

# Algebraic and discretization error estimation by equilibrated fluxes for discontinuous Galerkin methods on nonmatching grids\*

Vít Dolejší<sup>†</sup>

Ivana Šebestová<sup>†</sup>

Martin Vohralík<sup>§</sup>

June 14, 2014

## Abstract

We derive a posteriori error estimates for the discontinuous Galerkin method applied to the Poisson equation. We allow for a variable polynomial degree and simplicial meshes with hanging nodes and propose an approach allowing for simple (nonconforming) flux reconstructions in such a setting. We take into account the algebraic error stemming from the inexact solution of the associated linear systems and propose local stopping criteria for iterative algebraic solvers. An algebraic error flux reconstruction is introduced in this respect. Guaranteed reliability and local efficiency are proven. We next propose an adaptive strategy combining both adaptive mesh refinement and adaptive stopping criteria. At last, we detail a form of the estimates avoiding any practical reconstruction of a flux and only working with the approximate solution, which simplifies greatly their evaluation. Numerical experiments illustrate a tight control of the overall error, good prediction of the distribution of both the discretization and algebraic error components, and efficiency of the adaptive strategy.

**Key words:** linear diffusion problems, discontinuous Galerkin method, a posteriori estimate, flux reconstruction, distribution of the error, error components, stopping criteria, adaptivity, hanging nodes, simple evaluation

## 1 Introduction

We consider the second-order pure diffusion problem

$$-\Delta u = f \quad \text{in } \Omega, \tag{1.1a}$$

$$u = 0 \quad \text{on } \partial\Omega, \tag{1.1b}$$

where  $\Omega \subset \mathbb{R}^d$ ,  $d = 2$  or  $3$ , is a polygonal (polyhedral) domain and  $f$  a source term. The homogeneous Dirichlet boundary condition (1.1b) is only considered for the sake of simplicity. Hereafter,  $u$  is termed the *potential* and  $-\nabla u$  the *flux*. The model problem (1.1) admits a unique weak solution  $u$  provided that  $f \in H^{-1}(\Omega)$ . However, for the purpose of our a posteriori analysis, we assume that  $f \in L^2(\Omega)$ .

The total error in a computational approximation of (1.1) consists of two parts: the *discretization error*, which arises due to the transition from the infinite-dimensional mathematical model to a finite-dimensional numerical approximation, and the *algebraic error*, which arises due to inaccurate solution of the underlying algebraic systems. Despite a large number of papers dedicated to error estimates, most of them *do not* take into account the algebraic error. Among those that do, let us cite [9, 4, 8, 7, 23, 19, 17, 5]. As pointed out in these references, see also [6] and [22, Chapter 5], knowledge of the algebraic error is of significant importance

\*This work was supported by the ERC-CZ project MORE “MOdelling REvisited + MOdel REduction” LL1202. The research of V. Dolejší was supported by the Grant No. 13-00522S of the Czech Science Foundation and the membership in the Nečas Center for Mathematical Modeling (<http://ncmm.karlin.mff.cuni.cz>). The research of I. Šebestová was supported by the project MathMAC - University center for mathematical modeling, applied analysis and computational mathematics of the Charles University in Prague.

<sup>†</sup>Department of Numerical Mathematics, Charles University in Prague, Sokolovská 83, 186 75 Praha 8, Czech Republic ([dolejsi@karlin.mff.cuni.cz](mailto:dolejsi@karlin.mff.cuni.cz), [ivana.sebestova@karlin.mff.cuni.cz](mailto:ivana.sebestova@karlin.mff.cuni.cz)).

<sup>§</sup>INRIA Paris-Rocquencourt, B.P. 105, 78153 Le Chesnay, France ([martin.vohralik@inria.fr](mailto:martin.vohralik@inria.fr)).

for an *efficient numerical solution* of partial differential equations. The key idea is that of balancing of the discretization and algebraic errors through *a posteriori error estimates* and *stopping criteria* for iterative algebraic solvers.

Such an idea has already appeared in [9]. A posteriori error estimates involving both discretization and algebraic error in  $H^1$  and  $L^2$  norms have been derived there for the Poisson equation considering a piecewise linear finite element approximation together with a multigrid algebraic solver. A stopping criterion for the solver has been proposed. The approach is based on a strong stability and the orthogonality property. Numerical examples illustrate its reliability and efficiency. In [4], stopping criteria for the conjugate gradient method with respect to the finite element discretization have been studied. The result is based on a lower bound of the energy norm of the algebraic error of the conjugate gradient method. An extension for non-self-adjoint problems has been carried out in [8]. Goal-oriented a posteriori error analysis for a linear elliptic problem focusing on the multigrid method has been carried out in [23]. This concept has been later applied to the linear elliptic eigenvalue problem in [29]. A linear diffusion problem discretized by a finite volume method with a focus on the conjugate gradient solver is a subject of the study in [19]. Recently, a general framework for adaptive numerical solution of nonlinear partial differential equations of diffusion type has been given in [17]. A posteriori error estimates distinguishing the individual error components together with stopping criteria for both iterative linear and nonlinear solvers have been developed therein. Finally, the recent paper [5] deals with stopping criteria for iterative linear solvers in the context of adaptive finite elements. Guaranteed bounds on the energy norm of the algebraic error for the conjugate gradient method are proved and the (inexact) adaptive finite element method with global stopping criteria using these bounds is proved to converge, which is a very important result.

It has been illustrated in, e.g., [22] that even for simple model problems, the *local distribution* of the discretization and algebraic errors can differ significantly. It may happen that the overall discretization error dominates the overall algebraic one, whereas locally, it is just the opposite. Congruently, the stopping criteria proposed in [19, 17] are based on the *local balancing* of the discretization and algebraic errors. This typically leads to *local efficiency* of the estimates even in presence of the algebraic error. Thus, the adaptive computational process can be carried out safely including *adaptive mesh refinement*.

The presence of hanging nodes in the computational mesh is rather seldom allowed in a posteriori error analysis. In the context of the so-called equilibrated fluxes, it has been done by prescription of the local degrees of freedom of the flux for advection–diffusion–reaction problems on nonmatching simplicial meshes in [15]. Pure diffusion problems with non-uniform polynomial degree of the approximate solution are treated in [2, 3]. In [16], almost arbitrary polygonal/polyhedral meshes are considered and the flux is constructed by solving local Neumann problems. All these approaches require existence of a matching submesh of the given nonmatching mesh to obtain a  $\mathbf{H}(\text{div}, \Omega)$ -conforming flux reconstruction, whereas an approach not requiring a matching submesh is presented in [27, Section 6], following an idea from [35].

In this paper, we aim at extending the existing theory of equilibrated flux a posteriori error estimates for *discontinuous Galerkin methods* (DGMs) by including the *algebraic error*, permitting *hanging nodes* (without the necessity to construct any submesh), and allowing the *polynomial degree* of the approximate solution to *vary* locally. Our error estimates are derived in the broken  $H^1$ -seminorm with the aid of a flux reconstruction that is constructed in broken Raviart–Thomas–Nédélec (RTN) space; unlike the existing approaches in the literature, our proposed approach operates on the *original nonmatching mesh only*. Therefore, our flux reconstructions generally fail to belong to  $\mathbf{H}(\text{div}, \Omega)$  and flux-nonconformity estimators appear. We also focus on *facilitating the evaluation* of the estimators; at least for low-order approximations, we present final formulas that can be rewritten in a simple form only featuring the approximate solution, without any appearance of the flux reconstruction. Thus, the flux reconstruction can be considered as a theoretical tool that needs not be factually constructed in practice.

The reconstructed flux consists of the discretization and algebraic components. Following the approach introduced in [19, 17], the algebraic component is constructed directly from the discretization flux reconstruction by performing some additional steps of the iterative algebraic solver. Such a construction is not computationally expensive, as the forward iterations are used at the next step of the algebraic solver, but it does not lead to the exact equilibration property of the flux reconstruction. A remainder term appears which is treated as in [17].

This paper is organized as follows. We introduce the continuous and discrete settings in Section 2. A guaranteed a posteriori error estimate taking into account nonmatching meshes, varying polynomial degrees, and the algebraic error is derived in Section 3. Local adaptive stopping criteria are devised in

Section 4. Section 5 then proves local efficiency (up to a generic constant depending in particular on the shape-regularity and local quasi-uniformity of the mesh and on the degree of the employed polynomials) even in the considered complex setting. The discussion of simple practical implementation of the derived a posteriori estimates is presented in Section 6. Finally, Section 7 demonstrates the tight prediction of the distribution of both the discretization and algebraic errors even on meshes with hanging nodes via numerical experiments and Section 8 concludes the paper.

## 2 Continuous and discrete problems

We set up here our notation and introduce the continuous and discrete problems.

### 2.1 Continuous problem

We use standard notation for the Lebesgue and Sobolev spaces. Specifically, for a given domain  $M \subset \mathbb{R}^d$ ,  $L^2(M)$  denotes the space of square-integrable functions and  $H_0^1(M)$  the space of functions having square-integrable weak derivatives up to the first order and traces vanishing on the boundary. Further,  $(\cdot, \cdot)_M$  denotes the inner product in  $L^2(M)$  or  $[L^2(M)]^d$ ,  $\|\cdot\|_M$  denotes the induced norm, and  $(\cdot, \cdot)_{\partial M}$  denotes  $(d-1)$ -dimensional  $L^2(\partial M)$  inner product on  $\partial M$ . We will omit subscript  $M$  in case  $M = \Omega$ . By  $\mathbf{H}(\text{div}, M) := \{\mathbf{v} \in [L^2(M)]^d; \nabla \cdot \mathbf{v} \in L^2(M)\}$  we denote the space with square-integrable weak divergences, see, e.g., [10] or [28]. Let us introduce the weak formulation of the problem (1.1): Find  $u \in H_0^1(\Omega)$  such that

$$(\nabla u, \nabla v) = (f, v) \quad \forall v \in H_0^1(\Omega). \quad (2.1)$$

### 2.2 Meshes with hanging nodes

We consider a family  $\mathcal{T}_h$  ( $h > 0$ ) of partitions of the closure of  $\Omega$  into a finite number of closed simplices (triangles in 2D and tetrahedra in 3D). We suppose that the simplices have mutually disjoint interiors but we *admit* the presence of the so-called *hanging nodes*: the intersection of two different simplices  $K$  and  $K'$  may either be empty, form a vertex, or an entire edge or face of one of the simplices. We assume that any mesh  $\mathcal{T}_h$  was formed from some initial matching simplicial mesh (without hanging nodes) by subdividing some of its elements (repeatedly) into  $(d+1) + (d-1)^2$  congruent simplices. Thus, for each  $K \in \mathcal{T}_h$  with a face possessing a hanging node, there exists a simplex, called a macro-element, sharing this entire face. Note that macro-elements are not included in the mesh  $\mathcal{T}_h$ . Figure 1, left gives an illustration of an admissible mesh and an example of the macro-element.

Due to a possible presence of hanging nodes, we have to distinguish two types of faces. First, each simplex  $K \in \mathcal{T}_h$  has  $d+1$  faces  $\Gamma$  defining its boundary  $\partial K$ . Second, if a face  $\Gamma$  of some  $K \in \mathcal{T}_h$  contains (a) hanging node(s) then  $\Gamma$  can be split into several sub-faces  $\gamma \subset \Gamma$  where  $\gamma = \partial K \cap \partial K'$  for some  $K' \in \mathcal{T}_h$ ,  $K' \neq K$ . Hence, the symbol  $\Gamma$  denotes an entire face of some  $K \in \mathcal{T}_h$  whereas the symbol  $\gamma$  its part which is a common boundary between two neighboring elements. If  $\Gamma \subset \partial K$  does not contain a hanging node then there exists  $\gamma \subset \partial K$  such that  $\gamma = \Gamma$ , see Figure 1, right.

By  $\mathcal{E}_K$  we denote the set of all faces  $\Gamma \subset \partial K$ , by  $\mathcal{E}_K^{\text{HG}}$  those of them that contain at least one hanging node, by  $\mathcal{E}_K^{\text{I}}$  the faces of  $\mathcal{E}_K$  lying in the interior of  $\Omega$ , and by  $\mathcal{E}_K^{\text{B}}$  the faces of  $\mathcal{E}_K$  lying on the boundary of  $\Omega$ . Additionally, we set

$$\mathcal{E}_K^{\text{HG,N}} := \{\Gamma \in \mathcal{E}_K^{\text{I}} \setminus \mathcal{E}_K^{\text{HG}}, \Gamma \not\subset \Gamma' \in \mathcal{E}_{K'}^{\text{HG}}, K' \in \mathcal{T}_h\}, \quad (2.2)$$

which denotes the set of all faces of  $K \in \mathcal{T}_h$  which *do not* possess a hanging node, *neither* are a part of a face with a hanging node of the neighboring element, see Figure 2. Obviously, if  $\Gamma \in \mathcal{E}_K^{\text{HG,N}}$  then there exists  $\gamma = \partial K \cap \partial K'$  for some  $K' \in \mathcal{T}_h$  such that  $\gamma = \Gamma$ .

The previous notations give the identity

$$\bigcup_{K \in \mathcal{T}_h} \partial K = \bigcup_{K \in \mathcal{T}_h} \left\{ \left( \bigcup_{\Gamma \in \mathcal{E}_K^{\text{HG}}} \Gamma \right) \cup \left( \bigcup_{\Gamma \in \mathcal{E}_K^{\text{HG,N}}} \Gamma \right) \cup \left( \bigcup_{\Gamma \in \mathcal{E}_K^{\text{B}}} \Gamma \right) \right\}. \quad (2.3)$$

Obviously, all faces having a hanging node belong to the first union, all interior faces without hanging nodes and not being a part of a face with a hanging node of the neighboring element appear in the second

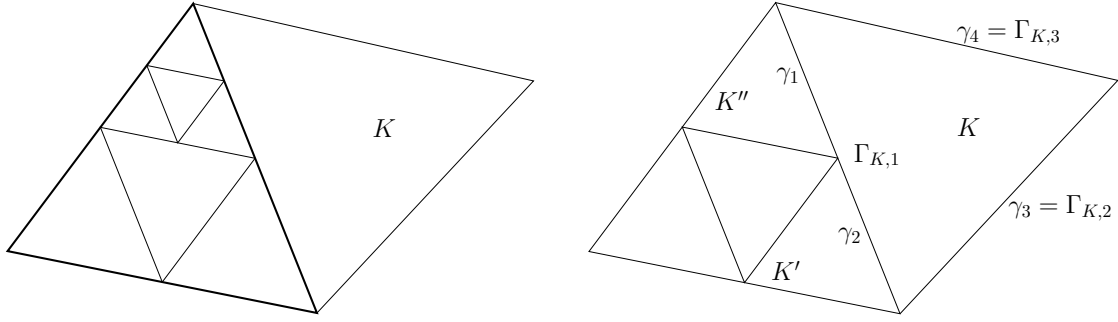


Figure 1: Example of  $K$  having a face with hanging nodes and the macro-element (bold) sharing this face (left). Notation of the symbols  $\Gamma$  and  $\gamma$ : entire faces  $\Gamma_{K,i}$ ,  $i = 1, 2, 3$ , of the element  $K$  with a hanging node and the sub-faces  $\gamma_j$ ,  $j = 1, \dots, 4$ ; obviously  $\Gamma_{K,1} = \gamma_1 \cup \gamma_2$  (right)

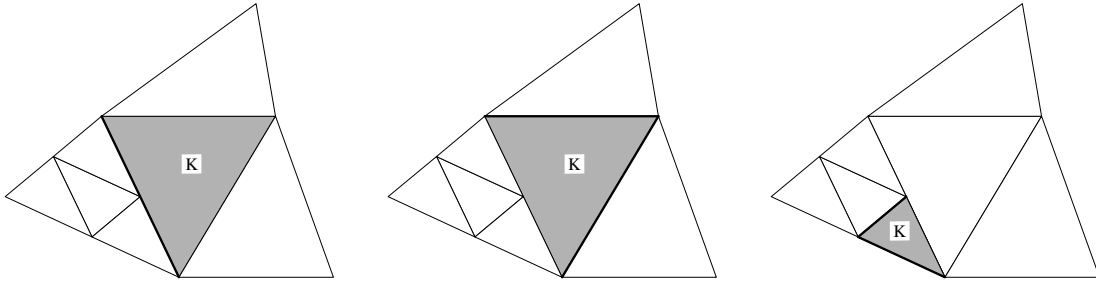


Figure 2: An example of the set of faces (bold lines)  $\mathcal{E}_K^{\text{HG}}$  of an element with a hanging node (left), the set of faces  $\mathcal{E}_K^{\text{HG,N}}$  of an element with a hanging node (center) and the set of faces  $\mathcal{E}_K^{\text{HG,N}}$  of an element whose vertex gives rise to a hanging node of the neighboring element (right)

union two times, and all boundary faces appear in the third union. On the other hand, in general,  $\partial K \neq (\cup_{\Gamma \in \mathcal{E}_K^{\text{HG}}} \Gamma) \cup (\cup_{\Gamma \in \mathcal{E}_K^{\text{HG,N}}} \Gamma) \cup (\cup_{\Gamma \in \mathcal{E}_K^{\text{B}}} \Gamma)$ , see Figure 2, right. Notice finally that boundary faces do not possess hanging nodes, i.e.  $\mathcal{E}_K^{\text{HG}} \cap \mathcal{E}_K^{\text{B}} = \emptyset$ ,  $K \in \mathcal{T}_h$ .

For each inner face  $\gamma$ , we use the notation  $K_\gamma^{\text{L}}$  and  $K_\gamma^{\text{R}}$  for the two elements, called neighbors hereafter, such that  $\gamma = \partial K_\gamma^{\text{L}} \cap \partial K_\gamma^{\text{R}}$ . Hence, for the example pictured in Figure 1, right, we have  $K_{\gamma_1}^{\text{L}} = K''$ ,  $K_{\gamma_1}^{\text{R}} = K$ ,  $K_{\gamma_2}^{\text{L}} = K'$ , and  $K_{\gamma_2}^{\text{R}} = K$ . We define a unit normal vector  $\mathbf{n}_\gamma$  to each  $\gamma$  so that it points out of  $K_\gamma^{\text{L}}$ . We assume that  $\mathbf{n}_\gamma$  for boundary faces  $\gamma$  coincides with the unit outward normal to  $\partial\Omega$ . Note that a face of an element that is divided into several parts due to the presence of (a) hanging node(s) has, in fact, several parallel normal vectors (possibly with different orientation). We denote the set of all sub-faces  $\gamma$  by  $\mathcal{F}_h$  and set  $h_\gamma := \text{diam}(\gamma)$ ,  $\gamma \in \mathcal{F}_h$ . We also use the notation  $h_K := \text{diam}(K)$  for  $K \in \mathcal{T}_h$ ,  $|K|$  for the Lebesgue measure of an element  $K$ , and  $|\partial K|$  for the  $(d-1)$ -dimensional Lebesgue measure of  $\partial K$ .

Let  $\mathcal{T}_K$  stand for the set of the element  $K$  itself, its neighbors, and neighbors are defined above with different meaning all elements of  $\mathcal{T}_h$  that are contained in the macro-elements sharing a complete face with the element  $K$  in case that  $K$  possesses (a) hanging node(s). Further,  $\mathcal{F}_K$  denotes all the faces in this patch and  $\tilde{\mathcal{F}}_K$  stands for the set of faces that share at least a point with  $K$ . All faces visualized in Figure 2 belongs to  $\mathcal{F}_K$ .

We assume that the following conditions are satisfied:

$$\text{shape regularity: } \exists C_s > 0; \frac{h_K}{\rho_K} \leq C_s \quad \forall K \in \mathcal{T}_h, \quad (2.4a)$$

$$\text{local quasi-uniformity: } \exists C_H > 0; h_K \leq C_H h_{K'} \quad \forall K, K' \in \mathcal{T}_h \text{ neighbors}, \quad (2.4b)$$

where  $\rho_K$  denotes the diameter of the largest  $d$ -dimensional ball inscribed into  $K$ .

### 2.3 Broken spaces

We define the so-called broken Sobolev space over the mesh  $\mathcal{T}_h$ ,

$$H^s(\Omega, \mathcal{T}_h) = \{v \in L^2(\Omega); v|_K \in H^s(K) \quad \forall K \in \mathcal{T}_h\}, \quad s \geq 1.$$

We equip it with the norm  $\|v\|_{H^s(\Omega, \mathcal{T}_h)}^2 := \sum_{K \in \mathcal{T}_h} \|v\|_{H^s(K)}^2$ . For  $v \in H^1(\Omega, \mathcal{T}_h)$ , we define the broken gradient  $\nabla_h v$  of  $v$  by  $(\nabla_h v)|_K := \nabla(v|_K)$  for all  $K \in \mathcal{T}_h$  and use the following notation:  $v_\gamma^L$  stands for the trace of  $v|_{K_\gamma^L}$  on  $\gamma$ ,  $v_\gamma^R$  is the trace of  $v|_{K_\gamma^R}$  on  $\gamma$ ,  $\langle v \rangle_\gamma := \frac{1}{2}(v_\gamma^L + v_\gamma^R)$ ,  $[v]_\gamma := v_\gamma^L - v_\gamma^R$ , for an inner face  $\gamma$ . Further, for boundary faces  $\gamma$ , we define  $v_\gamma^L$  as the trace of  $v|_{K_\gamma^L}$  on  $\gamma$ , and  $\langle v \rangle_\gamma := [v]_\gamma := v_\gamma^L$ . If  $[\cdot]_\gamma$  or  $\langle \cdot \rangle_\gamma$  appear in an integral of the form  $\int_\gamma \cdot dS$ , we will omit the subscript  $\gamma$  and write, respectively,  $[\cdot]$  and  $\langle \cdot \rangle$  instead.

To each  $K \in \mathcal{T}_h$  we assign an integer  $p_K \geq 1$  and set  $\mathbf{p} := \{p_K\}_{K \in \mathcal{T}_h}$ . Then, we define the space of discontinuous piecewise polynomials

$$S_h^{\mathbf{p}} = \{v \in L^2(\Omega); v|_K \in \mathbb{P}^{p_K}(K) \quad \forall K \in \mathcal{T}_h\},$$

where  $\mathbb{P}^{p_K}(K)$  is the space of polynomials on  $K$  of total degree at most  $p_K$ . We let  $N := \dim(S_h^{\mathbf{p}})$  and  $N_K := \dim(\mathbb{P}^{p_K}(K))$ .

### 2.4 The discontinuous Galerkin method

We discretize the problem (1.1) with the aid of the *interior penalty discontinuous Galerkin method*, see, e.g., [13] and the references therein. Hence, for  $u_h, v_h \in S_h^{\mathbf{p}}$ , we define the forms

$$a(u_h, v_h) := \sum_{K \in \mathcal{T}_h} (\nabla u_h, \nabla v_h)_K - \sum_{\gamma \in \mathcal{F}_h} (\langle \nabla u_h \rangle \cdot \mathbf{n}_\gamma, [v_h])_\gamma \quad (2.5a)$$

$$- \theta \sum_{\gamma \in \mathcal{F}_h} (\langle \nabla v_h \rangle \cdot \mathbf{n}_\gamma, [u_h])_\gamma + \sum_{\gamma \in \mathcal{F}_h} (\alpha_\gamma h_\gamma^{-1} [u_h], [v_h])_\gamma,$$

$$\ell(v_h) := (f, v_h), \quad (2.5b)$$

where  $\alpha_\gamma > 0$ ,  $\gamma \in \mathcal{F}_h$ , are (sufficiently large) penalty parameters, and the parameter  $\theta \in \{1, -1, 0\}$  corresponds to the symmetric, nonsymmetric, and incomplete variants of the interior penalty DGM, respectively. The discontinuous Galerkin method for problem (1.1) then reads:

$$\text{Find } u_h \in S_h^{\mathbf{p}} \text{ such that } a(u_h, v_h) = \ell(v_h) \quad \forall v_h \in S_h^{\mathbf{p}}. \quad (2.6)$$

### 2.5 Algebraic solution of the linear systems

Let  $\{\varphi_l\}_{l=1 \dots N}$  be a basis of the space  $S_h^{\mathbf{p}}$  such that support of each  $\varphi_l$ ,  $l = 1, \dots, N$ , is just one simplex  $K \in \mathcal{T}_h$ . Then, expressing the solution of (2.6) in this basis,  $u_h = \sum_{l=1}^N U_{h,l} \varphi_l$ , (2.6) can be rewritten in the matrix form as follows:

$$\text{Find } U_h \in \mathbb{R}^N \text{ such that } \mathbb{A}U_h = F, \quad (2.7)$$

where  $\mathbb{A} = \{\mathbb{A}_{kl}\}_{k,l=1 \dots N} := \{a(\varphi_l, \varphi_k)\}_{k,l=1 \dots N}$ ,  $U_h := \{U_{h,l}\}_{l=1 \dots N}$ , and  $F = \{F_k\}_{k=1 \dots N} := \{\ell(\varphi_k)\}_{k=1 \dots N}$ .

Using an iterative algebraic method, the linear algebraic system (2.7) is *not solved exactly*; at  $i$ -th iteration step, we have

$$\mathbb{A}U_h^i = F - R^i, \quad U_h^i, R^i \in \mathbb{R}^N, \quad (2.8)$$

where  $R^i$  is the algebraic residual vector associated with the available approximation  $U_h^i$ , given by

$$R^i := F - \mathbb{A}U_h^i. \quad (2.9)$$

Thus, the solution that we have at our disposal at step  $i$  solves the algebraic system with a perturbed right-hand side.

Let us define the residual function  $r_h^i \in S_h^{\mathbf{p}}$  by  $(r_h^i, \varphi_k) = R_k^i$  for  $k = 1 \dots N$ . Then the system (2.8) represents the following perturbed discontinuous Galerkin problem:

$$\text{Find } u_h^i \in S_h^{\mathbf{p}} \text{ such that } a(u_h^i, v_h) = \ell(v_h) - (r_h^i, v_h) \quad \forall v_h \in S_h^{\mathbf{p}}. \quad (2.10)$$

### 3 Guaranteed error upper bound

In this section, we derive a posteriori error estimate on the error between the approximation  $u_h^i$  available from (2.10) and the unknown weak solution  $u$  of (2.1). The weak solution  $u$  given by (2.1) satisfies  $u \in H_0^1(\Omega)$  and  $-\nabla u \in \mathbf{H}(\text{div}, \Omega)$ , whereas its approximate counterpart does not, i.e.,  $u_h^i \notin H_0^1(\Omega)$  and  $-\nabla u_h^i \notin \mathbf{H}(\text{div}, \Omega)$ . The presented error estimates are based on a potential reconstruction  $\mathcal{I}_{\text{Av}}(u_h^i) \in H_0^1(\Omega)$ ,  $\mathcal{I}_{\text{Av}}(u_h^i) \approx u_h^i$ , and a flux reconstruction  $\mathbf{t}_h^i \in \mathbf{RTN}_l(\mathcal{T}_h) \subset \mathbf{H}(\text{div}, \Omega)$ ,  $\mathbf{t}_h^i \approx -\nabla u_h^i$ , where  $\mathbf{RTN}_l(\mathcal{T}_h)$  is the broken Raviart–Thomas–Nédélec space given by (3.4) below.

#### 3.1 Potential reconstruction based on an averaging interpolation operator

We construct here a  $H_0^1(\Omega)$ -conforming piecewise polynomial interpolation of the discontinuous piecewise polynomial approximate solution  $u_h^i$  on the given nonmatching mesh  $\mathcal{T}_h$ . We follow the approach based on averaging from [20], where the construction has been done for uniform polynomial degree over a mesh possibly containing hanging nodes. In [3], an extension for a varying polynomial degree considering a matching submesh has been carried out. Our proposed approach operates on the *original nonmatching mesh only* and extends that of [20].

For  $v_h \in S_h^p$ , we intend to define a conforming interpolation  $\mathcal{I}_{\text{Av}}(v_h)$  having a polynomial degree in the interior of each element  $K \in \mathcal{T}_h$  equal to maximum of polynomial degrees of  $v_h$  in some neighborhood. Namely, for elements *without* a face being a part of a face with a hanging node, the maximum is taken over neighboring elements. For elements *with* such a face, the maximum is also taken over elements sharing the face with the hanging node. Precisely, for each  $\Gamma \in \mathcal{E}_K$ ,  $K \in \mathcal{T}_h$ , define the value

$$p_\Gamma := \max\{p_{K'}; |\Gamma \cap \partial K'| > 0, K' \in \mathcal{T}_h\}$$

denoting the maximal polynomial degree on elements sharing the face  $\Gamma$ . Then the polynomial degree of  $\mathcal{I}_{\text{Av}}(v_h)$  on  $K \in \mathcal{T}_h$  is set to

$$\tilde{p}_K := \max\{p_\Gamma; |\Gamma \cap \partial K| > 0, \Gamma \in \mathcal{E}_{K'}, K' \in \mathcal{T}_h\}. \tag{3.1}$$

Illustrations are given in Figure 3, left and center.

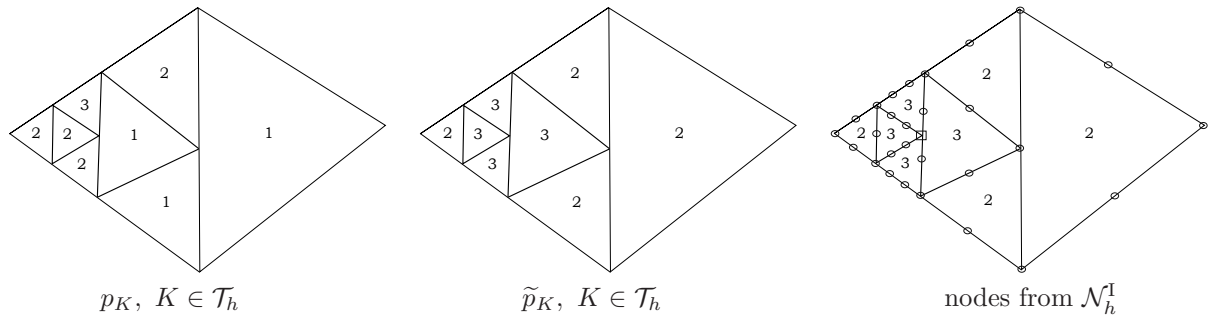


Figure 3: Example of a mesh with polynomial approximation degrees  $p_K$  (left) and the corresponding  $\tilde{p}_K$  (center). The nodes from the set  $\mathcal{N}_h^I$  are marked by  $\circ$ , the node marked by  $\square$  is in  $\tilde{\mathcal{N}}_h^I \setminus \mathcal{N}_h^I$ , excluded by condition i) (right)

Let  $K \in \mathcal{T}_h$  be an element with vertices  $V_K^i \in \mathbb{R}^d$ ,  $i = 0, \dots, d$ , and let  $\tilde{p}_K$  be the corresponding degree given by (3.1). The set of the Lagrange nodes of order  $\tilde{p}_K$  on  $K$  is given by

$$L_K = \left\{ V \in \mathbb{R}^d; V = \frac{i_0}{\tilde{p}_K} V_K^0 + \frac{i_1}{\tilde{p}_K} V_K^1 + \dots + \frac{i_d}{\tilde{p}_K} V_K^d, 0 \leq i_0, \dots, i_d \leq \tilde{p}_K \ \& \ i_0 + \dots + i_d = \tilde{p}_K \right\}, K \in \mathcal{T}_h.$$

In the studied case, due to a possible presence of hanging nodes and of variable polynomial degree, the above Lagrange nodes on interior faces of  $\mathcal{T}_h$  may not coincide. Therefore, the nodes defining  $\mathcal{I}_{\text{Av}}(v_h)$

have to be chosen carefully in order to ensure its  $H_0^1(\Omega)$ -conformity. For this purpose, we define the set of boundary Lagrange nodes  $\mathcal{N}_h^B = \{V \in L_K; V \in \partial\Omega, K \in \mathcal{T}_h\}$  and the set  $\tilde{\mathcal{N}}_h^I$  of Lagrange nodes such that

$$\tilde{\mathcal{N}}_h^I = \{V \in L_K \cap (K^\circ \cup \mathcal{E}_K^{\text{HG}} \cup \mathcal{E}_K^{\text{HG},\text{N}}), K \in \mathcal{T}_h\},$$

where  $K^\circ$  denotes the interior of  $K \in \mathcal{T}_h$ . This means that  $\tilde{\mathcal{N}}_h^I$  does not contain the Lagrange nodes lying on a face which is a part of face with a hanging node of a neighboring element. Next, define a subset of  $\tilde{\mathcal{N}}_h^I$ , denoted by  $\mathcal{N}_h^I$ , excluding from  $\tilde{\mathcal{N}}_h^I$  the nodes possibly violating the conformity of the interpolation:

- i) if  $V \in \tilde{\mathcal{N}}_h^I$  is a hanging node of  $K \in \mathcal{T}_h$  and  $V \notin L_K$  then  $V \notin \mathcal{N}_h^I$ ,
- ii) if  $\Gamma = K \cap K'$  and  $\tilde{p}_K < \tilde{p}_{K'}$ ,  $K \in \mathcal{T}_h$ ,  $K' \in \mathcal{T}_h$ , then  $\mathcal{N}_h^I$  does not contain Lagrange nodes from  $L_{K'} \cap \Gamma$ .

For illustration, in Figure 3, right, the nodes from  $\mathcal{N}_h^I$  are marked by  $\circ$  and the node marked by  $\square$  is excluded from  $\mathcal{N}_h^I$  by condition i) above. The condition ii) means that  $\mathcal{I}_{\text{Av}}(v_h)$  will be a polynomial of possibly *lower degree* on concerned faces  $\Gamma \in \mathcal{E}_{K'}$ . Finally, we define the set

$$\mathcal{N}_h^N = \{V \in L_K, K \in \mathcal{T}_h\} \setminus (\mathcal{N}_h^I \cup \mathcal{N}_h^B), \quad (3.2)$$

which is the subset of the Lagrange nodes lying in interior of some  $\Gamma$  which possible violate the  $H_0^1(\Omega)$ -conformity of the interpolation.

Let  $\mathcal{T}_V = \{K \in \mathcal{T}_h; V \in K\}$  for a given Lagrangian interpolation node  $V$  from  $\mathcal{N}_h^I \cup \mathcal{N}_h^B$ . Then we define the averaging operator  $\mathcal{I}_{\text{Av}}$  in  $V$  by

$$\mathcal{I}_{\text{Av}}(v_h)(V) = \begin{cases} \frac{1}{\text{card}(\mathcal{T}_V)} \sum_{K' \in \mathcal{T}_V} v_h|_{K'}(V), & V \in \mathcal{N}_h^I, \\ 0, & V \in \mathcal{N}_h^B. \end{cases} \quad (3.3)$$

The remaining degrees of freedom are defined as follows, cf. Figure 3, right. If  $V \in \mathcal{N}_h^N$  was not excluded by the condition i) above, then it belongs to some interior face  $\Gamma$ ,  $\Gamma \in \mathcal{E}_K$ ,  $K \in \mathcal{T}_h$ , such that  $\Gamma \subset \Gamma'$ ,  $\Gamma' \in \mathcal{E}_{K'}$ ,  $K' \in \mathcal{T}_h$ . The two situations that can appear are  $\Gamma = \Gamma'$  but  $\tilde{p}_K > \tilde{p}_{K'}$  or  $\Gamma \subsetneq \Gamma'$ .  $\Gamma'$  already contains all the nodes from  $L_{K'} \cap \Gamma'$ . Then we set  $\mathcal{I}_{\text{Av}}(v_h)|_\Gamma$  equal to the polynomial of degree  $\tilde{p}_{K'}$  on  $\Gamma'$  given by these nodal values. If  $V \in \tilde{\mathcal{N}}_h^I$  was excluded by the condition i) above, its value is given by the value  $\mathcal{I}_{\text{Av}}(v_h)|_K(V)$ .

### 3.2 Discretization flux reconstruction

Let  $\mathbf{l} = \{l_K\}_{K \in \mathcal{T}_h}$ ,  $\mathbf{l} = \mathbf{p}$  or  $\mathbf{l} = \mathbf{p} - \mathbf{1}$ , with  $\mathbf{p} - \mathbf{1} = \{p_K - 1\}_{K \in \mathcal{T}_h}$ . Let  $\mathbf{RTN}_{l_K}(K) := [\mathbb{P}^{l_K}(K)]^d + \mathbf{x}^{\mathbb{P}^{l_K}(K)}$  for  $K \in \mathcal{T}_h$ . Our flux reconstructions will be carried in the *broken* Raviart–Thomas–Nédélec space

$$\mathbf{RTN}_{\mathbf{l}}(\mathcal{T}_h) := \{\mathbf{v}_h \in [L^2(\Omega)]^d, \mathbf{v}_h|_K \in \mathbf{RTN}_{l_K}(K) \quad \forall K \in \mathcal{T}_h\}. \quad (3.4)$$

Recall that for  $\mathbf{v}_h \in \mathbf{RTN}_{\mathbf{l}}(\mathcal{T}_h)$ , we have  $\nabla \cdot \mathbf{v}_h|_K \in \mathbb{P}^{l_K}(K)$  and  $\mathbf{v}_h \cdot \mathbf{n}|_\Gamma \in \mathbb{P}^{l_\Gamma}(\Gamma)$ ,  $\Gamma \in \mathcal{E}_K$ , see [10] or [28].

**Remark 3.1** (Choice of  $\mathbf{l}$ ). *There are two possibilities how to choose the degree of the broken Raviart–Thomas–Nédélec space. The choice  $\mathbf{l} = \mathbf{p}$  leads to more accurate results and, as we will see later, to zero residual estimator for polynomial  $f$ . On the other hand, the choice  $\mathbf{l} = \mathbf{p} - \mathbf{1}$  is cheaper.*

The  $\mathbf{H}(\text{div}, \Omega)$ -conformity violation gives rise to additional estimators measuring the discontinuity of the normal components of the reconstructed fluxes in our estimates. It happens in two cases:

1. The polynomial degree of the approximate solution in two neighboring elements is different. We could maintain  $\mathbf{H}(\text{div}, \Omega)$ -conformity in this case by increasing the polynomial degree  $\mathbf{l}$ . However, we prefer to exploit the advantage of DGMs, namely the possibility of varying polynomial degrees, without any extra work for flux reconstructions.

2. The mesh  $\mathcal{T}_h$  contains hanging nodes. We could maintain  $\mathbf{H}(\text{div}, \Omega)$ -conformity in this case by introducing a matching simplicial submesh as in [15, 16, 2, 3], again increasing the algorithmic complexity.

We will construct separately a *discretization flux reconstruction*  $\mathbf{d}_h^i$  and an *algebraic error flux reconstruction*  $\mathbf{a}_h^i$ . The first one is prescribed as follows:

**Definition 3.2** (Discretization flux reconstruction). *Let  $u_h^i$  solve (2.10). The discretization flux reconstruction  $\mathbf{d}_h^i \in \mathbf{RTN}_l(\mathcal{T}_h)$  is defined as follows: For all  $K \in \mathcal{T}_h$ , all  $\Gamma \in \mathcal{E}_K$ , and all  $q_h \in \mathbb{P}^{l_K}(\Gamma)$ , we set*

$$(\mathbf{d}_h^i \cdot \mathbf{n}_\Gamma, q_h)_\Gamma := (-\langle \nabla u_h^i \cdot \mathbf{n}_\Gamma \rangle + \alpha h^{-1} [u_h^i], q_h)_\Gamma, \quad (3.5a)$$

and for all  $\mathbf{q}_h \in [\mathbb{P}^{l_K-1}(K)]^d$ , we set

$$(\mathbf{d}_h^i, \mathbf{q}_h)_K := (-\nabla u_h^i, \mathbf{q}_h)_K + \theta \sum_{\Gamma \in \mathcal{E}_K} w_\Gamma (\mathbf{q}_h \cdot \mathbf{n}_\Gamma, [u_h^i])_\Gamma, \quad (3.5b)$$

where  $w_\Gamma := \frac{1}{2}$  for interior faces and  $w_\Gamma := 1$  for boundary faces, the function  $\alpha : \mathcal{F}_h \rightarrow \mathbb{R}$  is defined piecewise by  $\alpha|_\gamma := \alpha_\gamma$ , and the function  $h : \mathcal{F}_h \rightarrow \mathbb{R}$  is defined by  $h|_\gamma := h_\gamma$ .

The reconstruction  $\mathbf{d}_h^i$  has the following property:

**Lemma 3.3** (Divergence of the discretization flux reconstruction). *Let  $K \in \mathcal{T}_h$  be arbitrary and  $\mathbf{d}_h^i$  be given by (3.5). Then*

$$\nabla \cdot \mathbf{d}_h^i|_K = \Pi_{l_K}(f|_K - r_h^i|_K),$$

where  $\Pi_{l_K}$  is the  $L^2(K)$ -orthogonal projection onto polynomials of degree  $l_K$ .

*Proof.* Let  $v_h \in S_h^l$ , with support on  $K$  only, be arbitrary. Using the Green theorem, Definition 3.2, (2.5), and (2.10), we obtain the sequence of equalities

$$(\nabla \cdot \mathbf{d}_h^i, v_h)_K = -(\mathbf{d}_h^i, \nabla v_h)_K + \sum_{\Gamma \in \mathcal{E}_K} (\mathbf{d}_h^i \cdot \mathbf{n}_K, v_h)_\Gamma = a(u_h^i, v_h) = (f, v_h)_K - (r_h^i, v_h)_K.$$

□

### 3.3 Algebraic error flux reconstruction

The algebraic error will be measured using the algebraic error flux reconstruction. We follow the recent work [17], see also the references therein.

**Definition 3.4** (Algebraic error flux reconstruction). *Consider the  $i$ -th step of the iterative algebraic solver, leading to (2.8). Perform additional  $\nu > 0$  steps of the algebraic solver. This gives (2.8) with  $i$  replaced by  $i + \nu$ . Let  $\mathbf{d}_h^i$  and  $\mathbf{d}_h^{i+\nu}$  be the discretization flux reconstructions given by Definition 3.2, with  $i$  replaced by  $i + \nu$  in the second case. We define the algebraic error flux reconstruction by*

$$\mathbf{a}_h^i := \mathbf{d}_h^{i+\nu} - \mathbf{d}_h^i. \quad (3.6)$$

Due to Definition 3.4, we have immediately for all  $K \in \mathcal{T}_h$

$$\nabla \cdot \mathbf{a}_h^i|_K = \begin{cases} \nabla \cdot \mathbf{d}_h^{i+\nu}|_K - \nabla \cdot \mathbf{d}_h^i|_K = \Pi_{p_K} f|_K - r_h^{i+\nu}|_K - \Pi_{p_K} f|_K + r_h^i|_K \\ \quad = r_h^i|_K - r_h^{i+\nu}|_K & \text{for } l = p, \\ \Pi_{l_K}(r_h^i|_K - r_h^{i+\nu}|_K) & \text{for } l = p - 1. \end{cases}$$

Let us finally define the *total flux reconstruction* as the sum of the discretization and the algebraic error flux reconstruction,

$$\mathbf{t}_h^i := \mathbf{d}_h^i + \mathbf{a}_h^i. \quad (3.7)$$

Then we have

$$\nabla \cdot \mathbf{t}_h^i|_K = \Pi_{l_K}(f|_K - r_h^{i+\nu}|_K) \quad \forall K \in \mathcal{T}_h, l \in \{p-1, p\}. \quad (3.8)$$



**Remark 3.5** (Exact equilibration). In [19, Section 7.3], another method for construction of the algebraic error flux reconstruction has been proposed. It is more precise, leading to the exact equilibration  $\nabla \cdot \mathbf{t}_h^i|_K = \Pi_{L^2} f|_K$  instead of (3.8), but is more costly. On the contrary, in the present approach, the algebraic error flux reconstruction is constructed simply by (3.6), while the information gained by performing some additional steps of the algebraic solver is used in the next algebraic solver iteration.

**Remark 3.6** (Rounding errors). The algebraic error comprises the error due to premature stopping of the algebraic solver and due to rounding errors arising in finite precision arithmetics. Let us stress that the effects of the latter are also considered in the analysis. Indeed, as in [19, 17], the present analysis does not hinge on any assumption of equality satisfied by the approximate solution  $u_h^i$  (and is also not linked to any specific linear solver); instead, for any approximation  $u_h^i$ , the residual vector  $R^i$  and consequently the residual function  $r_h^i$  are prescribed (see (2.9)) so that (2.10) is satisfied. The price we pay for this is the presence of the additional terms  $\eta_{\text{rem},K}^i$  defined by (3.13e) below and of the balancing criterion (4.1a). A rounding error a priori analysis for the preconditioned conjugate gradient method has been done in [33], whereas the numerical stability of the GMRES method based on the modified Gram-Schmidt process has been analyzed in [18, 25] and on the Householder orthogonalization in [14].

### 3.4 Guaranteed and fully computable a posteriori error estimate

In the sequel we will use the following inequalities: The Poincaré inequality reads

$$\forall K \in \mathcal{T}_h, \quad \|\varphi - \Pi_0 \varphi\|_K \leq \frac{h_K}{\pi} \|\nabla \varphi\|_K \quad \forall \varphi \in H^1(K), \quad (3.9)$$

where the constant  $\frac{1}{\pi}$  appears due to the convexity of simplices; see [26]. The Friedrichs inequality reads

$$\|\varphi\| \leq C_{F,\Omega} \|\nabla \varphi\| \quad \forall \varphi \in H_0^1(\Omega). \quad (3.10)$$

The constant  $C_{F,\Omega}$  can be estimated in the following way, see [30]:

$$C_{F,\Omega} \leq \frac{1}{\pi} \left( \frac{1}{a_1^2} + \dots + \frac{1}{a_d^2} \right)^{-\frac{1}{2}}, \quad (3.11)$$

where  $a_i, i = 1 \dots d$ , are the lengths of the edges of a cuboid in which the domain  $\Omega$  is contained. We will also use the trace inequality

$$\forall K \in \mathcal{T}_h, \forall \Gamma \in \mathcal{E}_K, \quad \|\varphi - \Pi_{0,\Gamma} \varphi\|_\Gamma \leq C_{\Gamma,K} h_\Gamma^{\frac{1}{2}} \|\nabla \varphi\|_K \quad \forall \varphi \in H^1(K), \quad (3.12)$$

where  $\Pi_{0,\Gamma}$  is the  $L^2(\Gamma)$ -orthogonal projection onto  $\mathbb{P}^0(\Gamma)$ . The constant  $C_{\Gamma,K}$  has been estimated in [24, Lemma 3.5] as follows:

$$C_{\Gamma,K} \leq \left( C_{s,d} \frac{|\Gamma| h_K^2}{|K| h_\Gamma} \right)^{\frac{1}{2}},$$

where  $C_{s,d} \approx 0,77708$  for a triangle and  $C_{s,d} \approx 3,84519$  for a tetrahedron.

Now, we are ready to state the main theorem concerning the error upper bound. First, we define different error estimators. Consider an  $i$ -th iteration step of the algebraic solver leading to (2.10). For an arbitrary  $K \in \mathcal{T}_h$ , define

$$H_0^1(\Omega)\text{-nonconformity estimator:} \quad \eta_{\text{PNC},K}^i := \|\nabla(u_h^i - \mathcal{I}_{\text{Av}}(u_h^i))\|_K, \quad (3.13a)$$

$$\text{residual estimator:} \quad \eta_{\text{R},K}^i := \frac{h_K}{\pi} \|f - \nabla \cdot \mathbf{t}_h^i - r_h^{i+\nu}\|_K, \quad (3.13b)$$

$$\mathbf{H}(\text{div}, \Omega)\text{-nonconformity estimator:} \quad \eta_{\text{FNC},K}^i := \sum_{\Gamma \in \mathcal{E}_K^{\text{HG}} \cup \mathcal{E}_K^{\text{HG},\text{N}}} \bar{w}_\Gamma C_{\Gamma,K} h_\Gamma^{\frac{1}{2}} \|[\mathbf{t}_h^i \cdot \mathbf{n}_\Gamma]\|_\Gamma, \quad (3.13c)$$

$$\text{flux estimator:} \quad \eta_{\text{F},K}^i := \|\nabla u_h^i + \mathbf{t}_h^i\|_K, \quad (3.13d)$$

$$\text{algebraic remainder estimator:} \quad \eta_{\text{rem},K}^i := C_{F,\Omega} \|r_h^{i+\nu}\|_K, \quad (3.13e)$$

where  $\bar{w}_\Gamma := 1$  for faces from  $\mathcal{E}_K^{\text{HG}}$  and  $\bar{w}_\Gamma := \frac{1}{2}$  for faces from  $\mathcal{E}_K^{\text{HG},\text{N}}$ .

**Remark 3.7** (Algebraic remainder). *Let us point out that the constant  $C_{F,\Omega}$  in (3.13e) can be quite large, scaling like  $h_\Omega$  for regularly-shaped domains  $\Omega$ , see (3.11). It, however, only appears in the algebraic remainder estimator, which will be made small enough (see (4.1a) and (4.2a) in Section 4).*

**Theorem 3.8** (Guaranteed and fully computable a posteriori error estimate). *Let  $u \in H_0^1(\Omega)$  be the weak solution given by (2.1). Let an  $i$ -th algebraic solver step be given and let  $u_h^i \in S_h^p$  be the DGM output given by (2.10). Consider  $\nu > 0$  additional algebraic solver steps and let  $\mathbf{t}_h^i$  be the total flux reconstruction given by (3.7). Then*

$$\begin{aligned} \|\nabla_h(u - u_h^i)\| &\leq \left\{ \sum_{K \in \mathcal{T}_h} (\eta_{\text{PNC},K}^i)^2 + \left\{ \left\{ \sum_{K \in \mathcal{T}_h} (\eta_{\text{R},K}^i + \eta_{\text{F},K}^i + \eta_{\text{FNC},K}^i)^2 \right\}^{\frac{1}{2}} + \left\{ \sum_{K \in \mathcal{T}_h} (\eta_{\text{rem},K}^i)^2 \right\}^{\frac{1}{2}} \right\}^2 \right\}^{\frac{1}{2}} \\ &=: \eta^i. \end{aligned} \quad (3.14)$$

In order to prove Theorem 3.8, we recall the abstract energy error estimate (see [21, Lemma 4.4]):

**Lemma 3.9** (Abstract energy norm estimate). *Let  $u$  be the solution of (2.1) and let  $u_h \in H^1(\Omega, \mathcal{T}_h)$  be arbitrary. Then*

$$\|\nabla_h(u - u_h)\|^2 \leq \inf_{s \in H_0^1(\Omega)} \|\nabla_h(u_h - s)\|^2 + \sup_{\varphi \in H_0^1(\Omega), \|\nabla\varphi\|=1} (\nabla_h(u - u_h), \nabla\varphi)^2. \quad (3.15)$$

*Proof of Theorem 3.8.* Taking  $u_h := u_h^i$  and bounding the first term on the right-hand side of (3.15) using  $\mathcal{I}_{\text{Av}}(u_h^i)$  together with (2.1) gives

$$\|\nabla_h(u - u_h^i)\|^2 \leq \|\nabla_h(u_h^i - \mathcal{I}_{\text{Av}}(u_h^i))\|^2 + \sup_{\varphi \in H_0^1(\Omega), \|\nabla\varphi\|=1} \left\{ \sum_{K \in \mathcal{T}_h} \{(f, \varphi)_K - (\nabla u_h^i, \nabla\varphi)_K\} \right\}^2. \quad (3.16)$$

Add and subtract  $\{(\mathbf{t}_h^i, \nabla\varphi)_K + (r_h^{i+\nu}, \varphi)_K\}$  in (3.16) and employ the Green theorem on each  $K \in \mathcal{T}_h$  to obtain

$$\begin{aligned} \|\nabla_h(u - u_h^i)\|^2 &\leq \|\nabla_h(u_h^i - \mathcal{I}_{\text{Av}}(u_h^i))\|^2 + \sup_{\varphi \in H_0^1(\Omega), \|\nabla\varphi\|=1} \left\{ \sum_{K \in \mathcal{T}_h} \{(f - \nabla \cdot \mathbf{t}_h^i - r_h^{i+\nu}, \varphi)_K \right. \\ &\quad \left. + (\mathbf{t}_h^i \cdot \mathbf{n}_K, \varphi)_{\partial K} + (r_h^{i+\nu}, \varphi)_K - (\nabla u_h^i + \mathbf{t}_h^i, \nabla\varphi)_K\} \right\}^2. \end{aligned} \quad (3.17)$$

Let us estimate the terms in (3.17) separately.

Using (3.8), the Cauchy–Schwarz inequality, the Poincaré inequality (3.9), (3.13b), and (3.13d) gives

$$\begin{aligned} |(f - \nabla \cdot \mathbf{t}_h^i - r_h^{i+\nu}, \varphi)_K - (\nabla u_h^i + \mathbf{t}_h^i, \nabla\varphi)_K| &\leq \|f - \nabla \cdot \mathbf{t}_h^i - r_h^{i+\nu}\|_K \|\varphi - \Pi_0\varphi\|_K + \|\nabla u_h^i + \mathbf{t}_h^i\|_K \|\nabla\varphi\|_K \\ &\leq \|f - \nabla \cdot \mathbf{t}_h^i - r_h^{i+\nu}\|_K \frac{h_K}{\pi} \|\nabla\varphi\|_K + \|\nabla u_h^i + \mathbf{t}_h^i\|_K \|\nabla\varphi\|_K = (\eta_{\text{R},K}^i + \eta_{\text{F},K}^i) \|\nabla\varphi\|_K. \end{aligned} \quad (3.18)$$

Applying the Cauchy–Schwarz inequality, the Friedrichs inequality (3.10), and (3.13e) yields

$$\sum_{K \in \mathcal{T}_h} (r_h^{i+\nu}, \varphi)_K \leq \|r_h^{i+\nu}\| \|\varphi\| \leq \|r_h^{i+\nu}\| C_{F,\Omega} \|\nabla\varphi\| = \left\{ \sum_{K \in \mathcal{T}_h} (\eta_{\text{rem},K}^i)^2 \right\}^{\frac{1}{2}} \|\nabla\varphi\|. \quad (3.19)$$

Finally, using (2.3), the fact that  $\varphi \in H_0^1(\Omega)$ , that  $([\mathbf{t}_h^i \cdot \mathbf{n}_\Gamma], \Pi_{0,\Gamma}\varphi)_\Gamma = 0$  for  $\Gamma \in \mathcal{E}_K^{\text{HG}} \cup \mathcal{E}_K^{\text{HG},\text{N}}$ ,  $K \in \mathcal{T}_h$ , see (3.5a) and (3.7), and that  $([\mathbf{t}_h^i \cdot \mathbf{n}_\Gamma], \varphi)_\Gamma = 0$  for  $\Gamma \in \mathcal{E}_K^{\text{B}}$ , we can write,

$$\begin{aligned} \sum_{K \in \mathcal{T}_h} (\mathbf{t}_h^i \cdot \mathbf{n}_K, \varphi)_{\partial K} &= \sum_{K \in \mathcal{T}_h} \left( \sum_{\Gamma \in \mathcal{E}_K^{\text{HG}}} ([\mathbf{t}_h^i \cdot \mathbf{n}_\Gamma], \varphi)_\Gamma + \sum_{\Gamma \in \mathcal{E}_K^{\text{HG},\text{N}}} \bar{w}_\Gamma ([\mathbf{t}_h^i \cdot \mathbf{n}_\Gamma], \varphi)_\Gamma \right) \\ &= \sum_{K \in \mathcal{T}_h} \left( \sum_{\Gamma \in \mathcal{E}_K^{\text{HG}}} ([\mathbf{t}_h^i \cdot \mathbf{n}_\Gamma], \varphi - \Pi_{0,\Gamma}\varphi)_\Gamma + \sum_{\Gamma \in \mathcal{E}_K^{\text{HG},\text{N}}} \bar{w}_\Gamma ([\mathbf{t}_h^i \cdot \mathbf{n}_\Gamma], \varphi - \Pi_{0,\Gamma}\varphi)_\Gamma \right). \end{aligned}$$

Further, using the Cauchy–Schwarz inequality, the trace inequality (3.12), and (3.13c), we obtain

$$\begin{aligned} \sum_{K \in \mathcal{T}_h} (\mathbf{t}_h^i \cdot \mathbf{n}_K, \varphi)_{\partial K} &\leq \sum_{K \in \mathcal{T}_h} \left\{ \sum_{\Gamma \in \mathcal{E}_K^{\text{HG}}} C_{\Gamma,K} h_{\Gamma}^{\frac{1}{2}} \|\mathbf{t}_h^i \cdot \mathbf{n}_{\Gamma}\|_{\Gamma} \|\nabla \varphi\|_K + \sum_{\Gamma \in \mathcal{E}_K^{\text{HG},\text{N}}} \bar{w}_{\Gamma} C_{\Gamma,K} h_{\Gamma}^{\frac{1}{2}} \|\mathbf{t}_h^i \cdot \mathbf{n}_{\Gamma}\|_{\Gamma} \|\nabla \varphi\|_K \right\} \\ &= \sum_{K \in \mathcal{T}_h} \eta_{\text{FNC},K}^i \|\nabla \varphi\|_K. \end{aligned} \quad (3.20)$$

Now, by using (3.18), (3.19), and (3.20) in (3.17) together with the Cauchy–Schwarz inequality, we conclude (3.14).  $\square$

Let us now distinguish the discretization and algebraic error components. We in particular define

$$\text{discretization estimator:} \quad \eta_{\text{disc},K}^i := \eta_{\text{PNC},K}^i + \eta_{\text{R},K}^i + \eta_{\text{FD},K}^i + \eta_{\text{FNCD},K}^i, \quad (3.21a)$$

$$\text{algebraic estimator:} \quad \eta_{\text{alg},K}^i := \eta_{\text{FA},K}^i + \eta_{\text{FNCA},K}^i, \quad (3.21b)$$

with

$$\text{discretization flux estimator:} \quad \eta_{\text{FD},K}^i := \|\nabla u_h^i + \mathbf{d}_h^i\|_K,$$

$$\text{algebraic flux estimator:} \quad \eta_{\text{FA},K}^i := \|\mathbf{a}_h^i\|_K,$$

$$\text{discretization flux nonconformity estimator:} \quad \eta_{\text{FNCD},K}^i := \sum_{\Gamma \in \mathcal{E}_K^{\text{HG}} \cup \mathcal{E}_K^{\text{HG},\text{N}}} \bar{w}_{\Gamma} C_{\Gamma,K} h_{\Gamma}^{\frac{1}{2}} \|[\mathbf{d}_h^i \cdot \mathbf{n}_{\Gamma}]\|_{\Gamma}, \quad (3.22)$$

$$\text{algebraic flux nonconformity estimator:} \quad \eta_{\text{FNCA},K}^i := \sum_{\Gamma \in \mathcal{E}_K^{\text{HG}} \cup \mathcal{E}_K^{\text{HG},\text{N}}} \bar{w}_{\Gamma} C_{\Gamma,K} h_{\Gamma}^{\frac{1}{2}} \|[\mathbf{a}_h^i \cdot \mathbf{n}_{\Gamma}]\|_{\Gamma}.$$

We also let  $\eta_{\star}^i = \left( \sum_{K \in \mathcal{T}_h} (\eta_{\star,K}^i)^2 \right)^{\frac{1}{2}}$  with  $\star \in \{\text{R}, \text{PNC}, \text{FD}, \text{FA}, \text{FNCD}, \text{FNCA}, \text{rem}, \text{disc}, \text{alg}\}$ , where the local indicators  $\eta_{\star,K}^i$  are defined by (3.13a), (3.13b), (3.13e), and (3.22). We have:

**Corollary 3.10** (A posteriori error estimate distinguishing contributions of the discretization and algebraic error). *Let the assumptions of Theorem 3.8 be satisfied. Then*

$$\|\nabla_h(u - u_h^i)\| \leq 2^{\frac{1}{2}} \eta_{\text{disc}}^i + \eta_{\text{alg}}^i + \eta_{\text{rem}}^i. \quad (3.23)$$

*Proof.* Using the inequalities  $\|\nabla u_h^i + \mathbf{t}_h^i\|_K \leq \|\nabla u_h^i + \mathbf{d}_h^i\|_K + \|\mathbf{a}_h^i\|_K$  and  $\|[\mathbf{t}_h^i \cdot \mathbf{n}_{\Gamma}]\|_{\Gamma} \leq \|[\mathbf{d}_h^i \cdot \mathbf{n}_{\Gamma}]\|_{\Gamma} + \|[\mathbf{a}_h^i \cdot \mathbf{n}_{\Gamma}]\|_{\Gamma}$  following from (3.7), the Cauchy–Schwarz inequality, (3.21) and (3.22), we obtain from (3.14) the assertion (3.23).  $\square$

## 4 Stopping criteria and the adaptive algorithm

We propose in this section our stopping criteria and the corresponding adaptive solution algorithm. As discussed in [19, 17], on a given mesh, there is no need to continue iterations of the algebraic solver when the algebraic error falls below the discretization error. The total error cannot be reduced anyway. Combining this concept with that of adaptive mesh refinement, we propose the following *adaptive solution algorithm*: Let parameters  $\gamma_{\text{rem}} > 0$ ,  $\gamma_{\text{alg}} > 0$ , and an integer  $\nu^* > 0$  be given. Let  $\mathcal{T}_1$  be an initial mesh,  $U_1^0 \in \mathbb{R}^N$  an initial guess for the iterative algebraic solver, and  $TOL$  a user-given tolerance (the subscript  $h$  from the previous sections is in this section replaced by  $j$ ).

**Algorithm 4.1** (Adaptive solution algorithm).

1. Set  $j := 1$ .
2. (a) Set  $i := \nu^*$ .
  - (b) Perform  $\nu^*$  steps of the algebraic solver starting with  $U_j^0$  to get a new approximation  $U_j^i$  solving (2.8).

- (c) i. Set  $\nu := \nu^*$ .  
 ii. Perform  $\nu^*$  additional steps of the algebraic solver starting with  $U_j^{i+\nu-\nu^*}$ , save  $U_j^{i+\nu}$ . Construct new  $\mathbf{a}_j^i$  following (3.6) and evaluate the estimators  $\eta_{\text{disc},K}^i$ ,  $\eta_{\text{alg},K}^i$ , and  $\eta_{\text{rem},K}^i$  for all  $K \in \mathcal{T}_j$ .  
 iii. Check whether

$$\eta_{\text{rem},K}^i \leq \gamma_{\text{rem}}(\eta_{\text{disc},K}^i + \eta_{\text{alg},K}^i) \quad \forall K \in \mathcal{T}_j. \quad (4.1a)$$

If not satisfied, set  $\nu := \nu + \nu^*$  and go back to step 2(c)ii.

- (d) Check whether

$$\eta_{\text{alg},K}^i \leq \gamma_{\text{alg}}\eta_{\text{disc},K}^i \quad \forall K \in \mathcal{T}_j. \quad (4.1b)$$

If not satisfied,  $i := i + \nu$  and go to step 2(c)i.

3. Check whether  $\eta^i \leq \text{TOL}$ . If satisfied, stop. Else refine  $\mathcal{T}_j$  adaptively to  $\mathcal{T}_{j+1}$ , interpolate the currently available  $U_j^i$  from  $\mathcal{T}_j$  to  $\mathcal{T}_{j+1}$  to get new  $U_{j+1}^0$ , set  $j := j + 1$ , and go to step 2a.

**Remark 4.2** (Global stopping criteria). One can also define the following global version of the criteria (4.1):

$$\eta_{\text{rem}}^i \leq \gamma_{\text{rem}}(\eta_{\text{disc}}^i + \eta_{\text{alg}}^i), \quad (4.2a)$$

$$\eta_{\text{alg}}^i \leq \gamma_{\text{alg}}\eta_{\text{disc}}^i. \quad (4.2b)$$

## 5 Local efficiency of the a posteriori error estimate

In this section, we will show that the estimators  $\eta_{\text{disc},K}^i$ ,  $\eta_{\text{alg},K}^i$ , and  $\eta_{\text{rem},K}^i$  also provide local lower bound for the error. This gives a theoretical justification of these estimators and of their usage in Algorithm 4.1. Recall that  $\mathcal{T}_K$  denotes the set of the element  $K$  itself with its neighbors (including all elements that are contained in the macro-element sharing a complete face with the element  $K$  in case  $K$  possesses a hanging node),  $\mathcal{F}_K$  denotes the faces  $\Gamma$  in this patch, and  $\tilde{\mathcal{F}}_K$  the set of faces  $\gamma$  that share at least a vertex with  $K$ . We denote the macro-element sharing the complete face  $\Gamma$  with the element  $K$  by  $K_\Gamma^{\text{NEI}}$ , see Figure 1, left.

**Theorem 5.1** (Local efficiency of the estimate). *Let  $u \in H_0^1(\Omega)$  be the weak solution given by (2.1). Let an  $i$ -th algebraic solver step be given and let  $u_h^i \in S_h^{\mathbf{p}}$  be given by (2.10). Let  $f$  be a piecewise polynomial of degree  $\mathbf{p}$ . Let finally the algebraic solver be stopped as soon as the local stopping criteria (4.1) hold. Then there exists a generic constant  $C$  depending only on the shape-regularity constant  $C_s$  of (2.4a), the local quasi-uniformity constant  $C_H$  of (2.4b), the given weights  $\gamma_{\text{rem}}$  and  $\gamma_{\text{alg}}$ , the space dimension  $d$ , the DGM penalty parameter  $\alpha_{\text{max}} := \max_{\gamma \in \mathcal{F}_h} \alpha_\gamma$ , and the polynomial degree  $\mathbf{p}$  of the function  $u_h^i$  such that, for all  $K \in \mathcal{T}_h$ ,*

$$\eta_{\text{disc},K}^i + \eta_{\text{alg},K}^i + \eta_{\text{rem},K}^i \leq C \left\{ \sum_{K' \in \mathcal{T}_h: \mathcal{T}_K \cap \mathcal{T}_{K'} \neq \emptyset} \|\nabla(u - u_h^i)\|_{K'}^2 \right\}^{\frac{1}{2}} + C \left\{ \sum_{\gamma \in \mathcal{F}_K \cup \tilde{\mathcal{F}}_K} h_\gamma^{-1} \| [u_h^i] \|_\gamma^2 \right\}^{\frac{1}{2}}. \quad (5.1)$$

*Proof.* Let  $K \in \mathcal{T}_h$  be arbitrary but fixed. Due to the imposed local stopping criteria (4.1), we have

$$\eta_{\text{disc},K}^i + \eta_{\text{alg},K}^i + \eta_{\text{rem},K}^i \leq C\eta_{\text{disc},K}^i. \quad (5.2)$$

First, analogously to [20], it can be shown that the operator  $\mathcal{I}_{\text{Av}}$  defined in Section 3.1 has the following approximation property:

$$\|\nabla(v_h - \mathcal{I}_{\text{Av}}(v_h))\|_K^2 \leq C \sum_{\gamma \in \tilde{\mathcal{F}}_K} h_\gamma^{-1} \| [v_h] \|_\gamma^2, \quad \forall v_h \in S_h^{\mathbf{p}}.$$

Thus, we have

$$\eta_{\text{PNC},K}^i \leq C \left\{ \sum_{\gamma \in \bar{\mathcal{F}}_K} h_\gamma^{-1} \| [u_h^i]_\gamma \|^2 \right\}^{\frac{1}{2}}. \quad (5.3)$$

Further, observe that if the flux reconstruction  $\mathbf{t}_h^i$  has the order  $\mathbf{l} = \mathbf{p}$ ,  $\eta_{\text{R},K}^i = \frac{h_K}{\pi} \| f - \nabla \cdot \mathbf{t}_h^i - r_h^{i+\nu} \|_K = 0$  due to (3.8) and the assumption made on  $f$ . In the case of  $\mathbf{l} = \mathbf{p} - \mathbf{1}$  we proceed as follows. By adding and subtracting  $\Delta u_h^i$  in  $\eta_{\text{R},K}^i$ , using the triangle inequality and the inverse inequality

$$\| \nabla v_h \|_K \leq C h_K^{-1} \| v_h \|_K \quad \forall v_h \in \mathbb{P}^{p_K}(K),$$

we have

$$\eta_{\text{R},K}^i \leq \frac{h_K}{\pi} \| f + \Delta u_h^i \|_K + C \| \nabla u_h^i + \mathbf{t}_h^i \|_K + \frac{h_K}{\pi} \| r_h^{i+\nu} \|_K.$$

Further, (3.7) and the triangle inequality give

$$\eta_{\text{R},K}^i \leq \frac{h_K}{\pi} \| f + \Delta u_h^i \|_K + C \| \nabla u_h^i + \mathbf{d}_h^i \|_K + C \| \mathbf{a}_h^i \|_K + \frac{h_K}{\pi} \| r_h^{i+\nu} \|_K. \quad (5.4)$$

Due to (3.21b) and (3.13e), the last two terms are bounded by  $\eta_{\text{alg},K}^i + \eta_{\text{rem},K}^i$ . The first term is a standard residual estimator known to satisfy (see [34])

$$h_K \| f + \Delta u_h^i \|_K \leq C \| \nabla(u - u_h^i) \|_K. \quad (5.5)$$

Next, we will estimate  $\| \nabla u_h^i + \mathbf{d}_h^i \|_K$ . According to [12, Lemma 3.5], we can write

$$\| \nabla u_h^i + \mathbf{d}_h^i \|_K^2 \leq C \left\{ h_K \sum_{\Gamma \in \mathcal{E}_K} \| (\nabla u_h^i + \mathbf{d}_h^i) \cdot \mathbf{n}_\Gamma \|_\Gamma^2 + \left( \sup_{\mathbf{q}_h \in [\mathbb{P}_{l_K-1}(K)]^d} \frac{(\nabla u_h^i + \mathbf{d}_h^i, \mathbf{q}_h)_K}{\| \mathbf{q}_h \|_K} \right)^2 \right\}. \quad (5.6)$$

For  $\mathbf{q}_h \in [\mathbb{P}^{l_K-1}(K)]^d$ , taking into account definition (3.5b), we have

$$(\nabla u_h^i + \mathbf{d}_h^i, \mathbf{q}_h)_K = \theta \sum_{\Gamma \in \mathcal{E}_K} w_\Gamma (\mathbf{q}_h \cdot \mathbf{n}_\Gamma, [u_h^i]_\Gamma).$$

Now, by using the Cauchy–Schwarz inequality and the inverse inequality  $\| \mathbf{q}_h \|_\Gamma \leq C h_K^{-1/2} \| \mathbf{q}_h \|_K$ , we obtain

$$(\nabla u_h^i + \mathbf{d}_h^i, \mathbf{q}_h)_K \leq C |\theta| h_K^{-1/2} \| \mathbf{q}_h \|_K \sum_{\Gamma \in \mathcal{E}_K} w_\Gamma \| [u_h^i]_\Gamma \|. \quad (5.7)$$

By putting (5.7) into (5.6) and using definition (3.5a), we get

$$\| \nabla u_h^i + \mathbf{d}_h^i \|_K^2 \leq C \left\{ h_K \sum_{\Gamma \in \mathcal{E}_K} \| \Pi_{l_K} [\nabla u_h^i \cdot \mathbf{n}_\Gamma] \|_\Gamma^2 + h_K \alpha_{\text{max}}^2 \sum_{\Gamma \in \mathcal{E}_K} h_\Gamma^{-2} \| \Pi_{l_K} [u_h^i] \|_\Gamma^2 + |\theta|^2 h_K^{-1} \sum_{\Gamma \in \mathcal{E}_K} w_\Gamma^2 \| [u_h^i]_\Gamma \|^2 \right\}.$$

Let  $\Gamma \in \mathcal{E}_K^I$ . With the aid of the edge bubble functions technique introduced by Verfürth, see [34], it can be shown that

$$h_\Gamma^{\frac{1}{2}} \| [\nabla u_h^i \cdot \mathbf{n}_\Gamma] \|_\Gamma \leq \begin{cases} C \sum_{K'' \in \{K_\gamma^L, K_\gamma^R\}} \| \nabla(u - u_h^i) \|_{K''} & \text{for } \gamma = \Gamma \in \mathcal{E}_K^I \setminus \mathcal{E}_K^{\text{HG}}, \\ C \left( \| \nabla(u - u_h^i) \|_K + \sum_{K' \subset K_\Gamma^{\text{NEI}}} \| \nabla(u - u_h^i) \|_{K'} \right) & \text{for } \Gamma \in \mathcal{E}_K^{\text{HG}}. \end{cases}$$

Therefore, by taking into account the estimate

$$\| \Pi_{l_K} [\nabla u_h^i \cdot \mathbf{n}_\Gamma] \|_\Gamma \leq \| [\nabla u_h^i \cdot \mathbf{n}_\Gamma] \|_\Gamma, \quad \Gamma \in \mathcal{E}_K^I, \quad (5.8)$$

we obtain

$$h_K \sum_{\Gamma \in \mathcal{E}_K^I} \|\Pi_{l_K}[\nabla u_h^i \cdot \mathbf{n}_\Gamma]\|_\Gamma^2 \leq h_K \sum_{\Gamma \in \mathcal{E}_K^I} \|\nabla u_h^i \cdot \mathbf{n}_\Gamma\|_\Gamma^2 \leq C \sum_{K \in \mathcal{T}_K} \|\nabla(u - u_h^i)\|_K^2.$$

Finally, using

$$\|\Pi_{l_K}[u_h^i]\|_\Gamma \leq \|[u_h^i]\|_\Gamma$$

yields

$$\|\nabla u_h^i + \mathbf{d}_h^i\|_K^2 \leq C \left\{ \sum_{K \in \mathcal{T}_K} \|\nabla(u - u_h^i)\|_K^2 + (\alpha_{\max}^2 + 1) \sum_{\Gamma \in \mathcal{E}_K} h_\Gamma^{-1} \|[u_h^i]\|_\Gamma^2 \right\}. \quad (5.9)$$

Now, it remains to estimate the last term of (3.21a). According to the estimate in [35, Theorem 3.3, (3.21)] and [27, Theorem 4.3], for a vector only piecewise in  $\mathbf{H}(\text{div}, \cdot)$ , we can write

$$h_\Gamma^{\frac{1}{2}} \|[d_h^i \cdot \mathbf{n}_\Gamma]\|_\Gamma \leq \begin{cases} C \sum_{K'' \in \{K_\gamma^L, K_\gamma^R\}} \|d_h^i + \nabla u\|_{K''} & \text{for } \gamma = \Gamma \in \mathcal{E}_K^{\text{HG}, \text{N}}, \\ C \left( \|d_h^i + \nabla u\|_K + \sum_{K' \subset K_\Gamma^{\text{NEI}}} \|d_h^i + \nabla u\|_{K'} \right) & \text{for } \Gamma \in \mathcal{E}_K^{\text{HG}}. \end{cases}$$

Now, adding and subtracting  $\nabla u_h^i$  in the norm to the above right-hand sides together with the triangle inequality yields

$$h_\Gamma^{\frac{1}{2}} \|[d_h^i \cdot \mathbf{n}_\Gamma]\|_\Gamma \leq \begin{cases} C \sum_{K'' \in \{K_\gamma^L, K_\gamma^R\}} (\|d_h^i + \nabla u_h^i\|_{K''} + \|\nabla(u - u_h^i)\|_{K''}) & \text{for } \gamma = \Gamma \in \mathcal{E}_K^{\text{HG}, \text{N}}, \\ C \left( \|d_h^i + \nabla u_h^i\|_K + \|\nabla(u - u_h^i)\|_K \right. \\ \left. + \sum_{K' \subset K_\Gamma^{\text{NEI}}} (\|d_h^i + \nabla u_h^i\|_{K'} + \|\nabla(u - u_h^i)\|_{K'}) \right) & \text{for } \Gamma \in \mathcal{E}_K^{\text{HG}}. \end{cases} \quad (5.10)$$

Now combining (5.2) with definition (3.21a) and (5.3), (5.4), (5.5), (5.9), and (5.10) gives

$$\eta_{\text{disc}, K}^i + \eta_{\text{alg}, K}^i + \eta_{\text{rem}, K}^i \leq C \left\{ \sum_{K' \in \mathcal{T}_h; \mathcal{T}_K \cap \mathcal{T}_{K'} \neq \emptyset} \|\nabla(u - u_h^i)\|_{K'}^2 \right\}^{\frac{1}{2}} + C \left\{ \sum_{\gamma \in \mathcal{F}_K \cup \tilde{\mathcal{F}}_K} h_\gamma^{-1} \|[u_h^i]\|_\gamma^2 \right\}^{\frac{1}{2}} + C(\eta_{\text{alg}, K}^i + \eta_{\text{rem}, K}^i).$$

Choosing  $\gamma_{\text{rem}}$  and  $\gamma_{\text{alg}}$  in (4.1) small enough allows to discard the contribution of  $\eta_{\text{alg}, K}^i$  and  $\eta_{\text{rem}, K}^i$  from the above right-hand side and to arrive at (5.1).  $\square$

**Remark 5.2** (Both-sided equivalence). *We remark that in Theorem 5.1 besides the broken  $H^1$ -seminorm, in which we measure the error, jump terms appear. In order to have a two-sided equivalence of the estimate and of the error, we can as usual augment the broken  $H^1$ -seminorm by these jump terms. Then Theorem 3.8*

and Theorem 5.1, together with the fact that  $[u] = 0$  for  $\gamma \in \mathcal{F}_h$ , lead to

$$\begin{aligned} & \|\nabla_h(u - u_h^i)\|^2 + \sum_{K \in \mathcal{T}_h} \sum_{\gamma \in \mathcal{F}_K \cup \tilde{\mathcal{F}}_K} h_\gamma^{-1} \| [u - u_h^i] \|_\gamma^2 \leq \sum_{K \in \mathcal{T}_h} (\eta_{\text{PNC}, K}^i)^2 + \left\{ \left\{ \sum_{K \in \mathcal{T}_h} (\eta_{\text{R}, K}^i + \eta_{\text{F}, K}^i + \eta_{\text{FNC}, K}^i)^2 \right\}^{\frac{1}{2}} \right. \\ & \left. + \left\{ \sum_{K \in \mathcal{T}_h} (\eta_{\text{rem}, K}^i)^2 \right\}^{\frac{1}{2}} \right\}^2 + \sum_{K \in \mathcal{T}_h} \sum_{\gamma \in \mathcal{F}_K \cup \tilde{\mathcal{F}}_K} h_\gamma^{-1} \| [u_h^i] \|_\gamma^2, \\ & \eta_{\text{disc}, K}^i + \eta_{\text{alg}, K}^i + \eta_{\text{rem}, K}^i + \left\{ \sum_{\gamma \in \mathcal{F}_K \cup \tilde{\mathcal{F}}_K} h_\gamma^{-1} \| [u_h^i] \|_\gamma^2 \right\}^{\frac{1}{2}} \\ & \leq C \left\{ \sum_{K' \in \mathcal{T}_h; \mathcal{T}_K \cap \mathcal{T}_{K'} \neq \emptyset} \|\nabla(u - u_h^i)\|_{K'}^2 \right\}^{\frac{1}{2}} + C \left\{ \sum_{\gamma \in \mathcal{F}_K \cup \tilde{\mathcal{F}}_K} h_\gamma^{-1} \| [u - u_h^i] \|_\gamma^2 \right\}^{\frac{1}{2}}, \end{aligned}$$

which gives the upper and lower estimates in the so-called DG-norm defined, for any  $v \in H^1(\Omega, \mathcal{T}_h)$ , by  $\|v\|_{\text{DG}}^2 := \|\nabla_h(v)\|^2 + \sum_{K \in \mathcal{T}_h} \sum_{\gamma \in \mathcal{F}_K \cup \tilde{\mathcal{F}}_K} h_\gamma^{-1} \| [v] \|_\gamma^2$ .

## 6 Simple evaluation of the a posteriori estimates

The estimators of Theorem 3.8 and of Corollary 3.10 may seem rather difficult to evaluate at a first sight. In particular the flux reconstructions in RTN spaces may be a little involved to code and work with. In this section, we show that, at least for low-order approximations (most frequently used in practice), our estimates lead to simple formulas not featuring the flux reconstructions which are easy to implement and evaluate.

### 6.1 First-order discretization

We start with the simplest case considering  $u_h^i \in S_h^1$ ,  $\mathbf{1} := \{1_K\}_{K \in \mathcal{T}_h}$ , and  $\mathbf{t}_h^i \in \mathbf{RTN}_0(\mathcal{T}_h)$ ,  $\mathbf{0} := \{0_K\}_{K \in \mathcal{T}_h}$ . Let us first provide an explicit definition of the  $\mathbf{RTN}_0(\mathcal{T}_h)$ -basis.

**Definition 6.1.** (Definition of  $\mathbf{RTN}_0(\mathcal{T}_h)$ -basis functions) *Let a simplex  $K \in \mathcal{T}_h$  be given. Let  $\Gamma_j, j = 1 \dots d+1$ , be its edges or faces and  $V_{\Gamma_j}, j = 1 \dots d+1$ , the opposite vertices, respectively. Let  $\mathbf{n}_{j, \partial K}$  denote the unit outward normal to  $K$  along  $\Gamma_j$ . The  $\mathbf{RTN}_0(\mathcal{T}_h)$ -basis functions are defined by*

$$\psi_{\Gamma_j}(\mathbf{x}) := \mathbf{n}_{\Gamma_j} \cdot \mathbf{n}_{j, \partial K} \frac{1}{d|K|} (\mathbf{x} - V_{\Gamma_j}) \quad \text{for } j = 1 \dots d+1, \mathbf{x} \in K.$$

Note that the volume  $|K|$  can be computed by the formula:

$$|K| = \frac{1}{d!} \det \begin{pmatrix} V_{\Gamma_1} & \dots & V_{\Gamma_{d+1}} \\ 1 & \dots & 1 \end{pmatrix}.$$

For any  $\mathbf{t}_h^i \in \mathbf{RTN}_0(\mathcal{T}_h)$ , we can write  $\mathbf{t}_h^i|_K = \sum_{\Gamma \in \mathcal{E}_K} (\mathbf{t}_h^i)_\Gamma \psi_\Gamma$ , where  $(\mathbf{t}_h^i)_\Gamma$  are the degrees of freedom associated with the basis of Definition 6.1. Recall that we distinguish the discretization and algebraic components, i.e.,  $\mathbf{t}_h^i = \mathbf{d}_h^i + \mathbf{a}_h^i$ . Let  $K \in \mathcal{T}_h$  be fixed. Then, for  $\Gamma \in \mathcal{E}_K$ , the coefficients  $(\mathbf{d}_h^i)_\Gamma$  are given by, see (3.5a),

$$(\mathbf{d}_h^i)_\Gamma := (\mathbf{d}_h^i \cdot \mathbf{n}_\Gamma, 1)_\Gamma = (-\langle \nabla u_h^i \cdot \mathbf{n}_\Gamma \rangle + \alpha h^{-1} [u_h^i], 1)_\Gamma, \quad (6.1a)$$

whereas the coefficients  $(\mathbf{a}_h^i)_\Gamma$  are given by, see (3.6),

$$(\mathbf{a}_h^i)_\Gamma := ((\mathbf{d}_h^{i+\nu} - \mathbf{d}_h^i) \cdot \mathbf{n}_\Gamma, 1)_\Gamma = (-\langle \nabla (u_h^{i+\nu} - u_h^i) \cdot \mathbf{n}_\Gamma \rangle + \alpha h^{-1} [u_h^{i+\nu} - u_h^i], 1)_\Gamma. \quad (6.1b)$$

Let  $\{\varphi_j\}_{j \in \mathcal{S}_K}$  be a basis of  $S_h^1|_K$  and  $\{\varphi_j\}_{j \in \mathcal{N}_K}$  a basis of  $S_h^1$  on  $K$  and all its neighbors. Expressing  $u_h^i$  in these bases yields

$$u_h^i|_K = \sum_{j \in \mathcal{S}_K} (u_h^i)_j \varphi_j, \quad u_h^i|_{K' \in \mathcal{T}_h; |\partial K \cap \partial K'| > 0} = \sum_{j \in \mathcal{N}_K} (u_h^i)_j \varphi_j. \quad (6.2)$$

Using (6.2) in (6.1a) and (6.1b) gives

$$(\mathbf{d}_h^i)_\Gamma = (\mathbf{d}_h^i \cdot \mathbf{n}_\Gamma, 1)_\Gamma = \left( - \sum_{j \in \mathcal{N}_K} (u_h^i)_j \langle \nabla \varphi_j \cdot \mathbf{n}_\Gamma \rangle + \alpha h^{-1} \sum_{j \in \mathcal{N}_K} (u_h^i)_j [\varphi_j], 1 \right)_\Gamma, \quad (6.3a)$$

$$(\mathbf{a}_h^i)_\Gamma = (\mathbf{a}_h^i \cdot \mathbf{n}_\Gamma, 1)_\Gamma = \left( - \sum_{j \in \mathcal{N}_K} (u_h^{i+\nu} - u_h^i)_j \langle \nabla \varphi_j \cdot \mathbf{n}_\Gamma \rangle + \alpha h^{-1} \sum_{j \in \mathcal{N}_K} (u_h^{i+\nu} - u_h^i)_j [\varphi_j], 1 \right)_\Gamma. \quad (6.3b)$$

Now, we are ready to provide explicit formulas for evaluation of the a posteriori error estimators in Corollary 3.10, avoiding the physical construction of the flux reconstructions  $\mathbf{d}_h^i$  and  $\mathbf{a}_h^i$ , as well as of the potential reconstruction  $\mathcal{I}_{\text{Av}}(u_h^i)$ .

Let us start with the estimator  $\|\nabla u_h^i + \mathbf{d}_h^i\|_K$ . First,  $(\nabla u_h^i + \mathbf{d}_h^i)|_K \in [\mathbb{P}^1(K)]^d$  holds. Thus, we will need a quadrature rule that is exact for quadratic polynomials such as

$$\int_K w(\mathbf{x}) \, d\mathbf{x} \approx \begin{cases} \frac{|K|}{3} \sum_{\Gamma \in \mathcal{E}_K} w(\mathbf{x}_\Gamma), & \text{for } d = 2, \\ \frac{|K|}{20} \sum_{k=1}^4 w(V_{\Gamma_k}) + \frac{4|K|}{5} w(\mathbf{x}_K), & \text{for } d = 3, \end{cases} \quad (6.4)$$

where  $\mathbf{x}_\Gamma$  are the mid-points of the sides of the triangle  $K$ ,  $\mathbf{x}_K$  is the barycentre of  $K$ , and  $V_{\Gamma_k}$ ,  $k = 1, 2, 3, 4$ , are the vertices of the tetrahedron  $K$ . Recall that  $\mathcal{E}_K$  denotes the set of the faces of the element  $K$ . With the aid of (6.4) and (6.3a), we have

$$\|\nabla u_h^i + \mathbf{d}_h^i\|_K^2 = \begin{cases} \left| \frac{|K|}{3} \sum_{\Gamma \in \mathcal{E}_K} \left| \sum_{j \in \mathcal{S}_K} (u_h^i)_j \nabla \varphi_j(\mathbf{x}_\Gamma) + \sum_{\Gamma' \in \mathcal{E}_K} (\mathbf{d}_h^i)_{\Gamma'} \psi_{\Gamma'}(\mathbf{x}_\Gamma) \right|^2 \right|, & \text{for } d = 2, \\ \left| \frac{|K|}{20} \sum_{k=1}^4 \left| \sum_{j \in \mathcal{S}_K} (u_h^i)_j \nabla \varphi_j(V_{\Gamma_k}) + \sum_{\Gamma' \in \mathcal{E}_K} (\mathbf{d}_h^i)_{\Gamma'} \psi_{\Gamma'}(V_{\Gamma_k}) \right|^2 \right. \\ \left. + \frac{4|K|}{5} \left| \sum_{j \in \mathcal{S}_K} (u_h^i)_j \nabla \varphi_j(\mathbf{x}_K) + \sum_{\Gamma' \in \mathcal{E}_K} (\mathbf{d}_h^i)_{\Gamma'} \psi_{\Gamma'}(\mathbf{x}_K) \right|^2 \right|, & \text{for } d = 3. \end{cases}$$

Analogously, as  $\mathbf{a}_h^i|_K \in [\mathbb{P}^1(K)]^d$ , we get the relation for evaluation of  $\|\mathbf{a}_h^i\|_K^2$ .

Another estimator that needs to be evaluated is  $\|\nabla(u_h^i - \mathcal{I}_{\text{Av}}(u_h^i))\|_K$ . As  $\nabla(u_h^i - \mathcal{I}_{\text{Av}}(u_h^i))|_K \in [\mathbb{P}^0(K)]^d$  holds, the following quadrature rule of the algebraic order 1 is sufficient:

$$\int_K w(\mathbf{x}) \, d\mathbf{x} \approx |K| w(\mathbf{x}_K).$$

With the aid of (6.2) and (3.3), we can write

$$\|\nabla(u_h^i - \mathcal{I}_{\text{Av}}(u_h^i))\|_K^2 = |K| \left| \sum_{j \in \mathcal{S}_K} (u_h^i)_j \nabla \varphi_j(\mathbf{x}_K) - \nabla \sum_{k=1}^{d+1} (\mathcal{I}_{\text{Av}}(u_h^i))_k \varphi_k(\mathbf{x}_K) \right|^2,$$

$$(\mathcal{I}_{\text{Av}}(u_h^i))_k := \begin{cases} \frac{1}{\text{card}(\mathcal{T}_{V_{\Gamma_k}})} \sum_{K' \in \mathcal{T}_{V_{\Gamma_k}}} \sum_{j \in \mathcal{S}_{K'}} (u_h^i)_j \varphi_j|_{K'}(V_{\Gamma_k}), & \text{for } V_{\Gamma_k} \notin \mathcal{N}_h^N, \\ \sum_{V_j \in \mathcal{N}_h^I \cup \mathcal{N}_h^B} \left( \frac{1}{\text{card}(\mathcal{T}_{V_j})} \sum_{K' \in \mathcal{T}_{V_j}} \sum_{l \in \mathcal{S}_{K'}} (u_h^i)_l \varphi_l|_{K'}(V_j) \right) \varphi_j(V_{\Gamma_k}), & \text{for } V_{\Gamma_k} \in \mathcal{N}_h^N, \end{cases}$$

where  $\mathcal{N}_h^N$  is defined in (3.2),  $V_{\Gamma_k}$ ,  $k = 1 \dots d+1$ , are the vertices of the element  $K$  and the basis functions  $\varphi_k$  of  $S_h^1|_K$  are ordered as the vertices.



Now, the estimators

$$\sum_{\Gamma \in \mathcal{E}_K^{\text{HG}}} C_{\Gamma,K} h_{\Gamma}^{\frac{1}{2}} \|[\mathbf{d}_h^i \cdot \mathbf{n}_{\Gamma}]\|_{\Gamma} \quad \text{and} \quad \sum_{\Gamma \in \mathcal{E}_K^{\text{HG},\text{N}}} \bar{w}_{\Gamma} C_{\Gamma,K} h_{\Gamma}^{\frac{1}{2}} \|[\mathbf{d}_h^i \cdot \mathbf{n}_{\Gamma}]\|_{\Gamma}$$

are to be evaluated. Due to the fact that  $[\mathbf{d}_h^i \cdot \mathbf{n}_{\Gamma}]_{\Gamma} \in \mathbb{P}^0(\Gamma)$ , it is quite easy. There are two cases to consider:

1. If the face  $\Gamma$  has no hanging node, then

$$[\mathbf{d}_h^i \cdot \mathbf{n}_{\Gamma}]_{\Gamma} = 0.$$

2. The face  $\Gamma$  possesses a hanging node. For simplicity, let us assume that only one hanging node is present. Let  $\gamma_1$  and  $\gamma_2$  be those parts which have the hanging node in common and which comprise the face  $\Gamma$ . Let  $K_1$  and  $K_2$  be the elements sharing their faces with  $K$  (see Figure 4). Then

$$\begin{aligned} \|[\mathbf{d}_h^i \cdot \mathbf{n}_{\Gamma}]\|_{\Gamma}^2 &= ([-\langle \nabla u_h^i \cdot \mathbf{n}_{\gamma_1} \rangle] + \alpha_{\gamma_1} h_{\gamma_1}^{-1} (\Pi_{0,\gamma_1} - \Pi_{0,\Gamma}|_{\gamma_1})[u_h^i])^2 |\gamma_1| \\ &\quad + ([-\langle \nabla u_h^i \cdot \mathbf{n}_{\gamma_2} \rangle] + \alpha_{\gamma_2} h_{\gamma_2}^{-1} (\Pi_{0,\gamma_2} - \Pi_{0,\Gamma}|_{\gamma_2})[u_h^i])^2 |\gamma_2|, \end{aligned}$$

where  $\Pi_{0,\Gamma_j}$ ,  $j = 1, 2$ , and  $\Pi_{0,\Gamma}$  are the  $L^2(\Gamma_j)$ - and  $L^2(\Gamma)$ -orthogonal projections onto  $\mathbb{P}^0(\Gamma_j)$  and  $\mathbb{P}^0(\Gamma)$ , respectively.

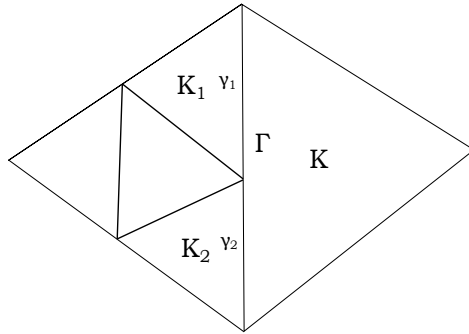


Figure 4: Components of a face  $\Gamma$

The estimators with  $\mathbf{a}_h^i$  are treated similarly, whereas  $\eta_{R,K}^i$  is easily evaluated using (3.8).

## 6.2 Second-order discretization

We continue by considering  $u_h^i \in S_h^2$ ,  $\mathbf{2} := \{2_K\}_{K \in \mathcal{T}_h}$ , and  $\mathbf{t}_h^i \in \mathbf{RTN}_1(\mathcal{T}_h)$ . Let us state the explicit definition of  $\mathbf{RTN}_1(\mathcal{T}_h)$ -basis.

**Definition 6.2.** (Definition of  $\mathbf{RTN}_1(\mathcal{T}_h)$ -basis functions) *Let a simplex  $K \in \mathcal{T}_h$  be given. Let  $\Gamma_j, j = 1 \dots d+1$ , be its edges or faces and  $V_{\Gamma_j}, j = 1 \dots d+1$ , the opposite vertices, respectively. Let  $\mathbf{n}_{j,\partial K}$  denote the unit outward normal to  $\partial K$  along  $\Gamma_j$ . Let  $\mathbf{e}_j, j = 1 \dots d$ , stand for the canonical basis of  $\mathbb{R}^d$  and  $S_{\Gamma_j}$  for the average value of  $S$  on  $\Gamma_j$ . Let finally  $\lambda_j, j = 1 \dots d+1$ , be the barycentric coordinates,  $\varphi_l$  be the basis functions of  $\mathbb{P}^2(K)$  of the type  $4\lambda_i\lambda_j, 1 \leq i < j \leq d+1$ , and  $\phi_l, l = 1 \dots 8$ , be defined by:*

$$\begin{bmatrix} x_2^2 \\ -x_1x_2 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ -x_2x_3 \\ x_2^2 \end{bmatrix}, \quad \begin{bmatrix} -x_1x_2 \\ x_1^2 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} -x_1x_3 \\ 0 \\ x_1^2 \end{bmatrix}, \quad \begin{bmatrix} x_3^2 \\ 0 \\ -x_1x_3 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ x_3^2 \\ -x_2x_3 \end{bmatrix}, \quad \begin{bmatrix} x_2x_3 \\ -x_1x_3 \\ 0 \end{bmatrix}, \quad \begin{bmatrix} 0 \\ x_1x_3 \\ -x_1x_2 \end{bmatrix}.$$

The  $\mathbf{RTN}_1(\mathcal{T}_h)$ -basis functions corresponding to the following degrees of freedom

$$\begin{aligned} \mathbf{q} &\mapsto \int_{\Gamma_j} \mathbf{q} \cdot \mathbf{n}_{\Gamma_j} \, dS && \text{for } j = 1 \dots d+1, \\ \mathbf{q} &\mapsto \int_{\Gamma_j} \mathbf{q} \cdot \mathbf{n}_{j, \partial K} (S - S_{\Gamma_j}) \, dS && \text{for } j = 1 \dots d+1, \\ \mathbf{q} &\mapsto \int_K \mathbf{q} \cdot \mathbf{e}_j \, d\mathbf{x} && \text{for } j = 1 \dots d, \end{aligned}$$

are defined as follows, for  $\mathbf{x} \in K$ ,

$$\begin{aligned} \psi_{\Gamma_j}(\mathbf{x}) &:= \mathbf{n}_{\Gamma_j} \cdot \mathbf{n}_{j, \partial K} \frac{1}{d|K|} (\mathbf{x} - V_{\Gamma_j}) && \text{for } j = 1 \dots d+1, \\ \psi_l(\mathbf{x}) &:= \begin{cases} \mathbf{curl} \varphi_l & \text{for } l = 1 \dots 3 \text{ if } d = 2, \\ \mathbf{curl} \phi_l & \text{for } l = 1 \dots 8 \text{ if } d = 3, \end{cases} \\ \psi_j(\mathbf{x}) &:= 4 \sum_{i=1}^{d+1} \lambda_i \frac{\partial \lambda_i}{\partial x_j} (V_{\Gamma_i} - \mathbf{x}) && \text{for } j = 1 \dots d. \end{aligned}$$

Enumerate the basis functions from Definition 6.2 as  $\psi_l$ ,  $l = 1 \dots 8$ , for  $d = 2$  and  $l = 1 \dots 15$  for  $d = 3$ . For any  $\mathbf{t}_h^i \in \mathbf{RTN}_1(\mathcal{T}_h)$ , we can write  $\mathbf{t}_h^i|_K = \sum_l (\mathbf{t}_h^i)_l \psi_l$  with  $(\mathbf{t}_h^i)_l$  the associated degrees of freedom. The evaluation of a posteriori error estimators in Corollary 3.10 can be done again without factual construction of reconstructions as in Section 6.1. In particular, one needs a quadrature rule that is exact for quartic polynomials. Let

$$\begin{aligned} V_{ij}^{(3,1,0)} &:= \frac{3}{4} V_{\Gamma_i} + \frac{1}{4} V_{\Gamma_j} && 1 \leq i, j \leq d+1, i \neq j, \\ V_{ij}^{(2,2,0)} &:= \frac{1}{2} V_{\Gamma_i} + \frac{1}{2} V_{\Gamma_j} && 1 \leq i < j \leq d+1, \\ V_{ijk}^{(2,1,1)} &:= \frac{1}{2} V_{\Gamma_i} + \frac{1}{4} V_{\Gamma_j} + \frac{1}{4} V_{\Gamma_k} && 1 \leq i, j, k \leq d+1, i \neq j < k. \end{aligned}$$

Then such a formula is, see [32],

$$\int_K w(\mathbf{x}) \, d\mathbf{x} \approx \begin{cases} \frac{4|K|}{45} \sum_{1 \leq i, j \leq 3, i \neq j} w(V_{ij}^{(3,1,0)}) - \frac{|K|}{45} \sum_{1 \leq i < j \leq 3} w(V_{ij}^{(2,2,0)}) \\ \quad + \frac{8|K|}{45} \sum_{1 \leq i, j, k \leq 3, i \neq j < k} w(V_{ijk}^{(2,1,1)}), && \text{for } d = 2, \\ \frac{-5|K|}{420} \sum_{k=1}^4 w(V_{\Gamma_k}) + \frac{16|K|}{420} \sum_{1 \leq i, j \leq 4, i \neq j} w(V_{ij}^{(3,1,0)}) \\ \quad - \frac{12|K|}{420} \sum_{1 \leq i < j \leq 4} w(V_{ij}^{(2,2,0)}) + \frac{16|K|}{420} \sum_{1 \leq i, j, k \leq 4, i \neq j < k} w(V_{ijk}^{(2,1,1)}) \\ \quad + \frac{128|K|}{420} w(\mathbf{x}_K), && \text{for } d = 3. \end{cases}$$

## 7 Numerical experiments

In this section we will illustrate the behavior of the error estimates introduced in Section 3 and of the adaptive solution algorithm introduced in Section 4. Our aim is to illustrate:

- i) the behavior of all contributions of the error estimate from Theorem 3.8;
- ii) the behavior of the discretization and algebraic errors and their estimators from Corollary 3.10;
- iii) the local distribution of the algebraic and discretization errors and of their estimators;
- iv) the comparison of the efficiency of the proposed stopping criteria with classical ones.

We employ the incomplete interior penalty DGM, i.e., (2.6) with  $\theta = 0$ , with the penalty parameter  $\alpha_\gamma = 20$  for all  $\gamma \in \mathcal{F}_h$  on triangular meshes possibly containing hanging nodes. This discretization leads to non-symmetric algebraic systems (2.7), which we solve by the GMRES method [31] with ILU(0) preconditioning. The approximate solution is sought in the space  $S_h^2$  and the flux reconstructions in the RTN space  $\mathbf{RTN}_2(\mathcal{T}_h)$ . Algorithm 4.1 is applied with parameters  $\nu^* = 15$  and  $\gamma_{\text{rem}} = \gamma_{\text{alg}} = 10^{-1}$ . We carried out the computations with two types of stopping criteria: *local stopping criteria* (4.1), and *classical stopping criteria*, where the GMRES method was let to converge to four different tolerances for the relative preconditioned algebraic residuum measured in the  $\ell^2$ -norm:  $10^{-4}$ ,  $10^{-6}$ ,  $10^{-9}$ , and  $10^{-12}$ . [—]

Meshes are generated adaptively according to the elementwise discretization estimator (3.21a). Since we aim at fulfilling the condition  $\eta_{\text{disc}}^i \leq \omega$  for some tolerance  $\omega$ , we require

$$\eta_{\text{disc},K}^i \leq \omega \text{card}(\mathcal{T}_h)^{-1/2} \tag{7.1}$$

to hold for all  $K \in \mathcal{T}_h$ , where  $\text{card}(\mathcal{T}_h)$  denotes the number of triangles in the current mesh. Therefore, triangles for those the condition (7.1) is violated are split into four smaller ones. The tolerance  $\omega$  has been set to  $1 \cdot 10^{-2}$  in the first example and to  $3.3 \cdot 10^{-3}$  in the second example in the computations.

### 7.1 Example 1: singular solution on an L-shaped domain

Similarly as in [1] and [11], we solve the Poisson equation (1.1a) with  $f = 0$  on the L-shaped domain  $\Omega = (-1, 1)^2 \setminus (0, 1)^2$  with a nonhomogeneous Dirichlet boundary condition, which is prescribed such that the exact solution has in the polar coordinates  $(r, \phi)$  the following form:

$$u(r, \phi) = r^{\frac{2}{3}} \sin\left(\frac{2\phi}{3}\right).$$

The computation is started with zero a initial guess, on a matching triangular grid with 96 elements. We carried out four levels of mesh adaptation.

Figure 5 shows the values of all error estimator contributions  $\eta_\star^i$  defined from (3.13a), (3.13b), (3.13c), and (3.22). Discretization and algebraic flux nonconformity estimators are not displayed for the first mesh as no hanging nodes are present and consequently they are zero (recall that we consider the same polynomial degree over the whole mesh). Similarly, the residual estimator is not displayed as its values are zero. Figure 5 confirms our expectations: whereas the algebraic error estimators are important in the first GMRES iterations, they drop rapidly, leaving the discretization error estimators to dominate. Many unnecessary GMRES iterations on each mesh level can be seen in the right part of Figure 5. Some description of the results...

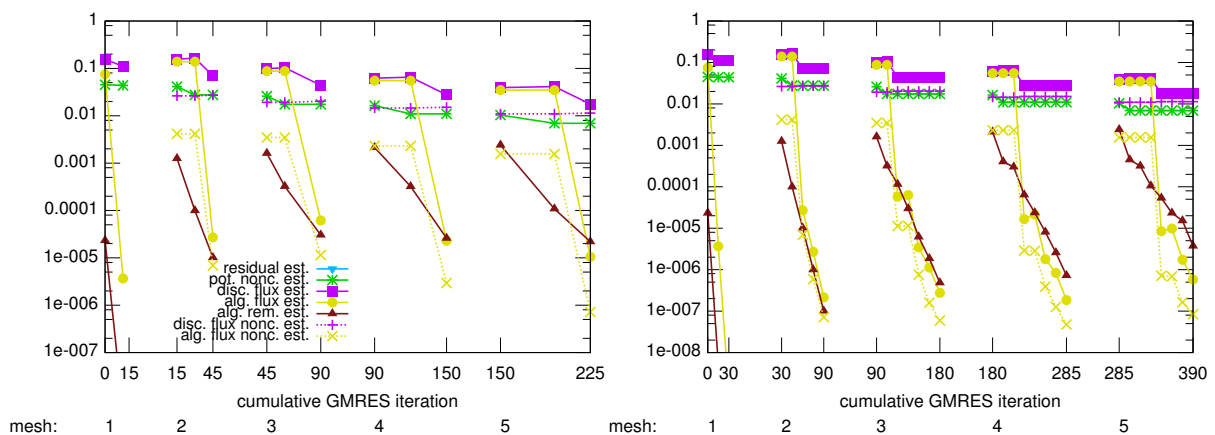


Figure 5: Example 1, development of the estimators on the individual meshes for local (left) and classical with  $10^{-6}$  relative tolerance (right) stopping criteria

Further, we present the evolution of the total, discretization, and algebraic errors as well as of the respective estimators  $\eta^i$ ,  $\eta_{\text{disc}}^i$ , and  $\eta_{\text{alg}}^i$ , where  $\eta^i$  is given by (3.14). Figure 6 shows the evolution of

these errors and estimators through the whole adaptation process. We observe that our estimators predict very precisely the overall error as well as its algebraic and discretization components. (The discretization estimator gives slightly higher values than the total estimator, which may happen due to the fact that we plot the right-hand side of (3.14) as the total estimator, not (3.23).)

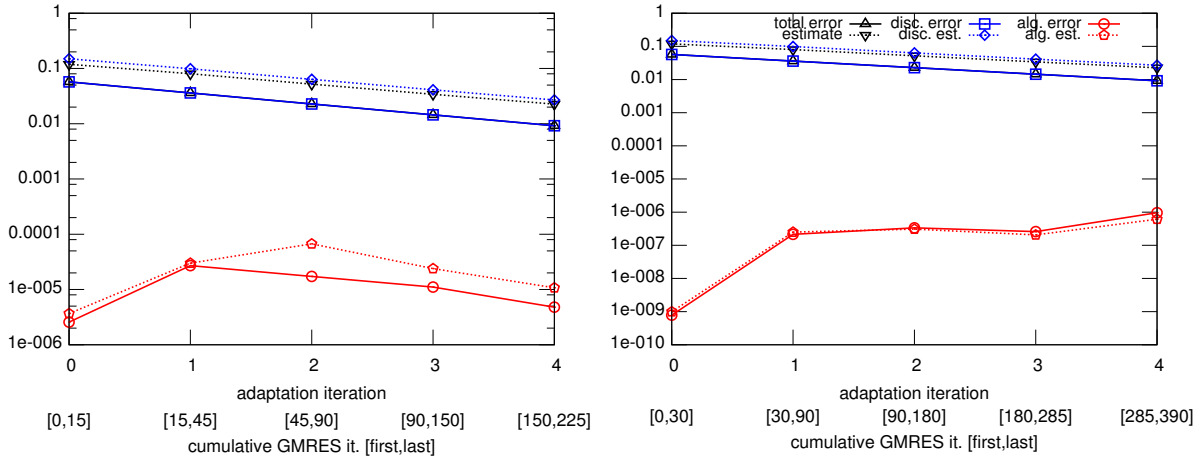


Figure 6: Example 1, development of the total, discretization, and algebraic errors and of their respective estimates during the adaptive process for local (left), and classical with  $10^{-6}$  relative tolerance (right) stopping criteria

Next, Figures 7–9 compare the actual and estimated distributions of the total, discretization, and algebraic errors. We observe that our prediction of *distribution* of the discretization error is sharp on meshes not including hanging nodes whereas such a sharpness is lost a little when hanging nodes appear. That is the price we pay for an economical computation of flux reconstructions as we do not construct a matching submesh of the original (nonmatching) mesh. Importantly, the singularity is well revealed by the estimators and consequently mesh in its vicinity is highly refined. The figures show that our estimates provide a very good prediction of the algebraic error distribution even on meshes with hanging nodes.

Finally, Figure 10 (left) compares the error measured in the broken  $H^1$ -seminorm of the computational solution  $u_h^i$  as a function of the cumulated GMRES iterations for the local stopping criteria and the classical criteria with the four relative tolerances  $10^{-4}$ ,  $10^{-6}$ ,  $10^{-9}$ , and  $10^{-12}$ . We observe that results for the local adaptive stopping criteria are comparable with those resulting from classical stopping criteria with the tolerance  $10^{-4}$ . On the other hand, the local adaptive stopping criteria lead to much fewer GMRES iterations compared to the classical stopping criteria with tolerances  $10^{-6}$ ,  $10^{-9}$ , and  $10^{-12}$ , with a minimal loss of accuracy. Figure 10 (right) then gives a similar comparison for the effectivity indices, given as the ratio  $\eta^i / \|\nabla_h(u - u_h^i)\|$ . Their value is stable around 2.3.

## 7.2 Example 2: steep gradient solution

We consider  $\Omega \in (0, 1) \times (0, 1)$  and solve the Poisson equation (1.1a) with a nonhomogeneous Dirichlet boundary condition. We prescribe the source term and the boundary condition such that the exact solution has the form

$$u(x_1, x_2) = \arctan(36x_1).$$

We employ the same initial setting as in the previous example. The computation is started on a matching triangular grid with 288 elements. As in the previous example, we show results for five successive meshes resulting from four levels of adaptation. Figure 11 shows the evolution of the estimators  $\eta_R^i$ ,  $\eta_{\text{PNC}}^i$ ,  $\eta_{\text{FD}}^i$ ,  $\eta_{\text{FA}}^i$ ,  $\eta_{\text{FNCD}}^i$ ,  $\eta_{\text{FNCA}}^i$ , and  $\eta_{\text{rem}}^i$ . The number of hanging nodes is much smaller in comparison with the Example 1. As a result, the flux nonconformity estimators are not dominant in this example even on meshes with hanging nodes. We observe that the residual estimator is by one to two orders of magnitude smaller than the (dominating) discretization flux estimator. The behavior of all the estimators and efficiency of the stopping criteria can again be appreciated.

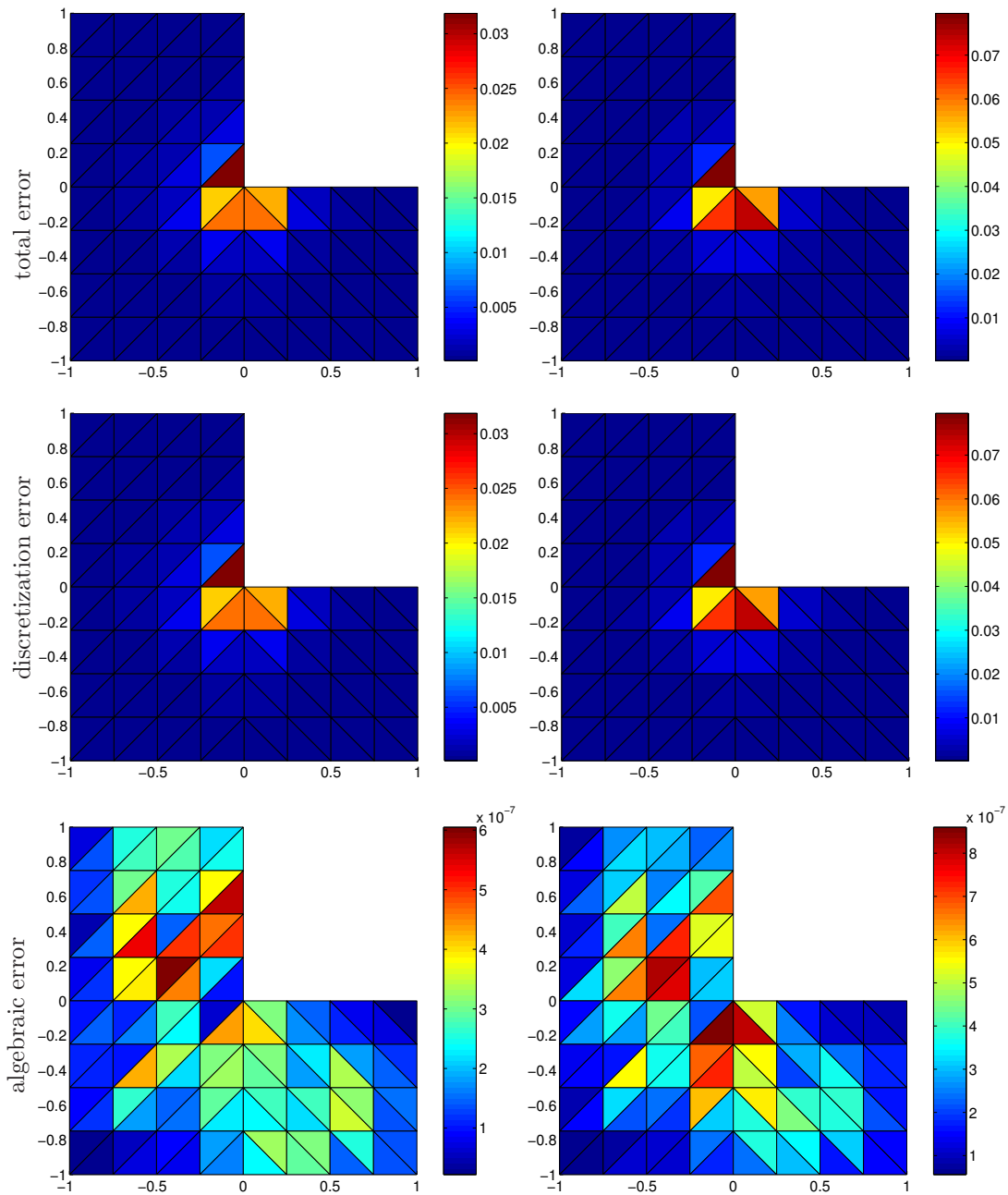


Figure 7: Example 1, distribution of the total, discretization, and algebraic errors (left) and of their estimates (right) on the initial mesh, local adaptive stopping criteria (4.1)

In Figure 12 the evolution of both the errors and the estimators (total, discretization, and algebraic) through the whole adaptation process is displayed. As in the previous section, our predictions match very precisely the reality. Figures 13–15 then show the actual and estimated distributions of the total, discretization, and algebraic errors when local stopping criteria (4.1) are applied. We observe that the steep gradient region is well predicted as the main error source. Unfortunately, the discretization error is overestimated in triangles with a hanging node on the final mesh; on the other hand, the estimates again provide a sharp prediction of the algebraic error distribution.

Finally, Figure 16 (left) compares the error measured in the broken  $H^1$ -seminorm and Figure 16 (right) the effectivity index as a function of the cumulated GMRES iterations for the local stopping criteria and

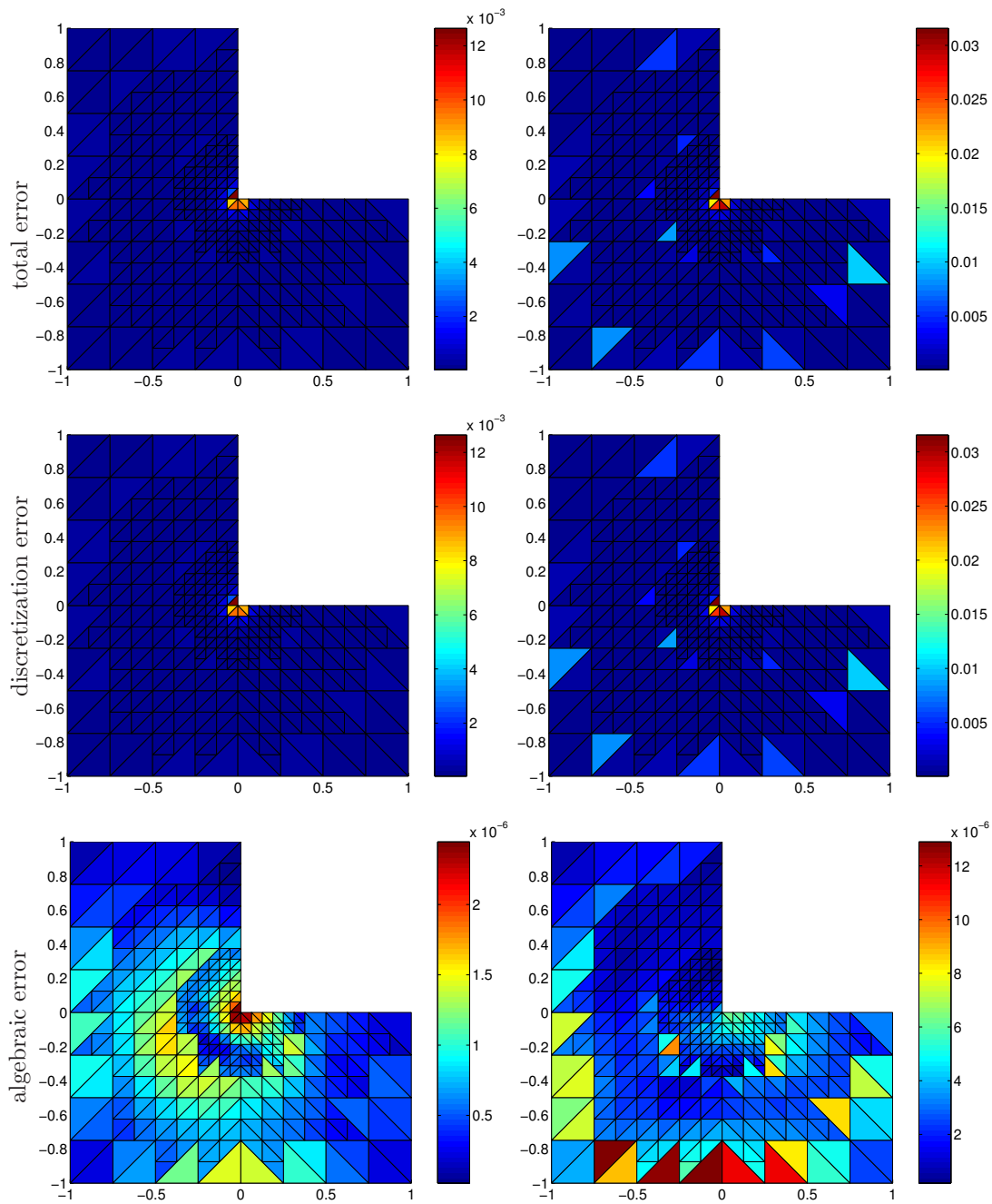


Figure 8: Example 1, distribution of the total, discretization, and algebraic errors (left) and of their estimates (right) after two levels of mesh refinement, local adaptive stopping criteria (4.1)

for the classical ones with the four relative tolerances. We observe that the local stopping criteria are more expensive here than the classical one with the relative tolerance  $10^{-4}$  and comparable to  $10^{-6}$ . This manifests the safeguard role of our criteria in a case of strong spatial singularity; note that the distributions of the discretization and algebraic errors are very much different here (the discretization error is highest in the boundary layer whereas the algebraic one on coarsest mesh elements). Important iterations savings are, however, still observed in comparison with the stronger classical criteria  $10^{-9}$  and  $10^{-12}$ . We also remark that the global stopping criteria (4.2) only require about half of the iterations in comparison with the local ones (4.1), with subtle, but present, precision loss (not shown). Finally, we can see that the variation of the

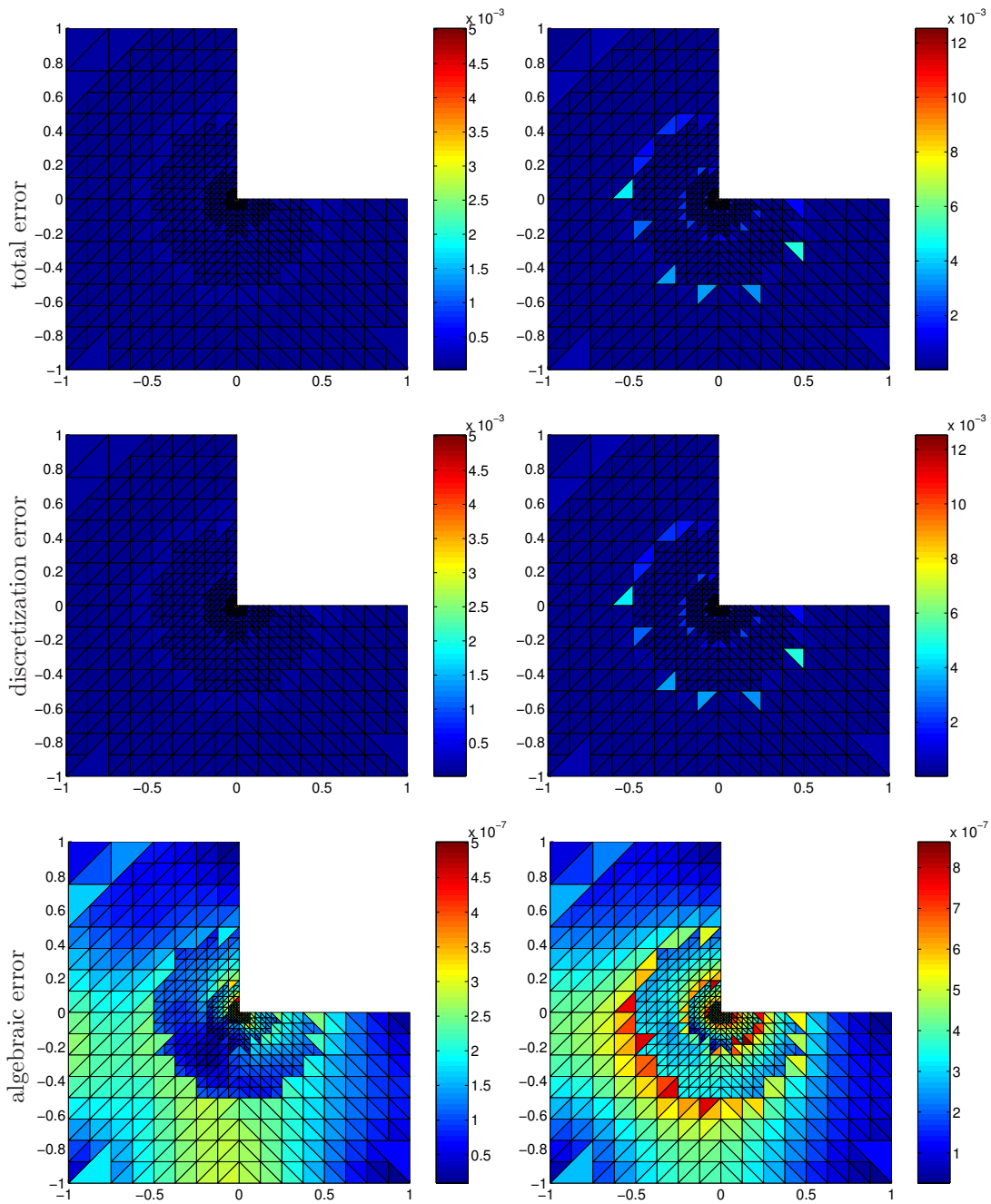


Figure 9: Example 1, distribution of the total, discretization, and algebraic errors (left) and of their estimates (right) after four levels of mesh refinement, local adaptive stopping criteria (4.1)

effectivity index is not much influenced by the presence of hanging nodes, as the number of hanging nodes is not substantial here in comparison with the number of triangles.

## 8 Conclusion

We have presented a posteriori error estimates for the discontinuous Galerkin method applied to the Poisson equation which include the algebraic error. Our technique is able to deal with meshes containing hanging

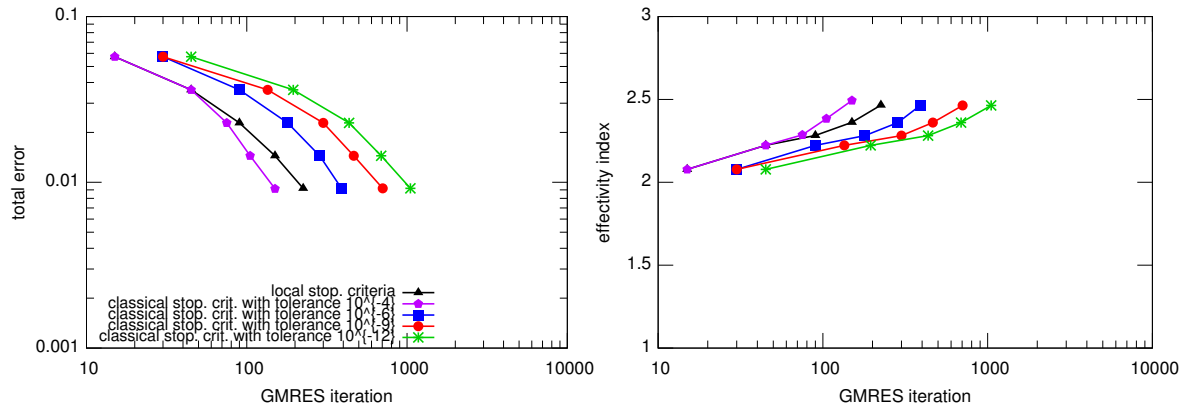


Figure 10: Example 1, comparison of the error in the broken  $H^1$ -seminorm (left) and of the effectivity indices (right) for the local stopping criteria (4.1) and for the four choices of tolerance for classical stopping criteria

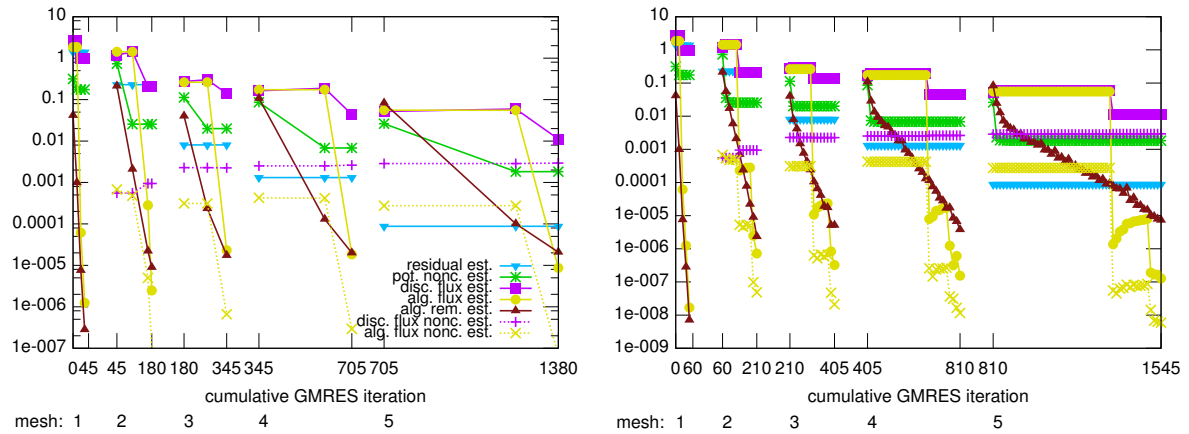


Figure 11: Example 2, development of the estimators on the individual meshes for local (left) and classical with  $10^{-6}$  relative tolerance (right) stopping criteria

nodes and with a variable polynomial approximation degree without the necessity to construct any matching submesh. The derived estimates are guaranteed and locally efficient for the error measured in the broken  $H^1$ -seminorm/in the DG-norm. We presented two numerical examples showing that the proposed Algorithm 4.1 gives approximate solutions that are not severely influenced by the algebraic error and ensures that the corresponding algebraic systems are not over-solved. Importantly, the derived estimates are able to predict the local distribution of the algebraic and discretization errors, albeit the prediction is not extremely sharp for meshes including hanging nodes.

## References

- [1] M. AINSWORTH, *Robust a posteriori error estimation for nonconforming finite element approximation*, SIAM J. Numer. Anal., 42 (2005), pp. 2320–2341.
- [2] M. AINSWORTH AND R. RANKIN, *Fully computable error bounds for discontinuous Galerkin finite element approximations on meshes with an arbitrary number of levels of hanging nodes*, SIAM J. Numer. Anal., 47 (2010), pp. 4112–4141.



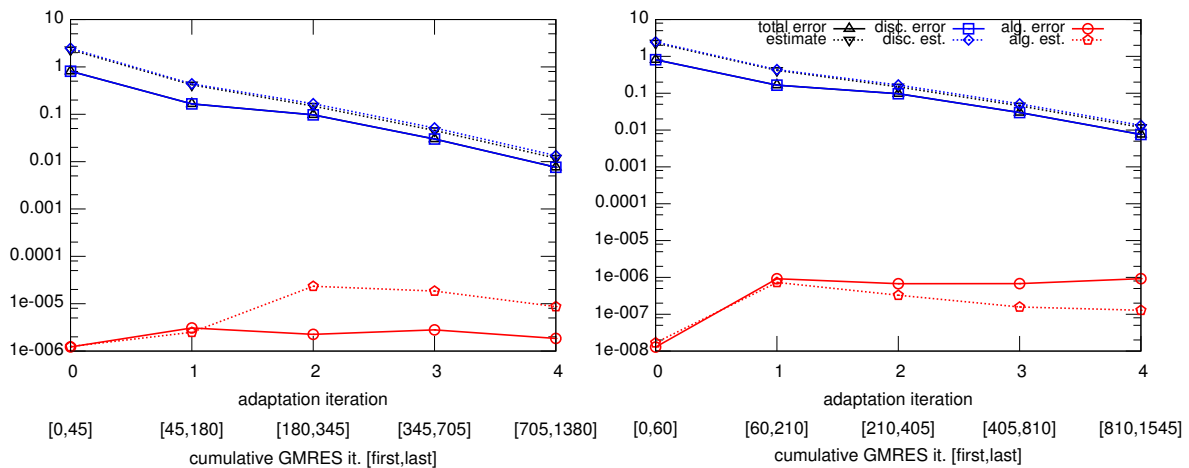


Figure 12: Example 2, development of the total, discretization, and algebraic errors and of their respective estimates during the adaptive process for local (left), and classical with  $10^{-6}$  relative tolerance (right) stopping criteria

- [3] —, *Constant free error bounds for nonuniform order discontinuous Galerkin finite-element approximation on locally refined meshes with hanging nodes*, IMA J. Numer. Anal., 31 (2011), pp. 254–280.
- [4] M. ARIOLI, *A stopping criterion for the conjugate gradient algorithm in a finite element method framework*, Numer. Math., 97 (2004), pp. 1–24.
- [5] M. ARIOLI, E. H. GEORGIOULIS, AND D. LOGHIN, *Stopping criteria for adaptive finite element solvers*, SIAM J. Sci. Comput., 35 (2013), pp. A1537–A1559.
- [6] M. ARIOLI, J. LIESEN, A. MIĘDLAR, AND Z. STRAKOŠ, *Interplay between discretization and algebraic computation in adaptive numerical solution of elliptic PDE problems*, GAMM-Mitt., 36 (2013), pp. 102–129.
- [7] M. ARIOLI AND D. LOGHIN, *Stopping criteria for mixed finite element problems*, Electron. Trans. Numer. Anal., 29 (2007/08), pp. 178–192.
- [8] M. ARIOLI, D. LOGHIN, AND A. J. WATHEN, *Stopping criteria for iterations in finite element methods*, Numer. Math., 99 (2005), pp. 381–410.
- [9] R. BECKER, C. JOHNSON, AND R. RANNACHER, *Adaptive error control for multigrid finite element methods*, Computing, 55 (1995), pp. 271–288.
- [10] F. BREZZI AND M. FORTIN, *Mixed and hybrid finite element methods*, vol. 15 of Springer Series in Computational Mathematics, Springer-Verlag, New York, 1991.
- [11] I. CHEDDADI, R. FUČÍK, M. I. PRIETO, AND M. VOHRALÍK, *Computable a posteriori error estimates in the finite element method based on its local conservativity: improvements using local minimization*, ESAIM Proc., 24 (2008), pp. 77–96.
- [12] S. COCHEZ-DHONDT AND S. NICAISE, *Equilibrated error estimators for discontinuous Galerkin methods*, Numer. Methods Partial Differential Equations, 24 (2008), pp. 1236–1252.
- [13] D. A. DI PIETRO AND A. ERN, *Mathematical aspects of discontinuous Galerkin methods*, vol. 69 of Mathématiques & Applications (Berlin) [Mathematics & Applications], Springer, Heidelberg, 2012.
- [14] J. DRKOŠOVÁ, A. GREENBAUM, M. ROZLOŽNÍK, AND Z. STRAKOŠ, *Numerical stability of GMRES*, BIT, 35 (1995), pp. 309–330.

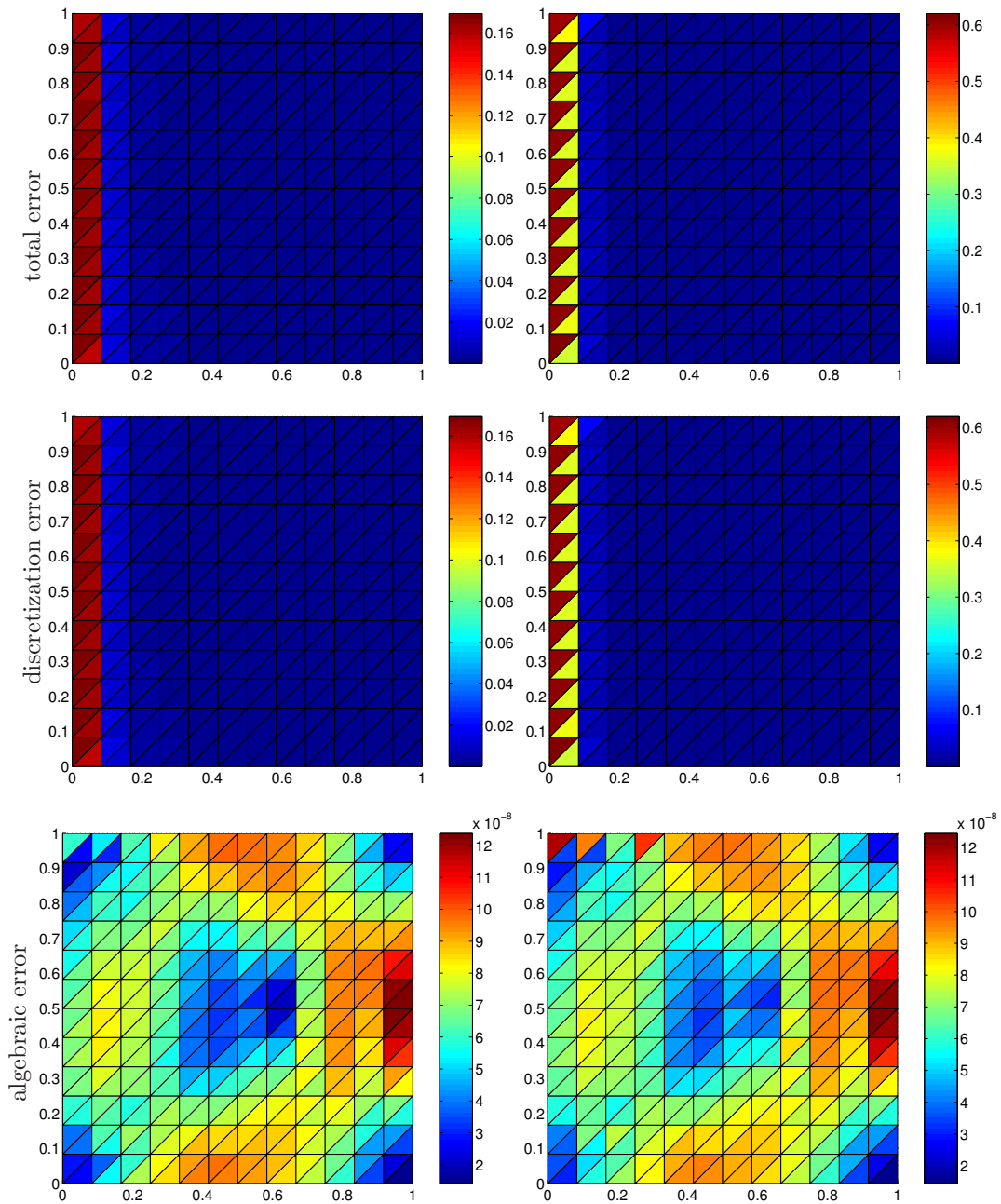


Figure 13: Example 2, distribution of the total, discretization, and algebraic errors (left) and of their estimates (right) on the initial mesh, local adaptive stopping criteria (4.1)

- [15] A. ERN, A. F. STEPHANSEN, AND M. VOHRALÍK, *Guaranteed and robust discontinuous Galerkin a posteriori error estimates for convection–diffusion–reaction problems*, J. Comput. Appl. Math., 234 (2010), pp. 114–130.
- [16] A. ERN AND M. VOHRALÍK, *Flux reconstruction and a posteriori error estimation for discontinuous Galerkin methods on general nonmatching grids*, C. R. Math. Acad. Sci. Paris, 347 (2009), pp. 441–444.
- [17] ———, *Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs*, SIAM J. Sci. Comput., 35 (2013), pp. A1761–A1791.

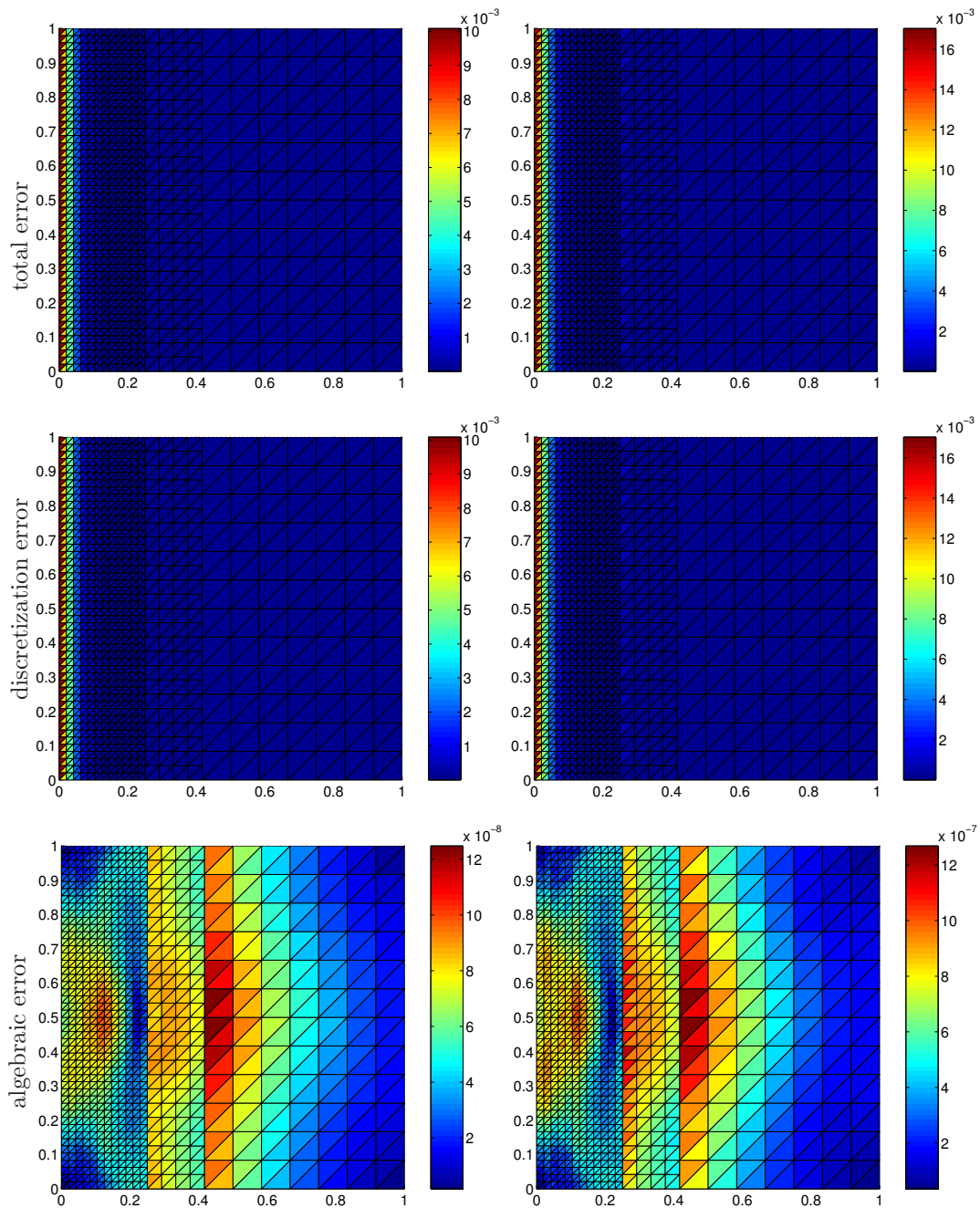


Figure 14: Example 2, distribution of the total, discretization, and algebraic errors (left) and of their estimates (right) after two levels of mesh refinement, local adaptive stopping criteria (4.1)

- [18] A. GREENBAUM, M. ROZLOŽNÍK, AND Z. STRAKOŠ, *Numerical behaviour of the modified Gram-Schmidt GMRES implementation*, BIT, 37 (1997), pp. 706–719.
- [19] P. JIRÁNEK, Z. STRAKOŠ, AND M. VOHRALÍK, *A posteriori error estimates including algebraic error and stopping criteria for iterative solvers*, SIAM J. Sci. Comput., 32 (2010), pp. 1567–1590.
- [20] 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 (2003), pp. 2374–2399.

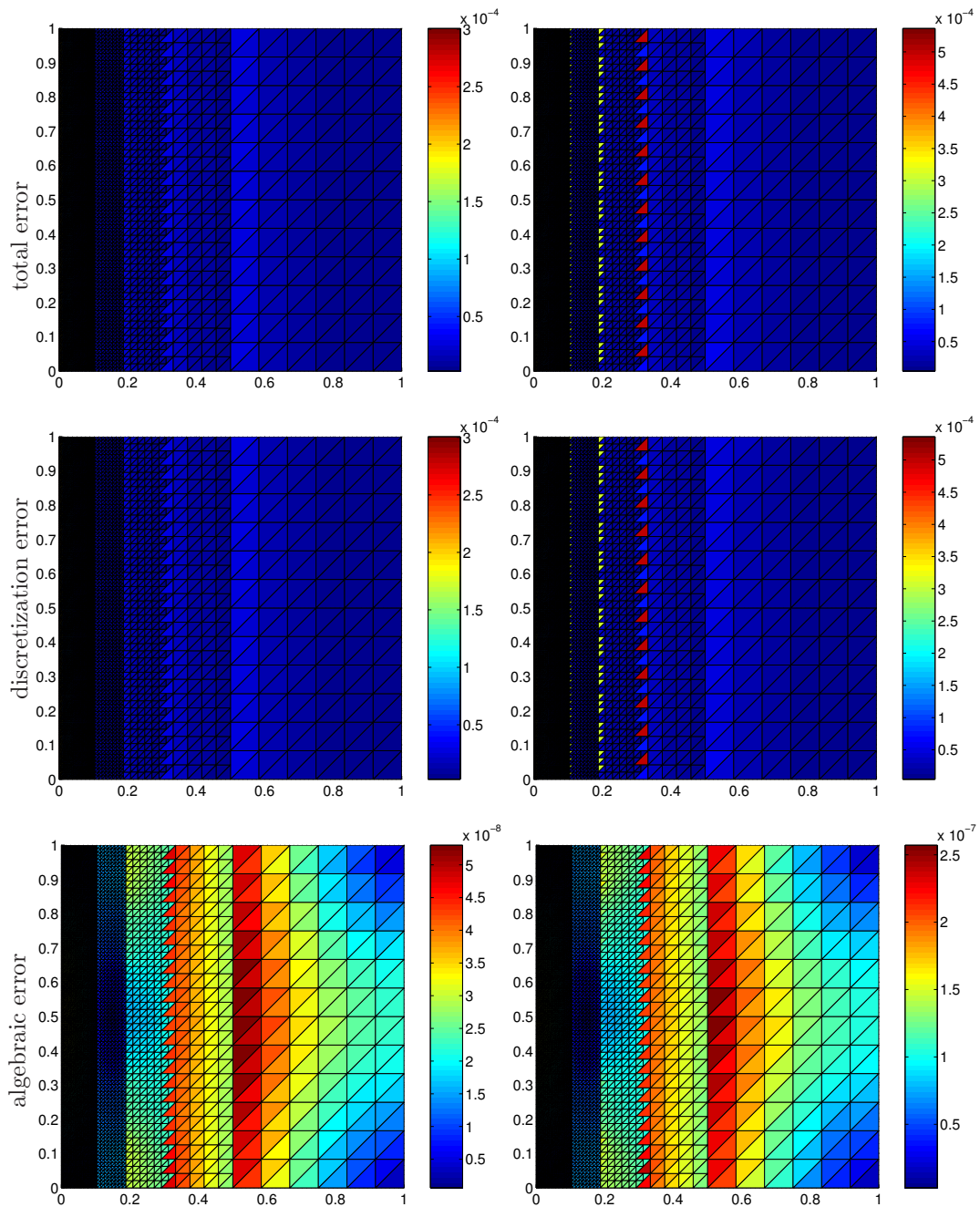


Figure 15: Example 2, distribution of the total, discretization, and algebraic errors (left) and of their estimates (right) after four levels of mesh refinement, local adaptive stopping criteria (4.1)

- [21] K. Y. KIM, *A posteriori error analysis for locally conservative mixed methods*, Math. Comp., 76 (2007), pp. 43–66.
- [22] J. LIESEN AND Z. STRAKOŠ, *Krylov Subspace Methods. Principles and Analysis*, Numerical Mathematics and Scientific Computation, Oxford University Press, Oxford, United Kingdom, 2013.
- [23] D. MEIDNER, R. RANNACHER, AND J. VIHAREV, *Goal-oriented error control of the iterative solution of finite element equations*, J. Numer. Math., 17 (2009), pp. 143–172.

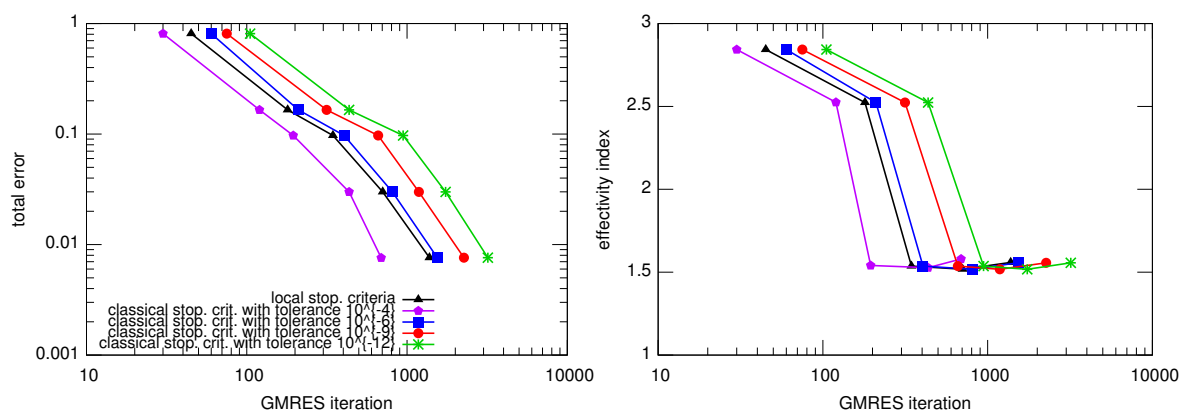


Figure 16: Example 2, comparison of the error in the broken  $H^1$ -seminorm (left) and of the effectivity indices (right) for the local stopping criteria (4.1) and for the four choices of tolerance for classical stopping criteria

- [24] S. NICAISE, *A posteriori error estimations of some cell-centered finite volume methods*, SIAM J. Numer. Anal., 43 (2005), pp. 1481–1503.
- [25] C. C. PAIGE, M. ROZLOŽNÍK, AND Z. STRAKOŠ, *Modified Gram–Schmidt (MGS), least squares, and backward stability of MGS-GMRES*, SIAM J. Matrix Anal. Appl., 28 (2006), pp. 264–284.
- [26] L. E. PAYNE AND H. F. WEINBERGER, *An optimal Poincaré inequality for convex domains*, Arch. Rational Mech. Anal., 5 (1960), pp. 286–292.
- [27] G. V. PENCHEVA, M. VOHRALÍK, M. F. WHEELER, AND T. WILDEY, *Robust a posteriori error control and adaptivity for multiscale, multinumercs, and mortar coupling*, SIAM J. Numer. Anal., 51 (2013), pp. 526–554.
- [28] A. QUARTERONI AND A. VALLI, *Numerical approximation of partial differential equations*, vol. 23 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1994.
- [29] R. RANNACHER, A. WESTENBERGER, AND W. WOLLNER, *Adaptive finite element solution of eigenvalue problems: balancing of discretization and iteration error*, J. Numer. Math., 18 (2010), pp. 303–327.
- [30] K. REKTORYS, *Variational methods in mathematics, science and engineering*, D. Reidel Publishing Co., Dordrecht, 1977. Translated from the Czech by Michael Basch.
- [31] Y. SAAD AND M. H. SCHULTZ, *GMRES: a generalized minimal residual algorithm for solving non-symmetric linear systems*, SIAM J. Sci. Statist. Comput., 7 (1986), pp. 856–869.
- [32] P. SILVESTER, *Symmetric quadrature formulae for simplexes*, Math. Comp., 24 (1970), pp. 95–100.
- [33] Z. STRAKOŠ AND P. TICHÝ, *Error estimation in preconditioned conjugate gradients*, BIT, 45 (2005), pp. 789–817.
- [34] R. VERFÜRTH, *A review of a posteriori error estimation and adaptive mesh-refinement techniques*, Teubner-Wiley, Stuttgart, 1996.
- [35] M. F. WHEELER AND I. YOTOV, *A posteriori error estimates for the mortar mixed finite element method*, SIAM J. Numer. Anal., 43 (2005), pp. 1021–1042.