IMA Journal of Numerical Analysis (2013) Page 1 of 38
doi:10.1093/imanum/drnxxx

Multigrid for an HDG Method

B. COCKBURN †
School of Mathematics, University of Minnesota, Vincent Hall, Minneapolis, Minnesota, USA.

O. DUBOIS ‡
École Polytechnique de Montréal, Québec, Canada.

J. GOPALAKRISHNAN§
Portland State University, Portland, Oregon, USA.

AND

S. TAN ¶
Department of Mathematics, Texas A&M University, College Station Texas, USA.

Dedicated to Raytcho Lazarov, on the occasion of his 70$^{\text{th}}$ birthday.

[Received on [Date]]

We analyze the convergence of a multigrid algorithm for the Hybridizable Discontinuous Galerkin (HDG) method for diffusion problems. We prove that a non-nested multigrid V-cycle, with a single smoothing step per level, converges at a mesh independent rate. Along the way, we study conditioning of the HDG method, prove new error estimates for it, and identify an abstract class of problems for which a non-nested two-level multigrid cycle with one smoothing step converges even when the prolongation norm is greater than one. Numerical experiments verifying our theoretical results are presented.

Keywords: multigrid methods; discontinuous Galerkin methods; hybrid methods.

1. Introduction

In this paper we present the *first* convergence study of a multigrid algorithm for the Hybridizable Discontinuous Galerkin (HDG) method introduced in (Cockburn *et al.*, 2009b). The main multigrid result of this paper is that a *non-nested multigrid V-cycle, with one smoothing per level, applied to the HDG method, converges at a mesh independent rate.* We are able to prove this result by using results that exploit the specific structure in the HDG method. For example, a technical tool that is available in the HDG case, and not in other DG methods, is a projection operator designed in (Cockburn *et al.*, 2010). The analysis of this paper uses it together with other interesting properties of the HDG scheme.

For perspective, the HDG methods were devised for efficient implementation (Cockburn *et al.*, 2009b). They are as competitive as mixed or other conforming methods. Indeed, the HDG methods were constructed in such a way that the only globally coupled degrees of freedom are the so-called

[†]Partially supported by the National Science Foundation (Grant DMS-0712955) and by the Minnesota Supercomputing Institute. Email:cockburn@math.umn.edu
[‡]Email: olivier.dubois@polymtl.ca
[§]Supported in part by the NSF under grant DMS-1211635. Email: gjay@pdx.edu
[¶]Email: stan232@bloomberg.net

"numerical traces" of the scalar variable. This property is shared by the hybridized version of classical mixed methods (see, e.g., Cockburn & Gopalakrishnan, 2004, and the references therein) and can be thought of as an extension of the well-known technique of "static condensation". The specific HDG method we are concerned with here is constructed by using the local discontinuous Galerkin (LDG) method on each element. For this reason, it is often called the LDG-H method, but here we refer to it simply as the HDG method.

The close relationship between HDG and the classical mixed Raviart-Thomas (RT) and Brezzi-Douglas-Marini (BDM) methods was highlighted in (Cockburn *et al.*, 2009b). Soon enough, this relationship was exploited to show that the HDG method shared the convergence and superconvergence properties of mixed method (Cockburn *et al.*, 2009a, 2008). Moreover, a specific HDG method was shown to have exactly the same stiffness matrix as the hybridized version of the RT and BDM methods of corresponding degree (Cockburn *et al.*, 2008). An earlier paper (Cockburn & Gopalakrishnan, 2004) had shown a similar result that the hybridized version of the RT and BDM methods had the same stiffness matrix. The multigrid results we present are valid for quite general HDG methods, so these results imply their applicability for the RT and BDM methods.

Further understanding of the similarities between the mixed and HDG methods was made in the recent error analysis of (Cockburn *et al.*, 2010) based on a new projection akin to the projections used in the analyses of the RT and BDM methods. This projection will be used extensively this paper. The error estimates we obtain here are new and complement those obtained in (Cockburn *et al.*, 2010). Moreover, they are the estimates needed in our multigrid estimates. The multigrid algorithm exhibits further evidence of the relationship between the RT and the HDG methods. Specifically, the intergrid transfer operators used in the multigrid method proposed in (Gopalakrishnan & Tan, 2009) for the hybridized version of the RT method are exactly the same operators we are using for the HDG method. However, the multigrid convergence results we obtain in this paper are better.

To elaborate, let us review some recent developments in analysis of solvers for DG methods. The first work to apply multigrid theory to a DG discretization (Gopalakrishnan & Kanschat, 2003b) considered the interior penalty DG method. It was generalized to other DG methods in (Gopalakrishnan & Kanschat, 2003a) using the unified analysis of (Arnold *et al.*, 0102). While (Gopalakrishnan & Kanschat, 2003a,b) considered the geometric multigrid setting, algebraic multigrid techniques were considered in (Kraus & Tomar, 2008b,a). All these works use multigrid techniques to develop an optimal preconditioner for use within nonlinear iterations such as conjugate gradients. In contrast, in this paper, we consider the use of a multigrid V-cycle as a linear iteration, which is less complex and slightly less expensive. (A linear iteration that reduces error uniformly can always be used as an optimal preconditioner (Bramble & Zhang, 2000), but not vice versa, in general.) Moreover, we apply the algorithm to an HDG method, a DG method that is not covered by the previous analyses.

A recurring technique in many works, which we will borrow, is the use of conforming or continuous subspaces of DG spaces to design DG solvers. One can find this in the application of domain decomposition techniques to DG methods (Antonietti & Ayuso de Dios, 2007; Feng & Karakashian, 2005). Iterative methods based on a decomposition of discontinuous approximation spaces as a conforming space or piecewise linear nonconforming space plus a remainder are considered in (Ayuso de Dios & Zikatanov, 2009) and (Dobrev *et al.*, 2006), respectively. Convergence of V-cycle, F-cycle, and W-cycle algorithms for non-conforming methods were proven in (Brenner & Sung, 2006) under the assumption that the number of smoothing steps is sufficiently large. Under the same assumption, multigrid for DG methods were considered in (Brenner & Zhao, 2005; Duan *et al.*, 2007). How large the number of smoothing steps should be in practice is difficult to determine from these analyses. In contrast, we are able to obtain convergence with just one post and pre-smoothing in a multilevel setting that uses spaces of continuous functions at coarser levels, i.e., our multilevel spaces are nested except at the finest level. This is also an improvement over the analysis of the multigrid method for the hybridized RT method in (Gopalakrishnan & Tan, 2009).

We identify an abstract class of problems for which one can prove convergence of a non-nested two-level V-cycle algorithm with just one smoothing per level. This two-level result can be extended to a multilevel result by simply imbedding a nested hierarchy of coarse spaces. Convergence results under the assumption that the number of smoothings increase geometrically as we go to coarser levels (Bramble *et al.*, 1991), or that a uniform but sufficiently large (practically unknown) number of smoothings be performed at each level (Brenner, 2004; Duan *et al.*, 2007) are already known. Nonetheless, it has often been observed that some nonnested algorithms when applied to DG methods converge even with one smoothing (see, e.g., Gopalakrishnan & Kanschat, 2003b). From this perspective, this paper's contribution is to identify a multilevel setting, particularly suitable for DG methods, where theory and practice meets. Namely, we give a multilevel setting with one pre-and-post-smoothing per level that yields uniform convergence theoretically and practically. We are also able, in this setting, to relax the often-made assumption that a certain norm of the intergrid transfer (prolongation) operator is at most one by admitting $O(h)$ perturbations from the unit norm.

In the course of arriving at the multigrid convergence result, we develop the following intermediate results which are independently interesting:

1. We give an error analysis of the HDG method. Although HDG error estimates were given in (Cockburn *et al.*, 2010), the new estimates obtained in this paper hold without the regularity assumptions used in (Cockburn *et al.*, 2010) and are proved without using a duality argument.

2. We study the condition number of (the unpreconditioned) HDG matrix system. The technique of the above mentioned error analysis proceeds by first proving local stability estimates for the HDG method. These results immediately yield the condition number estimates.

The rest of the paper is organized as follows. We begin by recalling the HDG method in Section 2 including the preferred hybridized form of the implementation, which will be the target of multigrid application. In Section 3, we prove the new stability, error, and conditioning estimates. The multigrid algorithm is discussed in Section 4, which also states two theorems in an abstract framework, potentially applicable to other hybrid bilinear forms. The convergence analysis of the multigrid algorithm applied to the HDG method appears in Section 5. The main result is Theorem 5.1. Section 6 details our numerical results. The proofs of the abstract multigrid theorems are in Appendix B.

2. The HDG method

In this section, we recall the HDG method and discuss several preliminary details, including its preferred form for implementation.

2.1 *The definition of the method*

The HDG method we now describe was first introduced in (Cockburn *et al.*, 2009b), where it was also called the LDG-H method. The method is applied here to the partial differential equation $-\nabla\cdot(a(\vec{x})\,\vec{\nabla}u) = f$ with the Dirichlet boundary condition $u = g$ on the boundary (where "$\vec{\nabla}$" and "$\nabla\cdot$" denotes the gradient and divergence, respectively). Introducing the "flux" $\vec{q}$, we can rewrite this boundary value problem as

$$\vec{q} + a(\vec{x})\,\vec{\nabla} u = 0 \qquad \text{on } \Omega, \tag{2.1a}$$
$$\nabla \cdot \vec{q} = f \qquad \text{on } \Omega, \tag{2.1b}$$
$$u = g \qquad \text{on } \partial\Omega. \tag{2.1c}$$

Here $\Omega \subset \mathbb{R}^n$ is a polyhedral domain ($n \geqslant 2$), $a : \Omega \to \mathbb{R}^{n\times n}$ denotes a variable matrix valued coefficient, which we assume to be symmetric and uniformly positive definite, f is in $L^2(\Omega)$ and $g \in H^{1/2}(\partial\Omega)$. The domain Ω is subdivided into simplices forming a mesh $\mathscr{T}_h$ satisfying the standard finite element conditions for geometrical conformity (Ciarlet, 1978). The HDG method defines a scalar approximation u_h to u and a vector approximation $\vec{q}_h$ to $\vec{q}$ in the following spaces, respectively:

$$W_h = \{w : \text{ for every mesh element } K, w|_K \in P_d(K)\}, \tag{2.2}$$
$$V_h = \{\vec{v} : \text{ for every mesh element } K, \vec{v}|_K \in P_d(K)^n\}. \tag{2.3}$$

Note that functions in these spaces need not be continuous across element interfaces. Above and elsewhere, we use $P_d(D)$ to denote the space of polynomials of degree at most $d \geqslant 0$ on some domain D. The subscript h denotes the mesh size defined as the maximum of the diameters of all mesh elements.

For any (scalar or vector) function q in V_h or W_h, the trace $q|_F$ is, in general, a double-valued function on any interior mesh face $F = \partial K^+ \cap \partial K^-$ shared by the mesh elements K^+ and K^-. Its two branches, denoted by $[q]_{K^+}$ and $[q]_{K^-}$, are defined by $[q]_{K^\pm}(\vec{x}) = \lim_{\varepsilon\downarrow 0} q(\vec{x} - \varepsilon\,\vec{n}_{K^\pm})$ for all $\vec{x}$ in F. Here and elsewhere, $\vec{n}$ denotes the double-valued function of unit normals on the element interfaces: on a face $F \subseteq \partial K$, its branch $[\vec{n}]_K$ equals the unit normal on ∂K pointing outward from K. For functions u and v in $L^2(D)$, we write $(u,v)_D = \int_D uv\,dx$ whenever D is a domain of $\mathbb{R}^n$, and $\langle u,v\rangle_D = \int_D uv\,dx$ whenever D is an $(n-1)$-dimensional domain. To simplify the notation, define

$$(v,w)_{\mathscr{T}_h} = \sum_{K\in\mathscr{T}_h} (v,w)_K \qquad \text{and} \qquad \langle v,w\rangle_{\partial\mathscr{T}_h} = \sum_{K\in\mathscr{T}_h} \langle v,w\rangle_{\partial K},$$

where in the latter, we understand that for double valued v and w, the integral $\langle v,w\rangle_{\partial K}$ is computed using the branches $[v]_K$ and $[w]_K$ from K. For vector functions $\vec{v}$ and $\vec{w}$, the notations are similarly defined with the integrand being the dot product $\vec{v}\cdot\vec{w}$.

In addition to the spaces V_h and W_h introduced above, our method also uses another discrete space M_h, consisting of functions defined on the domain $\cup_{K\in\mathscr{T}_h}\partial K$, namely

$$M_h = \{\mu : \text{ for every mesh face } F,\ \mu|_F \text{ is in } P_d(F), \text{ and if } F \subseteq \partial\Omega,\ \mu|_F = 0\}. \tag{2.4}$$

Clearly, a function in M_h is supported only on the interior mesh faces (or edges if $n = 2$). The HDG method defines the approximations $\vec{q}_h$, u_h, and λ_h, as the functions in V_h, W_h and M_h, respectively, satisfying

$$(c\,\vec{q}_h, \vec{r})_{\mathscr{T}_h} - (u_h, \nabla\cdot\vec{r})_{\mathscr{T}_h} + \langle \lambda_h, \vec{r}\cdot\vec{n}\rangle_{\partial\mathscr{T}_h} = -\langle g, \vec{r}\cdot\vec{n}\rangle_{\partial\Omega}, \quad \text{for all } \vec{r} \in V_h, \tag{2.5a}$$
$$-(\vec{q}_h, \vec{\nabla} w)_{\mathscr{T}_h} + \langle \widehat{q}_h\cdot\vec{n}, w\rangle_{\partial\mathscr{T}_h} = (f,w)_{\mathscr{T}_h} \quad \text{for all } w \in W_h, \tag{2.5b}$$
$$\langle \mu, \widehat{q}_h\cdot\vec{n}\rangle_{\partial\mathscr{T}_h} = 0 \quad \text{for all } \mu \in M_h, \tag{2.5c}$$

where $c = a^{-1}$ and $\widehat{q}_h$ is a double-valued vector function on mesh interfaces defined by

$$\widehat{q}_h = \vec{q}_h + \tau_K\big(u_h - \lambda_h\big)\vec{n}.$$

Note that this defines all branches, i.e., on the boundary ∂K of every mesh element K, the value of the branch of $\widehat{q}_h$ from K is $[\widehat{q}_h]_K = [\vec{q}_h]_K + [\tau]_K\big([u_h]_K - \lambda_h\big)[\vec{n}]_K$. Here τ is a non-negative penalty function. Note that τ is also a double valued function on the element interfaces and τ_K above denotes the branch of τ-values from K. For simplicity, we assume that any branch of τ is a constant function on each mesh edge. It is proved in (Cockburn *et al.*, 2009b) that the system (2.5) is uniquely solvable if τ_K is positive on at least one face of K for every element K. Such unique solvability results hold for other choices of $\widehat{q}_h$ which generate other HDG methods, as expounded in (Cockburn *et al.*, 2009b).

2.2 *Preferred form for implementation*

The main advantage of HDG methods is that, unlike many DG methods, we can eliminate the variables $\vec{q}_h$ and u_h from (2.5) to obtain a single equation for λ_h. Thus the often made criticism that DG methods have too many unknowns does not apply to HDG methods. Moreover, $\vec{q}_h$ and u_h can be locally recovered once λ_h is found. To precisely state this result, it will be notationally efficient to rewrite (2.5) in terms of the following operators. Define $\mathcal{A}: V_h \to V_h$, $\mathcal{B}: V_h \to W_h$, $\mathcal{C}: V_h \to M_h$, $\mathcal{R}: W_h \to W_h$, $\mathcal{S}: W_h \to M_h$, and $\mathcal{T}: M_h \to M_h$, by

$$(\mathcal{A}\vec{p},\vec{r})_{\mathcal{T}_h} = (c\,\vec{p},\vec{r})_{\mathcal{T}_h}, \qquad (\mathcal{B}\vec{r},w)_{\mathcal{T}_h} = -(w,\nabla\cdot\vec{r})_{\mathcal{T}_h}, \qquad \langle \mathcal{C}\vec{r},\mu\rangle_{\partial\mathcal{T}_h} = \langle \mu,\vec{r}\cdot\vec{n}\rangle_{\partial\mathcal{T}_h},$$
$$(\mathcal{R}w,v)_{\mathcal{T}_h} = -\langle \tau w,v\rangle_{\partial\mathcal{T}_h}, \qquad \langle \mathcal{S}w,\mu\rangle_{\partial\mathcal{T}_h} = \langle \tau w,\mu\rangle_{\partial\mathcal{T}_h}, \qquad \langle \mathcal{T}\mu,\eta\rangle_{\partial\mathcal{T}_h} = -\langle \tau\mu,\eta\rangle_{\partial\mathcal{T}_h},$$

for all $\vec{p},\vec{r}\in V_h$, $w,v\in W_h$, and $\mu,\eta\in M_h$. The HDG method generates operator equations of the form

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}^t & \mathcal{C}^t \\ \mathcal{B} & \mathcal{R} & \mathcal{S}^t \\ \mathcal{C} & \mathcal{S} & \mathcal{T} \end{pmatrix} \begin{pmatrix} \vec{q}_h \\ u_h \\ \lambda_h \end{pmatrix} = \begin{pmatrix} \vec{g}_h \\ f_h \\ 0 \end{pmatrix}, \tag{2.6}$$

for some $\vec{g}_h \in V_h$ and $f_h \in W_h$, where the superscript "t" denotes the adjoint with respect to $(\cdot,\cdot)_{\mathcal{T}_h}$ or $\langle\cdot,\cdot\rangle_{\partial\mathcal{T}_h}$ as appropriate. It is easy to see that (2.5) can be rewritten as the above system with $f_h = P_h^W f$, where P_h^W denotes the $L^2(\Omega)$-orthogonal projection into W_h, and $\vec{g}_h$ set to the unique function in V_h satisfying

$$(\vec{g}_h,\vec{r})_{\mathcal{T}_h} = -\langle g,\vec{r}\cdot\vec{n}\rangle_{\partial\Omega} \quad \text{for all } \vec{r}\in V_h. \tag{2.7}$$

Note that in the lowest order case $d=0$, the operator $\mathcal{B}$ is zero, but the system continues to be uniquely solvable.

The result on the above mentioned elimination can be described using additional "local" operators $\vec{\mathcal{Q}}_V: V_h \to V_h$, $\vec{\mathcal{Q}}_W: W_h \to V_h$, $\mathcal{U}_V: V_h \to W_h$, $\mathcal{U}_W: W_h \to W_h$, whose action is defined by solving the following systems

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}^t \\ \mathcal{B} & \mathcal{R} \end{pmatrix} \begin{pmatrix} \vec{\mathcal{Q}}_V\vec{g}_h \\ \mathcal{U}_V\vec{g}_h \end{pmatrix} = \begin{pmatrix} \vec{g}_h \\ 0 \end{pmatrix} \qquad\qquad \begin{pmatrix} \mathcal{A} & \mathcal{B}^t \\ \mathcal{B} & \mathcal{R} \end{pmatrix} \begin{pmatrix} \vec{\mathcal{Q}}_W f_h \\ \mathcal{U}_W f_h \end{pmatrix} = \begin{pmatrix} 0 \\ f_h \end{pmatrix} \tag{2.8}$$

for all $\vec{g}_h \in V_h$ and $f_h \in W_h$. Let $\vec{\mathcal{Q}}\mu = -\vec{\mathcal{Q}}_V(\mathcal{C}^t\mu) - \vec{\mathcal{Q}}_W(\mathcal{S}^t\mu)$ and $\mathcal{U}\mu = -\mathcal{U}_V(\mathcal{C}^t\mu) - \mathcal{U}_W(\mathcal{S}^t\mu)$. Note that all these operators are local – for example, $\vec{\mathcal{Q}}\mu$ and $\mathcal{U}\mu$ can be computed on an element K independently of all other elements, solely using the values of μ on ∂K, because (2.8) implies that

$$(c\,\vec{\mathcal{Q}}\mu,\vec{r})_K - (\mathcal{U}\mu,\nabla\cdot\vec{r})_K = -\langle \mu,\vec{r}\cdot\vec{n}\rangle_{\partial K} \qquad \text{for all } \vec{r}\in V_h \tag{2.9a}$$
$$(w,\nabla\cdot\vec{\mathcal{Q}}\mu)_K + \langle \tau(\mathcal{U}\mu-\mu),w\rangle_{\partial K} = 0 \qquad \text{for all } w\in W_h. \tag{2.9b}$$

The meaning of such operators is amply discussed in (Cockburn *et al.*, 2009b) – and in (Cockburn & Gopalakrishnan, 2004, 2005) in the context of mixed methods – so we will not repeat. We have the following theorem.

THEOREM 2.1 Suppose $\vec{q}_h$ and $\vec{g}_h$ are in V_h, u_h and f_h are in W_h, and λ_h is in M_h. Then $\vec{q}_h, u_h, \lambda_h$ satisfy (2.6) if and only if λ_h is the only function in M_h satisfying

$$a_h(\lambda_h, \mu) = b_h(\mu) \quad \text{for all } \mu \in M_h, \text{ and} \tag{2.10}$$

$$\vec{q}_h = \vec{\mathcal{Q}}\lambda_h + \vec{\mathcal{Q}}_W f_h + \vec{\mathcal{Q}}_V \vec{g}_h, \tag{2.11}$$

$$u_h = \mathcal{U}\lambda_h + \mathcal{U}_W f_h + \mathcal{U}_V \vec{g}_h, \tag{2.12}$$

where

$$a_h(\eta, \mu) = (c\,\vec{\mathcal{Q}}\eta, \vec{\mathcal{Q}}\mu)_{\mathscr{T}_h} + \langle \tau(\mathcal{U}\eta - \eta), (\mathcal{U}\mu - \mu)\rangle_{\partial\mathscr{T}_h}, \tag{2.13}$$
$$b_h(\mu) = (f_h, \mathcal{U}\mu)_{\mathscr{T}_h} - (\vec{g}_h, \vec{\mathcal{Q}}\mu)_{\mathscr{T}_h}.$$

Its proof proceeds exactly as a proof in (Cockburn *et al.*, 2009b) so we omit it (the differences are only in the additional terms involving $\vec{g}_h$, which creates no complications). Theorem 2.1 clearly demonstrates the previously mentioned advantages of the HDG method. It also shows the preferred form of implementation of the method. Indeed, we should not implement the method in the form (2.5). Instead, we should compute the solution of (2.5) by first solving for λ_h from (2.10), and then recovering $\vec{q}_h$ and u_h locally (element by element) using (2.11) and (2.12). Unlike (2.5), implementation of (2.10) results in a symmetric positive definite matrix system. Moreover, since (2.10) only involves λ_h, it gives a smaller system than (2.5).

The most computationally intensive step is the solution of equation (2.10), which results in a large sparse matrix system. To investigate the performance of iterative techniques applied to (2.10), we will need to study its conditioning, as done in the next section.

3. Estimates for the HDG method

This section is devoted to obtaining estimates on the stability, conditioning, and discretization errors of the HDG method. Our technique consists of first obtaining bounds for various local solution operators of the HDG method. The local bounds then imply global bounds, such as bounds for the discretization errors and the spectrum of the operator associated with the HDG (bilinear) form.

Before we begin, let us mention a few conventions in all the estimates of this paper. Let h_K denote the diameter of a mesh element K. Throughout, constants that do not depend on h_K are generically denoted by C. Their value may differ at different occurrences, and may depend on the the shape regularity of the mesh. While dependencies on the coefficient $c(x) = a(x)^{-1}$ will be absorbed into C, any dependencies on τ will always be explicitly mentioned. Finally, for any domain D we denote by $\|\cdot\|_D$ the $L^2(D)$-norm (or the product norm in $(L^2(D))^n$ for vector functions). Set $h_K = \text{diam}(K)$ and define

$$\|\lambda\|_{h,D} = \left(\sum_{K\in\mathscr{T}_h,\, K\subseteq\bar{D}} \|\lambda\|^2_{L^2(\partial K)} \frac{|K|}{|\partial K|} \right)^{1/2}. \tag{3.1}$$

Here, $|K|$ and $|\partial K|$ denote the n and $(n-1)$-dimensional measures of K and ∂K, respectively. When the domain under consideration is Ω, we drop Ω as a subscript in notations whenever no confusion can arise, e.g., we often use $\|\cdot\|_h$, and $\|\cdot\|$ to denote $\|\cdot\|_{h,\Omega}$, and $\|\cdot\|_\Omega$, respectively.

3.1 *Stability of the local HDG solutions*

In this subsection, we will establish the following result giving bounds on various local solution operators. Its proof will be completed at the end of this subsection.

THEOREM 3.1 The local solution operators obey the following bounds:

$$\|\vec{\mathcal{Q}}\mu\|_K \leqslant c_{h\tau}^K C h_K^{-1}\|\mu\|_{h,K}, \qquad \|\mathcal{U}\mu\|_K \leqslant c_{h\tau}^K C\|\mu\|_{h,K}, \tag{3.2}$$

$$\|\vec{\mathcal{Q}}_W f\|_K \leqslant d_{h\tau}^K C h_K\|f\|_K, \qquad \|\mathcal{U}_W f\|_K \leqslant (d_{h\tau}^K)^2 C h_K^2\|f\|_K, \tag{3.3}$$

$$\|\vec{\mathcal{Q}}_V \vec{g}\|_K \leqslant C\|\vec{g}\|_K, \qquad \|\mathcal{U}_V \vec{g}\|_K \leqslant d_{h\tau}^K C h_K\|\vec{g}\|_K. \tag{3.4}$$

where $c_{h\tau}^K = 1+(\tau_K^* h_K)^{1/2}$ and $d_{h\tau}^K = 1+(\tau_K^{\max} h_K)^{-1/2}$. Here $\tau_K^{\max}$ denotes the maximum value of τ_K on ∂K and τ_K^* denotes the maximum value of τ_K on $\partial K \setminus F_{\max}$ where $F_{\max}$ is any face at which $\tau_K = \tau_K^{\max}$.

In (Cockburn & Gopalakrishnan, 2005), we established similar estimates for the local solution operators of the Raviart-Thomas (RT) method. Since we shall use these, let us recall the Raviart-Thomas spaces. Let $R_d(K)$ denote the space of all polynomials of the form $\vec{p}_d + \vec{x} q_d$ for some $\vec{p}_d \in P_d(K)^n$ and some q_d in $P_d(K)$. Then define the local RT liftings $\vec{\mathcal{Q}}^{\rm RT}(\cdot)$ and $\mathcal{U}^{\rm RT}(\cdot)$ by

$$(c\,\vec{\mathcal{Q}}^{\rm RT}\mu, \vec{r})_K - (\mathcal{U}^{\rm RT}\mu, \nabla\cdot\vec{r})_K = -\langle \mu, \vec{r}\cdot\vec{n}\rangle_{\partial K} \qquad \text{for all } \vec{r}\in R_d(K), \tag{3.5a}$$

$$-(w, \nabla\cdot\vec{\mathcal{Q}}^{\rm RT}\mu)_K = 0 \qquad \text{for all } w\in P_d(K). \tag{3.5b}$$

The remaining solution operators $\vec{\mathcal{Q}}_V^{\rm RT}, \mathcal{U}_V^{\rm RT}, \vec{\mathcal{Q}}_W^{\rm RT}, \mathcal{U}_W^{\rm RT}$ are defined similarly, with the obvious modification of the right hand side. It is instructive to compare Theorem 3.1 with a similar result for the RT operators, as proved in (Cockburn & Gopalakrishnan, 2005, Lemma 3.3). For instance, one pair of inequalities of (Cockburn & Gopalakrishnan, 2005, Lemma 3.3) is

$$\|\vec{\mathcal{Q}}^{\rm RT}\mu\|_K \leqslant C h_K^{-1}\|\mu\|_{h,K}, \qquad \|\mathcal{U}^{\rm RT}\mu\|_K \leqslant C\|\mu\|_{h,K}, \tag{3.6}$$

which is comparable to (3.2), when τ is of unit size. More interestingly, cf. (3.3) with

$$\|\vec{\mathcal{Q}}_W^{\rm RT} f\|_K \leqslant C h_K\|f\|_K, \qquad \|\mathcal{U}_W^{\rm RT} f\|_K \leqslant C h_K^2\|f\|_K,$$

which is another pair of inequalities of (Cockburn & Gopalakrishnan, 2005, Lemma 3.3). Observe that if τ is of unit size, these local RT operators are more stable than the corresponding HDG ones. Indeed, while $\mathcal{U}_W^{\rm RT} f$ damps perturbations in f by $O(h^2)$, the corresponding HDG operator, namely $\mathcal{U}_W f$, damps it by only $O(h)$ because $(d_{h\tau}^K)^2 = O(1/h)$.

We will now develop a series of intermediate results to prove Theorem 3.1 in the remainder of this subsection.

LEMMA 3.1 For all λ in M_h,

$$\|\mathcal{U}^{\rm RT}\lambda - \lambda\|_{\partial K} \leqslant C h_K^{1/2}\|\vec{\mathcal{Q}}^{\rm RT}\lambda\|_K.$$

Proof. Integrating (3.5a) by parts,

$$(c\,\vec{\mathcal{Q}}^{\rm RT}\lambda + \vec{\nabla}\mathcal{U}^{\rm RT}\lambda, \vec{r})_K = \langle \mathcal{U}^{\rm RT}\lambda - \lambda, \vec{r}\cdot\vec{n}\rangle_{\partial K}. \tag{3.7}$$

There is an $\vec{r}$ in $R_d(K)$ such that $\vec{r}\cdot\vec{n} = \mathcal{U}^{\rm RT}\lambda - \lambda$ on ∂K and $(\vec{r}, \vec{p}_{d-1})_K = 0$ for all $\vec{p}_{d-1}$ in $P_{d-1}(K)^n$ (this is obvious from the well-known degrees of freedom of the space $R_d(K)$). Additionally, by a scaling argument it is immediate that

$$\|\vec{r}\|_K \leqslant C h_K^{1/2}\|\vec{r}\cdot\vec{n}\|_{\partial K}. \tag{3.8}$$

With this $\vec{r}$ in (3.7), we obtain

$$\|\mathcal{U}^{\mathrm{RT}}\lambda-\lambda\|_{\partial K}^2=(c\,\vec{\mathcal{Q}}^{\mathrm{RT}}\lambda,\vec{r})_K$$

from which the lemma follows by Cauchy-Schwarz inequality and (3.8). □

LEMMA 3.2 If F is any face of the simplex K,

$$C\|w\|_K\leqslant h_K\|\mathcal{B}^t w\|_K+h_K^{1/2}\|w\|_F,\qquad \forall\, w\in W_h.$$

Proof. On the unit simplex $\hat{K}$, we have

$$\hat{C}\|\hat{w}\|_{\hat{K}}\leqslant \sup_{\vec{r}\in P_d(\hat{K})}\frac{|(\hat{w},\nabla\cdot\vec{r})_{\hat{K}}|}{\|\vec{r}\|_{\hat{K}}}+\|\hat{w}\|_{\hat{F}},\qquad \forall\,\hat{w}\in P_d(K), \tag{3.9}$$

for any face $\hat{F}$ of $\hat{K}$. This follows by equivalence of norms. That the right hand side indeed defines a norm can be seen as follows: divergence is a surjective map from $P_d(K)^n$ to $P_{d-1}(K)$. Hence if the supremum is zero, then $\hat{w}$ is orthogonal to $P_{d-1}(K)$, in which case $\hat{w}$ is zero once it vanishes on any face $\hat{F}$ (see (Cockburn *et al.*, 2010, Lemma A.1)). The lemma follows by mapping (3.9) to any simplex K and using standard scaling arguments. □

LEMMA 3.3 For all $\mu\in M_h$, we have

$$\|\tau(\mathcal{U}\mu-\mu)\|_{\partial K}\leqslant C\sqrt{\tau_K^*}\,\|\mathcal{U}\mu-\mu\|_{\tau,\partial K},$$

where $\|\mu\|_{\tau,\partial K}=\langle\tau\mu,\mu\rangle_{\partial K}^{1/2}$.

Proof. Let $F_{\max}$ denote a face of K where $\tau=\tau_K^{\max}$. Then, since

$$\|\tau(\mathcal{U}\mu-\mu)\|_{\partial K}\leqslant\sqrt{\tau_K^*}\,\|\mathcal{U}\mu-\mu\|_{\tau,\partial K\setminus F_{\max}}+\tau_K^{\max}\,\|\mathcal{U}\mu-\mu\|_{F_{\max}},$$

we only have to show that

$$\tau_K^{\max}\,\|\mathcal{U}\mu-\mu\|_{F_{\max}}\leqslant C\sqrt{\tau_K^*}\,\|\mathcal{U}\mu-\mu\|_{\tau,\partial K}.$$

For this, we first note that, for $d>0$, there is a unique w in $P_d(K)$ such that $(w,p)_K=0$ for all $p\in P_{d-1}(K)$ and $w=\mathcal{U}\mu-\mu$ on $F_{\max}$; note that with this choice, we do have that $\|w\|_{\partial K}\leqslant C\,\|\mathcal{U}\mu-\mu\|_{F_{\max}}$. With this test function, the second equation defining the lifting, namely (2.9b), becomes

$$\langle\tau(\mathcal{U}\mu-\mu),w\rangle_{\partial K}=0=\tau_K^{\max}\,\|\mathcal{U}\mu-\mu\|_{F_{\max}}^2+\langle\tau\,(\mathcal{U}\mu-\mu),w\rangle_{\partial K\setminus F_{\max}}.$$

Then, by the Cauchy-Schwarz inequality,

$$\begin{aligned}\tau_K^{\max}\,\|\mathcal{U}\mu-\mu\|_{F_{\max}}^2 &\leqslant\sqrt{\tau_K^*}\,\|w\|_{\partial K\setminus F_{\max}}\,\|\mathcal{U}\mu-\mu\|_{\tau,\partial K\setminus F_{\max}}\\ &\leqslant C\sqrt{\tau_K^*}\,\|\mathcal{U}\mu-\mu\|_{F_{\max}}\,\|\mathcal{U}\mu-\mu\|_{\tau,\partial K}.\end{aligned}$$

This completes the proof in the $d>0$ case. The proof in the $d=0$ case proceeds similarly setting $w=(\mathcal{U}\mu-\mu)|_{F_{\max}}$. □

LEMMA 3.4 Let μ be any function in M_h. The following statements hold:

(i) If $\mu|_{\partial K} = v|_{\partial K}$ for some $v \in P_0(K)$, then

$$\mathcal{U}\mu|_{\partial K} = \mu|_{\partial K} \qquad \text{and} \qquad \vec{\mathcal{Q}}\mu = 0.$$

(ii) If $a(x)$ is constant on K, and $\mu|_{\partial K} = v|_{\partial K}$ for some $v \in P_1(K)$, then

$$\vec{\mathcal{Q}}^{\mathrm{RT}}\mu = -a\vec{\nabla}v.$$

This also holds when the condition $\mu|_{\partial K} = v|_{\partial K}$ is replaced by $\langle \mu - v, 1\rangle_F = 0$ for all faces F of ∂K, in the $d = 0$ case.

(iii) If $d > 0$, $a(x)$ is constant on K, and $\mu|_{\partial K} = v|_{\partial K}$ for some $v \in P_1(K)$, then

$$\mathcal{U}\mu = v \qquad \text{and} \qquad \vec{\mathcal{Q}}\mu = -a\vec{\nabla}v.$$

(iv) We have the following bounds:

$$\|\mathcal{U}\mu - \mu\|_{\tau,\partial K} \leqslant C\sqrt{\tau_K^* h_K}\,\|\vec{\mathcal{Q}}^{\mathrm{RT}}\mu\|_K, \tag{3.10}$$

$$\|\vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\mathrm{RT}}\mu\|_K \leqslant C\sqrt{\tau_K^* h_K}\,\|\vec{\mathcal{Q}}^{\mathrm{RT}}\mu\|_K, \tag{3.11}$$

$$\|\mathcal{U}\mu - \mathcal{U}^{\mathrm{RT}}\mu\|_K \leqslant Ch_K\,(1 + \sqrt{\tau_K^* h_K})\,\|\vec{\mathcal{Q}}^{\mathrm{RT}}\mu\|_K. \tag{3.12}$$

(v) If J_K is the orthogonal projection onto $\{\vec{r} \in P_d(K)^n : \nabla\cdot\vec{r} = 0\}$ with respect to the inner product $(c\,\cdot,\cdot)_K$, with corresponding norm $\|\vec{r}\|_{c,K} \equiv (c\,\vec{r},\vec{r})^{1/2}$, then

$$\vec{\mathcal{Q}}^{\mathrm{RT}}\mu = J_K\vec{\mathcal{Q}}\mu. \tag{3.13}$$

In particular,

$$\|\vec{\mathcal{Q}}^{\mathrm{RT}}\mu\|_{c,K} \leqslant \|\vec{\mathcal{Q}}\mu\|_{c,K}. \tag{3.14}$$

Proof. This proof proceeds by comparing the RT and HDG equations for the local solutions. Subtracting (3.5) from (2.9) we have

$$(c\,(\vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\mathrm{RT}}\mu),\vec{r})_K - (\mathcal{U}\mu - \mathcal{U}^{\mathrm{RT}}\mu, \nabla\cdot\vec{r})_K = 0 \tag{3.15a}$$

$$(\nabla\cdot(\vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\mathrm{RT}}\mu), w)_K + \langle \tau(\mathcal{U}\mu - \mathcal{U}^{\mathrm{RT}}\mu), w\rangle_{\partial K} = \langle \tau(\mu - \mathcal{U}^{\mathrm{RT}}\mu), w\rangle_{\partial K} \tag{3.15b}$$

for all $\vec{r} \in P_d(K)^n$ and for all w in $P_d(K)$. Note that since $\nabla\cdot\vec{\mathcal{Q}}^{\mathrm{RT}}\mu = 0$, the lifting $\vec{\mathcal{Q}}^{\mathrm{RT}}\mu$ is in fact in $P_d(K)^n$. Hence $\{\vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\mathrm{RT}}\mu, \mathcal{U}\mu - \mathcal{U}^{\mathrm{RT}}\mu\}$ forms the unique solution of (3.15).

First, let us prove the first assertion (i) of the lemma. Indeed, if μ takes a constant value on ∂K, then it is well known that $\mathcal{U}^{\mathrm{RT}}\mu$ equals the same constant (Gopalakrishnan, 2003, Lemma 2.1) and $\vec{\mathcal{Q}}^{\mathrm{RT}}\mu = 0$, so the right hand side of (3.15) vanishes. Hence $\mathcal{U}^{\mathrm{RT}}\mu - \mathcal{U}\mu$ and $\vec{\mathcal{Q}}^{\mathrm{RT}}\mu - \vec{\mathcal{Q}}\mu$ also vanish, thus proving (i).

The argument to prove the next statement (ii) is essentially contained in (Gopalakrishnan & Tan, 2009, Lemma 4.2), but we give it here for completeness. Equation (3.5a) implies

$$(c\,\vec{\mathcal{Q}}^{\mathrm{RT}}\mu, J_K\vec{r})_K = -\langle \mu, J_K\vec{r}\cdot\vec{n}\rangle_{\partial K} = -(\vec{\nabla}v, J_K\vec{r})_K.$$

Since (3.5b) implies that $\vec{\mathcal{Q}}^{\mathrm{RT}}\mu$ is in the range of J_K, and since $\vec{\nabla}v$ is obviously in the range of J_K, we have proved that $\vec{\mathcal{Q}}^{\mathrm{RT}}\mu = -(c^{-1})\,\vec{\nabla}v$.

The statement (iii) is proved by the same technique as (i). The only difference is that the analogous result for the RT case is less well known, so let us first show it, namely $\mathcal{U}^{\rm RT}\mu|_{\partial K} = \mu|_{\partial K}$ when $d > 0$ and $\mu|_{\partial K}$ equals the trace of some $v \in P_1(K)$. In light of (ii), equation (3.5a) becomes

$$-(\vec{\nabla} v, \vec{r})_K - (\mathcal{U}^{\rm RT}\mu, \nabla\cdot\vec{r})_K = -\langle v, \vec{r}\cdot\vec{n}\rangle_{\partial K} = -(\vec{\nabla} v, \vec{r})_K - (v, \nabla\cdot\vec{r})_K.$$

This implies that

$$(\mathcal{U}^{\rm RT}\mu - v, \nabla\cdot\vec{r})_K = 0 \qquad \forall \vec{r} \in R_d(K),$$

so that $\mathcal{U}^{\rm RT}\mu = v$. Thus, just as in the proof of item (i), the solution of (3.15) vanishes in this case also, and we have proven item (iii).

Next, let us prove the estimates. Setting $\vec{r} = \vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\rm RT}\mu$ and $w = \mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu$, we have

$$\|\vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\rm RT}\mu\|^2_{c,K} + \|\mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu\|^2_{\tau,\partial K} = \langle \tau(\mu - \mathcal{U}^{\rm RT}\mu), \mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu\rangle_{\partial K},$$

or, equivalently,

$$\|\vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\rm RT}\mu\|^2_{c,K} + \|\mathcal{U}\mu - \mu\|^2_{\tau,\partial K} = \langle \tau(\mathcal{U}\mu - \mu), \mathcal{U}^{\rm RT}\mu - \mu\rangle_{\partial K}.$$

Using the Cauchy-Schwarz inequality, we obtain

$$\begin{aligned}\|\vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\rm RT}\mu\|^2_{c,K} + \|\mathcal{U}\mu - \mu\|^2_{\tau,\partial K} &\leqslant \|\tau(\mathcal{U}\mu - \mu)\|_{\partial K}\|\mathcal{U}^{\rm RT}\mu - \mu\|_{\partial K}\\ &\leqslant C\sqrt{\tau_K^* h_K}\,\|\mathcal{U}\mu - \mu\|_{\tau,\partial K}\|\vec{\mathcal{Q}}^{\rm RT}\mu\|_{\partial K},\end{aligned}$$

by Lemmas 3.3 and 3.1. The estimates (3.10), and (3.11) immediately follow.

It remains only to prove (3.12). Let $F_{\max}$ denote a face of K where $\tau = \tau_K^{\max}$. Then

$$\tau_K^{\max}\|\mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu\|^2_{F_{\max}} = \|\mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu\|^2_{\tau,F_{\max}} \leqslant \|\mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu\|^2_{\tau,\partial K} \leqslant C\tau_K^{\max} h_K\|\vec{\mathcal{Q}}^{\rm RT}\mu\|^2_K,$$

so canceling off the common factor $\tau_K^{\max}$, we have

$$\|\mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu\|_{F_{\max}} \leqslant Ch_K^{1/2}\|\vec{\mathcal{Q}}^{\rm RT}\mu\|_K.$$

Hence using Lemma 3.2, we obtain

$$\begin{aligned}\|\mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu\|_K &\leqslant C\left(h_K\|\mathcal{B}^t(\mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu)\|_K + h_K^{1/2}\|\mathcal{U}\mu - \mathcal{U}^{\rm RT}\mu\|_{F_{\max}}\right)\\ &\leqslant C\left(h_K\|\vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\rm RT}\mu\|_K + h_K\|\vec{\mathcal{Q}}^{\rm RT}\mu\|_K\right),\end{aligned}$$

from which (3.12) follows. (This applies even if $d = 0$, in which case the term involving $\mathcal{B}^t$ is absent.) Thus we have proved item (iv).

For the final item (v),

$$\begin{aligned}(c\,J_K\vec{\mathcal{Q}}\mu, J_K\vec{r})_K &= -\langle \mu, J_K\vec{r}\cdot\vec{n}\rangle_{\partial K} && \text{by (2.9a)}\\ &= (c\,\vec{\mathcal{Q}}^{\rm RT}\mu, J_K\vec{r})_K, && \text{by (3.5a)},\end{aligned}$$

which proves the equality (3.13), as $\vec{\mathcal{Q}}^{\rm RT}\mu$ is in the range of J_K by (3.5b). The estimate (3.14) is then obvious as orthogonal projectors have unit norm. □

Proof of Theorem 3.1. First, we prove the bounds on $\vec{\mathcal{Q}}\mu, \mathcal{U}\mu$:

$$
\begin{aligned}
\|\vec{\mathcal{Q}}\mu\|_K &\leqslant \|\vec{\mathcal{Q}}^{\mathrm{RT}}\mu\|_K + \|\vec{\mathcal{Q}}\mu - \vec{\mathcal{Q}}^{\mathrm{RT}}\mu\|_K \\
&\leqslant \|\vec{\mathcal{Q}}^{\mathrm{RT}}\mu\|_K + C(\tau_K^* h_K)^{1/2}\|\vec{\mathcal{Q}}^{\mathrm{RT}}\mu\|_K && \text{by (3.11) of Lemma (3.4),} \\
&\leqslant c_{h\tau}^K C\|\mu\|_{h,K} && \text{by (3.6).}
\end{aligned}
$$

The bound for $\mathcal{U}\mu$ is proved similarly using (3.12) in place of (3.11).

Next, consider $\vec{\mathcal{Q}}_W f, \mathcal{U}_W f$. From their definitions, it is easy to see that

$$
(c\,\vec{\mathcal{Q}}_W f, \vec{\mathcal{Q}}_W f)_K + \langle \tau \mathcal{U}_W f, \mathcal{U}_W f\rangle_{\partial K} = -(\mathcal{U}_W f, f)_K. \tag{3.16}
$$

Let $F_{\max}$ denote a face of K where $\tau = \tau_K^{\max}$. Then,

$$
\begin{aligned}
\|\mathcal{U}_W f\|_K &\leqslant C(h_K\|\mathcal{B}^t\mathcal{U}_W f\|_K + h_K^{1/2}\|\mathcal{U}_W f\|_{F_{\max}}) && \text{by Lemma 3.2,} \\
&\leqslant C(h_K\|\vec{\mathcal{Q}}_W f\|_K + h_K^{1/2}\|\mathcal{U}_W f\|_{F_{\max}}) && \text{as } \mathcal{A}\vec{\mathcal{Q}}_W f + \mathcal{B}^t\mathcal{U}_W f = 0, \\
&\leqslant Ch_K(\|\vec{\mathcal{Q}}_W f\|_K + (\tau_K^{\max}h_K)^{-1/2}\|\mathcal{U}_W f\|_{\tau,\partial K}) \\
&\leqslant Ch_K((\mathcal{U}_W f, f)_K^{1/2} + (\tau_K^{\max}h_K)^{-1/2}(\mathcal{U}_W f, f)_K^{1/2}) && \text{by (3.16),} \\
&\leqslant Ch_K(1 + (\tau_K^{\max}h_K)^{-1/2})\|\mathcal{U}_W f\|_K^{1/2}\|f\|_K^{1/2},
\end{aligned}
$$

from which required bound on $\mathcal{U}_W f$ follows. Using this in (3.16), we immediately get the stated bound for $\vec{\mathcal{Q}}_W f$ as well.

Finally, to prove (3.4), we start from the following easy consequence of the definitions of $\vec{\mathcal{Q}}_V \vec{g}, \mathcal{U}_V \vec{g}$:

$$
\|\vec{\mathcal{Q}}_V\vec{g}\|_{c,K}^2 + \|\mathcal{U}_V\vec{g}\|_{\tau,\partial K}^2 = (\vec{g}, \vec{\mathcal{Q}}_V\vec{g})_K. \tag{3.17}
$$

Since it is immediate from the above that $\|\vec{\mathcal{Q}}_V\vec{g}\|_K \leqslant C\|\vec{g}\|_K$, it only remains to prove the bound for $\mathcal{U}_V\vec{g}$. By Lemma 3.2,

$$
\begin{aligned}
\|\mathcal{U}_V\vec{g}\|_K &\leqslant Ch_K\|\mathcal{B}^t\mathcal{U}_V\vec{g}\|_K + Ch_K^{1/2}\|\mathcal{U}_V\vec{g}\|_{F_{\max}} \\
&\leqslant Ch_K(\|\mathcal{A}\vec{\mathcal{Q}}_V\vec{g}\|_K + \|\vec{g}\|_K) + Ch_K(\tau_K^{\max}h_K)^{-1/2}\|\vec{\mathcal{Q}}_V\vec{g}\|_{\tau,\partial K},
\end{aligned}
$$

and the final inequality of the theorem follows by using (3.17) in the above. □

3.2 *Conditioning of the HDG method*

We now obtain bounds on the spectrum of the operator generated by $a_h(\cdot,\cdot)$. The main result of this subsection is the following.

THEOREM 3.2 Suppose $\mathscr{T}_h$ is quasiuniform and $h = \max\{h_K : K \in \mathscr{T}_h\}$. There are positive constants C_1 and C_2 independent of h such that

$$
C_1\|\mu\|_h^2 \leqslant a_h(\mu,\mu) \leqslant \gamma_{h\tau}^{(2)} C_2 h^{-2}\|\mu\|_h^2, \quad \text{for all } \mu \in M_h \tag{3.18}
$$

where $\gamma_{h\tau}^{(2)} = \max\{1 + (\tau_K^* h_K)^2 : K \in \mathscr{T}_h\}$.

Note that this result holds for $\tau_K^* = 0$ on $\partial\mathscr{T}_h$ which is the choice of the stabilization function τ that characterizes the SFH method (Cockburn *et al.*, 2008). For that specific HDG method, we thus see that the condition number is independent of the value of $\tau_K^{\max}$. This is not surprising since in (Cockburn *et al.*, 2008) it was proven that the matrix for the SFH method is *identical* to that of the hybridized RT and the hybridized BDM (if $d \geqslant 1$) methods of corresponding degrees. As a consequence, our multigrid results apply to those two methods as well.

The implication of this theorem for a condition number bound is as follows. Consider the stiffness matrix of $a_h(\cdot,\cdot)$, obtained through any standard local (face by face) finite element basis for M_h. Let κ be the spectral condition number of this stiffness matrix. Then standard arguments using the two-sided estimate of Theorem 3.2 imply

$$\kappa \leqslant \gamma_{h\tau}^{(2)} C h^{-2}.$$

In particular, note that for all choices of τ satisfying $h\tau \leqslant C$, the condition number grows like $O(h^{-2})$. (For the so-called "super-penalized" cases where τ is chosen to be $O(1/h^{\alpha})$ with $\alpha > 1$, it grows even faster.) The growth of the condition number implies a deterioration in the performance of many iterative techniques as h decreases. This motivates our development of efficient multigrid algorithms (in Section 4) that converge at an h-independent rate.

The proof of Theorem 3.2 relies on the two lemmas below. To state them, we need to introduce an additional norm, defined by

$$m_K(\lambda) = \frac{1}{|\partial K|}\int_{\partial K} \lambda \, ds, \tag{3.19}$$

$$|||\lambda|||_{h,D} = \left(\sum_{K\in\mathscr{T}_h,\, K\subseteq \bar{D}} \|\lambda - m_K(\lambda)\|^2_{L^2(\partial K)} \frac{1}{h_K} \right)^{1/2},$$

and $|||\cdot|||_h = |||\cdot|||_{h,\Omega}$. Note that when D is strictly contained in Ω, $|||\lambda|||_{h,D}$ is a semi-norm. However, when $D = \Omega$, since $\|\cdot\|_h$ is an L^2-like norm, $|||\cdot|||_h$ is an H^1-like norm, and since functions in M_h can be thought of as having zero boundary conditions on $\partial\Omega$, it is not surprising that the following Poincaré - type inequality holds:

LEMMA 3.5 There is a constant C_0 such that on all quasiuniform meshes

$$\|\mu\|_h \leqslant C_0 |||\mu|||_h \qquad \text{for all } \mu \in M_h. \tag{3.20}$$

Proof. See (Gopalakrishnan, 2003, Proof of Theorem 2.3). □

LEMMA 3.6 Let $\vec{\mathbb{Q}}(\cdot)$ denote the HDG flux lifting operator defined in (2.9). Then

$$C|||\mu|||_{h,K} \leqslant \|\vec{\mathbb{Q}}\mu\|_K \tag{3.21}$$

for all μ in M_h and all mesh elements K.

Proof. If we use the inequality $|||\lambda|||_{h,K} \leqslant C\|\vec{\mathbb{Q}}^{\mathrm{RT}}\lambda\|_K$ established in (Gopalakrishnan, 2003), the proof of lemma can be completed instantly by

$$|||\lambda|||_{h,K} \leqslant C\|\vec{\mathbb{Q}}^{\mathrm{RT}}\lambda\|_K \leqslant C\|\vec{\mathbb{Q}}\lambda\|_K$$

where we have used (3.14) of Lemma 3.4. However, to give a better idea of how the $|||\cdot|||_{h,K}$-norm enters the arena, we give a more direct proof below.

Let T_K be the affine isomorphism mapping the reference unit simplex $\hat{K}$ one-to-one onto K. It has the form $T_K(\hat{x}) = M_K\hat{x} + b$ for some $n \times n$ matrix M_K. We will also need the Piola map Φ_K mapping functions on K to $\hat{K}$, defined by $\Phi_K(\vec{r}) = (\det M_K)^{-1} M_K^{-1}\vec{r} \circ T_K$. We start by letting $\hat{\lambda} = \lambda \circ T_K$ and recalling that there is a function $\vec{r}_{\hat{\lambda}}$ in $P_d(K)^n$ such that

$$\begin{aligned} \nabla \cdot \vec{r}_{\hat{\lambda}} &= 0, && \text{on } \hat{K}, \\ \vec{r}_{\hat{\lambda}} \cdot \vec{n} &= \hat{\lambda} - m_{\hat{K}}(\hat{\lambda}), && \text{on } \partial\hat{K}, \text{ and} \\ \|\vec{r}_{\hat{\lambda}}\|_{\hat{K}} &\leqslant C\|\hat{\lambda} - m_{\hat{K}}(\hat{\lambda})\|_{\partial\hat{K}}. \end{aligned}$$

Such an $\vec{r}_{\hat{\lambda}}$ can be obtained, e.g., by the polynomial extension in (Demkowicz *et al.*, 2012) applied to $\hat{\lambda} - m_{\hat{K}}(\hat{\lambda})$, or even by more elementary observations. Next let $\vec{v}_\lambda = \Phi_K^{-1}(\vec{r}_{\hat{\lambda}})$. By the well known properties of the Piola map (Brezzi & Fortin, 1991), we know that

$$\nabla \cdot \vec{v}_\lambda = 0 \quad \text{and} \quad \langle \lambda, \vec{v}_\lambda \cdot \vec{n} \rangle_{\partial K} = \langle \hat{\lambda}, \vec{r}_{\hat{\lambda}} \cdot \vec{n} \rangle_{\partial\hat{K}}.$$

Setting $\vec{r}$ equal to $\vec{v}_\lambda$ in (2.9a), we get

$$\begin{aligned} (c\vec{\mathcal{Q}}\lambda, \vec{v}_\lambda)_K &= -\langle \lambda, \vec{v}_\lambda \cdot \vec{n} \rangle_{\partial K} = -\langle \hat{\lambda}, \vec{r}_{\hat{\lambda}} \cdot \vec{n} \rangle_{\partial\hat{K}} \\ &= -\langle \hat{\lambda} - m_{\hat{K}}(\hat{\lambda}), \vec{r}_{\hat{\lambda}} \cdot \vec{n} \rangle_{\partial\hat{K}} \\ &= -\|\hat{\lambda} - m_{\hat{K}}(\hat{\lambda})\|^2_{\partial\hat{K}}. \end{aligned}$$

This implies

$$\|\hat{\lambda} - m_{\hat{K}}(\hat{\lambda})\|^2_{\partial\hat{K}} \leqslant \|\vec{\mathcal{Q}}\lambda\|_K \|\vec{v}_\lambda\|_K \leqslant C\|\vec{\mathcal{Q}}\lambda\|_K \|\vec{r}_{\hat{\lambda}}\|_{\hat{K}} \leqslant C\|\vec{\mathcal{Q}}\lambda\|_K \|\hat{\lambda} - m_{\hat{K}}(\hat{\lambda})\|_{\partial\hat{K}},$$

so

$$\|\hat{\lambda} - m_{\hat{K}}(\hat{\lambda})\|_{\partial\hat{K}} \leqslant C\|\vec{\mathcal{Q}}\lambda\|_K.$$

Using the fact that $m_K(\lambda)$ is the best approximating constant on ∂K to λ, and using a scaling argument,

$$\|\lambda - m_K(\lambda)\|_{\partial K} \leqslant \|\lambda - m_{\hat{K}}(\hat{\lambda})\|_{\partial K} \leqslant C h_K^{1/2} \|\hat{\lambda} - m_{\hat{K}}(\hat{\lambda})\|_{\partial\hat{K}}.$$

Therefore,

$$|||\lambda|||_{h,K} = C h_K^{-1/2} \|\lambda - m_K(\lambda)\|_{\partial K} \leqslant C\|\vec{\mathcal{Q}}\lambda\|_K$$

and the lemma is proved. □

Proof of Theorem 3.2. For the upper bound, we use (3.10) and (3.14) of Lemma 3.4 to conclude that

$$\|\mathcal{U}\lambda - \lambda\|^2_{\tau,\partial K} \leqslant C\tau_K^* h_K \|\vec{\mathcal{Q}}^{\mathrm{RT}}\lambda\|^2_K \leqslant C\tau_K^* h_K \|\vec{\mathcal{Q}}\lambda\|^2_K.$$

Hence, summing over all elements, and denoting $\|\cdot\|_\tau = \langle \tau\cdot, \cdot \rangle^{1/2}_{\partial\mathcal{T}_h}$,

$$\begin{aligned} a_h(\lambda,\lambda) = \|\vec{\mathcal{Q}}\lambda\|^2_c + \|\mathcal{U}\lambda - \lambda\|^2_\tau &\leqslant C \sum_{K \in \mathcal{T}_h} (1 + \tau_K^* h_K) \|\vec{\mathcal{Q}}\lambda\|^2_K \\ &\leqslant C \sum_{K \in \mathcal{T}_h} (1 + \tau_K^* h_K)(c^K_{h\tau})^2 h_K^{-2} \|\lambda\|^2_h, \end{aligned}$$

where we have used Theorem 3.1. Thus, the upper bound follows.

For the lower bound, we combine the estimates of Lemmas 3.5 and 3.6 to obtain

$$\|\lambda\|^2_h \leqslant C_0 |||\lambda|||^2_h \leqslant C_0 C \|\vec{\mathcal{Q}}\lambda\|^2_{\mathcal{T}_h} \leqslant C C_0\, a_h(\lambda,\lambda),$$

so the proof is complete. □

3.3 *Remarks on preconditioning*

The increase in condition number as $h \to 0$, as given by Theorem 3.2, shows the importance of designing efficient solution strategies. One way to do this is by constructing preconditioners suitable for use in nonlinear iterative solvers like the conjugate gradient method. Let us note a simple consequence of our previous results that has implications in preconditioning the HDG matrix.

THEOREM 3.3 For all $\lambda \in M_h$,

$$C_3 |||\lambda|||_h^2 \leqslant a_h(\lambda,\lambda) \leqslant C_4 \gamma_{h\tau}^{(1)} |||\lambda|||_h^2.$$

where $\gamma_{h\tau}^{(1)} = \max\{1 + \tau_K^* h_K : K \in \mathscr{T}_h\}$.

Proof. The first inequality follows from Lemma 3.6, so it only remains to prove the upper bound. For this, note that

$$\begin{aligned} a_h(\lambda,\lambda) &= \|\vec{\mathcal{Q}}\lambda\|^2 + \|\mathcal{U}\lambda - \lambda\|_\tau^2 \\ &\leqslant 2\|\vec{\mathcal{Q}}\lambda - \vec{\mathcal{Q}}^{\mathrm{RT}}\lambda\|^2 + 2\|\vec{\mathcal{Q}}^{\mathrm{RT}}\lambda\|^2 + C\gamma_{h\tau}^{(1)}\|\vec{\mathcal{Q}}^{\mathrm{RT}}\lambda\|^2 && \text{by (3.10)} \\ &\leqslant C\gamma_{h\tau}^{(1)}\|\vec{\mathcal{Q}}^{\mathrm{RT}}\lambda\|^2 && \text{by (3.11).} \end{aligned}$$

Hence the upper bound follows from the inequality

$$\|\vec{\mathcal{Q}}^{\mathrm{RT}}\lambda\| \leqslant C|||\lambda|||_h$$

proved in (Gopalakrishnan, 2003). □

It is proved in (Cockburn & Gopalakrishnan, 2005; Gopalakrishnan, 2003) that the norms $|||\cdot|||_h$ and $a_h^{\mathrm{RT}}(\cdot,\cdot)^{1/2}$ are equivalent. Therefore, by Theorem 3.3, we find that a spectrally equivalent preconditioner B for the hybridized mixed method (for the form $a_h^{\mathrm{RT}}(\cdot,\cdot)$) will also yield a preconditioner for the HDG method's form $a_h(\cdot,\cdot)$. In particular, the Schwarz preconditioner in (Gopalakrishnan, 2003) or the multigrid preconditioner in (Gopalakrishnan & Tan, 2009), both originally intended for the HRT method, could be used for preconditioning the HDG method. In the next section we give a less expensive linear iterative solver that directly uses the HDG bilinear form and is more effective in practice.

3.4 *Error estimates for the HDG method*

Error estimates for the HDG method under consideration have been proved in (Cockburn *et al.*, 2010). Here, as an application of the estimates we proved in § 3.1, we prove two new error estimates not in (Cockburn *et al.*, 2010). We need the orthogonal projection into M_h defined by

$$\langle P_h^M u, \mu\rangle_{\partial\mathscr{T}_h} = \langle u, \mu\rangle_{\partial\mathscr{T}_h}, \quad \text{for all } \mu \in M_h. \tag{3.22}$$

We also need the special projection of (Cockburn *et al.*, 2010). This projection, denoted by $\Pi_h(\vec{q},u)$, is into the product space $V_h \times W_h$, and its domain is a subspace of $H(\mathrm{div},\Omega) \times L^2(\Omega)$ consisting of sufficiently regular functions, e.g., $H(\mathrm{div},\Omega) \cap H^s(\Omega)^n \times H^s(\Omega)$ for $s > 1/2$. When its components need to be identified, we also write $\Pi_h(\vec{q},u)$ as $(\Pi_h^V\vec{q}, \Pi_h^W u)$ where $\Pi_h^V\vec{q}$ and $\Pi_h^W u$ are the components of the projection in V_h and W_h, respectively. (Despite this notation, note that $\Pi_h^V\vec{q}$ depends not just on $\vec{q}$, but rather on both $\vec{q}$ and u. The same applies for $\Pi_h^W u$.) We omit the definition (Cockburn *et al.*, 2010, eq. (2.1)) and other details of the projection, but let us recall the following properties we need:

THEOREM 3.4 Let $s_u, s_q \in (1/2, d+1]$ and let $(\vec{q}, u) \in H(\text{div}, \Omega) \cap H^{s_q}(\Omega)^n \times H^{s_u}(\Omega)$. Then we have

$$\|\Pi_h^V \vec{q} - \vec{q}\|_K \leqslant C h_K^{s_q} |\vec{q}|_{H^{s_q}(K)} + C h_K^{s_u} \tau_K^* |u|_{H^{s_u}(K)} \tag{3.23a}$$

$$\|\Pi_h^W u - u\|_K \leqslant C h_K^{s_u} |u|_{H^{s_u}(K)} + C \frac{h_K^{s_q}}{\tau_K^{\max}} |\vec{q}|_{H^{s_q}(K)}, \tag{3.23b}$$

Let us recall that, see Theorem 3.1, $\tau_K^* := \max \tau|_{\partial K \setminus F_{\max}}$, where $F_{\max}$ is a face of K at which $\tau|_{\partial K}$ is maximum. Moreover, letting $\vec{\varepsilon}_h^{\,q} = \Pi_h^V \vec{q} - \vec{q}_h$, $\varepsilon_h^u = \Pi_h^W u - u_h$, and $\varepsilon_h^\lambda = P_h^M u - \lambda_h$, the identity

$$\begin{pmatrix} \mathcal{A} & \mathcal{B}^t & \mathcal{C}^t \\ \mathcal{B} & \mathcal{R} & \mathcal{S}^t \\ \mathcal{C} & \mathcal{S} & \mathcal{T} \end{pmatrix} \begin{pmatrix} \vec{\varepsilon}_h^{\,q} \\ \varepsilon_h^u \\ \varepsilon_h^\lambda \end{pmatrix} = \begin{pmatrix} \vec{e}_h \\ 0 \\ 0 \end{pmatrix}, \tag{3.24}$$

holds, where $\vec{e}_h$ is the unique function in V_h satisfying $(\vec{e}_h, \vec{r})_{\mathcal{T}_h} = (c\,(\Pi_h^V \vec{q} - \vec{q}), \vec{r})_{\mathcal{T}_h}$.

See Appendix A for a proof and references. When the approximation property (3.23a) is combined with the following theorem, we get optimal estimates for all variables of the HDG method. Let $\|\mu\|_a = a_h(\mu, \mu)^{1/2}$ and $\|\vec{r}\|_c = (c\,\vec{r}, \vec{r})^{1/2}$.

THEOREM 3.5 Let the exact solution satisfying (2.1) be $(\vec{q}, u)$, and the discrete solution satisfying (2.5) be $(\vec{q}_h, u_h, \lambda_h)$. Then, the following error estimates hold:

$$\|\vec{q} - \vec{q}_h\|_c \leqslant 2\|\vec{q} - \Pi_h^V \vec{q}\|_c, \tag{3.25}$$

$$\|P_h^M u - \lambda_h\|_a \leqslant \|\vec{q} - \Pi_h^V \vec{q}\|_c, \tag{3.26}$$

$$\|u - u_h\|_{\mathcal{T}_h} \leqslant C\|u - \Pi_h^W u\| + b_\tau C\|\vec{q} - \Pi_h^V \vec{q}\|_c, \tag{3.27}$$

where $b_\tau = \max\{1 + h_K \tau_K^* + h_K/\tau_K^{\max} : K \in \mathcal{T}_h\}$.

Proof. The first estimate is easy and is proved in (Cockburn *et al.*, 2010), so we only prove the remaining two.

To prove (3.26), we apply Theorem 2.1 to (3.24). Then, we find that ε_h^λ satisfies

$$a_h(\varepsilon_h^\lambda, \mu) = (-\vec{e}_h, \vec{\mathcal{Q}}\mu)_{\mathcal{T}_h} = (c\,(\Pi_h^V \vec{q} - \vec{q}), \vec{\mathcal{Q}}\mu)_{\mathcal{T}_h}$$

for all μ in M_h. Hence (3.26) follows by choosing $\mu = \varepsilon_h^\lambda$ and applying the Cauchy-Schwarz inequality.

To prove (3.27), we apply the local recovery equation (2.12) of Theorem 2.1 to (3.24), which gives $\varepsilon_h^u = \mathcal{U}\varepsilon_h^\lambda + \mathcal{U}_V \vec{e}_h$. Therefore,

$$\|\varepsilon_h^u\|_K \leqslant \|\mathcal{U}\varepsilon_h^\lambda\|_K + \|\mathcal{U}_V \vec{e}_h\|_K \leqslant c_{h\tau}^K C\|\varepsilon_h^\lambda\|_{h,K} + C d_{h\tau}^K h_K \|\vec{e}_h\|_K,$$

where we have used Theorem 3.1. Since $d_{h\tau}^K h_K \leqslant C h_K^{1/2} (\tau_K^{\max})^{-1/2}$,

$$\|u - u_h\|_K^2 \leqslant C(c_{h\tau}^K)^2 \|\varepsilon_h^\lambda\|_{h,K}^2 + C(\tau_K^{\max})^{-1} h_K \|\Pi_h^V \vec{q} - \vec{q}\|_K^2 + C\|u - \Pi_h^W u\|_K^2.$$

Summing over all mesh elements and using Theorem 3.2, we obtain

$$\|u - u_h\|_{\mathcal{T}_h}^2 \leqslant b'_\tau C\|\varepsilon_h^\lambda\|_a^2 + C b''_\tau \|\Pi_h^V \vec{q} - \vec{q}\|_{\mathcal{T}_h}^2 + C\|u - \Pi_h^W u\|_{\mathcal{T}_h}^2.$$

where $b'_\tau = \max\{1 + \tau_K^* h_K : K \in \mathcal{T}_h\}$ and $b''_\tau = \max\{h_K/\tau_K^{\max} : K \in \mathcal{T}_h\}$. Thus we can finish the proof of (3.27) using the previous estimate (3.26) for ε_h^λ. $\square$

REMARK 3.1 Stronger error estimates for u_h and λ_h are established in (Cockburn *et al.*, 2010) under additional regularity assumptions. The only regularity requirement for the estimates (3.27) and (3.26) to hold is that $(\vec{q}, u)$ is in the domain of Π_h, whereas the analysis in (Cockburn *et al.*, 2010) assumes in addition the full regularity condition needed for an Aubin-Nitsche type argument.

4. A multigrid algorithm

In this section we discuss some ways of applying multigrid techniques to efficiently solve matrix systems arising from methods like the HDG method. We consider an abstract sequence of two spaces and a general nonnested two-level algorithm on these spaces. Fitting the HDG application into this abstract setup is the purpose of the next section (so we emphasize that the generic forms and spaces in this section need not be those from the HDG method). We give a linear two level iteration for which we can prove convergence independent of mesh size. The abstract multigrid theorem we shall state here is an adaptation of the well-known results of (Bramble *et al.*, 1991; Xu, 1990).

4.1 *The non-nested two-level V-cycle*

Let M_1 and M_0 be two given Hilbert spaces. Suppose we want to solve for μ in a space M_1 satisfying

$$a_1(\mu, \eta) = (b, \eta)_1 \qquad \forall \eta \in M_1.$$

Here $b \in M_1$ is given and $(\cdot,\cdot)_1$ and $a_1(\cdot,\cdot)$ are two inner products in M_1. We want to construct an optimally convergent linear iteration of the form

$$\mu_{\ell+1} = \mu_\ell + B_1(b - A_1\mu_\ell), \qquad \ell = 1, 2, \ldots. \tag{4.1}$$

The iteration is started with any $\mu_0 \in M_1$ and the operator $A_1 : M_1 \to M_1$ is defined by

$$(A_1\eta, \mu)_1 = a_1(\eta, \mu) \qquad \forall \eta, \mu \in M_1.$$

The operator $B_1 : M_1 \to M_1$ needs to be suitably defined to achieve fast convergence. The idea is to use a 'nearby' problem for which optimal solvers are already known. (The same idea has been pursued in different directions by other researchers (Brenner, 1999; Xu, 1996).) This forms the "0th level", while the original problem forms "level 1" in the two-level algorithm we give below. The nearby problem uses inner products $a_0(\cdot,\cdot)$ and $(\cdot,\cdot)_0$ on another space M_0. Let $A_0 : M_0 \to M_0$ be defined by

$$(A_0 v, w)_0 = a_0(v, w) \qquad \forall v, w \in M_0.$$

That a good solver is available for the nearby problem is implied by the next assumption.

Assumption 4.1 We assume that there is a number $0 \leqslant \delta_0 < 1$, and an operator $B_0 : M_0 \to M_0$ that is self-adjoint in the $(\cdot,\cdot)_0$-inner product, such that

$$0 \leqslant a_0(v - B_0 A_0 v, v) \leqslant \delta_0\, a_0(v, v), \qquad \forall v \in M_0,$$

We construct the operator B_1 appearing in (4.1) using B_0 and two other ingredients. The first is a smoothing operator $R_1 : M_1 \to M_1$. The second is a grid transfer operator $I_1 : M_0 \to M_1$ that maps data between discretizations. Note that the spaces M_0 and M_1 are not assumed to be nested, i.e., $M_0 \not\subseteq M_1$ in general. (Specific examples of B_0, R_1, and I_1, will be given in § 4.2.) Define $Q_0 : M_1 \to M_0$ by

$$(Q_0\mu, w)_0 = (\mu, I_1 w)_1 \qquad \forall \mu \in M_1 \text{ and } w \in M_0.$$

Let R_1^t denote the adjoint of R_1 in the $(\cdot,\cdot)_1$-inner product. With these notations, the operator B_1 is defined below.

Algorithm 4.1 (2-level V-cycle) For any g in M_1 define $B_1 g$ by the following steps:

1. Smooth: $v_1 = R_1 g$.
2. Correct: $v_2 = v_1 + I_1 B_0 Q_0 (g - A_1 v_1)$.
3. Smooth: $v_3 = v_2 + R_1^t (g - A_1 v_2)$.

Set $B_1 g \equiv v_3$.

Now we describe a few conditions, taken from (Bramble *et al.*, 1991), under which one can prove optimal convergence of (4.1).

Assumption 4.2 For all v in M_0

$$a_1(I_1 v, I_1 v) \leqslant a_0(v,v).$$

Verifications of all assumptions listed here for the HDG application appear in the next section. In the lowest order case of the HDG method, we are not able to verify Assumption 4.2. Instead, as we shall see in Section 5, we can only verify the following.

Assumption 4.3 There is a constant $C_0 > 0$ such that for all v in M_0

$$a_1(I_1 v, I_1 v) \leqslant (1 + C_0 h_1) a_0(v,v).$$

Here h_1 is a mesh size parameter associated with M_1. The next assumption involves an operator $P_0 : M_1 \to M_0$ defined by

$$a_0(P_0\mu, v) = a_1(\mu, I_1 v), \qquad \forall \mu \in M_1 \text{ and } v \in M_0.$$

Assumption 4.4 There are constants $C_1 > 0$ and $0 < \alpha \leqslant 1$ such that for all $\mu \in M_1$,

$$a_1(\mu - I_1 P_0 \mu, \mu) \leqslant C_1 \left(\frac{\|A_1\mu\|_1^2}{\rho(A_1)} \right)^{\alpha} a_1(\mu,\mu)^{1-\alpha}.$$

where $\rho(A_1)$ is the spectral radius of A_1.

Here and elsewhere, the norms $\|\cdot\|_1$ and $\|\cdot\|_0$ are generated by the inner products $(\cdot,\cdot)_1$ and $(\cdot,\cdot)_0$, respectively. The final assumption is on the smoother.

Assumption 4.5 There is a number $\omega > 0$ such that

$$\omega \frac{\|\mu\|_1^2}{\rho(A_1)} \leqslant (\widetilde{R}_1 \mu, \mu)_1, \qquad \forall \mu \in M_1,$$

where $\widetilde{R}_1 = R_1 + R_1^t - R_1 A_1 R_1^t$.

We then have the following theorems.

THEOREM 4.2 If Assumptions 4.1, 4.2, 4.4 and 4.5 hold, then there is a constant $0 \leqslant \delta_1 < 1$ depending only on δ_0, C_1, ω, and α such that the iterates of (4.1) satisfy

$$\|\mu - \mu_\ell\|_{a_1} \leqslant \delta_1^\ell \|\mu - \mu_0\|_{a_1}.$$

Here $\|\cdot\|_{a_1} = a_1(\cdot,\cdot)^{1/2}$ and $\mu = A_1^{-1} b$ is the exact solution.

THEOREM 4.3 If Assumptions 4.1, 4.3, 4.4 and 4.5 hold, then there is a constant $0 \leqslant \delta_1 < 1$ depending only on δ_0, C_1, ω, and α, and a constant $H > 0$ depending only on δ_1 and C_0, such that whenever $h_1 < H$, the iterates of (4.1) satisfy

$$\|\mu - \mu_\ell\|_{a_1} \leqslant \delta_1^\ell \|\mu - \mu_0\|_{a_1}.$$

The proofs of Theorems 4.2 and 4.3 proceed by modifying certain standard multigrid arguments (Bramble, 1993) appropriately. We present the proofs in Appendix B.

4.2 *Application to the HDG method*

To apply Algorithm 4.1 to the HDG method, we need to specify the computational ingredients B_0, R_1, and I_1 that appear in Algorithm 4.1. To apply Theorems 4.2 and 4.3, we must then verify the above mentioned assumptions. These verifications are in Section 5. In this subsection, we only give the algorithmic ingredients.

We select the 0th level discretization to be the standard continuous piecewise linear finite elements, on the *same* mesh as the HDG method, i.e.,

$$M_0 = \{v : \Omega \to \mathbb{R} \big|\ v \text{ is continuous},\ v|_{\partial\Omega} = 0,\ v|_K \in P_1(K),\ \forall \text{ triangles } K \in \mathscr{T}_h\},$$
$$(v,w)_0 = \int_\Omega vw\,dx, \qquad a_0(v,w) = \int_\Omega a\,\vec{\nabla} v \cdot \vec{\nabla} w\,dx. \tag{4.2}$$

The level 1 discretization, where we need the solution, is of course given by the HDG method, i.e., with $a_h(\cdot,\cdot)$ and M_h as defined before, we set

$$M_1 = M_h, \qquad (\eta,\mu)_1 = \sum_{K\in\mathscr{T}_h} \frac{|K|}{|\partial K|}\langle \eta,\mu\rangle_{\partial K}, \qquad a_1(\eta,\mu) = a_h(\eta,\mu).$$

Here $a_h(\cdot,\cdot)$ is as defined in (2.13) and $(\cdot,\cdot)_1$ is the inner product corresponding to the $\|\cdot\|_h$-norm defined in (3.1). The smoothing operator R_1 is chosen so that the smoothing step coincides with one Gauss-Seidel sweep for A_1. (As usual, it can also be chosen to be a scaled Jacobi iteration.)

The intergrid transfer operator I_1 is defined by

$$I_1 v = P_h^{M_1} v \tag{4.3}$$

where $P_h^{M_1} = P_h^{M_h}$ is the L^2-orthogonal projection onto $M_1 = M_h$ defined in (3.22). Clearly, when $d > 0$, this means that $(I_1 v)|_F = v|_F$ since any $v \in M_0$ is linear on F and hence in $P_d(F)$. In the case $d = 0$, $(I_1 v)|_F$ equals the mean of v on F. This operator is the same as that used for the HRT method in (Gopalakrishnan & Tan, 2009) where other deceptively similar but numerically unsuccessful operators are also discussed. Let us emphasize that the case $d > 0$ is essentially different from the case $d = 0$, as will be apparent in what follows.

It only remains to specify the operator B_0. This can be any *domain decomposition* or *multigrid* operator for the *standard* linear finite element method (satisfying Assumption 4.1). Examples can be found in (Toselli & Widlund, 2005) and (Bramble, 1993). For definiteness, we now consider a geometric multiplicative multigrid operator in more detail.

In the multigrid setting, as usual we assume that the mesh $\mathscr{T}_h$ is obtained by a number (say J) of successive refinements of a coarse mesh. Denote the coarse mesh by $\mathscr{T}_{-J}$. In two dimensions, one refinement strategy to obtain the mesh $\mathscr{T}_{-k+1}$ from $\mathscr{T}_{-k}$ is by simply connecting the midpoints of all edges of $\mathscr{T}_{-k}$. The multilevel spaces, in addition to the previously defined M_0 and M_1, are now defined by $M_{-k} = \{v : \Omega \to \mathbb{R} \,|\, v \text{ is continuous, } v|_{\partial\Omega} = 0,\ v|_K \in P_1(K),\ \forall \text{ triangles } K \in \mathscr{T}_{-k}\}$, for $-k = -J, -J+1, \ldots, -1$. Clearly these are the same standard finite element spaces as M_0, but defined with respect to the coarse meshes and $M_{-J} \subseteq M_{-J+1} \ldots \subseteq M_0$.

The full multilevel algorithm for the HDG method, obtained by combining the standard V-cycle on levels $-J, -J+1, \ldots, 0$, with the previous two level algorithm (Algorithm 4.1), is as follows. For $k = J-1, \ldots 1$, we set $I_{-k} : M_{-k-1} \to M_{-k}$ to identity, and Q_{-k} to the $L^2(\Omega)$-orthogonal projection onto M_{-k}. For the same indices, the operators A_{-k} are generated by the form $a_{-k}(\cdot,\cdot) = a_0(\cdot,\cdot)$ and $(\cdot,\cdot)_{-k} = (\cdot,\cdot)_0$ defined by (4.2), and the smoothers R_{-k} are defined by Gauss-Seidel sweeps using A_{-k}. With these notations, the following algorithm is the textbook V-cycle.

Algorithm 4.4 (Multilevel V-cycle) For any g in M_1 we define $B_1^{\mathrm{mg}} g$ recursively. First, at the coarsest level, set $B_{-J}^{\mathrm{mg}} g = (A_{-J})^{-1} g$. For all $-J < k \leqslant 1$, define $B_k^{\mathrm{mg}} : M_k \to M_k$ by

1. Smooth: $v_1 = R_k g$.
2. Correct: $v_2 = v_1 + I_k B_{k-1}^{\mathrm{mg}} Q_k (g - A_k v_1)$.
3. Smooth: $v_3 = v_2 + R_k^t (g - A_k v_2)$.

Set $B_k^{\mathrm{mg}} g \equiv v_3$.

The convergence of this algorithm is studied next.

5. Multigrid convergence analysis

This section is devoted to proving the uniform convergence of the previously discussed multigrid algorithm for the HDG method. We will do so under a mild regularity assumption on the solutions of the boundary value problem (2.1).

Assumption 5.1 From now on we assume the following:

(i) The coefficient $a(x)$ is constant on each element of the finest mesh $\mathscr{T}_h$.

(ii) Problem (2.1) admits the following regularity estimate for its solution:

$$\|\vec{q}\|_{H^s(\Omega)^n} + \|u\|_{H^{1+s}(\Omega)} \leqslant C \|f\|_{H^{-1+s}(\Omega)} \tag{5.1}$$

for some number $1/2 < s \leqslant 1$.

Note that once (5.1) holds with $s > 1/2$, we can apply the projection Π_h to $(\vec{q}, u)$. The projection Π_h is required in multigrid analysis, hence the assumption is that $s > 1/2$. Note also that in the simple case of $a \equiv 1$ in a polygonal Ω (with no slits), it is well known (Dauge, 1988; Kellogg, 1971) that the estimate (5.1) holds with $s > 1/2$. The full regularity estimate with $s = 1$ is well known to hold when $a \equiv 1$ and Ω is a convex domain in two or three dimensions (Grisvard, 1985).

THEOREM 5.1 Suppose Assumption 5.1 holds. Given any $\mu_0 \in M_h$, suppose the iterates μ_ℓ for $\ell \geqslant 1$ are given by $\mu_{\ell+1} = \mu_\ell + B_1^{\mathrm{mg}}(b - A_1 \mu_\ell)$, where B_1^{mg} is defined by Algorithm 4.4. Then, for the HDG method with degree $d > 0$, there is a $\delta < 1$ independent of the fine mesh size h (but depending on τ and a) such that

$$\|\mu - \mu_\ell\|_a \leqslant \delta^\ell \|\mu - \mu_0\|_a, \tag{5.2}$$

where $\|\mu\|_a = a_h(\mu,\mu)^{1/2}$ with $a_h(\cdot,\cdot)$ defined by (2.13). For the $d = 0$ case, (5.2) holds provided $\kappa_{h\tau} := \max\{\tau_K^* h_K : K \in \mathscr{T}_h\}$ is sufficiently small.

Due to the theorem, we can expect the multigrid convergence rate to be h-independent if $\tau \equiv 1$ (the most common choice) in the $d > 0$ case. In the $d = 0$ case, the theorem says that the choice $\tau \equiv 1$ would result in uniform multigrid convergence, provided the *fine* mesh size is sufficiently small, which is a reasonable assumption in most practical situations.

This theorem obviously follows from the abstract statements of Theorems 4.2 and 4.3, once we verify its assumptions for the particular case of the HDG method. Note that Algorithm 4.4 is the same as Algorithm 4.1 with B_0 set to a standard multigrid operator, namely B_0^{mg} in our notation. Assumption 4.1 is well known (Bramble, 1993) to hold for B_0^{mg}. Furthermore, the assumption on the smoother R_1, namely Assumption 4.5 is also easily proved for the Gauss-Seidel operator based on any local basis. Although our form $a_h(\cdot,\cdot)$ is nonstandard, since it is local, few changes are needed from the standard smoothing analysis (Bramble, 1993). Hence we will only verify the remaining assumptions of Section 4. We begin with a preparatory lemma.

LEMMA 5.1 The following identities hold for all λ, η in M_1 and all w in M_0:

$$
\begin{aligned}
\vec{\mathcal{Q}}(I_1 w) &= -a\vec{\nabla} w, && (\forall d), && (5.3)\\
a_1(I_1 w, \eta) &= -(\vec{\nabla} w, \vec{\mathcal{Q}}\eta)_{\mathscr{T}_h}, && \text{if } d>0, && (5.4)\\
(\vec{\mathcal{Q}}\lambda + a\vec{\nabla} P_0\lambda, \vec{\nabla} w)_{\mathscr{T}_h} &= \langle \tau(\mathcal{U}\lambda - \lambda), \mathcal{U}(I_1 w) - w\rangle_{\partial\mathscr{T}_h}, && \text{if } d=0, && (5.5)\\
(\vec{\mathcal{Q}}\lambda + a\vec{\nabla} P_0\lambda, \vec{\nabla} w)_{\mathscr{T}_h} &= 0, && \text{if } d>0, && (5.6)\\
a_1(\lambda - I_1 P_0\lambda, \lambda) &= (c\,(\vec{\mathcal{Q}}\lambda + a\vec{\nabla} P_0\lambda), \vec{\mathcal{Q}}\lambda - a\vec{\nabla} P_0\lambda)_{\mathscr{T}_h} \\
&\quad + \|\mathcal{U}\lambda - \lambda\|_\tau^2, && \text{if } d=0, && (5.7)\\
a_1(\lambda - I_1 P_0\lambda, \lambda) &= \|\vec{\mathcal{Q}}\lambda + a\vec{\nabla} P_0\lambda\|_c^2 + \|\mathcal{U}\lambda - \lambda\|_\tau^2, && \text{if } d>0. && (5.8)
\end{aligned}
$$

Proof. If $d > 0$, then the first identity follows from Lemma 3.4 (iii) by virtue of Assumption 5.1 (i). If $d = 0$, it follows from the definition of $\vec{\mathcal{Q}}(I_1 w)$, namely (2.9a), which reduces to

$$(c\,\vec{\mathcal{Q}}(I_1 w), \vec{r})_K - 0 = -\langle I_1 w, \vec{r}\cdot\vec{n}\rangle_{\partial K} = -\langle w, \vec{r}\cdot\vec{n}\rangle_{\partial K} = -(\vec{\nabla} w, \vec{r})_K$$

for all constant vectors $\vec{r}$. Since $c\,\vec{\mathcal{Q}}(I_1 w)$ and $\vec{\nabla} v$ are constant vectors, (5.3) follows.

To prove (5.4), we use (5.3) in the definition of $a_1(\cdot,\cdot)$ to get

$$a_1(I_1 w, \eta) = (-\vec{\nabla} w, \vec{\mathcal{Q}}\eta)_{\mathscr{T}_h} + \langle \tau(\mathcal{U}(I_1 w) - w), \mathcal{U}\eta - \eta\rangle_{\partial\mathscr{T}_h} \tag{5.9}$$

and observe that the last term vanishes because of Lemma 3.4 (iii).

To prove (5.5) and (5.6), we again use (5.3) as well as the fact that $a = c^{-1}$, to get

$$
\begin{aligned}
(\vec{\mathcal{Q}}\lambda + a\vec{\nabla} P_0\lambda, \vec{\nabla} w)_{\mathscr{T}_h} &= (c\,\vec{\mathcal{Q}}\lambda, a\vec{\nabla} w)_{\mathscr{T}_h} + (a\vec{\nabla} P_0\lambda, \vec{\nabla} w)_{\mathscr{T}_h}\\
&= -(c\,\vec{\mathcal{Q}}\lambda, \vec{\mathcal{Q}}(I_1 w))_{\mathscr{T}_h} + a_0(P_0\lambda, w)\\
&= -(c\,\vec{\mathcal{Q}}\lambda, \vec{\mathcal{Q}}(I_1 w))_{\mathscr{T}_h} + a_1(\lambda, I_1 w) = \langle \tau(\mathcal{U}\lambda - \lambda), \mathcal{U}(I_1 w) - w\rangle_{\partial\mathscr{T}_h}.
\end{aligned}
$$

which holds for all d. In particular, for $d = 0$, this is (5.5). If $d > 0$, then the last term vanishes because of Lemma 3.4 (iii) and we get (5.6).

Finally, to prove (5.7) and (5.8),

$$\begin{aligned} a_1(\lambda - I_1P_0\lambda, \lambda) &= a_1(\lambda,\lambda) - a_0(P_0\lambda, P_0\lambda) \\ &= (c\,\vec{\mathcal{Q}}\lambda, \vec{\mathcal{Q}}\lambda)_{\mathcal{T}_h} - (a\,\vec{\nabla}P_0\lambda, \vec{\nabla}P_0\lambda)_{\mathcal{T}_h} + \|\mathcal{U}\lambda - \lambda\|_\tau^2 \\ &= (c\,(\vec{\mathcal{Q}}\lambda + a\,\vec{\nabla}P_0\lambda), (\vec{\mathcal{Q}}\lambda - a\,\vec{\nabla}P_0\lambda))_{\mathcal{T}_h} + \|\mathcal{U}\lambda - \lambda\|_\tau^2. \end{aligned}$$

This holds for either $d = 0$ or $d > 0$, so (5.7) is already proved. To see that (5.8) also holds, it suffices to note that the term $\vec{\mathcal{Q}}\lambda - a\vec{\nabla}P_0\lambda$ can be replaced by $\vec{\mathcal{Q}}\lambda + a\vec{\nabla}P_0\lambda$ due to (5.6) whenever $d > 0$. □

5.1 *Norm of prolongation*

In this subsection we prove the following result, which verifies Assumptions 4.2 and 4.3.

THEOREM 5.2 For all v in M_0, we have

$$\begin{aligned} &a_1(I_1v, I_1v) = a_0(v,v), && \text{if } d > 0, \\ &a_1(I_1v, I_1v) \leqslant (1 + C\kappa_{h\tau})a_0(v,v), && \text{if } d = 0, \end{aligned}$$

Thus, Assumption 4.2 holds for $d > 0$ and Assumption 4.3 holds for $d = 0$ case with $C_0h_1 = C\kappa_{h\tau}$.

LEMMA 5.2 If $d = 0$, then for all w in $P_1(K)$,

$$\|\mathcal{U}(I_1w) - I_1w\|_{\tau,\partial K} \leqslant C(\tau_K^* h_K)^{1/2}\|\vec{\nabla}w\|_K. \tag{5.10}$$

$$\|\mathcal{U}(I_1w) - w\|_K \leqslant Ch_K\|\vec{\nabla}w\|_K. \tag{5.11}$$

Proof. Because of (2.9b),

$$\langle \tau(\mathcal{U}(I_1w) - I_1w), \mathcal{U}(I_1w) - w_0\rangle_{\partial K} = 0,$$

for any constant w_0, and so

$$\begin{aligned} \|\mathcal{U}(I_1w) - I_1w\|_{\tau,\partial K}^2 &= \langle \tau(\mathcal{U}(I_1w) - I_1w), w_0 - I_1w\rangle_{\partial K} \\ &\leqslant \|\tau(\mathcal{U}(I_1w) - I_1w)\|_{\partial K}\|I_1w - w_0\|_{\partial K} \\ &\leqslant C\sqrt{\tau_K^*}\|\mathcal{U}(I_1w) - I_1w)\|_{\tau,\partial K}\|I_1w - w_0\|_{\partial K}, \end{aligned}$$

by Lemma 3.3. This implies that

$$\|\mathcal{U}(I_1w) - I_1w\|_{\tau,\partial K} \leqslant C\sqrt{\tau_K^*}\,\|I_1w - w_0\|_{\partial K},$$

and if we take w_0 to be the value of w at the barycenter of the simplex K, we immediately get that

$$\|\mathcal{U}(I_1w) - I_1w\|_{\tau,\partial K} \leqslant C\,(\tau_K^* h_K)^{1/2}\|\vec{\nabla}w\|_K.$$

This proves (5.10).

For (5.11), we note that, because of (2.9b),

$$\langle \tau(\mathcal{U}(I_1w) - w), \mathcal{U}(I_1w) - w_0\rangle_{\partial K} = 0,$$

for any constant w_0, and so

$$\|\mathcal{U}(I_1w) - w\|_{\tau,\partial K} \leqslant \|w - w_0\|_{\tau,\partial K}. \tag{5.12}$$

Next, we use a standard local estimate for linear functions,

$$C\|\mathcal{U}(I_1 w)-w\|_K \leqslant h_K\|\vec{\nabla}(\mathcal{U}(I_1 w)-w)\|_K + h_K^{1/2}\|\mathcal{U}(I_1 w)-w)\|_F$$

for any face F of K. Choosing $F = F_{\max}$, a face where τ assumes its maximum value,

$$\begin{aligned} C\|\mathcal{U}(I_1 w)-w\|_K &\leqslant h_K\|\vec{\nabla}w\|_K + h_K^{1/2}(\tau_K^{\max})^{-1/2}\|\mathcal{U}(I_1 w)-w\|_{\tau,F_{\max}} \\ &\leqslant h_K\|\vec{\nabla}w\|_K + h_K^{1/2}(\tau_K^{\max})^{-1/2}\|w-w_0\|_{\tau,\partial K} \\ &\leqslant h_K\|\vec{\nabla}w\|_K + h_K^{1/2}\|w-w_0\|_{\partial K}, \end{aligned}$$

where we have used (5.12). Choosing w_0 to be the mean of w on K, we get the inequality (5.11). This completes the proof. □

Proof of Theorem 5.2. To prove the $d > 0$ case, we use two identities of Lemma 5.1:

$$\begin{aligned} a_1(I_1 v, I_1 v) &= -(\vec{\nabla}v, \vec{\mathcal{Q}}(I_1 v))_{\mathscr{T}_h} && \text{by (5.4)} \\ &= (a\vec{\nabla}v, \vec{\nabla}v)_{\mathscr{T}_h} && \text{by (5.3).} \end{aligned}$$

To prove the next inequality for the $d = 0$ case, we again use (5.3) of Lemma 5.1, hence we can put $\vec{\mathcal{Q}}(I_1 v) = -a\vec{\nabla}v$ in the definition of $a_1(\cdot,\cdot)$ to get

$$\begin{aligned} a_1(I_1 v, I_1 v) &= (a\vec{\nabla}v, \vec{\nabla}v)_{\mathscr{T}_h} + \|\mathcal{U}(I_1 v) - I_1 v\|_\tau^2 \\ &\leqslant a_0(v,v) + C\kappa_{h\tau}\|\vec{\nabla}v\|^2, \end{aligned}$$

where the last step was due to (5.10) of Lemma 5.2. This proves the theorem. □

5.2 *Regularity and approximation property*

This subsection is devoted to proving Assumption 4.4. We begin with a simple consequence of Theorem 5.2.

LEMMA 5.3 For all μ in M_1, we have

$$a_0(P_0\mu, P_0\mu) \leqslant \begin{cases} \|\mu\|_a^2 & \text{if } d > 0 \\ (1 + C\kappa_{h\tau})\|\mu\|_a^2 & \text{if } d = 0. \end{cases}$$

Proof. The estimate follows from

$$\begin{aligned} (a\vec{\nabla}P_0\mu, \vec{\nabla}P_0\mu) = a_0(P_0\mu, P_0\mu) &= a_1(\mu, I_1 P_0\mu) \\ &\leqslant \|\mu\|_a \|I_1 P_0\mu\|_a, \end{aligned}$$

and applying Theorem 5.2 to the last term. Recall that $\|\cdot\|_a$ is the norm associated to a_1. □

The known techniques to prove the regularity and approximation property involve a duality argument that shows that $\mu - I_1 P_0\mu$ is small in appropriate norms. The usual difficulty is that μ is a finite element function on which higher order Sobolev norms cannot be put (in our case μ in M_h is in general discontinuous). One technique to overcome this difficulty proceeds by constructing an H^1-approximation to any given μ in M_h. To do so, we solve the boundary value problem with a specific right hand side

f_μ constructed by applying a discrete version of the exact partial differential operator to μ. The added difficulty in the HDG case is that μ is supported only on mesh element boundaries, so obtaining a proper f_μ within the element interiors requires some trickery. First, we introduce a local operator S_i^K.

Let λ be the restriction of a function in M_h on ∂K for some mesh element K. Let F_i denote the face of K opposite to the ith vertex of K. Then define $S_i^K\lambda$ in $P_{d+1}(K)$ by

$$\langle S_i^K\lambda, \eta\rangle_{F_i} = \langle \lambda, \eta\rangle_{F_i} \qquad \text{for all } \eta \in P_{d+1}(F_i), \tag{5.13a}$$

$$(S_i^K\lambda, v)_K = (\mathcal{U}\lambda, v)_K \qquad \text{for all } v \in P_d(K), \tag{5.13b}$$

and, considering all the $n+1$ faces of K, define

$$(\lambda,\mu)_S = \sum_{K\in\mathscr{T}_h} \frac{1}{n+1}\sum_{i=1}^{n+1}(S_i^K\lambda, S_i^K\mu)_K \qquad \text{and} \qquad \|\lambda\|_S^2 = (\lambda,\lambda)_S.$$

LEMMA 5.4 Equations (5.13a) and (5.13b) uniquely define a $S_i^K\lambda$ in $P_{d+1}(K)$. Furthermore, for all λ in M_h,

$$\mathcal{U}\lambda|_K = P_h^W(S_i^K\lambda), \tag{5.14}$$

$$C_5\|\lambda\|_1 \leqslant \|\lambda\|_S \leqslant C_6\sqrt{\gamma_{h\tau}^{(1)}}\|\lambda\|_1, \tag{5.15}$$

$$\|\vec{\nabla}(S_i^K\lambda)\|_K \leqslant C\sqrt{\gamma_{h\tau}^{(1)}}\|\vec{\mathcal{Q}}\lambda\|_K. \tag{5.16}$$

Recall that (see Theorem 3.3) $\gamma_{h\tau}^{(1)} = \max\{1+\tau_K^* h_K : K \in \mathscr{T}_h\}$.

Proof. Since (5.13) forms a square system for $S_i^K\lambda$, to show that it has a unique solution, it suffices to show that the only solution when the right hand sides are zero is the trivial solution. That this is indeed the case is an immediate consequence of (Cockburn *et al.*, 2010, Lemma A.1). The identity (5.14) is obvious from (5.13b). Let us prove the remaining assertions.

We prove (5.15) by a scaling argument. To this end, consider a fixed reference simplex $\hat{K}$, with an arbitrarily chosen face $\hat{F}$, and define a map $\Psi : P_{d+1}(\hat{F}) \times P_d(\hat{K}) \to P_{d+1}(\hat{K})$ by

$$\langle \Psi(\hat{\lambda},\hat{q}), \eta\rangle_{\hat{F}} = \langle\hat{\lambda},\eta\rangle_{\hat{F}} \qquad \text{for all } \eta \in P_{d+1}(\hat{F}),$$

$$(\Psi(\hat{\lambda},\hat{q}), v)_{\hat{K}} = (\hat{q}, v)_{\hat{K}} \qquad \text{for all } v \in P_d(\hat{K}).$$

It is easy to see that $(\hat{\lambda},\hat{q}) \mapsto \|\Psi(\hat{\lambda},\hat{q})\|_{\hat{K}}$ and $(\hat{\lambda},\hat{q}) \mapsto (\|\hat{\lambda}\|_{\hat{F}}^2 + \|\hat{q}\|_{\hat{K}}^2)^{1/2}$ are equivalent norms on $P_{d+1}(\hat{F}) \times P_d(\hat{K})$. Mapping to any element K such that $\hat{F}$ gets mapped to the face F_i of K, and relating $\Psi(\hat{\lambda},\hat{q})$ to $S_i^K\lambda$, we have

$$C_5(h_K\|\lambda\|_{F_i}^2 + \|\mathcal{U}\lambda\|_K^2) \leqslant \|S_i^K\lambda\|_K^2 \leqslant C(h_K\|\lambda\|_{F_i}^2 + \|\mathcal{U}\lambda\|_K^2)$$

Summing over all faces $F_i \subset \partial K$,

$$C_5(h_K\|\lambda\|_{\partial K}^2) \leqslant \sum_{i=1}^{n+1}\|S_i^K\lambda\|_K^2 \leqslant C\left(h_K\|\lambda\|_{\partial K}^2 + C\gamma_{h\tau}^{(1)}h_K\|\lambda\|_{\partial K}^2\right) \tag{5.17}$$

where we have used Theorem 3.1. Now, to obtain (5.15), we need only sum over all K.

To prove (5.16), first observe that if λ takes a constant value κ on the boundary of some mesh element ∂K, then $S_i^K\lambda \equiv \kappa$. This is because $\mathcal{U}\lambda \equiv \kappa$ by Lemma 3.4(i), so the function κ satisfies both the equations of (5.13). Therefore, by the unique solvability of (5.13), $S_i^K\lambda \equiv \kappa$. A consequence of this fact is that for any λ, we have

$$\vec{\nabla}(S_i^K(m_K(\lambda)) = 0$$

where $m_K(\lambda)$ is as in (3.19). Therefore,

$$\begin{aligned}
\|\vec{\nabla}(S_i^K\lambda)\|_K &= \|\vec{\nabla}S_i^K(\lambda - m_K(\lambda))\|_K, \\
&\leqslant Ch_K^{-1}\|S_i^K(\lambda - m_K(\lambda))\|_K \qquad \text{(by an inverse inequality)} \\
&\leqslant Ch_K^{-1}(1+(\tau_K^* h_K)^{1/2})h_K^{1/2}\|\lambda - m_K(\lambda)\|_{\partial K} \qquad \text{(by (5.17))} \\
&\leqslant C\sqrt{\gamma_{h\tau}^{(1)}}|\!|\!|\lambda - m_K(\lambda)|\!|\!|_{h,K},
\end{aligned}$$

since, see Theorem 3.3, $\gamma_{h\tau}^{(1)} = \max\{1+\tau_K^* h_K : K \in \mathscr{T}_h\}$. Thus, (5.16) follows from Lemma 3.6. □

Next, we define a map $\lambda \longmapsto \tilde{\lambda}$ from M_1 into M_1 as follows. First, given λ in M_1, let ϕ_λ be the unique function in M_1 satisfying

$$(\phi_\lambda, \mu)_S = a_h(\lambda, \mu), \quad \forall \mu \in M_1. \tag{5.18}$$

This equation is uniquely solvable for ϕ_λ in M_1, because if the right-hand side is zero, then by (5.15) of Lemma 5.4, we have that $\phi_\lambda = 0$. Next, let $f_\lambda = \mathcal{U}\phi_\lambda$ and define $\tilde{\lambda} \in M_1$ to be the unique solution of the equation

$$a_h(\tilde{\lambda}, \mu) = (f_\lambda, \mathcal{U}\mu), \quad \forall \mu \in M_1. \tag{5.19}$$

Recall that we have agreed to drop the subscript Ω from notations for norms and inner products, as e.g., in the right hand side above: $(f_\lambda, \mathcal{U}\mu)_\Omega \equiv (f_\lambda, \mathcal{U}\mu)$.

LEMMA 5.5 The following statements hold for all λ in M_1:

$$\|f_\lambda\| \leqslant \|\phi_\lambda\|_S \leqslant C\|A_1\lambda\|_1 \tag{5.20}$$

$$\|\lambda - \tilde{\lambda}\|_a \leqslant Ch\|A_1\lambda\|_1 \tag{5.21}$$

$$\|f_\lambda\|_{H^{-1}(\Omega)} \leqslant C\gamma_{h\tau}^{(1)}\|\lambda\|_a. \tag{5.22}$$

Proof. The proofs of (5.20) and (5.21) are similar to the proof of (Gopalakrishnan & Tan, 2009, Lemma 4.5). The only difference is that we now use the estimates of Lemma 5.4. To prove (5.20), first observe that

$$f_\lambda = P_h^W S_i^K \phi_\lambda$$

by (5.14) of Lemma 5.4. Therefore,

$$\|f_\lambda\|^2 = \sum_{K \in \mathscr{T}_h} \frac{1}{n+1}\sum_{i=1}^{n+1}\|P_h^W(S_i^K\phi_\lambda)\|_K^2 \leqslant \|\phi_\lambda\|_S^2,$$

which is the first of the inequalities in (5.20). Moreover,

$$\begin{aligned}
\|\phi_\lambda\|_S &= \sup_{\mu\in M_1}\frac{(\phi_\lambda,\mu)_S}{\|\mu\|_S} = \sup_{\mu\in M_1}\frac{a_h(\lambda,\mu)}{\|\mu\|_S} && \text{by (5.18),}\\
&= \sup_{\mu\in M_1}\frac{(A_1\lambda,\mu)_1}{\|\mu\|_S} \leqslant \sup_{\mu\in M_1}\frac{(A_1\lambda,\mu)_1}{C_5\|\mu\|_1} && \text{by (5.15) of Lemma 5.4,}\\
&\leqslant C\|A_1\lambda\|_1,
\end{aligned}$$

thus completing the proof of (5.20).

To prove (5.21), let us first note that we can rewrite (5.18) and (5.19) as follows:

$$\begin{aligned}
a_h(\lambda,\mu) &= \sum_{K\in\mathscr{T}_h}\frac{1}{n+1}\sum_{i=1}^{n+1}(S_i^K\phi_\lambda, S_i^K\mu)_K,\\
a_h(\tilde\lambda,\mu) &= \sum_{K\in\mathscr{T}_h}\frac{1}{n+1}\sum_{i=1}^{n+1}(P_h^W S_i^K\phi_\lambda, S_i^K\mu)_K.
\end{aligned}$$

To get the last identity, we have again used Lemma 5.4, whereby $f_\lambda = \mathcal{U}\phi_\lambda = \frac{1}{n+1}\sum_{i=1}^{n+1}P_h^W(S_i^K\phi_\lambda)$ on any element K. Subtracting, and setting $\mu=\lambda-\tilde\lambda$, we get

$$\begin{aligned}
\|\lambda-\tilde\lambda\|_a^2 &= \sum_{K\in\mathscr{T}_h}\frac{1}{n+1}\sum_{i=1}^{n+1}((I-P_h^W)S_i^K\phi_\lambda, S_i^K(\lambda-\tilde\lambda))_K\\
&= \sum_{K\in\mathscr{T}_h}\frac{1}{n+1}\sum_{i=1}^{n+1}(S_i^K\phi_\lambda, (I-P_h^W)S_i^K(\lambda-\tilde\lambda))_K.
\end{aligned}$$

Using the Friedrichs estimate $\|u-P_h^W u\|_{L^2(K)}\leqslant Ch|u|_{H^1(K)}$, we get

$$\begin{aligned}
\|\lambda-\tilde\lambda\|_a^2 &\leqslant C\|\phi_\lambda\|_S\left(\sum_{K\in\mathscr{T}_1}\frac{1}{n+1}\sum_{i=1}^{n+1}h^2|S_i^K(\lambda-\tilde\lambda)|^2_{H^1(K)}\right)^{1/2}\\
&\leqslant Ch\|\phi_\lambda\|_S\,\|\vec{\mathcal{Q}}(\lambda-\tilde\lambda)\|\\
&\leqslant Ch\|A_1\lambda\|_1\,\|\lambda-\tilde\lambda\|_a
\end{aligned}$$

by (5.20) and (5.16) of Lemma 5.4. Canceling the common factor above, we obtain (5.21).

Next, let us prove (5.22). To this end, given any ψ in $H_0^1(\Omega)$, let ψ_0 in M_0 denote a function satisfying

$$\|\vec\nabla\psi_0\|\leqslant C\|\vec\nabla\psi\| \qquad\text{and}\qquad \|\psi-\psi_0\|\leqslant Ch\|\vec\nabla\psi\|. \tag{5.23}$$

Such approximations are well known to exist (Scott & Zhang, 1990). Then,

$$\begin{aligned}
\|f_\lambda\|_{H^{-1}(\Omega)} &= \sup_{\psi\in H_0^1(\Omega)}\frac{(f_\lambda,\psi)}{\|\vec\nabla\psi\|}\\
&= \sup_{\psi\in H_0^1(\Omega)}\frac{(f_\lambda,(\psi-\psi_0)+(\psi_0-\mathcal{U}(I_1\psi_0))+\mathcal{U}(I_1\psi_0))}{\|\vec\nabla\psi\|}.
\end{aligned}$$

Now, since

$$\begin{aligned}
(f_\lambda, \psi - \psi_0) &\leqslant C\|f_\lambda\|h\|\vec{\nabla}\psi\|, && \text{by (5.23)},\\
(f_\lambda, \psi_0 - \mathcal{U}(I_1\psi_0)) &\leqslant C\|f_\lambda\|h\|\vec{\nabla}\psi\|, && \text{by Lemma 5.2, if } d=0, \text{ while}\\
(f_\lambda, \psi_0 - \mathcal{U}(I_1\psi_0)) &= 0, && \text{by Lemma 3.4(iii), if } d>0,\\
(f_\lambda, \mathcal{U}(I_1\psi_0)) &= a_h(\tilde{\lambda}, I_1\psi_0) && \text{by (5.19)},\\
&\leqslant \|\tilde{\lambda}\|_a C d_\tau \|\vec{\nabla}\psi_0\|, && \text{by Theorem 5.2},
\end{aligned}$$

where $d_\tau = 1 + (1-\delta_{d0})(\max(\tau)h)^{1/2}$, the terms in the supremum can be bounded accordingly to get that

$$\|f_\lambda\|_{H^{-1}(\Omega)} \leqslant C d_\tau \|\tilde{\lambda}\|_a + Ch\|f_\lambda\|.$$

Finally, since (5.20) implies that $\|f_\lambda\| \leqslant C\|A_1\lambda\|_1$, and since (5.21) implies

$$\|\tilde{\lambda}\|_a \leqslant \|\lambda\|_a + Ch\|A_1\lambda\|_1,$$

we have

$$\|f_\lambda\|_{H^{-1}(\Omega)} \leqslant C d_\tau \|\lambda\|_a + Ch\|A_1\lambda\|_1. \tag{5.24}$$

Now,

$$\|A_1\lambda\|_1^2 = (A_1 A_1^{1/2}\lambda, A_1^{1/2}\lambda)_1 \leqslant \rho(A_1)\,(A_1^{1/2}\lambda, A_1^{1/2}\lambda)_1 = \rho(A_1)\,\|\lambda\|_a^2.$$

By Theorem 3.2, $\rho(A_1) \leqslant \gamma_{h\tau}^{(2)} c_2 h^{-2}$, so we find that the last term in (5.24) satisfies $h\|A_1\lambda\|_1 \leqslant C\gamma_{h\tau}^{(1)}\|\lambda\|_a$. Thus, we have proved (5.22). □

Now, let $\tilde{u}$ be the unique function in $H_0^1(\Omega)$ that solves

$$(a\vec{\nabla}\tilde{u}, \vec{\nabla}v) = (f_\lambda, v), \quad \forall v \in H_0^1(\Omega). \tag{5.25}$$

This $\tilde{u}$ serves as the H^1-approximation to λ mentioned in the beginning of this subsection (see the remarks after Lemma 5.3). Let $\tilde{u}_0$ be the unique function in M_0 satisfying

$$(a\vec{\nabla}\tilde{u}_0, \vec{\nabla}v) = (f_\lambda, v), \quad \forall v \in M_0. \tag{5.26}$$

LEMMA 5.6 If $d > 0$, then $P_0\tilde{\lambda} - \tilde{u}_0 = 0$, whereas, if $d = 0$, then

$$\|\vec{\nabla}(P_0\tilde{\lambda} - \tilde{u}_0)\| \leqslant Ch\|A_1\lambda\|_1.$$

Proof. Observe that for all w in M_0,

$$\begin{aligned}
(a\vec{\nabla}P_0\tilde{\lambda}, \vec{\nabla}w) &= a_0(P_0\tilde{\lambda}, w) = a_h(\tilde{\lambda}, I_1 w)\\
&= (f_\lambda, \mathcal{U}(I_1 w)),
\end{aligned} \tag{5.27}$$

by (5.19). Now, if $d > 0$, by Lemma 3.4 (iii), we know that $\mathcal{U}(I_1 w) - w = 0$. Hence we have

$$(a\vec{\nabla}P_0\tilde{\lambda}, \vec{\nabla}w) = (f_\lambda, w) \qquad \forall w \in M_0,$$

which is the same equation satisfied by $\tilde{u}_0$. Hence $P_0\tilde{\lambda}$ and $\tilde{u}_0$ coincide if $d > 0$.

If $d = 0$, then we again proceed as above, noting that although $\mathcal{U}(I_1 w) - w$ may not vanish, it can be bounded using Lemma 5.2:

$$\begin{aligned}(a\vec{\nabla}P_0\tilde{\lambda}, \vec{\nabla}w) &= (f_\lambda, w) + (f_\lambda, \mathcal{U}(I_1 w) - w)\\ &= (a\vec{\nabla}\tilde{u}_0, \vec{\nabla}w) + (f_\lambda, \mathcal{U}(I_1 w) - w)\\ &\leqslant (a\vec{\nabla}\tilde{u}_0, \vec{\nabla}w) + \|f_\lambda\|Ch\|\vec{\nabla}w\|\end{aligned}$$

Hence

$$(a\vec{\nabla}(P_0\tilde{\lambda} - \tilde{u}_0), \vec{\nabla}w) \leqslant Ch\|A_1\lambda\|_1\,\|\vec{\nabla}w\|.$$

Choosing $w = P_0\tilde{\lambda} - \tilde{u}_0$, and applying (5.20) of Lemma 5.5, we finish the proof of the required inequality. □

LEMMA 5.7 If s is as in Assumption 5.1, for any λ in M_1,

$$\|\vec{\mathcal{Q}}\lambda + a\vec{\nabla}P_0\lambda\|^2_{\mathscr{T}_h} \leqslant Ch^2\|A_1\lambda\|_1^2 + C\gamma_\tau h^{2s} a_h(\lambda,\lambda)^{1-s}\|A_1\lambda\|_1^{2s}.$$

where $\gamma_\tau = (1 + \max_{K\in\mathscr{T}_h}(\tau_K^*)^2)(\gamma_{h\tau}^{(1)})^{(2-2s)}$.

Proof. First, let us split the term requiring estimation as

$$\vec{\mathcal{Q}}\lambda + a\,\vec{\nabla}(P_0\lambda) = \sum_{i=1}^{6} t_i,$$

where

$$\begin{aligned}t_1 &= \vec{\mathcal{Q}}(\lambda - \tilde{\lambda}),\\ t_2 &= \vec{\mathcal{Q}}(\tilde{\lambda} - P_h^M\tilde{u}),\\ t_3 &= \vec{\mathcal{Q}}(P_h^M\tilde{u} - I_1\tilde{u}_0),\\ t_4 &= \vec{\mathcal{Q}}(I_1\tilde{u}_0) - (-a\vec{\nabla}\tilde{u}_0),\\ t_5 &= a\vec{\nabla}P_0\tilde{\lambda} - a\vec{\nabla}\tilde{u}_0,\\ t_6 &= a\vec{\nabla}P_0(\lambda - \tilde{\lambda}).\end{aligned}$$

These terms are bounded as follows:

$$\begin{aligned}\|t_1\| &\leqslant \|\lambda - \tilde{\lambda}\|_a \leqslant Ch\|A_1\lambda\|_1 && \text{by Lemma 5.5,}\\ \|t_2\| &\leqslant \|\tilde{\lambda} - P_h^M\tilde{u}\|_a \leqslant \|\tilde{q} - \Pi_h^V\tilde{q}\|_c && \text{by (3.26) of Theorem 3.5,}\\ &\leqslant Ch^s(|\tilde{q}|_{H^s(\Omega)} + \max_{K\in\mathscr{T}_h}(\tau_K^*)|\tilde{u}|_{H^s(\Omega)}) && \text{by (3.23a),}\end{aligned}$$

where $\tilde{q} = -a\vec{\nabla}\tilde{u}$. For t_3, we use (3.2) of Theorem 3.1 to get that

$$\|\vec{\mathcal{Q}}(P_h^M\tilde{u} - \tilde{u}_0)\|^2 \leqslant C\gamma_{h\tau}^{(1)}h^{-2}\|P_h^M(\tilde{u} - \tilde{u}_0)\|_h^2 \leqslant C\gamma_{h\tau}^{(1)}h^{-2}\|\tilde{u} - \tilde{u}_0\|_h^2, \tag{5.28}$$

where $\gamma_{h\tau}^{(1)} = \max\{1 + h_K \tau_K^{\max} : K \in \mathscr{T}_h\}$. By a local trace inequality, we can estimate the mesh dependent norm above by interior norms as follows:

$$C\|\tilde{u} - \tilde{u}_0\|_{h,K}^2 \leqslant \|\tilde{u} - \tilde{u}_0\|_K^2 + h_K^2 \|\vec{\nabla}(\tilde{u} - \tilde{u}_0)\|_K^2. \tag{5.29}$$

Since $\tilde{u}_0$ is a standard Galerkin approximation (Ciarlet, 1978) of $\tilde{u}$, we have

$$\|\vec{\nabla}(\tilde{u} - \tilde{u}_0)\| \leqslant Ch^s |\tilde{u}|_{H^{1+s}(\Omega)}. \tag{5.30}$$

Furthermore, a standard duality argument (Ciarlet, 1978; Nitsche, 1968) yields

$$\|\tilde{u} - \tilde{u}_0\| \leqslant Ch^s \|\vec{\nabla}(\tilde{u} - \tilde{u}_0)\| \leqslant Ch^{1+s} |\tilde{u}|_{H^{1+s}(\Omega)}. \tag{5.31}$$

Summing (5.29) over all elements and using (5.30) and (5.31), we can estimate $\|\tilde{u} - \tilde{u}_0\|_h$. Returning to (5.28) and using this bound, we have

$$\|t_3\| \leqslant C(\gamma_{h\tau}^{(1)})^{1/2} h^s |u|_{H^{1+s}(\Omega)}.$$

Proceeding to the succeeding terms,

$$\begin{aligned}
&\|t_4\| = 0, && \text{by (5.3) of Lemma 5.1,}\\
&\|t_5\| \leqslant \begin{cases} 0, & \text{if } d > 0,\\ Ch\|A_1\lambda\|_1, & \text{if } d = 0,\end{cases} && \text{by Lemma 5.6,}\\
&\|t_6\| \leqslant C\|\lambda - \tilde{\lambda}\|_a \leqslant Ch\|A_1\lambda\|_1, && \text{by Lemmas 5.3 and 5.5.}
\end{aligned}$$

Combining these estimates for all t_i, we obtain

$$\begin{aligned}
\|\vec{\mathrm{Q}}\lambda + a\vec{\nabla}P_0\lambda\|^2 &\leqslant Ch^2\|A_1\lambda\|_1^2 + C\gamma_{h\tau}^{(1)} h^{2s} |\tilde{u}|_{H^{1+s}(\Omega)}^2 + Ch^{2s}(|\tilde{q}|_{H^s(\Omega)} + \max_{K\in\mathscr{T}_h} \tau_K^* |\tilde{u}|_{H^s(\Omega)})^2\\
&\leqslant Ch^2\|A_1\lambda\|_1^2 + C(1 + \max_{K\in\mathscr{T}_h} (\tau_K^*)^2) h^{2s} \|f_\lambda\|_{H^{-1+s}(\Omega)}^2
\end{aligned} \tag{5.32}$$

by the regularity assumption (5.1). Since $H^{-1+s}(\Omega)$ is an interpolation space (Bergh & Löfström, 1976) in the scale of intermediate spaces between $H^{-1}(\Omega)$ and $L^2(\Omega)$, we know that

$$\|f_\lambda\|_{H^{-1+s}(\Omega)} \leqslant \|f_\lambda\|_{H^{-1}(\Omega)}^{1-s} \|f_\lambda\|^s.$$

Therefore,

$$\|f_\lambda\|_{H^{-1+s}(\Omega)} \leqslant C\left(\gamma_{h\tau}^{(1)} \|\lambda\|_a\right)^{1-s} \|A_1\lambda\|_1^s,$$

by (5.22) and (5.20) of Lemma 5.5. Using this bound in (5.32), we obtain the estimate of the lemma. □

THEOREM 5.3 If Assumption 5.1 holds with some $s > 1/2$, then there is a constant C_τ independent of h, whose value increases with $\tau_{\max}^* := \max_{K\in\mathscr{T}_h} \tau_K^*$, such that

1. in the $d > 0$ case, Assumption 4.4 holds true with $\alpha = s$ and $C_1 = C_\tau$, and

2. in the $d = 0$ case, Assumption 4.4 holds true with $\alpha = s/2$ and $C_1 = C_\tau$.

Proof. By the identities (5.7) and (5.8) of Lemma 5.1, we know that

$$a_1((I-I_1P_0)\lambda,\lambda) \leqslant \begin{cases} C\|\vec{\mathcal{Q}}\lambda + a\vec{\nabla}P_0\lambda\|_c\|\lambda\|_a + \|\mathcal{U}\lambda-\lambda\|_\tau^2, & \text{if } d=0,\\ \|\vec{\mathcal{Q}}\lambda + a\vec{\nabla}P_0\lambda\|_c^2 + \|\mathcal{U}\lambda-\lambda\|_\tau^2, & \text{if } d>0. \end{cases} \tag{5.33}$$

Here, for the $d=0$ case, we have used the Cauchy-Schwarz inequality and Lemma 5.3.

Since the terms involving $\vec{\mathcal{Q}}\lambda + a\vec{\nabla}P_0\lambda$ can be bounded as in Lemma 5.7, let us first investigate the remaining term involving $\mathcal{U}\lambda-\lambda$. To this end, the following inequality will be helpful:

$$\|\mathcal{U}(\lambda-I_1P_0\lambda)-(\lambda-I_1P_0\lambda)\|_{\tau,\partial K} \leqslant C\sqrt{\tau_K^* h_K}\,\|\vec{\mathcal{Q}}^{\mathrm{RT}}(\lambda-I_1P_0\lambda)\|_K.$$

This is due to (3.10) of Lemma 3.4. By item (ii) of the same lemma, we also know that $\vec{\mathcal{Q}}^{\mathrm{RT}}(I_1P_0\lambda) = -a\vec{\nabla}P_0\lambda$. Thus

$$\begin{aligned}\|\mathcal{U}(\lambda-I_1P_0\lambda)-(\lambda-I_1P_0\lambda)\|_{\tau,\partial K} &= C\sqrt{\tau_K^* h_K}\,\|\vec{\mathcal{Q}}^{\mathrm{RT}}\lambda + a\vec{\nabla}(P_0\lambda)\|_K\\ &\leqslant C\sqrt{\tau_K^* h_K}\,\|\vec{\mathcal{Q}}\lambda + a\vec{\nabla}(P_0\lambda)\|_K,\end{aligned} \tag{5.34}$$

where the last inequality holds because of the identity $\vec{\mathcal{Q}}^{\mathrm{RT}}\lambda + a\vec{\nabla}(P_0\lambda) = J_K(\vec{\mathcal{Q}}\lambda + a\vec{\nabla}(P_0\lambda))$. This identity follows from Lemma 3.4(v), Assumption 5.1(i), and the observation that constant vector fields on K are in the range of J_K.

Now consider the case $d>0$. By Lemma 3.4 (iii), we know that $\mathcal{U}(I_1P_0\lambda) = P_0\lambda$, so

$$\begin{aligned}\|\mathcal{U}\lambda-\lambda\|_{\tau,\partial K} &= \|\mathcal{U}(\lambda-I_1P_0\lambda)-(\lambda-I_1P_0\lambda)\|_{\tau,\partial K}\\ &\leqslant C\sqrt{\tau_K^* h_K}\,\|\vec{\mathcal{Q}}\lambda + a\vec{\nabla}(P_0\lambda)\|_K,\end{aligned}$$

by (5.34). Using this in (5.33), we have

$$\begin{aligned}a_1((I-I_1P_0)\lambda,\lambda) &\leqslant C\gamma_{h\tau}^{(1)}\|\vec{\mathcal{Q}}\lambda + a\vec{\nabla}(P_0\lambda)\|^2\\ &\leqslant C\gamma_{h\tau}^{(1)}\big(h^2\|A_1\lambda\|_1^2 + C\gamma_\tau h^{2s}a_h(\lambda,\lambda)^{1-s}\|A_1\lambda\|_1^{2s}\big).\end{aligned}$$

by Lemma 5.7. Since Theorem 3.2 shows that

$$h^2 \leqslant \frac{\gamma_{h\tau}^{(2)}}{\rho(A_1)} \tag{5.35}$$

the above inequality after obvious manipulations, implies that

$$a_h((I-I_1P_0)\lambda,\lambda) \leqslant C_\tau\left(\frac{\|A_1\lambda\|_1^2}{\rho(A_1)}\right)^s a_h(\lambda,\lambda)^{1-s},$$

for some C_τ that is an increasing function of $\tau_{\max}^*$. This proves the inequality of Assumption 4.4 for $d>0$.

Finally, consider the $d=0$ case. Since $\mathcal{U}(I_1P_0\lambda)$ and $P_0\lambda$ do not coincide in general, we estimate $\mathcal{U}\lambda-\lambda$ differently as follows. By the inequalities (iv) and (v) of Lemma 3.4,

$$\begin{aligned}\|\mathcal{U}\lambda-\lambda\|_{\tau,\partial K} &\leqslant C\sqrt{\tau_K^* h_K}\,\|\vec{\mathcal{Q}}^{\mathrm{RT}}\lambda\|_K\\ &\leqslant C\sqrt{\tau_K^* h_K}\,\|\vec{\mathcal{Q}}\lambda\|_K\\ &\leqslant C\sqrt{\tau_K^* h_K}\,\big(\|\vec{\mathcal{Q}}\lambda + a\vec{\nabla}(P_0\lambda)\|_K + \|\vec{\nabla}(P_0\lambda)\|_K\big)\end{aligned}$$

Using this in (5.33) we obtain

$$a_1((I-I_1P_0)\lambda,\lambda) \leqslant C\|\vec{\mathrm{Q}}\lambda + a\vec{\nabla}P_0\lambda\|_c\|\lambda\|_a + C\gamma_{h\tau}^{(1)}\|\vec{\mathrm{Q}}\lambda + a\vec{\nabla}P_0\lambda\|^2 + C\tau^*_{\max}h\|\vec{\nabla}P_0\lambda\|^2.$$

The right hand side can be bounded using Lemma 5.7 and Lemma 5.3, to get

$$a_1((I-I_1P_0)\lambda,\lambda) \leqslant C\gamma_{h\tau}^{(1)}(r_1+r_2+r_3) \tag{5.36}$$

where

$$\begin{aligned} r_1 &= h\|A_1\lambda\|_1 a_h(\lambda,\lambda)^{1/2},\\ r_2 &= \gamma_\tau h^s a_h(\lambda,\lambda)^{(1-s)/2}\|A_1\lambda\|_1^s a_h(\lambda,\lambda)^{1/2},\\ r_3 &= \tau^*_{\max} h\, a_h(\lambda,\lambda). \end{aligned}$$

The first two terms can be bounded using (5.35) as follows.

$$\begin{aligned} r_1 &\leqslant C\left(\frac{\|A_1\lambda\|_1^2}{\rho(A_1)}\right)^{1/2} a_h(\lambda,\lambda)^{1/2}\\ r_2 &\leqslant C\gamma_\tau\left(\frac{\|A_1\lambda\|_1^2}{\rho(A_1)}\right)^{s/2} a_h(\lambda,\lambda)^{1-(s/2)}. \end{aligned}$$

To bound r_3, we use

$$\begin{aligned} h\,a_h(\lambda,\lambda) &\leqslant h\|\lambda\|_h\|A_1\lambda\|_h && \text{by the Cauchy-Schwarz inequality}\\ &\leqslant \left(\frac{c_2\gamma_{h\tau}^{(2)}}{\rho(A_1)}\right)^{1/2}\|A_1\lambda\|_h\,\|\lambda\|_h && \text{by the upper bound of Theorem 3.2}\\ &\leqslant C_\tau\left(\frac{\|A_1\lambda\|_h^2}{\rho(A_1)}\right)^{1/2} a_h(\lambda,\lambda)^{1/2} && \text{by the lower bound of Theorem 3.2} \end{aligned}$$

for some constant C_τ whose value increases with $\tau^*_{\max}$. Using these estimates in (5.36), we finish the proof. □

Proof of Theorem 5.1. Apply Theorem 4.2 for the $d>0$ case and Theorem 4.3 for the $d=0$ case. The assumptions of these theorems have been verified for the HDG setting in Theorems 5.2 and 5.3. □

6. Numerical results

In this section we report numerical experiments illustrating out theoretical results. We begin by displaying history of convergence plots that confirm that the approximations λ_h, u_h and $\vec{q}_h$ provided by the HDG method converge with order $d+1$ for fixed τ. We then explore the numerical efficacy of our multigrid algorithm in terms of the stabilization parameter τ and mesh size h.

For all the experiments, we started with a coarse mesh generated by a public domain meshing software TRIANGLE (Shewchuk, 1996), and then produced a sequence of refinements by connecting the midpoints of edges, as explained before. The domain and the first two meshes are shown in Figure 1.

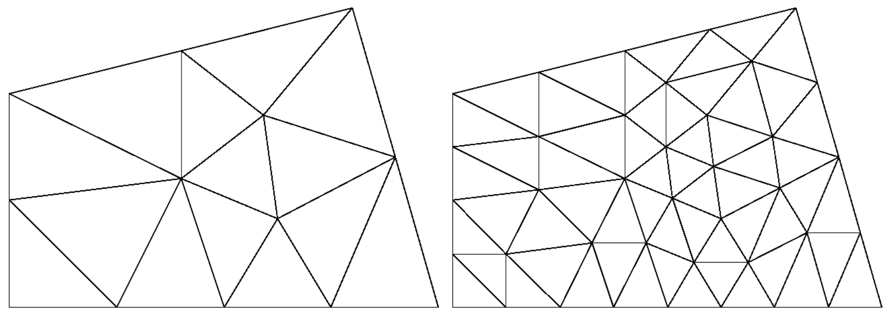

FIG. 1. The initial mesh (left), and one refinement (right). Corner coordinates are $(0,0),(1,0),(0.8,0.7)$ and $(0,0.5)$.

We consider numerically solving the Dirichlet problem (2.1) on the finest mesh level $\mathscr{T}_1$ for various choices of $-J$. The problem is chosen such that the solution is $u(x,y) = e^y \sin x$. As suggested by Theorem 2.1, we solve (2.10) for λ_h, and subsequently recover u_h and $\vec{q}_h$ through the local solvers. We consider two cases $d = 0$ and $d = 1$, i.e., the cases where λ_h is approximated by piecewise constant and piecewise linear functions, respectively, on mesh edges. The multigrid iteration is then carried out on the matrix system $Ax = b$ resulting from (2.10) for both cases.

In Figure 2, we display the history of convergence of the HDG method for different values of the stabilization function τ which we take constant on $\partial\mathscr{T}_h$. As expected from the results in Subsection 3.4, for all the choices of τ, we see first and second order of convergence is achieved for $d = 0$ and $d = 1$, respectively.

In order to study the iteration errors in our multigrid cycle we design the first experiment as follows. We set $b = 0$, so that the exact solution of $Ax = b$ is $x = 0$. The initial iterate x_0 in the multigrid iteration (Algorithm 4.4) on each multilevel space M_{-k}, $k = 1,\ldots,J$ is set to be $I_1 I_0 \cdots I_{-J+1} v$, where v is the function in the coarsest space M_{-J} which equals one on every interior mesh node (and of course, is linear on all mesh elements, is continuous across elements, and decreases to 0 on the boundary). We use one Gauss-Seidel sweep as the smoother. We use one Gauss-Seidel sweep for the pre-smoothing iteration in Algorithm 4.4 and another sweep in reversed order (the adjoint of Gauss-Seidel) for the post-smoothing. We stop iterations when $\|x_i - x_{i-1}\|_a \leqslant C_{bd}\|x_0 - x\|_a$ or when the iteration count reaches 99, whichever comes first, where $C_{bd} = 10^{-6}$ for $d = 0$ and 10^{-8} for $d = 1$.

The results for $d = 0$ for various choices of τ are presented in Table 1. As can be seen from both tables, for each fixed τ, the number of iterations quickly appears to approach to a constant number on all the subsequent meshes. This illustrates the efficacy of our multigrid algorithm. The corresponding average error reduction rates are also reported in Table 2. For those entries marked with "*", we report the average number for the first 99 iterations (the stopping criterion). The existence of such cases is in agreement with the smallness condition on τ for convergence of the multigrid method in Theorem 5.1.

The results for the case $d = 1$ are presented in Table 3 and in Table 4. Full agreement with Theorem 5.1 is observed.

A. Proof of Theorem 3.4

First of all, note that the identity (3.24) is proved in (Cockburn *et al.*, 2010, Lemma 3.1), so we need only prove the estimates of the theorem. When s_u and s_q are natural numbers the estimates have already

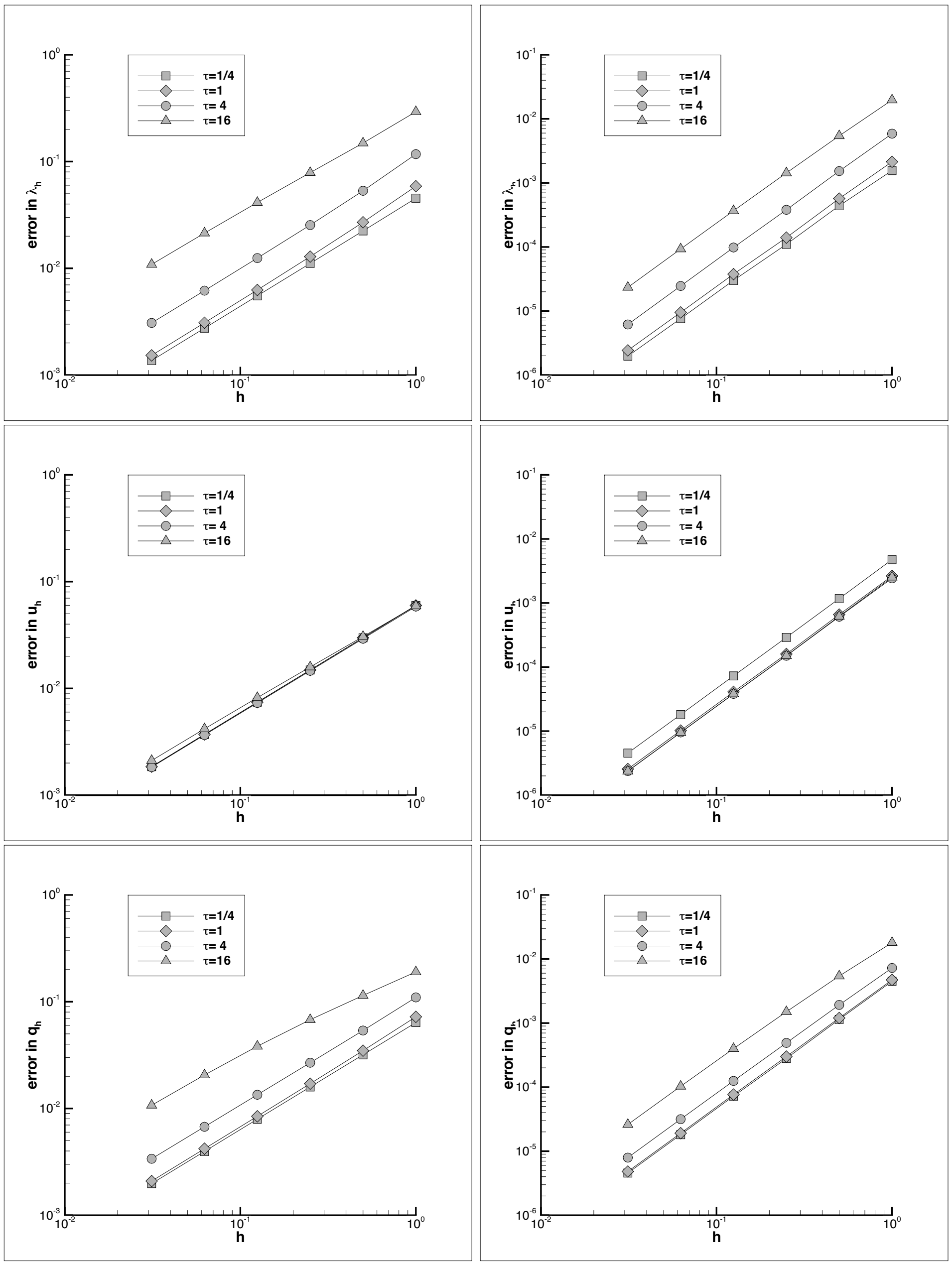

FIG. 2. History of convergence of the HDG method for $d = 0$ (left) and $d = 1$ (right) in terms of the stabilization parameter τ. Displayed are the quantities $\|Pu-\lambda_h\|_a$ (top), $\|u-u_h\|_{L^2}$ (middle) and $\|\vec{q}-\vec{q}_h\|_{L^2}$ (bottom). First and second order of convergence is achieved for each of these quantities for $d = 0$ and $d = 1$, respectively.

mesh	1	2	3	4	5	6
$\tau = 1/4$	13	13	14	16	15	14
$\tau = 1$	12	12	14	15	15	14
$\tau = 4$	11	11	14	15	15	14
$\tau = 16$	14	17	12	13	14	14
$\tau = 64$	*	*	*	31	12	12
$\tau = 256$	*	*	*	*	*	42

Table 1. The number of multigrid iterations ($m_k = 1$) for HDG(d=0) with different τ. The symbol * indicates excessively large number (or divergence).

mesh	1	2	3	4	5	6
$\tau = 1/4$	0.10	0.11	0.14	0.17	0.15	0.14
$\tau = 1$	0.09	0.10	0.14	0.16	0.15	0.14
$\tau = 4$	0.07	0.08	0.13	0.15	0.15	0.14
$\tau = 16$	0.14	0.19	0.09	0.12	0.13	0.13
$\tau = 64$	3.68	4.39	1.61	0.40	0.09	0.10
$\tau = 256$	63.3	74.0	28.3	8.34	2.35	0.52

Table 2. The average error reduction rate for HDG(d=0) with different τ.

mesh	1	2	3	4	5	6
$\tau = 1/4$	88	75	74	71	69	66
$\tau = 1$	86	75	74	71	69	66
$\tau = 4$	82	74	72	70	68	66
$\tau = 16$	73	69	69	68	67	65
$\tau = 64$	64	63	63	63	63	63
$\tau = 256$	59	58	56	56	56	57

Table 3. The number of multigrid iterations ($m_k = 1$) for HDG(d=1) with different τ.

mesh	1	2	3	4	5	6
$\tau = 1/4$	0.69	0.65	0.64	0.63	0.62	0.61
$\tau = 1$	0.68	0.64	0.64	0.63	0.62	0.61
$\tau = 4$	0.67	0.64	0.63	0.63	0.62	0.61
$\tau = 16$	0.64	0.62	0.62	0.62	0.61	0.60
$\tau = 64$	0.60	0.59	0.59	0.59	0.59	0.59
$\tau = 256$	0.57	0.57	0.56	0.55	0.56	0.56

Table 4. The average error reduction rate for HDG(d=1) with different τ.

been proved in (Cockburn *et al.*, 2010), so this appendix is devoted only to proving them for fractional s_u and s_q. Let us begin with an observation, whose simple proof we omit.

LEMMA A.1 If u_d denotes the $L^2(K)$-orthogonal projection of u into $P_d(K)$, then

$$\|u-u_d\|_{\partial K} \leqslant Ch_K^{s-1/2}\|u\|_{H^s(K)}$$

for all $u \in H^s(K)$ for all $s \in (1/2, d+1]$.

Proof of (3.23a) *and* (3.23b). From (Cockburn *et al.*, 2010, Proof of Proposition A.2), we know that with $\delta^u := \Pi_h^W u - u_d$, we have

$$\|\Pi_h^W u - u\|_K \leqslant \|u-u_d\|_K + \leqslant C\,\frac{h_K}{\tau_K^{\max}}\left(\|b_q\|+\|b_u\|\right). \tag{A.1}$$

where $b_q(w) := (\nabla\cdot\vec{q}, w)_K,\ b_u(w) := \langle \tau\,(u-u_d), w\rangle_{\partial K}$, and $\|b\| := sup_{w\in P_d^{\perp}(K)\setminus 0} b(w)/\|w\|_K$, where $P_d^{\perp} := w \in P_d(K) : (w,\zeta)_K = 0, \forall \zeta \in P_{d-1}(K)$.

Let us estimate $\|b_q\|$. It is proved in (Cockburn *et al.*, 2010) that

$$\|b_q\| \leqslant C\,h_K^{\ell_q}\,|\nabla\cdot\vec{q}|_{H^{\ell_q}(K)}, \tag{A.2}$$

for ℓ_q in $[0,d]$. We claim that (A.2) also holds for all $\ell_q \in (-1/2, 0]$. Indeed, for such ℓ_q,

$$b_q(w) = (\nabla\cdot\vec{q}, w)_K \leqslant \|\nabla\cdot\vec{q}\|_{H^{\ell_q}(K)}\|w\|_{H^{-\ell_q}(K)} \leqslant C\|\nabla\cdot\vec{q}\|_{H^{\ell_q}(K)} h_K^{\ell_q}\|w\|_K$$

where we have used a local inverse inequality for w. Hence (A.2) holds for all $\ell_q \in (-1/2, k]$. This, together with the fact that differentiation is a continuous operator from $H^{s+1}(K)$ into $H^s(K)$ for all real s (Grisvard, 1985), implies that

$$\|b_q\| \leqslant C\,h_K^{s_q-1}\,|\vec{q}|_{H^{s_q}(K)}, \tag{A.3}$$

for all $s_q \in (1/2, k+1]$ (identifying $\ell_q + 1 = s_q$).

For b_u, we first note that

$$b_u(w) \leqslant \tau_K^{\max}\,\|u-u_d\|_{\partial K}\,\|w\|_{\partial K} \leqslant C\,h_K^{-1/2}\,\tau_K^{\max}\,\|u-u_d\|_{\partial K}\,\|w\|_K.$$

By Lemma A.1,

$$\|b_u\| \leqslant C\,\tau_K^{\max}\,h_K^{s_u-1}\,|u|_{H^{s_u}(K)}, \tag{A.4}$$

Using (A.4) and (A.3) in (A.1), and a standard estimate for the L^2 projection u_d, we complete the proof of (3.23b).

To prove (3.23a), we again follow along the lines of (Cockburn *et al.*, 2010, Proposition A.3) to find that

$$\|\Pi_h^V\vec{q} - \vec{q}\,\|_K \leqslant \|B_h^V\vec{q} - \vec{q}\,\|_K + C\tau_K^* h_K^{1/2}\|\Pi_h^W u - u\|_{\partial K}$$

where B_h^V is the projection introduced in (Cockburn & Dong, 2007). Now, we can modify the proof of (Cockburn & Dong, 2007, Lemma 3.3) to extend the validity of the estimate

$$\|B_h^V\vec{q} - \vec{q}\,\|_K \leqslant\ C h_K^{s_q}\,|\vec{q}|_{H^{s_q}(K)} \qquad \forall\, s_q \in [1, d+1],$$

to $s_q \in (1/2, d+1]$, exactly as done above, and complete the proof using (3.23b) and the fact that $\tau_K^* \leqslant \tau_K^{\max}$. □

B. Proofs of Theorems 4.2 and 4.3

Proof of Theorem 4.2. This proof is a modification of one in (Bramble *et al.*, 1991), so we will be brief, highlighting only our modifications. Let $E_1 = I - B_1A_1$ and $E_0 = I - B_0A_0$. It is clear from Algorithm 4.1 that

$$E_1 = (I - R_1^t A_1)(I - I_1 B_0 Q_0 A_1)(I - R_1 A_1).$$

It is easy to see (Bramble, 1993) that $Q_0A_1 = A_0P_0$, the adjoint of $K = I - R_1A_1$ with respect to $a_1(\cdot,\cdot)$-inner product is $K^a = I - R_1^t A_1$, and consequently E_1 is self-adjoint with respect to $a_1(\cdot,\cdot)$. Since E_1 is the error reducing operator of Algorithm 4.1, it suffices to prove that

$$0 \leqslant a_1(E_1\mu,\mu) \leqslant \delta_1 a_1(\mu,\mu), \qquad \forall \mu \in M_1, \tag{A.1}$$

with δ_1 as stated in the theorem.

The starting point is the following identity:

$$\begin{aligned} a_1(E_1\mu,\mu) &= a_1((I - I_1B_0A_0P_0)K\mu, K\mu) \\ &= a_1((I - I_1P_0)K\mu, K\mu) + a_0(E_0P_0K\mu, P_0K\mu). \end{aligned} \tag{A.2}$$

By Assumption 4.1, the last term is non-negative. So is the first term on the right hand side due to Assumption 4.2. Hence the lower inequality of (A.1) is proved.

For the upper bound of (A.1), we use the well known consequence (see, e.g., Bramble, 1993) of Assumption 4.5 that

$$\frac{\|A_1K\mu\|_1^2}{\rho(A_1)} \leqslant \frac{1}{\omega}\big(a_1(\mu,\mu) - a_1(K\mu,K\mu)\big).$$

Combining with Assumption 4.4, we have

$$\begin{aligned} a_1((I - I_1P_0)K\mu, K\mu) &\leqslant C_1\left(\frac{1}{\omega}\big(a_1(\mu,\mu) - a_1(K\mu,K\mu)\big)\right)^{\alpha} a_1(K\mu,K\mu)^{1-\alpha} \\ &= f(t)\, a_1(\mu,\mu) \end{aligned} \tag{A.3}$$

where $t = a_1(K\mu,K\mu)/a_1(\mu,\mu)$ and $f(t) = \frac{C_1}{\omega^\alpha}(1-t)^\alpha t^{1-\alpha}$. Thus, by (A.2) and Assumption 4.1, we have for any number $0 \leqslant \delta < 1$,

$$\begin{aligned} a_1(E_1\mu,\mu) &= (1-\delta+\delta)a_1((I - I_1P_0)K\mu, K\mu) + a_0(E_0P_0K\mu, P_0K\mu) \\ &\leqslant (1-\delta)f(t)a_1(\mu,\mu) + \delta a_1(K\mu,K\mu) + (\delta_0 - \delta)a_0(P_0K\mu, P_0K\mu) \\ &= g(t)\, a_1(\mu,\mu) + (\delta_0 - \delta)a_0(P_0K\mu, P_0K\mu), \end{aligned} \tag{A.4}$$

where $g(t) = (1-\delta)f(t) + \delta t$.

Introducing a positive number ε to be chosen shortly, and using the arithmetic-geometric mean inequality,

$$\begin{aligned} g(t) &= (1-\delta)\frac{C_1}{\omega^\alpha}\big(\varepsilon^{-(1-\alpha)/\alpha}(1-t)\big)^{\alpha}\big(\varepsilon t\big)^{1-\alpha} + \delta t \\ &\leqslant (1-\delta)\frac{C_1}{\omega^\alpha}\big((1-\alpha)\varepsilon t + \alpha\varepsilon^{-(1-\alpha)/\alpha}(1-t)\big) + \delta t \end{aligned}$$

Since the right hand side is linear in t, and $0 \leqslant t \leqslant 1$ by the smoothing assumption, its maximum is achieved at $t = 0$ or $t = 1$. Thus

$$g(t) \leqslant \max\left((1-\delta)\frac{C_1}{\omega^\alpha}\alpha\varepsilon^{-(1-\alpha)/\alpha},\ \delta+(1-\delta)\frac{C_1}{\omega^\alpha}(1-\alpha)\varepsilon\right). \tag{A.5}$$

Now choose ε small enough so that

$$\frac{C_1}{\omega^\alpha}(1-\alpha)\varepsilon \equiv \eta < 1, \tag{A.6}$$

e.g., $\varepsilon \leqslant \omega^\alpha/2C_1(1-\alpha)$. Then, with this ε, let $C_2 = \frac{C_1}{\omega^\alpha}\alpha\varepsilon^{-(1-\alpha)/\alpha}$, so that (A.5) becomes

$$g(t) \leqslant \max\big((1-\delta)C_2,\ \delta+(1-\delta)\eta\big). \tag{A.7}$$

Next, set

$$\delta = \max(\delta_0, \frac{C_2}{1+C_2}).$$

Then $(1-\delta)C_2 \leqslant \delta$, so the maximum in (A.7) is achieved by the second argument. Furthermore, since $\delta \geqslant \delta_0$, the last term in (A.4) is negative. Consequently,

$$a_1(E_1\mu,\mu) \leqslant (\delta+(1-\delta)\eta)a_1(\mu,\mu).$$

Setting $\delta_1 = \delta+(1-\delta)\eta$, and noting that $\eta < 1$ by (A.6), we have proved (A.1) with $\delta_1 < 1$. □

Proof of Theorem 4.3. In the previous proof, we used Assumption 4.2 to obtain the lower inequality in (A.1). This is the only argument that needs modification, since we can now only assume Assumption 4.3 instead. The proof of the upper bound proceeds exactly as before yielding a $\delta_1 < 1$ such that $a_1(E_1\mu,\mu) \leqslant \delta_1 a_1(\mu,\mu)$.

We claim that if $h_1 < \delta_1/C_0 \equiv H$, then

$$-\delta_1 a_1(\mu,\mu) \leqslant a_1(E_1\mu,\mu) \leqslant \delta_1 a_1(\mu,\mu), \qquad \forall \mu \in M_1, \tag{A.8}$$

This is because by (A.2) and Assumption 4.1,

$$\begin{aligned} a_1(E_1\mu,\mu) &\geqslant a_1((I-I_1P_0)K\mu, K\mu) \\ &= a_1(K\mu,K\mu) - a_0(P_0K\mu, P_0K\mu). \end{aligned}$$

By Assumption 4.3, $a_0(P_0K\mu,P_0K\mu) \leqslant (1+C_0h_1)a_1(K\mu,K\mu)$, cf. Lemma 5.3, and by the smoothing properties, $a_1(K\mu,K\mu) \leqslant a_1(\mu,\mu)$. Thus,

$$a_1(E_1\mu,\mu) \geqslant -C_0h_1a_1(\mu,\mu),$$

and the claim follows, thus finishing the proof. □

REFERENCES

ANTONIETTI, P. F. & AYUSO DE DIOS, B. (2007) Schwarz domain decomposition preconditioners for discontinuous Galerkin approximations of elliptic problems: non-overlapping case. *M2AN Math. Model. Numer. Anal.*, **41**, 21–54.

ARNOLD, D. N., BREZZI, F., COCKBURN, B. & MARINI, L. D. (2001/02) Unified analysis of discontinuous Galerkin methods for elliptic problems. *SIAM J. Numer. Anal.*, **39**, 1749–1779 (electronic).

AYUSO DE DIOS, B. & ZIKATANOV, L. (2009) Uniformly convergent iterative methods for discontinuous Galerkin discretizations. *J. Sci. Comput.*, **40**, 4–36.

BERGH, J. & LÖFSTRÖM, J. (1976) *Interpolation Spaces. An Introduction*. New York: Springer-Verlag.

BRAMBLE, J. H., PASCIAK, J. E. & XU, J. (1991) The analysis of multigrid algorithms with nonnested spaces or noninherited quadratic forms. *Math. Comp.*, **56**, 1–34.

BRAMBLE, J. H. (1993) *Multigrid Methods*. Pitman research notes in mathematics series. Longman Scientific & Technical, Harlow, UK.

BRAMBLE, J. H. & ZHANG, X. (2000) The analysis of multigrid methods. Amsterdam: North-Holland, pp. 173–415.

BRENNER, S. C. (1999) Convergence of nonconforming multigrid methods without full elliptic regularity. *Math. Comp.*, **68**, 25–53.

BRENNER, S. C. (2004) Convergence of nonconforming V-cycle and F-cycle multigrid algorithms for second order elliptic boundary value problems. *Math. Comp.*, **73**, 1041–1066 (electronic).

BRENNER, S. C. & SUNG, L.-Y. (2006) Multigrid algorithms for C^0 interior penalty methods. *SIAM J. Numer. Anal.*, **44**, 199–223 (electronic).

BRENNER, S. C. & ZHAO, J. (2005) Convergence of multigrid algorithms for interior penalty methods. *Appl. Numer. Anal. Comput. Math.*, **2**, 3–18.

BREZZI, F. & FORTIN, M. (1991) *Mixed and Hybrid Finite Element Methods*. Springer Series in Computational Mathematics. New York: Springer-Verlag.

CIARLET, P. G. (1978) *The Finite Element Method for Elliptic Problems*. Amsterdam: North-Holland Publishing Company.

COCKBURN, B., DONG, B. & GUZMÁN, J. (2008) A superconvergent LDG-hybridizable Galerkin method for second-order elliptic problems. *Math. Comp.*, **77**, 1887–1916.

COCKBURN, B., GUZMÁN, J. & WANG, H. (2009a) Superconvergent discontinuous Galerkin methods for second-order elliptic problems. *Math. Comp.*, **78**, 1–24.

COCKBURN, B., GOPALAKRISHNAN, J. & LAZAROV, R. (2009b) Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. *SIAM Journal on Numerical Analysis*, **47**, 1319–1365.

COCKBURN, B., GOPALAKRISHNAN, J. & SAYAS, F.-J. (2010) A projection-based error analysis of HDG methods. *Math. Comp.*, **79**, 1351–1367.

COCKBURN, B. & DONG, B. (2007) An analysis of the minimal dissipation local discontinuous Galerkin method for convection-diffusion problems. *J. Sci. Comput.*, **32**, 233–262.

COCKBURN, B. & GOPALAKRISHNAN, J. (2004) A characterization of hybridized mixed methods for the Dirichlet problem. *SIAM J. Numer. Anal.*, **42**, 283–301.

COCKBURN, B. & GOPALAKRISHNAN, J. (2005) Error analysis of variable degree mixed methods for elliptic problems via hybridization. *Math. Comp.*, **74**, 1653–1677 (electronic).

DAUGE, M. (1988) *Elliptic boundary value problems on corner domains: smoothness and asymptotics of solutions*. Lecture notes in mathematics. Berlin: Springer-Verlag.

DEMKOWICZ, L., GOPALAKRISHNAN, J. & SCHÖBERL, J. (2012) Polynomial extension operators. Part III. *Math. Comp.*, **81**, 1289–1326.

DOBREV, V. A., LAZAROV, R. D., VASSILEVSKI, P. S. & ZIKATANOV, L. T. (2006) Two-level preconditioning of discontinuous Galerkin approximations of second-order elliptic equations. *Numer. Linear Algebra Appl.*, **13**, 753–770.

DUAN, H.-Y., GAO, S.-Q., TAN, R. C. E. & ZHANG, S. (2007) A generalized BPX multigrid framework covering nonnested V-cycle methods. *Math. Comp.*, **76**, 137–152 (electronic).

FENG, X. & KARAKASHIAN, O. A. (2005) Two-level non-overlapping Schwarz preconditioners for a discontinuous Galerkin approximation of the biharmonic equation. *J. Sci. Comput.*, **22/23**, 289–314.

GOPALAKRISHNAN, J. (2003) A Schwarz preconditioner for a hybridized mixed method. *Computational Methods in Applied Mathematics*, **3**, 116–134.

GOPALAKRISHNAN, J. & KANSCHAT, G. (2003a) Application of unified DG analysis to preconditioning DG methods. *Computational Fluid and Solid Mechanics 2003* (K. J. Bathe ed.). MIT: Elsevier, pp. 1943–1945.

GOPALAKRISHNAN, J. & KANSCHAT, G. (2003b) A multilevel discontinuous Galerkin method. *Numer. Math.*, **95**, 527–550.

GOPALAKRISHNAN, J. & TAN, S. (2009) A convergent multigrid cycle for the hybridized mixed method. *Numer. Linear Algebra Appl.*, **16**, 689–714.

GRISVARD, P. (1985) *Elliptic Problems in Nonsmooth Domains*. Monographs and Studies in Mathematics. Marshfield, Massachusetts: Pitman Advanced Publishing Program.

KELLOGG, R. B. (1971) Interpolation between subspaces of a Hilbert space. *Technical Note* BN-719. College Park, Maryland: University of Maryland.

KRAUS, J. K. & TOMAR, S. K. (2008a) A multilevel method for discontinuous Galerkin approximation of three-dimensional anisotropic elliptic problems. *Numer. Linear Algebra Appl.*, **15**, 417–438.

KRAUS, J. K. & TOMAR, S. K. (2008b) Multilevel preconditioning of two-dimensional elliptic problems discretized by a class of discontinuous Galerkin methods. *SIAM J. Sci. Comput.*, **30**, 684–706.

NITSCHE, J. (1968) Ein Kriterium für die Quasi-Optimalität des Ritzschen Verfahrens. *Numer. Math.*, **11**, 346–348.

SCOTT, L. R. & ZHANG, S. (1990) Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, **54**, 483–493.

SHEWCHUK, J. R. (1996) Triangle: Engineering a 2D Quality Mesh Generator and Delaunay Triangulator. *Applied Computational Geometry: Towards Geometric Engineering* (M. C. Lin & D. Manocha eds). Lecture Notes in Computer Science, vol. 1148. New York: Springer-Verlag, pp. 203–222.

TOSELLI, A. & WIDLUND, O. (2005) *Domain decomposition methods—algorithms and theory*. Springer Series in Computational Mathematics, vol. 34. Berlin: Springer-Verlag.

XU, J. (1990) Convergence estimates for some multigrid algorithms. *Third International Symposium on Domain Decomposition Methods for Partial Differential Equations (Houston, TX, 1989).* Philadelphia, PA: SIAM, pp. 174–187.

XU, J. (1996) The auxiliary space method and optimal multigrid preconditioning techniques for unstructured grids. *Computing*, **56**, 215–235.