Journal of Scientific Computing (2022) 90:92
https://doi.org/10.1007/s10915-022-01767-1

Error Analysis of an Unfitted HDG Method for a Class of Non-linear Elliptic Problems

Nestor Sánchez[1,2,4] · **Tonatiuh Sánchez-Vizuet**[3] · **Manuel Solano**[1,2]

Received: 12 May 2021 / Revised: 29 December 2021 / Accepted: 10 January 2022

Abstract
We study Hibridizable Discontinuous Galerkin (HDG) discretizations for a class of non-linear interior elliptic boundary value problems posed in curved domains where both the source term and the diffusion coefficient are non-linear. We consider the cases where the non-linear diffusion coefficient depends on the solution and on the gradient of the solution. To sidestep the need for curved elements, the discrete solution is computed on a polygonal subdomain that is not assumed to interpolate the true boundary, giving rise to an unfitted computational mesh. We show that, under mild assumptions on the source term and the computational domain, the discrete systems are well posed. Furthermore, we provide a priori error estimates showing that the discrete solution will have optimal order of convergence as long as the distance between the curved boundary and the computational boundary remains of the same order of magnitude as the mesh parameter.

Keywords Hybridizable discontinuous Galerkin · Non-linear boundary value problems · Curved boundary · Unfitted mesh · Transfer paths

Mathematics Subject Classification 65N08 · 65N30 · 65N85

1 Introduction

In this work we will study a discretization based on the hybridizable discontinous Galerkin (HDG) method [2] for a class of quasilinear elliptic boundary value problems of the form

$$-\nabla \cdot (\kappa \nabla u) = f(u) \qquad \text{in } \Omega \tag{1.1a}$$

✉ Tonatiuh Sánchez-Vizuet
tonatiuh@math.arizona.edu

1 Departamento de Ingeniería Matemática, Universidad de Concepción, Concepción, Chile

2 Centro de Investigación en Ingeniería Matemática (CI2MA), Universidad de Concepción, Concepción, Chile

3 Department of Mathematics, The University of Arizona, Tucson, USA

4 Instituto de Matemáticas, Unidad Juriquilla, Universidad Nacional Autónoma de México, Juriquilla, Mexico

Published online: 04 February 2022

$$u = g \qquad \text{on } \Gamma := \partial\Omega, \tag{1.1b}$$

where the domain $\Omega \subset \mathbb{R}^d$ $(d = 2, 3)$ is not necessarily polygonal/polyhedral, the diffusion coefficient, κ, is a positive function that depends on the solution, u, in one of the following functional forms

$$\kappa = \begin{cases} \kappa(u) \\ \kappa(\nabla u) \end{cases}, \tag{1.1c}$$

and that will be assumed to be a bounded and Lipschitz—in a sense that will be made precise in due time. In addition, the source function f will be taken to be a Lipschitz-continuous mapping from $L^2(\Omega)$ to $L^2(\Omega)$, so that there exists $L_f > 0$ such that

$$\|f(u_1) - f(u_2)\|_\Omega \leq L_f \|u_1 - u_2\|_\Omega \qquad \forall\, u_1, u_2 \in L^2(\Omega). \tag{1.2}$$

The authors became interested boundary value problems of this form through the study of magnetic equilibrium configurations in cylindrically symmetric fusion reactors. In the context of plasma equilibrium, if the location of the plasma within the reactor is assumed to be known, the equilibrium condition between magnetic and hydrostatic forces results in an equation like the one above, known as the Grad–Shafranov (or Grad–Shafranov–Schlüter) equation [11, 13, 19]. The domain Ω in which the equation holds, corresponds to the region of the reactor occupied by the plasma. This region in general has a non-polygonal, piecewise smooth Lipschitz boundary $\Gamma := \partial\Omega$ with a small number of corners—known in the plasma literature as *x-points*.

In this model, the unknown u is the stream function of the poloidal magnetic field, the source term f is a nonlinear function of u accounting for the effects of the hydrostatic pressure and total electric currents present in the device, and the coefficient κ encodes the magnetic properties of the system. The case where κ is a constant leads to a semi-linear equation for which an HDG discretization was proposed and implemented in [17, 18], and analyzed in detail in [16]. However, in the presence of ferroelectric materials, the permeability is affected by the total magnetic field $\mathbf{B}$—which is proportional to the gradient of u—and the coefficient then takes the form $\kappa = \kappa(\nabla u)$, leading to a quasi-linear equation that requires the more detailed treatment that will be the subject of this article. Some theoretical studies of the HDG method applied to quasilinear problems have been pursued recently [8–10], however these efforts are limited to polygonal domains. Moreover, the first reference does not consider non-linearities of the form $\kappa(\nabla u)$, while in [9, 10] the authors analyzed an augmented HDG discretization for a strictly quasi-linear problem arising from a non-linear Stokes flow using an approach based on a nonlinear version of the Babuska-Brezzi theory. A remarkable effort involving HDG and interpolatory methods for semilinear problems was recently carried out in [1, 6], as well as some recent contribution for the steady state incompressible Navier–Stokes equations [12]. An interesting application of HDG to some fully non-linear time dependent equations related to shallow water models was carried out in [14], while an HDG scheme for the equations of magnetohydrodynamics was analyzed in [15]. However in these works, either the domain is considered to be polygonal or non-linearities are restricted to the source term—or both. As we will show, our analysis will be valid for both quasi-linear and semi-linear problems, will not require an augmented formulation and the domain may be piecewise smooth.

Due to the non-polygonal nature of the domain of definition, any discretization scheme employed must handle the additional geometric complexity of the problem. The HDG scheme that we propose deals with the geometry using a method introduced within the context of HDG discretizations for linear elliptic equations in [7] and given firm theoretical justification in [5]. The idea consists of posing the discrete problem in a polygonal subdomain $\Omega_h \subset \Omega$

and transferring the boundary data to the computational boundary $\Gamma_h := \partial\Omega_h$ through a mechanism that involves a line integral of the numerical flux. This process requires reformulating the equation in the form of a first order system (mixed form), but greatly simplifies the computational implementation by allowing the use of a shape regular triangulation and avoiding the need for high order curved elements on the boundary. Moreover, in applications where the flux is a physically relevant variable—as is the case of the plasma problem—the direct discretization of the flux required by the mixed formulation is an additional advantage of the HDG formulation.

Depending on the location of the non-linearity within the equation, the analysis required to generalize [5, 7] into non-linear problems of the form (1.1) can be split into two independent parts. The case where κ is independent of the solution and only the source f is a function of u, has already been dealt with in [16]. The current communication will deal with situations where the source term is independent of u and κ takes one of the forms (1.1c), each of which must also be analyzed separately. We will start by introducing basic notation in Sect. 2, where and the idea of the extended domains and transfer paths will be presented. In this section we will also describe some geometric hypotheses on the the computational domain that will be useful for the error analysis. Having established the basic setting, we then proceed to study separately the HDG discretizations for the case when the diffusion coefficient is a function of u only (Sect. 3), and the case where the diffusion coefficient depends on ∇u (Sect. 4). In these sections the well posedness of the corresponding discrete HDG formulations are established, and a priori error analyses on the discretizations are performed.

2 Preliminaries

2.1 Computational Domains and Admissible Triangulations

Given the domain Ω where (1.1) is posed, we will define a family of polygonal subdomains and admissible triangulations approximating Ω where we will ultimately pose our discretization. First, consider a family of simply connected domains $\{\Omega_\alpha\}_{\alpha>0}$ such that, for every α the following conditions hold: (1) $\Omega_\alpha \subseteq \Omega$, (2) the boundary $\Gamma_\alpha := \partial\Omega_\alpha$ is a polygon, and (3) for every $\epsilon > 0$ there are infinitely many indices α such that $\lambda(\Omega\backslash\Omega_\alpha) < \epsilon$. In the preceding expression $\lambda(\cdot)$ denotes the Lebesgue measure. These conditions ensure that the family of subdomains $\{\Omega_\alpha\}_{\alpha>0}$ will exhaust Ω.

Having built the family $\{\Omega_\alpha\}_{\alpha>0}$ satisfying all the conditions above, the next step is to define a family of admissible simplicial triangulations $\{\mathcal{T}_h\}_{h>0}$. To be considered admissible, a triangulation $\mathcal{T}_h$ must be such that: (1) $\mathcal{T}_h$ is a triangulation for at least one $\Omega_h \in \{\Omega_\alpha\}_{\alpha>0}$ (we will identify both $\mathcal{T}_h$ and the respective domain Ω_h with the same subscript h, adding copies of Ω_h to $\{\Omega_\alpha\}_{\alpha>0}$ if necessary to account for different triangulations of the same domain) (2) it is shape regular, meaning that there exists $\beta > 0$ such that for all elements $T \in \mathcal{T}_h$ and all $h > 0$, $h_T/\rho_T \leq \beta$, where h_T is the diameter of T and ρ_T is the diameter of the largest ball contained in T, and (3) for every $T \in \mathcal{T}_h$ such that $T \cap \Gamma_h \neq \varnothing$, the maximum distance between $\boldsymbol{x} \in T \cap \Gamma_h$ and $\boldsymbol{y} \in \Gamma$ is of the same order of magnitude as the element diameter h_T. More precisely, if $d_{loc} := \max\{d(\boldsymbol{x}, \boldsymbol{y}) : \boldsymbol{x} \in T \cap \Gamma_h \text{ and } \boldsymbol{y} \in \Gamma\}$ then $d_{loc} = \mathcal{O}(h_T)$. This last requirement, which will be referred to as the *local proximity condition* and is depicted schematically in Fig. 1, is of key importance for the transfer process that will be defined later on.

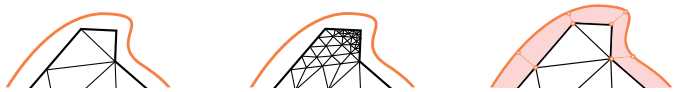

Fig. 1 Left and center: The proximity condition ensures that the distance between the computational and the physical boundaries remains always of the same order of magnitude as the local element diameter. The schematic shows close ups to the boundary of two triangulations of the same domain: an admissible triangulation satisfying the local proximity condition (left), and inadmissible one that violates the local proximity condition (center). Right: An admissible triangulation and one possible arrangement of extension patches T^e_{ext} (shaded in the figure) defined on the region $\Omega \backslash \Omega_h$

For every element T in a particular triangulation, we will denote by $\boldsymbol{n}_T$ the outward unit normal vector to T, or simply $\boldsymbol{n}$ instead of $\boldsymbol{n}_T$ whenever the context prevents any confusion. As it is conventional, we will denote the mesh parameter as $h := \max_{T \in \mathcal{T}_h} h_T$, which will be assumed to be smaller than one for the sake of simplicity. We will denote by e any face of a simplex and its length by h_e. Moreover, we will talk about an *interior face* e if there are two elements T^+ and T^- in the triangulation $\mathcal{T}_h$ such that $e = \partial T^+ \cap \partial T^-$. The set of all interior faces will be denoted by $\mathcal{E}_h^\circ$. In a similar manner, we will talk about a *boundary face* e if there is an element $T \in \mathcal{T}_h$ such that $e = \partial T \cap \Gamma_h$; the set of boundary faces will be denoted by $\mathcal{E}_h^\partial$. Note that with these definitions, the entirety of the faces of the triangulation denoted by $\mathcal{E}_h$ (often referred to as the *skeleton* of the mesh) can then be decomposed as $\mathcal{E}_h = \mathcal{E}_h^\circ \cup \mathcal{E}_h^\partial$.

Throughout the paper, we will be working with functions that in general will not be continuous across mesh elements. For a scalar-valued function we will use the symbol $[\![w]\!] := w^+ - w^-$ to refer to its jump across any given interior face. At the boundary faces, the jump will be defined as $[\![w]\!] := w - \varphi_h$, where φ_h is the approximation of the boundary data at Γ_h that will be defined later. In the case of vector-valued functions $\boldsymbol{v}$, we will be interested in the discontinuity of its normal component across interior faces, which will be denoted by $[\![\boldsymbol{v}]\!] := \boldsymbol{v}^+ \cdot \boldsymbol{n}^+ + \boldsymbol{v}^- \cdot \boldsymbol{n}^-$.

A remark on the local proximity condition and mesh refinement: The local proximity condition limits the minimum size that the elements near the boundary of an admissible triangulation can attain. Therefore, mesh refinement in this context must be understood as moving through a sequence of computational domains in the set $\{\Omega_h\}_{h>0}$ and their corresponding admissible triangulations in $\{\mathcal{T}_h\}_{h>0}$ as the parameter $h \to 0$. As it will be shown later, the error estimates will not depend on the particular domain Ω_h or triangulation $\mathcal{T}_h$ as long as the three requirements on the mesh stated above are satisfied. Possible ways of building sequences of admissible triangulations and computational domians have been detailed in [7] for uniform meshes and in [18] for adaptively refined triangulations.

2.2 The Extended Domain

Having defined the family of polygonal subdomains and admissible triangulations on which the discretization will be performend, we will now proceed to detail the process through which the boundary information will be transferred from the boundary into the computational domain. In order to do that we will have to tessellate the region enclosed between the two boundaries Γ and Γ_h as follows.

Given a triangulation $\mathcal{T}_h$ of the computational domain Ω_h and a boundary face $e \in \mathcal{E}_h^\partial$, we will denote by T^e the unique element of $\mathcal{T}_h$ such that $e \cap \overline{T^e} = e$. To every point $\boldsymbol{x} \in e$, we will associate—in a smooth fashion—a point $\overline{\boldsymbol{x}} \in \Gamma$ and set $l(\boldsymbol{x}) = |\boldsymbol{x} - \overline{\boldsymbol{x}}|$. We will define

the *extension patch* T^e_{ext} as

$$T^e_{ext} := \{\boldsymbol{x} + s\boldsymbol{t} : 0 \leq s \leq l(\boldsymbol{x}), \boldsymbol{x} \in e\},$$

where $\boldsymbol{t} = \boldsymbol{t}(\boldsymbol{x})$ is the unit vector anchored at $\boldsymbol{x}$ and pointing in the direction of $\overline{\boldsymbol{x}}$. With this notation, the line segment connecting $\boldsymbol{x}$ to $\overline{\boldsymbol{x}}$ can be parameterized by

$$\sigma_{\boldsymbol{t}}(\boldsymbol{x}) := \{\boldsymbol{x} + s\boldsymbol{t} : s \in [0, l(\boldsymbol{x})]\}.$$

The point $\overline{\boldsymbol{x}} \in \Gamma$ and therefore the vector $\boldsymbol{t}(\boldsymbol{x})$ can be specified in several ways. Here, we will consider that the point has been determined in such a way that

$$\boldsymbol{t}(\boldsymbol{x}) = \boldsymbol{n} \text{ for all } \boldsymbol{x} \in e. \tag{2.1}$$

This assumption is made with the sole purpose of making the analysis simpler, it can in fact be relaxed to the existence of a constant a_0 such that $0 < a_0 \leq \boldsymbol{t}(\boldsymbol{x}) \cdot \boldsymbol{n}$ for every $\boldsymbol{x}$ belonging to a boundary edge. The numerical method described here is remarkably robust with respect to the method used to choose $\boldsymbol{t}(\boldsymbol{x})$. Previously, the direction had been determined using the algorithm proposed by [7], which assigns $\overline{\boldsymbol{x}}$ in such a way that the three following conditions are satisfied: (1) $\overline{\boldsymbol{x}}$ is unique, (2) any two different line segments $\sigma_{\boldsymbol{t}}$ do not intersect each other inside T^e_{ext}, and (3) the segments $\sigma_{\boldsymbol{t}}$ do not intersect the interior of Ω_h. As it is proven in the aforementioned reference, these three conditions guarantee that the union of T^e_{ext} completely covers $\Omega^c_h := \Omega \backslash \overline{\Omega_h}$. An alternate method was used and tested numerically in [17, 18], where $\boldsymbol{t}(\boldsymbol{x})$ was determined using a weighted average of the normal vectors from neighboring boundary edges.

For an extended patch T^e_{ext} and a mesh element $T^e \in \mathcal{T}_h$ sharing a boundary face $e \in \mathcal{E}^{\partial}_h$, we denote by $h^{\perp}_e$ (resp. $H^{\perp}_e$) the largest distance between a point inside T_e (resp. T^e_{ext}) and the plane determined by the face e. The ratio between these two distances will be denoted by $r_e := H^{\perp}_e / h^{\perp}_e$ and the maximum such ratio taken over all the boundary edges will be denoted by $R_{\mathcal{T}_h} := \max_{e \in \mathcal{E}^{\partial}_h} r_e$. Note that the local proximity condition will ensure the existence of a constant R, that we will call the *proximity constant*, independent of the particular triangulation $\mathcal{T}_h$, such that

$$R := \sup_{\mathcal{T}_h} R_{\mathcal{T}_h} = \sup_{\mathcal{T}_h} \left\{ \max_{e \in \mathcal{E}^{\partial}_h} r_e \right\} < \infty. \tag{2.2}$$

We will also define the class of non-trivial vector-valued polynomials of degree at most k defined in and across both patches as

$$\mathcal{V}^k := \left\{ \boldsymbol{p} \in [\mathbb{P}_k(T^e_{ext} \cup T^e)]^2 \ : \ \boldsymbol{p} \cdot \boldsymbol{n}_e \neq \boldsymbol{0} \right\}.$$

We can then introduce, for all those elements with a non-empty intersection with the computational boundary, the element-wise constants

$$C^e_{ext} := \frac{1}{\sqrt{r_e}} \sup_{\boldsymbol{\chi} \in \mathcal{V}^k} \frac{\|\boldsymbol{\chi} \cdot \boldsymbol{n}_e\|_{T^e_{ext}}}{\|\boldsymbol{\chi} \cdot \boldsymbol{n}_e\|_{T^e}} \quad \text{and} \quad C^e_{inv} := h^{\perp}_e \sup_{\boldsymbol{\chi} \in \mathcal{V}^k} \frac{\|\nabla \boldsymbol{\chi} \cdot \boldsymbol{n}_e\|_{T^e}}{\|\boldsymbol{\chi} \cdot \boldsymbol{n}_e\|_{T^e}},$$

where, in abuse of notation, $\boldsymbol{n}_e$ is a constant vector field defined in $T^e_{ext} \cup T^e \cup e$ that coincides with the unit exterior normal vector associated to the face e and pointing in the direction of the extension patch. Above, the norms $\|\cdot\|_{T^e_{ext}}$ and $\|\cdot\|_{T^e}$ are the standard L^2 norms supported on the extension patch T^e_{ext} and its neighboring element T^e respectively. In [5], these constants were bounded in terms of the polynomial degree of the approximation, k, and the regularity

constant of the mesh, β, as

$$C^e_{ext} \leq C_1(k+1)^2(3\beta+2)^k\,, \qquad C^e_{inv} \leq C_2 k^2, \tag{2.3}$$

where C_1 and C_2 depend only on the mesh regularity. As we will see below, these constants will determine the magnitude of the proximity constant and therefore the maximum admissible gap between the computational and physical boundaries.

We will also need to make two technical assumptions relating the proximity constant to the and the diffusion coefficient κ and the degree of the polynomial approximation. For each $e \in \mathcal{E}^\partial_h$ we will require the following to hold:

$$H^\perp_e \leq \frac{1}{3}\underline{\kappa}\,\overline{\tau}^{-1}, \tag{2.4a}$$

$$\overline{\kappa}\,\underline{\kappa}^{-1}\, r_e^3\, (C^e_{ext}\, C^e_{inv})^2 \leq 1, \tag{2.4b}$$

where $\underline{\kappa}$ and $\overline{\kappa}$ are the lower and upper bounds of κ, resp., specified in (3.1), whereas $\overline{\tau}$ is the maximum of the stabilization parameter τ of the HDG scheme. The first of these two conditions, (2.4a), states the well known fact that for small values of the diffusivity, small scale behavior can be expected near the physical boundary, and therefore fine extension patches are required. However, it also provides the additional insight that the distance between the boundaries can be increased at the cost of accepting smaller values of the stabilization factor τ—and hence larger discontinuities in the discrete solution. In a similar vein, (2.4b) relates the range of values of the diffusion coefficient with the maximum separation between the computational and physical boundaries, thus making sure that the external patches are fine enough to resolve possible boundary behavior induced by large variations in diffusivity over the domain. Moreover, it sets a hard upper limit to the mechanism that allows for a larger separation by decreasing τ.

By combining (2.3) with (2.4) it is not hard to show that, for $k > 0$, $H^\perp_e$ must be bounded as

$$H^\perp_e \leq \min\left\{ \frac{h^\perp_e}{(C_1C_2)^2}\left(\frac{\overline{\kappa}^{-1}\underline{\kappa}}{k^4(k+1)^4(3\beta+1)^{2k}}\right)^{1/3},\ \frac{1}{3}\underline{\kappa}\,\overline{\tau}^{-1}\right\}.$$

This expression provides insight into the way in which the physics of the problem—through the range of values for κ—interacts with the discretization—through the parameters $H^\perp_e, h^\perp_e, k, \beta$, and τ—and determines the maximum separation between the physical boundary and that of an admissible triangulation. Of particular note is the role played by the polynomial degree of the approximation: for larger values of k the distance between the mesh and the boundary must decrease. The reason for this will become apparent soon, as we will resort to extrapolation to approximate some quantities over the extension patches.

2.3 An Equivalent Mixed Formulation and the Transfer Paths

Having established the requirements for an admissible triangulation, and defined the extension patches T^e_{ext} in such a way that for each of them there corresponds a single $T^e \in \mathcal{T}_h$, we can define a way to extend polynomial functions from the computational domain into Ω^c_h. This extension process will enable us to transfer the boundary condition from Γ into the computational boundary Γ_h. Let $p : T^e \to \mathbb{R}$ be a polynomial function and T^e_{ext} an extension patch associated to T^e. We will define the extension $\boldsymbol{E}_h(p)$ of p to T^e_{ext} by extrapolation as follows:

$$\begin{aligned} E_h : \quad & \mathbb{P}_k(T^e) \longrightarrow \mathbb{P}_k(T^e \cup T^e_{ext}) \\ & p(\boldsymbol{y}) \quad \forall\, \boldsymbol{y} \in T^e \longmapsto p(\boldsymbol{y}) \quad \forall\, \boldsymbol{y} \in T^e \cup T^e_{ext}. \end{aligned}$$

Where, to keep notation simple, a polynomial function p should be understood as its extrapolation $E_h(p)$ whenever an evaluation outside of Ω_h is required, which should be clear from the context. For vector-valued polynomial functions, the extension is defined similarly component by component.

To avoid computing on a domain with curved boundary, we wish to pose the boundary vaue problem (1.1) polygonal subdomain $\Omega_h \subset \Omega$. This simpler geometry can then be discretized by a uniform triangulation of size satisfying the conditions outlined in Sects. 2.1 and 2.2.

Recasting (1.1) in mixed form and restricting the resulting equivalent first order system to Ω_h leads to

$$\boldsymbol{q} + \kappa\, \nabla u = 0 \qquad \text{in } \Omega_h \tag{2.5a}$$

$$\nabla \cdot \boldsymbol{q} = f(u) \qquad \text{in } \Omega_h, \tag{2.5b}$$

$$u = \varphi \qquad \text{on } \Gamma_h := \partial\Omega_h, \tag{2.5c}$$

where the specific relation between κ, u and ∇u has not been made explicit, and the—a priori unknown—function φ encodes the restriction of u to the computational boundary Γ_h. We can recover φ following the method proposed by [4] (in one dimension) and extended to higher dimensions by [7]. The idea consists of transferring the Dirichlet data g from Γ to Γ_h along segments called *transfer paths* by computing a line integral of the flux $\boldsymbol{q}$.

To be precise, given $\boldsymbol{x} \in \Gamma_h$ and $\overline{\boldsymbol{x}} \in \Gamma$, equation (2.5a) can be integrated along the segment connecting them. Lets denote by $\boldsymbol{t}(\boldsymbol{x})$ the unit vector anchored at $\boldsymbol{x}$ pointing towards $\overline{\boldsymbol{x}}$, and by $l(\boldsymbol{x})$ the length of the segment connecting them. We then have the following representation for φ:

$$\varphi := g(\overline{\boldsymbol{x}}) + \int_0^{l(\boldsymbol{x})} (\kappa^{-1}\, \boldsymbol{q})(\boldsymbol{x} + \boldsymbol{t}(\boldsymbol{x})s) \cdot \boldsymbol{t}(\boldsymbol{x}) ds. \tag{2.6}$$

Note that φ depends on the values of either u or ∇u (through κ^{-1}), and $\boldsymbol{q}$ over the extended domain Ω_h^c. As such, we should write $\varphi = \varphi(u, \nabla u, \boldsymbol{q}, \boldsymbol{x})$ however, to keep notation simple, we will abstain from this and will write simply φ. In a similar fashion, $g(\overline{\boldsymbol{x}})$ is in fact a function of $\boldsymbol{x}$, since the point $\overline{\boldsymbol{x}}$ varies smoothly with $\boldsymbol{x}$. To avoid the use of cumbersome notation we will write either $g(\overline{\boldsymbol{x}})$ or simply $\overline{g} := g(\overline{\boldsymbol{x}}(\boldsymbol{x}))$.

2.4 Sobolev Space Notation

To denote spaces of functions we will make use of the standard notation and terminology from Sobolev space theory. Let $\mathcal{O}$ be a domain in $\mathbb{R}^d$, and Σ be either a Lipschitz curve (if $d = 2$) or surface (if $d = 3$); for scalar-valued functions and non zero real numbers s, we will use the spaces $H^s(\mathcal{O})$ and $H^s(\Sigma)$ with their usual definition, whereas for the case $s = 0$ we will write simply $L^2(\mathcal{O})$ and $L^2(\Sigma)$. The spaces of vector-valued functions will be denoted in bold face, therefore $\boldsymbol{H}^s(\mathcal{O}) := [H^s(\mathcal{O})]^d$ and $\boldsymbol{H}^s(\Sigma) := [H^s(\Sigma)]^d$.

The L^2 inner products for both scalar and vector-valued functions on volumes and surfaces will be denoted by $(\cdot,\cdot)_{\mathcal{O}}$ and $\langle\cdot,\cdot\rangle_{\Sigma}$ respectively. The associated norms will be denoted by $\|\cdot\|_{s,\mathcal{O}}$ and $\|\cdot\|_{s,\Sigma}$ and simply $\|\cdot\|_{\mathcal{O}}$ for the case $s = 0$. As is common, will we write $|\cdot|_{s,\mathcal{O}}$ for the $\boldsymbol{H}^s$ and H^s-semi norms.

Given a triangulation $\mathcal{T}_h$ we will define the following mesh-dependent inner products over elements and edges

$$(\cdot,\cdot)_{\mathcal{T}_h} := \sum_{T\in\mathcal{T}_h} (\cdot,\cdot)_T, \qquad \langle\cdot,\cdot\rangle_{\partial\mathcal{T}_h} := \sum_{T\in\mathcal{T}_h} \langle\cdot,\cdot\rangle_{\partial T} \quad \text{and} \quad \langle\cdot,\cdot\rangle_{\Gamma_h} := \sum_{e\in\mathcal{E}_h^\partial} \langle\cdot,\cdot\rangle_e.$$

These inner products induce mesh-dependent norms that will be denoted, respectively, by

$$\|\cdot\|_{\Omega_h} := \left(\sum_{T\in\mathcal{T}_h} \|\cdot\|_T^2\right)^{1/2}, \qquad \|\cdot\|_{\partial\mathcal{T}_h} := \left(\sum_{T\in\mathcal{T}_h} \|\cdot\|_{\partial T}^2\right)^{1/2} \quad \text{and}$$

$$\|\cdot\|_{\Gamma_h} := \left(\sum_{e\in\mathcal{E}_h^\partial} \|\cdot\|_e^2\right)^{1/2}.$$

In the forthcoming analysis, the expression $a \lesssim b$ should be understood as meaning $a \le Cb$ where C is a positive constant independent of h.

For the discrete formulations that will be introduced in the next sections, we will make use of the following finite dimensional spaces of piece-wise polynomial functions

$$\boldsymbol{V}_h := \{\boldsymbol{v} \in \boldsymbol{L}^2(\mathcal{T}_h) : \boldsymbol{v}|_T \in [\mathbb{P}_k(T)]^d, \quad \forall\, T \in \mathcal{T}_h\}, \tag{2.7a}$$

$$W_h := \{w \in L^2(\mathcal{T}_h) : w|_T \in \mathbb{P}_k(T), \quad \forall\, T \in \mathcal{T}_h\}, \tag{2.7b}$$

$$M_h := \{\mu \in L^2(\mathcal{E}_h) : \mu|_T \in \mathbb{P}_k(e), \quad \forall\, e \in \mathcal{E}_h\}, \tag{2.7c}$$

where, $\mathbb{P}_k(T)$ denotes the space of polynomials of degree at most k defined in $T \in \mathcal{T}_h$. Similarly, $\mathbb{P}_k(e)$ denotes the space of polynomials of degree at most k defined over a face $e \in \mathcal{E}_h$.

3 Non-linearities of the Form $\kappa(u)$

3.1 The HDG Formulation

We will first consider the case when the coefficient κ depends on the solution in the form

$$\begin{aligned} \kappa : L^2(\Omega) &\longrightarrow L^\infty(\overline{\Omega}) \\ u &\longmapsto \kappa(u). \end{aligned}$$

For the analysis, we will require the existence of positive constants $\underline{\kappa}$ and $\overline{\kappa}$ such that for all $u \in L^2(\Omega)$

$$\underline{\kappa} \le \kappa(u) \le \overline{\kappa} \quad \text{almost everywhere in } \Omega. \tag{3.1}$$

Moreover, κ will be assumed to be Lipschitz-continuous on $L^2(\Omega)$, i.e, there exists $\tilde{L} > 0$ such that

$$\|\kappa(u_1) - \kappa(u_2)\|_{L^\infty(\Omega)} \le \tilde{L}\|u_1 - u_2\|_{L^2(\Omega)} \qquad \forall\, u_1, u_2 \in L^2(\Omega). \tag{3.2}$$

The two conditions above, together, imply the existence of constants $\widehat{L}$ and L such that

$$\|\kappa^{1/2}(u_1) - \kappa^{1/2}(u_2)\|_{L^\infty(\Gamma)} \le \widehat{L}\|u_1 - u_2\|_{L^2(\Omega)} \qquad \forall\, u_1, u_2 \in L^2(\Omega), \tag{3.3}$$

$$\|\kappa^{-1}(u_1) - \kappa^{-1}(u_2)\|_{L^\infty(\overline{\Omega})} \le L\|u_1 - u_2\|_{L^2(\Omega)} \qquad \forall\, u_1, u_2 \in L^2(\Omega). \tag{3.4}$$

Note that all these assumptions imply the Lipschitz continuity of κ, $\kappa^{1/2}$, and κ^{-1} on the subdomain $\Omega_h \subset \Omega$ with corresponding Lipschitz constants equal to or smaller than those stated above.

Before introducing the discrete formulation we will recall here the mixed form (2.5), but now we make explicit the dependence $\kappa = \kappa(u)$

$$\boldsymbol{q} + \kappa(u)\,\nabla u = 0 \qquad \text{in } \Omega_h, \tag{3.5a}$$

$$\nabla \cdot \boldsymbol{q} = f(u) \qquad \text{in } \Omega_h, \tag{3.5b}$$

$$u = \varphi \qquad \text{on } \partial\Omega_h. \tag{3.5c}$$

The boundary data φ on the computational boundary Γ_h is transferred according to (2.6).

Taking an admissible triangulation $\mathcal{T}_h$ of the computational domain Ω_h, the HDG discretization of (3.5) reads: Find $(\boldsymbol{q}_h, u_h, \widehat{u}_h) \in \boldsymbol{V}_h \times W_h \times M_h$, such that

$$(\kappa^{-1}(u_h)\,\boldsymbol{q}_h, \boldsymbol{v})_{\mathcal{T}_h} - (u_h, \nabla \cdot \boldsymbol{v})_{\mathcal{T}_h} + \langle \widehat{u}_h, \boldsymbol{v} \cdot \boldsymbol{n} \rangle_{\partial\mathcal{T}_h} = 0, \tag{3.6a}$$

$$-(\boldsymbol{q}_h, \nabla w)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}, w \rangle_{\partial\mathcal{T}_h} = (f(u_h), w)_{\mathcal{T}_h}, \tag{3.6b}$$

$$\langle \widehat{u}_h, \mu \rangle_{\Gamma_h} = \langle \varphi_h(u_h), \mu \rangle_{\Gamma_h}, \tag{3.6c}$$

$$\langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}, \mu \rangle_{\partial\mathcal{T}_h \setminus \Gamma_h} = 0, \tag{3.6d}$$

for all $(\boldsymbol{v}, w, \mu) \in \boldsymbol{V}_h \times W_h \times M_h$. Here

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

with τ being a positive stabilization function, whose maximum will be denoted by $\overline{\tau}$. The approximate boundary condition φ_h on the right hand side of (3.6c) is given by the discrete counterpart of (2.6)

$$\varphi_h(u_h)(\boldsymbol{x}) := g(\overline{\boldsymbol{x}}) + \int_0^{l(\boldsymbol{x})} (\kappa^{-1}(u_h)\, E_h \boldsymbol{q}_h)(\boldsymbol{x} + \boldsymbol{t}(\boldsymbol{x})s) \cdot \boldsymbol{t}(\boldsymbol{x}))ds, \quad \text{for} \quad \boldsymbol{x} \in \Gamma_h. \tag{3.6e}$$

In the definition above, we have used the extrapolation $E_h \boldsymbol{q}_h$ due to the fact that the approximation $\boldsymbol{q}_h$ is available only inside of the computational domain Ω_h, but the transfer paths along which the integral is computed are defined over the complementary extended region Ω_h^c.

3.2 Well-Posedness

In this section we employ a Banach fixed-point argument to ensure the well-posedness of the discrete problem (3.6). To that end we will define an operator $\mathcal{J} : W_h \to W_h$ mapping ζ to the second component of the triplet $(\boldsymbol{q}, u, \widehat{u}) \in \boldsymbol{V}_h \times W_h \times M_h$ satisfying, for all $(\boldsymbol{v}, w, \mu) \in \boldsymbol{V}_h \times W_h \times M_h$, the HDG system (3.6) where the source has been evaluated at ζ, namely

$$(\kappa^{-1}(\zeta)\,\boldsymbol{q}, \boldsymbol{v})_{\mathcal{T}_h} - (u, \nabla \cdot \boldsymbol{v})_{\mathcal{T}_h} + \langle \widehat{u}, \boldsymbol{v} \cdot \boldsymbol{n} \rangle_{\partial\mathcal{T}_h} = 0, \tag{3.7a}$$

$$-(\boldsymbol{q}, \nabla w)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{q}} \cdot \boldsymbol{n}, w \rangle_{\partial\mathcal{T}_h} = (f(\zeta), w)_{\mathcal{T}_h}, \tag{3.7b}$$

$$\langle \widehat{u}, \mu \rangle_{\Gamma_h} = \langle \varphi(\zeta), \mu \rangle_{\Gamma_h}, \tag{3.7c}$$

$$\langle \widehat{\boldsymbol{q}} \cdot \boldsymbol{n}, \mu \rangle_{\partial\mathcal{T}_h \setminus \Gamma_h} = 0. \tag{3.7d}$$

Above, the term $\varphi(\zeta)$ corresponds to the boundary condition transferred to the computational domain by means of (3.6e). The mapping $\mathcal{J}$ is well defined, as the linearized system (3.7) is uniquely solvable as proven in [5].

The main result of this section—that the mapping $\mathcal{J}$ defined above is a contraction—relies on the validity of a particular inequality—estimate (3.8) below—but is otherwise a simple argument. Since the proof of (3.8) requires a sequence of technical arguments, in the interest of clarity (we will first prove the main theorem assuming that the aforementioned inequality is valid. After having established the well posedness of the discrete problem, the reminder of the section will be devoted to verifying the validity of (3.8). This will be finally established in Lemma 4, after a series of auxiliary results.

Theorem 1 (Well-posedness of the discrete problem) *Suppose that Assumptions* (2.4) *are satisfied and that additionally*

$$(\sqrt{3}\,\widehat{c} + \overline{\kappa}^{1/2}\,R^{1/2})\,\|l^{-1/2}\,\overline{g}\|_{\Gamma_h}\,\widehat{L}\,h^{1/2} < 1/4,$$
$$\max\{\widehat{c}^2 h,\,1\}\,L_f < 1/8,$$

where $\widehat{c}$ is the constant given in Lemma 4. Then $\mathcal{J}$ is a contraction operator.

Proof Let $\zeta_1, \zeta_2 \in W_h$ and define $u_1 := \mathcal{J}(\zeta_1)$ and $u_2 := \mathcal{J}(\zeta_2)$. Then u_1 and u_2 are the second components of solutions to (3.7) and Lemma 4 guarantees that

$$\|\mathcal{J}(\zeta_1) - \mathcal{J}(\zeta_2)\|_{\Omega_h} = \|u_1 - u_2\|_{\Omega_h} \leq 4\,\max\{\widehat{c}^2 h,\,1\}\,\|f(\zeta_1) - f(\zeta_2)\|_{\Omega_h} + 2\,(\sqrt{3}\,\widehat{c} + \overline{\kappa}^{1/2}\,R^{1/2})\,h^{1/2}\,\|(\kappa^{1/2}(\zeta_1) - \kappa^{1/2}(\zeta_2))\,l^{-1/2}\,\overline{g}\|_{\Gamma_h}. \tag{3.8}$$

Then, applying the Lipschitz-continuity of f and $\kappa^{1/2}$—given respectively in (1.2) and (3.2)—we get

$$\begin{aligned}
&\leq 4\,\max\{\widehat{c}^2 h,\,1\}\,L_f\,\|\zeta_1 - \zeta_2\|_{\Omega_h} + 2\,(\sqrt{3}\,\widehat{c} + \overline{\kappa}^{1/2}\,R^{1/2})\,h^{1/2}\,\|\kappa^{1/2}(\zeta_1) \\
&\quad - \kappa^{1/2}(\zeta_2)\|_{L^\infty(\Gamma_h)}\,\|l^{-1/2}\,\overline{g}\|_{\Gamma_h} \\
&\leq 4\,\max\{\widehat{c}^2 h,\,1\}\,L_f\,\|\zeta_1 - \zeta_2\|_{\Omega_h} + 2\,(\sqrt{3}\,\widehat{c} + \overline{\kappa}^{1/2}\,R^{1/2})\,\widehat{L}\,h^{1/2}\,\|\zeta_1 \\
&\quad - \zeta_2\|_{\Omega_h}\,\|l^{-1/2}\,\overline{g}\|_{\Gamma_h}.
\end{aligned}$$

The result follows from the hypothesis for $\widehat{L}$ and L_f. □

Combined with a standard fixed-point argument, the theorem above guarantees the existence and uniqueness of the solution to the discrete HDG system. We will now direct our efforts to showing that inequality (3.8) holds. To that end we will make use of the following auxiliary function and its properties listed in the lemma below—the proof of which can be found on [5,Lemma 5.2].

Lemma 1 *Consider $\boldsymbol{x} \in \Gamma_h$ and any smooth enough function $\boldsymbol{v}$ defined in $T^e \cup T^e_{ext}$, and define*

$$\delta_{\boldsymbol{v}}(\boldsymbol{x}) := \frac{1}{l(\boldsymbol{x})}\int_0^{l(\boldsymbol{x})} [\boldsymbol{v}(\boldsymbol{x} + \boldsymbol{n}s) - \boldsymbol{v}(\boldsymbol{x})]\cdot\boldsymbol{n}\,ds. \tag{3.9}$$

The following estimates hold for each $e \in \mathcal{E}_h^\partial$:

$$\|l^{1/2}\,\delta_{\boldsymbol{v}}\|_e \leq \frac{1}{\sqrt{3}}\,r_e^{3/2}\,C^e_{ext}\,C^e_{inv}\,\|\boldsymbol{v}\|_{T^e} \qquad \forall\,\boldsymbol{v} \in [\mathbb{P}_k(T)]^d, \tag{3.10a}$$

$$\|l^{1/2}\,\delta_{\boldsymbol{v}}\|_e \leq \frac{1}{\sqrt{3}}\,r_e\,\|h^\perp \partial_n \boldsymbol{v}\cdot\boldsymbol{n}\|_{T^e_{ext}} \qquad \forall\,\boldsymbol{v} \in [H^1(T)]^d. \tag{3.10b}$$

$$\|l^{1/2}\,\delta_{\boldsymbol{v}}\|_\infty \leq \frac{1}{\sqrt{3}}\,r_e\,\sup_{\boldsymbol{x}\in e}\|h_e^\perp\,\partial_n \boldsymbol{v}\cdot\boldsymbol{n}\|_{l(\boldsymbol{x})} \qquad \forall\,\boldsymbol{v} \in [H^1(T)]^d. \tag{3.10c}$$

For the subsequent analysis, we will also make use of the following norm. Given $(\boldsymbol{v}, w, \mu) \in \boldsymbol{V}_h \times W_h \times M_h$ and $\xi \in M_h$ we define

$$|||(\boldsymbol{v}, w, \mu)|||_\xi := \left(\|\kappa^{-1/2}(\xi)\boldsymbol{v}\|^2_{\Omega_h} + \|\tau^{1/2}\, w\|^2_{\partial\mathcal{T}_h} + \|\kappa^{1/2}(\xi)\, l^{-1/2}\, \mu(\xi)\|^2_{\Gamma_h}\right)^{1/2}. \quad (3.11)$$

The proof of (3.8) requires considering the solution ϕ to an auxiliary problem (c.f. (B.1)) and using a duality argument that connects a stability estimate for ϕ with our variables of interest. This will be done in Lemma 4. Lemmas 2 and 3 below establish estimates relating the norm (3.11) of the solution $(\boldsymbol{q}, u, \widehat{u})$ to problem data that will be used in the final step of Lemma 4.

Lemma 2 *Let φ be the transferred boundary condition appearing in (3.7) and suppose that assumptions (2.4) are satisfied. It holds*

$$\begin{aligned}
\langle\varphi(\zeta), \delta_{\boldsymbol{q}}\rangle_{\Gamma_h} &\le \frac{1}{6}\,\|\kappa^{1/2}(\zeta)\, l^{-1/2}\,\varphi(\zeta)\|^2_{\Gamma_h} + \frac{1}{2}\,\|\kappa^{-1/2}(\zeta)\,\boldsymbol{q}\|^2_{\Omega_h},\\
\langle\varphi(\zeta), \tau(u-\widehat{u})\rangle_{\Gamma_h} &\le \frac{1}{6}\|\kappa^{1/2}(\zeta)\, l^{-1/2}\,\varphi(\zeta)\|^2_{\Gamma_h} + \frac{1}{2}\|\tau^{1/2}(u-\widehat{u})\|^2_{\partial\mathcal{T}_h},\\
\langle\varphi(\zeta), \kappa(\zeta)\, l^{-1}\,\overline{g}\rangle_{\Gamma_h} &\le \frac{1}{6}\|\kappa^{1/2}(\zeta)\, l^{-1/2}\,\varphi(\zeta)\|^2_{\Gamma_h} + \frac{3}{2}\|\kappa^{1/2}(\zeta)\, l^{-1/2}\,\overline{g}\|^2_{\Gamma_h},
\end{aligned}$$

where $\overline{g}(\boldsymbol{x}) = g(\overline{\boldsymbol{x}}(\boldsymbol{x}))\quad \forall\, \boldsymbol{x} \in \Gamma_h$.

Proof The first inequality is obtained after applying Young's inequality and estimate (B.5a), whereas the second inequality follows from assumption (2.2) and (2.4a). The third inequality follows from Young's inequality exclusively. □

Lemma 3 *If Assumptions (2.4) hold, then*

$$|||(\boldsymbol{q}, u-\widehat{u}, \varphi)|||^2_\zeta \le 2\|f(\zeta)\|_{\Omega_h}\|u\|_{\Omega_h} + 3\|\kappa^{1/2}(\zeta)\, l^{-1/2}\,\overline{g}\|^2_{\Gamma_h}.$$

Proof Let $\zeta \in W_h$ and $u = \mathcal{J}(\zeta) \in W_h$. Since u defined this way is the solution to the discrete system (3.7), then testing (3.7) with

$$\boldsymbol{v} = \boldsymbol{q}, \quad w = u, \quad \mu := \begin{cases} -\widehat{\boldsymbol{q}}\cdot\boldsymbol{n} & \text{on } \Gamma_h,\\ -\widehat{u} & \text{on } \partial\mathcal{T}_h\backslash\Gamma_h \end{cases}$$

we deduce that

$$\|\kappa^{-1/2}(\zeta)\,\boldsymbol{q}\|^2_{\Omega_h} + \|\tau^{1/2}\,(u-\widehat{u})\|^2_{\partial\mathcal{T}_h} = -\,\langle\varphi(\zeta), \widehat{\boldsymbol{q}}\cdot\boldsymbol{n}\rangle_{\Gamma_h} + (f(\zeta), u)_{\mathcal{T}_h}. \quad (3.12)$$

On the other hand, we can use the definition of φ and $\delta_{\boldsymbol{q}}$ (cf. (3.6e) and (B.4)) to show that

$$\widehat{\boldsymbol{q}}\cdot\boldsymbol{n} = \kappa(\zeta)\, l^{-1}(\varphi(\zeta) - \overline{g}) - \delta_{\boldsymbol{q}} + \tau(u-\widehat{u}).$$

Substituting the expression for $\widehat{\boldsymbol{q}}\cdot\boldsymbol{n}$ above in (3.12), we obtain

$$\begin{aligned}
&\|\kappa^{-1/2}(\zeta)\,\boldsymbol{q}\|^2_{\Omega_h} + \|\tau^{1/2}\,(u-\widehat{u})\|^2_{\partial\mathcal{T}_h} + \|\kappa^{1/2}(\zeta)\, l^{-1/2}\,\varphi(\zeta)\|^2_{\Gamma_h}\\
&\quad = |||(\boldsymbol{q}, u-\widehat{u}, \varphi)|||^2_\zeta\\
&\quad \le |\langle\varphi(\zeta), \kappa(\zeta)\, l^{-1}\,\overline{g}\,\rangle_{\Gamma_h}| + |\langle\varphi(\zeta), \delta_{\boldsymbol{q}}\rangle_{\Gamma_h}| + |\langle\varphi(\zeta), \tau(u-\widehat{u})\rangle_{\Gamma_h}| + |(f(\zeta), u)_{\mathcal{T}_h}|.
\end{aligned}$$

Now, using Lemma 2 to estimate the first three terms in the right hand of this expression we obtain

$$\frac{1}{2}\,|||(\boldsymbol{q}, u-\widehat{u}, \varphi)|||^2_\zeta \le \|f(\zeta)\|_{\Omega_h}\|u\|_{\Omega_h} + \frac{3}{2}\|\kappa^{1/2}(\zeta)\, l^{-1/2}\,\overline{g}\|^2_{\Gamma_h},$$

whereupon the proof is concluded. □

For the following result, we will make use of the properties of the HDG projectors $\boldsymbol{\Pi}_{\boldsymbol{V}}$ and Π_W onto the discrete spaces $\boldsymbol{V}_h$ and W_h. This projection was first introduced in [3] and we include its definition and main properties in the "Appendix A". The L^2 projector onto the space M_h will be denoted by P_M, while Id_M will denote the identity on M_h.

Lemma 4 *Suppose that Assumptions (2.4) and the regularity (B.2) are satisfied. Then, there exists $\widehat{c} > 0$, independent of h such that*

$$\|u\|_{\Omega_h} \leq 4\,\max\{\widehat{c}^2 h, 1\}\,\|f(\zeta)\|_{\Omega_h} + 2\left(\sqrt{3}\,\widehat{c} + \overline{\kappa}^{1/2}\,R^{1/2}\right) h^{1/2}\,\|\kappa^{1/2}(\zeta)\,l^{-1/2}\,\overline{g}\|_{\Gamma_h}. \tag{3.13}$$

Proof Consider $\Theta \in L^2(\Omega)$ and let ϕ and ψ be the solutions to the dual problem (B.1) associated to Θ. If we define

$$\mathbb{T}_{\boldsymbol{q}} := (\kappa^{-1}(\zeta)\boldsymbol{q}, \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{\phi} - \boldsymbol{\phi})_{\mathcal{T}_h}, \quad \mathbb{T}_u := \langle \widehat{u}, P_M(\boldsymbol{\phi}\cdot\boldsymbol{n})\rangle_{\Gamma_h} - \langle \widehat{\boldsymbol{q}}\cdot\boldsymbol{n}, \Pi_W\psi\rangle_{\Gamma_h}$$
$$\text{and} \quad \mathbb{T}_f := (f(\zeta), \Pi_W\psi)_{\mathcal{T}_h},$$

it is possible to verify that

$$(u, \Theta)_{\mathcal{T}_h} = \mathbb{T}_{\boldsymbol{q}} + \mathbb{T}_f + \mathbb{T}_u. \tag{3.14}$$

The terms $\mathbb{T}_{\boldsymbol{q}}$ and $\mathbb{T}_f$ appearing on the expression above can be easily estimated by

$$|\mathbb{T}_{\boldsymbol{q}}| \lesssim \underline{\kappa}^{-1/2}\,h\|\kappa^{-1/2}(\zeta)\,\boldsymbol{q}\|_{\Omega_h}\,\|\Theta\|_{\Omega}, \quad \text{and} \quad |\mathbb{T}_f| \lesssim \|f(\zeta)\|_{\Omega_h}\,\|\Theta\|_{\Omega}. \tag{3.15}$$

In order to bound the final term of the decomposition of $(u, \Theta)_{\mathcal{T}_h}$, we rewrite $\mathbb{T}_u = \sum_{i=1}^{5}\mathbb{T}_u^i$ where

$$\mathbb{T}_u^1 := -\langle \kappa(\zeta) l^{-1}\varphi(\zeta), \psi + l\partial_n\psi\rangle_{\Gamma_h}, \quad \mathbb{T}_u^4 := -\langle \tau(u-\widehat{u}), P_M\psi\rangle_{\Gamma_h},$$
$$\mathbb{T}_u^2 := \langle \kappa(\zeta)\varphi(\zeta), (P_M - Id_M)\partial_n\psi\rangle_{\Gamma_h}, \ \mathbb{T}_u^5 := \langle \kappa(\zeta)\,l^{-1}\,\overline{g}, \psi\rangle_{\Gamma_h},$$
$$\mathbb{T}_u^3 := \langle \delta_{\boldsymbol{q}}, \psi\rangle_{\Gamma_h}.$$

It is not hard, if cumbersome, to verify that for the terms above the following estimates hold

$$|\mathbb{T}_u^1| \lesssim \overline{\kappa}^{1/2}\,R\,h\,\|\kappa^{1/2}(\zeta)\,l^{-1/2}\,\varphi(\zeta)\|_{\Gamma_h}\|\Theta\|_{\Omega}, \quad |\mathbb{T}_u^2| \lesssim \overline{\kappa}^{1/2}\,R^{1/2}\,h\,\|\kappa^{1/2}(\zeta)l^{-1/2}\varphi(\zeta)\|_{\Gamma_h}\|\Theta\|_{\Omega},$$
$$|\mathbb{T}_u^3| \lesssim \overline{\kappa}^{1/2}\,R^2\,h^{1/2}\|\kappa^{-1/2}(\zeta)\boldsymbol{q}\|_{\Omega_h}\|\Theta\|_{\Omega}, \quad |\mathbb{T}_u^4| \lesssim \overline{\tau}^{1/2}\,R\,h\,\|\tau^{1/2}(u-\widehat{u})\|_{\partial\mathcal{T}_h}\|\Theta\|_{\Omega},$$
$$|\mathbb{T}_u^5| \lesssim \overline{\kappa}^{1/2}\,(Rh)^{1/2}\,\|\kappa^{1/2}(\zeta)\,l^{-1/2}\,\overline{g}\|_{\Gamma_h}\|\Theta\|_{\Omega}.$$

Taking $\Theta = u$ in (3.14) and combining all of the above estimates with (3.15), we obtain

$$\|u\|_{\Omega_h} \leq \widehat{c}\,h^{1/2}|||(\boldsymbol{\sigma}, u-\widehat{u}, \varphi)|||_{\zeta} + \overline{\kappa}^{1/2}\,(Rh)^{1/2}\|\kappa^{1/2}(\zeta)\,l^{-1/2}\,\overline{g}\|_{\Gamma_h} + \|f(\zeta)\|_{\Omega_h},$$

where $\widehat{c} := C\,\max\{\underline{\kappa}^{-1/2}, \overline{\kappa}^{1/2}R, \overline{\kappa}^{1/2}R^2, \overline{\kappa}^{1/2}R^{1/2}, \overline{\tau}^{1/2}R\}$, and $C > 0$ is the constant hidden in the symbol $\lesssim$. Then, applying Lemma 3, we get

$$\begin{aligned}\|u\|_{\Omega_h} \leq\ & \widehat{c}\,h^{1/2}\left(\sqrt{2}\|f(\zeta)\|_{\Omega_h}^{1/2}\|u\|_{\Omega_h}^{1/2} + \sqrt{3}\|\kappa^{1/2}(\zeta)\,l^{-1/2}\,\overline{g}\|_{\Gamma_h}\right)\\ &+ \overline{\kappa}^{1/2}\,(Rh)^{1/2}\|\kappa^{1/2}(\zeta)\,l^{-1/2}\,\overline{g}\|_{\Gamma_h} + \|f(\zeta)\|_{\Omega_h}\\ \|u\|_{\Omega_h} \leq\ & 4\,\max\{\widehat{c}^2 h, 1\}\,\|f(\zeta)\|_{\Omega_h} + 2\,(\sqrt{3}\,\widehat{c} + \overline{\kappa}^{1/2}\,R^{1/2})\,h^{1/2}\,\|\kappa^{1/2}(\zeta)\,l^{-1/2}\,\overline{g}\|_{\Gamma_h},\end{aligned}$$

with which the proof is concluded. □

3.3 A Prior Error Analysis

We now provide the a priori error bounds for the discretization error. The main results of the section are Theorem 3.16 and Corollary 1 immediately after it. As we will see, several of the results leading to the main presented in this section can be proven by using similar arguments to those of Sect. 3.2 and we will omit some of the arguments. The analysis will be performed by decomposing the approximation errors in two components using the properties of the HDG projection (see "Appendix A"). The *projection of the errors* is defined as

$$\boldsymbol{\varepsilon}^{\boldsymbol{q}} := \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} - \boldsymbol{q}_h \quad \text{and} \quad \varepsilon^u := \Pi_W u - u_h,$$

and the *error of the projections* are given by

$$\boldsymbol{I}^{\boldsymbol{q}} := \boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} \quad \text{and} \quad I^u := u - \Pi_W u.$$

This allows to express the approximation errors as

$$\boldsymbol{q} - \boldsymbol{q}_h = \boldsymbol{\varepsilon}^{\boldsymbol{q}} + \boldsymbol{I}^{\boldsymbol{q}} \quad \text{and} \quad u - u_h = \varepsilon^u + I^u.$$

In addition, recalling that P_M is the L^2 projection into M_h, we define the projection error for the hybrid unknown $\widehat{u}_h$ as $\varepsilon^{\widehat{u}} := P_M u - \widehat{u}_h$. The L^2-projection of the error for the numerical flux on $\partial\mathcal{T}_h$ can be expressed as $\boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} \cdot \boldsymbol{n} = \boldsymbol{\varepsilon}^{\boldsymbol{q}} \cdot \boldsymbol{n} + \tau(\varepsilon^u - \varepsilon^{\widehat{u}})$. It is not difficult show that $(\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u, \varepsilon^{\widehat{u}})$ belongs to $\boldsymbol{V}_h \times W_h \times M_h$ and satisfies

$$(\kappa^{-1}(u_h)\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \boldsymbol{v})_{\mathcal{T}_h} - (\varepsilon^u, \nabla\cdot\boldsymbol{v})_{\mathcal{T}_h} + \langle \varepsilon^{\widehat{u}}, \boldsymbol{v}\cdot\boldsymbol{n}\rangle_{\partial\mathcal{T}_h} = -(\kappa^{-1}(u)\boldsymbol{I}^{\boldsymbol{q}}, v)_{\mathcal{T}_h} - \left((\kappa^{-1}(u) - \kappa^{-1}(u_h))\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}, v\right)_{\mathcal{T}_h}, \tag{3.16a}$$

$$-(\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \nabla w)_{\mathcal{T}_h} + \langle \boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}}\cdot\boldsymbol{n}, w\rangle_{\partial\mathcal{T}_h} = (f(u) - f(u_h), w)_{\mathcal{T}_h}, \tag{3.16b}$$

$$\langle \varepsilon^{\widehat{u}}, \mu\rangle_{\Gamma_h} = \langle \varphi(u) - \varphi_h(u_h), \mu\rangle_{\Gamma_h}, \tag{3.16c}$$

$$\langle \boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}}\cdot\boldsymbol{n}, \mu\rangle_{\partial\mathcal{T}_h\setminus\Gamma_h} = 0, \tag{3.16d}$$

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

To try and keep the notation compact, we will define the following two quantities involving only the errors in the projections $\boldsymbol{I}^{\boldsymbol{v}}$, $\boldsymbol{I}^{\boldsymbol{q}}$, and I^u measured in the three relevant domains Ω_h, Ω_h^c and Γ_h

$$\Lambda_{\boldsymbol{q}} := \left(\|\boldsymbol{I}^{\boldsymbol{q}}\|^2_{\Omega_h} + \|h^{\perp}\partial_n(\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n})\|^2_{\Omega_h^c} + \|(h^{\perp})^{1/2}\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n})\|^2_{\Gamma_h}\right)^{1/2}, \tag{3.17a}$$

$$\Lambda_u := \left(\|(h^{\perp})^{1/2}I^u\|_{\Gamma_h} + \|I^u\|_{\Omega_h}\right)^{1/2}. \tag{3.17b}$$

We note that, as pointed out in [3, 16] by using the properties of the projectors and scaling arguments, if $\boldsymbol{q} \in \boldsymbol{H}^{k+1}(\Omega)$, $u \in H^{k+1}(\Omega)$ and τ is of order one, then $\Lambda_{\boldsymbol{q}}$ and Λ_u are of order h^{k+1}. As stated in the theorem below, these quantities are in fact the key to estimating the approximation error of the method

Theorem 2 *If L is small enough, the regularity (B.2) holds and the discrete spaces are of polynomial degree $k \geq 1$, then there exists $h_0 > 0$ such that, for all $h \leq h_0$, we have*

$$|||(\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||^2_{u_h} \lesssim \Lambda^2_{\boldsymbol{q}} + \Lambda^2_u. \tag{3.18}$$

The proof of this result will follow straightforwardly from Lemmas 5 and 6 below.

Before setting out to prove these two lemmas (and therefore the theorem above) we first state the convergence order of the method—the main result of the section—which thanks to the remark made just above Theorem 2 follows as a corollary.

Corollary 1 (Order of convergence) *Suppose that the assumptions of Theorem 2 hold. If, in addition, $u \in H^{k+1}(\Omega)$ and $\boldsymbol{q} \in \boldsymbol{H}^{k+1}(\Omega)$, then*

$$\|\boldsymbol{q} - \boldsymbol{q}_h\|_\Omega + \|u - u_h\|_\Omega \leq Ch^{k+1} \left(|u|_{k+1,\Omega} + |\boldsymbol{q}|_{k+1,\Omega}\right).$$

Having stated the main results of the section, we now set out to prove the two lemmas leading to Theorem 2. The first part of the analysis will require using an energy argument on the error equations (3.16) and a meticulous study of the error contribution due to the transferred boundary conditions $\varphi_h(u_h)$. This will be done in the following

Lemma 5 *There exist positive constants, independent of h, such that*

$$|||(\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||_{u_h}^2 \leq 12 \max\{C_1 h, C_2\} L^2 \left(\|\varepsilon^u\|_{\Omega_h}^2 + \|I^u\|_{\Omega_h}^2\right) + C_3 \Lambda_{\boldsymbol{q}}^2. \quad (3.19)$$

Proof Starting from (3.16) and letting

$$\boldsymbol{v} = \boldsymbol{\varepsilon}^{\boldsymbol{q}}, \quad \text{and} \quad w = \varepsilon^u \quad \text{in } \mathcal{T}_h, \quad \text{and} \quad \mu := \begin{cases} -\boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} \cdot \boldsymbol{n} & \text{on } \Gamma_h \\ -\varepsilon^{\widehat{u}} & \text{on } \partial\mathcal{T}_h \backslash \Gamma_h \end{cases}$$

it follows that

$$\begin{aligned} \|\kappa^{-1/2}(u_h)\,\boldsymbol{\varepsilon}^{\boldsymbol{q}}\|_{\Omega_h}^2 + \|\tau^{1/2}(\varepsilon^u - \varepsilon^{\widehat{u}})\|_{\partial\mathcal{T}_h}^2 &= -(\kappa^{-1}(u)\boldsymbol{I}^{\boldsymbol{q}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}})_{\mathcal{T}_h} \\ &- ((\kappa^{-1}(u) - \kappa^{-1}(u_h))\Pi_{\boldsymbol{V}}\boldsymbol{q}, \boldsymbol{\varepsilon}^{\boldsymbol{q}})_{\mathcal{T}_h} + (f(u) - f(u_h), \varepsilon^u)_{\mathcal{T}_h} - \langle \varphi(u) - \varphi_h(u_h), \boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} \cdot \boldsymbol{n}\rangle_{\Gamma_h}. \end{aligned} \quad (3.20)$$

We will now manipulate the final term in the expression above to include a term involving the norm of the difference $\varphi(u) - \varphi_h(u_h)$, thus allowing us to estimate the transfer error. Using definitions of φ_h and φ (cf. (2.6) and (3.6e) respectively), as well as the definition of $\delta_{\boldsymbol{q}}$ it follows that

$$\varphi(u) - \overline{g} = \kappa^{-1}(u)\ell\left(\delta_{\boldsymbol{q}} + \boldsymbol{q} \cdot \boldsymbol{n}\right) \quad \text{and} \quad \varphi_h(u_h) - \overline{g} = \kappa^{-1}(u_h)\ell\left(\delta_{\boldsymbol{q}_h} + \boldsymbol{q}_h \cdot \boldsymbol{n}\right).$$

Subtracting the second expression from the first one and adding zero in the form of $\pm\kappa^{-1}(u_h)\left(\delta_{\boldsymbol{q}} - \boldsymbol{q} \cdot \boldsymbol{n}\right)$ it is possible to express the difference as

$$\begin{aligned} \varphi(u) - \varphi_h(u_h) &= \ell\left(\kappa^{-1}(u) - \kappa^{-1}(u_h)\right)(\delta_{\boldsymbol{q}} + \boldsymbol{q} \cdot \boldsymbol{n}) + \ell\,\kappa^{-1}(u_h)\left(\delta_{\boldsymbol{q}-\boldsymbol{q}_h} + (\boldsymbol{q} - \boldsymbol{q}_h) \cdot \boldsymbol{n}\right) \\ &= \ell\left(\kappa^{-1}(u) - \kappa^{-1}(u_h)\right)\left(\delta_{\boldsymbol{q}} + \boldsymbol{q} \cdot \boldsymbol{n}\right) + \ell\,\kappa^{-1}(u_h)\left(\delta_{\boldsymbol{\varepsilon}^{\boldsymbol{q}}} + \delta_{\boldsymbol{I}^{\boldsymbol{q}}} + \left(\boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} + \boldsymbol{I}^{\boldsymbol{q}}\right)\right) \\ &\quad \cdot \boldsymbol{n} - \tau\left(\varepsilon^u - \varepsilon^{\widehat{u}}\right) = \ell\left(\kappa^{-1}(u) - \kappa^{-1}(u_h)\right)\left(\delta_{\boldsymbol{I}^{\boldsymbol{q}}} + \delta_{\Pi_v\boldsymbol{q}} + \left(\boldsymbol{I}^{\boldsymbol{q}} + \Pi_{\boldsymbol{V}}\boldsymbol{q}\right) \cdot \boldsymbol{n}\right) \\ &\quad + \ell\kappa^{-1}(u_h)\left(\delta_{\boldsymbol{\varepsilon}^{\boldsymbol{q}}} + \delta_{\boldsymbol{I}^{\boldsymbol{q}}} + \left(\boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} + \boldsymbol{I}^{\boldsymbol{q}}\right) \cdot \boldsymbol{n} - \tau\left(\varepsilon^u - \varepsilon^{\widehat{u}}\right)\right), \end{aligned} \quad (3.21)$$

where the first equality comes from the substitutions

$$(\boldsymbol{q} - \boldsymbol{q}_h) \cdot \boldsymbol{n} = \left(\boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} + \boldsymbol{I}^{\boldsymbol{q}}\right) \cdot \boldsymbol{n} - \tau\left(\varepsilon^u - \varepsilon^{\widehat{u}}\right), \quad \text{and} \quad \delta_{\boldsymbol{q}+\boldsymbol{q}_h} = \delta_{\boldsymbol{q}} + \delta_{\boldsymbol{q}_h},$$

while the second one is obtained by replacing $\boldsymbol{q} = \boldsymbol{I}^{\boldsymbol{q}} + \Pi_{\boldsymbol{V}}\boldsymbol{q}$. The expression (3.21) allows us to write the term $\boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} \cdot \boldsymbol{n}$ in terms of the transfer error

$$\begin{aligned} \boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} \cdot \boldsymbol{n} &= \kappa(u_h)\, l^{-1}\left(\varphi(u) - \varphi_h(u_h)\right) - \kappa(u_h)\left(\kappa^{-1}(u) - \kappa^{-1}(u_h)\right) \\ &\quad (\delta_{\boldsymbol{I}^{\boldsymbol{q}}} + \delta_{\Pi_{\boldsymbol{V}}\boldsymbol{q}} + \boldsymbol{I}^{\boldsymbol{q}} \cdot \boldsymbol{n} + \Pi_{\boldsymbol{V}}\boldsymbol{q} \cdot \boldsymbol{n}) \\ &\quad - \delta_{\boldsymbol{\varepsilon}^{\boldsymbol{q}}} - \delta_{\boldsymbol{I}^{\boldsymbol{q}}} - \boldsymbol{I}^{\boldsymbol{q}} \cdot \boldsymbol{n} + \tau(\varepsilon^u - \varepsilon^{\widehat{u}}). \end{aligned}$$

Substituting this expression back into (3.20) and rearranging terms, it follows that

$$\begin{aligned}\|\kappa^{-1/2}(u_h)\,\boldsymbol{\varepsilon^q}\|^2_{\Omega_h}+\|\tau^{1/2}(\varepsilon^u-\varepsilon^{\widehat{u}})\|^2_{\partial\mathcal{T}_h}+\|\kappa^{1/2}(u_h)\,l^{-1/2}\,(\varphi(u)-\varphi_h(u_h))\|^2_{\Gamma_h}\\ \leq|(\kappa^{-1}(u)\boldsymbol{I^q},\boldsymbol{\varepsilon^q})_{\mathcal{T}_h}|+|((\kappa^{-1}(u)-\kappa^{-1}(u_h))\Pi_{\boldsymbol{V}}\boldsymbol{q},\boldsymbol{\varepsilon^q})_{\mathcal{T}_h}|\\ +|\mathbb{T}^f|+|\mathbb{T}_\varphi|,\end{aligned}\tag{3.22}$$

with $\mathbb{T}^f:=(f(u)-f(u_h),\varepsilon^u)_{\mathcal{T}_h}$ and $\mathbb{T}_\varphi:=\sum_{i=1}^8|\mathrm{T}^i_\varphi|$, where

$$\begin{array}{ll}\mathbb{T}^1_\varphi:=\langle\varphi(u)-\varphi_h(u_h),\delta_{\boldsymbol{\varepsilon^q}}\rangle_{\Gamma_h} & \mathbb{T}^5_\varphi:=\langle\varphi(u)-\varphi_h(u_h),\kappa(u_h)\,(\kappa^{-1}(u)-\kappa^{-1}(u_h))\,\delta_{\Pi_{\boldsymbol{V}}\boldsymbol{q}}\rangle_{\Gamma_h}\\ \mathbb{T}^2_\varphi:=-\langle\varphi(u)-\varphi_h(u_h),\tau(\varepsilon^u-\varepsilon^{\widehat{u}})\rangle_{\Gamma_h} & \mathbb{T}^6_\varphi:=\langle\varphi(u)-\varphi_h(u_h),\kappa(u_h)\,(\kappa^{-1}(u)-\kappa^{-1}(u_h))\,\delta_{\boldsymbol{I^q}}\rangle_{\Gamma_h}\\ \mathbb{T}^3_\varphi:=\langle\varphi(u)-\varphi_h(u_h),\boldsymbol{I^q}\cdot\boldsymbol{n}\rangle_{\Gamma_h} & \mathbb{T}^7_\varphi:=\langle\varphi(u)-\varphi_h(u_h),\kappa(u_h)\,(\kappa^{-1}(u)-\kappa^{-1}(u_h))\,\boldsymbol{I^q}\cdot\boldsymbol{n}\rangle_{\Gamma_h}\\ \mathbb{T}^4_\varphi:=\langle\varphi(u)-\varphi_h(u_h),\delta_{\boldsymbol{I^q}}\rangle_{\Gamma_h} & \mathbb{T}^8_\varphi:=\langle\varphi(u)-\varphi_h(u_h),\kappa(u_h)\,(\kappa^{-1}(u)-\kappa^{-1}(u_h))\,\Pi_{\boldsymbol{V}}\boldsymbol{q}\cdot\boldsymbol{n}\rangle_{\Gamma_h}.\end{array}$$

To determine upper bounds the terms in the right hand side of (3.22), we will make use of Young's inequality, the Lipschitz continuity of f and κ^{-1} and the fact that $\|v\|_{L^2(e)}\lesssim h_e^{1/2}\|v\|_e$, $\forall\, e\in\mathcal{E}_h^\partial,\forall\, v\in\mathbb{P}_k(e)$. A combination of these with arguments similar as those in [16,Lemma 5] results in the following

$$\begin{aligned}|\mathbb{T}^1_\varphi|&\leq\frac{1}{2\delta_1}\|\kappa^{1/2}(u_h)\,l^{-1/2}(\varphi(u)-\varphi_h(u_h))\|^2_{\Gamma_h}+\frac{\delta_1}{6}\|\kappa^{-1/2}(u_h)\boldsymbol{\varepsilon^q}\|^2_{\Omega_h},\\ |\mathbb{T}^2_\varphi|&\leq\frac{1}{2\delta_1}\|\kappa^{1/2}(u_h)\,l^{-1/2}(\varphi(u)-\varphi_h(u_h))\|^2_{\Gamma_h}+\frac{\delta_1}{6}\|\tau^{1/2}(\varepsilon^u-\varepsilon^{\widehat{u}})\|^2_{\partial\mathcal{T}_h},\\ |\mathbb{T}^3_\varphi|&\leq\frac{1}{2\delta_2}\|\kappa^{1/2}(u_h)\,l^{-1/2}(\varphi(u)-\varphi_h(u_h))\|^2_{\Gamma_h}+\frac{\delta_2}{2}R\underline{\kappa}^{-1}\|(h^\perp)^{1/2}\boldsymbol{I^q}\cdot\boldsymbol{n})\|^2_{\Gamma_h},\\ |\mathbb{T}^4_\varphi|&\leq\frac{1}{2\delta_2}\|\kappa^{1/2}(u_h)\,l^{-1/2}(\varphi(u)-\varphi_h(u_h))\|^2_{\Gamma_h}+\frac{\delta_2}{6}\underline{\kappa}^{-1}\max_{e\in\mathcal{E}_h^\partial}\{r_e^2\}\|h^\perp\partial_n(\boldsymbol{I^q}\cdot\boldsymbol{n})\|^2_{\Omega_h^c},\\ |\mathbb{T}^5_\varphi|&\leq\frac{1}{2\delta_2}\|\kappa^{1/2}(u_h)\,l^{-1/2}(\varphi(u)-\varphi_h(u_h))\|^2_{\Gamma_h}\\ &\quad+\frac{\delta_2}{6}\overline{\kappa}\,\|\Pi_{\boldsymbol{V}}\boldsymbol{q}\|^2_{L^\infty(\Omega_h)}\,h\,L^2\,\left(\|\varepsilon^u\|_{\Omega_h}+\|I^u\|_{\Omega_h}\right)^2,\\ |\mathbb{T}^6_\varphi|&\leq\frac{1}{2\delta_2}\|\kappa^{1/2}(u_h)\,l^{-1/2}(\varphi(u)-\varphi_h(u_h))\|^2_{\Gamma_h}+\frac{2\delta_2}{3}\overline{\kappa}\,\underline{\kappa}^{-2}\max_{e\in\mathcal{E}_h^\partial}r_e^2\,\|h^\perp\partial_n(\boldsymbol{I^q}\cdot\boldsymbol{n})\|^2_{\Omega_h^c},\\ |\mathbb{T}^7_\varphi|&\leq\frac{1}{2\delta_2}\|\kappa^{1/2}(u_h)\,l^{-1/2}(\varphi(u)-\varphi_h(u_h))\|^2_{\Gamma_h}+2\,\delta_2\,R\,\overline{\kappa}\,\underline{\kappa}^{-2}\,\|(h^\perp)^{1/2}\boldsymbol{I^q}\cdot\boldsymbol{n})\|^2_{\Gamma_h},\\ |\mathbb{T}^8_\varphi|&\leq\frac{1}{2\delta_2}\|\kappa^{1/2}(u_h)\,l^{-1/2}(\varphi(u)-\varphi_h(u_h))\|^2_{\Gamma_h}\\ &\quad+\frac{\delta_2}{2}\overline{\kappa}\,\|(h^\perp)^{1/2}\Pi_{\boldsymbol{V}}\boldsymbol{q}\cdot\boldsymbol{n}\|_{L^\infty(\Gamma_h)}L^2\,h\,\left(\|\varepsilon^u\|_{\Omega_h}+\|I^u\|_{\Omega_h}\right)^2,\end{aligned}$$

as well as

$$\begin{aligned}|(\kappa^{-1}(u)\boldsymbol{I^q},\boldsymbol{\varepsilon^q})_{\mathcal{T}_h}|&\leq\frac{1}{2\,\delta_3}\|\kappa^{-1/2}(u_h)\boldsymbol{\varepsilon^q}\|^2_{\Omega_h}+\frac{\delta_3}{2}\,\underline{\kappa}^{-2}\,\overline{\kappa}\|\boldsymbol{I^q}\|^2_{\Omega_h},\\ |(\kappa^{-1}(u)-\kappa^{-1}(u_h))\Pi_{\boldsymbol{V}}\boldsymbol{q},\boldsymbol{\varepsilon^q})_{\mathcal{T}_h}|&\leq\frac{1}{2\,\delta_3}\|\kappa^{-1/2}(u_h)\boldsymbol{\varepsilon^q}\|^2_{\Omega_h}\\ &\quad+\frac{\delta_3}{2}\|\Pi_{\boldsymbol{V}}\boldsymbol{q}\|^2_{L^\infty(\Omega_h)}\,\overline{\kappa}\,L^2\left(\|\varepsilon^u\|_{\Omega_h}+\|I^u\|_{\Omega_h}\right)^2.\\ |\mathbb{T}^f|&\leq L_f\,(\|\varepsilon^u\|_{\Omega_h}+\|I^u\|_{\Omega_h})\,\|\varepsilon^u\|_{\Omega_h},\end{aligned}$$

where $\delta_1, \delta_2, \delta_3$ are free positive parameters arising from applications of Young's inequality, $\overline{\kappa}$ and $\underline{\kappa}$ are the upper and lower bounds for the diffusivity, and L and L_f are the Lipschitz constants from κ^{-1} and f respectively.

If we let $\delta_1 = 4$, $\delta_2 = 12$, and $\delta_3 = 6$ in the above estimates and substitute back into (3.22) we obtain

$$\begin{aligned}
&|||(\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||_{u_h}^2 \\
&\quad \leq 12 \max\{C_1 h, C_2\} L^2 \left(\|\varepsilon^u\|_{\Omega_h}^2 + \|I^u\|_{\Omega_h}^2 \right) \\
&\qquad + C\, L_f (\|\varepsilon^u\|_{\Omega_h} + \|I^u\|_{\Omega_h}) \|\varepsilon^u\|_{\Omega_h} + C_3\, \Lambda_{\boldsymbol{q}}^2.
\end{aligned}$$

where C_1, C_2 and C_3 only depend on $\overline{\kappa}$, $\underline{\kappa}$, R, and the projections $\|(h^{\perp})^{1/2} \Pi_{\boldsymbol{V}} \boldsymbol{q} \cdot \boldsymbol{n}\|_{L^\infty(\Gamma_h)}^2$ and $\|\Pi_{\boldsymbol{V}} \boldsymbol{q}\|_{L^\infty(\Omega_h)}^2$. □

We now proceed to show that the approximation error in u can be indeed controlled by the errors in the approximation of the flux, the hybrid variable and the transfer error, modulo the approximation properties of the discrete spaces. To show that, in the next lemma we will build upon the ideas as in [16] and use a duality argument.

Lemma 6 *Assume that the Lipschitz constant is such that L_f is small enough, and consider the discrete spaces to be of polynomial degree $k \geq 1$. Then,*

$$\|\varepsilon^u\|_{\Omega_h} \lesssim h^{1/2} |||(\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||_{u_h} + (h^{1/2} + L\, h)\Lambda_{\boldsymbol{q}} + (L + h^{1/2})\Lambda_u. \tag{3.23}$$

Proof The first part of the proof follows very closely the argument used in the proof of Lemma 4. Given $\Theta \in L^2(\Omega)$ we will denote by $\boldsymbol{\phi}$ the solution to the dual problem (B.1) associated to Θ. Considering then the equations (3.16), together with the dual system, it is possible to show that

$$(\varepsilon^u, \Theta)_{\mathcal{T}_h} = \mathbb{T}_{\boldsymbol{q}}^1 + \mathbb{T}_{\boldsymbol{q}}^2 + \mathbb{T}_u + \mathbb{T}_f, \tag{3.24}$$

where

$$\begin{aligned}
&\mathbb{T}_{\boldsymbol{q}}^1 := (\kappa^{-1}(u_h)(\boldsymbol{q} - \boldsymbol{q}_h), \boldsymbol{\Pi}_{\boldsymbol{V}} \boldsymbol{\phi})_{\mathcal{T}_h} + (\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \nabla\psi)_{\mathcal{T}_h}, \qquad \mathbb{T}_{\boldsymbol{q}}^2 := (\kappa^{-1}(u) - \kappa^{-1}(u_h))(\boldsymbol{I}^{\boldsymbol{q}} + \Pi_V \boldsymbol{q}), \Pi_V \phi)_{\mathcal{T}_h}, \\
&\mathbb{T}_u := \langle \varepsilon^{\widehat{u}}, P_M(\boldsymbol{\phi} \cdot \boldsymbol{n}) \rangle_{\Gamma_h} - \langle \boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} \cdot \boldsymbol{n}, \Pi_W \psi \rangle_{\Gamma_h} \qquad \mathbb{T}_f := (f(u) - f(u_h), \Pi_W \psi)_{\mathcal{T}_h}.
\end{aligned}$$

To prove the result (3.23), we will bound each of the terms $\mathbb{T}_\star$, with $\star \in \{\boldsymbol{q}, u, f\}$ in the decomposition (3.24).

Bound for $\mathbb{T}_f^i$. The simplest term to bound is $\mathbb{T}_f$, for which an application of Cauchy-Schwartz, the properties of the HDG projector and the Lipschitz continuity of the source term f, together with the dual estimate (B.2) yield

$$|\mathbb{T}_f| \leq C_{\text{reg}}\, L_f\, (\|\varepsilon^u\|_{\Omega_h} + \|I^u\|_{\Omega_h}) \|\Theta\|_\Omega. \tag{3.25}$$

Where the constant C_{reg} is the stability constant from the dual problem (B.1).

Bound for $\mathbb{T}_{\boldsymbol{q}}^i$. Using the Lipschitz-continuity of κ (c.f. (3.2)) and following the arguments leading to equation (4.8) in [16,Lemma 5], the terms $\mathbb{T}_{\boldsymbol{q}}^1$ and $\mathbb{T}_{\boldsymbol{q}}^2$ can be bounded like

$$\begin{aligned}
|\mathbb{T}_{\boldsymbol{q}}^1| &\leq \underline{\kappa}^{-1/2} \|\kappa^{-1/2}(u_h)(\boldsymbol{\varepsilon}^{\boldsymbol{q}} + \boldsymbol{I}^{\boldsymbol{q}})\|_{\Omega_h} \|\boldsymbol{\Pi}_{\boldsymbol{V}} \phi - \phi\|_{\Omega_h} + \|\boldsymbol{I}^{\boldsymbol{q}}\|_{\Omega_h} \|\nabla(\psi - \psi_h)\|_{\Omega_h} \\
&\leq C\, C_{\text{reg}}\, \underline{\kappa}^{-1/2} h^{\min\{1,k\}} \|\kappa^{-1/2}(u_h)\boldsymbol{\varepsilon}^{\boldsymbol{q}}\|_{\Omega_h} \|\Theta\|_\Omega \\
&\quad + 2C\, C_{\text{reg}}\, \max\{\underline{\kappa}^{-1/2}\, \overline{\kappa}^{-1/2}, 1\} h^{\min\{1,k\}} \|\boldsymbol{I}^{\boldsymbol{q}}\|_{\Omega_h} \|\Theta\|_\Omega
\end{aligned} \tag{3.26a}$$

and

$$|\mathbb{T}_{\boldsymbol{q}}^2| \leq C_{\text{reg}}(\|\boldsymbol{I}^{\boldsymbol{q}}\|_\infty + \|\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}\|_\infty)\, L\, (\|\varepsilon^u\|_{\Omega_h} + \|I^u\|_{\Omega_h})\|\theta\|_\Omega. \tag{3.26b}$$

Bound for $\mathbb{T}_u$. To estimate this term we will have to decompose it and treat each of the parts separately. We will write then write $\mathbb{T}_u := \sum_{i=1}^{11} \mathbb{T}_u^i$, where:

$$\begin{aligned}
&\mathbb{T}_u^1 := -\langle \kappa(u_h)\, l^{-1}\, (\varphi(u) - \varphi(u_h)),\, \psi + l\partial_n\psi\rangle_{\Gamma_h}, \\
&\mathbb{T}_u^2 := -\langle \kappa(u_h)\, (\varphi(u) - \varphi_h(u_h)),\, (Id_M - P_M)\partial_n\psi\rangle_{\Gamma_h}, \\
&\mathbb{T}_u^3 := \langle \delta_{\boldsymbol{I}^{\boldsymbol{q}}},\, \psi\rangle_{\Gamma_h}, \\
&\mathbb{T}_u^4 := \langle \boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n},\, (Id_M - P_M)\psi\rangle_{\Gamma_h}, \\
&\mathbb{T}_u^5 := -\langle \tau\, P_M I^u,\, \psi\rangle_{\Gamma_h}, \\
&\mathbb{T}_u^6 := \langle \delta_{\boldsymbol{\varepsilon}^{\boldsymbol{q}}},\, \psi\rangle_{\Gamma_h}.
\end{aligned}
\qquad
\begin{aligned}
&\mathbb{T}_u^7 := -\langle \tau(\varepsilon^u - \widehat{\varepsilon^u}),\, P_M\psi\rangle_{\Gamma_h}, \\
&\mathbb{T}_u^8 := -\langle \kappa(u_h)\, (\varphi(u) - \varphi_h(u_h))\, \delta_{\boldsymbol{I}^{\boldsymbol{q}}},\, \psi\rangle_{\Gamma_h} \\
&\mathbb{T}_u^9 := -\langle \kappa(u_h)\, (\varphi(u) - \varphi_h(u_h))\, \delta_{\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}},\, \psi\rangle_{\Gamma_h}, \\
&\mathbb{T}_u^{10} := -\langle \kappa(u_h)\, (\varphi(u) - \varphi_h(u_h))\, \boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n},\, \psi\rangle_{\Gamma_h}. \\
&\mathbb{T}_u^{11} := -\langle \kappa(u_h)\, (\varphi(u) - \varphi_h(u_h))\, \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}\cdot\boldsymbol{n},\, \psi\rangle_{\Gamma_h},
\end{aligned}$$

Bounds for $\mathbb{T}_u^1 - \mathbb{T}_u^7$: These terms can be estimated by we applying the same techniques of [16,Lemma 6]. We will omit most of the the details here. Recalling that the length of the transfer path $l(\boldsymbol{x}) \leq c\, R\, h \quad \forall \boldsymbol{x} \in \Gamma_h$ and considering the constant $\tilde{c}$ from Lemma B1 [16,Lemma 6], we have

$$\begin{aligned}
&|\mathbb{T}_u^1| \leq c\,\tilde{c}\,\overline{\kappa}^{1/2}\, R\, h\, \|\kappa^{1/2}(u_h)\, l^{-1/2}\, (\varphi(u) - \varphi_h(u_h))\|_{\Gamma_h}\|\Theta\|_\Omega, \\
&|\mathbb{T}_u^2| \leq c\,\tilde{c}\,\overline{\kappa}^{1/2}\, R^{1/2}\, h\, \|\kappa^{1/2}(u_h)\, l^{-1/2}\, (\varphi(u) - \varphi_h(u_h))\|_{\Gamma_h}\|\Theta\|_\Omega, \\
&|\mathbb{T}_u^3| \leq \tfrac{1}{\sqrt{3}}\, c^{1/2}\,\tilde{c}\, R^{3/2}\, h^{1/2}\|h^\perp\partial_n\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n}\|_{\Omega_h^c}\|\Theta\|_\Omega, \\
&|\mathbb{T}_u^4| \leq \tilde{c}\, h\|(h^\perp)^{1/2}\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n}\|_{\Gamma_h}\|\Theta\|_\Omega, \\
&|\mathbb{T}_u^5| \leq \tilde{c}\,\overline{\tau}\, R\, h^{1/2}\|(h^\perp)^{1/2} I^u\|_{\Gamma_h}\|\Theta\|_\Omega, \\
&|\mathbb{T}_u^6| \leq \tfrac{1}{\sqrt{3}}\, c^{1/2}\,\tilde{c}\,\overline{\kappa}^{1/2}\, \max_{e\in\mathcal{E}_h^\partial}\{C_{\text{ext}}^e, C_{\text{inv}}^e\} R^2\, h^{1/2}\, \|\kappa^{-1/2}(u_h)\,\boldsymbol{\varepsilon}^{\boldsymbol{q}}\|_{\Omega_h}\|\Theta\|_\Omega, \\
&|\mathbb{T}_u^7| \leq c\,\tilde{c}\,\overline{\tau}^{1/2}\, R\, h\|\tau^{1/2}\,(\varepsilon^u - \widehat{\varepsilon^u})\|_{\partial\mathcal{T}_h}\|\Theta\|_\Omega.
\end{aligned}$$

Bounds for $\mathbb{T}_u^8 - \mathbb{T}_u^9$: Let us first notice that by definition of $\mathbb{T}_u^8$, we can obtain

$$|\mathbb{T}_u^8| = \left|\left\langle \kappa^{1/2}(u_h)\, l\, \kappa^{1/2}(u_h)\, l^{-1/2}(\varphi(u) - \varphi_h(u_h))\, l^{1/2}\, \delta_{\boldsymbol{I}^{\boldsymbol{q}}},\, l^{-1}\,\psi\right\rangle_{\Gamma_h}\right|.$$

Then, by Cauchy-Schwartz, the fact that $l(\boldsymbol{x}) \leq c\, R\, h \quad \forall \boldsymbol{x} \in \Gamma_h$ and the boundedness of κ, we can obtain

$$|\mathbb{T}_u^8| \leq c\,\overline{\kappa}^{1/2}\, R\, h\, \|\kappa^{1/2}(u_h)\, l^{-1/2}\, (\varphi(u) - \varphi_h(u_h))\, l^{1/2}\, \delta_{\boldsymbol{I}^{\boldsymbol{q}}}\|_{\Gamma_h}\|l^{-1}\,\psi\|_{\Gamma_h}.$$

Finally, a direct application of (B.5c) to the factor involving the function $\delta_{\boldsymbol{I}^{\boldsymbol{q}}}$, and using the estimation (B.3d) for the factor $\|l^{-1}\,\psi\|_{\Gamma_h}$, results in

$$|\mathbb{T}_u^8| \leq \tfrac{1}{\sqrt{3}}\, c\,\tilde{c}\,\overline{\kappa}^{1/2}\, R^2\, h\, \sup_{\boldsymbol{x}\in\Gamma_h}\|(h^\perp\,\partial_n\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n}\|_{l(\boldsymbol{x})}\, \|\kappa^{1/2}(u_h)\, l^{-1/2}\, (\varphi(u) - \varphi_h(u_h))\|_{\Gamma_h}\, \|\Theta\|_\Omega.$$

Analogously, we can show that

$$|\mathbb{T}_u^9| \leq \tfrac{1}{\sqrt{3}}\, c\,\tilde{c}\,\overline{\kappa}^{1/2}\, R^2\, h\, \sup_{\boldsymbol{x}\in\Gamma_h}\|(h^\perp\,\partial_n\boldsymbol{\Pi}_{\boldsymbol{v}}\boldsymbol{q}\cdot\boldsymbol{n}\|_{l(\boldsymbol{x})}\, \|\kappa^{1/2}(u_h)\, l^{-1/2}\, (\varphi(u) - \varphi_h(u_h))\|_{\Gamma_h}\, \|\Theta\|_\Omega.$$

Bounds for $\mathbb{T}_u^{10} - \mathbb{T}_u^{11}$: We start as in the case for $\mathbb{T}_u^{10} - \mathbb{T}_u^{11}$ by combining, Cauchy-Schwartz, the bounds for κ and $l \lesssim Rh$, to obtain

$$\begin{aligned}
|\mathbb{T}_u^{10}| &= \left|\left\langle \kappa^{1/2}(u_h)\, l\, \kappa^{1/2}(u_h)\, l^{-1/2}(\varphi(u) - \varphi_h(u_h))\, l^{1/2}\, \boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n},\, l^{-1}\,\psi\right\rangle_{\Gamma_h}\right| \\
&\leq c\,\overline{\kappa}^{1/2}\, R\, h\, \|\kappa^{1/2}(u_h)\, l^{-1/2}(\varphi(u) - \varphi_h(u_h))\, l^{1/2}\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n}\|_{\Gamma_h}\|l^{-1}\,\psi\|_{\Gamma_h}. \\
&\leq c\,\overline{\kappa}^{1/2}\, R\, h\, \|\kappa^{1/2}(u_h)\, l^{-1/2}(\varphi(u) - \varphi_h(u_h))\|_{\Gamma_h}\|l^{1/2}\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n}\|_{L^\infty(\Gamma_h)}\|l^{-1}\,\psi\|_{\Gamma_h}.
\end{aligned}$$

From here, we will use the inequality $l(\boldsymbol{x}) \leq r_e h_e^{\perp}$ together with the estimate (B.3d) for $\|l^{-1}\,\psi\|_{\Gamma_h}$, we get

$$|\mathbb{T}_u^{10}| \leq c\,\tilde{c}\,\overline{\kappa}^{1/2}\,R^{3/2}\,h\,\|\kappa^{1/2}(u_h)\,l^{-1/2}\,(\varphi(u)-\varphi_h(u_h))\|_{\Gamma_h}\,\|(h^{\perp})^{1/2}\,\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n}\|_{L^{\infty}(\Gamma_h)}\|\Theta\|_{\Omega}.$$

Similar arguments can be used to derive the following analogous bound

$$|\mathbb{T}_u^{11}| \leq c\,\tilde{c}\,\overline{\kappa}^{1/2}\,R^{3/2}\,h\,\|\kappa^{1/2}(u_h)\,l^{-1/2}\,(\varphi(u)-\varphi_h(u_h))\|_{\Gamma_h}\,\|(h^{\perp})^{1/2}\,\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}\cdot\boldsymbol{n}\|_{L^{\infty}(\Gamma_h)}\|\Theta\|_{\Omega}.$$

Finally, letting $\Theta = \varepsilon^u$ in Ω_h and $\Theta = 0$ in Ω_h^c and using the estimates derived above for all the terms $\mathbb{T}_\star$ in the decomposition (3.24), one arrives at the desired estimate:

$$\|\varepsilon^u\|_{\Omega_h} \lesssim h^{1/2}\,|||(\boldsymbol{\varepsilon}^{\boldsymbol{q}},\,\varepsilon^u-\widehat{\varepsilon^u},\,\varphi-\varphi_h)|||_{u_h} + (h^{1/2}+L\,h)\Lambda_{\boldsymbol{q}} + (L+h^{1/2})\Lambda_u,$$

□

This concludes the analysis of the discretization for problems with nonlinear diffusivities of the form $\kappa = \kappa(u)$. The reminder of the article will be devoted to the analysis of cases where the nonlinearity appears as a dependence to the gradient of the unknown. This functional dependence will require a different reformulation of the problem.

4 Non-linearities of the Form $\kappa(\nabla u)$

4.1 Problem Statement

In some applications, the diffusivity coefficient depends on the gradient of the solution, rather than on the solution itself. This is indeed the case, for instance, in the plasma equilibrium problem, where the coefficient is the inverse of the magnetic permeability. In ferromagnetic materials, the magnetic permeability becomes a function of the total magnetic field and therefore the coefficient has the functional dependence $\kappa = \kappa(\nabla u)$. In cases like this, we will be interested in boundary value problems of the form

$$-\nabla\cdot(\kappa(\nabla u)\nabla u) = f(u) \qquad \text{in } \Omega, \tag{4.1a}$$

$$u = g \qquad \text{on } \Gamma := \partial\Omega. \tag{4.1b}$$

where, just like in the previous section, the source term f will be assumed to be Lipschitz-continuous in Ω, with Lipschitz constant $L_f > 0$. We will also maintain the assumption (3.1) on boundedness of the permeability. Note that, since we will be searching for solutions with $H^1(\Omega)$ regularity, the hypothesis (3.1) will guarantee that the permeability remains bounded as a function of ∇u. The Lipschitz-continuity assumptions (3.2), (3.3), and (3.4) will be replaced by their following vector counterparts

$$\begin{aligned}
\|\kappa^{-1}(\boldsymbol{\sigma}_1)-\kappa^{-1}(\boldsymbol{\sigma}_2)\|_{L^2(\Gamma)} &\leq L\|\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2\|_{\boldsymbol{L}^2(\Omega)} \quad \forall\,\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\boldsymbol{L}^2(\Omega),\\
\|\kappa^{1/2}(\boldsymbol{\sigma}_1)-\kappa^{1/2}(\boldsymbol{\sigma}_2)\|_{L^{\infty}(\Gamma)} &\leq \widehat{L}\|\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2\|_{\boldsymbol{L}^2(\Omega)} \quad \forall\,\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\boldsymbol{L}^2(\Omega),\\
\|\kappa(\boldsymbol{\sigma}_1)-\kappa(\boldsymbol{\sigma}_2)\|_{L^2(\Omega)} &\leq \widetilde{L}\|\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2\|_{\boldsymbol{L}^2(\Omega)} \quad \forall\,\boldsymbol{\sigma}_1,\boldsymbol{\sigma}_2\in\boldsymbol{L}^2(\Omega).
\end{aligned} \tag{4.2}$$

Following the spirit of reformulating the problem in a mixed form, the functional dependence $\kappa(\nabla u)$ will require us to introduce a new auxiliary variable. Therefore, we introduce $\boldsymbol{\sigma} := \nabla u$ and will express the the flux as $\boldsymbol{q} := -\kappa(\boldsymbol{\sigma})\boldsymbol{\sigma}$, thus introducing two additional unknowns to the problem. With these definitions, it is possible to write (4.1) as the equivalent system

$$\boldsymbol{\sigma} - \nabla u = 0 \qquad \text{in } \Omega,$$

$$\begin{aligned} \boldsymbol{q} + \kappa(\boldsymbol{\sigma})\,\boldsymbol{\sigma} &= 0 && \text{in } \Omega, \\ \nabla \cdot \boldsymbol{q} &= f(u) && \text{in } \Omega, \\ u &= g && \text{on } \partial\Omega. \end{aligned}$$

We shall analyze the discretization of this system when restricted to the subdomain in Ω_h. In view of this, our target formulation is

$$\boldsymbol{\sigma} - \nabla u = 0 \qquad \text{in } \Omega_h, \tag{4.3a}$$

$$\boldsymbol{q} + \kappa(\boldsymbol{\sigma})\,\boldsymbol{\sigma} = 0 \qquad \text{in } \Omega_h, \tag{4.3b}$$

$$\nabla \cdot \boldsymbol{q} = f(u) \qquad \text{in } \Omega_h, \tag{4.3c}$$

$$u = \varphi \qquad \text{on } \Gamma_h = \partial\Omega_h. \tag{4.3d}$$

where the boundary conditions have been transferred by means of

$$\varphi(\boldsymbol{\sigma}, \boldsymbol{x}) := g(\overline{\boldsymbol{x}}) + \int_0^{l(\boldsymbol{x})} \big(\kappa^{-1}(\boldsymbol{\sigma})\boldsymbol{q}\big)(\boldsymbol{x} + \boldsymbol{t}(\boldsymbol{x})s) \cdot \boldsymbol{t}(\boldsymbol{x}) ds.$$

The HDG scheme associated to (4.3) reads: Find $(\boldsymbol{q}_h, \boldsymbol{\sigma}_h, u_h, \widehat{u}_h) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times M_h$, such that

$$(\boldsymbol{\sigma}_h, \boldsymbol{v})_{\mathcal{T}_h} + (u_h, \nabla \cdot \boldsymbol{v})_{\mathcal{T}_h} - \langle \widehat{u}_h, \boldsymbol{v} \cdot \boldsymbol{n} \rangle_{\partial\mathcal{T}_h} = 0, \tag{4.4a}$$

$$(\boldsymbol{q}_h, \boldsymbol{s})_{\mathcal{T}_h} + (\kappa(\boldsymbol{\sigma}_h)\,\boldsymbol{\sigma}_h, \boldsymbol{s})_{\mathcal{T}_h} = 0, \tag{4.4b}$$

$$-(\boldsymbol{q}_h, \nabla w)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}, w \rangle_{\partial\mathcal{T}_h} = (f(u_h), w)_{\mathcal{T}_h}, \tag{4.4c}$$

$$\langle \widehat{u}_h, \mu \rangle_{\Gamma_h} = \langle \varphi_h(\boldsymbol{\sigma}_h), \mu \rangle_{\Gamma_h}, \tag{4.4d}$$

$$\langle \widehat{\boldsymbol{q}}_h \cdot \boldsymbol{n}, \mu \rangle_{\partial\mathcal{T}_h \setminus \Gamma_h} = 0, \tag{4.4e}$$

for all $(\boldsymbol{v}, \boldsymbol{s}, w, \mu) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times M_h$. Here, the spaces $\boldsymbol{V}_h$, W_h, and M_h have been defined in (2.7), the restriction to the mesh skeleton of the numerical flux has been defined as

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

and the approximate boundary condition given by

$$\varphi_h(\boldsymbol{\sigma}_h, \boldsymbol{x}) := g(\overline{\boldsymbol{x}}) + \int_0^{l(\boldsymbol{x})} \big(\kappa^{-1}(\boldsymbol{\sigma}_h)\boldsymbol{q}_h\big)(\boldsymbol{x} + \boldsymbol{t}(\boldsymbol{x})s) \cdot \boldsymbol{t}(\boldsymbol{x}) ds. \tag{4.4f}$$

As before, the maximum value of the positive stabilization function τ will be denoted by $\overline{\tau}$.

In this section we will analyze an HDG scheme for problems of this form. We will first have to reformulate the problem in terms of a mixed system with one additional unknown when compared to the case analyzed in the previous section. Many of the arguments required for the analysis will be similar to those developed in the previous section, and the analysis technique is similar as well. We will therefore omit many of the technical details and indicate the main steps in the analysis, focusing on those that are different from the previous section.

4.2 Well-Posedness

The proof that the system (3.6) is well-posed will rely on a fixed point argument. As in the previous section, we define the operator $\mathcal{J}_2 : \boldsymbol{V}_h \times W_h \to \boldsymbol{V}_h \times W_h$ that maps a pair of

functions $(\boldsymbol{\eta}, \zeta)$ to the first and third component of the solution $(\boldsymbol{q}, \boldsymbol{\sigma}, u, \widehat{u}) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times M_h$ to the HDG system (4.4) where the source has been evaluated at $(\boldsymbol{\eta}, \zeta)$. Namely

$$(\boldsymbol{\sigma}, \boldsymbol{v})_{\mathcal{T}_h} + (u, \nabla \cdot \boldsymbol{v})_{\mathcal{T}_h} - \langle \widehat{u}, \boldsymbol{v} \cdot \boldsymbol{n} \rangle_{\partial \mathcal{T}_h} = 0, \tag{4.5a}$$

$$(\boldsymbol{q}, \boldsymbol{s})_{\mathcal{T}_h} + (\kappa(\boldsymbol{\eta})\, \boldsymbol{\sigma}, \boldsymbol{s})_{\mathcal{T}_h} = 0, \tag{4.5b}$$

$$-(\boldsymbol{q}, \nabla w)_{\mathcal{T}_h} + \langle \widehat{\boldsymbol{q}} \cdot \boldsymbol{n}, w \rangle_{\partial \mathcal{T}_h} = (f(\zeta), w)_{\mathcal{T}_h}, \tag{4.5c}$$

$$\langle \widehat{u}, \mu \rangle_{\Gamma_h} = \langle \varphi(\boldsymbol{\eta}), \mu \rangle_{\Gamma_h}, \tag{4.5d}$$

$$\langle \widehat{\boldsymbol{q}} \cdot \boldsymbol{n}, \mu \rangle_{\partial \mathcal{T}_h \backslash \Gamma_h} = 0, \tag{4.5e}$$

for all $(\boldsymbol{v}, \boldsymbol{s}, w, \mu) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times M_h$. Just as before, $\varphi(\zeta)$ accounts for the transferred boundary conditions and, since the discrete linearized system above is uniquely solvable [5], $\mathcal{J}_2$ is well defined.

Given a function $\boldsymbol{\eta} \in \boldsymbol{V}_h$, we define the following norm over the product space $\boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times M_h$

$$|||(\boldsymbol{s}, \boldsymbol{v}, w, \lambda)|||_{\boldsymbol{\eta}} := \left(\|\boldsymbol{s}\|^2_{\Omega_h} + \|\kappa^{1/2}(\boldsymbol{\eta})\boldsymbol{v}\|^2_{\Omega_h} + \|\tau^{1/2}\, w\|^2_{\partial \mathcal{T}_h} + \|\kappa^{1/2}(\boldsymbol{\eta})\, l^{-1/2}\, \lambda(\boldsymbol{\eta})\|^2_{\Gamma_h} \right)^{1/2}. \tag{4.6}$$

The general road map for the proof is as follows. Lemmas 7 and 8 below, will allow us to control $(\boldsymbol{q}, \boldsymbol{\sigma}, u - \widehat{u}, \varphi)$ by the linearized source term $f(\zeta)$ and the boundary condition *at the physical boundary*, $\overline{g}$. An application of these two results will then allow us—modulo some technical assumptions involving the bound of the diffusivity and the distance between the physical and computational domains—to use the Lipschitz continuity of f and κ to prove that the mapping is indeed a contraction. This will be done in Theorem 3, from which the well posedness of the HDG system (4.4) will follow as a simple corollary.

Lemma 7 *If Assumptions* (2.4) *hold, then*

$$|||(\boldsymbol{q}, \boldsymbol{\sigma}, u - \widehat{u}, \varphi)|||^2_{\boldsymbol{\eta}} \leq \max\{1, \overline{\kappa}\} \Big(4 \|f(\zeta)\|_{\Omega_h} \|u\|_{\Omega_h} + 6 \|\kappa^{1/2}(\boldsymbol{\eta})\, l^{-1/2}\, \overline{g}\|^2_{\Gamma_h} \Big). \tag{4.7}$$

Proof Note that if we let $\boldsymbol{s} = \boldsymbol{q}$ in (4.5b), we have

$$\|\boldsymbol{q}\|^2_{\Omega_h} \leq \overline{\kappa}\, \|\kappa^{1/2}(\boldsymbol{\eta})\, \boldsymbol{\sigma}\|^2_{\Omega_h}. \tag{4.8}$$

Then, following the process outlined in the proof of Lemma 3, we go back to (4.5) and choose the test functions as $\boldsymbol{v} = -\boldsymbol{q}$, $\boldsymbol{s} = \boldsymbol{\sigma}$, $w = u$, and

$$\mu = \begin{cases} -\widehat{u}, & \text{on } \partial \mathcal{T}_h \backslash \Gamma_h \\ -\widehat{\boldsymbol{q}} \cdot \boldsymbol{n}, & \text{on } \Gamma_h. \end{cases}$$

This leads to the equality

$$\begin{aligned} &\|\kappa^{1/2}(\boldsymbol{\eta})\boldsymbol{\sigma}\|^2_{\Omega_h} + \|\tau^{1/2}(u - \widehat{u})\|^2_{\partial \mathcal{T}_h} \\ &\quad = (f(\zeta), u)_{\mathcal{T}_h} - \langle \varphi(\boldsymbol{\eta}), \kappa(\boldsymbol{\eta}) l^{-1} \varphi(\boldsymbol{\eta}) \rangle_{\Gamma_h} + \langle \varphi(\boldsymbol{\eta}), \kappa(\boldsymbol{\sigma}) l^{-1} \overline{g} \rangle_{\Gamma_h} \\ &\qquad + \langle \varphi(\boldsymbol{\eta}), \overline{g} \rangle_{\Gamma_h} - \langle \varphi(\boldsymbol{\eta}), \tau(u - \widehat{u}) \rangle_{\Gamma_h}. \end{aligned}$$

The terms on the right hand side can be estimated by an application of Lemma 2, yielding

$$\begin{aligned} &\|\kappa^{1/2}(\boldsymbol{\eta})\boldsymbol{\sigma}\|^2_{\Omega_h} + \|\tau^{1/2}\,(u - \widehat{u})\|^2_{\partial \mathcal{T}_h} + \|\kappa^{1/2}(\eta)\, l^{-1/2}\, \varphi(\boldsymbol{\eta})\|^2_{\Gamma_h} \\ &\qquad \leq 2 \|f(\zeta)\|_{\Omega_h} \|u\|_{\Omega_h} + 3 \|\kappa^{1/2}(\boldsymbol{\eta})\, l^{-1/2}\, \overline{g}\|^2_{\Gamma_h}. \end{aligned}$$

Combining this estimate with (4.8), we obtain

$$|||(\boldsymbol{q},\boldsymbol{\sigma},u-\widehat{u},\varphi)|||_{\boldsymbol{\eta}}^2 \leq \max\{1,\overline{\kappa}\}\left(4\|f(\zeta)\|_{\Omega_h}\|u\|_{\Omega_h}+6\|\kappa^{1/2}(\boldsymbol{\eta})\,l^{-1/2}\,\overline{g}\|_{\Gamma_h}^2\right),$$

which concludes the proof. □

It only remains now to estimate the norm of u in terms of the sources and the boundary conditions. This will be done in the next lemma.

Lemma 8 *Suppose that Assumptions (2.4) and (B.2) are satisfied. Then, there exists $\widehat{c}>0$, independent of h such that*

$$\|u\|_{\Omega_h} \leq 4\,\max\{\widehat{c}^2h,\,1\}\,\|f(\zeta)\|_{\Omega_h}+2\,(\sqrt{3}\,\widehat{c}+\overline{\kappa}^{1/2}\,R^{1/2})\,h^{1/2}\,\|\kappa^{1/2}(\boldsymbol{\eta})\,l^{-1/2}\,\overline{g}\|_{\Gamma_h}. \quad (4.9)$$

Proof The proof of this result is follows, with small variations, the same process as that of Lemma 4. By using $\boldsymbol{\eta}$ instead of ζ in the dual problem given in (B.1), and splitting the duality product as

$$(u,\Theta)_{\mathcal{T}_h}=\mathbb{T}_{\boldsymbol{\sigma}}+\mathbb{T}_u+\mathbb{T}_f,$$

where

$$\mathbb{T}_{\boldsymbol{\sigma}}:=-(\boldsymbol{\sigma},\boldsymbol{\Pi}_V\boldsymbol{\phi}-\boldsymbol{\phi})_{\mathcal{T}_h},\quad \mathbb{T}_u:=\langle\widehat{u},P_M(\boldsymbol{\phi}\cdot\boldsymbol{n})\rangle_{\Gamma_h}-\langle\widehat{\boldsymbol{q}}\cdot\boldsymbol{n},\Pi_W\psi\rangle_{\Gamma_h}\quad\text{and}$$
$$\mathbb{T}_f:=(f(\zeta),\Pi_W\psi)_{\mathcal{T}_h}.$$

The terms $\mathbb{T}_{\boldsymbol{\sigma}}$ and $\mathbb{T}_f$ are bounded as

$$|\mathbb{T}_{\boldsymbol{\sigma}}|\lesssim\underline{\kappa}^{-1/2}\,h\|\kappa^{-1/2}(\boldsymbol{\eta})\,\boldsymbol{\sigma}\|_{\Omega_h}\,\|\Theta\|_{\Omega},\quad\text{and}\quad|\mathbb{T}_f|\lesssim\|f(\zeta)\|_{\Omega_h}\,\|\Theta\|_{\Omega},$$

and, we rewrite $\mathbb{T}_u$ as $\mathbb{T}_u=\sum_{i=1}^5\mathbb{T}_u^i$, with

$$\mathbb{T}_u^1:=-\langle\kappa(\boldsymbol{\eta})l^{-1}\varphi(\boldsymbol{\eta}),\psi+l\partial_n\psi\rangle_{\Gamma_h},\quad\mathbb{T}_u^4:=-\langle\tau(u-\widehat{u}),P_M\psi\rangle_{\Gamma_h},$$
$$\mathbb{T}_u^2:=\langle\kappa(\boldsymbol{\eta})\varphi(\boldsymbol{\eta}),(P_M-I_d)\partial_n\psi\rangle_{\Gamma_h},\quad\mathbb{T}_u^5:=\langle\kappa(\boldsymbol{\eta})\,l^{-1}\,\overline{g},\psi\rangle_{\Gamma_h},$$
$$\mathbb{T}_u^3:=\langle\delta_{\boldsymbol{\sigma}},\psi\rangle_{\Gamma_h}.$$

These terms can be bounded using arguments analogous to those in Lemma 4, yielding the desired estimate (4.9). □

The results in the two preceding lemmas can be combined to estimate $(\boldsymbol{q},\boldsymbol{\sigma},u-\widehat{u},\varphi)$ in terms of the source $f(\zeta)$ and the boundary data $\overline{g}$. This follows readily from an application of Lemma 8 to (4.7), yielding

$$\begin{aligned}|||(\boldsymbol{q},\boldsymbol{\sigma},u-\widehat{u},\varphi)|||_{\eta}^2\leq\;&\left(16\,\max\{1,\overline{\kappa}\}^2+8\,\max\{\widehat{c}^2h,1\}^2\right)\|f(\zeta)\|_{\Omega_h}^2\\&+\left(6\max\{1,\overline{\kappa}\}+2\,(\sqrt{3}\,\widehat{c}+\overline{\kappa}^{1/2}\,R^{1/2})^2\,h\right)\|\kappa^{1/2}(\boldsymbol{\eta})\,l^{-1/2}\,\overline{g}\|_{\Gamma_h}^2.\end{aligned} \quad (4.10)$$

In turn, (4.9) implies that

$$\|u\|_{\Omega_h}^2\leq 32\,\max\{\widehat{c}^2h,1\}^2\,\|f(\zeta)\|_{\Omega_h}^2+8\,(\sqrt{3}\,\widehat{c}+\overline{\kappa}^{1/2}\,R^{1/2})^2\,h\,\|\kappa^{1/2}(\boldsymbol{\eta})\,l^{-1/2}\,\overline{g}\|_{\Gamma_h}^2. \quad (4.11)$$

These two estimates will be used to prove the contractive properties of $\mathcal{J}_2$ as we will now show.

Theorem 3 *Suppose that the dual regularity* (B.2) *and the Assumptions* (2.4) *hold and suppose also that*

$$\left(16\,\max\{1,\overline{\kappa}\}^2+40\,\max\{\widehat{c}^2h,1\}^2\right)L_f^2<\frac{1}{4} \tag{4.12}$$

and

$$\left(6\max\{1,\overline{\kappa}\}+10\,(\sqrt{3}\,\widehat{c}+\overline{\kappa}^{1/2}\,R^{1/2})^2\,h\right)\widehat{L}^2\|l^{-1/2}\,\overline{g}\|_{\Gamma_h}^2<\frac{1}{4}. \tag{4.13}$$

are satisfied. Then $\mathcal{J}_2$ is a contraction operator.

Proof Let $(\boldsymbol{\eta}_i,\zeta_i)\in \boldsymbol{V}_h\times W_h$ and define $(\boldsymbol{\sigma}_i,u_i):=\mathcal{J}_2\left((\boldsymbol{\eta}_i,\zeta_i)\right)\in \boldsymbol{V}_h\times W_h$ for $i\in\{1,2\}$. Then,

$$\|\mathcal{J}_2(\boldsymbol{\eta}_1,\zeta_1)-\mathcal{J}_2(\boldsymbol{\eta}_2,\zeta_2)\|_{\Omega_h}=\|(\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2,u_1-u_2)\|_{\Omega_h}=\left(\|\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2\|_{\Omega_h}^2+\|u_1-u_2\|_{\Omega_h}^2\right)^{1/2}.$$

By applying the inequalities (4.10) and (4.11) respectively to $(\boldsymbol{\sigma}_1-\boldsymbol{\sigma}_2)$ and (u_1-u_2), we obtain

$$\|\mathcal{J}_2(\boldsymbol{\eta}_1,\zeta_1)-\mathcal{J}_2(\boldsymbol{\eta}_2,\zeta_2)\|_{\Omega_h}\le\Bigg(\left(16\,\max\{1,\overline{\kappa}\}^2+40\,\max\{\widehat{c}^2h,1\}^2\right)\|f(\zeta_1)-f(\zeta_2)\|_{\Omega_h}^2$$
$$+\left(6\max\{1,\overline{\kappa}\}+10\,(\sqrt{3}\,\widehat{c}+\overline{\kappa}^{1/2}\,R^{1/2})^2\,h\right)\|\kappa^{1/2}(\boldsymbol{\eta}_1)-\kappa^{1/2}(\boldsymbol{\eta}_2)\|_{L^\infty(\Gamma_h)}^2\|l^{-1/2}\,\overline{g}\|_{\Gamma_h}^2\Bigg)^{1/2}.$$

Then, using the Lipschitz continuity of f and $\kappa^{1/2}$, we get

$$\|\mathcal{J}_2(\boldsymbol{\eta}_1,\zeta_1)-\mathcal{J}_2(\boldsymbol{\eta}_2,\zeta_2)\|_{\Omega_h}\le\Bigg(\left(16\,\max\{1,\overline{\kappa}\}^2+40\,\max\{\widehat{c}^2h,1\}^2\right)L_f^2\|\zeta_1-\zeta_2\|_{\Omega_h}^2$$
$$+\left(6\max\{1,\overline{\kappa}\}+10\,(\sqrt{3}\,\widehat{c}+\overline{\kappa}^{1/2}\,R^{1/2})^2\,h\right)\widehat{L}^2\|\boldsymbol{\eta}_1-\boldsymbol{\eta}_2\|_{L^\infty(\Omega_h)}^2\|l^{-1/2}\,\overline{g}\|_{\Gamma_h}^2\Bigg)^{1/2}.$$

The proof is concluded, by applying the assumptions (4.12) and (4.13) to the right hand side of the preceding inequality. □

As a result we can then conclude this section with with the following

Corollary 2 *If the hypotheses of Theorem* 3 *are satisfied, the HDG system* (4.5) *is well posed.*

Having established the well posedness of the discrete system (4.5), we will concentrate our efforts in the next section on establishing the convergence properties of the HDG scheme.

4.3 A Priori Error Analysis

The study of the convergence properties and rates of our discretization follows similar steps as the ones laid out in Sect. 3.3, adapted for the extended system that arises from the introduction of the additional auxiliary variable $\boldsymbol{\sigma}=\nabla u$. To avoid unnecessary repetition of arguments, we will focus on the differences between these two cases and will omit most of the details that can be easily inferred from Sect. 3.3.

As before, we decompose the error with the aid of the HDG projection as:

$$\boldsymbol{q}-\boldsymbol{q}_h=\boldsymbol{\varepsilon}^{\boldsymbol{q}}+\boldsymbol{I}^{\boldsymbol{q}},\quad \boldsymbol{\sigma}-\boldsymbol{\sigma}_h=\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}+\boldsymbol{I}^{\boldsymbol{\sigma}},\quad \text{and}\quad u-u_h=\varepsilon^u+I^u,$$

where, similar to Sect. 3.3, we have defined

$$\boldsymbol{\varepsilon}^{\boldsymbol{q}}:=\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}-\boldsymbol{q}_h,\quad \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}:=\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{\sigma}-\boldsymbol{\sigma}_h,\quad \varepsilon^u:=\Pi_W u-u_h\qquad \textit{(Projection of the error)},$$

$$\boldsymbol{I}^{\boldsymbol{q}} := \boldsymbol{q} - \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}, \quad \boldsymbol{I}^{\boldsymbol{\sigma}} := \boldsymbol{\sigma} - \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{\sigma}, \quad I^{u} := u - \Pi_{W}u \quad \textit{(Error of the projection)}.$$

In addition, using the L^2 projection into M_h we have $\varepsilon^{\widehat{u}} := P_M u - \widehat{u}_h$. The vector of error projections $(\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \varepsilon^{u}, \varepsilon^{\widehat{u}})$ belongs to $\boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times M_h$ and satisfies the error equations

$$(\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \boldsymbol{v})_{\mathcal{T}_h} + (\varepsilon^{u}, \nabla \cdot \boldsymbol{v})_{\mathcal{T}_h} - \langle \varepsilon^{\widehat{u}}, \boldsymbol{v} \cdot \boldsymbol{n} \rangle_{\partial\mathcal{T}_h} = -(\boldsymbol{I}^{\boldsymbol{\sigma}}, \boldsymbol{v})_{\mathcal{T}_h} - (\boldsymbol{I}^{u}, \nabla \cdot \boldsymbol{v})_{\mathcal{T}_h} \tag{4.14a}$$

$$(\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \boldsymbol{s})_{\mathcal{T}_h} + \big(\kappa(\boldsymbol{\sigma}_h)\, \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \boldsymbol{s}\big)_{\mathcal{T}_h} = -(\boldsymbol{I}^{\boldsymbol{q}}, \boldsymbol{s})_{\mathcal{T}_h} - \big(\kappa(\boldsymbol{\sigma})\, \boldsymbol{I}^{\boldsymbol{\sigma}}, \boldsymbol{s}\big)_{\mathcal{T}_h} \tag{4.14b}$$

$$- ((\kappa(\boldsymbol{\sigma}) - \kappa(\boldsymbol{\sigma}_h))\, \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{\sigma}, s)_{\mathcal{T}_h}\,, \tag{4.14c}$$

$$- (\boldsymbol{\varepsilon}^{\boldsymbol{q}}, \nabla w)_{\mathcal{T}_h} + \langle \boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} \cdot \boldsymbol{n}, w \rangle_{\partial\mathcal{T}_h} = (f(u) - f(u_h), w)_{\mathcal{T}_h}, \tag{4.14d}$$

$$\langle \varepsilon^{\widehat{u}}, \mu \rangle_{\Gamma_h} = \langle \varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h), \mu \rangle_{\Gamma_h}, \tag{4.14e}$$

$$\langle \boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} \cdot \boldsymbol{n}, \mu \rangle_{\partial\mathcal{T}_h \setminus \Gamma_h} = 0, \tag{4.14f}$$

for all $(\boldsymbol{v}, \boldsymbol{s}, w, \mu) \in \boldsymbol{V}_h \times \boldsymbol{V}_h \times W_h \times M_h$. Here, as before, on $\partial\mathcal{T}_h$ we have $\boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}} \cdot \boldsymbol{n} = \boldsymbol{\varepsilon}^{\boldsymbol{q}} \cdot \boldsymbol{n} + \tau(\varepsilon^{u} - \varepsilon^{\widehat{u}})$.

Following the same arguments of Lemma 5 is possible to estimate the magnitude of $(\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^{u} - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)$, as measured by the norm $|||\cdot|||_{\boldsymbol{\sigma}}$ defined in (4.6). Choosing the vector of approximation errors both as test and trial in the error equations we obtain

$$\begin{aligned} \|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\|^2_{\Omega_h} + \|\tau^{1/2}(\varepsilon^{u} - \varepsilon^{\widehat{u}})\|^2_{\partial\mathcal{T}_h} + \|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, l^{-1/2}\, (\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h))\|^2_{\Gamma_h} \\ \leq |(\boldsymbol{I}^{\boldsymbol{q}}, \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}})_{\mathcal{T}_h}| + |(\boldsymbol{I}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}})_{\mathcal{T}_h}| + |(\kappa(\boldsymbol{\sigma})\boldsymbol{I}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}})_{\mathcal{T}_h}| \\ + |((\kappa(\boldsymbol{\sigma}) - \kappa(\boldsymbol{\sigma}_h))\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{\sigma}, \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}})_{\mathcal{T}_h}| + |\mathbb{T}^{f}| + |\mathbb{T}_{\varphi}|. \end{aligned} \tag{4.15}$$

The final two terms are defined as $\mathbb{T}^{f} := (f(u) - f(u_h), \varepsilon^{u})_{\mathcal{T}_h}$ and $\mathbb{T}_{\varphi} := \sum_{i=1}^{8} |\mathrm{T}^{i}_{\varphi}|$, where

$$\begin{aligned} \mathbb{T}^{1}_{\varphi} &:= \langle \varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h), \delta_{\boldsymbol{\varepsilon}^{\boldsymbol{q}}} \rangle_{\Gamma_h} \\ \mathbb{T}^{2}_{\varphi} &:= -\langle \varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h), \tau(\varepsilon^{u} - \varepsilon^{\widehat{u}}) \rangle_{\Gamma_h} \\ \mathbb{T}^{3}_{\varphi} &:= \langle \varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h), \boldsymbol{I}^{\boldsymbol{q}} \cdot \boldsymbol{n} \rangle_{\Gamma_h} \\ \mathbb{T}^{4}_{\varphi} &:= \langle \varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h), \delta_{\boldsymbol{I}^{\boldsymbol{q}}} \rangle_{\Gamma_h} \\ \mathbb{T}^{5}_{\varphi} &:= \langle \varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h), \kappa(\boldsymbol{\sigma}_h)\, (\kappa^{-1}(\boldsymbol{\sigma}) - \kappa^{-1}(\boldsymbol{\sigma}_h))\, \delta_{\boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q}} \rangle_{\Gamma_h} \\ \mathbb{T}^{6}_{\varphi} &:= \langle \varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h), \kappa(\boldsymbol{\sigma}_h)\, (\kappa^{-1}(\boldsymbol{\sigma}) - \kappa^{-1}(\boldsymbol{\sigma}_h))\, \delta_{\boldsymbol{I}^{\boldsymbol{q}}} \rangle_{\Gamma_h} \\ \mathbb{T}^{7}_{\varphi} &:= \langle \varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h), \kappa(\boldsymbol{\sigma}_h)\, (\kappa^{-1}(\boldsymbol{\sigma}) - \kappa^{-1}(\boldsymbol{\sigma}_h))\, \boldsymbol{I}^{\boldsymbol{q}} \cdot \boldsymbol{n} \rangle_{\Gamma_h} \\ \mathbb{T}^{8}_{\varphi} &:= \langle \varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h), \kappa(\boldsymbol{\sigma}_h)\, (\kappa^{-1}(\boldsymbol{\sigma}) - \kappa^{-1}(\boldsymbol{\sigma}_h))\, \boldsymbol{\Pi}_{\boldsymbol{V}}\boldsymbol{q} \cdot \boldsymbol{n} \rangle_{\Gamma_h}. \end{aligned}$$

By a combined use of arguments similar to those appearing in Lemma 5, it is possible to deduce

$$|\mathbb{T}^{1}_{\varphi}| \leq \frac{1}{2\delta_1} \|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, l^{-1/2}(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)\|^2_{\Gamma_h} + \frac{\delta_1}{6}\, \overline{\kappa}^{-1}\, \|\boldsymbol{\varepsilon}^{\boldsymbol{q}}\|^2_{\Omega_h},$$

$$|\mathbb{T}^{2}_{\varphi}| \leq \frac{1}{2\delta_2} \|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, l^{-1/2}(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)\|^2_{\Gamma_h} + \frac{\delta_2}{6}\, \|\tau^{1/2}(\varepsilon^{u} - \varepsilon^{\widehat{u}})\|^2_{\partial\mathcal{T}_h},$$

$$|\mathbb{T}^{3}_{\varphi}| \leq \frac{1}{2\delta_3} \|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, l^{-1/2}(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)\|^2_{\Gamma_h} + \frac{\delta_3}{2}\, R\, \underline{\kappa}^{-1}\, \|(h^{\perp})^{1/2} \boldsymbol{I}^{\boldsymbol{q}} \cdot \boldsymbol{n})\|^2_{\Gamma_h},$$

$$|\mathbb{T}^{4}_{\varphi}| \leq \frac{1}{2\delta_3} \|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, l^{-1/2}(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)\|^2_{\Gamma_h} + \frac{\delta_3}{6}\, \underline{\kappa}^{-1}\, R^2\, \|h^{\perp}\partial_n(\boldsymbol{I}^{\boldsymbol{q}} \cdot \boldsymbol{n})\|^2_{\Omega_h^c},$$

$$
\begin{aligned}
|\mathbb{T}_\varphi^5| &\le \frac{1}{2\delta_3}\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, l^{-1/2}(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)\|_{\Gamma_h}^2 \\
&\quad + \frac{\delta_3}{3}\,\overline{\kappa}\,R^2 \sup_{\boldsymbol{x}\in e\subseteq\Gamma_h} \|h^\perp \partial_n(\Pi_{\boldsymbol{V}}\boldsymbol{q}\cdot\boldsymbol{n})\|_{l(\boldsymbol{x})}^2\, L^2\, \left(\|\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2 + \|\boldsymbol{I}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2\right), \\
|\mathbb{T}_\varphi^6| &\le \frac{1}{2\delta_3}\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, l^{-1/2}(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)\|_{\Gamma_h}^2 + \frac{2\delta_3}{3}\,\overline{\kappa}\,\underline{\kappa}^{-2}\,R^2\,\|h^\perp\partial_n(\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n})\|_{\Omega_h^c}^2, \\
|\mathbb{T}_\varphi^7| &\le \frac{1}{2\delta_3}\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, l^{-1/2}(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)\|_{\Gamma_h}^2 + 2\,\delta_3\,R\,\overline{\kappa}\,\underline{\kappa}^{-2}\,\|(h^\perp)^{1/2}\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n})\|_{\Gamma_h}^2, \\
|\mathbb{T}_\varphi^8| &\le \frac{1}{2\delta_3}\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\, l^{-1/2}(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)\|_{\Gamma_h}^2 \\
&\quad + \delta_3\,\overline{\kappa}\,R\,\|(h^\perp)^{1/2}\,\Pi_{\boldsymbol{V}}\boldsymbol{q}\cdot\boldsymbol{n}\|_\infty^2\, L^2\, \left(\|\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2 + \|\boldsymbol{I}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2\right),
\end{aligned}
\tag{4.16}
$$

and

$$
\begin{aligned}
|(\boldsymbol{I}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}})_{\mathcal{T}_h}| &\le \frac{1}{2\,\delta_4}\|\boldsymbol{\varepsilon}^{\boldsymbol{q}}\|_{\Omega_h}^2 + \frac{\delta_4}{2}\,\|\boldsymbol{I}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2, \\
|(\boldsymbol{I}^{\boldsymbol{q}}, \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}})_{\mathcal{T}_h}| &\le \frac{1}{2\,\delta_5}\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2 + \frac{\delta_5}{2}\,\underline{\kappa}^{-1}\,\|\boldsymbol{I}^{\boldsymbol{q}}\|_{\Omega_h}^2, \\
|(\kappa(\boldsymbol{\sigma})\boldsymbol{I}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}})_{\mathcal{T}_h}| &\le \frac{1}{2\,\delta_5}\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2 + \frac{\delta_5}{2}\,\underline{\kappa}^{-1}\,\overline{\kappa}^2\|\boldsymbol{I}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2, \\
|(\kappa^{\backprime}(\boldsymbol{\sigma}) - \kappa(\boldsymbol{\sigma}_h))\,\Pi_{\boldsymbol{V}}\boldsymbol{\sigma}, \boldsymbol{\varepsilon}^{\boldsymbol{\sigma}})_{\mathcal{T}_h}| &\le \frac{1}{2\,\delta_5}\,\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2 \\
&\quad + \delta_5\,\underline{\kappa}^{-1}\,\|\Pi_{\boldsymbol{V}}\boldsymbol{q}\|_\infty^2\, L^2\, \left(\|\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2 + \|\boldsymbol{I}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2\right). \\
|\mathbb{T}^f| &\le L_f\,(\|\varepsilon^u\|_{\Omega_h} + \|I^u\|_{\Omega_h})\,\|\varepsilon^u\|_{\Omega_h}.
\end{aligned}
\tag{4.17}
$$

Then, taking the test function $s = \boldsymbol{\varepsilon}^{\boldsymbol{q}}$ in the second equation of (4.14), we get

$$
\begin{aligned}
\|\boldsymbol{\varepsilon}^{\boldsymbol{q}}\|_{\Omega_h}^2 &\le \left(4\,\overline{\kappa} + 8\,\underline{\kappa}^{-1}\,L^2\,\|\Pi_{\boldsymbol{V}}\boldsymbol{q}\|_\infty^2\right)\,\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\,\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2 + 4\,\|\boldsymbol{I}^{\boldsymbol{q}}\|_{\Omega_h}^2 \\
&\quad + 4\,\max\{\overline{\kappa}, 2\,L^2\,\|\Pi_{\boldsymbol{V}}\,\boldsymbol{\sigma}\|_\infty^2\}\,\|\boldsymbol{I}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2.
\end{aligned}
\tag{4.18}
$$

If the Lipschitz constants L and L_f are sufficiently small, a direct application of (4.18) in the first equations of (4.16) and (4.17), together with the choices $\delta_1 = 1$, $\delta_2 = 3$, $\delta_3 = 12$, $\delta_4 = 24/\overline{\kappa}$, $\delta_5 = 18$ yield the following estimate for the right hand side of (4.15)

$$
\begin{aligned}
\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\,\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\|_{\Omega_h}^2 + \|\tau^{1/2}(\varepsilon^u - \varepsilon^{\widehat{u}})\|_{\partial\mathcal{T}_h}^2 + \|\kappa^{1/2}(\boldsymbol{\sigma}_h)\,l^{-1/2}\,(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h))\|_{\Gamma_h}^2 \\
\lesssim\ \Lambda_{\boldsymbol{q}}^2 + \Lambda_{\boldsymbol{\sigma}}^2 + L_f\,(\|\varepsilon^u\|_{\Omega_h} + \|I^u\|_{\Omega_h})\,\|\varepsilon^u\|_{\Omega_h}.
\end{aligned}
\tag{4.19}
$$

In the expression above, the term $\Lambda_{\boldsymbol{\sigma}}$ has been defined according to (3.17a). By combining (4.18) and (4.19) we get

$$
|||(\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||_{\boldsymbol{\sigma}_h}^2 \lesssim \Lambda_{\boldsymbol{q}}^2 + \Lambda_{\boldsymbol{\sigma}}^2 + L_f\,(\|\varepsilon^u\|_{\Omega_h} + \|I^u\|_{\Omega_h})\,\|\varepsilon^u\|_{\Omega_h}. \tag{4.20}
$$

The following result allows us to estimate the term ε^u in (4.20) by means of a duality argument. The proof technique is analogous to the one used for Lemma 6.

Lemma 9 *Given the regularity condition (B.2), assume that the Lipschitz constant is such that L_f is small enough, and consider the discrete spaces to be of polynomial degree $k \ge 1$.*

Then,

$$\|\varepsilon^u\|^2_{\Omega_h} \lesssim 3\,h|||(\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||^2_{\boldsymbol{\sigma}_h} + 6\Big((h + L^2\,h^2)\Lambda^2_{\boldsymbol{q}} + (L^2 + h)\,\Lambda^2_u\Big). \tag{4.21}$$

Proof Consider the pair of functions function ϕ and ψ satisfying the dual problem (B.1). We will use them to define the following terms

$$\begin{aligned}
\mathbb{T}_{\boldsymbol{\sigma}} &:= -(\boldsymbol{\sigma} - \boldsymbol{\sigma}_h, \boldsymbol{\Pi_V}\boldsymbol{\phi} - \boldsymbol{\phi})_{\mathcal{T}_h} + ((\kappa(\boldsymbol{\sigma}_h) - \kappa(\boldsymbol{\sigma}))(\boldsymbol{I}^{\boldsymbol{\sigma}} + \boldsymbol{\Pi_V}\boldsymbol{\sigma}), \nabla\psi)_{\mathcal{T}_h},\\
\mathbb{T}_{\boldsymbol{q}} &:= -(\boldsymbol{I}^{\boldsymbol{q}}, \nabla\psi)_{\mathcal{T}_h},\\
\mathbb{T}_f &:= (f(u) - f(u_h), \Pi_W\psi)_{\mathcal{T}_h},\\
\mathbb{T}_u &:= \langle\varepsilon^{\widehat{u}}, P_M(\boldsymbol{\phi}\cdot\boldsymbol{n})\rangle_{\Gamma_h} - \langle\boldsymbol{\varepsilon}^{\widehat{\boldsymbol{q}}}\cdot\boldsymbol{n}, \Pi_W\psi\rangle_{\Gamma_h}.
\end{aligned}$$

With all the definitions above and the equations in (B.1), it is possible to decompose the inner product between ε^u and the function Θ appearing as the source term of the dual problem in the form

$$(\varepsilon^u, \Theta)_{\mathcal{T}_h} = \mathbb{T}_{\boldsymbol{\sigma}} + \mathbb{T}_{\boldsymbol{q}} + \mathbb{T}_f + \mathbb{T}_u, \tag{4.22}$$

Following arguments similar to the ones applied in Lemma 8, it is possible to bound each of the terms in the decomposition as

$$\begin{aligned}
|\mathbb{T}_{\boldsymbol{\sigma}}| &\lesssim h^{\min\{1,k\}}\,\|\kappa^{1/2}(\boldsymbol{\sigma}_h)(\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}} + \boldsymbol{I}^{\boldsymbol{\sigma}})\|_{\Omega_h}\,\|\Theta\|_\Omega + L\,(\|\kappa^{1/2}(\boldsymbol{\sigma}_h)\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}\| + \|\boldsymbol{I}^{\boldsymbol{\sigma}}\|)\,\|\Theta\|_\Omega\\
|\mathbb{T}_{\boldsymbol{q}}| &\lesssim h^{\min\{1,k\}}\|\boldsymbol{I}\|_{\Omega_h}\,\|\Theta\|_\Omega,\\
|\mathbb{T}_f| &\lesssim L_f\,\left(\|\varepsilon^u\|_{\Omega_h} + \|I^u\|_{\Omega_h}\right)\,\|\Theta\|_\Omega
\end{aligned}$$

The bound for the final term in (4.22) requires decomposing it in the form $\mathbb{T}_u := \sum_{i=1}^{10}\mathbb{T}^i_u$, where:

$$\begin{aligned}
\mathbb{T}^1_u &:= -\langle\kappa(\boldsymbol{\sigma}_h)\,l^{-1}\,(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)), \psi + l\partial_n\psi\rangle_{\Gamma_h}, \quad \mathbb{T}^7_u := -\langle\tau(\varepsilon^u - \varepsilon^{\widehat{u}}), P_M\psi\rangle_{\Gamma_h},\\
\mathbb{T}^2_u &:= -\langle\kappa(u_h)\,(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h)), (Id_M - P_M)\partial_n\psi\rangle_{\Gamma_h},\\
\mathbb{T}^8_u &:= -\langle\kappa(\boldsymbol{\sigma}_h)\,(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h))\,\delta_{\boldsymbol{I}^{\boldsymbol{q}}}, \psi\rangle_{\Gamma_h}\\
\mathbb{T}^3_u &:= \langle\delta_{\boldsymbol{I}^{\boldsymbol{q}}}, \psi\rangle_{\Gamma_h}, \quad \mathbb{T}^9_u := -\langle\kappa(\boldsymbol{\sigma}_h)\,(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h))\,\delta_{\boldsymbol{\Pi_V}\boldsymbol{q}}, \psi\rangle_{\Gamma_h},\\
\mathbb{T}^4_u &:= \left\langle\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n}, (Id_M - P_M)\psi\right\rangle_{\Gamma_h}, \quad \mathbb{T}^{10}_u := -\langle\kappa(\boldsymbol{\sigma}_h)\,(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h))\,\boldsymbol{I}^{\boldsymbol{q}}\cdot\boldsymbol{n}, \psi\rangle_{\Gamma_h}.\\
\mathbb{T}^5_u &:= -\left\langle\tau\,P_M I^u, \psi\right\rangle_{\Gamma_h}, \quad \mathbb{T}^{11}_u := -\langle\kappa(\boldsymbol{\sigma}_h)\,(\varphi(\boldsymbol{\sigma}) - \varphi_h(\boldsymbol{\sigma}_h))\,\boldsymbol{\Pi_V}\boldsymbol{q}\cdot\boldsymbol{n}, \psi\rangle_{\Gamma_h},\\
\mathbb{T}^6_u &:= \langle\delta_{\boldsymbol{\varepsilon}^{\boldsymbol{q}}}, \psi\rangle_{\Gamma_h}.
\end{aligned}$$

These terms can be estimated by arguments like the ones detailed in Lemma 6. Finally, taking $\Theta = \varepsilon^u$ in (4.22) and considering the estimates for the components of $\mathbb{T}^i_u$ it is possible to deduce that

$$\|\varepsilon^u\|^2_{\Omega_h} \lesssim 3\,h|||(\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||^2_{\boldsymbol{\sigma}_h} + 6\Big((h + L^2\,h^2)\Lambda^2_{\boldsymbol{q}} + (L^2 + h)\,\Lambda^2_u\Big).$$

□

The result of the previous Lemma allows us to estimate the error incurred by the HDG approximation by that of the HDG projection onto the discrete space, as we now show.

Theorem 4 *If L is small enough and the discrete spaces are of polynomial degree $k \geq 1$, then there exists $h_0 > 0$ such that, for all $h \leq h_0$, we have*

$$|||(\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||^2_{\boldsymbol{\sigma}_h} \lesssim \Lambda^2_{\boldsymbol{q}} + \Lambda^2_u + \Lambda^2_{\boldsymbol{\sigma}}. \tag{4.23}$$

Proof Using simple algebraic arguments, note that the term (4.20) can be rewritten as

$$|||(\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||^2_{\boldsymbol{\sigma}} \lesssim \Lambda^2_{\boldsymbol{q}} + \Lambda^2_{\boldsymbol{\sigma}} + \frac{3}{2} L_f \|\varepsilon^u\|^2_{\Omega_h} + \frac{1}{2} L_f \, \Lambda_u \|\varepsilon^u\|_{\Omega_h}$$

Combined the above with the estimate given in the Lemma 9, we can deduce

$$\left(1 - \frac{9}{2} L_f \, h \, c\right) |||(\boldsymbol{\varepsilon}^{\boldsymbol{\sigma}}, \boldsymbol{\varepsilon}^{\boldsymbol{q}}, \varepsilon^u - \varepsilon^{\widehat{u}}, \varphi - \varphi_h)|||^2_{\boldsymbol{\sigma}} \lesssim 9 \, L_f \left((h + L^2 h^2) \Lambda^2_{\boldsymbol{q}} + (L^2 + h) \, \Lambda^2_u \right)$$
$$+ \Lambda^2_{\boldsymbol{q}} + \Lambda^2_{\boldsymbol{\sigma}} + \frac{1}{2} L_f \Lambda^2_u,$$

where $c > 0$ is a constant independent of h arising from the symbol $\lesssim$. Assuming that L_f is small enough and considering that $h \leq h_0$, te proof is concluded. □

As a consequence of this theorem, it follows that the HDG approximation of the linearized systems will indeed achieve optimal order of convergence with respect to the degree of the polynomial basis, provided that the true solutions are smooth enough.

Corollary 3 *Suppose that assumptions of Theorem 4 hold. If $u \in H^{k+1}(\Omega)$ and $\boldsymbol{q} \in \boldsymbol{H}^{k+1}(\Omega)$, then*

$$\|\boldsymbol{\sigma} - \boldsymbol{\sigma}_h\|_\Omega + \|\boldsymbol{q} - \boldsymbol{q}_h\|_\Omega + \|u - u_h\|_\Omega \leq C h^{k+1} \left(|u|_{k+1,\Omega} + |\boldsymbol{q}|_{k+1,\Omega} + |\boldsymbol{\sigma}|_{k+1,\Omega} \right).$$

Acknowledgements The authors have no relevant financial or non-financial interests to disclose. All authors have contributed equally to the article and the order of authorship has been determined alphabetically. Nestor Sánchez is supported by the Scholarship Program of CONICYT-Chile. Manuel E. Solano was partially funded by CONICYT–Chile through FONDECYT project No. 1200569 and by Project AFB170001 of the PIA Program: Concurso Apoyo a Centros Científicos y Tecnológicos de Excelencia con Financiamiento Basal. Tonatiuh Sánchez–Vizuet was partially supported by the National Science Foundation throught the grant NSF-DMS-2137305 "LEAPS-MPS: Hybridizable discontinuous Galerkin methods for non-linear integro-differential boundary value problems in magnetic plasma confinement".

Appendix A: HDG Projection

The HDG projectors introduced by [3] and their properties have been used extensively throughout the text. Here we provide a quick definition and summary of the properties used in this article.

Consider constants $l_u, l_{\boldsymbol{q}} \in [0, k]$ and functions $(\boldsymbol{q}, u) \in H^{1+l_q}(T) \times H^{1+l_u}(T)$. Moreover, recall the discrete spaces

$$\boldsymbol{V}_h := \{\boldsymbol{v} \in \boldsymbol{L}^2(\mathcal{T}_h) : \boldsymbol{v}|_T \in [\mathbb{P}_k(T)]^d, \quad \forall \, T \in \mathcal{T}_h\},$$
$$W_h := \{w \in L^2(\mathcal{T}_h) : w|_T \in \mathbb{P}_k(T), \quad \forall \, T \in \mathcal{T}_h\},$$
$$M_h := \{\mu \in L^2(\mathcal{E}_h) : \mu|_T \in \mathbb{P}_k(e), \quad \forall \, e \in \mathcal{E}_h\},$$

defined in (2.7) in the text. We will denote by $\boldsymbol{\Pi}(\boldsymbol{q}, u) := (\boldsymbol{\Pi}_{\mathrm{v}}\boldsymbol{q}, \Pi_{\mathrm{w}} u)$ the projection over $\boldsymbol{V}_h \times W_h$ defined by the unique element-wise solutions of

$$(\boldsymbol{\Pi}_{\mathrm{v}}\boldsymbol{q}, \boldsymbol{v})_T = (\boldsymbol{q}, \boldsymbol{v})_T \qquad \forall \, \boldsymbol{v} \in [\mathbb{P}_{k-1}(T)]^d, \tag{A.1a}$$

$$(\Pi_{\mathrm{w}} u, w)_T = (u, w)_T \qquad \forall\, w \in \mathbb{P}_{k-1}(T), \tag{A.1b}$$

$$\langle \boldsymbol{\Pi}_{\mathrm{v}} \boldsymbol{q} \cdot \boldsymbol{n} + \tau \Pi_{\mathrm{w}} u, \mu \rangle_F = \langle \boldsymbol{q} \cdot \boldsymbol{n} + \tau u, \mu \rangle_F \qquad \forall\, \mu \in \mathbb{P}_k(F), \tag{A.1c}$$

for every element $T \in \mathcal{T}_h$, and $F \subset \partial T$. Will will denote the L^2 projector into M_h by P_M. It was proven in [3] that when the stabilization function is chosen so that $\tau_T^{\max} := \max \tau|_{\partial T} > 0$, then there is a constant $C > 0$ independent of T and τ such that

$$\|\boldsymbol{\Pi}_{\mathrm{v}} \boldsymbol{q} - \boldsymbol{q}\|_T \leq C h_T^{l_q+1} |\boldsymbol{q}|_{\boldsymbol{H}^{l_q+1}(T)} + C h_T^{l_u+1} \tau_T^* |u|_{H^{l_u+1}(T)}, \tag{A.2a}$$

$$\|\Pi_{\mathrm{w}} u - u\|_T \leq C h_T^{l_u+1} |u|_{H^{l_u+1}(T)} + C \frac{h_T^{l_q+1}}{\tau_T^{\max}} |\nabla \cdot \boldsymbol{q}|_{H^{l_q}(T)}. \tag{A.2b}$$

Here $\tau_T^* := \max \tau|_{\partial T \setminus F^*}$ and F^* is a face of T at which $\tau|_{\partial T}$ is maximum. As is customary, the symbol $|\cdot|_{H^s}$ is to be understood as the Sobolev semi norm of order $s \in \mathbb{R}$.

Appendix B: Auxiliary Estimates

Duality argument We will consider that, given $\Theta \in L^2(\Omega)$, the solution to the auxiliary problem

$$\kappa^{-1}(\zeta)\boldsymbol{\phi} + \nabla \psi = 0 \qquad \text{in } \Omega, \tag{B.1a}$$

$$\nabla \cdot \boldsymbol{\phi} = \Theta \qquad \text{in } \Omega, \tag{B.1b}$$

$$\psi = 0 \qquad \text{on } \partial\Omega, \tag{B.1c}$$

satisfies the regularity estimate

$$\|\boldsymbol{\phi}\|_{H^1(\Omega)} + \|\psi\|_{H^2(\Omega)} \leq C_{reg} \|\Theta\|_\Omega. \tag{B.2}$$

where $C_{reg} > 0$ depends on the domain Ω. Moreover, if Ω_h is a subdomain of Ω with boundary $\Gamma_h := \partial\Omega_h$, $l(x)$ is the length of the transfer path connecting $\Gamma := \partial\Omega$ to Γ_h as defined in Sect. 2.2, P_M is the L^2-projector onto the discrete space M_h defined in (2.7), and Id_M is the identity operator in M_h we have

Lemma B1 *[5,Lemma 5.5] Suppose Assumption (2.4b) and the elliptic regularity inequality (B.2) hold. Then, there exists a constant $\tilde{c} > 0$, such that:*

$$\|(h^{\perp})^{-1/2}(Id_M - P_M)\psi\|_{\Gamma_h} \leq \tilde{c}\, h \|\Theta\|_\Omega, \tag{B.3a}$$

$$\|l^{1/2}(Id_M - P_M)\partial_n \psi\|_{\Gamma_h} \leq \tilde{c}\, R^{1/2}\, h \|\Theta\|_\Omega, \tag{B.3b}$$

$$\|l^{-3/2}(\psi + l\partial_n \psi)\|_{\Gamma_h} \leq \tilde{c}\, \|\Theta\|_\Omega, \tag{B.3c}$$

$$\|l^{-1}\psi\|_{\Gamma_h} \leq \tilde{c}\, \|\Theta\|_\Omega. \tag{B.3d}$$

Function Delta For any smooth enough function $\boldsymbol{v}$ defined in $T^e \cup T^e_{ext}$ and $\boldsymbol{x} \in \Gamma_h$ we set

$$\delta_{\boldsymbol{v}}(\boldsymbol{x}) := \frac{1}{l(\boldsymbol{x})} \int_0^{l(\boldsymbol{x})} [\boldsymbol{v}(\boldsymbol{x} + \boldsymbol{n}s) - \boldsymbol{v}(\boldsymbol{x})] \cdot \boldsymbol{n} ds. \tag{B.4}$$

which hold for each $e \in \mathcal{E}_h^{\partial}$ (cf. [5,Lemma 5.2]):

$$\|l^{1/2}\, \delta_{\boldsymbol{v}}\|_e \leq \frac{1}{\sqrt{3}} r_e^{3/2}\, C^e_{ext}\, C^e_{inv}\, \|\boldsymbol{v}\|_{T^e} \qquad \forall\, \boldsymbol{v} \in [\mathbb{P}_k(T)]^d, \tag{B.5a}$$

$$\|l^{1/2}\,\delta_{\boldsymbol{v}}\|_e \leq \frac{1}{\sqrt{3}}\, r_e\, \|h^{\perp}\partial_n \boldsymbol{v}\cdot\boldsymbol{n}\|_{T^e_{ext}} \qquad \forall\, \boldsymbol{v}\in[H^1(T)]^d. \tag{B.5b}$$

$$\|l^{1/2}\,\delta_{\boldsymbol{v}}\|_\infty \leq \frac{1}{\sqrt{3}}\, r_e\, \sup_{\boldsymbol{x}\in e} \|h_e^{\perp}\,\partial_n \boldsymbol{v}\cdot\boldsymbol{n}\|_{l(\boldsymbol{x})} \qquad \forall\, \boldsymbol{v}\in[H^1(T)]^d. \tag{B.5c}$$

References

1. Chen, G., Cockburn, B., Singler, J., Zhang, Y.: Superconvergent interpolatory HDG methods for reaction diffusion equations I: an HDG$_k$ method. J. Sci. Comput. **81**(3), 2188–2212 (2019)
2. Cockburn, B., Gopalakrishnan, J., Lazarov, R.D.: Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. SIAM J. Numer. Anal. **47**, 1319–1365 (2009)
3. Cockburn, B., Gopalakrishnan, J., Sayas, F.-J.: A projection-based error analysis of HDG methods. Math. Comput. **79**, 1351–1367 (2010)
4. Cockburn, B., Gupta, D., Reitich, F.: Boundary-conforming discontinuous Galerkin methods via extensions from subdomains. J. Sci. Comput. **42**(1), 144 (2009)
5. Cockburn, B., Qiu, W., Solano, M.: A priori error analysis for HDG methods using extensions from subdomains to achieve boundary conformity. Math. Comput. **83**(286), 665–699 (2014)
6. Cockburn, B., Singler, J.R., Zhang, Y.: Interpolatory HDG method for parabolic semilinear PDEs. J. Sci. Comput. **79**(3), 1777–1800 (2019)
7. Cockburn, B., Solano, M.: Solving Dirichlet boundary-value problems on curved domains by extensions from subdomains. SIAM J. Sci. Comput. **34**(1), A497–A519 (2012)
8. Dautov, R.Z., Fedotov, E.M.: Abstract theory of hybridizable discontinuous Galerkin methods for second-order quasilinear elliptic problems. Comput. Math. Math. Phys. **54**(3), 474–490 (2014)
9. Gatica, G.N., Sequeira, F.A.: Analysis of an augmented HDG method for a class of Quasi-Newtonian Stokes flows. J. Sci. Comput. **65**(3), 1270–1308 (2015)
10. Gatica, G.N., Sequeira, F.A.: A priori and a posteriori error analyses of an augmented HDG method for a class of Quasi-Newtonian Stokes flows. J. Sci. Comput. **69**, 1192–1250 (2016)
11. Grad, H., Rubin, H.: Hydromagnetic equilibria and force-free fields. In: Proceedings of the Second international conference on the peaceful uses of atomic energy, Geneva, vol. 31,190, New York, United Nations (Oct 1958)
12. Leng, H.: Adaptive HDG methods for the steady-state incompressible Navier–Stokes equations. J. Sci. Comput. **87**(1), 1–26 (2021)
13. Lüst, R., Schlüter, A.: Axialsymmetrische magnetohydrodynamische Gleichgewichtskonfigurationen. Z. Naturf **12a**, 850–854 (1957)
14. Marche, F.: Combined hybridizable discontinuous Galerkin (HDG) and Runge–Kutta discontinuous Galerkin (RK-DG) formulations for Green–Naghdi equations on unstructured meshes. J. Comput. Phys. **418**, 109637 (2020)
15. Qiu, W., Shi, K.: A mixed DG method and an HDG method for incompressible magnetohydrodynamics. IMA J. Numer. Anal. **40**(2), 1356–1389, 01 (2019)
16. Sánchez, N., Sánchez-Vizuet, T., Solano, M.E.: A priori and a posteriori error analysis of an unfitted HDG method for semi-linear elliptic problems. Numer. Math. **148**(4), 919–958 (2021)
17. Sánchez-Vizuet, T., Solano, M.E.: A Hybridizable discontinuous Galerkin solver for the Grad-Shafranov equation. Comput. Phys. Commun. **235**, 120–132 (2019)
18. Sánchez-Vizuet, T., Solano, M.E., Cerfon, A.J.: Adaptive hybridizable discontinuous Galerkin discretization of the Grad–Shafranov equation by extension from polygonal subdomains. Comput. Phys. Commun. **255**, 107239 (2020)
19. Shafranov, V.D.: On magnetohydrodynamical equilibrium configurations. Sov. Phys. JETP **6**, 545–554 (1958)