^\ell(\hat{K}) : \int_{\hat{K}} p\,\mathrm{d}x = 0\}$ is a surjection. To prove this surjectivity, consider the function u satisfying $\Delta u = p$ on $\hat{K}$ and $\boldsymbol{n}\cdot\mathbf{grad}\,u = 0$ on $\partial\hat{K}$, for any given $p \in P^\ell(\hat{K})$. Let $\boldsymbol{q} \in \boldsymbol{x}P^\ell(\hat{K}) + P^\ell(\hat{K})^N$ be defined by $\boldsymbol{q}\cdot\boldsymbol{n} = 0$ on $\partial\hat{K}$ and $(\boldsymbol{q},\boldsymbol{r})_{\hat{K}} = (\mathbf{grad}\,u, \boldsymbol{r})_{\hat{K}}$ for all $\boldsymbol{r} \in P^{\ell-1}(\hat{K})^N$. In other words, $\boldsymbol{q}$ is the standard Raviart-Thomas projection of $\mathbf{grad}\,u$, and hence well defined; see, e.g., [6, Proposition III.3.3]. Then $(\mathrm{div}\,\boldsymbol{q}, s)_{\hat{K}} = -(\boldsymbol{q}, \mathbf{grad}\,s)_{\hat{K}} = -(\mathbf{grad}\,u, \mathbf{grad}\,s)_{\hat{K}} = (p, s)_{\hat{K}}$ for all $s \in P^\ell(\hat{K})$, so $\mathrm{div}\,\boldsymbol{q} = p$. This proves the surjectivity.

References

[1] T. Arbogast and Z. Chen, *On the implementation of mixed methods as nonconforming methods for second-order elliptic problems*, Math. Comp., 64 (1995), pp. 943–972. MR1303084 (95k:65102)

[2] D. N. Arnold and F. Brezzi, *Mixed and non-conforming finite element methods: implementation, post-processing and error estimates*, Modél. Math. Anal.Numér., 19 (1985), pp. 7–35. MR0813687 (87g:65126)

[3] P. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, vol. 15 of Texts in Applied Mathematics, Springer Verlag, 1994. MR1278258 (95f:65001)

[4] F. Brezzi, J. Douglas, Jr., and D. Marini, *Two families of mixed finite elements for second order elliptic problems*, Numer. Math., 47 (1985), pp. 217–235. MR0799685 (87g:65133)

[5] ———, *Variable degree mixed methods for second order elliptic problems*, Mat. Aplic. e Comp., 4 (1985), pp. 19–34. MR0808322 (87d:65129)

[6] F. Brezzi and M. Fortin, *Mixed and Hybrid finite element methods*, Springer Verlag, 1991. MR1115205 (92d:65187)

[7] B. Cockburn and J. Gopalakrishnan, *A characterization of hybridized mixed methods for the Dirichlet problem*, SIAM J. Numer. Anal., 42 (2004), pp. 283–301. MR2051067

[8] L. Demkowicz, P. Monk, L. Vardapetyan, and W. Rachowicz, *De Rham diagram for hp finite element spaces*, Comput. Math. Appl., 39 (2000), pp. 29–38. MR1746160 (2000m:78052)

[9] B. M. Fraejis de Veubeke, *Displacement and equilibrium models in the finite element method*, in Stress Analysis, O. Zienkiewicz and G. Holister, eds., Wiley, New York, 1977, pp. 145–197.

[10] G. N. Gatica, *Solvability and Galerkin approximations of a class of nonlinear operator equations*, Z. Anal. Anwendungen, 21 (2002), pp. 761–781. MR1929431 (2003h:65074)

[11] G. N. Gatica, N. Heuer, and S. Meddahi, *On the numerical analysis of nonlinear twofold saddle point problems*, IMA J. Numer. Anal., 23 (2003), pp. 301–330. MR1975268 (2004b:65183)

[12] J. Gopalakrishnan, *A Schwarz preconditioner for a hybridized mixed method*, Computational Methods in Applied Mathematics, 3 (2003), pp. 116–134. MR2002260 (2004g:65033)

[13] R. Kirby and C. Dawson, *Private Communication* (2003).

[14] I. Perugia and D. Schötzau, *An hp-analysis of the local discontinuous Galerkin method for diffusion problems*, J. Sci. Comput. (Special Issue: Proceedings of the Fifth International Conference on Spectral and High Order Methods (ICOSAHOM-01), Uppsala, Sweden), 17 (2002), pp. 561–571. MR1910752

[15] P. Raviart and J. Thomas, *A mixed method for second order elliptic problems*, in Mathematical Aspects of the Finite Element Method, I. Galligani and E. Magenes, eds., Lecture Notes in Math. 606, Springer-Verlag, New York, 1977. MR0483555 (58:3547)

[16] M. Suri, *On the stability and convergence of higher-order mixed finite element methods for second-order elliptic problems*, Math. Comp., 54 (1990), pp. 1–19. MR0990603 (90e:65164)

School of Mathematics, University of Minnesota, Vincent Hall, Minneapolis, Minnesota 55455

E-mail address: cockburn@math.umn.edu

Department of Mathematics, University of Florida, Gainesville, Florida 32611–8105

E-mail address: jayg@math.ufl.edu