IMA Journal of Numerical Analysis (2022) **42**, 1050–1086
https://doi.org/10.1093/imanum/draa102
Advance Access publication on 20 January 2021

An *a priori* error analysis of adjoint-based super-convergent Galerkin approximations of linear functionals

BERNARDO COCKBURN AND SHIQIANG XIA*
School of Mathematics, University of Minnesota, Minneapolis, MN 55455, USA
*Corresponding author: xiaxx268@umn.edu

[Received on 27 December 2019; revised on 19 August 2020]

We present the first *a priori* error analysis of a new method proposed in Cockburn & Wang (2017, Adjoint-based, superconvergent Galerkin approximations of linear functionals. *J. Comput. Sci.*, 73, 644–666), for computing adjoint-based, super-convergent Galerkin approximations of linear functionals. If $J(u)$ is a smooth linear functional, where u is the solution of a steady-state diffusion problem, the standard approximation $J(u_h)$ converges with order h^{2k+1}, where u_h is the Hybridizable Discontinuous Galerkin approximation to u with polynomials of degree $k > 0$. In contrast, numerical experiments show that the new method provides an approximation that converges with order h^{4k}, and can be computed by only using twice the computational effort needed to compute $J(u_h)$. Here, we put these experimental results in firm mathematical ground. We also display numerical experiments devised to explore the convergence properties of the method in cases not covered by the theory, in particular, when the solution u or the functional $J(\cdot)$ are not very smooth. We end by indicating how to extend these results to the case of general Galerkin methods.

Keywords: approximation of linear functionals; adjoint-based error correction; super-convergence; Galerkin methods; convolution; filtering.

1. Introduction

This is the second of a series of papers aiming to devise super-convergent Galerkin approximations to functionals. Here, we provide the very first *a priori* error analysis of the adjoint-based method proposed in the previous paper of this series, see Cockburn & Wang (2017).

In many scientific and engineering applications, one is often interested in the approximations of certain quantities that can be treated mathematically as functionals $J(\cdot)$. These functionals are typically determined by one or several field variables $u(x)$, which are governed by partial differential equations. For example, in fluid mechanics, the lift and drag forces of an object in a viscous incompressible fluid can be the quantities of interest. The fluid flow is governed by the Navier–Stokes equations and the lift and drag forces are surface integrals related to the stress tensor. Since the functional $J(\cdot)$ is used to quantify efficiency and performance of engineering design, its accurate approximation is of significant importance. In recent decades, numerous studies have been done on error control and error estimation for approximating functionals. In Giles *et al.* (1997), they considered the approximation of lift and drag coefficients. In Monk & Süli (1998), they studied a functional arising in electromagnetics. See also the review of Fidkowski & Darmofal (2011).

To approximate a functional $J(u)$, the standard way is to obtain a numerical approximation u_h of the exact solution u and then use $J(u_h)$ as an approximation of the functional. The adjoint-based method we study here improves the accuracy of these approximations in two ways. First, by increasing the accuracy of the numerical solution u_h by means of the filtering technique of Bramble & Schatz (1977). Then, by

© The Author(s) 2021. Published by Oxford University Press on behalf of the Institute of Mathematics and its Applications. All rights reserved.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

modifying the formula to approximate the functional by using the adoint-correction approach of Pierce & Giles (2000). Let us briefly describe these two components of the method.

The first component is a convolution filter that improves the accuracy of the Galerkin approximation. This local post-processing technique was first proposed by Bramble & Schatz (1977) for finite element methods for second-order elliptic equations. It takes advantage of the well-known fact that the Galerkin solution must oscillate around the exact solution in a certain pattern because of the Galerkin orthogonality property. Hence, convolving the Galerkin solution with a specific B-spine kernel filters out these oscillations and provides a more accurate solution. Bramble & Schatz (1977) showed how this takes place in a subdomain $\Omega_0 \subset\subset \Omega$ included in a set where the mesh is translation invariant. This approach was later extended to discontinuous Galerkin methods for hyperbolic equations in Cockburn *et al.* (2003).

Although this convolution filter has been proven to be effective only on a strict subdomain Ω_0 of the original domain Ω, or with periodic boundary conditions, one can still apply the convolution in the whole domain Ω, and even when the meshes are not locally translation invariant. Thus, Kirby, Ryan and their collaborators have further extended this technique to various types of meshes including unstructured triangular meshes, see Curtis *et al.* (2007), Mirzaee *et al.* (2011, 2013, 2012, 2014), King *et al.* (2012), Ji *et al.* (2014), Li *et al.* (2019). In their work, this technique is referred to as smoothness-increasing, accuracy-conserving (SIAC) filtering technique. Also, to apply the filtering in the whole domain, one-side kernel and position-dependent kernel approaches have been studied in Ryan *et al.* (2003), van Slingerland *et al.* (2011), Ji *et al.* (2014), Ryan *et al.* (2015). The adjoint-based method considered here keeps the original symmetric B-spline kernel employed by Bramble & Schatz (1977) to define a more accurate approximation in Ω_0. To extend it to the whole domain Ω, an auxiliary problem is solved on $\Omega \setminus \Omega_0$.

The second component of the adjoint-based method for approximating functionals more accurately is the adjoint-correction method of Pierce & Giles (2000), see the review in Giles & Süli (2002). Roughly speaking, this powerful technique consists in numerically solving the adjoint problem for the functional $J(\cdot)$ so that, an extra, computable correction term can be added to $J(u_h)$, which results in a much better approximation.

The idea of using the adjoint problem has been studied for decades. For example, the dual-weighted residual (DWR) method uses the adjoint problems to devise *a posteriori* error estimates and mesh adaptivity algorithms, see Eriksson *et al.* (1995); Becker & Rannacher (2001), and the references therein. Note that the DWR method and the adjoint-correction method we consider here are different. To obtain the *a posteriori* error for the DWR method, the exact adjoint solution or a very high-order accurate approximation (obtained by a higher-order method or a finer mesh) is needed. In contrast, the adjoint-correction method we consider here does not require a more accurate approximation for the adjoint problem, and is not used within an adaptive algorithm driven by an *a posteriori* error estimate.

In Cockburn & Wang (2017), these two component were put together, which resulted in the adjoint-error correction method we are analysing here. Therein, numerical experiments in one and two dimensional spaces were carried out. The hybridizable discontinuous Galerkin (HDG) method with polynomial degree $k > 0$ was used to obtain an approximation u_h of the exact solution u. The results indicate that the approximation defined by this new adjoint-based method converges to the functional $J(u)$ with order h^{4k}. Compared to the standard approximation $J(u_h)$, which converges with order h^{2k+1}, this new method essentially doubles the order of convergence by only doubling the computational effort for obtaining $J(u_h)$.

The use of the method under consideration with adaptive methods will be considered elsewhere. Here, we only study the convergence properties of the method when the meshes are translation invariant

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

in most of the *interior* of the domain Ω. Moreover, although the method we analyze here can be applied to quite general functionals, as argued in Pierce & Giles (2000); Giles & Süli (2002); Cockburn & Wang (2017), our results are for the following simple model problem. The functional we are considering here is

$$J(u) := \int_{\Omega} g(x)u(x)\,\mathrm{d}x, \tag{1.1}$$

where u is the solution of a steady-state diffusion equation

$$-\Delta u = f \quad \text{in } \Omega, \tag{1.2a}$$

$$u = u_D \text{ on } \partial\Omega \tag{1.2b}$$

on a bounded domain Ω with Lipschitz continuous boundary $\partial\Omega$. For the error analysis, we need to assume that the domain Ω is $(k+2)$-regular. The precise definition of this standard elliptic regularity property is given in §2.3.

The remainder of this paper is organized as follows. In Section 2, we define the adjoint-based method. We then state and briefly discuss our main result, the *a priori* error estimates of Theorem 2.6. Their proof is detailed in Section 3. In Section 4, we present numerical results designed to validate the theory, and to explore the convergence properties of the method in cases not covered by the theory. Finally, in Section 5, we conclude and describe possible extensions of our results.

2. Main results

This section is devoted to stating and briefly discussing our main results.

2.1 *The components of the methods*

We begin by defining the three components of the adjoint-based approximations, namely, the Galerkin method (which we take to be the HDG method), the adjoint-correction method, and the technique of filtering by convolution. We would then be ready to define the adjoint-based method.

2.1.1 *The HDG Method.* We use the HDG method here because it has a general structure that allows us to extend the results to a large class of Galerkin methods, including the mixed finite element method and the continuous Galerkin method.

To define the HDG method, we first partition the domain Ω into elements K forming a conforming mesh $\mathscr{T}_h$. We set $\partial\mathscr{T}_h := \{\partial K : K \in \mathscr{T}_h\}$ and let $\mathscr{F}_h$ denote the set of faces F of all the elements $K \in \mathscr{T}_h$. We also let $\mathscr{F}(K)$ denote the set of all faces F of the element $K \in \mathscr{T}_h$.

We rewrite the model elliptic problem (1.2) as a system of first-order equations:

$$\boldsymbol{q} + \nabla u = 0 \qquad \text{in } \Omega, \tag{2.1a}$$

$$\nabla \cdot \boldsymbol{q} = f \qquad \text{in } \Omega, \tag{2.1b}$$

$$u = u_D \qquad \text{on } \Omega. \tag{2.1c}$$

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

The HDG method seeks an approximation, $(\boldsymbol{q}_h, u_h, \widehat{u}_h)$, to the exact solution of the model problem (2.1a), $(\boldsymbol{q}|_{\Omega}, u|_{\Omega}, u|_{\mathscr{F}_h})$, in the finite dimensional space $\boldsymbol{V}_h \times W_h \times M_h$, where

$$
\begin{aligned}
\boldsymbol{V}_h &:= \{\boldsymbol{v} \in \boldsymbol{L}^2(\mathscr{T}_h) : \boldsymbol{v}|_K \in \boldsymbol{V}(K) \quad \forall K \in \mathscr{T}_h\},\\
W_h &:= \{w \in L^2(\mathscr{T}_h) : w|_K \in W(K) \quad \forall K \in \mathscr{T}_h\},\\
M_h &:= \{\mu \in L^2(\mathscr{F}_h) : \mu|_F \in M(F) \quad \forall F \in \mathscr{F}_h\}.
\end{aligned}
$$

If we use notation

$$
(u, v)_{\mathscr{T}_h} := \sum_{K \in \mathscr{T}_h} \int_K uv \ \mathrm{d}x \quad \text{and } \langle v\,,\, w\rangle_{\partial \mathscr{T}_h} := \sum_{K \in \mathscr{T}_h} \int_{\partial K} vu \ \mathrm{d}s,
$$

then the HDG approximation is determined as the solution of the following weak formulation:

$$
(\boldsymbol{q}_h, \boldsymbol{r})_{\mathscr{T}_h} - (u_h, \nabla \cdot \boldsymbol{r})_{\mathscr{T}_h} + \langle \widehat{u}_h\,,\, \boldsymbol{r} \cdot \boldsymbol{n}\rangle_{\partial \mathscr{T}_h} = 0, \tag{2.2a}
$$

$$
-(\boldsymbol{q}_h, \nabla w)_{\mathscr{T}_h} + \langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}\,,\, w\rangle_{\partial \mathscr{T}_h} = (f, w)_{\mathscr{T}_h}, \tag{2.2b}
$$

$$
\langle \widehat{u}_h\,,\, \mu\rangle_{\partial \Omega} = \langle u_D\,,\, \mu\rangle_{\partial \Omega}, \tag{2.2c}
$$

$$
\langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}\,,\, \mu\rangle_{\partial \mathscr{T}_h \setminus \partial \Omega} = 0, \tag{2.2d}
$$

for all $(\boldsymbol{r}, w, \mu) \in \boldsymbol{V}_h \times W_h \times M_h$, where

$$
\widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n} := \boldsymbol{q}_h \cdot \boldsymbol{n} + \tau(u_h - \widehat{u}_h) \ \text{ on } \partial \mathscr{T}_h.
$$

We obtain different methods by choosing different local spaces $\boldsymbol{V}(K)$, $W(K)$ and

$$
M(\partial K) := \{\mu \in \boldsymbol{L}^2(\partial K) : \mu|_F \in M(F) \quad \forall F \in \mathscr{F}(K)\},
$$

and stabilization functions τ. The hybridized version of mixed methods is obtained by simply taking τ equal to zero, as pointed out in Cockburn *et al.* (2009). Since we consider the HDG method in a fairly general setting, we make the following assumptions.

Assumption 2.1 When we say we use *the HDG method with polynomial approximation of degree* $k \geqslant 0$ for a collection of meshes $\{\mathscr{T}_h\}_{h>0}$, we assume we choose proper local spaces $\boldsymbol{V}(K) \times W(K) \times M(\partial K)$ and τ such that the approximation errors in L^2 norm, satisfy

$$
\|u - u_h\|_{\mathscr{T}_h} \leqslant Ch^{k+1}, \tag{2.3a}
$$

$$
\|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathscr{T}_h} \leqslant Ch^{k+1}, \tag{2.3b}
$$

for smooth enough solution u and $\boldsymbol{q}$, where C is a constant independent of h.

Table 1 *Local spaces $\boldsymbol{V}(K) \times W(K)$ admitting an $M(\partial K)$-decomposition*

Element	$\boldsymbol{V}(K)$	$W(K)$	$M(\partial K)$
Triangle	$\mathscr{P}_k(K)$	$\mathscr{P}_k(K)$	$\mathscr{P}_k(\partial K)$
Quadrilateral	$\mathscr{P}_k(K) \oplus \boldsymbol{curl}\ span\{\xi_4\lambda_3^k, \xi_4\lambda_4^k\}$		
Square	$\mathscr{Q}_k(K) \oplus \boldsymbol{curl}\ span\{x^{k+1}y, xy^{k+1}\}$	$\mathscr{Q}_k(K)$	$\mathscr{P}_k(\partial K)$
Tetrahedron	$\mathscr{P}_k(K)$	$\mathscr{P}_k(K)$	$\mathscr{P}_k(\partial K)$
Prisms	$\mathscr{P}_k(K) \oplus \boldsymbol{curl}\ span\left\{\begin{array}{c} z^{k+1}(x\nabla y - y\nabla x), \\ z\widetilde{\mathscr{P}}_k(x,y)(x\nabla y - y\nabla x) \end{array}\right\}$	$\mathscr{P}_k(K)$	$\mathscr{P}_k(\partial K)$
Cube	$\mathscr{Q}_k(K) \oplus \boldsymbol{curl}\ span\left\{\begin{array}{c} x^k y z^{k+1}\nabla x, \\ x^{k+1}z\nabla y, \\ x^{k+1}y^k z\nabla y \\ (1-x)x(1-z)z^k\nabla y, \\ (1-x)x(1-y)y^k\nabla z, \\ (1-x)x(1-y)y^k z^k\nabla z \end{array}\right\}$	$\mathscr{Q}_k(K)$	$\mathscr{Q}_k(\partial K)$

Assumption 2.2 For the HDG method with polynomial approximation of degree k, there exists an element-wise auxiliary projection $\Pi_h(\boldsymbol{q}, u) := (\boldsymbol{\Pi_V q}, \Pi_W u) \in \boldsymbol{V}(K) \times W(K)$, called *the HDG projection*, such that the so-called *weak commutativity property*

$$(\nabla \cdot \boldsymbol{q}, w)_K = (\nabla \cdot \boldsymbol{\Pi_V q}, w)_K + \langle \tau(\Pi_W u - u)\,,\, w\rangle_{\partial K} \quad \forall\, w \in W(K), \tag{2.4a}$$

and the following property

$$(u, \nabla \cdot \boldsymbol{r})_K = (\Pi_W u, \nabla \cdot \boldsymbol{r}) \quad \forall \boldsymbol{r} \in \boldsymbol{V}(K), \tag{2.4b}$$

are satisfied, and the HDG projection has the optimal approximation properties

$$\|u - \Pi_W u\|_K \leqslant Ch^{k+1}, \tag{2.5a}$$

$$\|u - \Pi_W u\|_{\partial K} \leqslant Ch^{k+\frac{1}{2}}, \tag{2.5b}$$

$$\|\boldsymbol{q} - \boldsymbol{\Pi_V q}\|_K \leqslant Ch^{k+1}, \tag{2.5c}$$

$$\|(\boldsymbol{q} - \boldsymbol{\Pi_V q}) \cdot \boldsymbol{n}\|_{\partial K} \leqslant Ch^{k+\frac{1}{2}}, \tag{2.5d}$$

for smooth enough solution u and $\boldsymbol{q}$, where C is a constant independent of h.

These two assumptions do hold for local spaces that admit an M-decomposition, see Cockburn *et al.* (2017). For the sake of completeness, we give in Appendix A.1 the definition of an M-decomposition and in Appendix A.2 the definition of the HDG projection. For a few, simple shapes of elements, we give the local spaces that admit an M-decomposition in Table 1.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

In Table 1, $\mathscr{P}_k(K)$ denotes the space of polynomials in K with degree at most k, and $\mathscr{Q}_k(K)$ denotes the space of tensor product polynomials of degree at most k in each variable. Finally, $\widetilde{\mathscr{P}}_k(x,y)$ denotes the space of homogeneous polynomials of degree k in x and y. The same notation is used for the three-dimensional case. For a quadrilateral element K, let $\{\boldsymbol{v}_i\}_{i=1}^4$ be the set of vertices and let $\{\boldsymbol{e}_i\}_{i=1}^4$ be the set of edges, where the edge $\boldsymbol{e}_i$ connects $\boldsymbol{v}_i$ and $\boldsymbol{v}_{i+1}$, where we set $\boldsymbol{v}_5 = \boldsymbol{v}_1$. Then, for $1 \leqslant i \leqslant 4$, we define λ_i to be the linear function that vanishes on edge $\boldsymbol{e}_i$ and reaches maximum value 1 in the closure of K, and let ξ_i be a rational function such that $\xi_i|_{\boldsymbol{e}_i} \in \mathscr{P}(\boldsymbol{e}_i)$ and $\xi_i(\boldsymbol{v_j}) = \delta_{ij}$, where δ_{ij} is the Kronecker delta. More details and examples about the construction of M-decomposition spaces in two and three dimensions are provided in Cockburn & Fu (2017a,b); Cockburn *et al.* (2017).

2.1.2 *The adjoint-correction method.* The adjoint-correction method was proposed by Pierce & Giles (2000) for approximating functionals $J(u)$. Rather than simply using $J(u_h)$ as an approximation, this method obtains a new approximation $J_h(u_h)$ by adding a carefully devised computable term, AC_h, called *the adjoint-correction term*. Roughly speaking, this adjoint-correction term is zero or significantly smaller than the error $|J(u) - J(u_h)|$ when the numerical solution u_h is obtained by a Galerkin method. However, for a solution u_h^* that is not a Galerkin solution, including this term results in a better approximation to $J(u)$. Let us describe the adjoint-recovery method for the functional (1.1). We follow Cockburn & Wang (2017).

Let $(\boldsymbol{q}_h, u_h, \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}, \widehat{u}_h) \in \boldsymbol{L}^2(\mathscr{T}_h) \times H^1(\mathscr{T}_h) \times L^2(\partial\mathscr{T}_h) \times L^2(\mathscr{F}_h)$ be any approximation to $(\boldsymbol{q}, u, \boldsymbol{q} \cdot \boldsymbol{n}|_{\partial\mathscr{T}_h}, u|_{\partial\mathscr{T}_h})$ in the model problem (2.1a) such that $\langle \widehat{u}_h\,,\,\mu\rangle_{\partial\Omega} = \langle u_D\,,\,\mu\rangle_{\partial\Omega}$ for any $\mu \in L^2(\mathscr{F}_h)$. Similarly, consider the adjoint problem

$$\boldsymbol{p} + \nabla v = 0 \qquad \text{in } \Omega \tag{2.6a}$$

$$\nabla \cdot \boldsymbol{p} = g \qquad \text{in } \Omega \tag{2.6b}$$

$$v = 0 \qquad \text{on } \partial\Omega \tag{2.6c}$$

and let $(\boldsymbol{p}_h, v_h, \widehat{\boldsymbol{p}}_h \cdot \boldsymbol{n}, \widehat{v}_h) \in \boldsymbol{L}^2(\mathscr{T}_h) \times H^1(\mathscr{T}_h) \times L^2(\partial\mathscr{T}_h) \times L^2(\mathscr{F}_h)$ be any approximation to $(\boldsymbol{p}, v, \boldsymbol{p} \cdot \boldsymbol{n}|_{\partial\mathscr{T}_h}, v|_{\partial\mathscr{T}_h})$ in the adjoint problem (2.6) such that $\widehat{v}_h = 0$ on $\partial\Omega$. Next we define a new approximation of $J(u)$ as

$$J_h(u_h) = J(u_h) + AC_h, \tag{2.7}$$

where the adjoint-correction term is as follows

$$\begin{aligned} AC_h :=& (f, v_h)_{\mathscr{T}_h} + (\boldsymbol{q}_h, \nabla v_h)_{\mathscr{T}_h} - \langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}\,,\, v_h\rangle_{\partial\mathscr{T}_h} \\ &+ (\boldsymbol{q}_h + \nabla u_h, \boldsymbol{p}_h)_{\mathscr{T}_h} - \langle u_h - \widehat{u}_h\,,\, \boldsymbol{p}_h \cdot \boldsymbol{n}\rangle_{\partial\mathscr{T}_h} \\ &+ \langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}\,,\, \widehat{v}_h\rangle_{\partial\mathscr{T}_h\backslash\partial\Omega} \\ &+ \langle u_h - \widehat{u}_h\,,\, (\boldsymbol{p}_h - \widehat{\boldsymbol{p}}_h) \cdot \boldsymbol{n}\rangle_{\partial\mathscr{T}_h}. \end{aligned}$$

The derivation of AC_h is presented in Cockburn & Wang (2017), Theorem 2.1. For general integral functionals, including linear boundary integral functionals and nonlinear domain integral functionals, the derivation of AC_h is discussed in Pierce & Giles (2000); Giles & Süli (2002).

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

2.1.3 *Filtering the oscillations of Galerkin Approximations.* It is a well-known fact that numerical solutions defined by a Galerkin method must oscillate around the exact solutions in a certain pattern. This is the direct result of the so-called Galerkin orthogonality property. Bramble & Schatz (1977), showed that it is possible to filter out these oscillations by convolving the Galerkin solution with a B-spline kernel under the assumption that the test function spaces are translation invariant. As a result, a new approximation with a faster convergence rate can be obtained. We use this filtering technique to post-process a Galerkin solution u_h and obtain a more accurate approximation u_h^*.

Let us first recall the definition of the B-spline kernel in one dimension. Let χ be the function that is one on the interval $\left(-\frac{1}{2}, \frac{1}{2}\right)$ and zero outside of it, and set $\psi^{(0)} = \delta$, where δ is the Dirac delta function. The nth order B-spline $\psi^{(n)}$ is defined as the convolution of $\psi^{(n-1)}$ and χ, namely, $\psi^{(n)} = \psi^{(n-1)} * \chi$ for $n \geqslant 1$. The B-spline kernel is defined as $K_h^{2k}(x) := \frac{1}{h} K^{2k}\left(\frac{x}{h}\right)$, where

$$K^{2k}(x) := \sum_{r=-k}^{k} C_r \psi^{(k+1)}(x - r).$$

Here h is the size of the diameters of the locally-invariant mesh and k is a fixed integer, usually the polynomial degree used in the Galerkin method. The coefficients of the kernel C_r are determined by requiring that $p * K^{2k} = p$ for all polynomials p of degree at most $2k$.

Note that the support of $\psi^{(k+1)}(\cdot)$ is $\left[-\frac{k+1}{2}, \frac{k+1}{2}\right]$. Therefore, the support of the kernel K_h^{2k} is $\left[-\frac{3k+1}{2}h, \frac{3k+1}{2}h\right]$. This means that for a given point x, the filtering post-processing technique $u_h^*(x) = K_h^{2k} * u_h(x)$ involves only a fixed number of neighboring elements. It also shows that in general we can only apply this technique to regions that are at least $\frac{3k+1}{2}h$ away from the domain boundary. In the next subsection, we present how we overcome this limitation in the adjoint-based method. For the N-dimensional kernel, we simply define

$$K_h^{2k}(x) := \frac{1}{h^N} K_N^{2k}\left(\frac{x}{h}\right) \quad \text{where} \quad K_N^{2k}(x) := \prod_{i=1}^{N} K^{2k}(x_i).$$

2.2 *The two adjoint-based methods*

We are now ready to define the adjoint-based approximation of the functional $J(u)$. Let us start with two assumptions for the mesh.

2.2.1 *Assumptions for the meshes.*

ASSUMPTION 2.3 Let $\{\mathscr{T}_h\}_{h>0}$ be a family of shape-regular meshes of Ω. We assume that there exist subdomains $\Omega_0 \subset\subset \Omega_0' \subset\subset \Omega$ such that

1. $\mathscr{T}_h \cap \Omega_0'$ is a translation-invariant mesh with element size h_0.
2. Any element $K \in \mathscr{T}_h$ is fully contained in either Ω_0 or $\Omega_1 := \Omega \backslash \Omega_0$.
3. Any boundary element is allowed to have one curved face.

We denote $h := \min\{h_K | K \in \mathscr{T}_h\}$, where h_K is the diameter of an element $K \in \mathscr{T}_h$. To get the convergence results in Section 2.3, we also need the following assumptions on the elements lying on

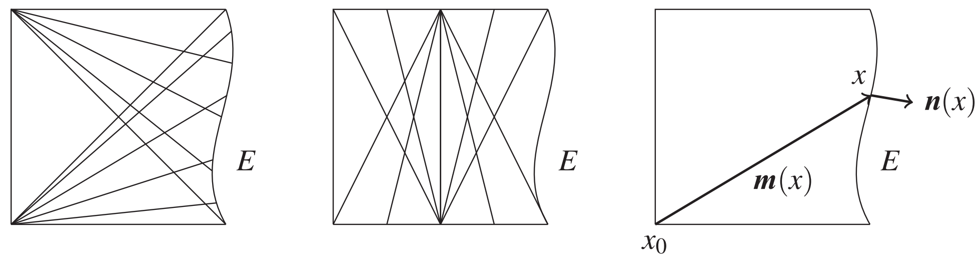

FIG. 1. Illustration of star-shapedness (left, middle) and shape-regularity (right).

the boundary; see Fig. 1 for an illustration of such elements. With these assumptions, the classic trace and inverse estimates for elements with flat faces can be extended to elements with curved faces; see Cangiani *et al.* (2018, 2019) for details.

ASSUMPTION 2.4 Let $\partial\Omega$ be Lipschitz continuous. Let K be a boundary element with one possibly curved face $E = K \cap \partial\Omega$. We assume that:

1. (*Star-shapedness*) The element K is star-shaped with respect to all vertices opposite to the face E. We assume K is also star-shaped with respect to all the midpoints of the edges sharing a common vertex with the face E.
2. (*Shape-regularity*) There is a constant c such that $\boldsymbol{m}(\boldsymbol{x}) \cdot \boldsymbol{n}(\boldsymbol{x}) \geqslant c|\boldsymbol{m}(\boldsymbol{x})|$ uniformly across the mesh, for every vector $\boldsymbol{m}(\boldsymbol{x}) = \boldsymbol{x} - \boldsymbol{x}_0$, with $\boldsymbol{x} \in E$ and $\boldsymbol{x}_0$ any vertex opposite E, and $\boldsymbol{n}(\boldsymbol{x})$ the unit outward normal vector of E at $\boldsymbol{x}$. We further assume $|\boldsymbol{m}(\boldsymbol{x})| = \mathscr{O}(h_K)$ uniformly for all boundary elements.

2.2.2 *The adjoint-based method.* To define the adjoint-based method, we need the following notation. We denote by $\mathscr{F}_h$ the set of all faces F of all elements $K \in \mathscr{T}_h$. Set, for $i = 0, 1$,

$$
\begin{aligned}
\mathscr{T}_{ih} &:= \{K \in \mathscr{T}_h \mid K \subset \Omega_i\},\\
\partial\mathscr{T}_{ih} &:= \{\partial K \in \mathscr{T}_h \mid K \subset \Omega_i\},\\
\mathscr{F}_{ih} &:= \{F \in \mathscr{F}(K) \mid K \in \mathscr{T}_{ih}\}.
\end{aligned}
$$

DEFINITION 2.5 We define the adjoint-based method in four steps:

1. On the whole mesh $\mathscr{T}_h$, let $(\boldsymbol{q}_h^k, u_h^k, \widehat{\boldsymbol{q}}_h^k \cdot \boldsymbol{n}, \widehat{u}_h^k)$ and $(\boldsymbol{p}_h^k, v_h^k, \widehat{\boldsymbol{p}}_h^k \cdot \boldsymbol{n}, \widehat{v}_h^k)$ be the approximations of the model problem (1.2) and adjoint problem (2.6), respectively, by the HDG method with polynomial degree $k \geqslant 1$.
2. Next, on $\mathscr{T}_{0h}$, we use the filtering technique described in Subsection 2.1.3 to obtain $u_h^* := K_h^{2k} * u_h^k$ and $v_h^* := K_h^{2k} * v_h^k$.
3. On $\mathscr{T}_{1h}$, let $(\boldsymbol{q}_h^{2k}, u_h^{2k}, \widehat{\boldsymbol{q}}_h^{2k} \cdot \boldsymbol{n}, \widehat{u}_h^{2k})$ and $(\boldsymbol{p}_h^{2k}, v_h^{2k}, \widehat{\boldsymbol{p}}_h^{2k} \cdot \boldsymbol{n}, \widehat{v}_h^{2k})$ be the approximations of the model problem and adjoint problem, respectively, by the HDG method with polynomial degree $2k$. To provide the boundary conditions on $\partial\Omega_1 \backslash \partial\Omega$, we define $\widehat{u}_h^{2k} := K_h^{2k} * u_h^k$ and $\widehat{v}_h^{2k} := K_h^{2k} * v_h^k$ to be the Dirichlet boundary conditions for the model and adjoint problems. Then we set the approximations $u_h^* := u_h^{2k}$ and $v_h^* := v_h^{2k}$ on $\mathscr{T}_{1h}$.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

4. Finally, we compute $J_h(u_h^*, v_h^*) = J(u_h^*) + AC_h$ as the adjoint-based approximation to $J(u)$, where

$$u_h^* := \begin{cases} K_h^{2k} * u_h^k & \text{in } K \in \mathscr{T}_{0h}, \\ u_h^{2k} & \text{in } K \in \mathscr{T}_{1h}, \end{cases}$$

and

$$v_h^* := \begin{cases} K_h^{2k} * v_h^k & \text{in } K \in \mathscr{T}_{0h}, \\ v_h^{2k} & \text{in } K \in \mathscr{T}_{1h}. \end{cases}$$

Note that the adjoint-correction term AC_h defined in Subsection 2.1.2 depends not only on the scalar approximations u_h^*, v_h^*, but also on their gradients and traces on each face. In the domain Ω_0, we use $(\nabla u_h^*, u_h^*, \nabla u_h^*|_{\partial \mathscr{T}_{0h}} \cdot \boldsymbol{n}, u_h^*|_{\mathscr{F}_{0h}})$ and $(\nabla v_h^*, v_h^*, \nabla v_h^*|_{\partial \mathscr{T}_{0h}} \cdot \boldsymbol{n}, v_h^*|_{\mathscr{F}_{0h}})$. In the domain Ω_1, we have two options:

2.2.3 *Method 1: Use the piecewise gradients in Ω_1.* The first method uses $(\nabla u_h^{2k}, u_h^{2k}, \widehat{\boldsymbol{q}}_h^{2k} \cdot \boldsymbol{n}, \widehat{u}_h^{2k})$ and $(\nabla v_h^{2k}, v_h^{2k}, \widehat{\boldsymbol{p}}_h^{2k} \cdot \boldsymbol{n}, \widehat{v}_h^{2k})$ on Ω_1. Since we use piecewise gradients, let us denote this method with superscript 'G'. We have

$$J_h^G(u_h^*, v_h^*) = J(u_h^*) + AC_h^G, \tag{2.8}$$

where

$$\begin{aligned} AC_h^G :=& (f, v_h^*)_{\mathscr{T}_h} - (\nabla u_h^*, \nabla v_h^*)_{\mathscr{T}_h} \\ &+ \langle \widehat{\boldsymbol{q}}_h^{2k} \cdot \boldsymbol{n} \,,\, \widehat{v}_h^{2k} - v_h^{2k} \rangle_{\partial \mathscr{T}_{1h}} \\ &+ \langle \widehat{u}_h^{2k} - u_h^{2k} \,,\, \widehat{\boldsymbol{p}}_h^{2k} \cdot \boldsymbol{n} \rangle_{\partial \mathscr{T}_{1h}}. \end{aligned}$$

2.2.4 *Method 2: Use the approximate fluxes in Ω_1.* For our second method, we use $(\boldsymbol{q}_h^{2k}, u_h^{2k}, \widehat{\boldsymbol{q}}_h^{2k} \cdot \boldsymbol{n}, \widehat{u}_h^{2k})$ and $(\boldsymbol{p}_h^{2k}, v_h^{2k}, \widehat{\boldsymbol{p}}_h^{2k} \cdot \boldsymbol{n}, \widehat{v}_h^{2k})$ on Ω_1, which come from our Galerkin approximation. Let us denote it by the superscript 'F'. We have

$$J_h^F(u_h^*, v_h^*) = J(u_h^*) + AC_h^F, \tag{2.9}$$

where

$$\begin{aligned} AC_h^F :=& (f, v_h^*)_{\mathscr{T}_{0h}} - (\nabla u_h^*, \nabla v_h^*)_{\mathscr{T}_{0h}} \\ &+ \langle \widehat{\boldsymbol{q}}_h^{2k} \cdot \boldsymbol{n} \,,\, \widehat{v}_h^{2k} \rangle_{\partial \Omega_1 \setminus \partial \Omega} \\ &+ \langle \widehat{u}_h^{2k} - u_h^{2k} \,,\, \boldsymbol{p}_h^{2k} - \widehat{\boldsymbol{p}}_h^{2k} \cdot \boldsymbol{n} \rangle_{\partial \mathscr{T}_{1h}}. \end{aligned}$$

For a detailed deduction of the above formulas, see Cockburn & Wang (2017).

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

2.3 *The* A Priori *Error Estimates*

We are now ready to present the *a priori* error estimates for the adjoint-based approximations. To do that, let us first introduce the notion of $(s+2)$-regularity of the domain Ω. For any integer $s \geqslant 0$, we say Ω is $(s+2)$-regular if the adjoint problem (2.6) is uniquely solvable for $g \in L^2(\Omega)$ and

$$\|\boldsymbol{p}\|_{s+1,\Omega} + \|v\|_{s+2,\Omega} \leqslant C\|g\|_{s,\Omega}$$

for all $g \in C_0^\infty(\Omega)$, where C only depends on the domain Ω. Note that this inequality does not necessarily hold for $g \in H^s(\Omega)$, but only for $g \in C_0^\infty(\Omega)$. For example, square domains are $(s+2)$-regular according to this definition, see (Nitsche & Schatz, 1974, Section 7, Example 3).

We can now state our main result.

THEOREM 2.6 Suppose that $k \geqslant 1$ and that Ω is $(k+2)$-regular. Let $u, v \in H^{2k+3}(\Omega)$. Let the HDG method satisfy Assumptions 2.1 and 2.2, where $\{\mathscr{T}_h\}_{h>0}$ is a family of meshes satisfying Assumptions 2.3 and 2.4. Finally, let $J_h^G(u_h^*, v_h^*)$ and $J_h^F(u_h^*, v_h^*)$ be the approximations of $J(u) = \int_\Omega g(x)u(x)\,\mathrm{d}x$ given by the adjoint-based methods (2.8) and (2.9), respectively. Then, for h small enough, there exists a constant C such that

$$|J(u) - J_h^G(u_h^*, v_h^*)| \leqslant C\,h^{4k},$$
$$|J(u) - J_h^F(u_h^*, v_h^*)| \leqslant C\,h^{4k},$$

where C is independent of h, but depends on u, v, k and the subdomain Ω_0.

Let us briefly discuss this result. First, we must note that, it is possible to show that

$$|J(u) - J(u_h)| \leqslant C\,h^{2k+1},$$

just by using the smoothness of the functional J, and by using that, for big enough values of $\ell \in \mathbb{N}$, the $H^{-\ell}(\Omega)$-norm of the error $u - u_h$ is of order h^{2k+1}. There is no need to assume that the meshes have to be translation invariant inside a fixed subdomain of Ω. However, if we do assume this, the above theorem states that, by computing an approximation v_h to the adjoint solution v, and effectively doubling the computational complexity, we can obtain an order of convergence of h^{4k}.

It is worth noting that the numerical experiments in Cockburn & Wang (2017) indicate that the order of convergence of $J(u_h)$, h^{2k+1}, and that of $J_h^G(u_h^*, v_h^*)$ and $J_h^G(u_h^*, v_h^*)$, h^{4k}, are actually *sharp*. The numerical experiments carried out there used HDG methods with local spaces admitting M-decompositions for triangular elements. Here, we use HDG methods with local spaces admitting M-decompositions for square elements. We verify that the theoretical results do hold for this case, even though in some cases, the observed orders of convergence are higher than the ones predicted.

3. *A priori* error analysis

In this section, we provide the proof of the error estimates of Theorem 2.6. We proceed in several steps. First, we recall the formula for the error from Cockburn & Wang (2017), and use it to get explicit expressions for the errors of the two methods we are considering. We then estimate the terms of the error associated to the *interior* domain Ω_0 and those associated to the *exterior* domain Ω_1. We conclude by putting those estimates together.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Step 1: A formula of the error

We begin by recalling a result that gives us an explicit formula of the approximation error.

THEOREM 3.1 (Cockburn & Wang (2017)). Let $J(u)$ be the linear functional defined in (1.1). Let $J_h(u_h)$ be the approximation defined in (2.7). Then we have that

$$J(u) = J_h(u_h) + E_h,$$

where

$$\begin{aligned} E_h := & (\boldsymbol{q} - \boldsymbol{q}_h, \boldsymbol{p} - \boldsymbol{p}_h)_{\mathscr{T}_h} \\ & + (\boldsymbol{q} - \boldsymbol{q}_h, \boldsymbol{p}_h + \nabla v_h)_{\mathscr{T}_h} + (\boldsymbol{q}_h + \nabla u_h, \boldsymbol{p} - \boldsymbol{p}_h)_{\mathscr{T}_h} \\ & + \langle (\widehat{\boldsymbol{q}}_h - \boldsymbol{q}) \cdot \boldsymbol{n} \,,\, v_h - \widehat{v}_h \rangle_{\partial \mathscr{T}_h} + \langle u_h - \widehat{u}_h \,,\, (\widehat{\boldsymbol{p}}_h - \boldsymbol{p}) \cdot \boldsymbol{n} \rangle_{\partial \mathscr{T}_h}. \end{aligned}$$

If we apply this result to the method (2.8), for which $J_h(u_h) = J_h^G(u_h^*, v_h^*)$, we get that

$$\begin{aligned} E_h^G := & J(u) - J_h^G(u_h^*, v_h^*) \\ = & (\boldsymbol{q} + \nabla K_h^{2k} * u_h^k, \boldsymbol{p} + \nabla K_h^{2k} * v_h^k)_{\mathscr{T}_{0h}} \\ & + (\boldsymbol{q} - \boldsymbol{q}_h^{2k}, \boldsymbol{p} - \boldsymbol{p}_h^{2k})_{\mathscr{T}_{1h}} \\ & + \langle (\widehat{\boldsymbol{q}}_h^{2k} - \boldsymbol{q}) \cdot \boldsymbol{n} \,,\, v_h^{2k} - \widehat{v}_h^{2k} \rangle_{\partial \mathscr{T}_{1h}} + \langle u_h^{2k} - \widehat{u}_h^{2k} \,,\, (\widehat{\boldsymbol{p}}_h^{2k} - \boldsymbol{p}) \cdot \boldsymbol{n} \rangle_{\partial \mathscr{T}_{1h}}, \end{aligned}$$

and if we apply it to the method (2.9), for which $J_h(u_h) = J_h^F(u_h^*, v_h^*)$, we obtain

$$\begin{aligned} E_h^F := & J(u) - J_h^G(u_h^*, v_h^*) \\ = & (\boldsymbol{q} + \nabla K_h^{2k} * u_h^k, \boldsymbol{p} + \nabla K_h^{2k} * v_h^k)_{\mathscr{T}_{0h}} \\ & + (\boldsymbol{q} - \boldsymbol{q}_h^{2k}, \boldsymbol{p} - \boldsymbol{p}_h^{2k})_{\mathscr{T}_{1h}} \\ & + (\boldsymbol{q} - \boldsymbol{q}_h^{2k}, \boldsymbol{p}_h^{2k} + \nabla v_h^{2k})_{\mathscr{T}_{1h}} + (\boldsymbol{q}_h^{2k} + \nabla u_h^{2k}, \boldsymbol{p} - \boldsymbol{p}_h^{2k})_{\mathscr{T}_{1h}} \\ & + \langle (\widehat{\boldsymbol{q}}_h^{2k} - \boldsymbol{q}) \cdot \boldsymbol{n} \,,\, v_h^{2k} - \widehat{v}_h^{2k} \rangle_{\partial \mathscr{T}_{1h}} + \langle u_h^{2k} - \widehat{u}_h^{2k} \,,\, (\widehat{\boldsymbol{p}}_h^{2k} - \boldsymbol{p}) \cdot \boldsymbol{n} \rangle_{\partial \mathscr{T}_{1h}}. \end{aligned}$$

Step 2: A basic result to get estimates involving convolutions

From the previous step, it is clear that we need to obtain estimates of functions of the form $w - K_h^{2k} * w_h$. To state the basic result we want, we need to introduce some standard notation.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

For any integer $m \geqslant 0$, open bounded set $\mathscr{D} \subset \mathbb{R}^N$ and sufficiently smooth function $u : \mathscr{D} \to \mathbb{R}$, we set

$$\|u\|_{m,\mathscr{D}} := \left(\sum_{0 \leqslant \ell \leqslant m} |u|^2_{\ell,\mathscr{D}} \right)^{\frac{1}{2}}, \quad |u|_{\ell,\mathscr{D}} := \left(\sum_{|\alpha|=\ell} \int_{\mathscr{D}} |D^\alpha u|^2 \, \mathrm{d}x \right)^{\frac{1}{2}},$$

$$\|u\|_{-m,\mathscr{D}} := \sup_{v \in C_0^\infty(\mathscr{D})} \frac{\int_{\mathscr{D}} uv \, \mathrm{d}x}{\|v\|_{m,\mathscr{D}}},$$

where $C_0^\infty(\mathscr{D})$ denotes the space of infinitely differentiable functions on $\mathscr{D}$ with compact support in $\mathscr{D}$. If $m = 0$, we simply write $\|u\|_{0,\mathscr{D}}$ as $\|u\|_{\mathscr{D}}$.

For any multi-index $\boldsymbol{\alpha} = (\alpha_1, ..., \alpha_N)$ and $h > 0$, let us define the difference quotient as

$$\partial_h^{\boldsymbol{\alpha}} u = \partial_{h,1}^{\alpha_1} \cdots \partial_{h,N}^{\alpha_N} u,$$

where

$$\partial_{h,j} u(x) = \frac{1}{h} \left(u\left(x + \frac{h}{2}\boldsymbol{e}_j\right) - u\left(x - \frac{h}{2}\boldsymbol{e}_j\right)\right),$$

and $\boldsymbol{e}_j$ is the unit vector with one in the jth entry and zero otherwise.

We are now ready to recall the following key result.

THEOREM 3.2 (Bramble & Schatz, 1977; Cockburn *et al.*, 2003). For a fixed integer $k \geqslant 1$, define $K_h^{2k}(x) := \frac{1}{h^N} K_N^{2k}(x/h)$ as described in 2.1.3. Let U be a function in $L^2(\Omega)$, where Ω is an open set in $\mathbb{R}^N$, and $u \in H^{2k+1}(\Omega)$. Let Ω_0 be an open subset of Ω, and there exists h_0 such that $\Omega_0 + 2supp(K_h^{2k}) \subset\subset \widetilde{\Omega}_0 \subset \Omega$ for any $h < h_0$. Then for any $h < h_0$, we have

$$\|u - K_h^{2k} * U\|_{\Omega_0} \leqslant C_1 h^{2k+1} |u|_{2k+1,\widetilde{\Omega}_0} + C_2 \sum_{|\boldsymbol{\alpha}| \leqslant k} \|\partial_h^{\boldsymbol{\alpha}}(u - U)\|_{-k,\widetilde{\Omega}_0},$$

where C_1 and C_2 are independent of u and h.

For this estimate to be useful, we need to obtain an estimate of negative-order norms of the difference quotient $\partial_h^{\boldsymbol{\alpha}}(u - U)$ for the case in which U is the component u_h of the approximation provided by the HDG method. The estimate we need is contained in the following result. Its proof follows from the duality argument and the observation that $\partial_h^{\boldsymbol{\alpha}} u_h$ satisfies the local HDG equations. The proof will be omitted, as it is a variation of that of the straight-faced elements.

LEMMA 3.3 Suppose $k \geqslant 1$ and Ω is $(k+2)$-regular. Let $\boldsymbol{\alpha}$ be a fixed multi-index, and assume the exact solution $u \in H^{k+2+|\boldsymbol{\alpha}|}(\Omega)$. Consider an HDG method with polynomial degree k satisfying Assumptions 2.1 and 2.2, for a family of meshes of Ω, $\{\mathscr{T}_h\}_{h>0}$, satisfying Assumptions 2.3 and 2.4. Then for h small enough we have

$$\|\partial_h^{\boldsymbol{\alpha}}(u - u_h)\|_{-k,\widetilde{\Omega}_0} \leqslant Ch^{k+l+1}(\|u\|_{l+2+|\boldsymbol{\alpha}|,\Omega} + \|u\|_{l+2,\Omega}),$$

where $\Omega_0 \subset\subset \widetilde{\Omega}_0 \subset\subset \Omega_0' \subset\subset \Omega$ and $0 \leqslant l \leqslant k$.

Step 3: Estimates of the terms defined in Ω_0.

The following result contains the estimates of the terms associated to the subdomain Ω_0.

LEMMA 3.4 We have

$$\|\boldsymbol{q} + \nabla K_h^{2k} * u_h^k\|_{\mathscr{T}_{0h}} \leqslant Ch^{2k+1}\|u\|_{2k+3,\Omega},$$

where C only depends on k, Ω_0, Ω_0' and Ω.

Proof. To prove the inequality, we begin by noting that,

$$\|\boldsymbol{q} + \nabla K_h^{2k} * u_h^k\|_{\mathscr{T}_{0h}} = \|\nabla u - \nabla K_h^{2k} * u_h^k\|_{\mathscr{T}_{0h}} \leqslant I + II,$$

where $I = \|\nabla u - K_h^{2k} * \nabla u\|_{\mathscr{T}_{0h}}$ and $II = \|\nabla(K_h^{2k} * (u - u_h^k))\|_{\mathscr{T}_{0h}}$. To estimate I, we apply Theorem 3.2 with $u := U := \partial_{x_i} u$, for each component $i = 1, \ldots, d$, to get that

$$\|\nabla u - K_h^{2k} * \nabla u\|_{\mathscr{T}_{0h}} \leqslant Ch^{2k+1}\|\nabla u\|_{2k+1,\Omega} \leqslant Ch^{2k+1}\|u\|_{2k+2,\Omega}.$$

It remains to get estimate for $II_i := \|\partial_{x_i}(K_h^{2k} * (u - u_h^k))\|_{\mathscr{T}_{0h}}$ for $i = 1, ..., d$. We have

$$\begin{aligned} II_i &\leqslant C \sum_{|\boldsymbol{\alpha}| \leqslant k} \|D^{\boldsymbol{\alpha}} \partial_{x_i}(K_h^{2k} * (u - u_h^k))\|_{-k,\widetilde{\Omega}_0} \quad \text{by (Bramble \& Schatz, 1977, Lemma 2.2)}, \\ &= C \sum_{|\boldsymbol{\alpha}| \leqslant k} \|D^{\boldsymbol{\alpha}+\boldsymbol{e}_i} K_h^{2k} * (u - u_h^k)\|_{-k,\widetilde{\Omega}_0}, \\ &\leqslant C \sum_{|\boldsymbol{\alpha}| \leqslant k} \|\partial_h^{\boldsymbol{\alpha}+\boldsymbol{e}_i}(u - u_h^k)\|_{-k,\widetilde{\Omega}_0'}, \end{aligned}$$

by (Bramble & Schatz, 1977, Lemma 5.3). Here, $\Omega_0 \subset\subset \widetilde{\Omega}_0 \subset\subset \widetilde{\Omega}_0' \subset\subset \Omega_0'$. Now we use the negative-order norm estimate of Lemma 3.3 for the HDG method with $l := k$ and $\boldsymbol{\alpha} := \boldsymbol{\alpha} + \boldsymbol{e}_i$ to get that, for h is small enough,

$$\|\partial_h^{\boldsymbol{\alpha}+\boldsymbol{e}_i}(u - u_h^k)\|_{-k,\widetilde{\Omega}_0'} \leqslant Ch^{2k+1}(\|u\|_{2k+3,\Omega} + \|u\|_{k+2,\Omega}),$$

since $|\boldsymbol{\alpha}| \leqslant k + 1$. This completes the proof. □

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Step 4: Estimates of the terms defined in Ω_1.

Let us now estimate the terms of the errors defined in Ω_1. In the following result, we gather all the estimates we need. It is stated in terms of norms we define next:

$$\|w\|_{\mathscr{T}_{1h}} := \left(\sum_{K\in\mathscr{T}_{1h}} \|w\|_K^2\right)^{1/2} \qquad \forall\, w \in L^2(\mathscr{T}_{1h}),$$

$$\|\mu\|_{h^\alpha,\partial\mathscr{T}_{1h}} := \left(\sum_{K\in\mathscr{T}_{1h}} h_K^\alpha\, \|\mu\|_{\partial K}^2\right)^{1/2} \qquad \forall\, \mu \in L^2(\partial\mathscr{T}_{1h}),$$

$$\|\mu\|_{h_0^\alpha,\partial\Omega_0} := \left(\sum_{F\in\partial\Omega_0} h_0^\alpha\, \|\mu\|_F^2\right)^{1/2} \qquad \forall\, \mu \in L^2(\partial\Omega_0),$$

where h_0 is the diameter of the elements of the translation-invariant mesh $\mathscr{T}_{0h}$ in Ω_0.

LEMMA 3.5 Let $(u_h^{2k}, \boldsymbol{q}_h^{2k}, \widehat{\boldsymbol{q}}_h^{2k}\cdot\boldsymbol{n}, \widehat{u}_h^{2k})$ be the HDG approximation with polynomial degree $2k$ in the domain Ω_1 as described in Definition 2.5. Let $\Pi_h(u,\boldsymbol{q}) = (\Pi_W u, \boldsymbol{\Pi_V q})$ be the corresponding HDG projection. Then for $k \geqslant 1$, we have the estimates of the errors

$$\|\boldsymbol{q} - \boldsymbol{q}_h^{2k}\|_{\mathscr{T}_{1h}} \leqslant C\,\Theta_1, \tag{3.1a}$$

$$\|\boldsymbol{q}\cdot\boldsymbol{n} - \widehat{\boldsymbol{q}}_h^{2k}\cdot\boldsymbol{n}\|_{h,\partial\mathscr{T}_{1h}} \leqslant C(\Theta_1 + \Theta_2 + \Theta_3), \tag{3.1b}$$

and the estimates of the residuals

$$\|\boldsymbol{q}_h^{2k} + \nabla u_h^{2k}\|_{\mathscr{T}_{1h}} \leqslant C(\Theta_1 + \Theta_2), \tag{3.1c}$$

$$\|u_h^{2k} - \widehat{u}_h^{2k}\|_{h^{-1},\partial\mathscr{T}_{1h}} \leqslant C(\Theta_1 + \Theta_2), \tag{3.1d}$$

where

$$\Theta_1 := \|\boldsymbol{\Pi_V q} - \boldsymbol{q}\|_{\mathscr{T}_{1h}} + \|P_M u - K_h^{2k} * u_h^k\|_{h_0^{-1},\partial\Omega_0},$$

$$\Theta_2 := \|u - \Pi_W u\|_{h^{-1},\partial\mathscr{T}_{1h}},$$

$$\Theta_3 := \|(\boldsymbol{\Pi_V q} - \boldsymbol{q})\cdot\boldsymbol{n}\|_{h,\partial\mathscr{T}_{1h}}.$$

Here $P_M u$ is the L^2 projection of u into M_h and the constant C depends on $k, u, \max_{K\in\mathscr{T}_{1h}}\{\tau_K^{max} h_K\}, \Omega_0$ and Ω_1.

The proof of this result entails carrying out a small modification of the standard *a priori* error analysis of the HDG method. The modification is due to the fact that the boundary condition on $\partial\Omega_0$ is given by the trace of $K_h^{2k} * u_h^k$ from Ω_0, not by the exact solution u. Since the proof is fairly long, we divide it in several steps.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Step i: The equations for the projection of the errors.

We begin by finding the equations satisfied by the projection of the errors, namely, $\varepsilon_h^q := \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} - \boldsymbol{q}_h^{2k}$, $\varepsilon_h^u := \Pi_W u - u_h^{2k}$ and $\varepsilon_h^{\widehat{u}} := u - \widehat{u}_h^{2k}$.

Using the properties of the HDG projection (2.4), by (Cockburn *et al.*, 2010, Lemma 3.1), we have the following error equations,

$$(\varepsilon_h^q, \boldsymbol{r})_{\mathscr{T}_{1h}} - (\varepsilon_h^u, \nabla\cdot\boldsymbol{r})_{\mathscr{T}_{1h}} + \langle \varepsilon_h^{\widehat{u}}, \boldsymbol{r}\cdot\boldsymbol{n}\rangle_{\partial\mathscr{T}_{1h}} = (\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} - \boldsymbol{q}, \boldsymbol{r})_{\mathscr{T}_h}, \tag{3.2a}$$

$$-(\varepsilon_h^q, \nabla w)_{\mathscr{T}_{1h}} + \langle \varepsilon_h^{\widehat{q}}\cdot\boldsymbol{n}, w\rangle_{\mathscr{T}_{1h}} = 0, \tag{3.2b}$$

$$\langle \varepsilon_h^{\widehat{u}}, \mu\rangle_{\partial\Omega_1} = \langle P_M u - K_h^{2k} * u_h^k, \mu\rangle_{\partial\Omega_0}, \tag{3.2c}$$

$$\langle \varepsilon_h^{\widehat{q}}\cdot\boldsymbol{n}, \mu\rangle_{\partial\mathscr{T}_{1h}\setminus\partial\Omega_1} = 0, \tag{3.2d}$$

for all $\boldsymbol{r}\in\boldsymbol{V}_h$, $w\in W_h$ and $\mu\in M_h$, where

$$\varepsilon_h^{\widehat{q}}\cdot\boldsymbol{n} = \varepsilon_h^q\cdot\boldsymbol{n} + \tau(\varepsilon_h^u - \varepsilon_h^{\widehat{u}}) \quad \text{on } \partial\mathscr{T}_{1h}. \tag{3.2e}$$

Note that the right-hand side of (3.2c) is not zero due to the fact that the Dirichlet condition on $\partial\Omega_0$ is given by the trace of $K_h^{2k} * u_h^k$ from Ω_0, not by the exact solution.

Step ii: Energy Argument.

To obtain the estimates of the errors, we use a standard energy argument. So, taking $\boldsymbol{r} := \varepsilon_h^q$ in (3.2a), $w := \varepsilon_h^u$ in (3.2b), $\mu := -\varepsilon_h^{\widehat{q}}\cdot\boldsymbol{n}$ in (3.2c) and $\mu := -\varepsilon_h^{\widehat{u}}$ in (3.2d) and adding the resulting four equations, we obtain

$$E_h := (\varepsilon_h^q, \varepsilon_h^q)_{\mathscr{T}_{1h}} + \langle \tau(\varepsilon_h^u - \varepsilon_h^{\widehat{u}}), (\varepsilon_h^u - \varepsilon_h^{\widehat{u}})\rangle_{\partial\mathscr{T}_{1h}} = T_1 + T_2,$$

where $T_1 := (\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} - \boldsymbol{q}, \varepsilon_h^q)_{\mathscr{T}_{1h}}$ and $T_2 := \langle P_M u - K_h^{2k} * u_h^k, -\varepsilon_h^{\widehat{q}}\cdot\boldsymbol{n}\rangle_{\partial\Omega_0}$.

Step iii: Estimates of the errors.

We estimate the first term as follows:

$$T_1 \leqslant \|\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} - \boldsymbol{q}\|_{\mathscr{T}_{1h}} \|\varepsilon_h^q\|_{\mathscr{T}_{1h}} \leqslant \|\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} - \boldsymbol{q}\|_{\mathscr{T}_{1h}} E_h^{1/2},$$

since τ is non-negative.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Let us now estimate the second term. We have

$$
\begin{aligned}
T_2 &\leqslant \left(\sum_{F \in \partial\Omega_0} h_0^{-1} \| P_M u - K_h^{2k} * u_h^k \|_F^2 \right)^{\frac{1}{2}} \left(\sum_{F \in \partial\Omega_0} h_0 \| \varepsilon_h^{\widehat{q}} \cdot \boldsymbol{n} \|_F^2 \right)^{\frac{1}{2}}, \\
&\leqslant \| P_M u - K_h^{2k} * u_h^k \|_{h_0^{-1}, \partial\Omega_0} \left(\sum_{K \in \Omega_1, \partial K \cap \partial\Omega_0 = F} h_K \| \varepsilon_h^{\widehat{q}} \cdot \boldsymbol{n} \|_F^2 \right)^{\frac{1}{2}}, \\
&\leqslant \| P_M u - K_h^{2k} * u_h^k \|_{h_0^{-1}, \partial\Omega_0} \| \varepsilon_h^{\widehat{q}} \cdot \boldsymbol{n} \|_{h, \partial\mathscr{T}_{1h}}.
\end{aligned}
$$

By the definition of $\varepsilon_h^{\widehat{q}} \cdot \boldsymbol{n}$, (3.2e), we have that

$$
\begin{aligned}
\| \varepsilon_h^{\widehat{q}} \cdot \boldsymbol{n} \|_{h, \partial\mathscr{T}_{1h}} &= \| \varepsilon_h^{q} \cdot \boldsymbol{n} + \tau (\varepsilon_h^u - \varepsilon_h^{\widehat{u}}) \|_{h, \partial\mathscr{T}_{1h}} \\
&\leqslant \| \varepsilon_h^q \|_{h, \partial\mathscr{T}_{1h}} + \max_{K \in \mathscr{T}_{1h}} (h_K \tau_K^{max})^{\frac{1}{2}} \langle \tau (\varepsilon_h^u - \varepsilon_h^{\widehat{u}}), (\varepsilon_h^u - \varepsilon_h^{\widehat{u}}) \rangle_{\mathscr{T}_{1h}}^{\frac{1}{2}} \\
&\leqslant C_{1,\tau} (\| \varepsilon_h^q \|_{\mathscr{T}_{1h}} + \langle \tau (\varepsilon_h^u - \varepsilon_h^{\widehat{u}}), (\varepsilon_h^u - \varepsilon_h^{\widehat{u}}) \rangle_{\mathscr{T}_{1h}}^{\frac{1}{2}}) = C_{1,\tau} E_h^{1/2},
\end{aligned}
$$

where $C_{1,\tau} = C \max\{1, (h_K \tau_K^{max})^{\frac{1}{2}}, K \in \mathscr{T}_{1h}\}$. Here we use the inverse inequality for polynomials. As a consequence, we get

$$
T_2 \leqslant C_{1,\tau} \| P_M u - K_h^{2k} * u_h^k \|_{h^{-1}, \partial\Omega_0} E_h^{1/2},
$$

and the first estimate (3.1a) immediately follows.

Let us obtain the second estimate. By the definition of $\varepsilon_h^{\widehat{q}} \cdot \boldsymbol{n}$, (3.2e), we have

$$
\begin{aligned}
\| \boldsymbol{q} \cdot \boldsymbol{n} - \widehat{\boldsymbol{q}}_h^{2k} \cdot \boldsymbol{n} \|_{h, \partial\mathscr{T}_{1h}} &\leqslant \| \boldsymbol{q} \cdot \boldsymbol{n} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q} \cdot \boldsymbol{n} - \tau (\Pi_W u - u) \|_{h, \partial\mathscr{T}_{1h}} + \| \varepsilon_h^{\widehat{q}} \cdot \boldsymbol{n} \|_{h, \partial\mathscr{T}_{1h}} \\
&\leqslant \| (\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}) \cdot \boldsymbol{n} \|_{h, \partial\mathscr{T}_{1h}} + \| \tau (\Pi_W u - u) \|_{h, \partial\mathscr{T}_{1h}} + \| \varepsilon_h^{\widehat{q}} \cdot \boldsymbol{n} \|_{h, \partial\mathscr{T}_{1h}} \\
&\leqslant \| (\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}) \cdot \boldsymbol{n} \|_{h, \partial\mathscr{T}_{1h}} + C_{2,\tau} \| \Pi_W u - u \|_{h^{-1}, \partial\mathscr{T}_{1h}} + \| \varepsilon_h^{\widehat{q}} \cdot \boldsymbol{n} \|_{h, \partial\mathscr{T}_{1h}},
\end{aligned}
$$

where $C_{2,\tau} = \max\{1, (h_K \tau_K^{max}), K \in \mathscr{T}_{1h}\}$. Then the estimate (3.1b) follows easily.

Step iv: Estimates of the residuals.

It remains to prove the estimates of the residuals. Integrating by parts in the first equation defining the HDG method in Ω_1, (2.2a), we get

$$
(\boldsymbol{q}_h^{2k} + \nabla u_h^{2k}, \boldsymbol{v})_{\mathscr{T}_{1h}} = \langle u_h^{2k} - \widehat{u}_h^{2k}, \boldsymbol{v} \cdot \boldsymbol{n} \rangle_{\partial\mathscr{T}_{1h}}.
$$

Taking $\boldsymbol{v} := \boldsymbol{q}_h^{2k} + \nabla u_h^{2k}$ on each element $K \in \mathscr{T}_{1h}$ and zero elsewhere, we obtain that

$$\begin{aligned}\|\boldsymbol{q}_h^{2k} + \nabla u_h^{2k}\|_K^2 &\leqslant \|u_h^{2k} - \widehat{u}_h^{2k}\|_{\partial K}\|(\boldsymbol{q}_h^{2k} + \nabla u_h^{2k}) \cdot \boldsymbol{n}\|_{\partial K}\\ &\leqslant Ch_K^{-\frac{1}{2}}\|u_h^{2k} - \widehat{u}_h^{2k}\|_{\partial K}\|\boldsymbol{q}_h^{2k} + \nabla u_h^{2k}\|_K,\end{aligned}$$

by an inverse inequality. Hence,

$$\|\boldsymbol{q}_h^{2k} + \nabla u_h^{2k}\|_K \leqslant Ch_K^{-\frac{1}{2}}\|u_h^{2k} - \widehat{u}_h^{2k}\|_{\partial K}.$$

And also note that

$$\begin{aligned}\|u_h^{2k} - \widehat{u}_h^{2k}\|_{h^{-1},\partial\mathscr{T}_{1h}} &= \|u_h^{2k} - \Pi_W u + \Pi_W u - u + u - \widehat{u}_h^{2k}\|_{h^{-1},\partial\mathscr{T}_{1h}}\\ &\leqslant \|\varepsilon_h^u - \varepsilon_h^{\widehat{u}}\|_{h^{-1},\partial\mathscr{T}_{1h}} + \|\Pi_W u - u\|_{h^{-1},\partial\mathscr{T}_{1h}}\\ &\leqslant \|\boldsymbol{q} - \boldsymbol{q}_h\|_{\mathscr{T}_{1h}} + \|\Pi_W u - u\|_{h^{-1},\partial\mathscr{T}_{1h}},\end{aligned}$$

where the estimate for $\|\varepsilon_h^u - \varepsilon_h^{\widehat{u}}\|_{h^{-1},\partial\mathscr{T}_{1h}}$ in the last inequality is proven in detail in Appendix A.4.

The estimates of the residuals easily follow. This completes the proof of Lemma 3.5.

Step v: Estimates of the auxiliary quantities Θ_1, Θ_2, and Θ_3.

To be able to conclude, it only remains to estimate the three terms that define Θ_1, Θ_2, and Θ_3 in the lemma of the previous step. The estimates are displayed in the following result.

LEMMA 3.6 We have

$$\begin{aligned}\|\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} - \boldsymbol{q}\|_{\mathscr{T}_{1h}} &\leqslant Ch^{2k+1}\|u\|_{2k+2,\Omega_1},\\ \|(\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} - \boldsymbol{q}) \cdot \boldsymbol{n}\|_{h,\partial\mathscr{T}_{1h}} &\leqslant Ch^{2k+1}\|u\|_{2k+2,\Omega_1},\\ \|u - \Pi_W u\|_{h^{-1},\partial\mathscr{T}_{1h}} &\leqslant Ch^{2k}\|u\|_{2k+2,\Omega_1},\\ \|P_M u - K_h^{2k} * u_h^k\|_{h_0^{-1},\partial\Omega_0} &\leqslant Ch_0^{2k}\|u\|_{2k+2,\Omega},\end{aligned}$$

where C depends on k, Ω_0, Ω_0' and Ω.

Proof. The first three inequalities follow from Assumption 2.1 and 2.2 about the HDG method and the properties of the HDG projection; see Appendix A.2. Let us prove the last estimate.

For each face F of the element $K \in \mathscr{T}_h \subset \Omega_0$, let $U_{F,K}$ be the function in $\mathscr{P}_{2k}(K)$ such that

$$\begin{aligned}&U_{F,K} = P_M u && \text{on } F,\\ &(U_{F,K} - u, w)_K = 0 && \forall\, w \in \mathscr{P}_{2k}(K) : w = 0 \text{ on } F.\end{aligned}$$

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Then,

$$
\begin{aligned}
T &:= \left(\sum_{F\in\partial\Omega_0} h_0 \, \|P_M u - K_h^{2k} * u_h^k\|_F^2 \right)^{\frac{1}{2}} = \left(\sum_{F\in\partial\Omega_0} h_0 \, \|U_{F,K} - K_h^{2k} * u_h^k\|_F^2 \right)^{\frac{1}{2}} \\
&\leqslant C \left(\sum_{K\in\Omega_0, \partial K\cap\partial\Omega_0 = F} \|U_{F,K} - K_h^{2k} * u_h^k\|_K^2 \right)^{\frac{1}{2}},
\end{aligned}
$$

by an inverse inequality. Then $T \leqslant T_1 + T_2$, where

$$
\begin{aligned}
T_1 &:= C \left(\sum_{K\in\Omega_0, \partial K\cap\partial\Omega_0 = F} \|U_{F,K} - u\|_K^2 \right)^{\frac{1}{2}}, \\
T_2 &:= C \left(\sum_{K\in\Omega_0} \|u - K_h^{2k} * u_h^k\|_K^2 \right)^{\frac{1}{2}}.
\end{aligned}
$$

A standard approximation theory gives us that

$$
T_1 \leqslant C h^{2k+1} \|u\|_{2k+1,\Omega_0},
$$

and, by Theorem 3.2 and by Lemma 3.3 with $U := u_h$, the approximation of u given by the HDG method, we get that

$$
T_2 = \|u - K_h^{2k} * u_h^k\|_{\Omega_0} \leqslant C h^{2k+1} \|u\|_{2k+2,\Omega}.
$$

This completes the proof. □

Step vi: Conclusion

We are now ready to prove Theorem 2.6. We only have to recall the formulas of the errors E_h^G and E_h^F in Step 1, apply the Cauchy–Schwarz inequality to each of the terms of those formulas and use the estimates obtained in the previous Steps. Since each of those terms is of the order of at least h^{2k}, we get that both errors are of order h^{4k}. This completes the proof of Theorem 2.6.

4. Computational results

In this section, we design several numerical experiments to explore the convergence properties of the method under consideration. The numerical experiments in Cockburn & Wang (2017), carried out with HDG methods using piecewise-polynomials of degree $k \geqslant 0$ suggest that the orders of convergence given by Theorem 2.6 are sharp. Here, we use HDG methods defined in squares and explore how close Ω_0 can be to the boundary $\partial\Omega$, how the smoothness of the solution u affects the convergence rate, and how the smoothness of the functional J affects the convergence rate.

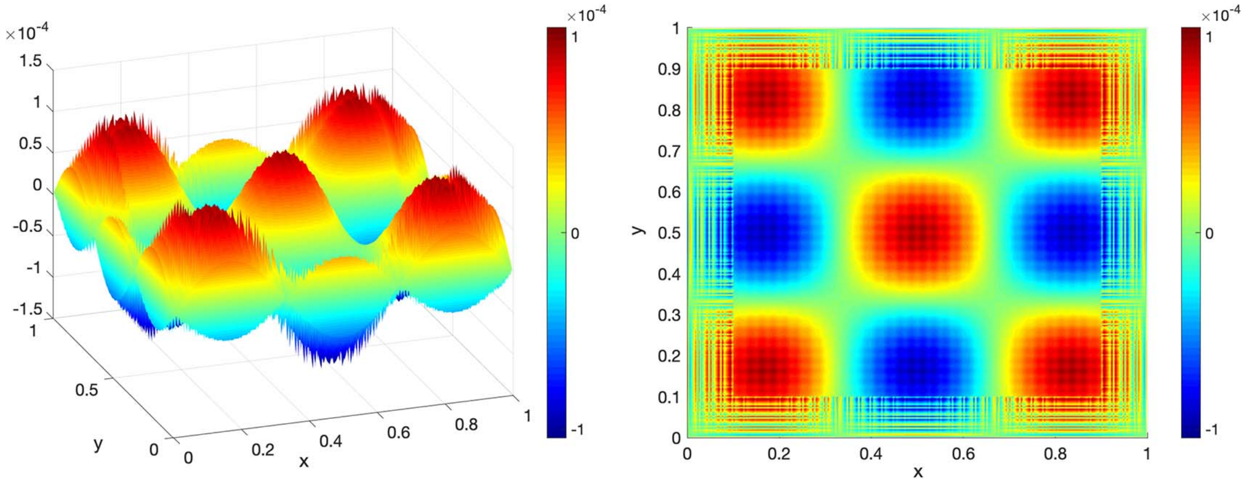

FIG. 2. Approximation error $u - u_h^*$ when $k = 1$ and $N = 40$. The distance between Ω_0 and $\partial\Omega$ is 0.1. Note that the magnitude of the error in $\Omega\backslash\Omega_0$ is essentially the same as that in Ω_0.

We consider a unit square domain $\Omega = (0,1)^2$. This domain has been shown to be $(k+2)$-regular for $k \geqslant 0$ in (Nitsche & Schatz, 1974, Section 7, Example 3). For each natural number N, we obtain the mesh by dividing Ω into N^2 uniform squares with $h = \frac{1}{N}$. Then for each element K, we choose the local spaces

$$\boldsymbol{V}(K) := \mathscr{Q}_k(K) \oplus \boldsymbol{curl}span\{x^{k+1}y, xy^{k+1}\},\ W(K) = \mathscr{Q}_k(K),\ M(\partial K) = \mathscr{Q}_k(\partial K),$$

so that Assumptions 2.1 and 2.2 are satisfied, see Table 1.

We implemented the numerical methods in MATLAB R2019a in a personal laptop (MacBook Pro 2019). To show the high-order accuracy of the method, we used the Multi-precision Computing Toolbox and select 32 digits of accuracy.

4.1 *Illustration of the approximation error $u - u_h^*$*

We start by displaying the approximation error $u - u_h^*$ in the case in which the linear functional is given by (1.1) with $g(x,y) := 18\pi^2 \sin(3\pi x)\sin(3\pi y)$. We choose the exact solution to be $u(x,y) := \sin(3\pi x)\sin(3\pi y)$, and determine the boundary conditions and the source term f accordingly. We choose Ω_0 to be region such that the distance between Ω_0 and the boundary $\partial\Omega$ is 0.1.

In Fig. 2, we display the approximation error plot $u - u_h^*$ when $k = 1$ and $N = 40$. In Fig. 3, we show 1D slices of the approximation error plot at $x = 0.50117$ and $y = 0.20117$, respectively. From these figures, we can see that the size of the error in region $\Omega\backslash\Omega_0$, using the HDG method with polynomial degree $2k$, matches the size of the error in Ω_0. Similar error plots are obtained for the following cases, so we omit them. We refer the reader to Cockburn & Wang (2017) for more approximation error plots in 1D and 2D triangular meshes.

4.2 *How close Ω_0 can be to the boundary $\partial\Omega$*

Next, we explore how close we can actually choose Ω_0 to the boundary and how the convergence rates might be effected. In our main theorem 2.6, we tacitly assume Ω_0 is independent of the mesh and of the

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

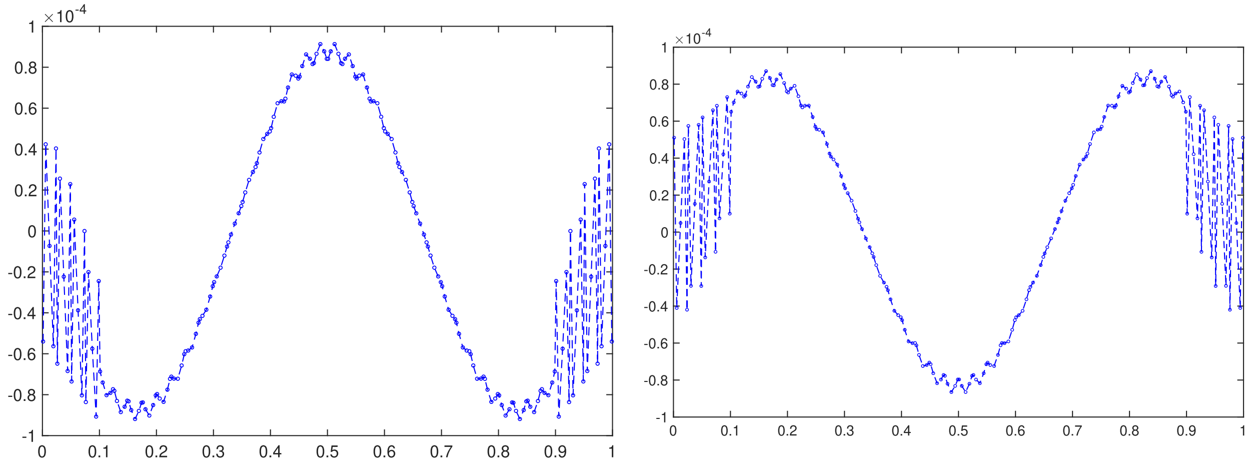

FIG. 3. Slices of the approximation error $u - u_h^*$ at $x = 0.50117$ (left) and $y = 0.20117$ (right) when $k = 1$ and $= 40$. In each element, we evaluate $u - u_h^*$ at 5 Gauss quadrature points.

TABLE 2 *History of convergence when the distance between Ω_0 and $\partial\Omega$ is fixed and equal to* 0.1

k	N	$\|u - u_h^*\|_{L^2(\Omega)}$	order	$\|J(u) - J_h^F(u_h^*, v_h^*)\|$	order	$\|J(u) - J_h^G(u_h^*, v_h^*)\|$	order
1	40	4.82e–05	–	9.95e–07	–	8.50e–05	–
	50	2.17e–05	3.59	2.58e–07	6.04	3.47e–05	4.02
	60	1.14e–05	3.55	8.91e–08	5.84	1.67e–05	4.01
	70	6.61e–06	3.51	3.70e–08	5.69	9.02e–06	4.00
2	40	5.4e–07	–	5.47e–11	–	6.69e–11	–
	50	1.44e–07	5.93	3.85e–12	11.89	5.84e–12	10.93
	60	4.87e–08	5.94	4.39e–13	11.91	8.91e–13	10.31
	70	1.94e–08	5.95	6.99e–14	11.92	2.00e–13	9.71

polynomial degree k. In other words, our theory holds if we fix the domain Ω_0. In Cockburn & Wang (2017), however, numerical experiments in one- and two-space dimensions were performed with Ω_0 just $(2k+1)$ elements away from the boundary. The theoretical orders of convergence rate h^{4k}, for $k = 1, 2, 3$ were actually achieved when the solutions u and v are very smooth. From a practical standpoint, the larger $\Omega_0 \subset\subset \Omega$ is, the less computational efforts are needed. This is because the convolution filtering technique is fairly inexpensive when compared to solving the boundary-value problem on Ω_1 with the HDG method. Therefore, we want to choose Ω_0 to be as large as possible. On the other hand, based on Theorem 3.2, in order to use the convolution filtering technique, Ω_0 has to have at least $2k$ elements outside in each direction. In the following numerical experiment, we consider the test problem of the previous subsection. In this way, both solutions u and v are very smooth. In this ideal case, we compare two cases: (1) Ω_0 is a fixed domain such that the distance between Ω_0 and the boundary $\partial\Omega$ is 0.1, (2) Ω_0 is defined by removing a boundary layer of $2k$ elements from Ω.

Let us report the history of convergence for these two cases. The results of the first case, where Ω_0 is a fixed domain, is shown in Table 2. We see that the convergence rate of u_h^* is of order $\mathcal{O}(h^{2k+1})$, and that the accuracy of the two approximations $J_h^F(u_h^*, v_h^*)$ and $J_h^G(u_h^*, v_h^*)$ are at least $\mathcal{O}(h^{4k})$ for $k = 1, 2$. These results are consistent with our theoretical results.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

TABLE 3 *History of convergence when the distance between Ω_0 and $\partial\Omega$ is, for each mesh, is* $2k$ *elements*

k	N	$\|u-u_h^*\|_{L^2(\Omega)}$	order	$\|J(u)-J_h^F(u_h^*,v_h^*)\|$	order	$\|J(u)-J_h^G(u_h^*,v_h^*)\|$	order
1	40	4.79e–05	–	8.46e–07	–	4.57e–05	–
	50	2.11e–05	3.67	1.94e–07	6.59	1.50e–05	5.00
	60	1.08e–05	3.66	5.95e–08	6.50	6.03e–06	5.00
	70	6.16e–06	3.65	2.21e–08	6.42	2.79e–06	5.00
2	40	5.61e–07	–	5.70e–11	–	6.73e–11	–
	50	1.50e–07	5.91	4.03e–12	11.87	5.45e–12	11.26
	60	5.07 e–08	5.95	4.59e–13	11.92	7.38e–13	10.96
	70	2.02 e–08	5.97	7.28e–14	11.93	1.43e–13	10.63

We display a history of convergence when Ω_0 varies with the mesh and k in Table 3. This case is not covered in the theory, but we can see that we still have super-convergence. This implies that the constant C in our main theorem 2.6 can remain bounded when we vary Ω_0. Moreover, these results suggest that we should select Ω_0 only $\mathcal{O}(h)$ away from the boundary since the computational effort is smaller and that the results are better. The results also indicate that $J_h^F(u_h^*,v_h^*)$ is more accurate than $J_h^G(u_h^*,v_h^*)$. This is most probably due to the better approximation to the gradient used in the HDG method of degree $2k$ used for $J_h^F(u_h^*,v_h^*)$.

Finally, in Table 4 we report a CPU time, error and efficiency comparison of using the adjoint-based approximation $J_h^F(u_h^*,v_h^*)$ and using the standard approximation $J(u_h)$ with the *HDG* method. Here t_{Adj}/t_{HDG} denotes the ratio of the CPU time spent on computing $J_h^F(u_h^*,v_h^*)$ to that of computing $J(u_h)$, and e_{HDG}/e_{Adj} stands for the ratio of the error $|J(u)-J(u_h)|$ to the error $|J(u)-J_h^F(u_h^*,v_h^*)|$. To measure the efficiency of the method, we use the quantity $\omega := 1/(te)$. As we see in Table 4, the running time of the adjoint-based method is about 3 times as much as that of the standard method. However, the error of the adjoint-based method is several orders of magnitude (from 10^3 to 10^5 times) smaller than that of the standard method. This is reflected in the fact that the ratio of the efficiencies of the methods $\omega_{Adj}/\omega_{HDG}$ is several orders of magnitude bigger than one, which shows the advantage of the adjoint-based method. In Fig. 4, we illustrate this advantage from another perspective. There we compare the efficiencies of the adjoint-based method $J_h^F(u_h^*,v_h^*)$ and the standard method $J(u_h)$ when $k=1$. We can see that to achieve the given error tolerance 9×10^{-5}, the adjoint-based method only takes 0.9s with a mesh with $N=20$, while the standard method takes 49s with a mesh with $N=90$. In other words, in this case, the adjoint-based method is more than 50 times faster than the standard method.

4.3 *Effect of the smoothness of the solution for variable* Ω_0

In this section, we explore how the smoothness of the solution can affect the convergence rate of the adjoint-based method. To do that, we compare the errors of $J(u_h)$ and $J_h^F(u_h^*,v_h^*)$ for solutions with different smoothness. Here we choose Ω_0 to be the subdomain of Ω, which lies $2k$ elements away from the boundary for k = 1, 2, 3.

4.3.1 *Smooth solution u.* As a first example, we consider the same problem as in the previous section. We report the errors and the history of convergence in Table 5. As we can see, when the exact solution is smooth, the HDG solution u_h with polynomial degree k converges at the rate of $\mathcal{O}(h^{k+1})$, the

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

TABLE 4 *CPU time, error and efficiency comparison of the adjoint-based method $J_h^F(u_h^*, v_h^*)$ and the standard method $J(u_h)$ when u is a smooth function*

k	N	Ω_0 fixed			Ω_0 varies with k and h		
		t_{Adj}/t_{HDG}	e_{HDG}/e_{Adj}	$\omega_{Adj}/\omega_{HDG}$	t_{Adj}/t_{HDG}	e_{HDG}/e_{Adj}	$\omega_{Adj}/\omega_{HDG}$
1	40	3.23	1.07e+03	331	3.11	1.27e+03	408
	50	3.41	2.15e+03	630	2.72	2.86e+03	1,051
	60	3.43	3.58e+03	1,043	2.89	5.38e+03	1,861
	70	3.43	5.38e+03	1,569	2.73	9.01e+03	3,300
2	40	4.16	1.27e+04	3,053	3.87	1.22e+04	3,152
	50	3.62	5.56e+04	15,359	3.13	5.32e+04	16,997
	60	2.94	1.87e+05	63,605	2.67	1.79e+05	67,041
	70	3.07	5.26e+05	171,336	2.53	5.05e+05	199,604

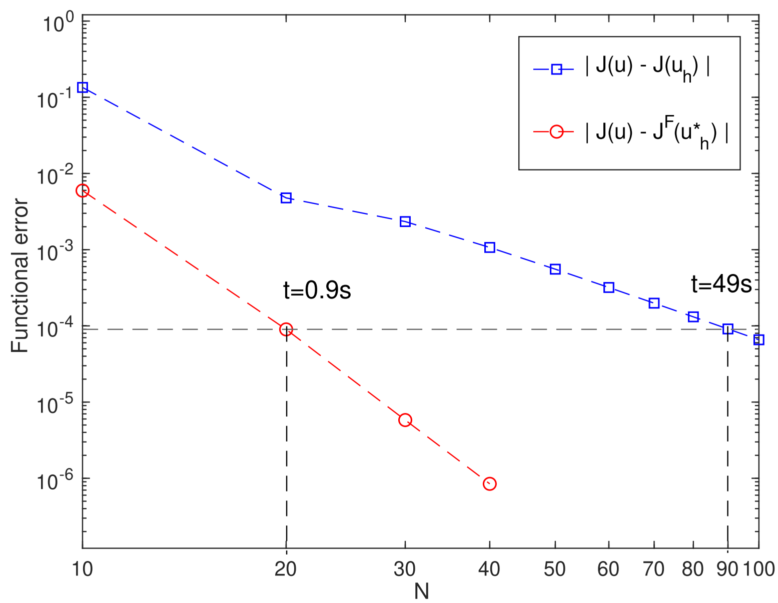

FIG. 4. Log scale error comparison of the adjoint-based method $J_h^F(u_h^*, v_h^*)$ ($\circ$) and the standard method $J(u_h)$ ($\square$) when u is a smooth function and $k = 1$. Note that to achieve the given error tolerance 9×10^{-5}, the adjoint-based method only takes 0.9s with a mesh $N = 20$, while the standard method takes 49s with a mesh $N = 90$.

approximation $J(u_h)$ converges at the rate of $\mathcal{O}(h^{2k+1})$, the post-processed solution u_h^* converges at the rate of $\mathcal{O}(h^{2k+1})$, and the adjoint-based method $J_h^F(u_h^*, v_h^*)$ has the accuracy of at least $\mathcal{O}(h^{4k})$. We thus recover the orders of convergence for the case in which the set Ω_0 is fixed.

4.3.2 *Solution u with corner singularity.* In this second example, we consider the same linear functional (1.1) and $g(x, y) := e^{x+y}$, but we choose $u(r, \theta) := r^{\frac{2}{3}} \sin\left(\frac{2}{3}\theta\right)$ in polar coordinates so that the exact solution u is singular at the origin. We define the boundary conditions and the source term f according to u. We use the adjoint-based method (2.9) to approximate the functional and report the

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

TABLE 5 *History of convergence when the solution u is smooth and $k = 1, 2, 3$. The distance between Ω_0 and $\partial\Omega$ is, for each mesh, $2k$ elements*

k	N	$\|u-u_h\|_{L^2(\Omega)}$	Order	$\|J(u)-J(u_h)\|$	Order	$\|u-u_h^*\|_{L^2(\Omega)}$	Order	$\|J(u)-J_h^F(u_h^*,v_h^*)\|$	Order
1	40	1.69e–03	–	1.07e–03	–	4.79e–05	–	8.46e–07	–
	50	1.03e–03	2.21	5.54e–04	2.94	2.11e–05	3.67	1.94e–07	6.59
	60	6.97e–04	2.15	3.19e–04	3.03	1.08e–05	3.66	5.95e–08	6.50
	70	5.03e–04	2.11	1.99e–04	3.07	6.16e–06	3.65	2.21e–08	6.42
2	20	2.33e–04	–	2.88e–05	–	2.43e–05	–	1.38e–07	–
	30	6.90e–05	3.00	3.23e–06	5.39	2.95e–06	5.21	1.65e–09	10.90
	40	2.91e–05	3.00	6.97e–07	5.34	5.61e–07	5.77	5.70e–11	11.70
	50	1.49e–05	3.00	2.14e–07	5.29	1.50e–07	5.91	4.03e–12	11.87
3	20	6.90e–06	–	1.22e–08	–	8.70e–07	–	1.74e–10	–
	30	1.37e–06	4.00	4.89e–10	7.93	6.69e–08	6.32	1.04e–12	12.63
	40	4.32e–07	4.00	4.94e–11	7.97	8.70e–09	7.09	1.49e–14	14.75
	50	1.77e–07	4.00	8.46e–12	7.91	1.59e–09	7.63	4.67e–16	15.53

TABLE 6 *History of convergence when the solution u has a corner singularity and $k = 1, 2, 3$. The distance between Ω_0 and $\partial\Omega$ is, for each mesh, $2k$ elements*

k	N	$\|u-u_h\|_{L^2(\Omega)}$	Order	$\|J(u)-J(u_h)\|$	Order	$\|u-u_h^*\|_{L^2(\Omega)}$	Order	$\|J(u)-J_h^F(u_h^*,v_h^*)\|$	Order
1	50	3.38e–05	–	4.68e–08	–	9.99e–06	–	8.60e–09	–
	60	2.50e–05	1.66	2.88e–08	2.66	7.38e–06	1.66	5.32e–09	2.64
	70	1.94e–05	1.66	1.91e–08	2.66	5.71e–06	1.66	3.54e–09	2.64
	80	1.55e–05	1.66	1.34e–08	2.65	4.57e–06	1.66	2.49e–09	2.64
2	50	1.01e–05	–	8.42e–09	–	2.43e–06	–	1.01e–09	–
	60	7.48e–06	1.66	5.18e–09	2.66	1.80e–06	1.66	6.20e–10	2.65
	70	5.79e–06	1.66	3.44e–09	2.66	1.39e–06	1.66	4.12e–10	2.65
	80	4.64e–06	1.66	2.41e–09	2.66	1.11e–06	1.66	2.89e–10	2.65
3	50	4.68e–06	–	3.04e–09	–	9.15e–07	–	1.52e–10	–
	60	3.45e–06	1.66	1.87e–09	2.66	6.76e–07	1.66	9.38e–11	2.65
	70	2.67e–06	1.66	1.24e–09	2.66	5.24e–07	1.66	6.22e–11	2.66
	80	2.14e–06	1.66	8.70e–10	2.66	4.20e–07	1.66	4.37e–11	2.65

history of convergence in Table 6. As we can see, due to the lack of smoothness of u, the convergence rate of the post-processed solution u_h^* and the adjoint approximation $J_h^F(u_h^*, v_h^*)$ do not improve as the polynomial degree is increased. In fact, in this case it can be shown that the HDG solution with polynomial degree $k \geqslant 1$ converges in $L^2(\Omega)$ with order $\mathscr{O}(h^{\frac{5}{3}})$. Contrast this behavior against the case of smooth solutions in the previous example in Table 5.

Let us also report the CPU time, error and efficiency comparison for this case in Table 7. Even though the convergence rates of the adjoint-based approximation $J_h^F(u_h^*, v_h^*)$ is the same as the standard approximation $J(u_h)$, the error $|J(u) - J_h^F(u_h^*, v_h^*)|$ is still much smaller (from 5 to 20 times) than $|J(u) - J(u_h)|$ for $k = 1, 2, 3$ (see e_{HDG}/e_{Adj} in Table 7). In addition, the running time of the adjoint-based approximation is only about twice that of the standard method (see t_{Adj}/t_{HDG} in Table 7). Finally, the

TABLE 7 *CPU time, error and efficiency comparison of the adjoint-based method $J_h^F(u_h^*, v_h^*)$ and the standard method $J(u_h)$ when u has a corner singularity. The set Ω_0 varies with k and h*

k	N	t_{Adj}/t_{HDG}	e_{HDG}/e_{Adj}	$\omega_{Adj}/\omega_{HDG}$
1	50	1.63	5.43	3.33
	60	1.63	5.41	3.32
	70	1.57	5.41	3.45
	80	1.70	5.41	3.18
2	50	1.62	8.40	5.19
	60	1.86	8.33	4.48
	70	1.99	8.33	4.19
	80	2.17	8.33	3.84
3	50	2.29	20.0	8.73
	60	2.20	19.6	8.91
	70	2.24	19.6	8.75
	80	2.44	19.6	8.03

ratio of efficiencies $\omega_{Adj}/\omega_{HDG}$ shows that the adjoint-based method is around 3 times more efficient than the standard method for $k = 1$, 4 times for $k = 2$ and 8 times for $k = 3$. This indicates that even though the exact solution u displays a corner singularity, we can still benefit from using the adjoint-based method.

4.4 *Effect of the smoothness of the functional*

The smoothness of a functional is reflected by the smoothness of the solution of the adjoint problem. In this section, we consider a functional on the boundary

$$J(u) := \langle \nabla u \cdot \boldsymbol{n}, \psi \rangle_{\partial\Omega} =< -\boldsymbol{q} \cdot \boldsymbol{n}, \psi >_{\partial\Omega}, \tag{4.1}$$

where u is the solution of the model problem (1.2) and $\boldsymbol{q} = -\nabla u$. If the dual problem

$$-\Delta v = 0 \quad \text{in } \Omega, \quad v = \psi \quad \text{on } \partial\Omega,$$

has a weak solution $v \in H^1(\Omega)$, then the functional can be written as

$$J(u, v) = (\nabla u, \nabla v)_\Omega - (f, v)_\Omega. \tag{4.2}$$

Therefore, we can simply define an approximation $J(u_h, v_h) = (\boldsymbol{q}_h, \boldsymbol{p}_h) - (f, v_h)$, where $\boldsymbol{q}_h$ and $\boldsymbol{p}_h$ are the approximations to $-\nabla u$ and $-\nabla v$, respectively. We can now design an adjoint-correction term AC_h and define a new approximation

$$J_{Ih} = J(u_h, v_h) + AC_h,$$

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

TABLE 8 *History of convergence for the boundary integral functional with* $v = \frac{\pi-\theta}{\pi}$

k	N	$\lvert J(u) - J_{Bh}\rvert$	Order	$\lvert J(u) - J_h^*\rvert$	Order
1	14	2.85e–05	–	3.30e–07	–
	22	7.82e–06	2.86	5.26e–08	4.06
	30	3.18e–06	2.91	1.43e–08	4.20
	38	1.59e–06	2.93	5.29e–09	4.20
2	14	3.34e–07	–	1.17e–10	–
	22	4.89e–08	4.25	7.69e–12	6.02
	30	1.33e–08	4.20	1.17e–12	6.07
	38	4.98e–09	4.16	2.62e–13	6.32
3	14	4.42e–09	–	7.97e–15	–
	22	4.37e–10	5.12	1.50e–16	8.79
	30	9.05e–11	5.08	1.03e–17	8.63
	38	2.72e–11	5.08	1.39e–18	8.46

where

$$\begin{aligned} AC_h = &- (\boldsymbol{q}_h, \boldsymbol{p}_h) + (u_h, \nabla \boldsymbol{p}_h) - \langle \widehat{u}_h, \boldsymbol{p}_h \cdot \boldsymbol{n} \rangle \\ &- (\boldsymbol{q}_h, \boldsymbol{p}_h + \nabla v_h) - \langle \widehat{v}_h - v_h \,, \boldsymbol{p}_h \cdot \boldsymbol{n} \rangle \\ &+ \langle (\boldsymbol{q}_h - \widehat{\boldsymbol{q}}_h) \cdot \boldsymbol{n}, \widehat{v}_h - v_h \rangle. \end{aligned}$$

After a simple calculation, we get that the error term is

$$\begin{aligned} E_h &:= J(u) - J_{Ih} \\ &= (\boldsymbol{q}_h - \boldsymbol{q}, \boldsymbol{p}_h + \nabla v_h) + (u - u_h, \nabla \cdot (\boldsymbol{p}_h - \boldsymbol{p})) \\ &\quad + \langle (\widehat{\boldsymbol{q}}_h - \boldsymbol{q}) \cdot \boldsymbol{n} \,, v - v_h \rangle + \langle (\boldsymbol{p}_h - \boldsymbol{p}) \cdot \boldsymbol{n} \,, \widehat{u}_h - u \rangle. \end{aligned}$$

In the following test cases, we compare two approximations of the functional defined by (4.1), namely,

$$\begin{aligned} J_{Bh} &:=< -\widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}, \psi >_{\partial\Omega}, \\ J_{Ih}^* &:= J(u_h^*, v_h^*) + AC_h. \end{aligned}$$

The first approximation only uses the HDG approximate solutions. The second uses the approximations u_h^* and v_h^*, and the HDG flux approximation to compute AC_h in the same manner as (2.9).

4.4.1 *Smooth solution u, but nonsmooth solution v.* In this first case, we choose the exact solution to be a smooth function $u = \sin(\pi x)\sin(\pi y)$ and let $v = \frac{\pi-\theta}{\pi}$, where θ is the angle component of the polar coordinates at $\left(\frac{1}{8}, 0\right)$. We define the boundary conditions accordingly. Note that ψ is a discontinuous function on boundary since $\psi(x,0) = 1$ when $x > \frac{1}{8}$ and $\psi(x,0) = 0$ when $x < \frac{1}{8}$.

As shown in Table 8, the convergence rates of J_{Bh} is $\mathscr{O}(h^{k+2})$ and the convergence rate of J_{Ih}^* is $\mathscr{O}(h^{2k+2})$. Compared with the case in Section 4.2, where the functional is an integral over the whole

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

TABLE 9 *History of convergence of approximations to* $u = \sin(\pi x)\sin(\pi y)$ *and to* $v = \frac{\pi-\theta}{\pi}$

k	N	$\|u - u_h\|_{L^2(\Omega)}$	order	$\|u - u_h^*\|_{L^2(\Omega)}$	order	$\|v - v_h\|_{L^2(\Omega)}$	order	$\|v - v_h^*\|_{L^2(\Omega)}$	order
1	14	1.34e–03	–	5.97e–05	–	5.43e–03	–	3.48e–03	
	22	5.41e–04	2.02	1.24e–05	3.48	3.46e–03	1.00	2.20e–03	1.01
	30	2.90e–04	2.01	4.29e–06	3.41	2.54e–03	1.00	1.63e–03	0.98
	38	1.81e–04	2.01	1.94e–06	3.36	2.01e–03	1.00	1.29e–03	0.99
2	14	2.52e–05	–	4.13e–07	–	3.99e–03	–	2.44e–03	–
	22	6.48e–06	3.00	2.94e–08	5.84	2.56e–03	0.98	1.56e–03	0.99
	30	2.56e–06	3.00	4.68e–09	5.93	1.89e–03	0.99	1.15e–03	0.99
	38	1.26e–06	3.00	1.15e–09	5.93	1.49e–03	0.99	9.09e–04	0.99
3	14	3.56e–07	–	4.44e–09	–	2.70e–03	–	1.39e–03	–
	22	5.83e–08	4.00	1.74e–10	7.17	1.72e–03	0.99	8.89e–04	0.99
	30	1.69e–08	4.00	1.51e–11	7.88	1.26e–03	1.00	6.53e–04	1.00
	38	6.55e–09	4.00	2.30e–12	7.96	9.99e–04	1.00	5.16e–04	1.00

TABLE 10 *CPU time, error and efficiency comparison of the adjoint-based approximation* $J_h^F(u_h^*, v_h^*)$ *and the standard approximation* $J(u_h)$ *for the boundary integral when* $u = \sin(\pi x)\sin(\pi y)$ *and to* $v = \frac{\pi-\theta}{\pi}$. *The set* Ω_0 *varies with k and h*

k	N	t_{Adj}/t_{HDG}	e_{HDG}/e_{Adj}	$\omega_{Adj}/\omega_{HDG}$
1	14	6.84	8.63e+01	13
	22	5.25	1.49e+02	28
	30	4.01	2.23e+02	56
	38	3.61	3.00e+02	83
2	14	6.61	2.86e+03	433
	22	7.23	6.36e+03	880
	30	4.77	1.14e+04	1,333
	38	4.38	1.90e+04	4,338
3	14	14.35	5.55e+05	38,676
	22	11.57	2.91e+06	251,513
	30	8.30	8.79e+06	1,059,036
	38	5.98	1.96e+07	3,277,592

domain Ω and the approximation accuracy is $\mathscr{O}(h^{2k+1})$ with HDG method, and $\mathscr{O}(h^{4k})$ with the adjoint-based method, we see that the accuracy of our approximation decreases. This result is reasonable because in this case the solution of the adjoint problem v is not in $H^1(\Omega)$ and the accuracy of our approximation solution v_h can only be $\mathscr{O}(h)$ as shown in Table 9.

We also report the CPU time, error and efficiency comparison in Table 10. We can see that as we refine the mesh, the running time of the adjoint-based approximation is only about 3–5 times that of the standard method (see t_{Adj}/t_{HDG}); however, the error of the adjoint-based method is significantly smaller (from 10^2 to 10^7 times) than that of the standard method (see e_{HDG}/e_{Adj}). The values of the ratio of efficiencies $\omega_{Adj}/\omega_{HDG}$ are always bigger than one by, in most cases, several orders of magnitude, showing that the adjoint-based method is clearly more efficient.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

TABLE 11 *History of convergence of the boundary integral functional with $u = r^{\frac{2}{3}}\sin(\frac{2}{3}\theta)$ and not smooth v*

k	N	$\lvert J(u) - J_{Bh}\rvert$	Order	$\lvert J(u) - J^*_{Ih}\rvert$	Order
1	20	1.24e–02	–	8.96e–03	–
	30	9.51e–03	0.655	6.86e–03	0.661
	40	7.87e–03	0.659	5.67e–03	0.663
	50	6.79e–03	0.660	4.89e–03	0.663
2	20	9.12e–03	–	4.32e–03	–
	30	6.97e–03	0.661	3.30e–03	0.665
	40	5.76e–03	0.663	2.73e–03	0.666
	50	4.97e–03	0.664	2.35e–03	0.666
3	20	6.09e–03	–	2.68e–03	–
	30	4.65e–03	0.664	2.05e–03	0.666
	40	3.84e–03	0.665	1.69e–03	0.666
	50	3.31e–03	0.665	1.46e–03	0.666

TABLE 12 *CPU time, error and efficiency comparison of the adjoint-based approximation $J_h^F(u_h^*, v_h^*)$ and the standard approximation $J(u_h)$ for the boundary integral when $u = r^{\frac{2}{3}}\sin\left(\frac{2}{3}\theta\right)$ and is not smooth v. The set Ω_0 varies with k and h*

k	N	t_{Adj}/t_{HDG}	e_{HDG}/e_{Adj}	$\omega_{Adj}/\omega_{HDG}$
1	20	3.89	1.38	0.35
	30	3.49	1.39	0.40
	40	3.42	1.39	0.41
	50	2.87	1.39	0.48
2	20	4.49	2.11	0.47
	30	4.33	2.11	0.49
	40	3.28	2.11	0.64
	50	2.98	2.11	0.71
3	20	7.58	2.27	0.30
	30	7.23	2.27	0.31
	40	4.97	2.27	0.46
	50	3.68	2.27	0.62

4.4.2 *Solution u with corner singularity and nonsmooth solution v.* In this second case, we let the exact solution be a function with a singularity at the origin, namely, $u = r^{\frac{2}{3}}\sin\left(\frac{2}{3}\theta\right)$ and set $\psi = e^x\chi_{[0,1]\times[0]}$. A simple calculation gives us that

$$J(u) = \langle \nabla u \cdot \boldsymbol{n}, \psi \rangle = -\frac{2}{3}\int_0^1 x^{-\frac{1}{3}} e^x \, \mathrm{d}x.$$

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

This integral is related to the Gamma function and its numerical value can be easily computed to any precision (for example, using Mathematica).

In Table 11, we see that we have less accuracy due to the lack of smoothness of the solutions u and v. In Table 12, we see that the error of the adjoint-based approximation is only about 2 times smaller than the standard approximation, but the running time of the adjoint-based approximation is at least 3 times longer than the standard approximation. Since the values of the ratio of efficiencies $\omega_{Adj}/\omega_{HDG}$ are less than one, it is not advantageous to use the adjoint-based approximation. This suggests the need of incorporating adaptive algorithms to deal with the lack of smoothness of the solutions u and v.

5. Extensions and concluding remarks

By combining the adjoint-correction method Pierce & Giles (2000) and the technique of filtering by convolution Bramble & Schatz (1977), two adjoint-based methods with super-convergent properties were obtained in Cockburn & Wang (2017). In this paper, we provide the first *a priori* error analysis for these methods when the solutions u and v are smooth enough. We also displayed extensive numerical results showing the advantages of this method over the standard approach, and showed the need of incorporating adaptivity algorithms whenever the solutions u and v lack sufficient smoothness.

Although we used the HDG method as the Galerkin method in our analysis, similar proofs can be carried out for other Galerkin methods, like the mixed methods and the continuous Galerkin method, as these methods typically provide a very small negative-order norm estimate of the error.

The application of the adjoint-based method to other (linear or nonlinear) functionals of solutions of (linear or nonlinear) partial differential equations, and the incorporation of adaptivity into the method, constitute subjects of ongoing work.

Acknowledgements

The authors would like to thank the referees for constructive criticism leading to a better presentation of the material in this paper.

Funding

U.S. National Science Foundation (grant DMS-1912646 to B.C.).

References

Becker, R. & Rannacher, R. (2001) An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numerica*, **10**, 1–102.

Bramble, J. & Schatz, A. (1977) Higher order local accuracy by averaging in the finite element method. *Math. Comp.*, **31**, 94–111.

Brezzi, F., Douglas, J. & Marini, L. D. (1985) Two families of mixed finite elements for second order elliptic problems. *Numer. Math.*, **47**, 217–235.

Cangiani, A., Dong, Z. & Georgoulis, E. H. (2019) *hp*-version discontinuous Galerkin methods on essentially arbitrarily-shaped elements. arXiv preprint, arXiv:1906.01715.

Cangiani, A., Georgoulis, E. & Sabawi, Y. (2018) Adaptive discontinuous Galerkin methods for elliptic interface problems. *Math. Comp.*, **87**, 2675–2707.

Cockburn, B. & Fu, G. (2017a) Superconvergence by M-decompositions. Part II: construction of two-dimensional finite elements. *ESAIM Math. Model. Numer. Anal.*, **51**, 165–186.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Cockburn, B. & Fu, G. (2017b) Superconvergence by M-decompositions. Part III: construction of three-dimensional finite elements. *ESAIM Math. Model. Numer. Anal.*, **51**, 365–398.

Cockburn, B., Fu, G. & Sayas, F. (2017) Superconvergence by M-decompositions. Part I: general theory for HDG methods for diffusion. *Math. Comp.*, **86**, 1609–1641.

Cockburn, B., Gopalakrishnan, J. & Lazarov, R. (2009) Unified hybridization of discontinuous Galerkin, mixed and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.*, **47**, 1319–1365.

Cockburn, B., Gopalakrishnan, J. & Sayas, F.-J. (2010) A projection-based error analysis of HDG methods. *Math. Comp.*, **79**.

Cockburn, B., Luskin, M., Shu, C.-W. & Süli, E. (2003) Enhanced accuracy by post-processing for finite element methods for hyperbolic equations. *Math. Comp.*, **72**, 577–606 (electronic).

Cockburn, B. & Wang, Z. (2017) Adjoint-based, superconvergent Galerkin approximations of linear functionals. *J. Sci. Comput.*, **73**, 644–666.

Curtis, S., Kirby, R. M., Ryan, J. K. & Shu, C.-W. (2007) Postprocessing for the discontinuous Galerkin method over nonuniform meshes. *SIAM J. Sci. Comput.*, **30**, 272–289.

Eriksson, K., Estep, D., Hansbo, P. & Johnson, C. (1995) Introduction to adaptive methods for differential equations. *Acta Numerica*, **4**, 105–158.

Fidkowski, K. J. & Darmofal, D. L. (2011) Review of output-based error estimation and mesh adaptation in computational fluid dynamics. *AIAA journal*, **49**, 673–694.

Giles, M., Larson, M. G., Levenstam, M & Süli, E. (1997) *Adaptive error control for finite element approximations of the lift and drag coefficients in viscous flow*. Technical report NA-97/06. Oxford, UK: Oxford University Computing Laboratory, Wolfson Building, Parks Road, Oxford OX1 3QD.

Giles, M. B. & Süli, E. (2002) Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality. *Acta Numerica*, **11**, 145–236.

Ji, L., Van Slingerland, P., Ryan, J. K. & Vuik, K. (2014) Superconvergent error estimates for position-dependent smoothness-increasing accuracy-conserving (siac) post-processing of discontinuous Galerkin solutions. *Math. Comp.*, 2239–2262.

King, J., Mirzaee, H., Ryan, J. K. & Kirby, R. M. (2012) Smoothness-increasing accuracy-conserving (SIAC) filtering for discontinuous Galerkin solutions: improved errors versus higher-order accuracy. *J. Sci. Comput.*, **53**, 129–149.

Li, X., Ryan, J. K., Kirby, R. M. & Vuik, K. (2019) Smoothness-increasing accuracy-conserving (SIAC) filtering for discontinuous Galerkin solutions over nonuniform meshes: superconvergence and optimal accuracy. *J. Sci. Comput.*, 1–31.

Mirzaee, H., Ji, L., Ryan, J. K. & Kirby, R. M. (2011) Smoothness-increasing accuracy-conserving (SIAC) postprocessing for discontinuous Galerkin solutions over structured triangular meshes. *SIAM J. Numer. Anal.*, **49**, 1899–1920.

Mirzaee, H., King, J., Ryan, J. K. & Kirby, R. M. (2013) Smoothness-increasing accuracy-conserving filters for discontinuous Galerkin solutions over unstructured triangular meshes. *SIAM J. Sci. Comput.*, **35**, A212–A230.

Mirzaee, H., Ryan, J. K. & Kirby, R. M. (2012) Efficient implementation of smoothness-increasing accuracy-conserving (siac) filters for discontinuous Galerkin solutions. *J. Sci. Comput.*, **52**, 85–112.

Mirzaee, H., Ryan, J. K. & Kirby, R. M. (2014) Smoothness-increasing accuracy-conserving (SIAC) filters for discontinuous Galerkin solutions: application to structured tetrahedral meshes. *J. Sci. Comput.*, **58**, 690–704.

Monk, P. & Süli, E. (1998) The adaptive computation of far-field patterns by a posteriori error estimation of linear functionals. *SIAM J. on Numer. Anal.*, **36**, 251–274.

Nitsche, J. A. & Schatz, A. H. (1974) Interior estimates for Ritz-Galerkin methods. *Math. Comp.*, **28**, 937–958.

Pierce, N. A. & Giles, M. B. (2000) Adjoint recovery of superconvergent functionals from PDE approximations. *SIAM Rev.*, **42**, 247–264.

Ryan, J. & Shu, C.-W. (2003) On a one-sided post-processing technique for the discontinuous Galerkin methods. *Methods Appl. Anal.*, **10**, 295–308.

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Ryan, J. K., Li, X., Kirby, R. M. & Vuik, K. (2015) One-sided position-dependent smoothness-increasing accuracy-conserving (siac) filtering over uniform and non-uniform meshes. *J. Sci. Comput.*, **64**, 773–817.
van Slingerland, P., Ryan, J. K. & Vuik, C. (2011) Position-dependent smoothness-increasing accuracy-conserving (SIAC) filtering for improving discontinuous Galerkin solutions. *SIAM J. Sci. Comput.*, **33**, 802–825.

A. Appendix

A.1 *The M-decompositions*

The concept of M-decomposition was introduced in Cockburn *et al.* (2017) to systematically construct HDG and hybridized mixed methods with superconvergence properties on unstructured meshes. In what follows, for each element K, we consider local finite dimensional spaces $\boldsymbol{V} \subset \boldsymbol{H}(div, K)$, $W \subset H^1(K)$ and $M \subset L^2(\partial K)$.

Definition A.1 We say $\boldsymbol{V} \times W$ admits an M-decomposition if

1. $tr(\boldsymbol{V} \times W) \subset M$,

and there exists a subspace $\widetilde{\boldsymbol{V}} \times \widetilde{W} \subset \boldsymbol{V} \times W$ such that

1. $\nabla W \times \nabla \cdot \boldsymbol{V} \subset \widetilde{\boldsymbol{V}} \times \widetilde{W}$,
2. $tr : \widetilde{\boldsymbol{V}}^{\perp} \times \widetilde{W}^{\perp} \to M$ is an isomorphism.

Here $\widetilde{\boldsymbol{V}}^{\perp}$ and $\widetilde{W}^{\perp}$ are the L^2-orthogonal complements of $\widetilde{\boldsymbol{V}}$ in $\boldsymbol{V}$ and $\widetilde{W}$ in W, respectively, and tr is the trace operator defined as follows

$$\begin{aligned} tr : \boldsymbol{V} \times W &\to L^2(\partial K), \\ (\boldsymbol{v}, w) &\mapsto (\boldsymbol{v} \cdot \boldsymbol{n} + w)|_{\partial K}. \end{aligned}$$

Note that the choice of $\widetilde{\boldsymbol{V}} \times \widetilde{W}$ is not unique. The canonical (orthogonal) M-decomposition is

$$\widetilde{\boldsymbol{V}} = \nabla W \oplus \boldsymbol{V}_{sbb} \quad \text{and} \quad \widetilde{W} = \nabla \cdot \boldsymbol{V},$$

where $\boldsymbol{V}_{sbb} := \{v \in \boldsymbol{V} : \nabla \cdot v = 0, v \cdot \boldsymbol{n} = 0\}$.

A.2 *The HDG Projection*

In this section we define the HDG projection and state its approximation properties. Since we deal with a general domain Ω, we may have two types of elements. The first type is a regular polygonal or polyhedral element and the second type is a boundary polygonal or polyhedral element with one curved face. Let us start with a regular polygonal or polyhedral element K and define the standard HDG projection.

Definition A.2 Let $(\boldsymbol{q}, u)$ be smooth enough so that their boundary traces are in $L^2(\partial K)$. Assume that local spaces $\boldsymbol{V} \times W$ admits an M-decomposition on K. Then the HDG projection $\Pi_h(\boldsymbol{q}, u) := (\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}, \Pi_W u)$ is defined by the following equations:

$$(u - \Pi_W u, w)_K = 0 \quad \forall\, w \in \widetilde{W}, \tag{A.1a}$$

$$(\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}, \boldsymbol{v})_K = 0 \quad \forall\, \boldsymbol{v} \in \widetilde{\boldsymbol{V}}, \tag{A.1b}$$

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

$$\langle (\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}) \cdot \boldsymbol{n} + \tau (u - \Pi_W u) \,,\, \mu \rangle_{\partial K} = 0 \quad \forall\, \mu \in M. \tag{A.1c}$$

Next, we define the HDG projection for elements that have one curved face.

DEFINITION A.3 Let K be a polygonal or polyhedral element with one curved face E. With the same assumption in Definition A.2, we define the HDG projection $\Pi_h(\boldsymbol{q}, u) := (\boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}, \Pi_W u)$ by the following equations:

$$(u - \Pi_W u, w)_K = 0 \quad \forall\, w \in \widetilde{W}, \tag{A.2a}$$

$$(\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}, \boldsymbol{v})_K = 0 \quad \forall\, \boldsymbol{v} \in \boldsymbol{S}, \tag{A.2b}$$

$$\langle (\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}) \cdot \boldsymbol{n} + \tau (u - \Pi_W u) \,,\, \mu \rangle_F = 0 \quad \forall\, \mu \in M(F) \quad F \neq E \tag{A.2c}$$

$$(\nabla \cdot (\boldsymbol{q} - \Pi_{\boldsymbol{V}} \boldsymbol{q}), w) + \langle \tau (u - \Pi_W u) \,,\, w \rangle_{\partial K} = 0 \quad \forall\, w \in W, \tag{A.2d}$$

where $\boldsymbol{S} := \boldsymbol{S}_1 \oplus \boldsymbol{S}_2$, $\boldsymbol{S}_1 = \nabla \widetilde{W}^{\perp}$ and

$$\boldsymbol{S}_2 = \{\boldsymbol{v} \in V \mid \nabla \cdot \boldsymbol{v} = 0 \text{ and } \boldsymbol{v} \cdot \boldsymbol{n}_F = 0 \text{ on } F \neq E \text{ and } (\boldsymbol{v}, \boldsymbol{p}) = 0\, \forall \boldsymbol{p} \in \boldsymbol{S}_1\}.$$

For both cases, the HDG projection depends on the stabilization parameter τ. Let us introduce the quantity

$$\tau_{\widetilde{W}^{\perp}} := \begin{cases} \inf_{\mu \in \widetilde{W}^{\perp} \setminus \{0\}} \frac{\langle \tau \mu, \mu \rangle_{\partial K}}{\|\mu\|^2_{\partial K}} & \text{if } \widetilde{W}^{\perp} \neq 0, \\ \infty & \text{if } \widetilde{W}^{\perp} = 0. \end{cases}$$

It can be shown that, if $\tau_{\widetilde{W}^{\perp}} > 0$, then the HDG projections (A.1) and (A.2) are well-defined. Let us show the approximation properties for the HDG projections. The proof for the regular element case is given in Cockburn *et al.* (2017), and the proof for the curved element case is displayed in Appendix A.3.2.

THEOREM A.4 Suppose $V \times W$ admits an M-decomposition and let the stabilization function τ satisfy the condition

$$\tau_{\widetilde{W}^{\perp}} > 0.$$

Let the element K satisfy Assumption 2.4. Then the HDG projection given by (A.1) or (A.2) is well-defined and has the following approximation properties

$$\begin{aligned} \|\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}\|_K &\leqslant \|(Id - P_{\boldsymbol{V}})\boldsymbol{q}\|_K + C h_K^{\frac{1}{2}} \|(Id - P_{\boldsymbol{V}})\boldsymbol{q} \cdot \boldsymbol{n}\|_{\partial K} + \Theta, \\ \|u - \Pi_W u\|_K &\leqslant \|(Id - P_W)u\|_K + \Theta, \end{aligned}$$

where $\Theta = C h_K^{\frac{1}{2}} \|(Id - P_W)u\|_{\partial K} + C h_K \|(Id - P_{\widetilde{W}})\nabla \cdot \boldsymbol{q}\|_K$ and P is the L^2-projection operator.

A.3 *The HDG projection for curved elements*

In this section, we prove that the HDG projection for curved elements is well-defined and has the optimal approximation properties as stated in Theorem A.4.

For any $K \in \mathscr{T}_h$, let the local space $\boldsymbol{V}(K) \times W(K)$ admit an $M(\partial K)$–decomposition. If K is a polygonal or polyhedral element with one curved face E, we define the HDG projection by equations (A.2).

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

We assume that the stabilization function τ is nonnegative on each face $F \in \partial K$ such that

$$w \in \widetilde{W}^{\perp}: \quad \langle \tau(w)\,,\, w\rangle_{\partial K} = 0 \implies w = 0. \tag{A.3}$$

We can see that $\tau_{\widetilde{W}^{\perp}} > 0$ implies (A.3).

REMARK A.5 If $\widetilde{W} = W(K)$ and $\tau = 0$, the projection becomes:

$$\begin{aligned}
(\Pi_W u, w)_K &= (u, w)_K && \forall\, w \in W(K),\\
(\nabla \cdot \Pi_{\boldsymbol{V}} \boldsymbol{q}, w) &= (\nabla \cdot \boldsymbol{q}, w) && \forall\, w \in W(K),\\
\langle \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q} \cdot \boldsymbol{n}\,,\, \mu\rangle_F &= \langle \boldsymbol{q} \cdot \boldsymbol{n}\,,\, \mu\rangle_F && \forall\, \mu \in M(F) \quad F \neq E\\
(\boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}, \boldsymbol{v})_K &= (\boldsymbol{q}, \boldsymbol{v})_K && \forall\, \boldsymbol{v} \in \boldsymbol{S}.
\end{aligned}$$

Here $\boldsymbol{S} := \{\boldsymbol{v} \in \boldsymbol{V}(K) \mid \nabla \cdot \boldsymbol{v} = 0,\ \boldsymbol{v} \cdot \boldsymbol{n}_F = 0 \text{ on } F \neq E\}$. This is the same projection proposed in Brezzi *et al.* (1985) for the BDM mixed method.

REMARK A.6 In general, even if E is flat, the projection is different from the standard HDG projection. However, if the local spaces satisfy the assumption that $\widetilde{W}^{\perp}|_E = M(E)$, then the projection is equivalent to the standard HDG projection. For example, it is easy to show that for simplicial elements, the projection is the same as the standard HDG projection. In this case, we choose $\boldsymbol{V}(K) := [\mathscr{P}_k(K)]^d$, $W(K) := \mathscr{P}_k(K)$ and $M(F) := \mathscr{P}_F(K)$ for all faces $F \in \partial K$ and $\widetilde{\boldsymbol{V}} := \mathscr{P}_{k-1}(K)$ and $\widetilde{W} := \mathscr{P}_{k-1}(K)$. Then, the assumption $\widetilde{W}^{\perp}|_E = \mathscr{P}_k^{\perp}(K)|_E = \mathscr{P}_k(E) = M(E)$ is satisfied.

Let us write the projection (A.2) in a more detailed manner:

$$(u - \Pi_W u, w)_K = 0 \quad \forall\, w \in \widetilde{W}, \tag{A.4a}$$

$$(\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}, \nabla w)_K = 0 \quad \forall\, w \in \widetilde{W}^{\perp}, \tag{A.4b}$$

$$(\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}, \boldsymbol{v})_K = 0 \quad \forall\, \boldsymbol{v} \in \boldsymbol{S}_2, \tag{A.4c}$$

$$\langle (\boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{q}) \cdot \boldsymbol{n} + \tau(u - \Pi_W u)\,,\, \mu\rangle_F = 0 \quad \forall\, \mu \in M(F) \quad F \neq E, \tag{A.4d}$$

$$(\nabla \cdot (\boldsymbol{q} - \Pi_{\boldsymbol{V}} \boldsymbol{q}), w) + \langle \tau(u - \Pi_W u)\,,\, w\rangle_{\partial K} = 0 \quad \forall\, w \in W(K). \tag{A.4e}$$

A.3.1 *Well-posedness.*

THEOREM A.7 Suppose $\boldsymbol{V} \times W$ admits an M–decomposition and let the stabilization function τ satisfy the condition (A.3). Then the HDG projection (A.4) is well-defined.

Proof. First notice that we can decouple the projection component $\Pi_W u$. If we take $w \in \widetilde{W}^{\perp}$ in (A.4e), then the equations (A.4a) and (A.4e) read

$$(u - \Pi_W u, w)_K = 0 \quad \forall\, w \in \widetilde{W}, \tag{A.5a}$$

$$\langle \tau(u - \Pi_W u)\,,\, w\rangle_{\partial K} = -(\nabla \cdot \boldsymbol{q}, w) \quad \forall\, w \in \widetilde{W}^{\perp}. \tag{A.5b}$$

Clearly, this is a square system for $\Pi_W u$. So to show the existence and uniqueness, we only need to show that, if $\boldsymbol{q} = 0$ and $u = 0$, then $\Pi_W u = 0$. From the first equation, we know that $\Pi_W u \in \widetilde{W}^{\perp}$ and

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

from the second, that $\Pi_W u = 0$ on ∂K. Since $\boldsymbol{V} \times W$ admits an M-decomposition, we also know that $\nabla \cdot (\nabla \Pi_W u) \in \widetilde{W}$. This implies that

$$(\nabla \Pi_W u, \nabla \Pi_W u)_K = \langle \Pi_W u, \nabla \Pi_W u \cdot \boldsymbol{n} \rangle_{\partial K} - (\Pi_W u, \nabla \cdot (\nabla \Pi_W u))K = 0.$$

Therefore, $\Pi_W u$ is a constant and vanishes on ∂K. So $\Pi_W u = 0$.

Next, let us consider the component $\boldsymbol{\Pi_V q}$. The equations (A.4e), (A.4d), (A.4b) and (A.4c), read

$$(\nabla \cdot (\boldsymbol{q} - \Pi_{\boldsymbol{V}} \boldsymbol{q}), w) + \langle \tau (u - \Pi_W u)\,,\, w \rangle_{\partial K} = 0 \quad \forall\, w \in \widetilde{W}, \tag{A.6a}$$

$$\langle (\boldsymbol{q} - \boldsymbol{\Pi_V q}) \cdot \boldsymbol{n} + \tau (u - \Pi_W u)\,,\, \mu \rangle_F = 0 \quad \forall\, \mu \in M(F) \quad F \neq F_0, \tag{A.6b}$$

$$(\boldsymbol{q} - \boldsymbol{\Pi_V q}, \nabla w)_K = 0 \quad \forall\, w \in \widetilde{W}^{\perp}, \tag{A.6c}$$

$$(\boldsymbol{q} - \boldsymbol{\Pi_V q}, v)_K = 0 \quad \forall\, \boldsymbol{v} \in \boldsymbol{S}_2. \tag{A.6d}$$

Using the fact that $\nabla \cdot \boldsymbol{V} = \widetilde{W}$, we can easily see this is a square system for $\boldsymbol{\Pi_V q}$. Now set $\boldsymbol{q} = 0$ and $u = 0$. Since we already proved that $\Pi_W u = 0$, we get that

$$\begin{aligned}
(\nabla \cdot \Pi_{\boldsymbol{V}} \boldsymbol{q}, w) &= 0 \quad \forall\, w \in \widetilde{W}, \\
\langle \boldsymbol{\Pi_V q} \cdot \boldsymbol{n}\,,\, \mu \rangle_F &= 0 \quad \forall\, \mu \in M(F) \quad F \neq F_0, \\
(\boldsymbol{\Pi_V q}, \nabla w)_K &= 0 \quad \forall\, w \in \widetilde{W}^{\perp}, \\
(\boldsymbol{\Pi_V q}, v)_K &= 0 \quad \forall\, \boldsymbol{v} \in \boldsymbol{S}_2.
\end{aligned}$$

Since, by definition, $\boldsymbol{S}_2 := \{\boldsymbol{v} \in \boldsymbol{V}(K) \mid \nabla \cdot \boldsymbol{v} = 0,\ \boldsymbol{v} \cdot \boldsymbol{n}_F = 0 \text{ on } F \neq F_0 \text{ and } (\boldsymbol{v}, \nabla w) = 0\, \forall\, w \in \widetilde{W}^{\perp}\}$, we can see that $\boldsymbol{\Pi_V q} = 0$. Therefore, the projection is well-posed. □

A.3.2 *Approximation properties.* Let us first define the following constants

$$\|\tau\| = \sup_{\lambda,\mu \in L^2(\partial K)} \frac{\langle \tau(\lambda)\,,\, \mu \rangle_{\partial K}}{\|\lambda\|_{\partial K} \|\mu\|_{\partial K}},$$

$$C_{\widetilde{W}^{\perp}} = \sup_{0 \neq w \in \widetilde{W}^{\perp}} \frac{h_K^{-\frac{1}{2}} \|w\|_K}{\|w\|_{\partial K}}.$$

We are now ready to show the proof of Theorem A.4 for the HDG projection for curved elements.

Proof. Let us first prove the estimate for $\Pi_W u$. Let $P_W u$ be the standard L^2-orthogonal projection of u on W. Since

$$\|u - \Pi_W u\|_K \leqslant \|(Id - P_W) u\|_K + \|P_W u - \Pi_W u\|_K,$$

we only need to estimate $\delta_k := P_W u - \Pi_W u$.

By the first of the equations (A.5), $(\delta_k, w)_K = (P_W u - \Pi_W u, w)_K = (u - \Pi_W u)_k = 0\, \forall\, w \in \widetilde{W}$, and so $\delta_k \in \widetilde{W}^{\perp}$. This means that we have

$$\|\delta_k\|_K \leqslant C_{\widetilde{W}^{\perp}} h_K^{\frac{1}{2}} \|\delta_k\|_{\partial K},$$

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

and that we can take $w := \delta_k$ in the second of the equations (A.5) to get

$$\langle \tau\delta_k\,,\,\delta_k\rangle_{\partial K} = (\nabla\cdot\boldsymbol{q},\delta_k)_K + \langle \tau(u-P_W u)\,,\,\delta_k\rangle_{\partial K}.$$

Then,

$$\begin{aligned}\|\delta_k\|^2_{\partial K} &\leqslant \frac{1}{\tau_{\widetilde{W}^\perp}}\langle \tau\delta_k\,,\,\delta_k\rangle_{\partial K}\\ &\leqslant \frac{1}{\tau_{\widetilde{W}^\perp}}(\|\nabla\cdot(\boldsymbol{q}) - P_{\widetilde{W}}\nabla\cdot\boldsymbol{q}\|_K\|\delta_k\|_K + \|\tau\|\|u-P_W u\|_{\partial K}\|\delta_k\|_{\partial K})\\ &\leqslant \frac{C_{\widetilde{W}^\perp}}{\tau_{\widetilde{W}^\perp}}h_K^{\frac{1}{2}}\|(Id-P_{\widetilde{W}})\nabla\cdot\boldsymbol{q}\|_K\|\delta_k\|_{\partial K} + \frac{\|\tau\|}{\tau_{\widetilde{W}^\perp}}\|u-P_W u\|_{\partial K}\|\delta_k\|_{\partial K},\end{aligned}$$

where $P_{\widetilde{W}}$ is the L^2 projection onto $\widetilde{W}$. Thus,

$$\|\delta_k\|_{\partial K} \leqslant \frac{C_{\widetilde{W}^\perp}}{\tau_{\widetilde{W}^\perp}}h_K^{\frac{1}{2}}\|(Id-P_{\widetilde{W}})\nabla\cdot\boldsymbol{q}\|_K + \frac{\|\tau\|}{\tau_{\widetilde{W}^\perp}}\|u-P_W u\|_{\partial K}, \tag{A.7}$$

and

$$\|\delta_k\|_K \leqslant \frac{C^2_{\widetilde{W}^\perp}}{\tau_{\widetilde{W}^\perp}}h_K\|(Id-P_{\widetilde{W}})\nabla\cdot\boldsymbol{q}\|_K + \frac{C_{\widetilde{W}^\perp}\|\tau\|}{\tau_{\widetilde{W}^\perp}}h_K^{\frac{1}{2}}\|u-P_W u\|_{\partial K}.$$

This completes the proof of the second inequality.

Now let us prove the first inequality. Let $P_{\boldsymbol{V}}\boldsymbol{q}$ be the standard L^2-orthogonal projection of $\boldsymbol{q}$ on V. Then

$$\|\boldsymbol{q}-\Pi_{\boldsymbol{V}}\boldsymbol{q}\|_K \leqslant \|(Id-P_{\boldsymbol{V}})\boldsymbol{q}\|_K + \|P_{\boldsymbol{V}}\boldsymbol{q}-\Pi_{\boldsymbol{V}}\boldsymbol{q}\|_K.$$

To estimate the second term, let us define norm

$$|||\boldsymbol{v}|||_K := \|P_{S_1}(\boldsymbol{v})\|_K + \|P_{S_2}(\boldsymbol{v})\|_K + h_K\|\nabla\cdot\boldsymbol{v}\|_K + \sum_{F\neq E} h_K^{\frac{1}{2}}\|\boldsymbol{v}\cdot\boldsymbol{n}\|_F.$$

By the definition of space S_1 and S_2, it is easy to see that $|||\cdot|||$ is a norm on $\mathscr{P}_k(K)$. Since all norms are equivalent infinite dimensional spaces, we have that $\|\boldsymbol{v}\|_K \leqslant C|||\boldsymbol{v}|||_K$ for any $\boldsymbol{v}\in V$. By a simple scaling argument, we know $C>0$ is independent of h_K. Therefore, we only need to estimate $\left|\left|\left|\Pi_{\boldsymbol{V}}\boldsymbol{q}-P_{\boldsymbol{V}}\boldsymbol{q}\right|\right|\right|_K$.

If we set $\boldsymbol{\xi_k} := \Pi_{\boldsymbol{V}}\boldsymbol{q} - P_{\boldsymbol{V}}\boldsymbol{q}$, the equations (A.6) can be written as follows:

$$\begin{aligned}(\nabla\cdot\boldsymbol{\xi_k},w)_K &= (\nabla\cdot(\boldsymbol{q}-P_{\boldsymbol{V}}\boldsymbol{q}),w)_K + \langle\tau(u-\Pi_W u)\,,\,w\rangle_{\partial K}\\ &= \langle(\boldsymbol{q}-P_{\boldsymbol{V}}\boldsymbol{q})\cdot\boldsymbol{n}\,,\,w\rangle_{\partial K} + \langle\tau(u-\Pi_W u)\,,\,w\rangle_{\partial K} && \forall\, w\in W\\ \langle\boldsymbol{\xi_k}\cdot\boldsymbol{n}\,,\,\mu\rangle_F &= \langle(\boldsymbol{q}-P_{\boldsymbol{V}}\boldsymbol{q})\cdot\boldsymbol{n} + \tau(u-\Pi_W u)\,,\,\mu\rangle_F && \forall\,\mu\in\mathscr{P}_k(F)\quad F\neq E,\\ (\boldsymbol{\xi_k},\boldsymbol{v})_K &= 0 && \forall\,\boldsymbol{v}\in\boldsymbol{S}=\boldsymbol{S}_1\cup\boldsymbol{S}_2.\end{aligned}$$

Taking $w := \nabla \cdot \boldsymbol{\xi_k}$ in the first equation, $\mu := \boldsymbol{\xi_k} \cdot \boldsymbol{n}$ in the second, and using an inverse estimate, we get

$$
\begin{aligned}
\|\nabla \cdot \boldsymbol{\xi_k}\|_K &\leqslant C h_K^{-\frac{1}{2}} \|(\boldsymbol{q} - \boldsymbol{P_V q}) \cdot \boldsymbol{n}\|_{\partial K} + C h_K^{-\frac{1}{2}} \|\tau\| \|u - \Pi_W u\|_{\partial K}, \\
\|\boldsymbol{\xi_k} \cdot \boldsymbol{n}\|_F &\leqslant \|(\boldsymbol{q} - \boldsymbol{P_V q}) \cdot \boldsymbol{n}\|_F + \|\tau\| \|u - \Pi_W u\|_F, \\
\|\mathbb{P}_{S_1}(\boldsymbol{\xi_k})\|_K &= 0 \\
\|\mathbb{P}_{S_2}(\boldsymbol{\xi_k})\|_K &= 0.
\end{aligned}
$$

Therefore, we have

$$
\begin{aligned}
\left\|\left|\xi_k\right|\right\| &= \|\mathbb{P}_{S_1}(\boldsymbol{\xi_k})\|_K + \|\mathbb{P}_{S_2}(\boldsymbol{\xi_k})\|_K + h_K \|\nabla \cdot \boldsymbol{\xi_k}\|_K + \sum_{F \neq E} h_K^{\frac{1}{2}} \|\boldsymbol{\xi_k} \cdot \boldsymbol{n}\|_F \\
&\leqslant C h_K^{\frac{1}{2}} \|(Id - \boldsymbol{P_V})\boldsymbol{q} \cdot \boldsymbol{n}\|_{\partial K} + C h_K^{\frac{1}{2}} \|\tau\| \|u - \Pi_W u\|_{\partial K}.
\end{aligned}
$$

By (A.7), we have

$$
\|u - \Pi_W u\|_{\partial K} \leqslant \|u - P_W u\|_{\partial K} + \|\delta_k\|_{\partial K} \leqslant C \|(Id - P_W)u\|_{\partial K} + C h_K^{\frac{1}{2}} \|(Id - P_{\widetilde{W}})\nabla \cdot \boldsymbol{q}\|_{\partial K},
$$

and we get

$$
\|\xi_k\|_K \leqslant C(h_K^{\frac{1}{2}} \|(Id - \boldsymbol{P_V})\boldsymbol{q} \cdot \boldsymbol{n}\|_{\partial K} + h_K \|(Id - P_{\widetilde{W}})\nabla \cdot \boldsymbol{q}\|_{\partial K}) + h_K^{\frac{1}{2}} \|(Id - P_W)u\|_{\partial K}).
$$

This completes the proof. □

A.4 *Estimate of the jump $\varepsilon_h^u - \varepsilon_h^{\widehat{u}}$*

Here, we give the following estimate of the jump $\varepsilon_h^u - \varepsilon_h^{\widehat{u}}$.

LEMMA A.8 Let K be an element of the mesh $\mathscr{T}_h$. Assume that the local space $\boldsymbol{V}(K) \times W(K)$ admits an $M(\partial K)$-decomposition. Then for the approximate solution of the HDG method (2.2), we have the following local stability estimate

$$
\|\varepsilon_h^u - \varepsilon_h^{\widehat{u}}\|_{\partial K} \leqslant C h_K^{\frac{1}{2}} \|\boldsymbol{q} - \boldsymbol{q}_h\|_K.
$$

In the proof, we are going to use the adjoint HDG Projection which is defined as follows.

DEFINITION A.9 (Cockburn *et al.* (2017)). Let K be a regular polygonal or polyhedral element. Assume that $\boldsymbol{V}(K) \times W(K)$ admits an $M(\partial K)$-decomposition. Let $d := (\boldsymbol{d_V}, d_w, d_\mu) \in \boldsymbol{V}(K) \times W(K) \times M(\partial K)$. Then the adjoint-HDG projection $\Pi_h^* d := (\boldsymbol{\Pi}_{\boldsymbol{V}}^* d, \Pi_W^* d) \in \boldsymbol{V} \times W$ is defined by equations

$$
\begin{aligned}
(\boldsymbol{\Pi}_{\boldsymbol{V}}^* d, v)_K &= (\boldsymbol{d_V}, v)_K \quad && \forall\, \boldsymbol{v} \in \widetilde{\boldsymbol{V}}(K), \\
(\Pi_W^* d, w)_K &= (d_w, w)_K \quad && \forall\, w \in \widetilde{W}(K), \\
\langle \boldsymbol{\Pi}_{\boldsymbol{V}}^* d \cdot \boldsymbol{n} - \tau \Pi_W^* d\,,\, \mu \rangle_F &= \langle d_\mu\,,\, \mu \rangle_F \quad && \forall\, \mu \in M(F).
\end{aligned}
$$

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023

DEFINITION A.10 With the same assumption in Definition A.9, let K be an element with one curved face E. Then the adjoint-HDG projection $\Pi_h^* d := (\boldsymbol{\Pi}_{\boldsymbol{V}}^* d, \Pi_W^* d) \in \boldsymbol{V}(K) \times W(K)$ is defined by equations

$$\begin{aligned}
(\boldsymbol{\Pi}_{\boldsymbol{V}}^* d, v)_K &= (\boldsymbol{d}_{\boldsymbol{V}}, v)_K && \forall\, \boldsymbol{v} \in \boldsymbol{S}(K),\\
(\Pi_W^* d, w)_K &= (d_w, w)_K && \forall\, w \in \widetilde{W}(K),\\
\langle \boldsymbol{\Pi}_{\boldsymbol{V}}^* d \cdot \boldsymbol{n} - \tau \Pi_W^* d\,,\, \mu\rangle_F &= \langle d_\mu\,,\, \mu\rangle_F && \forall\, \mu \in M(F), \forall\, F \neq E,\\
\langle \boldsymbol{\Pi}_{\boldsymbol{V}}^* d \cdot \boldsymbol{n} - \tau \Pi_W^* d\,,\, w\rangle_{\partial K} - (\boldsymbol{\Pi}_{\boldsymbol{V}}^* d, \nabla w)_K &= \langle d_\mu\,,\, w\rangle_{\partial K} - (\boldsymbol{d}_{\boldsymbol{V}}, \nabla w)_K && \forall\, w \in W(K).
\end{aligned}$$

Proof. We rewrite the first two error equations (3.2a) and (3.2b) on the element K as

$$\begin{aligned}
(\nabla \varepsilon_h^u, \boldsymbol{v})_K - \langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \boldsymbol{v}\cdot\boldsymbol{n}\rangle_{\partial K} &= (\boldsymbol{q}_h - \boldsymbol{q}, \boldsymbol{v}),\\
(\nabla \cdot \varepsilon_h^q, w)_K + \langle \tau(\varepsilon_h^u - \varepsilon_h^{\widehat{u}})\,,\, w\rangle_{\partial K} &= 0,
\end{aligned}$$

for all $(\boldsymbol{v}, w) \in \boldsymbol{V}(K) \times W(K)$. Adding these equalities and using the fact that the stabilization parameter τ is self-adjoint, we get

$$(\nabla \cdot \varepsilon_h^q, w)_K + (\nabla \varepsilon_h^u, \boldsymbol{v})_K - \langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \boldsymbol{v}\cdot\boldsymbol{n} - \tau w\rangle_{\partial K} = (\boldsymbol{q}_h - \boldsymbol{q}, \boldsymbol{v})_K.$$

Now we take $(\boldsymbol{v}, w) := (\boldsymbol{\Pi}_V^* d, \Pi_W^* d)$, where $d = -(\boldsymbol{0}, 0, \varepsilon_h^u - \varepsilon_h^{\widehat{u}})$ and $\boldsymbol{\Pi}^*$ is the adjoint-HDG projection. Note that for any interior face F, we have $\varepsilon_h^u - \varepsilon_h^{\widehat{u}} \in M(F)$ and so

$$\langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \boldsymbol{\Pi}_{\boldsymbol{V}}^* d \cdot \boldsymbol{n} - \tau \Pi_W^* d\rangle_F = \langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\rangle_F,$$

by definition. If K is a regular element, we have

$$-\langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\rangle_{\partial K} = (\boldsymbol{q}_h - \boldsymbol{q}, \boldsymbol{\Pi}_V^* d).$$

On the other hand, notice that $\varepsilon_h^u - \varepsilon_h^{\widehat{u}} = \varepsilon_h^u$ on $\partial\Omega$. Let K be a boundary element with one curved face E and then we have

$$\begin{aligned}
\langle \varepsilon_h^u\,,\, \boldsymbol{\Pi}_{\boldsymbol{V}}^* d \cdot \boldsymbol{n} - \tau \Pi_W^* d\rangle_E &= \langle \varepsilon_h^u\,,\, \boldsymbol{\Pi}_{\boldsymbol{V}}^* d \cdot \boldsymbol{n} - \tau \Pi_W^* d\rangle_{\partial K} - \langle \varepsilon_h^u\,,\, \boldsymbol{\Pi}_{\boldsymbol{V}}^* d \cdot \boldsymbol{n} - \tau \Pi_W^* d\rangle_{\partial K\setminus E}\\
&= \langle d_\mu\,,\, \varepsilon_h^u\rangle_{\partial K} - (\boldsymbol{d}_{\boldsymbol{V}} - \boldsymbol{\Pi}_{\boldsymbol{V}}^* d, \nabla \varepsilon_h^u)_K - \langle d_\mu\,,\, \varepsilon_h^u\rangle_{\partial K\setminus E}\\
&= \langle d_\mu\,,\, \varepsilon_h^u\rangle_E - (\boldsymbol{d}_{\boldsymbol{V}} - \boldsymbol{\Pi}_{\boldsymbol{V}}^* d, \nabla \varepsilon_h^u)_K\\
&= \langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\rangle_E + (\boldsymbol{\Pi}_{\boldsymbol{V}}^* d, \nabla \varepsilon_h^u)_K.
\end{aligned}$$

Therefore,

$$\langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \boldsymbol{\Pi}_{\boldsymbol{V}}^* d \cdot \boldsymbol{n} - \tau \Pi_W^* d\rangle_{\partial K} = \langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\rangle_{\partial K} + (\boldsymbol{\Pi}_{\boldsymbol{V}}^* d, \nabla \varepsilon_h^u)_K.$$

So if K is an element with one curved face, we have

$$(\nabla \varepsilon_h^u, \boldsymbol{\Pi}_V^* d)_K - \langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\rangle_{\partial K} - (\boldsymbol{\Pi}_{\boldsymbol{V}}^* d, \nabla \varepsilon_h^u)_K = (\boldsymbol{q}_h - \boldsymbol{q}, \boldsymbol{\Pi}_V^* d)_K.$$

Hence, in both cases, we get

$$\langle \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\,,\, \varepsilon_h^u - \varepsilon_h^{\widehat{u}}\rangle_{\partial K} = (\boldsymbol{q} - \boldsymbol{q}_h, \boldsymbol{\Pi}_V^* d)_K.$$

By the approximation properties of the adjoint HDG projection, we have that for $d = -(\boldsymbol{0}, 0, \varepsilon_h^u - \varepsilon_h^{\widehat{u}})$,

$$\|\boldsymbol{\Pi}_V^* d\|_K \leqslant C h_K^{\frac{1}{2}} \|d_\mu\|_{\partial K}.$$

This estimate was shown in Cockburn *et al.* (2017) for the case of regular elements K. For the case of elements with a curved face, the proof is similar.

Therefore, after a simple application of the Cauchy–Schwarz inequality, we have

$$\|\varepsilon_h^u - \varepsilon_h^{\widehat{u}}\|_{\partial K} \leqslant C h_K^{\frac{1}{2}} \|\boldsymbol{q} - \boldsymbol{q}_h\|_K.$$

This completes the proof. □

Downloaded from https://academic.oup.com/imajna/article/42/2/1050/6104058 by University of Minnesota Libraries - Twin Cities user on 22 March 2023