FULLY COMPUTABLE ERROR ESTIMATION OF A NONLINEAR, POSITIVITY-PRESERVING DISCRETIZATION OF THE CONVECTION-DIFFUSION-REACTION EQUATION[*]

ALEJANDRO ALLENDES[†], GABRIEL R. BARRENECHEA[‡], AND RICHARD RANKIN[§]

Abstract. This work is devoted to the proposal, analysis, and numerical testing of a fully computable a posteriori error bound for a class of nonlinear discretizations of the convection-diffusion-reaction equation. The type of discretization we consider is nonlinear, since it has been built with the aim of preserving the discrete maximum principle. Under mild assumptions on the stabilizing term, we obtain an a posteriori error estimator that provides a certified upper bound on the norm of the error. Under the additional assumption that the stabilizing term is both Lipschitz continuous and linearity preserving, the estimator is shown to be locally efficient. We present examples of discretizations that satisfy these two requirements, and the theory is illustrated by several numerical experiments in two and three space dimensions.

Key words. a posteriori error estimation; shock-capturing method; fully computable error bound; algebraic flux correction scheme.

AMS subject classifications. 65N15, 65N30, 65N50.

1. Introduction.

In this work we address the problem of how to design an adaptive algorithm for a positivity-preserving discretization of the convection-diffusion-reaction equation. More precisely, we present a fully computable upper bound for the discretization error, which is also shown to be locally efficient. We start by discussing the type of discretization we consider. In this work we focus on nonlinear discretizations, referred to as *shock-capturing* methods. The reason these methods are nonlinear comes from the fact that linear monotone methods are usually highly diffusive, and that leads to non-optimal convergence (see [21]). Hence, various nonlinear shock-capturing methods have been developed over the years. These shock-capturing schemes have been designed, originally, to approximate conservation laws within a finite volume context. Nevertheless, there are many examples of this sort of discretization in the finite element context, especially aimed at the solution of the convection-dominated convection-diffusion equation (see [23, 24], and the references therein, for extensive reviews). Some of the above-mentioned methods satisfy the, very desirable, property of preserving positivity, or, in other words, they respect the discrete maximum principle (DMP). Up to our knowledge, the first nonlinear discretization of the convection-diffusion equation satisfying the DMP is the one proposed in [32]. More recent approaches have been presented in the works [13, 14, 19, 7], amongst others. The particular type of discretization considered in this paper is the Algebraic Flux-Correction (AFC) method. The origins of this method can be tracked back to [12, 40], and more recently it has been revisited and revamped, especially by D. Kuzmin and co-workers (see, e.g., [29, 28, 27, 30]).

[*] Submitted March 6, 2017.

Funding: AA is supported by the USM grant 116.12.1. GRB is partially funded by CONICYT-Chile, through Project MEC 80140118. RR is supported by BASAL PFB03 CMM project, Universidad de Chile.

[†] Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile. (alejandro.allendes@usm.cl).

[‡] Department of Mathematics and Statistics, University of Strathclyde, 26 Richmond Street, Glasgow, G1 1XH. (gabriel.barrenechea@strath.ac.uk).

[§] Departamento de Matemática, Universidad Técnica Federico Santa María, Valparaíso, Chile. (richard.rankin@usm.cl).

The a posteriori error estimation for the convection-diffusion equation has been a problem that has received a lot of attention over the last two decades. An exhaustive review of the different estimators proposed over the last years is beyond the scope of this work, and we will only mention [37, 6, 34, 25] as examples of estimators obtained using different techniques. The robustness of the estimator, this is, to be able to prove that the equivalence constants between the error and the estimator do not depend on how convection-dominated the problem is, has only been achieved by modifying the norm in which the error is measured. The first result in this direction is [38], where residual error estimators were proven to be robust with respect to a norm that includes a dual norm of the convective derivative of the error (see also [35] for a more recent advance). Now, because of the presence of unknown constants in the upper bounds, none of the above references addresses the problem of producing a fully computable error bound. This is, an estimator whose numerical value is an upper bound for the actual error. This sort of estimator has been proposed for different problems, but their application to convection-diffusion-type equations is more recent (see [20, 1]). Finally, fewer a posteriori error estimators have been proposed for shock-capturing type methods. For example, in [16] the focus was to approximate conservation laws, while in [17, 26] the problem of interest was the convection-diffusion equation. In these last works the fact that the respective estimator is an upper bound (up to an unknown constant) for the error was proved, but no local lower bound was shown. As a matter of fact, as far as we are aware of, no a posteriori error estimator has been proved to be equivalent to the error for a nonlinear, positivity-preserving discretization of the convection-diffusion equation.

The purpose of this work is to bridge the gap mentioned at the end of the last paragraph. In addition, the a posteriori bound presented in this work is fully computable. We limit the analysis to the case of piecewise linear discretization since the analysis of positivity-preserving methods is restricted to the lowest order case. We impose some basic hypotheses on the discretization under which certified upper bounds and local efficiency are proved. More precisely, if we write the discrete method in the usual way, this is, as the sum of the Galerkin part, plus a stabilizing term, then the stabilizing term is supposed to be locally Lipschitz continuous and linearity preserving, which are properties that are considered desirable for this type of scheme. This last property has been linked to enhanced accuracy in smooth regions, but, to the best of our knowledge, no proof of a result of this kind has been given. The only exception, as far as we are aware, is the work [8], where the combination of Lipschitz continuity and linearity preservation was used to prove optimal convergence of the method proposed therein. Therefore, one aim of this work is to contribute to the understanding of why linearity preservation is a desirable property for a scheme of this kind to satisfy. As a matter of fact, that property is at the heart of our proof of the local efficiency of the estimator. However, as is standard for estimators of the energy norm of the error, the local lower bounds show a dependency on the local Péclet number. For estimators satisfying similar local lower bounds, see, e.g., [37, 11]. This behaviour has been referred to as semi-robust.

Finally, there are methods that satisfy the properties required for the analysis presented herein. In fact, our analysis will be applied to the methods proposed recently in [8, 9], which satisfy our assumptions (although we keep the presentation as general as possible). In addition, these methods satisfy the DMP in meshes that fullfil a standard hypothesis. More precisely, the hypothesis imposed in [8] for the validity of the DMP is the satisfaction of Xu and Zikatanov's condition (see [39]). In two space dimensions this reduces to imposing that the mesh is Delaunay (in 3D it is slightly more technical).

The rest of the manuscript is organised as follows. In Section 2 we present the notations to be used throughout. Section 3 is devoted to the presentation of the problem of interest, and a prototype of a nonlinear discretization of the kind considered in this work. Also, in that section we present the main hypotheses that the considered discretization needs to satisfy. The main results of this paper, namely the construction and analysis of the a posteriori error estimator, are given in Section 4. In Section 5 we specify the nonlinear discretization, and several numerical results showing the performance of the estimator are presented in Section 6. Finally, some conclusions are drawn.

2. Preliminaries. We shall use standard notation for Sobolev and Lebesgue spaces, norms, and inner products. Namely, for a bounded domain $G \subset \mathbb{R}^d$ where $d = 2, 3$: $L^2(G)$ denotes the space of square integrable functions over G, $H^1(G)$ is the usual Sobolev space and $H_0^1(G)$ denotes the subspace of $H^1(G)$ consisting of functions whose trace is zero on the boundary of G; $(\cdot, \cdot)_G$ denotes the inner product in $L^2(G)$ (or in $L^2(G)^d$ when necessary). The norm (seminorm) of the space $H^m(G)$ is denoted by $\| \cdot \|_{m,G}$ ($| \cdot |_{m,G}$) and the norm of the Lebesgue space $L^2(G)$ is denoted by $\| \cdot \|_{0,G}$. Finally, for $\ell \geq 0$, $\mathbb{P}_\ell(G)$ denotes the space of polynomials on G of total degree at most ℓ.

For convenience, we shall summarize all the notation used throughout the manuscript related to the partition of the domain. The problem of interest will be posed over a domain $\Omega \subset \mathbb{R}^d, d = 2, 3$, which is open, bounded, and polygonal/polyhedral, and has Lipschitz boundary. For a fixed triangulation $\mathscr{T}$ of $\bar{\Omega}$, belonging to a shape regular family of triangulations (in the sense of Ciarlet [15]), based on elements K that can be triangles or tetrahedrons, let

- $\mathcal{F}$ denote the set of all element edges(2D)/faces(3D), $\mathcal{F}_I \subset \mathcal{F}$ denote the set of interior edges(2D)/faces(3D), and $\mathcal{F}_{\partial\Omega} \subset \mathcal{F}$ denote the set of boundary edges(2D)/faces(3D);
- $\mathcal{E}$ denote the set of all edges in 2D or 3D, and $\mathcal{E}_I \subset \mathcal{E}$ denote the set of all interior edges in 2D or 3D;
- $\mathcal{V}$ index the set $\{\boldsymbol{x}_n\}_{n \in \mathcal{V}}$ of all the vertices in the triangulation;
- $\mathcal{V}_\Omega$ index the set $\{\boldsymbol{x}_n\}_{n \in \mathcal{V}_\Omega}$ of all the interior vertices in the triangulation;
- $\omega_n := \{K \in \mathscr{T} : \ \boldsymbol{x}_n \in K\}$ which is the set containing the elements for which $\boldsymbol{x}_n$ is a vertex;
- $\mathcal{F}_n$ denote the set of element edges(2D)/faces(3D) that have $\boldsymbol{x}_n$ as a vertex.

For elements $K \in \mathscr{T}$, let

- $\mathcal{F}_K$ denote the set containing the edges(2D)/faces(3D) of element K;
- $\mathcal{E}_K$ denote the set containing the edges in 2D or 3D of element K;
- $\omega_K := \{K' \in \mathscr{T} : \ K' \cap K \neq \emptyset\}$;
- $\mathcal{V}_K$ index the set $\{\boldsymbol{x}_n\}_{n \in \mathcal{V}_K}$ of all the vertices of element K;
- $|K|$ denote the area/volume of element K, and h_K denote the diameter of K;
- $\boldsymbol{n}_\gamma^K$ denote the unit exterior normal vector to the edge/face $\gamma \in \mathcal{F}_K$;
- $v_{|K}$ denote the restriction of v to the element K.

For edges(2D)/faces(3D) $\gamma \in \mathcal{F}$, let:

- $\omega_\gamma := \{K \in \mathscr{T} : \ \gamma \in \mathcal{F}_K\}$;
- $\mathcal{V}_\gamma$ index the set $\{\boldsymbol{x}_n\}_{n \in \mathcal{V}_\gamma}$ of all the vertices of the edge/face γ;

For edges(2D or 3D) $E \in \mathcal{E}$, let:

- $\mathcal{V}_E$ index the set $\{\boldsymbol{x}_n\}_{n \in \mathcal{V}_E}$ of all the vertices of the edge E;
- $|E|$ denote the length of the edge E;

- $\boldsymbol{t}_E$ denote a unit tangent vector to edge E. Its direction is of no importance.

For vertices in $\mathcal{V}$, let:

- $\mathcal{V}_i$ index the set $\{\boldsymbol{x}_n\}_{n\in\mathcal{V}_i}$ of all vertices which share an edge with vertex $\boldsymbol{x}_i$.

We note that in 2D, $\mathcal{E} = \mathcal{F}$, $\mathcal{E}_I = \mathcal{F}_I$ and $\mathcal{E}_K = \mathcal{F}_K$ but that this is not the case in 3D. For $K \in \mathscr{T}$ we define $\Pi_K : L^2(K) \to \mathbb{P}_1(K)$ to be the orthogonal projection operator characterized as

$$(f - \Pi_K(f), p)_K = 0 \quad \text{for all } p \in \mathbb{P}_1(K), \tag{1}$$

and we define $\bar{\mathsf{v}}_K := |K|^{-1} \int_K \mathsf{v}$. For a given partition $\mathscr{T}$ of the domain Ω, we make use of the following piecewise linear finite element spaces

$$\mathbb{W}(\mathscr{T}) := \{\mathsf{v} \in \mathcal{C}^0(\bar{\Omega}) : \ \mathsf{v}_{|K} \in \mathbb{P}_1(K) \ \forall \ K \in \mathscr{T}\} \quad , \quad \mathbb{V}(\mathscr{T}) := \mathbb{W}(\mathscr{T}) \cap H_0^1(\Omega). \tag{2}$$

For $n \in \mathcal{V}$, we let λ_n denote the usual continuous piecewise linear basis function associated to the vertex $\boldsymbol{x}_n$, characterized by the conditions $\lambda_n \in \mathbb{W}(\mathscr{T})$ and $\lambda_n(\boldsymbol{x}_m) = \delta_{nm}$ for all $m \in \mathcal{V}$, where δ_{nm} denotes the Kronecker delta. We note that we will abuse notation by using sets such as ω_K to denote the region $\cup_{K'\in\omega_K} K'$ when we write expressions such as $\mathbb{P}_1(\omega_K)$. Finally, throughout the manuscript we shall use C to denote any positive constant which is independent of any mesh size or any physical parameter related with the problem.

3. Model problem and nonlinear stabilized discretizations. We consider the following convection–reaction–diffusion problem

$$\begin{cases} -\varepsilon\Delta \mathsf{u} + \mathbf{b}\cdot\nabla\mathsf{u} + \kappa\mathsf{u} = \mathsf{f} & \text{in } \Omega, \\ \qquad\qquad \mathsf{u} = \mathsf{u}_D & \text{on } \partial\Omega. \end{cases} \tag{3}$$

The weak formulation of (3) reads as follows: *find* $\mathsf{u} \in H^1(\Omega)$ *such that* $\mathsf{u} = \mathsf{u}_D$ *on* $\partial\Omega$ *and*

$$\mathcal{B}(\mathsf{u},\mathsf{v}) = (\mathsf{f},\mathsf{v})_\Omega \qquad \text{for all } \mathsf{v} \in H_0^1(\Omega), \tag{4}$$

where the bilinear form is given by $\mathcal{B}(\mathsf{u},\mathsf{v}) := \varepsilon(\nabla\mathsf{u},\nabla\mathsf{v})_\Omega + (\kappa\mathsf{u} + \mathbf{b}\cdot\nabla\mathsf{u},\mathsf{v})_\Omega$. For simplicity of the presentation we will suppose that ε and κ are constants and that $\mathsf{u}_{D|\gamma} \in \mathbb{P}_1(\gamma)$ for all $\gamma \in \mathcal{F}_{\partial\Omega}$. We also suppose that $\varepsilon > 0$, $\kappa \geq 0$, $\mathbf{b} \in \boldsymbol{W}^{1,\infty}(\Omega)$ is a solenoidal field (that is $\mathbf{div}\ \mathbf{b} = 0$), $\mathsf{f} \in L^2(\Omega)$, and $\mathsf{u}_D \in \mathcal{C}^0(\partial\Omega)$. From these assumptions the existence and uniqueness of a solution to (4) follows by standard arguments (see [18]). To approximate (4) we consider a nonlinear discretization that reads as follows: *find* $\mathsf{u}_{\mathscr{T}} \in \mathbb{W}(\mathscr{T})$ *such that* $\mathsf{u} = \mathsf{u}_D$ *on* $\partial\Omega$ *and*

$$\mathcal{B}(\mathsf{u}_{\mathscr{T}},\mathsf{v}_{\mathscr{T}}) + \mathcal{S}(\mathsf{u}_{\mathscr{T}};\mathsf{u}_{\mathscr{T}},\mathsf{v}_{\mathscr{T}}) = (\mathsf{f},\mathsf{v}_{\mathscr{T}})_\Omega \qquad \text{for all } \mathsf{v}_{\mathscr{T}} \in \mathbb{V}(\mathscr{T}), \tag{5}$$

where $\mathcal{S}$ is a nonlinear stabilization term. We will suppose this problem has at least one solution $\mathsf{u}_{\mathscr{T}} \in \mathbb{W}(\mathscr{T})$. Despite the nonlinearity of the form $\mathcal{S}$, we will suppose, as it happens in practice, that it is linear in its third argument. In addition, in order to derive fully computable upper bounds, we assume two properties that are satisfied by most stabilized schemes. These are, that the nonlinear stabilization term can be written as a sum of local contributions in such a way that

$$\mathcal{S}(\mathsf{v}_{\mathscr{T}};\mathsf{w}_{\mathscr{T}},\mathsf{z}_{\mathscr{T}}) = \sum_{K\in\mathscr{T}} \mathcal{S}_K(\mathsf{v}_{\mathscr{T}};\mathsf{w}_{\mathscr{T}},\mathsf{z}_{\mathscr{T}|K}) \qquad \text{for all } \mathsf{v}_{\mathscr{T}},\mathsf{w}_{\mathscr{T}},\mathsf{z}_{\mathscr{T}} \in \mathbb{W}(\mathscr{T}), \tag{6}$$

and that

$$\mathcal{S}_K(\mathsf{v}_{\mathscr{T}};\mathsf{w}_{\mathscr{T}},1)=0 \text{ for all } K\in\mathscr{T} \quad \text{for all } \mathsf{v}_{\mathscr{T}},\mathsf{w}_{\mathscr{T}}\in\mathbb{W}(\mathscr{T}). \tag{7}$$

We now list two assumptions that will be the key to establishing the local efficiency of the a posteriori error estimator presented in the next section.

Assumption 1: Local Lipschitz continuity The local contributions, $\mathcal{S}_K(\cdot;\cdot,\cdot)$, are Lipschitz continuous. More precisely, there exists a constant C_{lip}, independent of any mesh size, such that, for all $\mathsf{v}_{\mathscr{T}},\mathsf{w}_{\mathscr{T}},\mathsf{z}_{\mathscr{T}}\in\mathbb{W}(\mathscr{T})$, and all elements $K\in\mathscr{T}$, the following holds

$$|\mathcal{S}_K(\mathsf{v}_{\mathscr{T}};\mathsf{v}_{\mathscr{T}},\mathsf{z}_{\mathscr{T}})-\mathcal{S}_K(\mathsf{w}_{\mathscr{T}};\mathsf{w}_{\mathscr{T}},\mathsf{z}_{\mathscr{T}})|\le \mathsf{C}_{lip}h_K\,|\mathsf{v}_{\mathscr{T}}-\mathsf{w}_{\mathscr{T}}|_{1,\omega_K}\,|\mathsf{z}_{\mathscr{T}}|_{1,\omega_K}\,. \tag{8}$$

Assumption 2: Local linearity preservation For any $K\in\mathscr{T}$ and any $\tilde{\mathsf{u}}\in\mathbb{P}_1(\omega_K)$, the following holds

$$\mathcal{S}_K(\tilde{\mathsf{u}};\mathsf{v}_{\mathscr{T}},\mathsf{z}_{\mathscr{T}})=0 \quad \text{for all } \mathsf{v}_{\mathscr{T}},\mathsf{z}_{\mathscr{T}}\in\mathbb{W}(\mathscr{T}). \tag{9}$$

In Section 5 we will present examples of discretizations that satisfy these requirements.

4. Fully computable a posteriori error estimation.

In order to perform the error analysis, we measure the error in the norm

$$\|\!|\mathsf{v}|\!\|_{\Omega}^2:=\varepsilon\,|\mathsf{v}|_{1,\Omega}^2+\kappa\|\mathsf{v}\|_{0,\Omega}^2. \tag{10}$$

To derive the a posteriori error bound, we follow closely the approach presented in [1, 3] for linear stabilized methods. For this, we introduce a set of equilibrated fluxes $\{g_{\gamma,K}\in\mathbb{P}_1(\gamma),\gamma\in\mathcal{F}_K,K\in\mathscr{T}\}$ satisfying the following two conditions (see [2]):

- **Consistency**:

$$g_{\gamma,K}+g_{\gamma,K'}=0,\quad \text{if } \gamma\in\mathcal{F}_K\cap\mathcal{F}_{K'},\ K,K'\in\mathscr{T},\ K\neq K'\,. \tag{11}$$

- **First order equilibration**: for all $K\in\mathscr{T}$ and all $\lambda\in\mathbb{P}_1(K)$

$$0=(\mathsf{f},\lambda)_K-\mathcal{B}_K\left(\mathsf{u}_{\mathscr{T}},\lambda\right)+\sum_{\gamma\in\mathcal{F}_K}\left(g_{\gamma,K},\lambda\right)_{\gamma}-\mathcal{S}_K\left(\mathsf{u}_{\mathscr{T}};\mathsf{u}_{\mathscr{T}},\lambda\right), \tag{12}$$

where $\mathcal{B}_K\left(\mathsf{u}_{\mathscr{T}},\lambda\right)=\varepsilon(\nabla\mathsf{u}_{\mathscr{T}},\nabla\lambda)_K+(\mathbf{b}\cdot\nabla\mathsf{u}_{\mathscr{T}}+\kappa\mathsf{u}_{\mathscr{T}},\lambda)_K$.

The existence of these fluxes follows from assumption (6) imposed on $\mathcal{S}_K\left(\cdot;\cdot,\cdot\right)$, along with its linearity in its third argument (their construction follows the same lines as in [3, § 6.4]). As a consequence, thanks to the assumption (7) imposed on $\mathcal{S}_K\left(\cdot;\cdot,\cdot\right)$, the element and edge residuals given by

$$\left\{\begin{array}{rcl}\mathscr{R}_K & := & \Pi_K(\mathsf{f})-\Pi_K(\mathbf{b})\cdot\nabla\mathsf{u}_{\mathscr{T}|K}-\kappa\mathsf{u}_{\mathscr{T}|K},\\ \mathscr{R}_{\gamma,K} & := & g_{\gamma,K}-\varepsilon\nabla\mathsf{u}_{\mathscr{T}|K}\cdot\boldsymbol{n}_{\gamma}^{K},\end{array}\right. \tag{13}$$

are compatible in the sense that the problem

$$\left\{\begin{array}{rcll}-\mathbf{div}\,\boldsymbol{\sigma}_K & = & \mathscr{R}_K & \text{in } K,\\ \boldsymbol{\sigma}_K\cdot\boldsymbol{n}_{\gamma}^{K} & = & \mathscr{R}_{\gamma,K} & \text{on } \gamma \text{ for all } \gamma\in\mathcal{F}_K,\end{array}\right. \tag{14}$$

has a solution. In fact, this problem has an infinite number of solutions. The one that will be used to build the a posteriori error estimator is detailed in § 4.1 below.

We are now in a position to derive the estimator. As usual, our starting point is the error equation associated to (4)-(5). Integrating by parts and using (1) and that $\boldsymbol{\sigma}_K$ solves (14), we obtain that, for all $\mathsf{v} \in H_0^1(\Omega)$,

$$\begin{aligned}
\mathcal{B}(\mathsf{u}-\mathsf{u}_{\mathscr{T}},\mathsf{v}) &= \sum_{K\in\mathscr{T}} \left((\mathscr{R}_K,\mathsf{v})_K + \sum_{\gamma\in\mathcal{F}_K} (\mathscr{R}_{\gamma,K},\mathsf{v})_\gamma + (\mathrm{osc}_K,\mathsf{v})_K \right) \\
&= \sum_{K\in\mathscr{T}} \left((-\mathbf{div}\,\boldsymbol{\sigma}_K,\mathsf{v})_K + \sum_{\gamma\in\mathcal{F}_K} (\boldsymbol{\sigma}_K\cdot\boldsymbol{n}_\gamma^K,\mathsf{v})_\gamma + (\mathrm{osc}_K,\mathsf{v}-\bar{\mathsf{v}}_K)_K \right) \\
&= \sum_{K\in\mathscr{T}} \Big((\boldsymbol{\sigma}_K,\nabla\mathsf{v})_K + (\mathrm{osc}_K,\mathsf{v}-\bar{\mathsf{v}}_K)_K \Big) \\
&\le \sum_{K\in\mathscr{T}} \Big(\|\boldsymbol{\sigma}_K\|_{0,K}\,\|\nabla\mathsf{v}\|_{0,K} + \|\mathrm{osc}_K\|_{0,K}\,\|\mathsf{v}-\bar{\mathsf{v}}_K\|_{0,K} \Big) .
\end{aligned}$$

Here, the oscillation term is given by

$$\mathrm{osc}_K := \mathsf{f} - \Pi_K(\mathsf{f}) - (\mathbf{b} - \Pi_K(\mathbf{b}))\cdot\nabla\mathsf{u}_{\mathscr{T}|K}. \tag{15}$$

Hence, the use of the Poincaré inequality, as given in [33, 10], gives

$$\|\mathsf{v}-\bar{\mathsf{v}}_K\|_{0,K} \le \mathsf{C}_{osc}|\!|\!|\mathsf{v}|\!|\!|_K \quad \text{with} \quad \mathsf{C}_{osc} := \min\left\{ \frac{h_K}{\pi\sqrt{\varepsilon}}, \frac{1}{\sqrt{\kappa}} \right\} , \tag{16}$$

which leads us to the following bound for the error equation

$$\mathcal{B}(\mathsf{u}-\mathsf{u}_{\mathscr{T}},\mathsf{v}) \le \left(\sum_{K\in\mathscr{T}} \left(\frac{1}{\sqrt{\varepsilon}}\|\boldsymbol{\sigma}_K\|_{0,K} + \mathsf{C}_{osc}\|\mathrm{osc}_K\|_{0,K} \right)^2 \right)^{\frac{1}{2}} |\!|\!|\mathsf{v}|\!|\!|_\Omega.$$

Finally, taking $\mathsf{v} = \mathsf{u}-\mathsf{u}_{\mathscr{T}} \in H_0^1(\Omega)$ in the last bound, and using the fact that $\mathcal{B}(\mathsf{v},\mathsf{v}) = |\!|\!|\mathsf{v}|\!|\!|_\Omega^2$ for all $\mathsf{v}\in H_0^1(\Omega)$, we prove the following upper bound for the error.

Theorem 1. *For* u *and* $\mathsf{u}_{\mathscr{T}}$*, solutions to* (4) *and* (5)*, respectively, the following fully computable error bound holds*

$$|\!|\!|\mathsf{u}-\mathsf{u}_{\mathscr{T}}|\!|\!|_\Omega \le \eta := \left(\sum_{K\in\mathscr{T}} \eta_K^2 \right)^{\frac{1}{2}} . \tag{17}$$

Here, the error indicators on each $K\in\mathscr{T}$ *are given by*

$$\eta_K := \frac{1}{\sqrt{\varepsilon}}\|\boldsymbol{\sigma}_K\|_{0,K} + \mathsf{C}_{osc}\|\mathrm{osc}_K\|_{0,K}. \tag{18}$$

4.1. An explicit formula for $\boldsymbol{\sigma}_K$. In this section we detail the solution to (14) that will be used in (18). Its construction improves the results from [1, 3] slightly. We recall that, for $n\in\mathcal{V}$, λ_n is such that $\lambda_n \in \mathbb{W}(\mathscr{T})$ and $\lambda_n(\boldsymbol{x}_m) = \delta_{nm}$ for all $m\in\mathcal{V}$. Let $K\in\mathscr{T}$ and let $\gamma_{K,i}$ be the edge(2D)/face(3D) of element K which is opposite to vertex $\boldsymbol{x}_i$ for $i \in \mathcal{V}_K$. Also, let the tangent vector $\boldsymbol{t}_{ij} = \boldsymbol{x}_j - \boldsymbol{x}_i$ for $i, j \in \mathcal{V}_K$ and let $\overline{\boldsymbol{x}}_K = \dfrac{1}{d+1} \sum_{i \in \mathcal{V}_K} \boldsymbol{x}_i$. Let us consider the function

$$\boldsymbol{\sigma}_{K,0} = \sum_{i \in \mathcal{V}_K} \left(\sum_{j \in \mathcal{V}_{\gamma_{K,i}}} \left(\mathscr{R}_{\gamma_{K,i},K}, \lambda_j\right)_{\gamma_{K,i}} \boldsymbol{\psi}_{i,j} + \left(|K| \nabla(\mathscr{R}_K) \cdot (\boldsymbol{x}_i - \overline{\boldsymbol{x}}_K)\right) \boldsymbol{\psi}_{K,i} \right).$$

Here, in the 2D case

$$\boldsymbol{\psi}_{i,j} = \frac{1}{2|K|} \left((3\lambda_j(\lambda_k - \lambda_i) + 4\lambda_j)\boldsymbol{t}_{ij} + (3\lambda_k(\lambda_i - \lambda_j) - 2\lambda_k)\boldsymbol{t}_{ik} \right), \tag{19}$$

$$\boldsymbol{\psi}_{K,i} = -\frac{\lambda_i}{3|K|}(\lambda_j \boldsymbol{t}_{ij} + \lambda_k \boldsymbol{t}_{ik}), \tag{20}$$

for distinct $i, j, k \in \mathcal{V}_K$, and, in 3D,

$$\begin{aligned} \boldsymbol{\psi}_{i,j} = & \frac{1}{4|K|} \Big((12\lambda_j + 19\lambda_k + 19\lambda_l - 2\lambda_i)\lambda_j \boldsymbol{t}_{ij} \\ & + (3\lambda_i - 4\lambda_k - 4\lambda_l - 11\lambda_j)\lambda_k \boldsymbol{t}_{ik} + (3\lambda_i - 4\lambda_k - 4\lambda_l - 11\lambda_j)\lambda_l \boldsymbol{t}_{il} \Big), \end{aligned} \tag{21}$$

$$\boldsymbol{\psi}_{K,i} = -\frac{1}{4|K|} \lambda_i \Big(\lambda_j \boldsymbol{t}_{ij} + \lambda_k \boldsymbol{t}_{ik} + \lambda_l \boldsymbol{t}_{il} \Big), \tag{22}$$

for distinct $i, j, k, l \in \mathcal{V}_K$. The function $\boldsymbol{\sigma}_{K,0}$ satisfies $-\mathbf{div}\ \boldsymbol{\sigma}_{K,0} = \mathscr{R}_K$ in K and $\boldsymbol{\sigma}_{K,0} \cdot \boldsymbol{n}_\gamma^K = \mathscr{R}_{\gamma,K}$ on γ for all $\gamma \in \mathcal{F}_K$ (see [3, Theorem 6.3]).

With the aim of sharpening the bound, we define

$$\boldsymbol{\psi}_{0,l} = \frac{1}{4\,|K|} \left(\lambda_i \lambda_j \boldsymbol{t}_{ij} + \lambda_k \lambda_i \boldsymbol{t}_{ki} + \lambda_j \lambda_k \boldsymbol{t}_{jk} \right),$$

for distinct $i, j, k \in \mathcal{V}_K$ and $l = 1$ when $d = 2$ and for distinct $i, j, k \in \mathcal{V}_\gamma$ for three distinct $\gamma \in \mathcal{F}_K$ which each correspond to a distinct $l = 1, 2, 3$ when $d = 3$. The functions $\boldsymbol{\psi}_{0,l}$ satisfy $-\mathbf{div}\ \boldsymbol{\psi}_{0,l} = 0$ in K and $\boldsymbol{\psi}_{0,l} \cdot \boldsymbol{n}_\gamma^K = 0$ on γ for all $\gamma \in \mathcal{F}_K$.

Then, when $d = 2$, the best possible solution in $\mathbb{P}_2(K)^2$ to (14) is

$$\boldsymbol{\sigma}_K = \boldsymbol{\sigma}_{K,0} - \frac{\left(\boldsymbol{\sigma}_{K,0}, \boldsymbol{\psi}_{0,1}\right)_K}{\left(\boldsymbol{\psi}_{0,1}, \boldsymbol{\psi}_{0,1}\right)_K} \boldsymbol{\psi}_{0,1}. \tag{23}$$

When $d = 3$, the best possible solution in $\mathbb{P}_2(K)^3$ to (14) is

$$\boldsymbol{\sigma}_K = \boldsymbol{\sigma}_{K,0} - \sum_{l=1}^{3} \alpha_l \boldsymbol{\psi}_{0,l}, \tag{24}$$

where $[\alpha_1\ \alpha_2\ \alpha_3]^T = \boldsymbol{G}^{-1}\boldsymbol{g}$, with the entries of the matrix $\boldsymbol{G}$ and vector $\boldsymbol{g}$ being given by $(\boldsymbol{G})_{ij} = \left(\boldsymbol{\psi}_{0,i}, \boldsymbol{\psi}_{0,j}\right)_K$ and $(\boldsymbol{g})_i = \left(\boldsymbol{\sigma}_{K,0}, \boldsymbol{\psi}_{0,i}\right)_K$.

Finally, it is important to mention that, following essentially the same arguments as in [1, 3], the $\boldsymbol{\sigma}_K$ given by (23) and (24) satisfy

$$\|\boldsymbol{\sigma}_K\|_{0,K} \leq C \left(h_K^{\frac{1}{2}} \sum_{\gamma \in \mathcal{F}_K} \|\mathscr{R}_{\gamma,K}\|_{0,\gamma} + h_K \|\mathscr{R}_K\|_{0,K} \right), \tag{25}$$

where $C > 0$ is a constant independent of the size of the elements in the mesh.

4.2. Local efficiency of the estimator. Our starting point is (18), which after applying (25) leads to

$$\eta_K \le C\left(\frac{h_K}{\sqrt{\varepsilon}}\|\mathscr{R}_K\|_{0,K} + \sum_{\gamma\in\mathcal{F}_K}\left(\frac{h_K}{\varepsilon}\right)^{\frac{1}{2}}\|\mathscr{R}_{\gamma,K}\|_{0,\gamma}\right) + \mathsf{C}_{osc}\|\mathrm{osc}_K\|_{0,K}.$$

We start with the decomposition $J_{\gamma,K} := \varepsilon\nabla\mathsf{u}_{\mathscr{T}|K}\cdot\boldsymbol{n}_\gamma^K = [\![J_\gamma]\!] + \langle J_{\gamma,K}\rangle$, where

$$[\![J_\gamma]\!] := \begin{cases} \frac{1}{2}(J_{\gamma,K}+J_{\gamma,K'}) & \text{if } \gamma\in\mathcal{F}_K\cap\mathcal{F}_{K'},\ K,K'\in\mathscr{T},\ K\neq K', \\ 0 & \text{if } \gamma\in\mathcal{F}_K\cap\mathcal{F}_{\partial\Omega},\ K\in\mathscr{T}, \end{cases}$$

and

$$\langle J_{\gamma,K}\rangle := \begin{cases} \frac{1}{2}(J_{\gamma,K}-J_{\gamma,K'}) & \text{if } \gamma\in\mathcal{F}_K\cap\mathcal{F}_{K'},\ K,K'\in\mathscr{T},\ K\neq K', \\ J_{\gamma,K} & \text{if } \gamma\in\mathcal{F}_K\cap\mathcal{F}_{\partial\Omega},\ K\in\mathscr{T}. \end{cases}$$

Hence, $\mathscr{R}_{\gamma,K} = g_{\gamma,K} - \langle J_{\gamma,K}\rangle - [\![J_\gamma]\!]$, which gives

$$\begin{aligned}\eta_K \le{}& C\Bigg(\frac{h_K}{\sqrt{\varepsilon}}\|\mathscr{R}_K\|_{0,K} \\ &+ \sum_{\gamma\in\mathcal{F}_K}\left(\frac{h_K}{\varepsilon}\right)^{\frac{1}{2}}\left(\|[\![J_\gamma]\!]\|_{0,\gamma} + \|g_{\gamma,K}-\langle J_{\gamma,K}\rangle\|_{0,\gamma}\right)\Bigg) + \mathsf{C}_{osc}\|\mathrm{osc}_K\|_{0,K}. \end{aligned} \tag{26}$$

In order to show that the error indicator η_K is locally efficient, we must bound the right hand side of the above inequality. We start by noticing that the error equation can be written as

$$\sum_{K\in\mathscr{T}}\left((\mathscr{R}_K,\mathsf{v})_K - \sum_{\gamma\in\mathcal{F}_K}([\![J_\gamma]\!],\mathsf{v})_\gamma\right) = \mathcal{B}(\mathsf{u}-\mathsf{u}_{\mathscr{T}},\mathsf{v}) - \sum_{K\in\mathscr{T}}(\mathrm{osc}_K,\mathsf{v})_K.$$

By applying standard bubble function arguments to the previous error equation (see [36, 2]), it can be proved that, for $K\in\mathscr{T}$ and $\gamma\in\mathcal{F}_K$,

$$\frac{h_K}{\sqrt{\varepsilon}}\|\mathscr{R}_K\|_{0,K} \le C\left(\mathsf{C}_K|\!|\!|\mathsf{u}-\mathsf{u}_{\mathscr{T}}|\!|\!|_K + \frac{h_K}{\sqrt{\varepsilon}}\|\mathrm{osc}_K\|_{0,K}\right), \tag{27}$$

$$\left(\frac{h_K}{\varepsilon}\right)^{\frac{1}{2}}\|[\![J_\gamma]\!]\|_{0,\gamma} \le C\sum_{K'\in\omega_\gamma}\left(\mathsf{C}_{K'}|\!|\!|\mathsf{u}-\mathsf{u}_{\mathscr{T}}|\!|\!|_{K'} + \frac{h_{K'}}{\sqrt{\varepsilon}}\|\mathrm{osc}_{K'}\|_{0,K'}\right), \tag{28}$$

where

$$\mathsf{C}_K := \max\left\{1, \frac{\|\boldsymbol{b}\|_{\infty,K}h_K}{\varepsilon}, \sqrt{\kappa}\frac{h_K}{\sqrt{\varepsilon}}\right\}. \tag{29}$$

Moreover, following similar steps as in [1, 3], the remaining term can be bounded as

$$\begin{aligned}\left(\frac{h_K}{\varepsilon}\right)^{\frac{1}{2}}\|g_{\gamma,K}-\langle J_{\gamma,K}\rangle\|_{0,\gamma} \le C\sum_{n\in\mathcal{V}_\gamma}\sum_{K'\in\omega_n}\Bigg(\frac{h_{K'}}{\sqrt{\varepsilon}}\|\mathscr{R}_{K'}\|_{0,K'} \\ + \sum_{\gamma'\in\mathcal{F}_{K'}\cap\mathcal{F}_n}\left(\frac{h_{K'}}{\varepsilon}\right)^{\frac{1}{2}}\|[\![J_{\gamma'}]\!]\|_{0,\gamma'} + \left(\frac{h_{K'}^{2-d}}{\varepsilon}\right)^{\frac{1}{2}}|\mathcal{S}_{K'}(\mathsf{u}_{\mathscr{T}};\mathsf{u}_{\mathscr{T}},\lambda_n)|\Bigg),\end{aligned}$$

which, together with (26), (27) and (28), gives

$$\begin{aligned}\eta_K \le C \sum_{n\in\mathcal{V}_K}\sum_{K'\in\omega_n}\Bigg(&\mathsf{C}_{K'}\vert\!\vert\!\vert\mathsf{u}-\mathsf{u}_{\mathscr{T}}\vert\!\vert\!\vert_{K'} + \frac{h_{K'}}{\sqrt{\varepsilon}}\|\mathrm{osc}_{K'}\|_{0,K'} \\ &+ \left(\frac{h_{K'}^{2-d}}{\varepsilon}\right)^{\frac{1}{2}}|\mathcal{S}_{K'}(\mathsf{u}_{\mathscr{T}};\mathsf{u}_{\mathscr{T}},\lambda_n)|\Bigg).\end{aligned} \tag{30}$$

Thus far, the nonlinear character of the stabilization term has not affected the derivations and results. We now proceed to bound this nonlinear term. In order to do this, we define, for every $K\in\mathscr{T}$, $\tilde{\mathsf{u}}_K$ as the only solution, in $\mathbb{P}_1(\omega_K)$, of the following problem:

$$\left\{\begin{array}{rcl}(\nabla\tilde{\mathsf{u}}_K,\nabla\psi)_{\omega_K} & = & (\nabla\mathsf{u},\nabla\psi)_{\omega_K} \quad \text{for all } \psi\in\mathbb{P}_1(\omega_K),\\ (\tilde{\mathsf{u}}_K,1)_{\omega_K} & = & 0.\end{array}\right. \tag{31}$$

Using this projection, along with Assumptions 1 and 2, and the fact that $|\lambda_n|_{1,\omega_K}\le Ch_K^{\frac{d}{2}-1}$, leads to the bound

$$\begin{aligned}\left(\frac{h_{K'}^{2-d}}{\varepsilon}\right)^{\frac{1}{2}}|\mathcal{S}_{K'}(\mathsf{u}_{\mathscr{T}};\mathsf{u}_{\mathscr{T}},\lambda_n)| &= \left(\frac{h_{K'}^{2-d}}{\varepsilon}\right)^{\frac{1}{2}}|\mathcal{S}_{K'}(\mathsf{u}_{\mathscr{T}};\mathsf{u}_{\mathscr{T}},\lambda_n)-\mathcal{S}_{K'}(\tilde{\mathsf{u}}_{K'};\tilde{\mathsf{u}}_{K'},\lambda_n)| \\ &\le \left(\frac{h_{K'}^{2-d}}{\varepsilon}\right)^{\frac{1}{2}}\mathsf{C}_{lip}h_{K'}|\mathsf{u}_{\mathscr{T}}-\tilde{\mathsf{u}}_{K'}|_{1,\omega_{K'}}|\lambda_n|_{1,\omega_{K'}} \\ &\le C\left(\frac{h_{K'}}{\varepsilon}\vert\!\vert\!\vert\mathsf{u}-\mathsf{u}_{\mathscr{T}}\vert\!\vert\!\vert_{\omega_{K'}}+\frac{h_{K'}}{\sqrt{\varepsilon}}|\mathsf{u}-\tilde{\mathsf{u}}_{K'}|_{1,\omega_{K'}}\right).\end{aligned} \tag{32}$$

Then, gathering (30) and (32), we arrive at the following local efficiency result for η_K.

THEOREM 2. *There exists $C>0$, independent of the size of the elements in the mesh $\mathscr{T}$, such that, for every $K\in\mathscr{T}$, the following local lower bound holds*

$$\begin{aligned}\eta_K \le\ & C\left(\max_{K'\in\omega_K}\mathsf{C}_{K'}+\frac{h_K}{\varepsilon}\right)\vert\!\vert\!\vert\mathsf{u}-\mathsf{u}_{\mathscr{T}}\vert\!\vert\!\vert_{\hat{\omega}_K} \\ &+C\sum_{K'\in\omega_K}\frac{h_{K'}}{\sqrt{\varepsilon}}\left(\|\mathrm{osc}_{K'}\|_{0,K'}+|\mathsf{u}-\tilde{\mathsf{u}}_{K'}|_{1,\omega_{K'}}\right),\end{aligned}$$

where $\mathsf{C}_{K'}$ is defined by (29), and

$$\hat{\omega}_K := \bigcup_{K'\in\omega_K}\omega_{K'}.$$

Remark 3. The term $h_{K'}|\mathsf{u}-\tilde{\mathsf{u}}_{K'}|_{1,\omega_{K'}}$ in the last result is unusual. Now, it is important to remark that, since the exact solution of the continuous problem belongs to $H^{1+\delta}(\Omega)$, with $\delta\ge\frac{1}{2}$ (see, e.g., [22]), this term is an oscillation.

Remark 4. The estimator is not robust with respect to ε. However, this is the usual case for a posteriori error estimators for the error measured in the norm $\vert\!\vert\!\vert\cdot\vert\!\vert\!\vert$. In [38, 35] residual based a posteriori estimators for the error in linear discretizations were proved to be robust with respect to a norm that includes a dual norm of the convective term. However, the application of the techniques from [38, 35] to nonlinear discretizations such as the ones considered in this paper does not seem to be feasible. Resolving this issue will be the subject of future research.

Remark 5. We end this section by noticing that if one of the conditions we have imposed on the stabilization term is not satisfied, then, at least a weaker result can be obtained. More precisely, if we suppose that the linearity preservation from Assumption 2 is not valid, but the Lipschitz continuity from Assumption 1 is, then, supposing that $\mathcal{S}_K(0;0,\mathsf{z}_{\mathscr{T}})=0$ for all $\mathsf{z}_{\mathscr{T}}\in\mathbb{W}(\mathscr{T})$ (which is always the case), and proceeding as in (32), we arrive at

$$\left(\frac{h_{K'}^{2-d}}{\varepsilon}\right)^{\frac{1}{2}}|\mathcal{S}_{K'}(\mathsf{u}_{\mathscr{T}};\mathsf{u}_{\mathscr{T}},\lambda_n)|\le C\left(\frac{h_{K'}}{\varepsilon}\|\!|\mathsf{u}-\mathsf{u}_{\mathscr{T}}\|\!|_{\omega_{K'}}+\frac{h_{K'}}{\sqrt{\varepsilon}}|\mathsf{u}|_{1,\omega_{K'}}\right).$$

The last term in the above estimate is not an oscillation, but is a term that decays with an optimal rate. Also, considering that for most of the methods of a shock-capturing type the a priori error estimates usually give an $O(\sqrt{h})$ convergence rate, then this last term may be expected to behave like an oscillation.

5. The nonlinear edge diffusion schemes. In this section we present discretizations satisfying the assumptions from Section 3. We have chosen to focus on algebraic flux correction schemes (see, e.g., [28]) as they have been the subject of extensive studies in the last few years. Our presentation will follow the lines of the rewriting of these schemes given recently in [8]. More precisely, the local contribution is given by

$$\mathcal{S}_K(\mathsf{z}_{\mathscr{T}};\mathsf{w}_{\mathscr{T}},\mathsf{v}_{\mathscr{T}}):=\sum_{E\in\mathcal{E}_K\cap\mathcal{E}_I}\mathsf{C}_E^{-1}\tau_E(\mathsf{z}_{\mathscr{T}})\,(\nabla\mathsf{w}_{\mathscr{T}}\cdot\boldsymbol{t}_E,\nabla\mathsf{v}_{\mathscr{T}}\cdot\boldsymbol{t}_E)_E\,, \tag{33}$$

where $\mathsf{C}_E:=\#\{K\in\mathscr{T}:\ K\cap E=E\}$. Using this stabilization term, we will use two different definitions of the parameter τ_E:

Method F-BBK: Our first choice is the definition given in [8, Remark 1]. For $\mathsf{v}\in\mathbb{W}(\mathscr{T})$, we define

$$\xi_{\mathsf{v}}(\boldsymbol{x}_i):=\begin{cases}\dfrac{\left|\displaystyle\sum_{j\in\mathcal{V}_i}\beta_{ij}(\mathsf{v}(\boldsymbol{x}_i)-\mathsf{v}(\boldsymbol{x}_j))\right|}{\displaystyle\sum_{j\in\mathcal{V}_i}\beta_{ij}|\mathsf{v}(\boldsymbol{x}_i)-\mathsf{v}(\boldsymbol{x}_j)|} & \text{if } \displaystyle\sum_{j\in\mathcal{V}_i}\beta_{ij}|\mathsf{v}(\boldsymbol{x}_i)-\mathsf{v}(\boldsymbol{x}_j)|\neq 0,\\ 0 & \text{otherwise,}\end{cases}$$

for every interior vertex $\boldsymbol{x}_i$, and $\xi_{\mathsf{v}}(\boldsymbol{x}_i):=0$ for all $\boldsymbol{x}_i\in\partial\Omega$ (the value of β_{ij} will be defined below). Then, we define $\tau_E(\mathsf{v})$ as

$$\tau_E(\mathsf{v}):=\mathsf{C}_0|E|^d\max_{i\in\mathcal{V}_E}[\xi_{\mathsf{v}}(\boldsymbol{x}_i)]^p\quad\text{with}\quad p\in[1,+\infty), \tag{34}$$

where $\mathsf{C}_0>0$ needs to be large enough to ensure the validity of the discrete maximum principle (see [8, Theorem 2] for details). Concerning the choice of the coefficients β_{ij}, we use the definition from [8, Remark 1]. These coefficients are chosen in such a way that $\tau_E(\mathsf{v})=0$ if $\mathsf{v}\in\mathbb{P}_1(\omega_E)$, thus implying the linearity preservation (9). This requirement can be written in an equivalent way as follows: recalling that $\boldsymbol{t}_{ij}=\boldsymbol{x}_j-\boldsymbol{x}_i$, for all $\mathsf{v} \in \mathbb{P}_1(\omega_i)$,

$$\sum_{j\in\mathcal{V}_i} \beta_{ij}\big(\mathsf{v}(x_j) - \mathsf{v}(x_i)\big) = \sum_{j\in\mathcal{V}_i} \beta_{ij} \nabla\mathsf{v}\cdot \boldsymbol{t}_{ij} = \nabla\mathsf{v}\cdot\left(\sum_{j\in\mathcal{V}_i} \beta_{ij}\boldsymbol{t}_{ij}\right) = 0,$$

which reduces to imposing

$$\sum_{j\in\mathcal{V}_i} \beta_{ij}\boldsymbol{t}_{ij} = \mathbf{0}\,. \tag{35}$$

The equation (35) is a first restriction that the β_{ij} have to satisfy. Another natural restriction is

$$\beta_{ij} \geq \beta_0 > 0, \tag{36}$$

where the value of β_0 is of no great importance. Finally, in the case when the mesh is symmetric with respect to its interior nodes (see [8] for a precise definition), $\beta_{ij} = 1$, for all i, j, should be an acceptable (and preferred) choice. Then, we find β_{ij} as the solution of the following problem: For all internal node $\boldsymbol{x}_i$, find

$$\big(\beta_{ij}\big)_{j\in\mathcal{V}_i} = \operatorname{argmin}\left\{\sum_{j\in\mathcal{V}_i} |\delta_{ij}-1|^2 : \{\delta_{ij}\} \text{ satisfies the restrictions (35), (36)}\right\}. \tag{37}$$

This method satisfies all the assumptions required in this work (see [8, Remark 1] for details).

Finally, in order to avoid the computation of the coefficients β_{ij}, we have set them as 1, leading to the *simplified* BBK method (referred to as S-BBK). This method, also presented in [8], also satisfies Assumption 1, but not the linearity preservation Assumption 2 if the mesh is not symmetric with respect to its interior nodes. Then, the results from Remark 5 are applicable to it.

If the mesh $\mathscr{T}$ satisfies the hypothesis of Xu and Zikatanov (see [39] and [8, Assumption 1]) then the following result holds for these methods (see [8] for the proof).

Theorem 6 (The Discrete Maximum Principle). *Let us suppose that $\kappa = 0$. Then, if $\mathsf{f} \geq 0$ (≤ 0) then $\mathsf{u}_{\mathscr{T}}$ attains its minimum (maximum) on the boundary of Ω. If $\kappa > 0$, then if both f and u_D are greater than (less than), or equal to, zero, then $\mathsf{u}_{\mathscr{T}} \geq 0$ (≤ 0) in Ω.*

Method BJK: This is a recent alternative introduced in [9] with the aim of satisfying the discrete maximum principle and the linearity preservation on general meshes. The Lipschitz continuity from Assumption 1 can be proved using similar arguments as in [8, Lemma 2] under the assumption that $\mathscr{T}$ satisfies the hypothesis of Xu and Zikatanov. First, we denote by $\mathbf{A}$ the finite element matrix with entries

$$(\mathbf{A})_{ij} = \mathbf{A}_{ij} := \mathcal{B}(\lambda_j, \lambda_i) \quad \text{for } i, j \in \mathcal{V},$$

which is the Galerkin matrix where the boundary conditions have not been included. Let $\mathsf{v} \in \mathbb{W}(\mathscr{T})$. Then, for $i \in \mathcal{V}_\Omega$ we define

$$\mathsf{d}_{i,j} := -\max\{0, \mathbf{A}_{ij}, \mathbf{A}_{ji}\} \quad \text{and} \quad \mathsf{f}_{i,j} := \mathsf{d}_{i,j}(\mathsf{v}(\boldsymbol{x}_j) - \mathsf{v}(\boldsymbol{x}_i)),$$

for $j \in \mathcal{V}_i$. With the previous notation, we also define

$$\mathsf{P}_i^+ := \sum_{\substack{j\in\mathcal{V}_i:\\ \mathsf{f}_{i,j}>0}} \mathsf{f}_{i,j} \quad \text{and} \quad \mathsf{P}_i^- := \sum_{\substack{j\in\mathcal{V}_i:\\ \mathsf{f}_{i,j}<0}} \mathsf{f}_{i,j}.$$

Also, recalling that $\gamma_{K,i}$ is the edge(2D)/face(3D) of element K that is opposite to the vertex $\boldsymbol{x}_i$, we define

$$\Psi_i := \max_{K\in\omega_i} \frac{|\gamma_{K,i}|}{d|K|}, \quad \mathsf{g}_i := \Psi_i \max_{j\in\mathcal{V}_i}\{|\boldsymbol{x}_i - \boldsymbol{x}_j|\} \quad \text{and} \quad \mathsf{q}_i := \mathsf{g}_i \sum_{j\in\mathcal{V}_i} \mathsf{d}_{i,j}.$$

We also take

$$\mathsf{Q}_i^+ := \mathsf{q}_i(\mathsf{v}(\boldsymbol{x}_i) - \mathsf{v}_i^{max}) \quad \text{and} \quad \mathsf{Q}_i^- := \mathsf{q}_i(\mathsf{v}(\boldsymbol{x}_i) - \mathsf{v}_i^{min}),$$

where

$$\mathsf{v}_i^{max} := \max_{j\in\mathcal{V}_i\cup\{i\}}\{\mathsf{v}(\boldsymbol{x}_j)\} \quad \text{and} \quad \mathsf{v}_i^{min} := \min_{j\in\mathcal{V}_i\cup\{i\}}\{\mathsf{v}(\boldsymbol{x}_j)\}.$$

These definitions allow us to define

$$\mathsf{R}_i^+ := \begin{cases} \min\left\{1, \frac{\mathsf{Q}_i^+}{\mathsf{P}_i^+}\right\} & \text{if } \mathsf{P}_i^+ \neq 0, \\ 1 & \text{otherwise,} \end{cases} \quad \text{and} \quad \mathsf{R}_i^- := \begin{cases} \min\left\{1, \frac{\mathsf{Q}_i^-}{\mathsf{P}_i^-}\right\} & \text{if } \mathsf{P}_i^- \neq 0, \\ 1 & \text{otherwise,} \end{cases}$$

and in turn

$$\alpha_{i,j} := \begin{cases} \mathsf{R}_i^+ & \text{if } \mathsf{f}_{i,j} > 0, \\ \mathsf{R}_i^- & \text{if } \mathsf{f}_{i,j} < 0, \\ 1 & \text{otherwise.} \end{cases}$$

With all the previous notation at hand, we consider the following stabilization parameter

$$\tau_E(\mathsf{v}) := |E| \max_{\substack{i\in\mathcal{V}_\Omega\cap\mathcal{V}_E,\\ j\in\mathcal{V}_E,\\ j\neq i}} (1-\alpha_{i,j})\,|\mathsf{d}_{i,j}|. \tag{38}$$

Theorem 6 holds for this method without the need for the mesh to satisfy the hypothesis of Xu and Zikatanov. However, as previously mentioned, it is under this assumption that the Lipschitz continuity from Assumption 1 can be proved.

Remark 7. We finish this section by making some statements about the order of convergence of the methods more precise. As it was mentioned before, for most of the methods of the type we consider in this work, the error estimate that can be proved is an estimate of the type

$$|\!|\!| \mathsf{u} - \mathsf{u}_{\mathscr{T}} |\!|\!|_\Omega \leq C\sqrt{h}\,|\mathsf{u}|_{2,\Omega}\,,$$

where $C > 0$ does not depend on h, or ε. This estimate is valid, in particular, for Method S-BBK (see [8]). This explains why we stated in Remark 5 that the term $h|\mathsf{u}|_{1,\hat{\omega}_K}$ may be seen as a term that, in general, decays faster than the error. Now, under Assumptions 1 and 2, in [8] the following estimate has been proven

$$|\!|\!|\mathsf{u} - \mathsf{u}_{\mathscr{T}}|\!|\!|_{\Omega} \le C\, h \left(|\mathsf{u}|_{2,\Omega} + \frac{1}{\sqrt{\varepsilon}} \|\mathsf{u}\|_{1,\Omega} \right),$$

where $C > 0$ does not depend on h or ε. This estimate is valid for both Methods F-BBK and BJK.

6. Numerical examples. In this section we report the results of a series of numerical examples in two and three space dimensions. All matrices have been assembled exactly. The right hand sides and approximation errors were computed using quadrature formulas which are exact for polynomials of degree 19 on triangles and degree 14 on tetrahedrons. The only non-standard procedure used in the calculation of integrals was that, for some examples, to avoid a big underestimation of the error, we computed the error using these quadrature rules in a mesh which is a refinement of the particular mesh we were using at the time. All linear systems were solved using the multifrontal massively parallel sparse direct solver (MUMPS) [4, 5].

The results were obtained using Algorithm 1, with the exception of some of the results for the first example which were obtained by uniformly refining the mesh. The damped fixed-point algorithm used to solve each nonlinear problem was introduced in [24, Figure 12]. This algorithm has a dynamic damping parameter which has been shown to yield better numerical performance than a predetermined choice. For this algorithm, we defined the residual vector $\mathbf{R}(\mathsf{v})$ by

$$(\mathbf{R}(\mathsf{v}))_l := \begin{cases} \mathcal{B}(\mathsf{v}, \lambda_l) + \mathcal{S}(\mathsf{v}; \mathsf{v}, \lambda_l) - (\mathsf{f}, \lambda_l)_{\Omega} & \text{if } l \in \mathcal{V}_{\Omega}, \\ 0 & \text{if } l \in \mathcal{V} \setminus \mathcal{V}_{\Omega}, \end{cases} \tag{39}$$

and stopped the damped fixed-point algorithm when $\|\mathbf{R}(\mathsf{u}^i_{\mathscr{T}})\|_{\mathsf{Euc}}$ was less than a tolerance which we took to be 10^{-8}. Here, $\|\cdot\|_{\mathsf{Euc}}$ denotes the usual Euclidean norm and $\mathsf{u}^i_{\mathscr{T}}$ is the ith damped iterative solution. We took the initial guess $\mathsf{u}^0_{\mathscr{T}_0} \in \mathbb{W}(\mathscr{T}_0)$ to be such that $\mathsf{u}^0_{\mathscr{T}_0} = \mathsf{u}_D$ on $\partial\Omega$ and $\mathcal{B}(\mathsf{u}^0_{\mathscr{T}_0}, \mathsf{v}_{\mathscr{T}_0}) = (\mathsf{f}, \mathsf{v}_{\mathscr{T}_0})_{\Omega}$ for all $\mathsf{v}_{\mathscr{T}_0} \in \mathbb{V}(\mathscr{T}_0)$. There are other possibilities for starting the iterative process, such as the solution of a SUPG method. Nevertheless, since this is only used for starting the algorithm on the initial mesh, the differences in performance are negligible. For computational efficiency, elements were marked for refinement using an average strategy. That is, an element $K \in \mathscr{T}$ is marked if $\eta_K^2 \ge \eta^2/\#\mathscr{T}$. Nevertheless, Algorithm 1 could also be implemented with another strategy such as a maximum strategy or a bulk criterion.

Finally, in all the numerical examples presented below the domain is $\Omega = (0,1)^d$, $d = 2, 3$, and, as it is usual, when the exact error is available, we measure the performance of the estimator by computing the effectivity index given by

$$\Upsilon := \frac{\eta}{|\!|\!|\mathsf{u} - \mathsf{u}_{\mathscr{T}}|\!|\!|_{\Omega}}.$$

As mentioned in Remark 4, the efficiency of the estimator degenerates with the local Péclet number. In all our examples, this dependency is only visible for relatively coarse meshes, for which the effectivity indices are rather high. Once the mesh is sufficiently refined, then the effectivity index reduces to a value close to 2 in 3D, and even less in 2D.

In two space dimensions, the meshes need to be Delaunay in order for the hypothesis of Xu and Zikatanov discussed in the previous section to hold. We have chosen as a mesh refinement strategy the same longest edge bisection algorithm used in MATLAB [31], which does not always produce Delaunay meshes and hence, we have modified this algorithm slightly. More precisely, the mesh was first refined using the above-mentioned longest edge refinement algorithm. The resulting mesh was then checked to see if it is a Delaunay mesh. All of the interior edges for which this condition was not satisfied were then marked for refinement. A new mesh was then created by refining the elements with at least one edge that had been marked for refinement such that only these marked edges were refined. The steps described in the previous three sentences were then repeated until a Delaunay mesh was reached. This was always the case for the two dimensional examples we considered.

Algorithm 1 Adaptive algorithm

```
Input:  An initial mesh $\mathscr{T}_0$, an initial guess $\mathsf{u}^0_{\mathscr{T}_0} \in \mathbb{W}(\mathscr{T}_0)$, a tolerence tol and data $\varepsilon$, $\mathbf{b}$, $\kappa$, $\mathsf{f}$ and $\mathsf{u}_D$;
1:      Set $j = 0$, $\mathsf{w}_{min} = 0.01$, $\mathsf{w}_{max} = 1$, $\mathsf{c}_1 = 1.001$, $\mathsf{c}_2 = 1.1$, $\mathsf{c}_3 = 1.001$ and $\mathsf{c}_4 = 0.9$;
Computation of the approximation:
2:      Set $i = 0$, $\mathsf{w}^0 = \mathsf{w}_{max}$ and compute $\|\mathbf{R}(\mathsf{u}^0_{\mathscr{T}_j})\|_{\mathsf{Euc}}$;
3:      While $\|\mathbf{R}(\mathsf{u}^i_{\mathscr{T}_j})\|_{\mathsf{Euc}} \geq$ tol do
            Compute $\tilde{\mathsf{u}}^{i+1}_{\mathscr{T}_j} \in \mathbb{W}(\mathscr{T}_j)$ such that $\tilde{\mathsf{u}}^{i+1}_{\mathscr{T}_j} = \mathsf{u}_D$ on $\partial\Omega$ and
            $\mathcal{B}(\tilde{\mathsf{u}}^{i+1}_{\mathscr{T}_j}, \mathsf{v}_{\mathscr{T}_j}) + \mathcal{S}(\mathsf{u}^i_{\mathscr{T}_j}; \tilde{\mathsf{u}}^{i+1}_{\mathscr{T}_j}, \mathsf{v}_{\mathscr{T}_j}) = (\mathsf{f}, \mathsf{v}_{\mathscr{T}_j})_\Omega$ for all $\mathsf{v}_{\mathscr{T}_j} \in \mathbb{V}(\mathscr{T}_j)$;
            Set $\mathsf{w}^{i+1} = \mathsf{w}^i$;
            Take $\mathsf{f\_d} = 1$ and $\mathsf{quit} = 0$;
            While $\mathsf{quit} = 0$
                Update to $\mathsf{u}^{i+1}_{\mathscr{T}_j} = \mathsf{u}^i_{\mathscr{T}_j} + \mathsf{w}^{i+1}(\tilde{\mathsf{u}}^{i+1}_{\mathscr{T}_j} - \mathsf{u}^i_{\mathscr{T}_j})$;
                Compute $\|\mathbf{R}(\mathsf{u}^{i+1}_{\mathscr{T}_j})\|_{\mathsf{Euc}}$;
                    If $\|\mathbf{R}(\mathsf{u}^{i+1}_{\mathscr{T}_j})\|_{\mathsf{Euc}} \leq \|\mathbf{R}(\mathsf{u}^i_{\mathscr{T}_j})\|_{\mathsf{Euc}}$ or $\mathsf{w}^{i+1} \leq \mathsf{c}_1\mathsf{w}_{min}$ then
                        If $\|\mathbf{R}(\mathsf{u}^{i+1}_{\mathscr{T}_j})\|_{\mathsf{Euc}} \leq \|\mathbf{R}(\mathsf{u}^i_{\mathscr{T}_j})\|_{\mathsf{Euc}}$ and $\mathsf{f\_d} = 1$ then
                            $\mathsf{w}_{max} = \min\{1, \mathsf{c}_3\mathsf{w}_{max}\}$ and $\mathsf{w}^{i+1} = \min\{\mathsf{w}_{\max}, \mathsf{c}_2\mathsf{w}^{i+1}\}$;
                        end If
                        $\mathsf{quit} = 1$;
                    else
                        $\mathsf{w}^{i+1} := \max\{\mathsf{w}_{min}, \mathsf{w}^{i+1}/2\}$;
                        If $\mathsf{f\_d} = 1$ then
                            $\mathsf{w}_{max} = \max\{\mathsf{w}_{min}, \mathsf{c}_4\mathsf{w}_{max}\}$ and $\mathsf{f\_d} = 0$;
                        end If
                    end If
            end While
            Set $i \leftarrow i + 1$;
        end While
Adaptive procedure:
4:      For each element $K \in \mathscr{T}_j$, compute $\eta_K$ defined by (18) with $\boldsymbol{\sigma}_K$ given by (23) or (24);
        Mark elements in the mesh for refinement to obtain $\mathscr{T}_{j+1}$;
        Set $\mathsf{u}^0_{\mathscr{T}_{j+1}} = \mathsf{Int}_{\mathscr{T}_{j+1}}\left(\mathsf{u}^{i+1}_{\mathscr{T}_j}\right)$, which is the interpolant of $\mathsf{u}^{i+1}_{\mathscr{T}_j}$ on the new mesh $\mathscr{T}_{j+1}$;
5:      Set $j \leftarrow j + 1$ and return to step 2.
```

Example 1 (A known two dimensional solution with a boundary layer). We consider $\varepsilon = 10^{-3}$, $\mathbf{b} = (2,1)^T$, $\kappa = 1$, $\mathsf{u}_D = 0$ and the right-hand side f such that the exact solution is given by

$$\mathsf{u}(x,y) = y(1-y)\left(x - \frac{e^{\frac{-(1-x)}{\varepsilon}} - e^{-\frac{1}{\varepsilon}}}{1 - e^{-\frac{1}{\varepsilon}}}\right). \tag{40}$$

The initial mesh for this case is the square Ω divided into two triangles by the straight line $y = x$. In Figure 1 we show the results that we obtained for the BJK method as well as both the full and simplified BBK methods. We can notice in Figure 1 (top) that the estimators for the full and simplified BBK methods are very close to each other, despite the fact that the S-BBK method is not linearity preserving in general meshes. This is explained by Remark 5, that states that the extra term appearing in the lower bound decays with an optimal rate. From Figure 1 we can also appreciate the advantage of performing adaptive refinement over uniform refinement. Moreover, the errors, and estimators, for all methods decrease with the optimal rate, once the mesh is sufficiently refined.

Next, in Figure 2 we can see that the refinement is being concentrated about the boundary layer and that, like the true solution, the approximate solutions are nonnegative. For the simulations presented in Figures 1 and 2 we used the parameters $\mathsf{C}_0 = 3$ and $p = 4$ for both the BBK methods. For a fixed mesh, the parameter p plays a role in the quality of the numerical solution. In fact, the larger the value of p, the steeper the boundary and inner layers get. Now, we have tested the adaptive procedure for different values of p and have observed that the estimator, and the resulting adapted meshes, are not dramatically affected by this value. To show this, in Figures 3 and 4 we depict the results obtained when implementing the F-BBK method for different values of p, both in uniformly refined and adaptive meshes. As it can be observed from those figures, the estimator presents a robust behavior with respect to p. The meshes and the numerical solutions show that the quality of the discrete solution does not change dramatically with p. Hence, the adaptive procedure, as a whole, seems to be robust with respect to the value of p.

Example 2 (A two dimensional problem with three boundary layers). We consider $\varepsilon = 10^{-3}$, $\mathbf{b} = (1, 0)$, $\kappa = 1$, $\mathsf{u}_D = 0$ and $\mathsf{f} = 1$.

The true solution to this problem is not known, but its qualitative behavior is. More precisely, the solution of this example has two parabolic boundary layers along the lines $y = 0$ and $y = 1$, and one exponential boundary layer along the line $x = 1$. We used the same initial mesh as in Example 1 and, for the BBK methods, we used the parameters $\mathsf{C}_0 = 3$ and $p = 4$. The results are shown in Figures 5 and 6 from which we can see that the estimators for all three methods decrease at the optimal rate. In addition, the mesh refinement is being concentrated about the boundary layers and the discrete maximum principle is not violated. The discrete solutions given by the three methods seem to be qualitatively similar, although the exit boundary layer seems to be sharper for the BJK method than the two BBK methods. This is reflected in the mesh refinement, which seems more restricted to the boundaries for the BJK method, while is more spread for both the BBK methods.

Example 3 (A two-dimensional example with internal and boundary layers). Here, $\varepsilon = 10^{-4}$, $\mathbf{b} = (\cos(-\pi/3), \sin(-\pi/3))^T$, $\kappa = 0$, $\mathsf{f} = 0$, and the Dirichlet datum is

$$\mathsf{u}_D = \begin{cases} 1 & \text{if } y = 1 \text{ or } x = 0 \text{ and } y \geq 0.7, \\ (y - 0.6999)/\varepsilon & \text{if } x = 0 \text{ and } 0.6999 < y < 0.7, \\ (y - 0.9999)/\varepsilon & \text{if } x = 1 \text{ and } 0.9999 < y < 1, \\ 0 & \text{elsewhere on } \partial\Omega. \end{cases} \tag{41}$$

For this example we report the results obtained using the full BBK method with $\mathsf{C}_0 = 3$ and $p = 8$, and the BJK method. The results are depicted in Figure 7, where at the top we show the performance of the estimator and can observe that it decays with an optimal rate. We then show, for the full BBK method, the initial

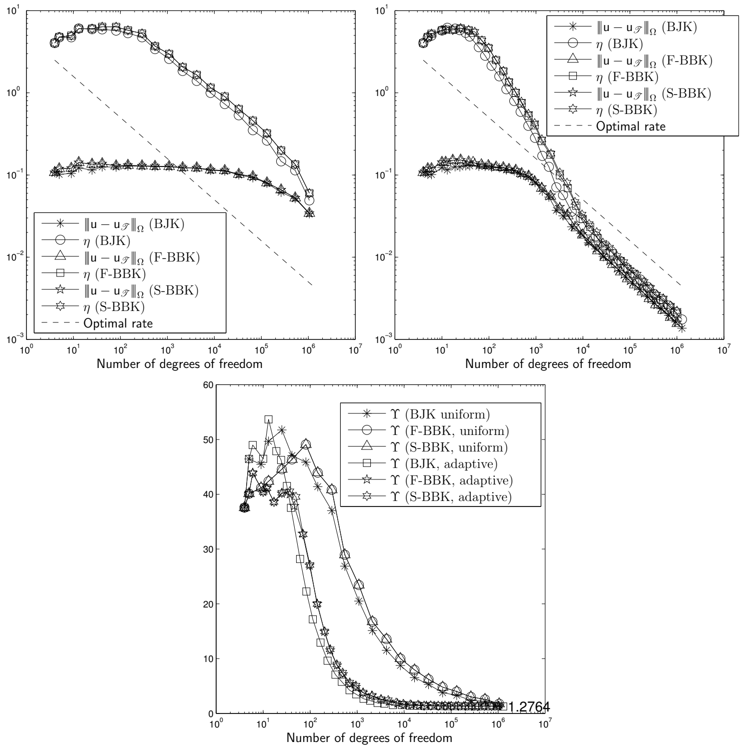

Fig. 1. *Example 1: The behaviour of the error $\|\mathsf{u}-\mathsf{u}_{\mathscr{T}}\|_\Omega$ and estimator η, for all the different methods, for uniform refinement (top left) and adaptive refinement (top right), and the effectivity indices Υ for both refinement strategies (bottom).*

mesh and approximate solution on this mesh in the center and the adapted mesh and discrete solution obtained after 15 steps of the adaptive alogorithm at the bottom. We can observe that the discrete solution always has values belonging to $[0,1]$, thus the discrete maximum principle is satisfied.

Example 4 (A two dimensional problem with a rotating convection field). The data for this example are as follows: $\varepsilon = 10^{-4}$, $\mathbf{b} = (-y,x)^T$, $\kappa = 0$, $\mathsf{f} = 0$, and

$$\mathsf{u}_D = \begin{cases} 1 & \text{if } 0.0001 \le x \le 0.4999 \text{ and } y = 0, \\ x/\varepsilon & \text{if } 0 < x < 0.0001 \text{ and } y = 0, \\ (0.5-x)/\varepsilon & \text{if } 0.4999 < x < 0.5 \text{ and } y = 0, \\ 0 & \text{elsewhere on } \partial\Omega. \end{cases} \tag{42}$$

For this example we report the results obtained using the full BBK method with $\mathsf{C}_0 = 3$ and $p = 8$, and the BJK method. Similar comments to the ones made for the previous example can be made for this case. More precisely, the results are depicted in Figure 8, following the same order as was done for the previous example, but with the meshes and approximate solutions shown being for the BJK method. We can observe a small undershoot in the numerical solutions, which is of the same order as the tolerance for the nonlinear fixed point solver, and so is not caused by a violation of the discrete maximum principle.

Example 5 (A known three dimensional solution). For this example $\varepsilon = 10^{-1}$, $\mathbf{b} = (1,1,1)^T$, $\kappa = 1$, $\mathsf{u}_D = 0$ and the right-hand side f is such that the exact solution

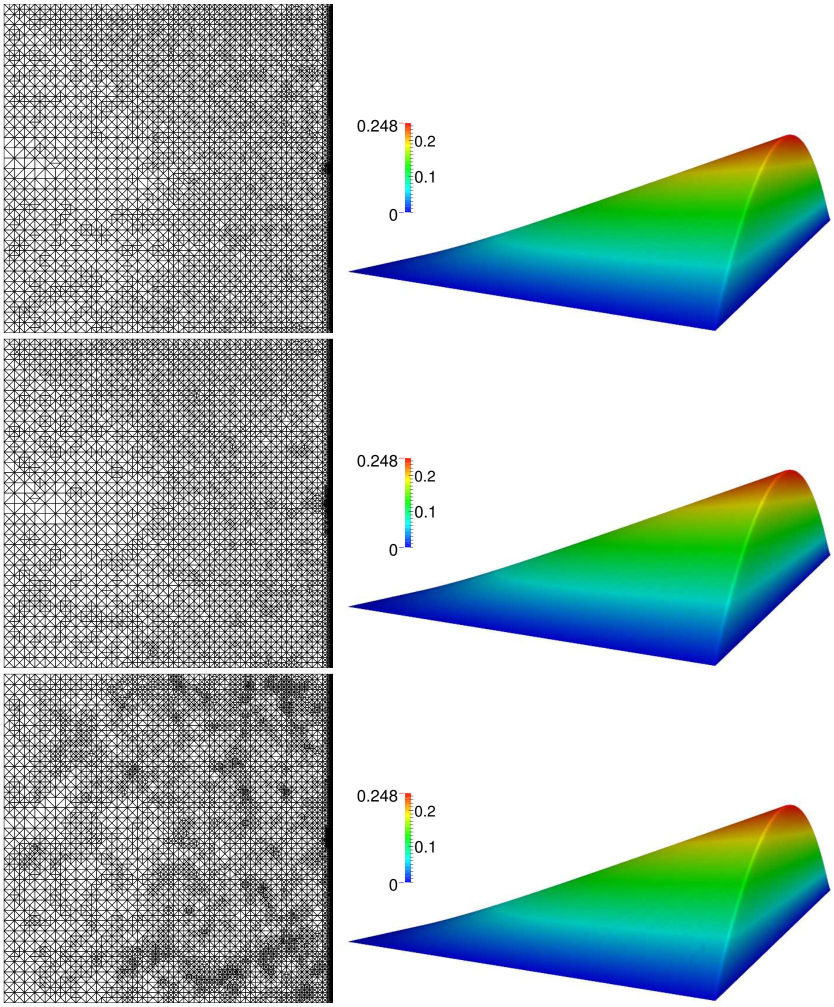

Fig. 2. *Example 1: The 29*th *adaptively refined meshes and approximations obtained on these meshes, using the BJK method (top), F-BBK method (center) and S-BBK method (bottom).*

is given by

$$\mathsf{u}(x,y,z) = xyz(1-x)(1-y)(1-z)\,. \tag{43}$$

We report the results for this problem using the F-BBK method with $\mathsf{C}_0 = 3$ and $p = 4$, and the BJK method. From Figure 9 we can see that the estimator and error follow a very similar pattern, and get closer as the mesh gets refined. This is confirmed on the right of Figure 9, where we depict the effectivity indices for this case. The initial mesh, and the approximate solution obtained after 23 steps of the adaptive scheme (which is, like the exact solution, non-negative), are depicted in Figure 10.

Example 6 (A known three dimensional solution with a boundary layer). The data for this example are as follows: $\varepsilon = 10^{-3}$, $\mathbf{b} = (1,0,0)^T$, $\kappa = 0$, $\mathsf{u}_D = 0$, and the

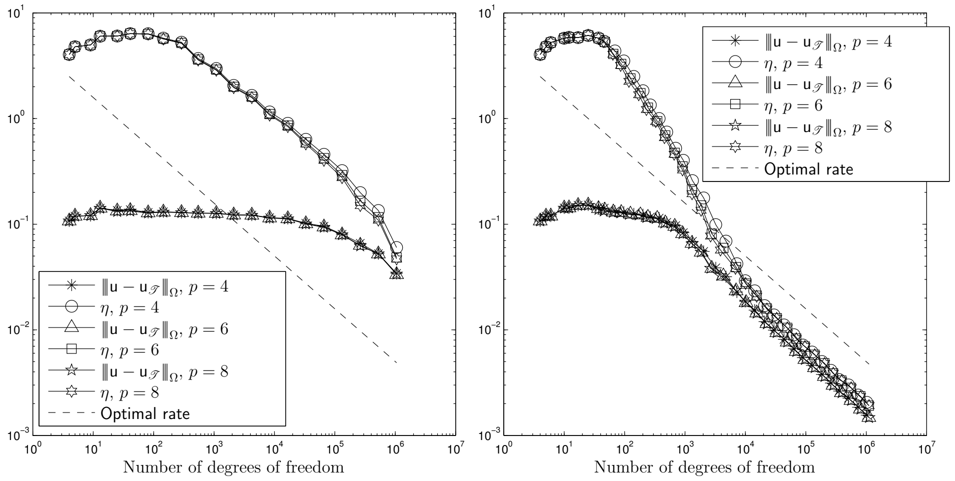

FIG. 3. *Example 1: The behaviour of the error $\|u - u_{\mathscr{T}}\|_\Omega$ and estimator η, for the F-BBK method, for uniform refinement (left) and adaptive refinement (right), for different values of p.*

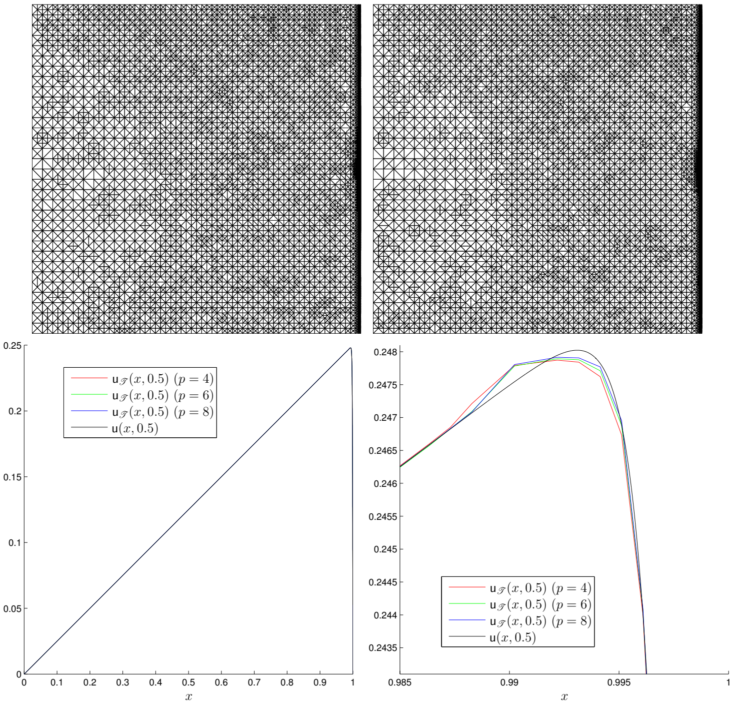

FIG. 4. *Example 1: The 29th adaptively refined meshes using the F-BBK method with $p = 6$ (top left) and $p = 8$ (top right). At the bottom, a cross section along the line $y = 0.5$ of the discrete solutions obtained using $p = 4, 6, 8$ and the exact solution of this problem. At the bottom right, a close up of the top right corner of this plot. We can see that the final discrete solutions do not differ dramatically between them.*

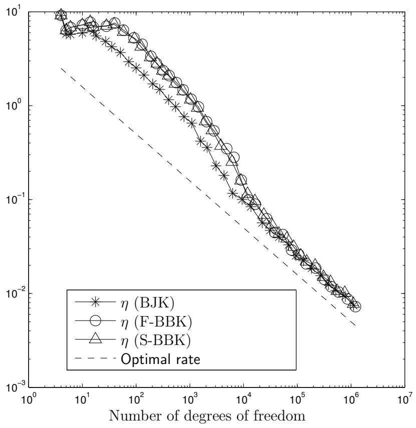

FIG. 5. *Example 2: Performance of the estimator.*

right-hand side f is such that the exact solution is given by

$$\mathsf{u}(x,y,z) = yz(1-y)(1-z)\left(x - \frac{e^{\frac{-(1-x)}{\varepsilon}} - e^{-\frac{1}{\varepsilon}}}{1-e^{-\frac{1}{\varepsilon}}}\right). \tag{44}$$

We report the results obtained using the F-BBK method with $\mathsf{C}_0 = 3$ and $p = 10$, and the BJK method, where the intial mesh is the same one as for the last example. In Figure 11 we depict the error and estimator, and effectivity indices. We can observe that the effectivity index does depend on the Péclet number, but the error and estimator get to a good agreement once the mesh is refined enough. In Figure 12 we depict the adapted meshes and the approximate solutions for different adaptive steps. We can observe that the mesh refinement is concentrated in the boundary layer region, and the solution respects the discrete maximum principle.

7. Concluding remarks. In this work we proposed and tested numerically a fully computable a posteriori error estimator for a shock-capturing like discretization of the convection-diffusion-reaction equation. The discretizations considered here are particular AFC schemes, but the presentation is general enough to accommodate any discretization that satisfies some basic hypotheses. More precisely, we have required the stabilization terms to be locally Lipschitz continuous, and locally linearity preserving. These two assumptions have been previously used to prove optimal convergence. Interestingly, they were not needed to prove the fact that the estimator is a computable upper bound for the error, but they are essential for proving the local lower bounds. Future avenues of research include the study of robustness of the estimator with respect to ε, and the search for more effective ways to solve the resulting nonlinear system.

Acknowledgments. The authors would like to thank Diego Paredes and Petr Knobloch for many interesting discussions and suggestions.

REFERENCES

[1] M. Ainsworth, A. Allendes, G. R. Barrenechea, and R. Rankin, *Fully computable a posteriori error bounds for stabilised fem approximations of convection–reaction–diffusion*

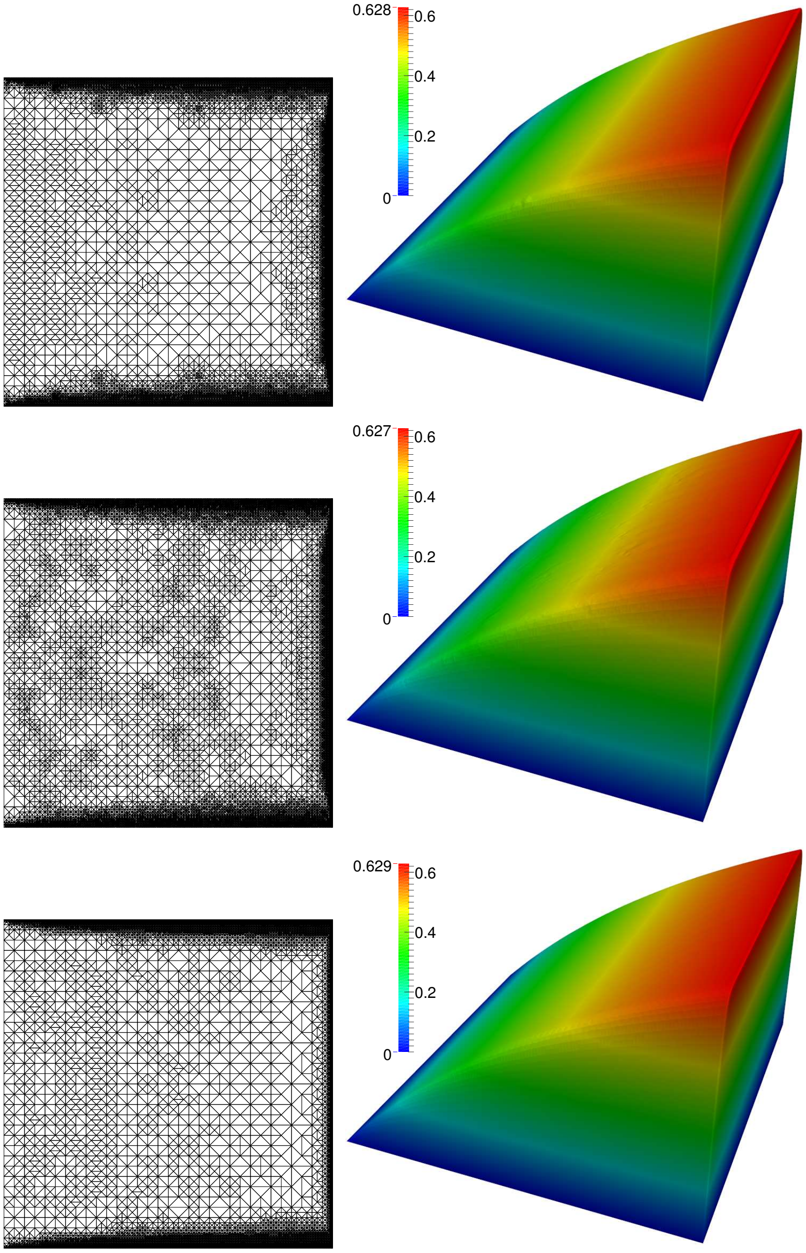

Fig. 6. *Example 2: The 30th adaptively refined meshes and approximations obtained on these meshes, for the F-BBK method (top), S-BBK method (center) and BJK method (bottom).*

problems in three dimensions, International Journal for Numerical Methods in Fluids, 73 (2013), pp. 765–790.

[2] M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Pure and Applied Mathematics (New York), Wiley-Interscience, New York, 2000.

[3] A. Allendes, F. Durán, and R. Rankin, *Error estimation for low-order adaptive finite element approximations for fluid flow problems*, IMA J. Numer. Anal., 36 (2016), pp. 1715–1747.

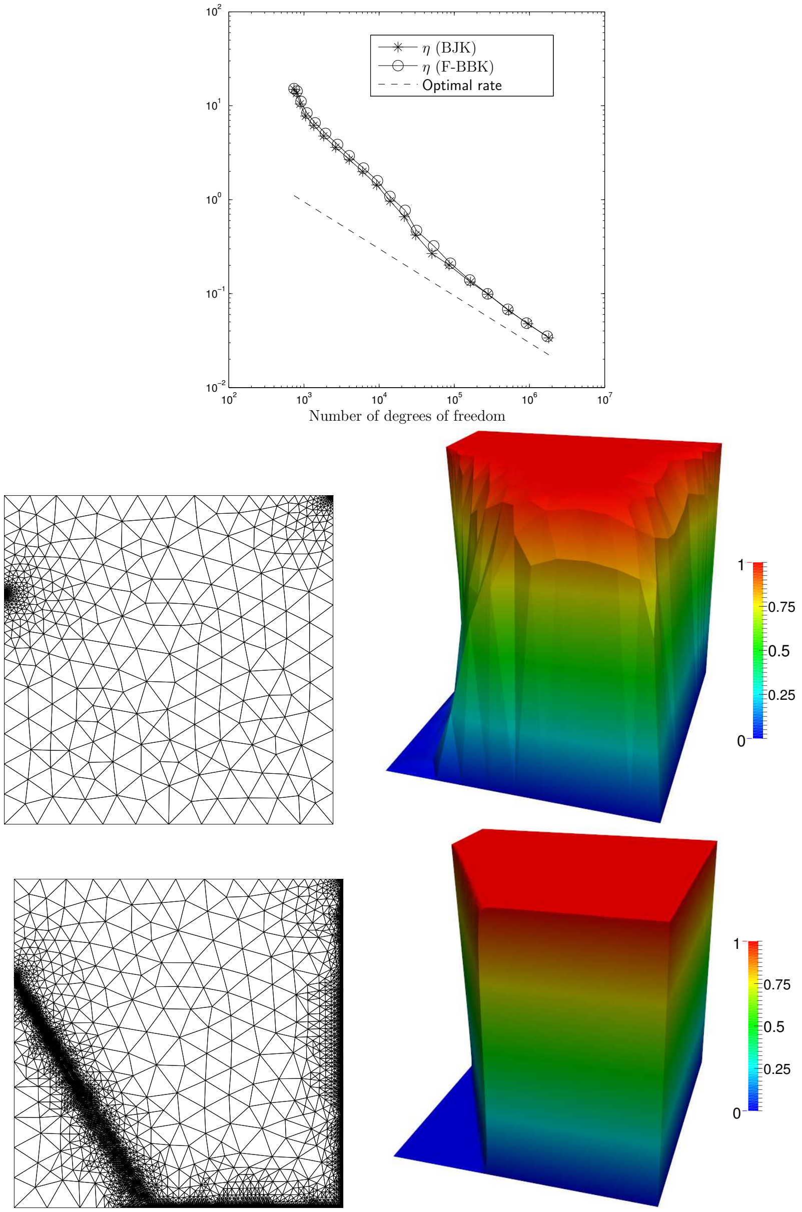

FIG. 7. *Example 3: The evolution of the estimator through the mesh refinement process (top), and, for the full BBK method, the initial mesh and approximate solution on this mesh (center) and the mesh and approximate solution after 15 adaptive steps (bottom).*

[4] P. R. Amestoy, I. S. Duff, J.-Y. L'Excellent, and J. Koster, *A fully asynchronous multifrontal solver using distributed dynamic scheduling*, SIAM J. Matrix Anal. Appl., 23 (2001), pp. 15–41 (electronic).

[5] P. R. Amestoy, A. Guermouche, J.-Y. L'Excellent, and S. Pralet, *Hybrid scheduling for the parallel solution of linear systems*, Parallel Comput., 32 (2006), pp. 136–156.

[6] R. Araya, A. H. Poza, and E. P. Stephan, *A hierarchical a posteriori error estimate for an advection-diffusion-reaction problem*, Math. Models Methods Appl. Sci., 15 (2005), pp. 1119–1139.

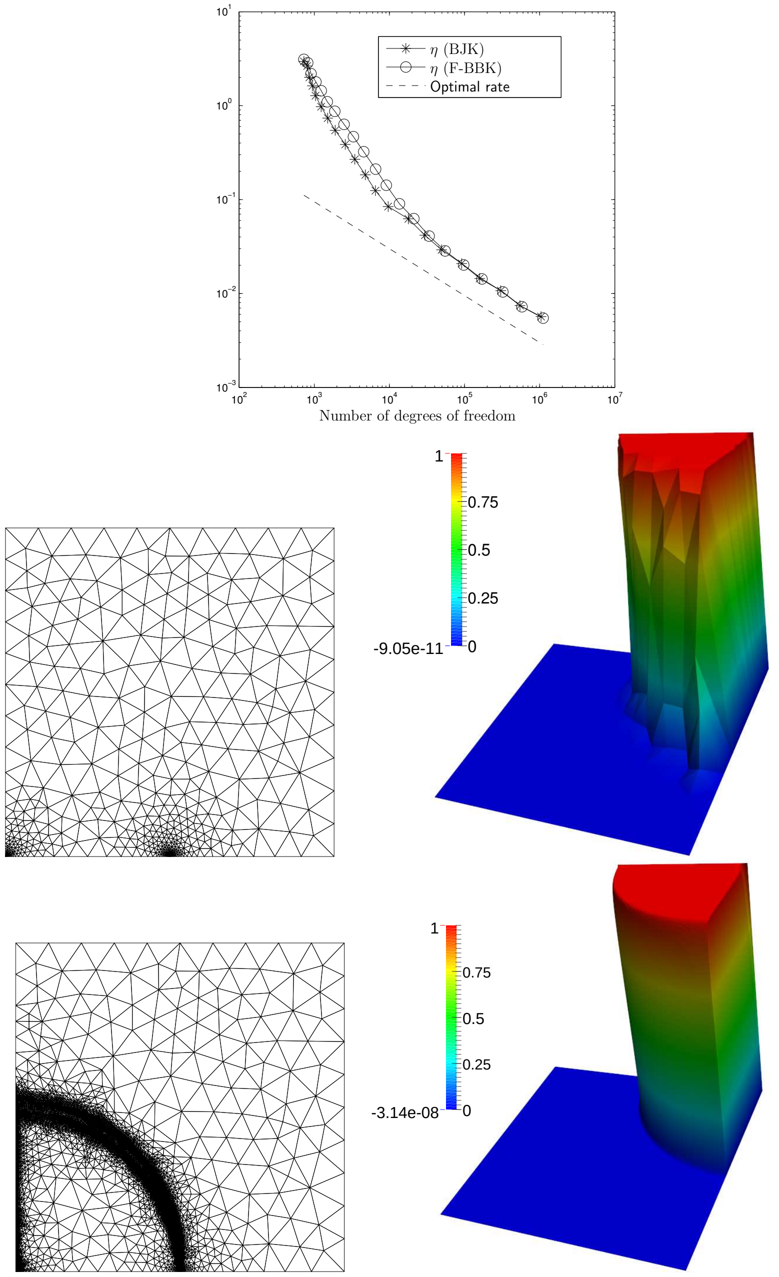

Fig. 8. *Example 4: The evolution of the estimator through the mesh refinement process (top), and, for the BJK method, the initial mesh and approximate solution on this mesh (center) and the mesh and approximate solution after 17 adaptive steps (bottom).*

[7] S. Badia and A. Hierro, *On monotonicity-preserving stabilized finite element approximations of transport problems*, SIAM J. Sci. Comput., 36 (2014), pp. A2673–A2697.

[8] G. R. Barrenechea, E. Burman, and F. Karakatsani, *Edge-based nonlinear diffusion for finite element approximations of convection-diffusion equations and its relation to algebraic flux-correction schemes*, Numer. Math., (to appear).

[9] G. R. Barrenechea, V. John, and P. Knobloch, *An algebraic flux correction scheme satis-*

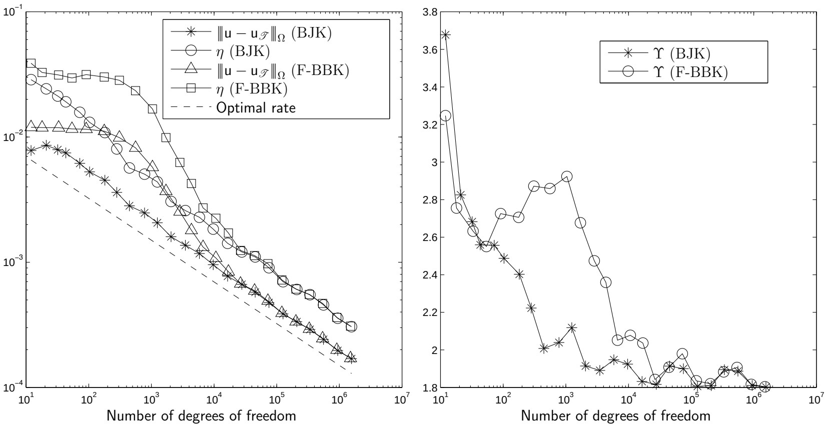

FIG. 9. *Example 5: The behaviour of the error $\|\mathsf{u}-\mathsf{u}_{\mathscr{T}}\|_\Omega$ and estimator η (left) and the effectivity indices Υ (right).*

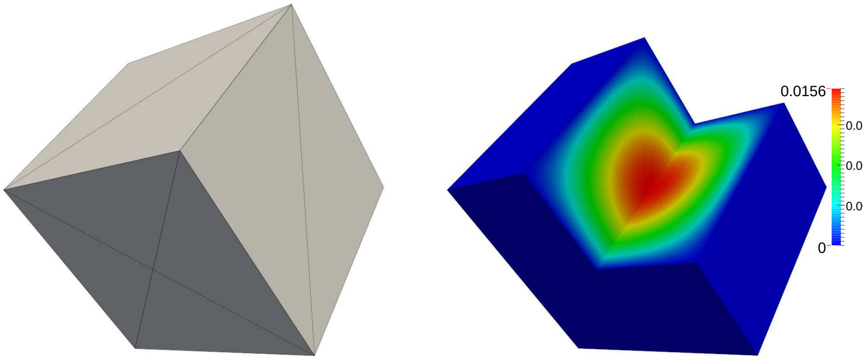

FIG. 10. *Example 5: Initial mesh (left) and the approximation obtained on the 23rd adaptively refined mesh using the full BBK method (right).*

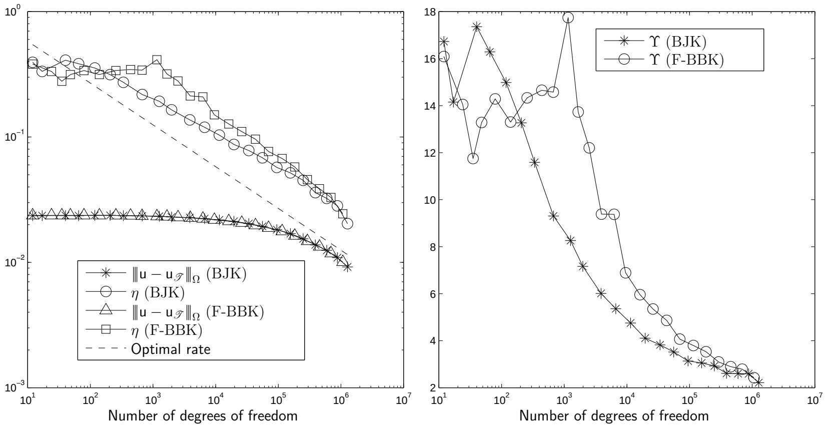

FIG. 11. *Example 6: The behaviour of the error $\|\mathsf{u}-\mathsf{u}_{\mathscr{T}}\|_\Omega$ and estimator η (left) and the effectivity indices Υ (right).*

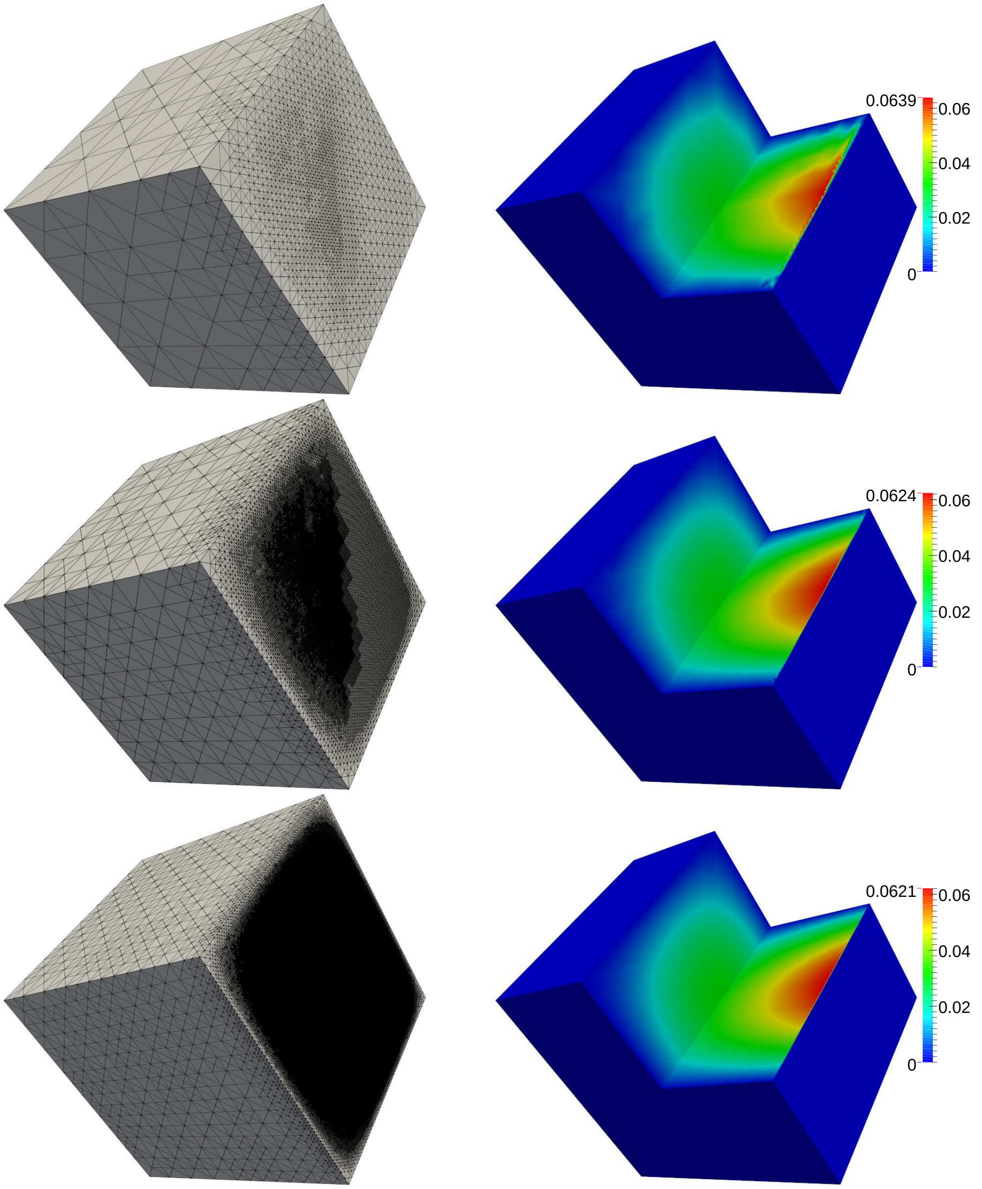

Fig. 12. *Example 6: The 12*th *(top), 17*th *(center) and 22*nd *(bottom) adaptively refined meshes (left) and approximate solutions (right) obtained on these meshes using the BJK method.*

fying the discrete maximum principle and linearity preservation on general meshes, Necas Center for Mathematical Modeling preprint NCMM/2016/06, (2016).

[10] M. Bebendorf, *A note on the Poincaré inequality for convex domains*, Z. Anal. Anwendungen, 22 (2003), pp. 751–756.

[11] S. Berrone, *Robustness in a posteriori error analysis for FEM flow models*, Numer. Math., 91 (2002), pp. 389–422.

[12] J. P. Boris and D. L. Book, *Flux-corrected transport. I. SHASTA, a fluid transport algorithm that works [J. Comput. Phys.* **11** *(1973), no. 1, 38–69]*, J. Comput. Phys., 135 (1997), pp. 170–186. With an introduction by Steven T. Zalesak, Commemoration of the 30th anniversary {of J. Comput. Phys.}.

[13] E. Burman and A. Ern, *Nonlinear diffusion and discrete maximum principle for stabilized Galerkin approximations of the convection–diffusion-reaction equation*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 3833–3855.

[14] E. Burman and A. Ern, *Stabilized Galerkin approximation of convection–diffusion–reaction equations: discrete maximum principle and convergence*, Math. Comp., 74 (2005), pp. 1637–1652.

[15] P. G. Ciarlet, *The finite element method for elliptic problems*, SIAM, Philadelphia, PA, 2002.

[16] B. Cockburn and P.-A. Gremaud, *Error estimates for finite element methods for scalar conservation laws*, SIAM J. Numer. Anal., 33 (1996), pp. 522–554.

[17] K. Eriksson and C. Johnson, *Adaptive streamline diffusion finite element methods for stationary convection-diffusion problems*, Math. Comp., 60 (1993), pp. 167–188, S1–S2.

[18] A. Ern and J.-L. Guermond, *Theory and practice of finite elements*, vol. 159, Springer Science & Business Media, 2013.

[19] A. Ern and J.-L. Guermond, *Weighting the edge stabilization*, SIAM J. Numer. Anal., 51 (2013), pp. 1655–1677.

[20] 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.

[21] S. K. Godunov, *A difference method for numerical calculation of discontinuous solutions of the equations of hydrodynamics*, Mat. Sb. (N.S.), 47 (89) (1959), pp. 271–306.

[22] P. Grisvard, *Elliptic problems in nonsmooth domains*, vol. 69 of Classics in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2011. Reprint of the 1985 original [MR0775683], With a foreword by Susanne C. Brenner.

[23] V. John and P. Knobloch, *On spurious oscillations at layers diminishing (SOLD) methods for convection–diffusion equations: Part I – A review*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 2197–2215.

[24] V. John and P. Knobloch, *On spurious oscillations at layers diminishing (SOLD) methods for convection–diffusion equations: Part II – Analysis for P_1 and Q_1 finite elements*, Comput. Methods Appl. Mech. Engrg., 197 (2008), pp. 1997–2014.

[25] V. John and J. Novo, *A robust SUPG norm a posteriori error estimator for stationary convection-diffusion equations*, Comput. Methods Appl. Mech. Engrg., 255 (2013), pp. 289–305.

[26] T. Knopp, G. Lube, and G. Rapin, *Stabilized finite element methods with shock capturing for advection-diffusion problems*, Comput. Methods Appl. Mech. Engrg., 191 (2002), pp. 2997–3013.

[27] D. Kuzmin, *Algebraic flux correction for finite element discretizations of coupled systems*, in Proceedings of the Int. Conf. on Computational Methods for Coupled Problems in Science and Engineering, M. Papadrakakis, E. Oñate, and B. Schrefler, eds., CIMNE, Barcelona, 2007, pp. 1–5.

[28] D. Kuzmin and M. Möller, *Algebraic flux correction. I. Scalar conservation laws*, in Flux-corrected transport, Sci. Comput., Springer, Berlin, 2005, pp. 155–206.

[29] D. Kuzmin, M. Möller, and S. Turek, *High-resolution FEM-FCT schemes for multidimensional conservation laws*, Comput. Methods Appl. Mech. Engrg., 193 (2004), pp. 4915–4946.

[30] D. Kuzmin and J. N. Shadid, *A new approach to enforcing discrete maximum principles in continuous Galerkin methods for convection-dominated transport equations*, tech. report, UA Ruhr Zentrum für partielle Differentialgleichungen, 2015.

[31] MATLAB, *(R2013a)*, The MathWorks Inc., Natick, Massachusetts, 2013.

[32] A. Mizukami and T. J. R. Hughes, *A Petrov-Galerkin finite element method for convection-dominated flows: an accurate upwinding technique for satisfying the maximum principle*, Comput. Methods Appl. Mech. Engrg., 50 (1985), pp. 181–193.

[33] L. E. Payne and H. F. Weinberger, *An optimal Poincaré inequality for convex domains*, Arch. Rational Mech. Anal., 5 (1960), pp. 286–292 (1960).

[34] G. Sangalli, *Robust a-posteriori estimator for advection-diffusion-reaction problems*, Math. Comp., 77 (2008), pp. 41–70.

[35] L. Tobiska and R. Verfürth, *Robust a posteriori error estimates for stabilized finite element methods*, IMA J. Numer. Anal., 35 (2015), pp. 1652–1671.

[36] R. Verfürth, *A review of a posteriori error estimation and adaptive mesh-refinement techniques*, John Wiley & Sons Inc, 1996.

[37] R. Verfürth, *A posteriori error estimators for convection-diffusion equations*, Numer. Math., 80 (1998), pp. 641–663.

[38] R. Verfürth, *Robust a posteriori error estimates for nonstationary convection-diffusion equa-*

tions, SIAM J. Numer. Anal., 43 (2005), pp. 1783–1802 (electronic).

[39] J. Xu and L. Zikatanov, *A monotone finite element scheme for convection-diffusion equations*, Math. Comp, 68 (1999), pp. 1429–1446.

[40] S. T. Zalesak, *Fully multidimensional flux-corrected transport algorithms for fluids*, J. Comput. Phys., 31 (1979), pp. 335–362.