An introduction to Hybrid High-Order methods

Daniele Antonio Di Pietro, Roberta Tittarelli

An introduction to Hybrid High-Order methods

Daniele A. Di Pietro and Roberta Tittarelli

Abstract This chapter provides an introduction to Hybrid High-Order (HHO) methods. These are new generation numerical methods for PDEs with several advantageous features: the support of arbitrary approximation orders on general polyhedral meshes, the reproduction at the discrete level of relevant continuous properties, and a reduced computational cost thanks to static condensation and compact stencil. After establishing the discrete setting, we introduce the basics of HHO methods using as a model problem the Poisson equation. We describe in detail the construction, and prove a priori convergence results for various norms of the error as well as a posteriori estimates for the energy norm. We then consider two applications: the discretization of the nonlinear p-Laplace equation and of scalar diffusion-advection-reaction problems. The former application is used to introduce compactness analysis techniques to study the convergence to minimal regularity solution. The latter is used to introduce the discretization of first-order operators and the weak enforcement of boundary conditions. Numerical examples accompany the exposition.

1 Introduction

This chapter provides an introduction to Hybrid High-Order (HHO) methods. The material is closely inspired by a series of lectures given by the first author at Institut Henri Poincaré in September 2016 within the thematic quarter *Numerical Methods for PDEs* (see `http://tinyurl.com/IHP-quarter-nmpdes`).

HHO methods, introduced in [19, 22], are discretization methods for Partial Differential Equations (PDEs) with relevant features that set them apart from classical Finite Elements or Finite Volumes schemes. These include, in particular:

Daniele A. Di Pietro · Roberta Tittarelli
Institut Montpelliérain Alexander Grothendieck, Université de Montpellier, place Eugène Bataillon, 34095, Montpellier (France). e-mail: `daniele.di-pietro@umontpellier.fr`, e-mail: `roberta.tittarelli@umontpellier.fr`

(i) The support of general polytopal meshes in arbitrary space dimension, paving the way to a seamless treatment of complex geometric features and unified 1d-2d-3d implementations;
(ii) The possibility to select the approximation order which, possibly combined with adaptivity, leads to a reduction of the simulation cost for a given precision or better precision for a given cost;
(iii) The compliance with the physics, including robustness with respect to the variations of physical coefficients and reproduction at the discrete level of key continuous properties such as local balances and flux continuity;
(iv) A reduced computational cost thanks to their compact stencil along with the possibility to perform static condensation.

As of today, HHO methods have been successfully applied to the discretization of several linear and nonlinear problems of engineering interest including: variable diffusion [20, 22, 23], quasi incompressible linear elasticity [18, 19], locally degenerate diffusion-advection-reaction [15], poroelasticity [4], creeping flows [1] possibly driven by volumetric forces with large irrotational part [24], electrostatics [27], phase separation problems governed by the Cahn–Hilliard equation [8], Leray–Lions type elliptic problems [13, 14]. More recent applications also include steady incompressible flows governed by the Navier–Stokes equations [25] and nonlinear elasticity [6]. Generalizations of HHO methods and comparisons with other (new generation or classical) discretization methods for PDEs can be found in [5, 10]. Implementation tools based on advanced programming techniques have been recently discussed in [9].

For the sake of simplicity, the introduction provided in this chapter focuses on scalar model problems. We start in Section 2 by presenting the discrete setting: we introduce the notion of polytopal mesh (Section 2.1), formulate assumptions on the way meshes are refined that are suitable to carry out a h-convergence analysis (Section 2.2), introduce the local polynomial spaces (Section 2.3) and projectors (Section 2.4) that lie at the heart of the HHO construction.

In Section 3 we present the basic principles of HHO methods using as a model problem the Poisson equation. While the material in this section is mainly adapted from [22], some results are new and the arguments have been shortened or made more elegant. In Section 3.1 we introduce the local space of degrees of freedom (DOFs) and discuss the main ingredients upon which HHO methods rely, namely:

(i) Reconstructions of relevant quantities obtained by solving small, embarassingly parallel problems on each element;
(ii) High-order stabilization terms obtained by penalizing cleverly designed residuals.

In Section 3.2 we show how to combine these ingredients to formulate local contributions, which are then assembled element-by-element as in standard Finite Elements. The construction is conceived so that only face-based DOFs are globally coupled, which paves the way to efficient practical implementations where element-based DOFs are statically condensed in a preliminary step. In Sections 3.3 and 3.4 we discuss, respectively, optimal a priori estimates for various norms and seminorms

of the error, and residual-based a posteriori estimates for the energy-norm of the error. Finally, some numerical examples are provided in Section 3.5 to demonstrate the theoretical results.

In Section 4 we consider the HHO discretization of the p-Laplace equation. The material is inspired by [13, 14], where more general Leray–Lions operators are considered. When dealing with nonlinear problems, regularity for the exact solution is often difficult to prove and can entail stringent assumptions on the data. For this reason, the h-convergence analysis can be carried out in two steps: in a first step, convergence to minimal regularity solutions is proved by a compactness argument; in a second step, convergence rates are estimated for smooth solutions (and smooth data). Convergence by compactness typically requires discrete counterparts of functional analysis results relevant for the study of the continuous problem. In our case, two sets of discrete functional analysis results are needed: discrete Sobolev embeddings (Section 4.1) and compactness for sequences of HHO functions uniformly bounded in a $W^{1,p}$-like seminorm (Section 4.2). The interest of both results goes beyond the specific method and problem considered here. As an example, in [25] they are used for the analysis of a HHO discretization of the steady incompressible Navier–Stokes equations. The HHO method for the p-Laplacian stated in Section 4.3 is designed according to similar principles as for the Poisson problem. Convergence results are stated in Section 4.4, and numerical examples are provided in Section 4.5.

Following [15], in Section 5 we extend the HHO method to diffusion-advection-reaction problems. In this context, a crucial property from the numerical point of view is the robustness in the advection-dominated regime. In Section 5.1 we modify the diffusive bilinear form introduced in Section 3.2 to incorporate weakly enforced boundary conditions. The weak enforcement of boundary conditions typically improves the behaviour of the method in the presence of boundary layers, since the discrete solution is not constrained to a fixed value on the boundary. In Section 5.2 we introduce the HHO discretization of first-order terms based on two novel ingredients: a local advective derivative reconstruction and an upwind penalty term. The former is used to formulate the consistency terms, while the role of the latter is to confer suitable stability properties to the advective-reactive bilinear form. The HHO discretization is finally obtained in Section 5.3 combining the diffusive and advective-reactive contributions, and its stability with respect to an energy-like norm including an advective derivative contribution is studied. In Section 5.4 we state an energy-norm error estimate which accounts for the dependence of the error contribution of each mesh element on a local Péclet number. A numerical illustration is provided in Section 5.5.

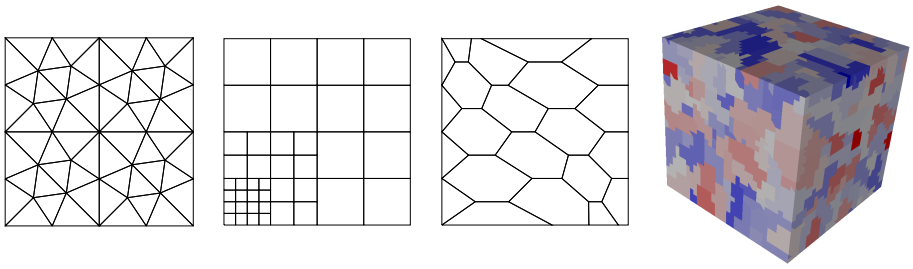
2 Discrete setting

Let $\Omega \subset \mathbb{R}^d$, $d \in \mathbb{N}^*$, denote a bounded connected open polyhedral domain with Lipschitz boundary and outward normal $\mathbf{n}$. We assume that Ω does not have cracks,

i.e., it lies on one side of its boundary. In what follows, we introduce the notion of polyhedral mesh of Ω, formulate assumptions on the way meshes are refined that enable to prove useful geometric and functional results, and introduce functional spaces and projectors that will be used in the construction and analysis of HHO methods.

2.1 Polytopal mesh

The following definition enables the treatment of meshes as general as the ones depicted in Fig. 1.



(a) Matching triangular (b) Nonconforming (c) Polygonal (d) Agglomerated

Fig. 1: Examples of polytopal meshes in two and three space dimensions. The triangular and nonconforming meshes are taken from the FVCA5 benchmark [31], the polygonal mesh family from [26, Section 4.2.3], and the agglomerated polyhedral mesh from [27].

Definition 1 (Polytopal mesh). A polytopal mesh of Ω is a couple $\mathscr{M}_h = (\mathscr{T}_h, \mathscr{F}_h)$ where:

(i) The set of *mesh elements* $\mathscr{T}_h$ is a finite collection of nonempty disjoint open polytopes T with boundary ∂T and diameter h_T such that the *meshsize* h satisfies $h = \max_{T\in\mathscr{T}_h} h_T$ and it holds that $\overline{\Omega} = \bigcup_{T\in\mathscr{T}_h} \overline{T}$.

(ii) The set of *mesh faces* $\mathscr{F}_h$ is a finite collection of disjoint subsets of $\overline{\Omega}$ such that, for any $F \in \mathscr{F}_h$, F is an open subset of a hyperplane of $\mathbb{R}^d$, the $(d-1)$-dimensional Hausdorff measure of F is strictly positive, and the $(d-1)$-dimensional Hausdorff measure of its relative interior $\overline{F}\setminus F$ is zero. Moreover, (a) for each $F \in \mathscr{F}_h$, either there exist two distinct mesh elements $T_1, T_2 \in \mathscr{T}_h$ such that $F \subset \partial T_1 \cap \partial T_2$ and F is called an *interface* or there exists one mesh element $T \in \mathscr{T}_h$ such that $F \subset \partial T \cap \partial\Omega$ and F is called a *boundary face*; (b) the set of faces is a partition of the mesh skeleton, i.e., $\bigcup_{T\in\mathscr{T}_h} \partial T = \bigcup_{F\in\mathscr{F}_h} \overline{F}$.

Interfaces are collected in the set $\mathscr{F}_h^{\mathrm{i}}$ and boundary faces in $\mathscr{F}_h^{\mathrm{b}}$, so that $\mathscr{F}_h = \mathscr{F}_h^{\mathrm{i}} \cup \mathscr{F}_h^{\mathrm{b}}$. For any mesh element $T \in \mathscr{T}_h$,

$$\mathscr{F}_T := \{F \in \mathscr{F}_h \mid F \subset \partial T\}$$

denotes the set of faces contained in ∂T. Similarly, for any mesh face $F \in \mathscr{F}_h$,

$$\mathscr{T}_F := \{T \in \mathscr{T}_h \mid F \subset \partial T\}$$

is the set of mesh elements sharing F. Finally, for all $F \in \mathscr{F}_T$, $\mathbf{n}_{TF}$ is the unit normal vector to F pointing out of T.

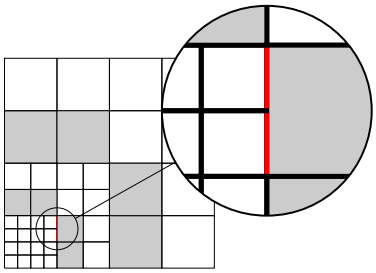


Fig. 2: Treatment of a nonconforming junction (red) as multiple coplanar faces. Gray elements are pentagons with two coplanar faces, white elements are squares.

Remark 1 (Nonconforming junctions). Meshes including nonconforming junctions such as the one depicted in Fig. 2 are naturally supported provided that each face containing hanging nodes is treated as multiple coplanar faces.

2.2 Regular mesh sequences

When studying the convergence of HHO methods with respect to the meshsize h, one needs to make assumptions on how the mesh is refined. The ones provided here are closely inspired by [17, Chapter 1], and refer to the case of isotropic meshes with non-degenerate faces. Isotropic means here that we do not consider the case of elements that become more and more stretched when refining. Non-degenerate faces means, on the other hand, that the diameter of each mesh face is uniformly comparable to that of the element(s) it belongs to; see (2) below.

Definition 2 (Matching simplicial submesh). Let $\mathscr{M}_h = (\mathscr{T}_h, \mathscr{F}_h)$ be a polytopal mesh of Ω. We say that $\mathfrak{T}_h$ is a matching simplicial submesh of $\mathscr{M}_h$ if (i) $\mathfrak{T}_h$ is a matching simplicial mesh; (ii) for all simplices $\tau \in \mathfrak{T}_h$, there is only one mesh element $T \in \mathscr{T}_h$ such that $\tau \subset T$; (iii) for all $\sigma \in \mathfrak{F}_h$, the set collecting the simplicial faces of $\mathfrak{T}_h$, there is at most one face $F \in \mathscr{F}_h$ such that $\sigma \subset F$.

If $\mathscr{T}_h$ itself is matching simplicial and $\mathscr{F}_h$ collects the corresponding simplicial faces, we can simply take $\mathfrak{T}_h = \mathscr{T}_h$, so that $\mathfrak{F}_h = \mathscr{F}_h$. The notion of regularity for refined mesh sequences is made precise by the following

Definition 3 (Regular mesh sequence). Denote by $\mathscr{H} \subset \mathbb{R}_*^+$ a countable set of meshsizes having 0 as its unique accumulation point. A sequence of refined meshes $(\mathscr{M}_h)_{h\in\mathscr{H}}$ is said to be *regular* if there exists a real number $\rho \in (0,1)$ such that, for all $h \in \mathscr{H}$, there exists a matching simplicial submesh $\mathfrak{T}_h$ of $\mathscr{M}_h$ and (i) for all simplices $\tau \in \mathfrak{T}_h$ of diameter h_τ and inradius r_τ, $\rho h_\tau \le r_\tau$; (ii) for all mesh elements $T \in \mathscr{T}_h$ and all simplices $\tau \in \mathfrak{T}_h$ such that $\tau \subset T$, $\rho h_T \le h_\tau$.

Remark 2 (Role of the simplicial submesh). The simplicial submesh introduced in Definition 3 is merely a theoretical tool, and needs not be constructed in practice.

Geometric bounds on regular mesh sequences can be proved as in [17, Section 1.4.2] (the definition of mesh face is slightly different therein since planarity is not required, but the proofs are based on the matching simplicial submesh and one can check that they carry out unchanged). We recall here, in particular, that the number of faces of one mesh element is uniformly bounded: There is $N_\partial \ge d+1$ such that

$$\max_{h\in\mathscr{H}} \max_{T\in\mathscr{T}_h} \mathrm{card}(\mathscr{F}_T) \le N_\partial. \tag{1}$$

Morevover, according to [17, Lemma 1.42], for all $h \in \mathscr{H}$, all $T \in \mathscr{T}_h$, and all $F \in \mathscr{F}_T$

$$\rho^2 h_T \le h_F \le h_T. \tag{2}$$

Discrete functional analysis results on regular mesh sequences can be found in [17, Chapter 1] and [13, 14].

2.3 Local and broken spaces

Throughout the rest of this chapter, for any $X \subset \overline{\Omega}$, we denote by $(\cdot,\cdot)_X$ and $\|\cdot\|_X$ the standard $L^2(X)$-product and norm, with the convention that the subscript is omitted whenever $X = \Omega$. The same notation is used for the vector-valued space $L^2(X)^d$.

Let now the set X be a mesh element or face. For an integer $l \ge 0$, we denote by $\mathbb{P}^l(X)$ the space spanned by the restriction to X of scalar-valued, d-variate polynomials of total degree l. We note the following trace inequality (see [17, Lemma 1.46]): There is a real number $C > 0$ only depending on d, ρ, and l such that, for all $h \in \mathscr{H}$, all $T \in \mathscr{T}_h$, all $v \in \mathbb{P}^l(T)$, and all $F \in \mathscr{F}_T$,

$$\|v\|_F \le C h_T^{-1/2} \|v\|_T. \tag{3}$$

At the global level, we define the broken polynomial space

$$\mathbb{P}^l(\mathscr{T}_h) := \left\{ v_h \in L^2(\Omega) \mid v_{h|T} \in \mathbb{P}^l(T) \quad \forall T \in \mathscr{T}_h \right\}.$$

Functions in $\mathbb{P}^l(\mathscr{T}_h)$ belong to the broken Sobolev space

$$W^{1,1}(\mathscr{T}_h) := \left\{ v \in L^1(\Omega) \mid v_{|T} \in W^{1,1}(T) \quad \forall T \in \mathscr{T}_h \right\}.$$

We denote by $\nabla_h : W^{1,1}(\mathscr{T}_h) \to L^1(\Omega)^d$ the usual broken gradient operator such that, for all $v \in W^{1,1}(\mathscr{T}_h)$,

$$(\nabla_h v)_{|T} = \nabla v_{|T} \qquad \forall T \in \mathscr{T}_h.$$

2.4 Projectors on local polynomial spaces

Projectors on local polynomial spaces play a key role in the design and analysis of HHO methods.

2.4.1 L^2-orthogonal projector

Let X denote a mesh element or face. The L^2-orthogonal projector (in short, L^2-projector) $\pi_X^{0,l} : L^1(X) \to \mathbb{P}^l(X)$ is defined as follows: For all $v \in L^1(X)$, $\pi_X^{0,l}$ is the unique polynomial in $\mathbb{P}^l(X)$ that satisfies

$$(\pi_X^{0,l} v - v, w)_X = 0 \qquad \forall w \in \mathbb{P}^l(X). \tag{4}$$

Existence and uniqueness of $\pi_X^{0,l} v$ follow from the Riesz representation theorem in $\mathbb{P}^l(X)$ for the standard $L^2(X)$-inner product. Moreover, we have the following characterization:

$$\pi_X^{0,l} v = \operatorname*{arg\,min}_{w \in \mathbb{P}^l(X)} \|w - v\|_X^2.$$

In what follows, we will also need the vector-valued L^2-projector denoted by $\boldsymbol{\pi}_X^{0,l}$ and obtained by applying $\pi_X^{0,l}$ component-wise. The following H^s-boundedness result is a special case of [13, Corollary 3.7]: For any $s \in \{1, \ldots, l+1\}$, there exists a real number $C > 0$ depending only on d, ρ, l, and s such that, for all $h \in \mathscr{H}$, all $T \in \mathscr{T}_h$, and all $v \in H^s(T)$,

$$|\pi_T^{0,l} v|_{H^s(T)} \le C |v|_{H^s(T)}. \tag{5}$$

At the global level, we denote by $\pi_h^{0,l} : L^1(\Omega) \to \mathbb{P}^l(\mathscr{T}_h)$ the L^2-projector on the broken polynomial space $\mathbb{P}^l(\mathscr{T}_h)$ such that, for all $v \in L^1(\Omega)$,

$$(\pi_h^{0,l} v)_{|T} := \pi_T^{0,l} v_{|T}.$$

2.4.2 Elliptic projector

For any mesh element $T \in \mathscr{T}_h$, we also define the elliptic projector $\pi_T^{1,l} : W^{1,1}(T) \to \mathbb{P}^l(T)$ as follows: For all $v \in W^{1,1}(T)$, $\pi_T^{1,l} v$ is the unique polynomial in $\mathbb{P}^l(T)$ that satisfies

$$(\nabla(\pi_T^{1,l} v - v), \nabla w)_T = 0 \qquad \forall w \in \mathbb{P}^l(T). \tag{6a}$$

By the Riesz representation theorem in $\nabla\mathbb{P}^l(T)$ for the $L^2(T)^d$-inner product, this relation defines a unique element $\nabla\pi_T^{1,l} v$, and thus a polynomial $\pi_T^{1,l} v$ up to an additive constant. This constant is fixed by writing

$$(\pi_T^{1,l} v - v, 1)_T = 0. \tag{6b}$$

Observing that (6a) is trivially verified when $l = 0$, it follows from (6b) that $\pi_T^{1,0} = \pi_T^{0,0}$. Finally, the following characterization holds:

$$\pi_T^{1,l} v = \underset{w \in \mathbb{P}^l(T),\, (w-v,1)_T = 0}{\arg\min} \|\nabla(w - v)\|_{L^2(T)^d}^2.$$

2.4.3 Approximation properties

On regular mesh sequences, both $\pi_T^{0,l}$ and $\pi_T^{1,l}$ have optimal approximation properties in $\mathbb{P}^l(T)$, as summarized by the following result (for a proof, see Theorem 1, Theorem 2, and Lemma 13 in [13]): For any $\alpha \in \{0,1\}$ and $s \in \{\alpha, \ldots, l+1\}$, there exists a real number $C > 0$ depending only on d, ρ, l, α, and s such that, for all $h \in \mathscr{H}$, all $T \in \mathscr{T}_h$, and all $v \in H^s(T)$,

$$|v - \pi_T^{\alpha,l} v|_{H^m(T)} \le C h_T^{s-m} |v|_{H^s(T)} \qquad \forall m \in \{0, \ldots, s\}, \tag{7a}$$

and, if $s \ge 1$,

$$|v - \pi_T^{\alpha,l} v|_{H^m(\mathscr{F}_T)} \le C h_T^{s-m-\frac{1}{2}} |v|_{H^s(T)} \qquad \forall m \in \{0, \ldots, s-1\}, \tag{7b}$$

where $H^m(\mathscr{F}_T) := \big\{ v \in L^2(\partial T) \mid v_{|F} \in H^m(F) \quad \forall F \in \mathscr{F}_T \big\}$.

3 Basic principles of Hybrid High-Order methods

To fix the main ideas and notation, we study in this section the HHO discretization of the Poisson problem: Find $u : \Omega \to \mathbb{R}$ such that

$$-\triangle u = f \qquad \text{in } \Omega, \tag{8a}$$

$$u = 0 \qquad \text{on } \partial\Omega, \tag{8b}$$

where $f \in L^2(\Omega)$ is a given volumetric source term. More general boundary conditions can replace (8b), but we restrict the discussion to the homogeneous Dirichlet case for the sake of simplicity. A detailed treatment of more general boundary conditions including also variable diffusion coefficients can be found in [23].

The starting point to devise a HHO discretization is the following weak formulation of problem (8): Find $u \in H^1_0(\Omega)$ such that

$$a(u,v) = (f,v) \qquad \forall v \in H^1_0(\Omega), \tag{9}$$

where the bilinear form $a : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ is such that

$$a(u,v) := (\nabla u, \nabla v). \tag{10}$$

In what follows, the quantities u and $-\nabla u$ will be referred to, respectively, as the potential and the flux.

3.1 Local construction

Throughout this section, we fix a polynomial degree $k \geq 0$ and a mesh element $T \in \mathscr{T}_h$. We introduce the local ingredients underlying the HHO construction: the DOFs, the potential reconstruction operator, and the discrete counterpart of the restriction to T of the global bilinear form a defined by (10).

3.1.1 Computing the local elliptic projection from L^2-projections

Consider a function $v \in H^1(T)$. We note the following integration by parts formula, valid for all $w \in C^\infty(\overline{T})$:

$$(\nabla v, \nabla w)_T = -(v, \triangle w)_T + \sum_{F \in \mathscr{F}_T} (v, \nabla w \cdot \mathbf{n}_{TF})_F. \tag{11}$$

Specializing (11) to $w \in \mathbb{P}^{k+1}(T)$, we obtain

$$(\nabla \pi_T^{1,k+1} v, \nabla w)_T = -(\pi_T^{0,k-1} v, \triangle w)_T + \sum_{F \in \mathscr{F}_T} (\pi_F^{0,k} v, \nabla w \cdot \mathbf{n}_{TF})_F, \tag{12a}$$

where we have used (6) to insert $\pi_T^{1,k+1}$ into the left-hand side and (4) to insert $\pi_T^{0,k-1}$ and $\pi_F^{0,k}$ into the right-hand side after observing that $\triangle w \in \mathbb{P}^{k-1}(T) \subset \mathbb{P}^k(T)$ and $(\nabla w)_{|F} \cdot \mathbf{n}_{TF} \in \mathbb{P}^k(F)$ for all $F \in \mathscr{F}_T$. Moreover, recalling (6b) and using definition (4) of the L^2-projector, we infer that

$$(v - \pi_T^{0,0} v, 1)_T = (\pi_T^{1,k+1} v - \pi_T^{0,\max(0,k-1)} v, 1)_T = 0. \tag{12b}$$

The relations (12) show that computing the elliptic projection $\pi_T^{1,k+1}v$ does not require a full knowledge of the function v. All that is required is

(i) $\pi_T^{0,\max(0,k-1)}v$, the L^2-projection of v on the polynomial space $\mathbb{P}^{\max(0,k-1)}(T)$. Clearly, one could also choose $\pi_T^{0,k}v$ instead, which has the advantage of not requiring a special treatment of the case $k=0$;
(ii) for all $F \in \mathscr{F}_T$, $\pi_F^{0,k}v_{|F}$, the L^2-projection of the trace of v on F on the polynomial space $\mathbb{P}^k(F)$.

3.1.2 Local space of degrees of freedom

The remark at the end of the previous section motivates the introduction of the following space of DOFs (see Fig. 3):

$$\underline{U}_T^k := \mathbb{P}^k(T) \times \left(\bigtimes_{F\in\mathscr{F}_T} \mathbb{P}^k(F) \right). \tag{13}$$

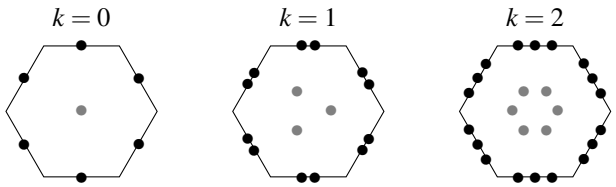


Fig. 3: DOFs in $\underline{U}_T^k$ for $k \in \{0,1,2\}$. Internal DOFs (in grey) can be eliminated by static condensation (see Section 3.2.4).

Observe that naming $\underline{U}_T^k$ space of DOFs involves a shortcut: the actual DOFs can be chosen in several equivalent ways (polynomial moments, point values, etc.), and the specific choice does not affect the following discussion. For a generic vector of DOFs in $\underline{U}_T^k$, we use the underlined notation $\underline{v}_T = (v_T, (v_F)_{F\in\mathscr{F}_T})$. On $\underline{U}_T^k$, we define the H^1-like seminorm $\|\cdot\|_{1,T}$ such that, for all $\underline{v}_T \in \underline{U}_T^k$,

$$\|\underline{v}_T\|_{1,T}^2 := \|\nabla v_T\|_T^2 + |\underline{v}_T|_{1,\partial T}^2, \qquad |\underline{v}_T|_{1,\partial T}^2 := \sum_{F\in\mathscr{F}_T} h_F^{-1}\|v_F - v_T\|_F^2, \tag{14}$$

where h_F denotes the diameter of F. The negative power of h_F in the second term ensures that both contributions have the same scaling. The DOFs corresponding to a smooth function $v \in W^{1,1}(T)$ are obtained via the reduction map $\underline{I}_T^k : W^{1,1}(T) \to \underline{U}_T^k$ such that

$$\underline{I}_T^k v := (\pi_T^{0,k}v, (\pi_F^{0,k}v_{|F})_{F\in\mathscr{F}_T}). \tag{15}$$

3.1.3 Potential reconstruction operator

Inspired by formula (12), we introduce the potential reconstruction operator $p_T^{k+1} : \underline{U}_T^k \to \mathbb{P}^{k+1}(T)$ such that, for all $\underline{v}_T \in \underline{U}_T^k$,

$$(\nabla p_T^{k+1}\underline{v}_T, \nabla w)_T = -(v_T, \triangle w)_T + \sum_{F\in\mathscr{F}_T} (v_F, \nabla w\cdot\mathbf{n}_{TF})_F \quad \forall w \in \mathbb{P}^{k+1}(T) \tag{16a}$$

and

$$(p_T^{k+1}\underline{v}_T - v_T, 1)_T = 0. \tag{16b}$$

Notice that $p_T^{k+1}\underline{v}_T$ is a polynomial function on T one degree higher than the element-based DOFs v_T. By definition, for all $v \in W^{1,1}(T)$ it holds that

$$(p_T^{k+1} \circ \underline{I}_T^k)v = \pi_T^{1,k+1} v, \tag{17}$$

i.e., the composition of the potential reconstruction operator with the reduction map gives the elliptic projector on $\mathbb{P}^{k+1}(T)$. An immediate consequence of (17) together with (7) is that $p_T^{k+1} \circ \underline{I}_T^k$ has optimal approximation properties in $\mathbb{P}^{k+1}(T)$.

3.1.4 Local contribution

We approximate the restriction $a_{|T} : H^1(T)\times H^1(T) \to \mathbb{R}$ to T of the continuous bilinear form a defined by (10) by the discrete bilinear form $\mathrm{a}_T : \underline{U}_T^k \times \underline{U}_T^k \to \mathbb{R}$ such that

$$\mathrm{a}_T(\underline{u}_T, \underline{v}_T) := (\nabla p_T^{k+1}\underline{u}_T, \nabla p_T^{k+1}\underline{v}_T)_T + \mathrm{s}_T(\underline{u}_T, \underline{v}_T), \tag{18}$$

where the first term in the right-hand side is the usual Galerkin contribution, while the second is a stabilization contribution for which we consider the following design conditions originally proposed in [5]:

Assumption 1 (Local stabilization bilinear form s_T) *The local stabilization bilinear form $\mathrm{s}_T : \underline{U}_T^k \times \underline{U}_T^k \to \mathbb{R}$ satisfies the following properties:*

(S1) Symmetry and positivity. *s_T is symmetric and positive semidefinite;*

(S2) Stability. *There is a real number $\eta > 0$ independent of h and of T, but possibly depending on d, ρ, and k, such that*

$$\eta^{-1}\|\underline{v}_T\|_{1,T}^2 \le \|\underline{v}_T\|_{\mathrm{a},T}^2 := \mathrm{a}_T(\underline{v}_T, \underline{v}_T) \le \eta\|\underline{v}_T\|_{1,T}^2 \qquad \forall \underline{v}_T \in \underline{U}_T^k; \tag{19}$$

(S3) Polynomial consistency. *For all $w \in \mathbb{P}^{k+1}(T)$ and all $\underline{v}_T \in \underline{U}_T^k$, it holds that*

$$\mathrm{s}_T(\underline{I}_T^k w, \underline{v}_T) = 0. \tag{20}$$

These requirements suggest that s_T can be obtained penalizing in a least square sense residuals that vanish for reductions of polynomial functions in $\mathbb{P}^{k+1}(T)$. Paradigmatic examples of such residuals are provided by the operators $\delta_T^k : \underline{U}_T^k \to$

$\mathbb{P}^k(T)$ and, for all $F \in \mathscr{F}_T$, $\delta_{TF}^k : \underline{U}_T^k \to \mathbb{P}^k(F)$ such that, for all $\underline{v}_T \in \underline{U}_T^k$,

$$\delta_T^k \underline{v}_T := \pi_T^{0,k}(p_T^{k+1}\underline{v}_T - v_T), \qquad \delta_{TF}^k \underline{v}_T := \pi_F^{0,k}(p_T^{k+1}\underline{v}_T - v_F) \quad \forall F \in \mathscr{F}_T. \tag{21}$$

To check that δ_T^k vanishes when $\underline{v}_T = \underline{I}_T^k w$ with $w \in \mathbb{P}^{k+1}(T)$, we observe that

$$\delta_T^k \underline{I}_T^k w = \pi_T^{0,k}(p_T^{k+1}\underline{I}_T^k w - \pi_T^{0,k} w) = \pi_T^{0,k}(\pi_T^{1,k+1} w - w) = \pi_T^{0,k}(w - w) = 0,$$

where we have used the definition of δ_T^k in the first equality, the relation (17) to replace $p_T^{k+1}\underline{I}_T^k$ by $\pi_T^{1,k+1}$ and the fact that $\pi_T^{0,k} w \in \mathbb{P}^k(T)$ to cancel $\pi_T^{0,k}$ from the second term in parentheses, and the fact that $\pi_T^{1,k+1}$ leaves polynomials of total degree up to $(k+1)$ unaltered as a projector to conclude. A similar argument shows that $\delta_{TF}^k \underline{I}_T^k w = 0$ for all $F \in \mathscr{F}_T$ whenever $w \in \mathbb{P}^{k+1}(T)$.

Accounting for dimensional homogeneity with the Galerkin term, one possible expression for s_T is thus

$$\mathrm{s}_T(\underline{u}_T, \underline{v}_T) := h_T^{-2}(\delta_T^k \underline{u}_T, \delta_T^k \underline{v}_T)_T + \sum_{F \in \mathscr{F}_T} h_F^{-1}(\delta_{TF}^k \underline{u}_T, \delta_{TF}^k \underline{v}_T)_F. \tag{22}$$

This choice, inspired by the Virtual Element literature [3], differs from the original HHO stabilization of [22], where the following expression is considered instead:

$$\mathrm{s}_T(\underline{u}_T, \underline{v}_T) := \sum_{F \in \mathscr{F}_T} h_F^{-1}(\delta_{TF}^k \underline{u}_T - \delta_T^k \underline{u}_T, \delta_{TF}^k \underline{v}_T - \delta_T^k \underline{v}_T)_F. \tag{23}$$

In this case, only quantities at faces are penalized. Both of the above expressions match the design conditions (S1)–(S3). A detailed proof for s_T as in (23) can be found in [22, Lemma 4]. Yet another example of stabilization bilinear form used in the context of HHO methods is provided by [1, Eq. (3.24)]. This expression results from the hybridization of the Mixed High-Order method of [20].

Remark 3 (Original HDG stabilization). The following stabilization bilinear form is used in the original Hybridizable Discontinuous Galerkin (HDG) method of [7, 11]:

$$\mathrm{s}_T(\underline{u}_T, \underline{v}_T) = \sum_{F \in \mathscr{F}_T} h_F^{-1}(u_F - u_T, v_F - v_T)_F.$$

While this choice obviously satisfies the properties (S1)-(S2), it fails to satisfy (S3) (it is only consistent for polynomials of degree up to k). As a result, up to one order of convergence is lost with respect to the estimates of Theorems 1 and 2 below. For a discussion including fixes that restore optimal orders of convergence in HDG see [10].

3.1.5 Consistency properties of the stabilization for smooth functions

In the following proposition we study the consistency properties of s_T when its arguments are reductions of a smooth function. We give a detailed proof since this

result is a new extension of the bound in [22, Theorem 8] (see, in particular, Eq. (45) therein) to more general stabilization bilinear forms.

Proposition 1 (Consistency of s_T). *Let s_T denote a stabilization bilinear form satisfying assumptions (S1)–(S3). Then, there is a real number $C>0$ independent of h, but possibly depending on d, ρ, and k, such that, for all $T\in\mathscr{T}_h$ and all $v\in H^{k+2}(T)$, it holds that*

$$\mathrm{s}_T(\underline{I}_T^k v,\underline{I}_T^k v)^{1/2}\le Ch_T^{k+1}\|v\|_{H^{k+2}(T)}. \tag{24}$$

Proof. We set, for the sake of brevity, $\check{v}_T:=\pi_T^{1,k+1}v$ and abridge as $A\lesssim B$ the inequality $A\le cB$ with multiplicative constant $c>0$ having the same dependencies as C in (24). Using (S2) and (S3) we infer that

$$\mathrm{s}_T(\underline{I}_T^k v,\underline{I}_T^k v)^{1/2}=\mathrm{s}_T(\underline{I}_T^k(v-\check{v}_T),\underline{I}_T^k(v-\check{v}_T))^{1/2}\le\eta\|\underline{I}_T^k(v-\check{v}_T)\|_{1,T}. \tag{25}$$

Recalling (14), we have that

$$\begin{aligned}\|\underline{I}_T^k(v-\check{v}_T)\|_{1,T}^2=&\\ &\|\nabla\pi_T^{0,k}(v-\check{v}_T)\|_T^2+\sum_{F\in\mathscr{F}_T}h_F^{-1}\|\pi_F^{0,k}(v-\check{v}_T-\pi_T^{0,k}(v-\check{v}_T))\|_F^2.\end{aligned} \tag{26}$$

Using the $H^1(T)$-boundedness of $\pi_T^{0,k}$ resulting from (5) with $l=k$, and $s=1$ followed by the optimal approximation properties (7a) of $\check{v}_T$ (with $\alpha=1$, $l=k+1$, $s=k+2$, and $m=1$), it is inferred that

$$\|\nabla\pi_T^{0,k}(v-\check{v}_T)\|_T\lesssim\|\nabla(v-\check{v}_T)\|_T\lesssim h^{k+1}\|v\|_{H^{k+2}(T)}. \tag{27}$$

On the other hand, for all $F\in\mathscr{F}_T$ it holds that

$$\begin{aligned}h_F^{-1/2}\|\pi_F^{0,k}(v-\check{v}_T-\pi_T^{0,k}(v-\check{v}_T))\|_F&\lesssim h_T^{-1}\|v-\check{v}_T-\pi_T^{0,k}(v-\check{v}_T)\|_T\\ &\lesssim\|\nabla(v-\check{v}_T)\|_T\\ &\lesssim h_T^{k+1}\|v\|_{H^{k+2}(T)},\end{aligned} \tag{28}$$

where we have used the $L^2(F)$-boundedness of $\pi_F^{0,k}$ together with (2) and the discrete trace inequality (3) in the first line, a local Poincaré inequality resulting from (7a) with $\alpha=0$, $l=k$, $s=1$, and $m=0$ to pass to the second line, and the optimal approximation properties of $\check{v}_T$ expressed by (7a) with $\alpha=1$, $l=k$, $s=k+2$, and $m=1$ to conclude. Plugging (27) and (28) into (26), recalling that $\mathrm{card}(\mathscr{F}_T)\lesssim 1$ (see (1)), and using the resulting bound to estimate (25), (24) follows. □

3.2 Discrete problem

We now show how to formulate the discrete problem from the local contributions introduced in the previous section.

3.2.1 Global spaces of degrees of freedom

We define the following global space of DOFs with single-valued interface unknowns:

$$\underline{U}_h^k := \left(\bigtimes_{T\in\mathscr{T}_h} \mathbb{P}^k(T) \right) \times \left(\bigtimes_{F\in\mathscr{F}_h} \mathbb{P}^k(F) \right).$$

Notice that single-valued means here that interface values match from one element to the adjacent one. For a generic element $\underline{v}_h \in \underline{U}_h^k$, we use the underlined notation $\underline{v}_h = ((v_T)_{T\in\mathscr{T}_h}, (v_F)_{F\in\mathscr{F}_h})$ and, for all $T \in \mathscr{T}_h$, we denote by $\underline{v}_T = (v_T, (v_F)_{F\in\mathscr{F}_T}) \in \underline{U}_T^k$ its restriction to T. We also define the broken polynomial function $v_h \in \mathbb{P}^k(\mathscr{T}_h)$ such that

$$v_{h|T} := v_T \qquad \forall T \in \mathscr{T}_h.$$

The DOFs corresponding to a smooth function $v \in W^{1,1}(\Omega)$ are obtained via the reduction map $\underline{I}_h^k : W^{1,1}(\Omega) \to \underline{U}_h^k$ such that

$$\underline{I}_h^k v := ((\pi_T^{0,k} v_{|T})_{T\in\mathscr{T}_h}, (\pi_F^{0,k} v_{|F})_{F\in\mathscr{F}_h}).$$

We define on $\underline{U}_h^k$ the seminorm $\|\cdot\|_{1,h}$ such that, for all $\underline{v}_h \in \underline{U}_h^k$,

$$\|\underline{v}_h\|_{1,h}^2 := \sum_{T\in\mathscr{T}_h} \|\underline{v}_T\|_{1,T}^2,$$

with local seminorm $\|\cdot\|_{1,T}$ defined by (14). To account for the homogeneous Dirichlet boundary condition (8b) in a strong manner, we introduce the subspace

$$\underline{U}_{h,0}^k := \left\{ \underline{v}_h \in \underline{U}_h^k \mid v_F \equiv 0 \quad \forall F \in \mathscr{F}_h^{\mathrm{b}} \right\}.$$

We recall the following discrete Poincaré inequality proved in [13, Proposition 5.4]: There exists a real number $C_{\mathrm{P}} > 0$ independent of h, but possibly depending on Ω, ρ, and k, such that, for all $\underline{v}_h \in \underline{U}_{h,0}^k$,

$$\|v_h\| \le C_{\mathrm{P}} \|\underline{v}_h\|_{1,h}. \tag{29}$$

Proposition 2 (Norm $\|\cdot\|_{1,h}$). *The map $\|\cdot\|_{1,h}$ defines a norm on $\underline{U}_{h,0}^k$.*

Proof. The seminorm property being evident, it suffices to prove that, for all $\underline{v}_h \in \underline{U}_{h,0}^k$, $\|\underline{v}_h\|_{1,h} = 0 \implies \underline{v}_h = \underline{0}_h$. Let $\underline{v}_h \in \underline{U}_{h,0}^k$ be such that $\|\underline{v}_h\|_{1,h} = 0$. By (29), we have $\|v_h\| = 0$, hence $v_T \equiv 0$ for all $T \in \mathscr{T}_h$. From the definition (14) of the norm $\|\cdot\|_{1,T}$, we also have that $\|v_F - v_T\|_F = 0$ for all $T \in \mathscr{T}_h$ and all $F \in \mathscr{F}_T$, hence $v_F = v_T \equiv 0$. Since any mesh face belongs to the set $\mathscr{F}_T$ for at least one mesh element $T \in \mathscr{T}_h$, this concludes the proof. □

3.2.2 Global bilinear form

We define the global bilinear forms $\mathrm{a}_h : \underline{U}_h^k \times \underline{U}_h^k \to \mathbb{R}$ and $\mathrm{s}_h : \underline{U}_h^k \times \underline{U}_h^k \to \mathbb{R}$ by element-by-element assembly setting, for all $\underline{u}_h, \underline{v}_h \in \underline{U}_h^k$,

$$\mathrm{a}_h(\underline{u}_h, \underline{v}_h) := \sum_{T\in\mathscr{T}_h} \mathrm{a}_T(\underline{u}_T, \underline{v}_T), \qquad \mathrm{s}_h(\underline{u}_h, \underline{v}_h) := \sum_{T\in\mathscr{T}_h} \mathrm{s}_T(\underline{u}_T, \underline{v}_T). \tag{30}$$

Lemma 1 (Properties of a_h). *The bilinear form a_h enjoys the following properties:*

(i) Stability. *For all $\underline{v}_h \in \underline{U}_{h,0}^k$ it holds with η as in* (19) *that*

$$\eta^{-1}\|\underline{v}_h\|_{1,h}^2 \le \|\underline{v}_h\|_{\mathrm{a},h}^2 := \mathrm{a}_h(\underline{v}_h, \underline{v}_h) \le \eta\|\underline{v}_h\|_{1,h}^2. \tag{31}$$

(ii) Consistency. *There is a real number $C > 0$ independent of h, but possibly depending on d, ρ, and k, such that, for all $w \in H_0^1(\Omega)\cap H^{k+2}(\Omega)$,*

$$\sup_{\underline{v}_h\in\underline{U}_{h,0}^k, \|\underline{v}_h\|_{1,h}=1} \mathscr{E}_h(w;\underline{v}_h) \le Ch^{k+1}\|w\|_{H^{k+2}(\Omega)}, \tag{32}$$

with linear form $\mathscr{E}_h(w;\cdot) : \underline{U}_h^k \to \mathbb{R}$ representing the conformity error such that, for all $\underline{v}_h \in \underline{U}_h^k$,

$$\mathscr{E}_h(w;\underline{v}_h) := -(\triangle w, v_h) - \mathrm{a}_h(\underline{I}_h^k w, \underline{v}_h). \tag{33}$$

Proof. (i) *Stability.* Summing inequalities (19) over $T \in \mathscr{T}_h$, (31) follows.

(ii) *Consistency.* Let $\underline{v}_h \in \underline{U}_{h,0}^k$ be such that $\|\underline{v}_h\|_{1,h} = 1$. Throughout the proof, we abridge as $A \lesssim B$ the inequality $A \le cB$ with multiplicative constant $c > 0$ having the same dependecies as C in (32). For the sake of brevity, we also let $\check{w}_T := p_T^{k+1}\underline{I}_T^k w = \pi_T^{1,k+1} w$ (cf. (17)) for all $T \in \mathscr{T}_h$. Integrating by parts element-by-element, we infer that

$$-(\triangle w, v_h) = \sum_{T\in\mathscr{T}_h}\left((\nabla w, \nabla v_T)_T + \sum_{F\in\mathscr{F}_T}(\nabla w\cdot\mathbf{n}_{TF}, v_F - v_T)_F\right). \tag{34}$$

To insert v_F into the second term in parentheses in (34), we have used the fact that $v_F \equiv 0$ for all $F \in \mathscr{F}_h^{\mathrm{b}}$ while, for all $F \in \mathscr{F}_h^{\mathrm{i}}$ such that $F \subset \partial T_1 \cap \partial T_2$ for distinct mesh elements $T_1, T_2 \in \mathscr{T}_h$, $(\nabla w)_{|T_1}\cdot\mathbf{n}_{T_1F} + (\nabla w)_{|T_2}\cdot\mathbf{n}_{T_2F} = 0$ (since $w \in H^{k+2}(\Omega)$), so that

$$\sum_{T\in\mathscr{T}_h}\sum_{F\in\mathscr{F}_T}(\nabla w\cdot\mathbf{n}_{TF}, v_F)_F = \sum_{F\in\mathscr{F}_h^{\mathrm{i}}}\Big(\sum_{T\in\mathscr{T}_T}(\nabla w)_{|T}\cdot\mathbf{n}_{TF}, v_F\Big)_F + \sum_{F\in\mathscr{F}_h^{\mathrm{b}}}(\nabla w\cdot\mathbf{n}, v_F)_F = 0.$$

On the other hand, plugging the definition (18) of a_T into (30), and expanding $p_T^{k+1}\underline{v}_T$ according to (16) with $w = \check{w}_T$, it is inferred that

$$\mathrm{a}_h(\underline{I}_h^k w, \underline{v}_h) = \sum_{T\in\mathscr{T}_h}\left((\nabla \check{w}_T, \nabla v_T)_T + \sum_{F\in\mathscr{F}_T}(\nabla \check{w}_T\cdot\mathbf{n}_{TF}, v_F - v_T)_F + \mathrm{s}_T(\underline{I}_T^k w, \underline{v}_T)\right). \tag{35}$$

Subtracting (35) from (34), using the definition (6) of $\pi_T^{1,k+1}$ to cancel the first terms in parentheses, and taking absolute values, we get

$$\begin{aligned}
|\mathscr{E}_h(w;\underline{v}_h)| &= \left|\sum_{T\in\mathscr{T}_h}\left(\sum_{F\in\mathscr{F}_T}(\nabla(w-\check{w}_T)\cdot\mathbf{n}_{TF}, v_F - v_T)_F + \mathrm{s}_T(\underline{I}_T^k w, \underline{v}_T)\right)\right| \\
&\le \left[\sum_{T\in\mathscr{T}_h}\left(h_T\|\nabla(w-\check{w}_T)\|_{\partial T}^2 + \mathrm{s}_T(\underline{I}_T^k w, \underline{I}_T^k w)\right)\right]^{1/2} \\
&\quad\times\left[\sum_{T\in\mathscr{T}_h}\left(|\underline{v}_T|_{1,\partial T}^2 + \mathrm{s}_T(\underline{v}_T,\underline{v}_T)\right)\right]^{1/2}.
\end{aligned}$$

Using (7b) with $\alpha = 1$, $l = k+1$, $s = k+2$, and $m = 1$ together with (24) for the first factor, and the seminorm equivalence (19) together with the fact that $\|\underline{v}_h\|_{1,h} = 1$ for the second, we infer the bound

$$|\mathscr{E}_h(w;\underline{v}_h)| \lesssim h^{k+1}\|w\|_{H^{k+2}(\Omega)}.$$

Since $\underline{v}_h$ is arbitrary, this yields (32). □

3.2.3 Discrete problem and well-posedness

The discrete problem reads: Find $\underline{u}_h \in \underline{U}_{h,0}^k$ such that

$$\mathrm{a}_h(\underline{u}_h, \underline{v}_h) = (f, v_h) \qquad \forall \underline{v}_h \in \underline{U}_{h,0}^k. \tag{36}$$

Lemma 2 (Well-posedness). *Problem* (36) *is well-posed, and we have the following a priori bound for the unique discrete solution* $\underline{u}_h \in \underline{U}_{h,0}^k$*:*

$$\|\underline{u}_h\|_{1,h} \le \eta C_{\mathrm{P}}\|f\|.$$

Proof. We check the assumptions of the Lax–Milgram lemma [33]. Clearly, $\underline{U}_{h,0}^k$ equipped with the norm $\|\cdot\|_{1,h}$ is a Hilbert space. The bilinear form a_h is coercive and continuous owing to (31) with coercivity constant equal to η^{-1}. The linear form $\underline{v}_h \mapsto (f, v_h)$ is continuous owing to (29) with continuity constant equal to C_{P}. □

3.2.4 Implementation

Let a basis $\mathscr{B}_h$ for the space $\underline{U}_{h,0}^k$ be fixed such that every basis function is supported by only one mesh element or face. For a generic element $\underline{v}_h \in \underline{U}_{h,0}^k$, denote by V_h

the corresponding vector of coefficients in $\mathscr{B}_h$ partitioned as

$$\mathsf{V}_h = \left[\begin{array}{c} \mathsf{V}_{\mathscr{T}_h} \\ \hdashline \mathsf{V}_{\mathscr{F}_h} \end{array}\right],$$

where the subvectors $\mathsf{V}_{\mathscr{T}_h}$ and $\mathsf{V}_{\mathscr{F}_h}$ collect the coefficients associated to element-based and face-based DOFs, respectively. Denote by A_h the matrix representation of the bilinear form a_h and by B_h the vector representation of the linear form $\underline{v}_h \mapsto (f, v_h)$, both partitioned in a similar way. The algebraic problem corresponding to (36) reads

$$\underbrace{\left[\begin{array}{c:c} \mathsf{A}_{\mathscr{T}_h\mathscr{T}_h} & \mathsf{A}_{\mathscr{T}_h\mathscr{F}_h} \\ \hdashline \mathsf{A}_{\mathscr{T}_h\mathscr{F}_h}^{\mathrm{T}} & \mathsf{A}_{\mathscr{F}_h\mathscr{F}_h} \end{array}\right]}_{\mathsf{A}_h} \underbrace{\left[\begin{array}{c} \mathsf{U}_{\mathscr{T}_h} \\ \hdashline \mathsf{U}_{\mathscr{F}_h} \end{array}\right]}_{\mathsf{U}_h} = \underbrace{\left[\begin{array}{c} \mathsf{B}_{\mathscr{T}_h} \\ \hdashline 0_{\mathscr{F}_h} \end{array}\right]}_{\mathsf{B}_h}. \tag{37}$$

The submatrix $\mathsf{A}_{\mathscr{T}_h\mathscr{T}_h}$ is block-diagonal and symmetric positive definite, and is therefore inexpensive to invert. In the practical implementation, this remark can be exploited by solving the linear system (37) in two steps (see, e.g., [10, Section 2.4]):

(i) First, element-based coefficients in $\mathsf{U}_{\mathscr{T}_h}$ are expressed in terms of $\mathsf{B}_{\mathscr{T}_h}$ and $\mathsf{U}_{\mathscr{F}_h}$ by the inexpensive solution of the first block equation:

$$\mathsf{U}_{\mathscr{T}_h} = \mathsf{A}_{\mathscr{T}_h\mathscr{T}_h}^{-1}\left(\mathsf{B}_{\mathscr{T}_h} - \mathsf{A}_{\mathscr{T}_h\mathscr{F}_h}\mathsf{U}_{\mathscr{F}_h}\right). \tag{38a}$$

This step is referred to as *static condensation* in the Finite Element literature;

(ii) Second, face-based coefficients in $\mathsf{U}_{\mathscr{F}_h}$ are obtained solving the global skeletal (i.e., involving unknowns attached to the mesh skeleton) problem

$$\left(\mathsf{A}_{\mathscr{F}_h\mathscr{F}_h} - \mathsf{A}_{\mathscr{T}_h\mathscr{F}_h}^{\mathrm{T}}\mathsf{A}_{\mathscr{T}_h\mathscr{T}_h}^{-1}\mathsf{A}_{\mathscr{T}_h\mathscr{F}_h}\right)\mathsf{U}_{\mathscr{F}_h} = \mathsf{A}_{\mathscr{T}_h\mathscr{F}_h}^{\mathrm{T}}\mathsf{A}_{\mathscr{T}_h\mathscr{T}_h}^{-1}\mathsf{B}_{\mathscr{T}_h}. \tag{38b}$$

This computationally more intensive step requires to invert the matrix in parentheses in the above expression. This symmetric positive definite matrix, whose stencil is the same as that of $\mathsf{A}_{\mathscr{F}_h\mathscr{F}_h}$ and only involves neighbours through faces, has size $N_{\mathrm{dof}} \times N_{\mathrm{dof}}$ with

$$N_{\mathrm{dof}} = \mathrm{card}(\mathscr{F}_h^{\mathrm{i}}) \times \binom{k+d-1}{k}. \tag{38c}$$

3.2.5 Local conservation and flux continuity

At the continuous level, the solution of problem (9) satisfies the following local balance for all $T \in \mathscr{T}_h$ and all $v_T \in \mathbb{P}^k(T)$:

$$(\nabla u, \nabla v_T)_T - \sum_{F\in\mathscr{F}_T} (\nabla u\cdot\mathbf{n}_{TF}, v_T)_F = (f, v_T)_T, \tag{39a}$$

and the normal flux traces are continuous in the sense that, for all $F \in \mathscr{F}_h^{\mathrm{i}}$ such that $F \subset \partial T_1 \cap \partial T_2$ with distinct mesh elements $T_1, T_2 \in \mathscr{T}_h$, it holds (see, e.g., [17, Lemma 4.3])

$$(\nabla u)_{|T_1} \cdot \mathbf{n}_{T_1 F} + (\nabla u)_{|T_2} \cdot \mathbf{n}_{T_2 F} = 0. \tag{39b}$$

We show in this section that a discrete counterpart of the relations (39) holds for the discrete solution. This property is relevant both from the engineering and mathematical points of view, and it can be exploited to derive a posteriori error estimators by flux equilibration. It was originally highlighted in [18] and, using different techniques, in [10] for the stabilization bilinear form s_T defined by (23). Here, using yet a different approach, we extend these results to more general stabilization bilinear forms.

Let a mesh element $T \in \mathscr{T}_h$ be fixed. We define the space

$$\underline{D}_{\partial T}^k := \bigtimes_{F \in \mathscr{F}_T} \mathbb{P}^k(F), \tag{40}$$

as well as the boundary difference operator $\underline{\Delta}_{\partial T}^k : \underline{U}_T^k \to \underline{D}_{\partial T}^k$ such that, for all $\underline{v}_T \in \underline{U}_T^k$,

$$\underline{\Delta}_{\partial T}^k \underline{v}_T = (\Delta_{TF}^k \underline{v}_T)_{F \in \mathscr{F}_T} := (v_F - v_{T|F})_{F \in \mathscr{F}_T}. \tag{41}$$

A useful remark is that, for all $\underline{v}_T \in \underline{U}_T^k$, it holds

$$\underline{v}_T - \underline{I}_T^k v_T = (v_T - \pi_T^{0,k} v_T, (v_F - \pi_F^{0,k} v_{T|F})_{F \in \mathscr{F}_T}) = (0, \underline{\Delta}_{\partial T}^k \underline{v}_T), \tag{42}$$

where the conclusion follows observing that, for all $T \in \mathscr{T}_h$ and all $F \in \mathscr{F}_T$, $\pi_T^{0,k} v_T = v_T$ and $\pi_F^{0,k} v_{T|F} = v_{T|F}$ since $v_T \in \mathbb{P}^k(T)$ and $v_{T|F} \in \mathbb{P}^k(F)$.

We show in the next proposition that any stabilization bilinear form with a suitable dependence on its arguments can be reformulated in terms of boundary differences.

Proposition 3 (Reformulation of the stabilization bilinear form). *Let $T \in \mathscr{T}_h$, and assume that s_T is a stabilization bilinear form that satisfies assumptions (S1)–(S3) and that depends on its arguments only through the residuals defined by* (21)*. Then, it holds for all $\underline{u}_T, \underline{v}_T \in \underline{U}_T^k$ that*

$$\mathrm{s}_T(\underline{u}_T, \underline{v}_T) = \mathrm{s}_T((0, \underline{\Delta}_{\partial T}^k \underline{u}_T), (0, \underline{\Delta}_{\partial T}^k \underline{v}_T)). \tag{43}$$

Proof. It suffices to show that, for all $\underline{v}_T \in \underline{U}_T^k$,

$$\delta_T^k \underline{v}_T = \delta_T^k(0, \underline{\Delta}_{\partial T}^k \underline{v}_T), \qquad \delta_{TF}^k \underline{v}_T = \delta_{TF}^k(0, \underline{\Delta}_{\partial T}^k \underline{v}_T) \quad \forall F \in \mathscr{F}_T.$$

Let us start by δ_T^k. Since $v_T \in \mathbb{P}^k(T)$, $p_T^{k+1} \underline{I}_T^k v_T = \pi_T^{1,k+1} v_T = v_T$. Hence,

$$\begin{aligned}\delta_T^k \underline{v}_T &= \pi_T^{0,k}(p_T^{k+1}\underline{v}_T - v_T)\\ &= \pi_T^{0,k}(p_T^{k+1}\underline{v}_T - p_T^{k+1}\underline{I}_T^k v_T)\\ &= \pi_T^{0,k}p_T^{k+1}(\underline{v}_T - \underline{I}_T^k v_T) = \delta_T^k(0,\underline{\Delta}_{\partial T}^k\underline{v}_T),\end{aligned}$$

where we have used the linearity of p_T^{k+1} to pass to the third line and (42) to conclude. Let now $F \in \mathscr{F}_T$ and consider δ_{TF}^k. We have

$$\begin{aligned}\delta_{TF}^k \underline{v}_T &= \pi_F^{0,k}(p_T^{k+1}\underline{v}_T - v_F)\\ &= \pi_F^{0,k}(p_T^{k+1}\underline{v}_T - p_T^{k+1}\underline{I}_T^k v_T + v_T - v_F)\\ &= \pi_F^{0,k}(p_T^{k+1}(0,\underline{\Delta}_{\partial T}^k\underline{v}_T) - \Delta_{TF}^k\underline{v}_T) = \delta_{TF}^k(0,\underline{\Delta}_{\partial T}^k\underline{v}_T),\end{aligned}$$

where we have introduced $v_T - p_T^{k+1}\underline{I}_T^k v_T = 0$ in the second line, used the linearity of p_T^{k+1} together with (42) and the definition (40) of $\underline{\Delta}_{\partial T}^k$ in the third line, and concluded recalling the definition (21) of δ_{TF}^k. □

Define the boundary residual operator $\underline{R}_{\partial T}^k : \underline{U}_T^k \to \underline{D}_{\partial T}^k$ such that, for all $\underline{v}_T \in \underline{U}_T^k$, $\underline{R}_{\partial T}^k\underline{v}_T = (R_{TF}^k\underline{v}_T)_{F\in\mathscr{F}_T}$ satisfies for all $\underline{\alpha}_{\partial T} = (\alpha_{TF})_{F\in\mathscr{F}_T} \in \underline{D}_{\partial T}^k$

$$-\sum_{F\in\mathscr{F}_T}(R_{TF}^k\underline{v}_T,\alpha_{TF})_F = \mathrm{s}_T((0,\underline{\Delta}_{\partial T}^k\underline{v}_T),(0,\underline{\alpha}_{\partial T})). \tag{44}$$

Problem (44) is well-posed, and computing $R_{TF}^k\underline{v}_T$ requires to invert the boundary mass matrix.

Lemma 3 (Local balance and flux continuity). *Under the assumptions of Proposition 3, denote by $\underline{u}_h \in \underline{U}_{h,0}^k$ the unique solution of problem* (36) *and, for all $T \in \mathscr{T}_h$ and all $F \in \mathscr{F}_T$, define the numerical trace of the flux*

$$S_{TF}(\underline{u}_T) := -\nabla p_T^{k+1}\underline{u}_T\cdot\mathbf{n}_{TF} + R_{TF}^k\underline{u}_T$$

with R_{RF}^k defined by (44)*. Then, for all $T \in \mathscr{T}_h$ we have the following discrete counterpart of the local balance* (39a)*: For all $v_T \in \mathbb{P}^k(T)$,*

$$(\nabla p_T^{k+1}\underline{u}_T,\nabla v_T)_T + \sum_{F\in\mathscr{F}_T}(S_{TF}(\underline{u}_T),v_T)_F = (f,v_T)_T, \tag{45a}$$

and, for any interface $F \in \mathscr{F}_h^{\mathrm{i}}$ such that $F \subset \partial T_1 \cap \partial T_2$ with distinct mesh elements $T_1, T_2 \in \mathscr{T}_h$, the numerical fluxes are continuous in the sense that (compare with (39b)*):*

$$S_{T_1F}(\underline{u}_{T_1}) + S_{T_2F}(\underline{u}_{T_2}) = 0. \tag{45b}$$

Proof. Let $\underline{v}_h \in \underline{U}_{h,0}^k$. Plugging the definition (18) of a_T into (30), using for all $T \in \mathscr{T}_h$ the definition of $p_T^{k+1}\underline{v}_T$ with $w = p_T^{k+1}\underline{u}_T$, and recalling the reformulation (43) of s_T together with the definition (44) of $\underline{R}_{\partial T}^k$ to write

$$\mathrm{s}_T(\underline{u}_T, \underline{v}_T) = -\sum_{F \in \mathscr{F}_T} (R_{TF}^k \underline{u}_T, v_F - v_T)_F \qquad \forall T \in \mathscr{T}_h, \tag{46}$$

we infer from the discrete problem (36) that

$$\sum_{T \in \mathscr{T}_h} \left((\nabla p_T^{k+1}\underline{u}_T, \nabla v_T)_T + \sum_{F \in \mathscr{F}_T} (\nabla p_T^{k+1}\underline{u}_T \cdot \mathbf{n}_{TF} - R_{TF}^k \underline{u}_T, v_F - v_T)_F \right) = (f, v_h).$$

Selecting $\underline{v}_h$ such that v_T spans $\mathbb{P}^k(T)$ for a selected mesh element $T \in \mathscr{T}_h$ while $v_{T'} \equiv 0$ for all $T' \in \mathscr{T}_h \setminus \{T\}$ and $v_F \equiv 0$ for all $F \in \mathscr{F}_h$, we obtain (45a). On the other hand, selecting $\underline{v}_h$ such that $v_T \equiv 0$ for all $T \in \mathscr{T}_h$, v_F spans $\mathbb{P}^k(F)$ for a selected interface $F \in \mathscr{F}_h^{\mathrm{i}}$ such that $F \subset \partial T_1 \cap \partial T_2$ for distinct mesh elements $T_1, T_2 \in \mathscr{T}_h$, and $v_{F'} \equiv 0$ for all $F' \in \mathscr{F}_h \setminus \{F\}$ yields (45b).

Remark 4 (Interpretation of the discrete problem). Lemma 3 and its proof provide further insight into the structure of the discrete problem (36), which consists of the local balances (45a) (corresponding to the local block equations (38a)) and a global transmission condition enforcing the continuity (45b) of numerical fluxes (corresponding to the global skeletal problem (38b)).

3.3 A priori error analysis

Having proved that the discrete problem (36) is well-posed, it remains to determine the convergence of the discrete solution towards the exact solution, which is precisely the goal of this section.

3.3.1 Energy error estimate

We start by deriving a basic convergence results. The error is measured as the difference between the exact solution and the global reconstruction obtained from the discrete solution through the operator $p_h^{k+1} : \underline{U}_h^k \to \mathbb{P}^{k+1}(\mathscr{T}_h)$ such that, for all $\underline{v}_h \in \underline{U}_h^k$,

$$(p_h^{k+1}\underline{v}_h)_{|T} := p_T^{k+1}\underline{v}_T \qquad \forall T \in \mathscr{T}_h. \tag{47}$$

Theorem 1 (Energy error estimate). *Let a polynomial degree $k \geq 0$ be fixed. Let $u \in H_0^1(\Omega)$ denote the unique solution to* (9)*, for which we assume the*

additional regularity $u \in H^{k+2}(\Omega)$. Let $\underline{u}_h \in \underline{U}_{h,0}^k$ denote the unique solution to (36) *with stabilization bilinear form s_T in* (18) *satisfying assumptions (S1)–(S3) for all $T \in \mathcal{T}_h$. Then, there exists a real number $C > 0$ independent of h, but possibly depending on d, ρ, and k, such that*

$$\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\| + |\underline{u}_h|_{\mathrm{s},h} \le Ch^{k+1}\|u\|_{H^{k+2}(\Omega)}, \tag{48}$$

where $|\cdot|_{\mathrm{s},h}$ is the seminorm defined by the bilinear form s_h on $\underline{U}_h^k$.

Proof. Let, for the sake of brevity, $\hat{\underline{u}}_h := \underline{I}_h^k u$ and $\check{u}_h := p_h^{k+1}\hat{\underline{u}}_h$. We abridge as $A \lesssim B$ the inequality $A \le cB$ with multiplicative constant $c > 0$ having the same dependencies as C in (48). Using the triangle and Cauchy–Schwarz inequalities, it is readily inferred that

$$\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\| + |\underline{u}_h|_{\mathrm{s},h} \le \underbrace{\|\underline{u}_h - \hat{\underline{u}}_h\|_{\mathrm{a},h}}_{\mathfrak{T}_1} + \underbrace{\Big(\|\nabla_h(\check{u}_h - u)\|^2 + |\hat{\underline{u}}_h|_{\mathrm{s},h}^2\Big)^{1/2}}_{\mathfrak{T}_2}. \tag{49}$$

We have that

$$\begin{aligned}\mathfrak{T}_1^2 &= \mathrm{a}_h(\underline{u}_h, \underline{u}_h - \hat{\underline{u}}_h) - \mathrm{a}_h(\hat{\underline{u}}_h, \underline{u}_h - \hat{\underline{u}}_h)\\ &= (f, u_h - \hat{u}_h) - \mathrm{a}_h(\hat{\underline{u}}_h, \underline{u}_h - \hat{\underline{u}}_h) = \mathscr{E}_h(u; \underline{u}_h - \hat{\underline{u}}_h),\end{aligned}$$

where we have used the definition (31) of the $\|\cdot\|_{\mathrm{a},h}$-norm together with the linearity of a_h in its first argument in the first line, the discrete problem (36) to pass to the second line, and the definition (33) of the conformity error to conclude. As a consequence, assuming $\underline{u}_h \neq \hat{\underline{u}}_h$ (the other case is trivial), we have that

$$|\mathfrak{T}_1| \le \mathscr{E}_h\left(u; \frac{\underline{u}_h - \hat{\underline{u}}_h}{\|\underline{u}_h - \hat{\underline{u}}_h\|_{\mathrm{a},h}}\right) \le \eta^{1/2}\mathscr{E}_h\left(u; \frac{\underline{u}_h - \hat{\underline{u}}_h}{\|\underline{u}_h - \hat{\underline{u}}_h\|_{1,h}}\right) \le \eta^{1/2} \sup_{\underline{v}_h \in \underline{U}_{h,0}^k, \|\underline{v}_h\|_{1,h}=1} \mathscr{E}_h(u; \underline{v}_h),$$

where we have used the linearity of $\mathscr{E}_h(u;\cdot)$, the first bound in (31), and a passage to the supremum to conclude. Recalling (32), we arrive at

$$|\mathfrak{T}_1| \lesssim h^{k+1}\|u\|_{H^{k+2}(\Omega)}. \tag{50}$$

On the other hand, using the approximation properties (7a) of $\check{u}_T$ with $\alpha = 1$, $l = k+1$, $s = k+2$, and $m = 1$ together with the approximation properties (24) of s_T, it is inferred for the second term

$$|\mathfrak{T}_2| \lesssim h^{k+1}\|u\|_{H^{k+2}(\Omega)}. \tag{51}$$

Using (50) and (51) to bound the right-hand side of (49), (48) follows. □

3.3.2 Convergence of the jumps

Functions in $H^1(\mathscr{T}_h)$ are in $H^1_0(\Omega)$ if their jumps vanish a.e. at interfaces and their trace is zero a.e. on $\partial\Omega$; see, e.g., [17, Lemma 1.23]. Thus, a measure of the nonconformity is provided by the jump seminorm $|\cdot|_{\mathrm{J},h}$ such that, for all $v \in H^1(\mathscr{T}_h)$,

$$|v|^2_{\mathrm{J},h} := \sum_{F\in\mathscr{F}_h} h_F^{-1}\|\pi_F^{0,k}[v]_F\|^2_F, \tag{52}$$

where $[\cdot]_F$ denotes the usual jump operator such that, for all faces $F \in \mathscr{F}_h$ and all functions $v : \bigcup_{T\in\mathscr{T}_F} T \to \mathbb{R}$ smooth enough,

$$[v]_F := \begin{cases} v_{|T_1} - v_{|T_2} & \forall F \in \mathscr{F}_{T_1}\cap\mathscr{F}_{T_2}, \\ v & \forall F \in \mathscr{F}_h^{\mathrm{b}}. \end{cases} \tag{53}$$

A natural question is whether the jump seminorm of $p_h^{k+1}\underline{u}_h$ convergences to zero. The answer is provided by the following lemma.

Lemma 4 (Convergence of the jumps). *Under the assumptions and notations of Theorem 1, and further supposing, for the sake of simplicity, that the local stabilization bilinear form* s_T *is given by* (22)*, there is a real number* $C > 0$ *independent of* h*, but possibly depending on* d*,* ρ*, and* k*, such that*

$$|p_h^{k+1}\underline{u}_h|_{\mathrm{J},h} \le Ch^{k+1}\|u\|_{H^{k+2}(\Omega)}. \tag{54}$$

Proof. Inserting u_F inside the jump and using the triangle inequality for every interface $F \in \mathscr{F}_h^{\mathrm{i}}$, and recalling that $v_F = 0$ on every boundary face $F \in \mathscr{F}_h^{\mathrm{b}}$, it is inferred that

$$\begin{aligned}\sum_{F\in\mathscr{F}_h} h_F^{-1}\|\pi_F^{0,k}[p_h^{k+1}\underline{u}_h]_F\|^2_F &\le 2\sum_{F\in\mathscr{F}_h}\sum_{T\in\mathscr{T}_F} h_F^{-1}\|\pi_F^{0,k}(p_T^{k+1}\underline{u}_T - u_F)\|^2_F \\ &\le 2\sum_{T\in\mathscr{T}_h}\sum_{F\in\mathscr{F}_T} h_F^{-1}\|\pi_F^{0,k}(p_T^{k+1}\underline{u}_T - u_F)\|^2_F \le 2|\underline{u}_h|^2_{\mathrm{s},h}.\end{aligned}$$

Using (48) to bound the right-hand side yields (54).

3.3.3 L^2-error estimate

To close this section, we state a result concerning the convergence of the error in the L^2-norm. Optimal error estimates require in this context further regularity for the continuous operator. More precisely, we assume that, for all $g \in L^2(\Omega)$, the unique solution of the problem: Find $z \in H^1_0(\Omega)$ such that

$$a(z,v) = (g,v) \qquad \forall v \in H^1_0(\Omega)$$

satisfies the a priori estimate

$$\|z\|_{H^2(\Omega)} \le C\|g\|,$$

with real number C depending only on Ω. Elliptic regularity holds when the domain Ω is convex; see, e.g., [30]. The following result, whose detailed proof is omitted, can be obtained using the arguments of [22, Theorem 10] and [1, Corollary 4.6].

Theorem 2 (L^2-error estimate). *Under the assumptions and notations of Theorem 1, and further assuming elliptic regularity and that $f \in H^1(\Omega)$ if $k = 0$, $f \in H^k(\Omega)$ if $k \ge 1$, there exists a real number $C > 0$ independent of h, but possibly depending on Ω, d, ρ, and k, such that*

$$\|p_h^{k+1}\underline{u}_h - u\| \le \begin{cases} Ch^2\|f\|_{H^1(\Omega)} & \text{if } k = 0, \\ Ch^{k+2}\left(\|u\|_{H^{k+2}(\Omega)} + \|f\|_{H^k(\Omega)}\right) & \text{if } k \ge 1. \end{cases} \tag{55}$$

Remark 5 (Supercloseness of element DOFs). An intermediate step in the proof of the estimate (55) (see [22, Theorem 10]) consists in showing that the element DOFs are superclose to the L^2-projection of the exact solution on $\mathbb{P}^k(\mathscr{T}_h)$:

$$\|\pi_h^{0,k}u - u_h\| \le \begin{cases} Ch^2\|f\|_{H^1(\Omega)} & \text{if } k = 0, \\ Ch^{k+2}\left(\|u\|_{H^{k+2}(\Omega)} + \|f\|_{H^k(\Omega)}\right) & \text{if } k \ge 1. \end{cases} \tag{56}$$

This is done adapting to the HHO framework the classical Aubin–Nitsche technique.

3.4 A posteriori error analysis

For smooth enough exact solutions, it is classically expected that increasing the polynomial degree k will reduce the computational time required to achieve a desired precision; see, e.g., the numerical test in Section 3.5.2 below and, in particular, Fig. 6. However, when the regularity requirements detailed in Theorems 1 and 2 are not met, the order of convergence is limited by the regularity of the solution instead of the polynomial degree. sTo restore optimal orders of convergence, local mesh adaptation is required. This is typically done using a posteriori error estimators to mark the elements where the error is larger, and locally refine the computational mesh based on this information. Here, we present energy-norm upper and lower bounds for the HHO method (36) inspired by the residual-based approach of [27].

3.4.1 Error upper bound

We start by proving an upper bound of the discretization error in terms of quantities whose computation does not require the knowledge of the exact solution. We will need the following local Poincaré and Friedrichs inequalities, valid for all $T \in \mathscr{T}_h$ and all $\varphi \in H^1(T)$:

$$\|\varphi - \pi_T^{0,0}\varphi\|_T \le C_{\mathrm{P},T} h_T \|\nabla\varphi\|_T, \tag{57}$$

$$\|\varphi - \pi_T^{0,0}\varphi\|_{\partial T} \le C_{\mathrm{F},T}^{1/2} h_T^{1/2} \|\nabla\varphi\|_T. \tag{58}$$

In (57), $C_{\mathrm{P},T}$ is a constant equal to π^{-1} if T is convex [2, 36], and for which upper bounds on nonconvex elements can be found in [38]. In (58), $C_{\mathrm{F},T}$ is a constant which, if T is a simplex, can be estimated as $C_{\mathrm{F},T} = C_{\mathrm{P},T}(h_T|\partial T|_{d-1}/|T|_d)(2/d + C_{\mathrm{P},T})$ (see [17, Section 5.6.2.2]).

Theorem 3 (A posteriori error upper bound). *Let $u \in H_0^1(\Omega)$ and $\underline{u}_h \in \underline{U}_{h,0}^k$ denote the unique solutions to problems* (9) *and* (36)*, respectively, with local stabilization bilinear form s_T satisfying the assumptions of Proposition 3 for all $T \in \mathscr{T}_h$. Let u_h^* be an arbitrary function in $H_0^1(\Omega)$. Then, it holds that*

$$\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\| \le \left[\sum_{T\in\mathscr{T}_h} \left(\eta_{\mathrm{nc},T}^2 + (\eta_{\mathrm{res},T} + \eta_{\mathrm{sta},T})^2\right)\right]^{1/2}, \tag{59}$$

with local nonconformity, residual, and stabilization estimators such that, for all $T \in \mathscr{T}_h$,

$$\eta_{\mathrm{nc},T} := \|\nabla(p_T^{k+1}\underline{u}_T - u_h^*)\|_T, \tag{60a}$$

$$\eta_{\mathrm{res},T} := C_{\mathrm{P},T} h_T \|(f + \triangle p_T^{k+1}\underline{u}_T) - \pi_T^{0,0}(f + \triangle p_T^{k+1}\underline{u}_T)\|_T, \tag{60b}$$

$$\eta_{\mathrm{sta},T} := C_{\mathrm{F},T}^{1/2} h_T^{1/2} \left(\sum_{F\in\mathscr{F}_T} \|R_{TF}^k \underline{u}_T\|_F^2\right)^{1/2}, \tag{60c}$$

where, for all $F \in \mathscr{F}_T$, the boundary residual R_{TF}^k is defined by (44)*.*

Remark 6 (Nonconformity estimator). To compute the estimator $\eta_{\mathrm{nc},T}$, we can obtain a $H_0^1(\Omega)$-conforming function u_h^* by applying a node-averaging operator to $p_h^{k+1}\underline{u}_h$. Let an integer $l \ge 1$ be fixed. When $\mathscr{T}_h$ is a matching simplicial mesh and $\mathscr{F}_h$ is the corresponding set of simplicial faces, the node-averaging operator $\mathscr{I}_h^l : \mathbb{P}^l(\mathscr{T}_h) \to \mathbb{P}^l(\mathscr{T}_h) \cap H_0^1(\Omega)$ is defined by setting for each (Lagrange) interpolation node N

$$\mathscr{I}_h^l v_h(N) := \begin{cases} \frac{1}{\mathrm{card}(\mathscr{T}_N)} \sum_{T\in\mathscr{T}_N} (v_h)_{|T}(N) & \text{if } N\in\Omega, \\ 0 & \text{if } N\in\partial\Omega, \end{cases}$$

where the set $\mathscr{T}_N \subset \mathscr{T}_h$ collects the simplices to which N belongs. We then set

$$u_h^* := \mathscr{I}_h^{k+1} p_h^{k+1} \underline{u}_h. \tag{61}$$

The generalization to polytopal meshes can be realized applying the node averaging operator to $p_h^{k+1}\underline{u}_h$ on a simplicial submesh of $\mathscr{T}_h$ (whose existence is guaranteed for regular mesh sequences, see Definition 3).

Proof. Let the equation residual $\mathscr{R} \in H^{-1}(\Omega)$ be such that, for all $\varphi \in H_0^1(\Omega)$, $\langle \mathscr{R}, \varphi\rangle_{-1,1} := (f,\varphi) - (\nabla_h p_h^{k+1}\underline{u}_h, \nabla\varphi)$. The following abstract error estimate descends from [17, Lemma 5.44] and is valid for any function $u_h^* \in H_0^1(\Omega)$:

$$\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\|^2 \le \|\nabla_h(p_h^{k+1}\underline{u}_h - u_h^*)\|^2 + \left(\sup_{\varphi\in H_0^1(\Omega),\|\nabla\varphi\|=1} \langle\mathscr{R},\varphi\rangle_{-1,1}\right)^2. \tag{62}$$

Denote by $\mathfrak{T}_1$ and $\mathfrak{T}_2$ the addends in the right-hand side of (62).

(i) *Bound of* $\mathfrak{T}_1$. Recalling the definition (60a) of the nonconformity estimator, it is readily inferred that

$$\mathfrak{T}_1 = \sum_{T\in\mathscr{T}_h} \eta_{\mathrm{nc},T}^2. \tag{63}$$

(ii) *Bound of* $\mathfrak{T}_2$. We bound the argument of the supremum in $\mathfrak{T}_2$ for a generic function $\varphi \in H_0^1(\Omega)$. Using an element-by-element integration by parts, we obtain

$$\langle\mathscr{R},\varphi\rangle_{-1,1} = \sum_{T\in\mathscr{T}_h}\left((f+\triangle p_T^{k+1}\underline{u}_T,\varphi)_T - \sum_{F\in\mathscr{F}_T}(\nabla p_T^{k+1}\underline{u}_T\cdot\mathbf{n}_{TF},\varphi)_F\right). \tag{64}$$

Let now $\underline{\varphi}_h \in \underline{U}_{h,0}^k$ be such that $\varphi_T = \pi_T^{0,0}\varphi$ for all $T\in\mathscr{T}_h$ and $\varphi_F = \pi_F^{0,k}\varphi_{|F}$ for all $F\in\mathscr{F}_h$. We have that

$$\begin{aligned}
\sum_{T\in\mathscr{T}_h}(\pi_T^{0,0}(f+\triangle p_T^{k+1}\underline{u}_T),\varphi)_T &= \sum_{T\in\mathscr{T}_h}(f+\triangle p_T^{k+1}\underline{u}_T,\varphi_T)_T \\
&= \sum_{T\in\mathscr{T}_h}\left(\mathrm{a}_T(\underline{u}_T,\underline{\varphi}_T) + \sum_{F\in\mathscr{F}_T}(\nabla p_T^{k+1}\underline{u}_T\cdot\mathbf{n}_{TF},\varphi_T)_F\right) \\
&= \sum_{T\in\mathscr{T}_h}\left(\mathrm{s}_T(\underline{u}_T,\underline{\varphi}_T) + \sum_{F\in\mathscr{F}_T}(\nabla p_T^{k+1}\underline{u}_T\cdot\mathbf{n}_{TF},\varphi)_F\right),
\end{aligned} \tag{65}$$

where we have used definition (4) of $\pi_T^{0,0}$ in the first line, the discrete problem (36) with $\underline{v}_h = \underline{\varphi}_h$ and an element-by-element integration by parts together with the fact that $\nabla\varphi_T \equiv 0$ for all $T\in\mathscr{T}_h$ in the second line. In order to pass to the third line, we have expanded a_T according to its definition (18) and used (16a) with $\underline{v}_T = \underline{\varphi}_T$

and $w = p_T^{k+1}\underline{u}_T$ for the consistency term (in the boundary integral, we can write φ instead of φ_F using the definition (4) of $\pi_F^{0,k}$).
Summing (65) and (64), and rearranging the terms, we obtain

$$\langle \mathscr{R}, \varphi \rangle_{-1,1} = \sum_{T\in\mathscr{T}_h} \left((f + \triangle p_T^{k+1}\underline{u}_T - \pi_T^{0,0}(f + \triangle p_T^{k+1}\underline{u}_T), \varphi - \varphi_T)_T + \mathrm{s}_T(\underline{u}_T, \underline{\varphi}_T) \right), \tag{66}$$

where we have used the definition (4) of $\pi_T^{0,0}$ to insert φ_T into the first term. Let us estimate the addends inside the summation, hereafter denoted by $\mathfrak{T}_{2,1}(T)$ and $\mathfrak{T}_{2,2}(T)$. Using the Cauchy–Schwarz and local Poincaré (57) inequalities, and recalling the definition (60b) of the residual estimator, we readily infer, for all $T \in \mathscr{T}_h$, that

$$|\mathfrak{T}_{2,1}(T)| \le \eta_{\mathrm{res},T} \|\nabla\varphi\|_T. \tag{67}$$

On the other hand, recalling the reformulation (46) of the local stabilization bilinear form s_T we have, for all $T \in \mathscr{T}_h$,

$$|\mathfrak{T}_{2,2}(T)| = \left| \sum_{F\in\mathscr{F}_T} (R_{TF}^k \underline{u}_T, \varphi - \varphi_T)_F \right| \le \eta_{\mathrm{sta},T} \|\nabla\varphi\|_T, \tag{68}$$

where we have used the fact that $\varphi_F = \pi_F^{0,k}\varphi$ and $R_{TF}^k\underline{u}_T \in \mathbb{P}^k(F)$ together with the definition (4) of $\pi_F^{0,k}$ to write φ instead of φ_F inside the boundary term, and the Cauchy–Schwarz and local Friedrichs (58) inequalities followed by definition (60c) of the stability estimator to conclude. Using (67) and (68) to estimate the right-hand side of (66) followed by a Cauchy–Schwarz inequality, and plugging the resulting bound inside the supremum in $\mathfrak{T}_2$, we arrive at

$$\mathfrak{T}_2 \le \sum_{T\in\mathscr{T}_h} (\eta_{\mathrm{res},T} + \eta_{\mathrm{sta},T})^2. \tag{69}$$

(iii) *Conclusion.* Plugging (63) and (69) into (62), the conclusion follows. □

3.4.2 Error lower bound

In practice, one wants to make sure that the error estimators are able to correctly localize the error (for use, e.g., in adaptive mesh refinement) and that they do not unduly overestimate it. We prove in this section that the error estimators defined in Theorem 3 are *locally efficient*, i.e., they are locally controlled by the error. This shows that they are suitable to drive mesh refinement. Moreover, they are also *globally efficient*, i.e., the right-hand side of (59) is (uniformly) controlled by the discretization error, so that it cannot depart from it.

Let a mesh element $T \in \mathscr{T}_h$ be fixed and define the following sets of faces and elements sharing at least one node with T:

$$\mathscr{F}_{\mathscr{N},T} := \{F \in \mathscr{F}_h \mid \overline{F} \cap \partial T \neq \emptyset\}, \qquad \mathscr{T}_{\mathscr{N},T} := \{T' \in \mathscr{T}_h \mid \overline{T}' \cap \overline{T} \neq \emptyset\}.$$

Let an integer $l \geq 1$ be fixed. The following result is proved in [32] for standard meshes: There is a real number $C > 0$ independent of h, but possibly depending on d, ρ, and l, such that, for all $v_h \in \mathbb{P}^l(\mathscr{T}_h)$ and all $T \in \mathscr{T}_h$,

$$\|v_h - \mathscr{I}_h^l v_h\|_T^2 \leq C \sum_{F \in \mathscr{F}_{\mathscr{N},T}} h_F \|[v_h]_F\|_F^2, \tag{70}$$

with jump operator defined by (53). Following [17, Section 5.5.2], (70) still holds on regular polyhedral meshes when the nodal interpolator is defined on the matching simplicial submesh of Definition 3. We also note the following technical result:

Proposition 4 (Estimate of boundary oscillations). *Let an integer $l \geq 0$ be fixed. There is a real number $C > 0$ independent of h, but possibly depending on d, ρ, and l, such that, for all mesh elements $T \in \mathscr{T}_h$ and all functions $\varphi \in H^1(T)$,*

$$h_F^{-1/2} \|\varphi - \pi_F^{0,l}\varphi\|_F \leq C \|\nabla\varphi\|_T. \tag{71}$$

Proof. We abridge as $A \lesssim B$ the inequality $A \leq cB$ with multiplicative constant $c > 0$ having the same dependencies as C in (71). Let $F \in \mathscr{F}_T$ and observe that

$$\begin{aligned} \|\varphi - \pi_F^{0,l}\varphi\|_F &\leq \|\varphi - \pi_T^{0,l}\varphi\|_F + \|\pi_F^{0,l}(\pi_T^{0,l}\varphi - \varphi)\|_F \\ &\leq 2\|\varphi - \pi_T^{0,l}\varphi\|_F \lesssim h_T^{1/2}\|\nabla\varphi\|_T, \end{aligned} \tag{72}$$

where we have inserted $\pm\pi_T^{0,l}\varphi$ and used the triangle inequality to infer the first bound, we have used the $L^2(F)$-boundedness of $\pi_F^{0,l}$ to infer the second, and invoked (7b) with $\alpha = 0$, $m = 0$, and $s = 1$ to conclude. Using the fact that $h_T/h_F \lesssim 1$ owing to (2) gives the desired result. □

Theorem 4 (A posteriori error lower bound). *Under the assumptions of Theorem 3, and further assuming, for the sake of simplicity, (i) that the local stabilization bilinear form s_T is given by (22) for all $T \in \mathscr{T}_h$, (ii) that $u_h^\star$ is obtained applying the node-averaging operator to $p_h^{k+1}\underline{u}_h$ on $\mathscr{T}_h$ if $\mathscr{T}_h$ is matching simplicial or on the simplicial submesh of Definition 3 if this is not the case, and (iii) that $f \in \mathbb{P}^{k+1}(\mathscr{T}_h)$, it holds for all $T \in \mathscr{T}_h$,*

$$\eta_{\mathrm{nc},T} \leq C\left(\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\|_{\mathscr{N},T} + |\underline{u}_h|_{\mathrm{s},\mathscr{N},T}\right), \tag{73a}$$

$$\eta_{\mathrm{res},T} \leq C\|\nabla(p_T^{k+1}\underline{u}_T - u_{|T})\|_T, \tag{73b}$$

$$\eta_{\mathrm{sta},T} \leq C|\underline{u}_T|_{\mathrm{s},T}, \tag{73c}$$

where $C > 0$ is a real number possibly depending on d, ρ, and on k but independent of both h and T. For all $T \in \mathscr{T}_h$, $\|\cdot\|_{\mathscr{N},T}$ denotes the L^2-norm on the union of the elements in $\mathscr{T}_{\mathscr{N},T}$ and we have set

$$|\underline{u}_T|_{\mathrm{s},T} = \mathrm{s}_T(\underline{u}_T,\underline{u}_T)^{1/2}, \qquad |\underline{u}_h|^2_{\mathrm{s},\mathscr{N},T} := \sum_{T'\in\mathscr{T}_{\mathscr{N},T}} |\underline{u}_T|^2_{\mathrm{s},T'}.$$

Proof. Let a mesh element $T\in\mathscr{T}_h$ be fixed. In the proof, we abridge as $A\lesssim B$ the inequality $A\le cB$ with multiplicative constant $c>0$ having the same dependencies as C in (73).

(i) *Bound* (73a) *on the nonconformity estimator.* Using a local inverse inequality (see, e.g., [17, Lemma 1.44]) and the relation (70), we infer from (60a) that

$$\eta^2_{\mathrm{nc},T} \lesssim h_T^{-2}\|p_T^{k+1}\underline{u}_T - u_h^*\|^2_T \lesssim \sum_{F\in\mathscr{F}_{\mathscr{N},T}} h_F^{-1}\|[p_h^{k+1}\underline{u}_h]_F\|^2_F, \tag{74}$$

where we have used the fact that, owing to mesh regularity, $h_F\lesssim h_T$ for all $F\in\mathscr{F}_{\mathscr{N},T}$. Using the fact $[u]_F=0$ for all $F\in\mathscr{F}_h$ (see, e.g., [17, Lemma 4.3]) to write $[p_h^{k+1}\underline{u}_h-u]_F$ instead of $[p_h^{k+1}\underline{u}_h]_F$, inserting $\pi_F^{0,k}[p_h^{k+1}\underline{u}_h]_F-\pi_F^{0,k}[p_h^{k+1}\underline{u}_h-u]_F=0$ inside the norm, and using the triangle inequality, we have for all $F\in\mathscr{F}_{\mathscr{N},T}$,

$$\begin{aligned}\|[p_h^{k+1}\underline{u}_h]_F\|_F &\le \|[p_h^{k+1}\underline{u}_h-u]_F-\pi_F^{0,k}[p_h^{k+1}\underline{u}_h-u]_F\|_F + \|\pi_F^{0,k}[p_h^{k+1}\underline{u}_h]_F\|_F\\ &\le \sum_{T\in\mathscr{T}_F}\|(p_T^{k+1}\underline{u}_T-u)-\pi_F^{0,k}(p_T^{k+1}\underline{u}_T-u)\|_F + \|\pi_F^{0,k}[p_h^{k+1}\underline{u}_h]_F\|_F,\end{aligned}$$

where we have expanded the jump according to its definition (53) and used a triangle inequality to pass to the second line. Plugging the above bound into (74), and using multiple times (71) with $\varphi = p_T^{k+1}\underline{u}_T-u$ for $T\in\mathscr{T}_{\mathscr{N},T}$, we arrive at

$$\eta^2_{\mathrm{nc},T}\lesssim \|\nabla(p_T^{k+1}\underline{u}_T-u)\|^2_{\mathscr{N},T} + \sum_{F\in\mathscr{F}_{\mathscr{N},T}} h_F^{-1}\|\pi_F^{0,k}[p_h^{k+1}\underline{u}_h]_F\|^2_F.$$

To conclude, we proceed as in the proof of Lemma 4 to prove that the last term is bounded by $|\underline{u}_h|^2_{\mathrm{s},\mathscr{N},T}$ up to a constant independent of h.

(ii) *Bound* (73b) *on the residual estimator.* We use classical bubble function techniques, see e.g. [37]. For the sake of brevity, we let $r_T := f_{|T}+\triangle p_T^{k+1}\underline{u}_T$. Denote by $\mathfrak{T}_h$ the simplicial submesh of $\mathscr{T}_h$ introduced in Definition 3, and let $\mathfrak{T}_T := \{\tau\in\mathfrak{T}_h \mid \tau\subset T\}$, the set of simplices contained in T. For all $\tau\in\mathfrak{T}_T$, we denote by $b_\tau\in H^1_0(\tau)$ the element bubble function equal to the product of barycentric coordinates of τ and rescaled so as to take the value 1 at the center of gravity of τ. Letting $\psi_\tau := b_\tau r_T$ for all $\tau\in\mathfrak{T}_T$, the following properties hold [37]:

$$\psi_\tau = 0 \text{ on } \partial\tau, \quad (75a) \qquad \|r_T\|^2_\tau \lesssim (r_T,\psi_\tau)_\tau, \quad (75b) \qquad \|\psi_\tau\|_\tau \lesssim \|r_T\|_\tau. \quad (75c)$$

We have that

$$\begin{aligned}\|r_T\|_T^2 &= \sum_{\tau\in\mathfrak{T}_T}\|r_T\|_\tau^2 \lesssim \sum_{\tau\in\mathfrak{T}_T}(r_T,\psi_\tau)_\tau\\ &= \sum_{\tau\in\mathfrak{T}_T}(\nabla(u-p_T^{k+1}\underline{u}_T),\nabla\psi_\tau)_\tau\\ &\le \|\nabla(u-p_T^{k+1}\underline{u}_T)\|_T\left(\sum_{\tau\in\mathfrak{T}_T}h_\tau^{-2}\|\psi_\tau\|_\tau^2\right)^{1/2}\\ &\lesssim h_T^{-1}\|\nabla(u-p_T^{k+1}\underline{u}_T)\|_T\|r_T\|_T,\end{aligned}\tag{76}$$

where we have used property (75b) in the first line, the fact that $f=-\triangle u$ together with an integration by parts and property (75a) to pass to the second line, the Cauchy–Schwarz inequality together with a local inverse inequality (see, e.g., [17, Lemma 1.44]) to pass to the third line, and (75c) together with the fact that $h_\tau^{-1}\le(\rho h_T)^{-1}$ for all $\tau\in\mathfrak{T}_T$ (see Definition 3) to conclude. Recalling the definition (60b) of the residual estimator, observing that $\|r_T-\pi_T^{0,0}r_T\|_T\le\|r_T\|_T$ as a result of the triangle inequality followed by the $L^2(T)$-boundedness of $\pi_T^{0,0}$, and using (76), the bound (73b) follows.

(iii) *Bound* (73c) *on the stabilization estimator.* Using the definition (44) of the boundary residual operator $\underline{R}_{\partial T}^k$ with $\underline{v}_T=\underline{u}_T$ and $\underline{\alpha}_{\partial T}=-h_T\underline{R}_{\partial T}^k\underline{u}_T=(-h_TR_{TF}^k\underline{u}_T)_{F\in\mathscr{F}_T}$, the stabilization estimator (60c) can be bounded as follows:

$$\eta_{\mathrm{sta},T}^2=C_{\mathrm{F},T}\mathrm{s}_T(\underline{u}_T,(0,-h_T\underline{R}_{\partial T}^k\underline{u}_T))\lesssim|\underline{u}_T|_{\mathrm{s},T}|(0,-h_T\underline{R}_{\partial T}^k\underline{u}_T)|_{\mathrm{s},T}.\tag{77}$$

On the other hand, from property (S2) in Assumption 1, the relation (2), and the definition (60c) of $\eta_{\mathrm{sta},T}$, it is inferred that

$$|(0,-h_T\underline{R}_{\partial T}^k\underline{u}_T)|_{\mathrm{s},T}\le\eta^{1/2}\left(\sum_{F\in\mathscr{F}_T}h_F^{-1}\|h_TR_{TF}^k\underline{u}_T\|_F^2\right)^{1/2}\le\eta^{1/2}\rho^{-1}C_{\mathrm{F},T}^{-1/2}\eta_{\mathrm{sta},T}.$$

Using this estimate to bound the right-hand side of (77), (73c) follows. □

Corollary 1 (Global lower bound). *Under the assumptions of Theorem 4, there exists a constant C independent of h, but possibly depending on d, ρ and k, such that*

$$\left[\sum_{T\in\mathscr{T}_h}\left(\eta_{\mathrm{nc},T}^2+(\eta_{\mathrm{res},T}+\eta_{\mathrm{sta},T})^2\right)\right]^{1/2}\le C\left(\|\nabla_h(p_h^{k+1}\underline{u}_h-u)\|+|\underline{u}_h|_{\mathrm{s},h}\right).$$

3.5 Numerical examples

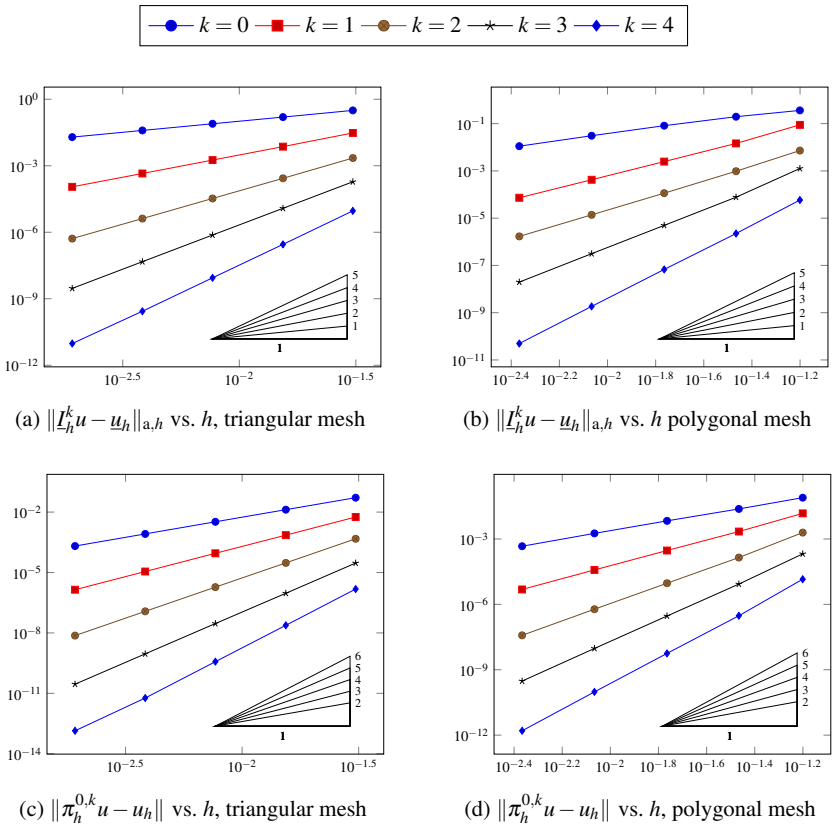
We illustrate the numerical performance of the HHO method on a set of model problems.

3.5.1 Two-dimensional test case

The first test case, taken from [22], aims at demonstrating the estimated orders of convergence in two space dimensions. We solve the Dirichlet problem in the unit square $\Omega = (0,1)^2$ with

$$u(\mathbf{x}) = \sin(\pi x_1)\sin(\pi x_2), \tag{78}$$

and corresponding right-hand side $f(\mathbf{x}) = 2\pi^2 \sin(\pi x_1)\sin(\pi x_2)$ on the the triangular and polygonal meshes of Fig. 1a and 1c. Fig. 4 displays convergence results for both mesh families and polynomial degrees up to 4. Recalling (50) and (56), we measure the energy- and L^2-errors by the quantities $\|\underline{I}_h^k u - \underline{u}_h\|_{\mathrm{a},h}$ and $\|\pi_h^{0,k} u - u_h\|$, respectively. In all cases, the numerical results show asymptotic convergence rates that match those predicted by the theory.



(a) $\|\underline{I}_h^k u - \underline{u}_h\|_{\mathrm{a},h}$ vs. h, triangular mesh

(b) $\|\underline{I}_h^k u - \underline{u}_h\|_{\mathrm{a},h}$ vs. h polygonal mesh

(c) $\|\pi_h^{0,k} u - u_h\|$ vs. h, triangular mesh

(d) $\|\pi_h^{0,k} u - u_h\|$ vs. h, polygonal mesh

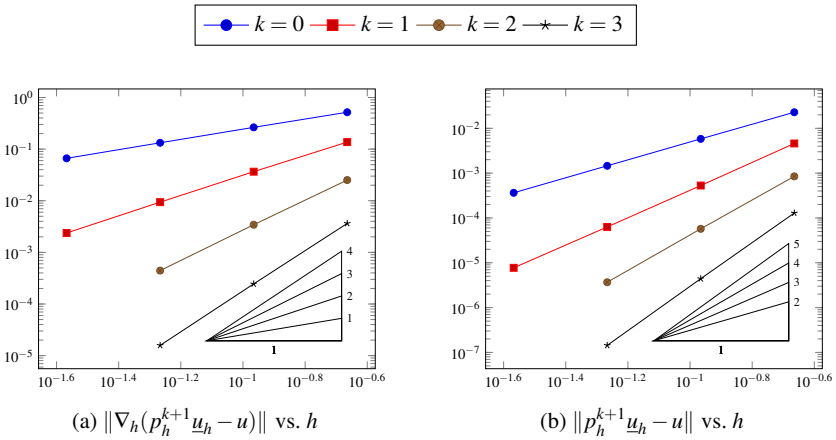
Fig. 4: Error vs. h for the test case of Section 3.5.1.

3.5.2 Three-dimensional test case

The second test case, taken from [27], demonstrates the orders of convergence in three space dimensions. We solve the Dirichlet problem in the unit cube $\Omega = (0,1)^3$ with

$$u(\mathbf{x}) = \sin(\pi x_1)\sin(\pi x_2)\sin(\pi x_3),$$

and corresponding right-hand side $f(\mathbf{x}) = 3\pi^2 \sin(\pi x_1)\sin(\pi x_2)\sin(\pi x_3)$ on a matching simplicial mesh family for polynomial degrees up to 3. The numerical results displayed in Fig. 5 show asymptotic convergence rates that match those predicted by (48) and (55). In Fig. 6 we display the error versus the total computational time t_{tot} (including the pre-processing, solution, and post-processing). It can be seen that the energy- and L^2-errors optimally scale as $t_{\mathrm{tot}}^{(k+1)/d}$ and $t_{\mathrm{tot}}^{(k+2)/d}$ (with $d = 3$), respectively.



(a) $\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\|$ vs. h (b) $\|p_h^{k+1}\underline{u}_h - u\|$ vs. h

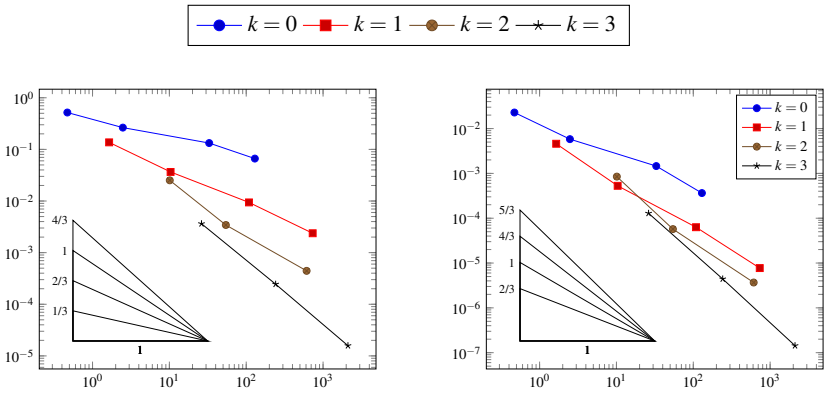
Fig. 5: Error vs. h for the test case of Section 3.5.2.

3.5.3 Three-dimensional case with adaptive mesh refinement

The third test case, known as Fichera corner benchmark, is taken from [27] and is based on the exact solution of [29] on the etched three-dimensional domain $\Omega = (-1,1)^3 \setminus [0,1]^3$:

$$u(\mathbf{x}) = \sqrt[4]{x_1^2 + x_2^2 + x_3^2},$$

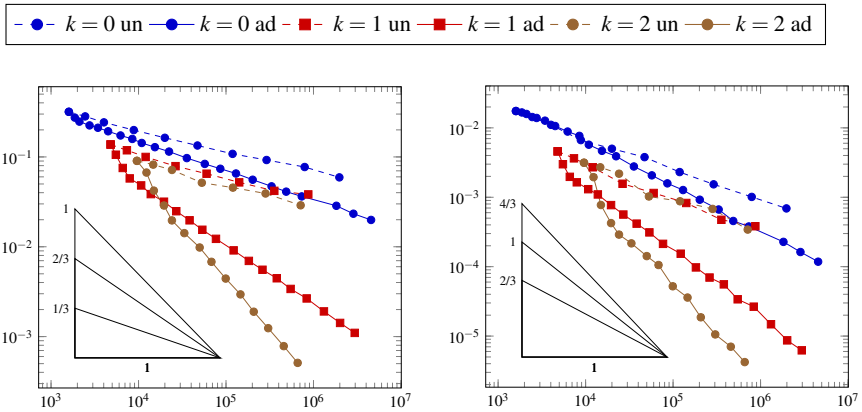
with right-hand side $f(\mathbf{x}) = -3/4(x_1^2 + x_2^2 + x_3^2)^{-3/4}$. In this case, the gradient of the solution has a singularity in the origin which prevents the method from attaining optimal convergence rates even for $k = 0$. In Fig. 7 we show a computation comparing the numerical error versus N_{dof} (cf. (38c)) for the Fichera problem on uniformly



(a) $\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\|$ vs. t_{tot} (b) $\|p_h^{k+1}\underline{u}_h - u\|$ vs. t_{tot}

Fig. 6: Error vs. total computational time for the test case of Section 3.5.2.

and adaptively refined mesh sequences for polynomial degrees up to 3. Clearly, the order of convergence is limited by the solution regularity when using uniformly refined meshes, while using adaptively refined meshes we recover optimal orders of convergence of $N_{\mathrm{dof}}^{(k+1)/d}$ and $N_{\mathrm{dof}}^{(k+2)/d}$ (with $d = 3$) for the energy- and L^2-errors, respectively.



(a) Energy-error vs. N_{dof} (b) L^2-error vs. N_{dof}

Fig. 7: Error vs. N_{dof} for the test case of Section 3.5.3.

4 A nonlinear example: The p-Laplace equation

We consider in this section an extension of the HHO method to the p-Laplace equation. This problem will be used to introduce the techniques for the discretization and analysis of nonlinear operators, as well as a set of functional analysis results of independent interest. An additional interesting point is that the p-Laplace problem is naturally posed in a non-Hilbertian setting. This will require to emulate a Sobolev structure at the discrete level.

Let $p \in (1,+\infty)$ be fixed, and set $p' := \frac{p}{p-1}$. The p-Laplace problem reads: Find $u : \Omega \to \mathbb{R}$ such that

$$\begin{aligned} -\nabla\cdot(\sigma(\nabla u)) &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega, \end{aligned} \tag{79}$$

where $f \in L^{p'}(\Omega)$ is a volumetric source term and the function $\sigma : \mathbb{R}^d \to \mathbb{R}^d$ is such that

$$\sigma(\tau) := |\tau|^{p-2}\tau. \tag{80}$$

The p-Laplace equation is a generalization of the Poisson problem considered in Section 3, which corresponds to the choice $p = 2$.

Classically, the weak formulation of problem (79) reads: Find $u \in W_0^{1,p}(\Omega)$ such that, for all $v \in W_0^{1,p}(\Omega)$,

$$a(u,v) = \int_\Omega f(\mathbf{x})v(\mathbf{x})\mathrm{d}\mathbf{x}, \tag{81}$$

where the function $a : W^{1,p}(\Omega)\times W^{1,p}(\Omega) \to \mathbb{R}$ is such that

$$a(u,v) := \int_\Omega \sigma(\nabla u(\mathbf{x}))\cdot\nabla v(\mathbf{x})\mathrm{d}\mathbf{x}. \tag{82}$$

From this point on, to alleviate the notation, we omit both the dependence of the integrand on $\mathbf{x}$ and the differential from integrals.

4.1 Discrete $W^{1,p}$-norms and Sobolev embeddings

In Section 3, the discrete space $\underline{U}_{h,0}^k$ and the norm $\|\cdot\|_{1,h}$ have played the role of the Hilbert space $H_0^1(\Omega)$ and of the seminorm $|\cdot|_{H^1(\Omega)}$, respectively (notice that $|\cdot|_{H^1(\Omega)}$ is a norm on $H_0^1(\Omega)$ by virtue of the continuous Poincaré inequality). For the p-Laplace equation, $\underline{U}_{h,0}^k$ will replace at the discrete level the Sobolev space $W_0^{1,p}$. A good candidate for the role of the corresponding seminorm $|\cdot|_{W^{1,p}(\Omega)}$ is the map $\|\cdot\|_{1,p,h}$ such that, for all $\underline{v}_h \in \underline{U}_h^k$,

$$\|\underline{v}_h\|_{1,p,h}^p := \sum_{T\in\mathcal{T}_h} \|\underline{v}_T\|_{1,p,T}^p, \tag{83}$$

where, for all $T \in \mathscr{T}_h$,

$$\|\underline{v}_T\|_{1,p,T}^p := \|\nabla v_T\|_{L^p(T)^d}^p + \sum_{F\in\mathscr{F}_T} h_F^{1-p}\|v_F - v_T\|_{L^p(F)}^p. \tag{84}$$

The power of h_F in the second term ensures that both contributions have the same scaling.

The following discrete Sobolev embeddings are proved in [13, Proposition 5.4]. The proof hinges on the results of [16, Theorem 6.1] for broken polynomial spaces (based, in turn, on the techniques originally developed in [28] in the context of Finite Volume methods). Their role in the analysis of HHO methods for problem (81) is discussed in Remark 9.

Theorem 5 (Discrete Sobolev embeddings). *Let a polynomial degree $k \geq 0$ and an index $p \in (1,+\infty)$ be fixed. Let $(\mathscr{M}_h)_{h\in\mathscr{H}}$ denote a regular sequence of meshes in the sense of Definition 3. Let $1 \leq q \leq \frac{dp}{d-p}$ if $1 \leq p < d$ and $1 \leq q < +\infty$ if $p \geq d$. Then, there exists a real number $C > 0$ only depending on Ω, ρ, l, p, and q such that, for all $\underline{v}_h \in \underline{U}_{h,0}^k$,*

$$\|v_h\|_{L^q(\Omega)} \leq C\|\underline{v}_h\|_{1,p,h}. \tag{85}$$

Remark 7 (Discrete Poincaré inequality). The discrete Poincaré inequality (29) is a special case of Theorem 5 corresponding to $p = q = 2$ (this choice is possible in any space dimension).

4.2 Discrete gradient and compactness

The analysis of numerical methods for linear problems is usually carried out in the spirit of the Lax–Richtmyer equivalence theorem: "For a consistent numerical method, stability is equivalent to convergence"; see for instance [12] for a rigorous proof in the case of linear Cauchy problems. When dealing with nonlinear problems, however, some form of compactness is also required; cf. Remark 10 for further insight into this point. In order to achieve it for problem (81), we need to introduce a local gradient reconstruction slightly richer than ∇p_T^{k+1}; see (16).

Let a mesh element $T \in \mathscr{T}_h$ be fixed. By the principles illustrated in Section 3.1.1, we define the local gradient reconstruction $\mathbf{G}_T^k : \underline{U}_T^k \to \mathbb{P}^k(T)^d$ such that, for all $\underline{v}_T \in \underline{U}_T^k$,

$$(\mathbf{G}_T^k\underline{v}_T, \boldsymbol{\tau})_T = -(v_T, \nabla\cdot\boldsymbol{\tau})_T + \sum_{F\in\mathscr{F}_T}(v_F, \boldsymbol{\tau}\cdot\mathbf{n}_{TF})_F \quad \forall\boldsymbol{\tau} \in \mathbb{P}^k(T)^d. \tag{86}$$

Notice that here we reverted to the L^2-product notation instead of using integrals to emphasize the fact that the definition of $\mathbf{G}_T^k$ is inherently L^2-based.

Remark 8 (Relation between $\mathbf{G}_T^k$ and p_T^{k+1}). Taking $\tau = \nabla w$ with $w \in \mathbb{P}^{k+1}(T)$ in (86) and comparing with (16a), it is readily inferred that

$$(\mathbf{G}_T^k \underline{v}_T - \nabla p_T^{k+1} \underline{v}_T, \nabla w)_T = 0 \qquad \forall w \in \mathbb{P}^{k+1}(T), \tag{87}$$

i.e., $\nabla p_T^{k+1} \underline{v}_T$ is the L^2-projection of $\mathbf{G}_T^k \underline{v}_T$ on $\nabla \mathbb{P}^{k+1}(T) \subset \mathbb{P}^k(T)^d$. In passing, we observe that for $k = 0$, using the fact that $\nabla \mathbb{P}^1(T) = \mathbb{P}^0(T)^d$, (87) implies that $\mathbf{G}_T^0 \underline{v}_T = \nabla p_T^1 \underline{v}_T$.

Choosing a larger arrival space for $\mathbf{G}_T^k$ has the effect of modifying the commuting property as follows (compare with (17)): For all $v \in W^{1,1}(T)$,

$$(\mathbf{G}_T^k \circ \underline{I}_T^k) v = \pi_T^{0,k}(\nabla v). \tag{88}$$

At the global level, we define the operator $\mathbf{G}_h^k : \underline{U}_h^k \to \mathbb{P}^k(\mathscr{T}_h)^d$ such that, for all $\underline{v}_h \in \underline{U}_h^k$,

$$(\mathbf{G}_h^k \underline{v}_h)_{|T} := \mathbf{G}_T^k \underline{v}_T \qquad \forall T \in \mathscr{T}_h. \tag{89}$$

The commuting property (88) is used in conjunction with the properties of the L^2-projector to prove the following lemma, which states the compactness of sequences of HHO functions uniformly bounded in a discrete Sobolev norm.

Lemma 5 (Discrete compactness). *Let a polynomial degree $k \geq 0$ and an index $p \in (1, +\infty)$ be fixed. Let $(\mathscr{M}_h)_{h \in \mathscr{H}}$ denote a regular sequence of meshes in the sense of Definition 3. Let $(\underline{v}_h)_{h \in \mathscr{H}} \in (\underline{U}_{h,0}^k)_{h \in \mathscr{H}}$ be a sequence for which there exists a real number $C > 0$ independent of h such that*

$$\|\underline{v}_h\|_{1,p,h} \leq C \qquad \forall h \in \mathscr{H}.$$

Then, there exists $v \in W_0^{1,p}(\Omega)$ such that, up to a subsequence, as $h \to 0$,

(i) $v_h \to v$ and $p_h^{k+1} \underline{v}_h \to v$ strongly in $L^q(\Omega)$ for all $1 \leq q < \frac{dp}{d-p}$ if $1 \leq p < d$ and $1 \leq q < +\infty$ if $p \geq d$;
(ii) $\mathbf{G}_h^k \underline{v}_h \to \nabla v$ weakly in $L^p(\Omega)^d$.

4.3 Discrete problem and well-posedness

The discrete counterpart of the function a defined by (82) is the function $\mathrm{a}_h : \underline{U}_h^k \times \underline{U}_h^k \to \mathbb{R}$ such that, for all $\underline{u}_h, \underline{v}_h \in \underline{U}_h^k$,

$$\mathrm{a}_h(\underline{u}_h,\underline{v}_h) := \int_\Omega \sigma(\mathbf{G}_h^k\underline{u}_h)\cdot\mathbf{G}_h^k\underline{v}_h + \sum_{T\in\mathscr{T}_h}\mathrm{s}_T(\underline{u}_T,\underline{v}_T). \tag{90}$$

Here, for all $T\in\mathscr{T}_h$, $\mathrm{s}_T:\underline{U}_T^k\times\underline{U}_T^k\to\mathbb{R}$ is a local stabilization function which can be obtained, e.g., by generalizing (23) to the non-Hilbertian setting:

$$\mathrm{s}_T(\underline{u}_T,\underline{v}_T) := \sum_{F\in\mathscr{F}_T} h_F^{1-p}\int_F |\delta_{TF}^k\underline{u}_T-\delta_T^k\underline{u}_T|^{p-2}(\delta_{TF}^k\underline{u}_T-\delta_T^k\underline{u}_T)(\delta_{TF}^k\underline{v}_T-\delta_T^k\underline{v}_T). \tag{91}$$

The discrete problem reads: Find $\underline{u}_h\in\underline{U}_{h,0}^k$ such that

$$\mathrm{a}_h(\underline{u}_h,\underline{v}_h) = \int_\Omega f v_h \qquad \forall\underline{v}_h\in\underline{U}_{h,0}^k. \tag{92}$$

The following result summarizes [13, Theorem 4.5, Remark 4.7, and Proposition 6.1].

Lemma 6 (Well-posedness). *Problem* (92) *admits a unique solution, and there exists a real number $C>0$ independent of h, but possibly depending on Ω, d, ρ, and k, such that, denoting by $p' := \frac{p}{p-1}$ the dual exponent of p, it holds that*

$$\|\underline{u}_h\|_{1,p,h} \le C\|f\|_{L^{p'}(\Omega)}^{\frac{1}{p-1}}. \tag{93}$$

Remark 9 (Role of the discrete Sobolev embeddings). The discrete Sobolev embedding (85) with $q=p$ is used in the proof of the a priori bound (93) to estimate the right-hand side of the discrete problem (92) after selecting $\underline{v}_h=\underline{u}_h$ and using Hölder's inequality:

$$\int_\Omega f u_h \le \|f\|_{L^{p'}(\Omega)}\|u_h\|_{L^p(\Omega)} \le \|f\|_{L^{p'}(\Omega)}\|\underline{u}_h\|_{1,p,h}.$$

4.4 Convergence and error analysis

The following theorem states the convergence of the sequence of solutions to problem (92) on a regular mesh sequence. Notice that convergence is proved for exact solutions that display only the minimal regularity $u\in W_0^{1,p}(\Omega)$ required by the weak formulation (81). This is an important point when dealing with nonlinear problems, for which further regularity can be hard to prove, and possibly requires assumptions on the data too strong to be matched in practical situations.

Theorem 6 (Convergence). *Let a polynomial degree $k \geq 0$ and an index $p \in (1,+\infty)$ be fixed. Let $(\mathcal{M}_h)_{h\in\mathscr{H}}$ denote a regular sequence of meshes in the sense of Definition 3. Let $u \in W_0^{1,p}(\Omega)$ denote the unique solution to* (81)*, and denote by $(\underline{u}_h)_{h\in\mathscr{H}} \in (\underline{U}_{h,0}^k)_{h\in\mathscr{H}}$ the sequence of solutions to* (92) *on $(\mathscr{T}_h)_{h\in\mathscr{H}}$. Then, as $h \to 0$, it holds*

(i) $u_h \to u$ and $p_h^{k+1}\underline{u}_h \to u$ strongly in $L^q(\Omega)$ for all $1 \leq q < \frac{dp}{d-p}$ if $1 \leq p < d$ and $1 \leq q < +\infty$ if $p \geq d$;
(ii) $\mathbf{G}_h^k \underline{u}_h \to \nabla u$ strongly in $L^p(\Omega)^d$.

Remark 10 (Role of compactness). The first step in the proof of Theorem 6 consists in proving the existence of a limit for the sequence of discrete solutions. This is done using the compactness result of Lemma 5 in conjunction with the a priori bound (93). The following steps consist in showing that this limit solves (81) (which is done adapting the techniques of [34, 35]) and in proving the strong convergence of the gradient.

When dealing with high-order methods, it is also important to determine the convergence rates attained when the solution is regular enough (or when adaptive mesh refinement is used, cf. Section 3.5.3). The answer to this question is provided by the following result, proved in [14, Theorem 7 and Corollary 10].

Theorem 7 (Energy error estimate). *Under the assumptions and notations of Theorem 6, and further assuming the regularity $u \in W^{k+2,p}(\Omega)$ and $\sigma(\nabla u) \in W^{k+1,p'}(\Omega)^d$ with $p' := \frac{p}{p-1}$, there exists a real number $C > 0$ independent of h such that the following holds: If $p \geq 2$,*

$$\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\|_{L^p(\Omega)^d} + |\underline{u}_h|_{\mathrm{s},h} \leq C\left[h^{k+1}|u|_{W^{k+2,p}(\Omega)} + h^{\frac{k+1}{p-1}}\left(|u|^{\frac{1}{p-1}}_{W^{k+2,p}(\Omega)} + |\sigma(\nabla u)|^{\frac{1}{p-1}}_{W^{k+1,p'}(\Omega)^d}\right)\right], \quad (94a)$$

while, if $p < 2$,

$$\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\|_{L^p(\Omega)^d} + |\underline{u}_h|_{\mathrm{s},h} \leq C\left(h^{(k+1)(p-1)}|u|^{p-1}_{W^{k+2,p}(\Omega)} + h^{k+1}|\sigma(\nabla u)|_{W^{k+1,p'}(\Omega)^d}\right), \quad (94b)$$

where, recalling the definition (91) *of the local stabilization function, we have introduced the seminorm on $\underline{U}_h^k$ such that, for all $\underline{v}_h \in \underline{U}_h^k$, $|\underline{v}_h|^p_{\mathrm{s},h} := \sum_{T\in\mathscr{T}_h} \mathrm{s}_T(\underline{v}_T, \underline{v}_T)$.*

Remark 11 (Order of convergence). The asymptotic scaling for the approximation error in the left-hand side of (94) is determined by the leading terms in the right-hand side. Using the Bachmann–Landau notation,

$$\|\nabla_h(p_h^{k+1}\underline{u}_h - u)\|_{L^p(\Omega)^d} + |\underline{u}_h|_{\mathrm{s},h} = \begin{cases} \mathscr{O}(h^{\frac{k+1}{p-1}}) & \text{if } p \geq 2, \\ \mathscr{O}(h^{(k+1)(p-1)}) & \text{if } p < 2. \end{cases} \tag{95}$$

For a discussion of these orders of convergence and a comparison with other methods studied in the literature, we refer the reader to [14, Remark 3.3].

4.5 Numerical example

To illustrate the performance of the HHO method, we solve the p-Laplace problem corresponding to the exact solution

$$u(\mathbf{x}) = \exp(x_1 + \pi x_2)$$

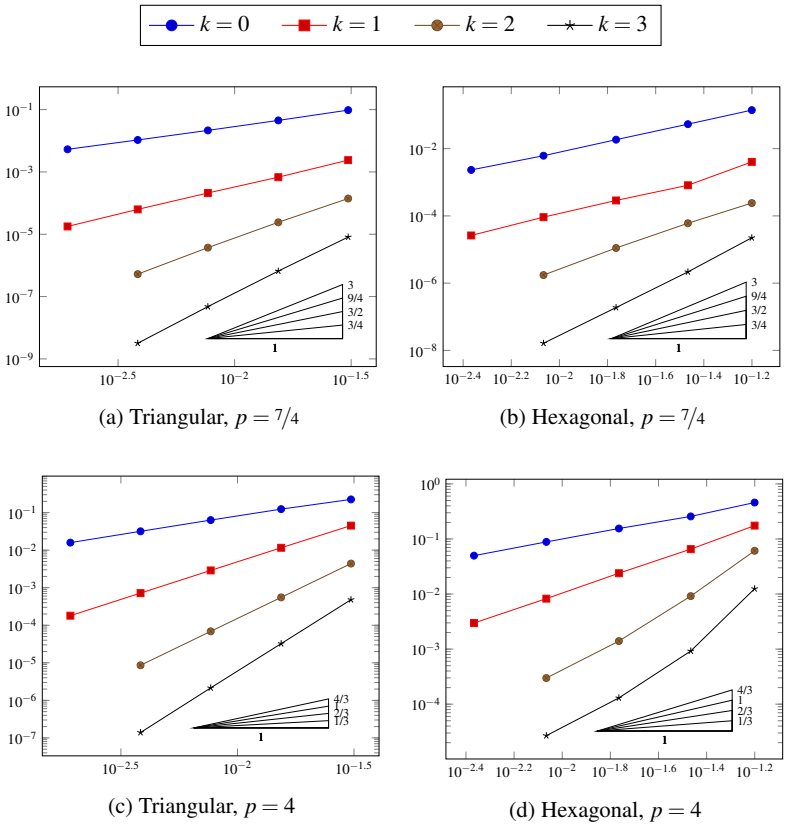
for $p \in \{7/4, 4\}$. This test is taken from [13, Section 4.4] and [14, Section 3.5]. The domain is again the unit square $\Omega = (0,1)^2$, and the volumetric source term f is inferred from (79). The convergence results for the same triangular and polygonal mesh families of Section 3.5.1 (see Fig. 1a and 1c) are displayed in Fig. 8. Here, the error is measured by the quantity $\|\underline{I}_h^k u - \underline{u}_h\|_{1,p,h}$, for which analogous estimates as those in Theorem 7 hold. The error estimate seem sharp for $p = 7/4$, and the asymptotic orders of convergence match the one predicted by the theory. For $p = 4$, better orders of convergence than the asymptotic ones in (95) are observed. One possible explanation is that the lowest-order terms in the right-hand side of (94) are not yet dominant for the specific problem data and mesh. Another possibility is that compensations occur among terms that are separately estimated in the proof.

5 Diffusion-advection-reaction

In this section we extend the HHO method to the scalar diffusion-advection-reaction problem: Find $u : \Omega \to \mathbb{R}$ such that

$$\begin{aligned} \nabla\cdot(-\kappa\nabla u + \beta u) + \mu u &= f \qquad \text{in } \Omega, \\ u &= 0 \qquad \text{on } \partial\Omega, \end{aligned}$$

where (i) $\kappa : \Omega \to \mathbb{R}_+^*$ is the diffusion coefficient, which we assume piecewise constant on a fixed partition of the domain $P_\Omega = \{\omega\}$ and uniformly elliptic; (ii) $\beta \in \mathrm{Lip}(\Omega)^d$ (hence, in particular, $\beta \in W^{1,\infty}(\Omega)^d$) is the advective velocity field, for which we additionally assume, for the sake of simplicity, $\nabla\cdot\beta \equiv 0$;



(a) Triangular, $p = 7/4$

(b) Hexagonal, $p = 7/4$

(c) Triangular, $p = 4$

(d) Hexagonal, $p = 4$

Fig. 8: $\|\underline{I}_h^k u - \underline{u}_h\|_{1,p,h}$ vs. h for the test case of Section 4.5.

(iii) $\mu \in L^\infty(\Omega)$ is the reaction coefficient such that $\mu \geq \mu_0 > 0$ a.e. in Ω for some real number μ_0; (iv) $f \in L^2(\Omega)$ is the volumetric source term.

Having assumed κ uniformly elliptic, the following weak formulation classically holds: Find $u \in H_0^1(\Omega)$ such that

$$a_{\kappa,\beta,\mu}(u,v) = (f,v) \qquad \forall v \in H_0^1(\Omega), \tag{97}$$

where the bilinear form $a_{\kappa,\beta,\mu} : H^1(\Omega) \times H^1(\Omega) \to \mathbb{R}$ is such that

$$a_{\kappa,\beta,\mu}(u,v) := a_\kappa(u,v) + a_{\beta,\mu}(u,v),$$

and the diffusive and advective-reactive contributions are respectively defined by

$$a_\kappa(u,v) := (\kappa\nabla u, \nabla v), \qquad a_{\beta,\mu}(u,v) := \tfrac{1}{2}(\beta\cdot\nabla u, v) - \tfrac{1}{2}(u, \beta\cdot\nabla v) + (\mu u, v).$$

The first novel ingredient introduced in this section is the robust HHO discretization of first-order terms. Problem (97) is characterized by the presence of spatially varying coefficients, which can give rise to different regimes in different regions of the domain. In practice, one is typically interested in numerical methods that handle in a robust way locally dominant advection, corresponding to large values of a local Péclet number. As pointed out in [21], this requires that the discrete counterpart of the bilinear form $a_{\beta,\mu}$ satisfies a stability condition that guarantees well-posedness even in the absence of diffusion. This is realized here combining a reconstruction of the advective derivative obtained in the HHO spirit with an upwind stabilization that penalizes the differences between face- and element-based DOFs.

The second novelty introduced in this section is a formulation of diffusive terms with weakly enforced boundary conditions. A relevant feature of problem (97) is that boundary layers can appear in the vicinity of the outflow portion of $\partial\Omega$ when the diffusion coefficient takes small values. To improve the numerical approximation in this situation, one can resort to weakly enforced boundary conditions, which do not constrain the numerical solution to a fixed boundary value.

The following material is closely inspired by [15], where locally vanishing diffusion is treated (see Remark 14), and more general formulations for the advective stabilization term are considered.

5.1 Discretization of diffusive terms with weakly enforced boundary conditions

To avoid dealing with jumps of the diffusion coefficient inside mesh elements when writing the HHO discretization of problem (97) on a mesh $\mathscr{M}_h = (\mathscr{T}_h, \mathscr{F}_h)$, we make the following

Assumption 2 (Compatible mesh) *The mesh $\mathscr{M}_h = (\mathscr{T}_h, \mathscr{F}_h)$ is compatible with the diffusion coefficient, i.e., for all $T \in \mathscr{T}_h$, there exists a unique subdomain $\omega \in P_\Omega$ such that $T \subset \omega$. For all $T \in \mathscr{T}_h$ we set, for the sake of brevity, $\kappa_T := \kappa_{|T}$.*

Letting $\zeta > 0$ denote a user-dependent boundary penalty parameter, we define the discrete diffusive bilinear form $\mathrm{a}_{\kappa,h} : \underline{U}_h^k \times \underline{U}_h^k \to \mathbb{R}$ such that

$$\begin{aligned}\mathrm{a}_{\kappa,h}(\underline{u}_h, \underline{v}_h) &:= \sum_{T\in\mathscr{T}_h} \kappa_T \mathrm{a}_T(\underline{u}_T, \underline{v}_T)\\ &+ \sum_{F\in\mathscr{F}_h^{\mathrm{b}}} \left\{ -(\kappa_{T_F}\nabla p_{T_F}^{k+1}\underline{u}_T, v_F)_F + (u_F, \kappa_{T_F}\nabla p_{T_F}^{k+1}\underline{v}_T)_F + \frac{\zeta\,\kappa_{T_F}}{h_F}(u_F, v_F)_F \right\},\end{aligned} \tag{98}$$

where, for all mesh elements $T \in \mathscr{T}_h$, a_T is the local diffusive bilinear form defined by (18) and, for all boundary faces $F \in \mathscr{F}_h^{\mathrm{b}}$, T_F denotes the unique mesh element such that $F \subset \partial T_F$. The terms in the second line of (98) are responsible for the weak enforcement of boundary conditions à la Nitsche.

Define the diffusion-weighted norm on $\underline{U}_h^k$ such that, for all $\underline{v}_h \in \underline{U}_h^k$,

$$\|\underline{v}_h\|_{\kappa,h}^2 := \sum_{T\in\mathscr{T}_h} \kappa_T \|\underline{v}_T\|_{\mathrm{a},T}^2 + \sum_{F\in\mathscr{F}_h^{\mathrm{b}}} \frac{\kappa_{T_F}}{h_F}\|v_F\|_F^2,$$

with seminorm $\|\cdot\|_{\mathrm{a},T}$ defined by (19). It is a simple matter to check that, for all $\zeta \geq 1$, we have the following coercivity property for $\mathrm{a}_{\kappa,h}$: For all $\underline{v}_h \in \underline{U}_h^k$,

$$\|\underline{v}_h\|_{\kappa,h}^2 \leq \mathrm{a}_{\kappa,h}(\underline{v}_h,\underline{v}_h). \tag{99}$$

5.2 Discretization of advective terms with upwind stabilization

We introduce the ingredients for the discretization of first-order terms: a local advective derivative reconstruction and an upwind stabilization term penalizing the differences between face- and element-based DOFs.

5.2.1 Local contribution

Let a mesh element $T \in \mathscr{T}_h$ be fixed. By the principles illustrated in Section 3.1.1, we define the local discrete advective derivative reconstruction $\mathbf{G}_{\beta,T}^k : \underline{U}_T^k \to \mathbb{P}^k(T)$ such that, for all $\underline{v}_T \in \underline{U}_T^k$,

$$(\mathbf{G}_{\beta,T}^k \underline{v}_T, w)_T = -(v_T, \beta\cdot\nabla w)_T + \sum_{F\in\mathscr{F}_T} ((\beta\cdot\mathbf{n}_{TF})v_F, w)_F \quad \forall w \in \mathbb{P}^k(T).$$

The local advective-reactive bilinear form $\mathrm{a}_{\beta,\mu,T} : \underline{U}_T^k \times \underline{U}_T^k \to \mathbb{R}$ is defined as follows:

$$\mathrm{a}_{\beta,\mu,T}(\underline{u}_T,\underline{v}_T) := \frac{1}{2}(\mathbf{G}_{\beta,T}^k\underline{u}_T, v_T)_T - \frac{1}{2}(u_T, \mathbf{G}_{\beta,T}^k\underline{v}_T)_T + (\mu u_T, v_T)_T + \mathrm{s}_{\beta,T}(\underline{u}_T,\underline{v}_T), \tag{100}$$

where the bilinear form

$$\mathrm{s}_{\beta,T}(\underline{u}_T,\underline{v}_T) := \frac{1}{2}\sum_{F\in\mathscr{F}_T} (|\beta\cdot\mathbf{n}_{TF}|(u_F-u_T), v_F - v_T)_F, \tag{101}$$

can be interpreted as an upwind stabilization term.

Remark 12 (Element-face upwind stabilization). Upwinding is realized here by penalizing the difference between face- and element-based DOFs. This is a relevant difference with respect to Finite Volume and Discontinuous Galerkin methods, where jumps of element-based DOFs are considered instead. With the choice (101) for the stabilization term, the stencil remains the same as for a pure diffusion problem, and static condensation of element-based DOFs in the spirit of Section 3.2.4 remains possible.

To express the stability properties of $\mathrm{a}_{\beta,\mu,T}$, we define the local seminorm such that, for all $\underline{v}_T \in \underline{U}_T^k$,

$$\|\underline{v}_T\|_{\beta,\mu,T}^2 := \frac{1}{2}\sum_{F\in\mathscr{F}_T} \||\boldsymbol{\beta}\cdot\mathbf{n}_{TF}|^{1/2}(v_F - v_T)\|_F^2 + \hat{\tau}_T^{-1}\|v_T\|_T^2,$$

where, letting $\mathrm{L}_{\beta,T} := \max_{1\le i\le d}\|\nabla\beta_i\|_{L^\infty(T)^d}$, we have introduced the reference time

$$\hat{\tau}_T := \{\max(\|\mu\|_{L^\infty(T)}, \mathrm{L}_{\beta,T})\}^{-1}.$$

Notice that the map $\|\cdot\|_{\beta,\mu,T}$ is actually a norm on $\underline{U}_T^k$ provided that $\beta_{|F}\cdot\mathbf{n}_{TF}$ is nonzero a.e. on each $F\in\mathscr{F}_T$. For all $\underline{v}_T\in\underline{U}_T^k$, letting $\underline{u}_T = \underline{v}_T$ in (100), t can be easily checked that the following coercivity property holds:

$$\min(1,\hat{\tau}_T\mu_0)\|\underline{v}_T\|_{\beta,\mu,T}^2 \le \mathrm{a}_{\beta,\mu,T}(\underline{v}_T,\underline{v}_T). \tag{102}$$

5.2.2 Global advective-reactive bilinear form

The global advective-reactive bilinear form is given by

$$\mathrm{a}_{\beta,\mu,h}(\underline{u}_h,\underline{v}_h) := \sum_{T\in\mathscr{T}_h}\mathrm{a}_{\beta,\mu,T}(\underline{u}_T,\underline{v}_T) + \frac{1}{2}\sum_{F\in\mathscr{F}_h^{\mathrm{b}}}(|\boldsymbol{\beta}\cdot\mathbf{n}|u_F,v_F)_F, \tag{103}$$

where the first term results from the assembly of elementary contributions, while the second term is responsible for the enforcement of the boundary condition on the inflow portion of $\partial\Omega$. Define the global advective-reactive norm such that, for all $\underline{v}_h\in\underline{U}_h^k$,

$$\|\underline{v}_h\|_{\beta,\mu,h}^2 := \sum_{T\in\mathscr{T}_h}\|\underline{v}_T\|_{\beta,\mu,T}^2 + \frac{1}{2}\sum_{F\in\mathscr{F}_h^{\mathrm{b}}}\||\boldsymbol{\beta}\cdot\mathbf{n}|^{1/2}v_F\|_F^2.$$

The following coercivity result for $\mathrm{a}_{\beta,\mu,h}$ follows from (102): For all $\underline{v}_h\in\underline{U}_h^k$

$$\min_{T\in\mathscr{T}_h}(1,\hat{\tau}_T\mu_0)\|\underline{v}_h\|_{\beta,\mu,h}^2 \le \mathrm{a}_{\beta,\mu,h}(\underline{v}_h,\underline{v}_h). \tag{104}$$

5.3 Global problem and inf-sup stability

We can now define the global bilinear form $\mathrm{a}_{\kappa,\beta,\mu,h}:\underline{U}_h^k\times\underline{U}_h^k\to\mathbb{R}$ combining the diffusive and advective-reactive contributions defined above:

$$\mathrm{a}_{\kappa,\beta,\mu,h}(\underline{u}_h,\underline{v}_h) := \mathrm{a}_{\kappa,h}(\underline{u}_h,\underline{v}_h) + \mathrm{a}_{\beta,\mu,h}(\underline{u}_h,\underline{v}_h).$$

The HHO approximation of (97) then reads: Find $\underline{u}_h\in\underline{U}_h^k$ such that, for all $\underline{v}_h\in\underline{U}_h^k$,

$$\mathrm{a}_{\kappa,\beta,\mu,h}(\underline{u}_h,\underline{v}_h)=(f,v_h). \tag{105}$$

Let us examine stability. In view of (99) and (104), the bilinear form $\mathrm{a}_{\kappa,\beta,\mu,h}$ is clearly coercive with respect to the norm

$$\|\underline{v}_h\|_{\flat,h}^2 := \|\underline{v}_h\|_{\kappa,h}^2 + \|\underline{v}_h\|_{\beta,\mu,h}^2,$$

which guarantees that problem (105) has a unique solution. This norm, however, does not convey any information on the discrete advective derivative. A stronger stability result is stated in the following lemma, where we consider the augmented norm

$$\|\underline{v}_h\|_{\sharp,h}^2 := \|\underline{v}_h\|_{\flat,h}^2 + \sum_{T\in\mathscr{T}_h,\hat{\beta}_T\neq 0} h_T\hat{\beta}_T^{-1}\|\mathbf{G}_{\beta,\underline{v}_T}^k\|_T^2,$$

with $\hat{\beta}_T := \|\beta\|_{L^\infty(T)^d}$ denoting the reference velocity on T.

Lemma 7 (Inf-sup stability of $\mathrm{a}_{\kappa,\beta,\mu,h}$). *Assume that $\zeta\geq 1$ and that, for all $T\in\mathscr{T}_h$,*

$$h_T\max(\mathrm{L}_{\beta,T},\mu_0)\leq\hat{\beta}_T. \tag{106}$$

Then, there exists a real number $C>0$, independent of h,κ,β and μ, but possibly depending on d, ρ, and k such that, for all $\underline{w}_h\in\underline{U}_h^k$,

$$C\min_{T\in\mathscr{T}_h}(1,\hat{\tau}_T\mu_0)\|\underline{w}_h\|_{\sharp,h}\leq\sup_{\underline{v}_h\in\underline{U}_h^k\setminus\{\underline{0}_h\}}\frac{\mathrm{a}_{\kappa,\beta,\mu,h}(\underline{w}_h,\underline{v}_h)}{\|\underline{v}_h\|_{\sharp,h}}.$$

Remark 13 (Condition (106)*).* Condition (106) means (i) that the advective field is well-resolved by the mesh and (ii) that reaction is not dominant.

5.4 Convergence

For each mesh element $T\in\mathscr{T}_h$, we introduce the local Péclet number such that

$$\mathrm{Pe}_T := \max_{F\in\mathscr{F}_T}\frac{h_F\|\beta_{|F}\cdot\mathbf{n}_{TF}\|_{L^\infty(F)}}{\kappa_F},$$

where $\kappa_F := \min_{T\in\mathscr{T}_F}\kappa_T$. For the mesh elements where diffusion dominates we have $\mathrm{Pe}_T\leq h_T$, for those where advection dominates we have $\mathrm{Pe}_T\geq 1$, while intermediate regimes correspond to $\mathrm{Pe}_T\in(h_T,1)$.

The following error estimate accounts for the variation of the convergence rate according to the value of the local Péclet number, showing that diffusion-dominated elements contribute with a term in $\mathscr{O}(h_T^{k+1})$ (as for a pure diffusion problem), whereas convection-dominated elements contribute with a term in $\mathscr{O}(h_T^{k+1/2})$ (as for a pure advection problem).

Theorem 8 (Energy error estimate). *Let u solve* (97) *and $\underline{u}_h$ solve* (105). *Under the assumptions of Lemma 7, and further assuming the regularity $u_{|T} \in H^{k+2}(T)$ for all $T \in \mathcal{T}_h$, there exists a real number $C > 0$ independent of h, κ, β, and μ, but possibly depending on ρ, d, and k, such that*

$$C \min_{T\in\mathcal{T}_h}(1, \hat{\tau}_T\mu_0)\|\hat{\underline{u}}_h - \underline{u}_h\|_{\sharp,h} \le \left\{ \sum_{T\in\mathcal{T}_h} \left[\left(\kappa_T \|u\|^2_{H^{k+2}(T)} + \hat{\tau}_T^{-1}\|u\|^2_{H^{k+1}(T)} \right) h_T^{2(k+1)} \right.\right.$$
$$\left.\left. + \hat{\beta}_T \min(1, \mathrm{Pe}_T)\|u\|^2_{H^{k+1}(T)} h_T^{2k+1} \right] \right\}^{1/2}.$$

Remark 14 (Extension to locally vanishing diffusion). It has been showed in [15] that the error estimate of Theorem 8 extends to locally vanishing diffusion provied that we conventionally set $\mathrm{Pe}_T = +\infty$ for any element $T \in \mathcal{T}_h$ such that $\kappa_F = 0$ for some $F \in \mathcal{F}_T$.

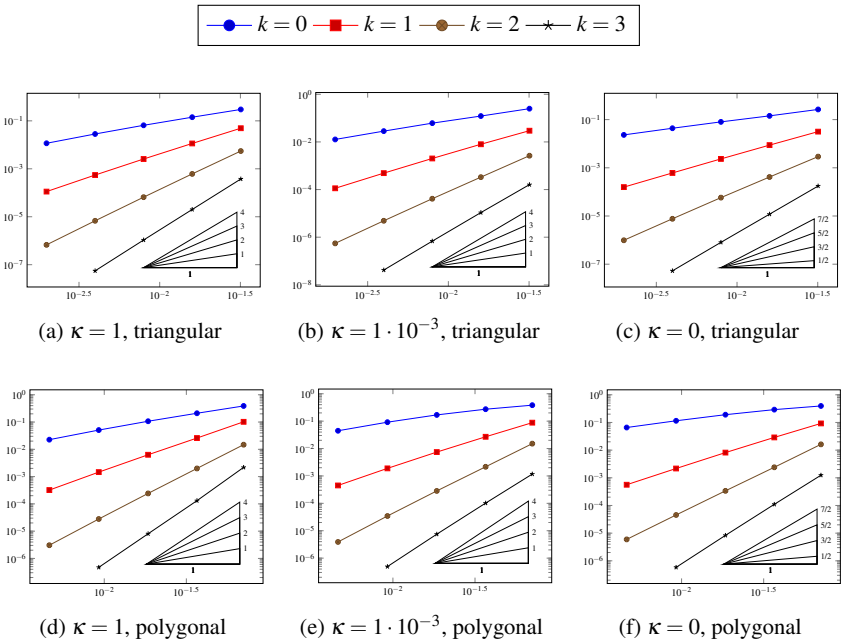
5.5 Numerical example

To illustrate the performance of the HHO method, we solve in the unit square $\Omega = (0,1)^2$ the Dirichlet problem corresponding to the solution (78) with $\beta(\mathbf{x}) = (1/2 - x_2, x_1 - 1/2)$, $\mu \equiv 1$, and a uniform diffusion coefficient κ taking values in $\{1, 1\cdot 10^{-3}, 0\}$. We take triangular and predominantly hexagonal meshes, as depicted in Figures 1a and 1c respectively. The convergence results are depicted in Figure 5.5. We observe that the convergence rate decreases with κ, slightly less than the half order predicted by the error estimate of Theorem 8.

Acknowledgements This work was funded by Agence Nationale de la Recherche grant HHOMM (ref. ANR-15-CE40-0005-01).

References

1. J. Aghili, S. Boyaval, and D. A. Di Pietro. Hybridization of mixed high-order methods on general meshes and application to the Stokes equations. *Comput. Meth. Appl. Math.*, 15(2):111–134, 2015.
2. M. Bebendorf. A note on the Poincaré inequality for convex domains. *Z. Anal. Anwendungen*, 22(4):751–756, 2003.
3. L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. D. Marini, and A. Russo. Basic principles of virtual element methods. *Math. Models Methods Appl. Sci.*, 199(23):199–214, 2013.
4. D. Boffi, M. Botti, and D. A. Di Pietro. A nonconforming high-order method for the Biot problem on general meshes. *SIAM J. Sci. Comput.*, 38(3):A1508–A1537, 2016.



(a) $\kappa = 1$, triangular (b) $\kappa = 1 \cdot 10^{-3}$, triangular (c) $\kappa = 0$, triangular

(d) $\kappa = 1$, polygonal (e) $\kappa = 1 \cdot 10^{-3}$, polygonal (f) $\kappa = 0$, polygonal

Fig. 9: $\|\underline{I}_h^k u - \underline{u}_h\|_{\sharp,h}$ vs. h for the test case of Section 5.5.

5. D. Boffi and D. A. Di Pietro. Unified formulation and analysis of mixed and primal discontinuous skeletal methods on polytopal meshes, 2016. Preprint arXiv:1609.04601 [math.NA].
6. M. Botti, D. A. Di Pietro, and P. Sochala. A Hybrid High-Order method for nonlinear elasticity, 2016. Submitted.
7. P. Castillo, B. Cockburn, I. Perugia, and D. Schötzau. An a priori error analysis of the local discontinuous Galerkin method for elliptic problems. *SIAM J. Numer. Anal.*, 38:1676–1706, 2000.
8. F. Chave, D. A. Di Pietro, F. Marche, and F. Pigeonneau. A hybrid high-order method for the Cahn–Hilliard problem in mixed form. *SIAM J. Numer. Anal.*, 54(3):1873–1898, 2016.
9. M. Cicuttin, D. A. Di Pietro, and A. Ern. Implementation of Discontinuous Skeletal methods on arbitrary-dimensional, polytopal meshes using generic programming, 2017. Submitted.
10. B. Cockburn, D. A. Di Pietro, and A. Ern. Bridging the Hybrid High-Order and Hybridizable Discontinuous Galerkin methods. *ESAIM: Math. Model. Numer. Anal.*, 50(3):635–650, 2016.
11. B. Cockburn, J. Gopalakrishnan, and R. Lazarov. Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems. *SIAM J. Numer. Anal.*, 47(2):1319–1365, 2009.
12. G. Dahlquist. Convergence and stability in the numerical integration of ordinary differential equations. *Math. Scand.*, 4:33–53, 1956.
13. D. A. Di Pietro and J. Droniou. A Hybrid High-Order method for Leray–Lions elliptic equations on general meshes. *Math. Comp.*, 2016. Published online. DOI: 10.1090/mcom/3180.
14. D. A. Di Pietro and J. Droniou. $W^{s,p}$-approximation properties of elliptic projectors on polynomial spaces, with application to the error analysis of a Hybrid High-Order discretisation of Leray–Lions problems. *Math. Models Methods Appl. Sci.*, 2017. Published online. DOI: 10.1142/S0218202517500191.
15. D. A. Di Pietro, J. Droniou, and A. Ern. A discontinuous-skeletal method for advection-diffusion-reaction on general meshes. *SIAM J. Numer. Anal.*, 53(5):2135–2157, 2015.

16. D. A. Di Pietro and A. Ern. Discrete functional analysis tools for discontinuous Galerkin methods with application to the incompressible Navier-Stokes equations. *Math. Comp.*, 79(271):1303–1330, 2010.
17. D. A. Di Pietro and A. Ern. *Mathematical aspects of discontinuous Galerkin methods*, volume 69 of *Mathématiques & Applications*. Springer-Verlag, Berlin, 2012.
18. D. A. Di Pietro and A. Ern. Equilibrated tractions for the Hybrid High-Order method. *C. R. Acad. Sci. Paris, Ser. I*, 353:279–282, 2015.
19. D. A. Di Pietro and A. Ern. A hybrid high-order locking-free method for linear elasticity on general meshes. *Comput. Meth. Appl. Mech. Engrg.*, 283:1–21, 2015.
20. D. A. Di Pietro and A. Ern. Arbitrary-order mixed methods for heterogeneous anisotropic diffusion on general meshes. *IMA J. Numer. Anal.*, 37(1):40–63, 2016.
21. D. A. Di Pietro, A. Ern, and J.-L. Guermond. Discontinuous Galerkin methods for anisotropic semidefinite diffusion with advection. *SIAM J. Numer. Anal.*, 46(2):805–831, 2008.
22. D. A. Di Pietro, A. Ern, and S. Lemaire. An arbitrary-order and compact-stencil discretization of diffusion on general meshes based on local reconstruction operators. *Comput. Meth. Appl. Math.*, 14(4):461–472, 2014.
23. D. A. Di Pietro, A. Ern, and S. Lemaire. *Building bridges: Connections and challenges in modern approaches to numerical partial differential equations*, chapter A review of Hybrid High-Order methods: formulations, computational aspects, comparison with other methods, pages 205–236. Springer, 2016.
24. D. A. Di Pietro, A. Ern, A. Linke, and F. Schieweck. A discontinuous skeletal method for the viscosity-dependent Stokes problem. *Comput. Meth. Appl. Mech. Engrg.*, 306:175–195, 2016.
25. D. A. Di Pietro and S. Krell. A Hybrid High-Order method for the steady incompressible Navier–Stokes problem, 2016. Preprint arXiv:1607.08159 [math.NA].
26. D. A. Di Pietro and S. Lemaire. An extension of the Crouzeix–Raviart space to general meshes with application to quasi-incompressible linear elasticity and Stokes flow. *Math. Comp.*, 84(291):1–31, 2015.
27. D. A. Di Pietro and R. Specogna. An a posteriori-driven adaptive Mixed High-Order method with application to electrostatics. *J. Comput. Phys.*, 326(1):35–55, 2016.
28. R. Eymard, T. Gallouët, and R. Herbin. Discretization of heterogeneous and anisotropic diffusion problems on general nonconforming meshes. SUSHI: a scheme using stabilization and hybrid interfaces. *IMA J. Numer. Anal.*, 30(4):1009–1043, 2010.
29. G. Fichera. Asymptotic behaviour of the electric field and density of the electric charge in the neighbourhood of singular points of a conducting surface. *Russian Math. Surveys*, 30(3):107, 1975.
30. P. Grisvard. *Singularities in Boundary Value Problems*. Masson, Paris, 1992.
31. R. Herbin and F. Hubert. Benchmark on discretization schemes for anisotropic diffusion problems on general grids. In R. Eymard and J.-M. Hérard, editors, *Finite Volumes for Complex Applications V*, pages 659–692. John Wiley & Sons, 2008.
32. O. A. Karakashian and F. Pascal. A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems. *SIAM J. Numer. Anal.*, 41(6):2374–2399, 2003.
33. P. D. Lax and A. N. Milgram. Parabolic equations. In *Contributions to the theory of partial differential equations*, Annals of Mathematics Studies, no. 33, pages 167–190. Princeton University Press, Princeton, N. J., 1954.
34. J. Leray and J.-L. Lions. Quelques résultats de Višik sur les problèmes elliptiques non linéaires par les méthodes de Minty-Browder. *Bull. Soc. Math. France*, 93:97–107, 1965.
35. G. J. Minty. On a "monotonicity" method for the solution of non-linear equations in Banach spaces. *Proc. Nat. Acad. Sci. U.S.A.*, 50:1038–1041, 1963.
36. L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5:286–292 (1960), 1960.
37. R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques*. Stuttgart, 1996.
38. M. Vohralík. A posteriori error estimates for lowest-order mixed finite element discretizations of convection-diffusion-reaction equations. *SIAM J. Numer. Anal.*, 45(4):1570–1599, 2007.