Stabilization and a posteriori error analysis of a mixed FEM for convection-diffusion problems with mixed boundary conditions *

María González† and Magdalena Strugaru‡

March 5, 2020

Abstract

We introduce a new augmented dual-mixed finite element method for the linear convection-diffusion equation with mixed boundary conditions. The approach is based on adding suitable residual type terms to a dual-mixed formulation of the problem. We prove that for appropriate values of the stabilization parameters, that depend on the diffusivity and the magnitude of the convective velocity, the new variational formulation and the corresponding Galerkin scheme are well-posed and a Céa estimate can be derived. We establish the rate of convergence when the flux and the concentration are approximated, respectively, by Raviart-Thomas/Brezzi-Douglas-Marini and continuous piecewise polynomials. In addition, we develop an a posteriori error analysis of residual type. We derive a simple a posteriori error indicator and prove that it is reliable and locally efficient. Finally, we provide some numerical experiments that illustrate the performance of the method.

Mathematics Subject Classifications (2010): 65N30; 65N12; 65N15

Key words: convection-diffusion, mixed finite element, stabilization, a posteriori error estimates.

*The research of the first author was partially supported by MICINN grant MTM2016-76497-R. The research of the second author was supported by the Basque Government through the BERC 2014-2017 programme and by the Spanish Ministry of Economy and Competitivity through the BCAM Severo Ochoa excellence accreditation SEV-2013-0323.

†Departamento de Matemáticas, Universidade da Coruña, Campus de Elviña s/n, 15071, A Coruña, Spain. E-mail: maria.gonzalez.taboada@udc.es.

‡Basque Center of Applied Mathematics, Alameda Mazarredo 14, 48009, Bilbao, Spain. E-mail: mstrugaru@bcamath.org

1 Introduction

Convection-diffusion problems are used to simulate the transport of quantities, such as temperature or concentration. The numerical solution of this kind of problems has attracted great attention, specially in the case when convection is dominant. Indeed, in that case standard finite element methods may lead to numerical solutions with unphysical oscillations, due to a lack of stability. One way to circumvent these difficulties is by adding some artificial diffusion to the discretization. Different stabilization methods have been proposed in the literature; among them, the streamline-upwind Petrov Galerkin (SUPG) method or streamline-diffusion method (SDM) [26, 30] is probably the most popular one.

In general, the solution of a convection-diffusion equation possesses small subregions, named layers, where the derivatives of the solution are very large. Therefore, the use of adaptive algorithms based on robust a posteriori error estimators becomes crucial in the numerical approximation. Many authors have developed and analyzed a posteriori error estimators for different discretizations of the convection-diffusion equation. For instance, for the SUPG method, Verfürth [40] presented in 1998 three reliable and locally efficient a posteriori error estimators for the error measured in the energy norm. Those estimates are optimal provided that the local mesh Peclet number is sufficiently small. Numerical studies of different a posteriori error estimators can be found in [28, 34]. In 2005, Verfürth [41] incorporated to the usual energy norm a dual norm of the convective derivative and proved that the proposed a posteriori error estimators are fully robust in the sense that the ratio of the upper and lower bounds is uniformly bounded with respect to the mesh-size, to the diffusivity and to the size of the convection. A hierarchical a posteriori error estimator was proposed in Achchab et al. [1] and a robust residual-based a posteriori error estimator for the error in the natural SUPG norm was proposed by John and Novo [29]. Fully computable upper bounds for the error measured in the energy norm were introduced by Ainsworth et al. [3]. More recently, Du and Zhang [16] introduced a novel dual norm under which the error estimator is robust with respect to the diffusivity parameter and presented impressive numerical results. Lastly, Eigel and Merdon [17] presented reliable and efficient a posteriori error estimators for the error in the energy norm based on the reconstruction of equilibrated fluxed.

In this paper we are interested in the simultaneous approximation of the concentration and the flux in a linear convection-diffusion equation with mixed boundary conditions using mixed finite element methods. In [13], Douglas and Roberts introduced a mixed finite element discretization of a convection-diffusion-reaction equation based on the Raviart-Thomas-Nédélec space and obtained $L^2(\Omega)$-estimates for the error. Negative norm and $L^\infty(\Omega)$ estimates can be derived by applying the same arguments as in [14]. However, that formulation leads to a centered discretization of the convective term that is numerically unstable when convection is important. Moreover, this formulation cannot be extended to certain classes of nonlinear problems (see [27]). In [27], Jaffre proposed a numerical scheme where the diffusion term is ap-

proximated by Raviart-Thomas mixed finite elements and the convective term is approximated by the Lesaint-Raviart upwind scheme for discontinuous finite elements. The error analysis of that scheme showed that the convection term is approximated less precisely than the diffusion term. Then, a modification of the initial formulation that allows to balance errors was proposed. Thomas [37] introduced and analysed a dual-mixed variational formulation with artificial diffusion that contains the one proposed by Jaffre [27]. It is well adapted to convection-dominated flow problems and he obtained optimal convergence rates $\mathcal{O}(h^k)$ provided that the flux is approximated by Raviart-Thomas elements of order k and the concentration is approximated by discontinuous piecewise polynomials of order k. In [19] the authors presented a mixed-hybrid finite element method for scalar convection-diffusion problems. The diffusive term is approximated with the lowest order Raviart-Thomas finite element whereas the convective term is treated by means of a Lagrange multiplier. Numerical results are very accurate for local Peclet numbers up to 5. However, no convergence analysis is provided.

Concerning the a posteriori error analysis of mixed finite element methods, we refer to the residual a posteriori error estimates derived by Vohralík [43] for lowest-order Raviart-Thomas mixed finite element discretizations of convection-diffusion-reaction equations. Later, Kim and Park [31] presented a posteriori error estimates for the error in the L^2-norm. Finally, Du [15] introduced new techniques to derive residual-based a posteriori error estimates over the stress and scalar displacement error for the lowest-order Raviart-Thomas mixed finite element.

In this work we present novel augmented mixed finite element methods for the scalar convection-diffusion equation and developed an a posteriori error analysis of residual type. The use of augmented mixed finite element methods allows to avoid the inf-sup condition in the analysis and, as a consequence, one can use a wider set of finite element subspaces in the discretization. Several augmented mixed finite element methods have been proposed in last years for different problems. Indeed, in [33] Masud and Hughes introduced an augmented mixed finite element method for Darcy flow. Later on, this method was extended by Gatica [20] to the linear elasticity problem with homogeneous Dirichlet boundary conditions. The case of nonhomogeneous Dirichlet boundary conditions was analyzed in [21] (see also [23]) and nonhomogeneous mixed boundary conditions were considered in [24]. This technique has been applied later to the generalized Stokes problem [5, 6], anisotropic porous media flow [7, 8], the linear convection-diffusion equation [25] and the Oseen problem [9]. However, in the papers [5, 25, 9] only homogeneous boundary conditions of Dirichlet type were treated.

The aim of this paper is to propose and analyze augmented mixed finite element methods for the linear convection-diffusion equation with mixed boundary conditions. We also develop a residual based a posteriori error analysis and derive simple a posteriori error indicators for the two and three dimensional cases. We prove that these a posteriori error indicators are reliable and locally efficient, and thus, can be used to drive a mesh adaptation process. We do not consider the case of dominant convection, that will be the subject of a forthcoming paper.

The paper is organized as follows. In Section 2 we derive a dual-mixed variational for-

mulation of the model problem. Then, in Section 3 we introduce and analyze the augmented dual-mixed variational formulation. In Section 4 we study the stabilized mixed finite element method. We prove the stability of the Galerkin scheme and derive the rate of convergence for some particular discretizations, using Raviart-Thomas or Brezzi-Douglas-Marini elements to approximate the flux and continuous piecewise polynomials to approximate the scalar variable. In Section 5, we introduce new a posteriori error estimators both in two and three dimensions, and prove that they are reliable and locally efficient. Some numerical experiments are reported in Section 6. Finally, in Section 7 we draw conclusions.

Throughout the paper, we use the usual notations for Sobolev spaces and norms. In particular, C or c, with or without subscripts, will denote generic constants independent of the discretization parameter.

2 Dual-mixed variational formulation

Let Ω be a bounded connected open subset of $\mathbb{R}^d$ $(d = 2, 3)$, with a Lipschitz-continuous boundary Γ. We assume that Γ consists of two disjoint parts, Γ_D and Γ_N, such that $\Gamma = \overline{\Gamma_D} \cup \overline{\Gamma_N}$ and $|\Gamma_D| > 0$. Let $\mathcal{K} \in [L^\infty(\Omega)]^{d \times d}$ be a symmetric and uniformly positive definite tensor, that is, $\mathcal{K}$ satisfies

$$(\mathcal{K}(\mathbf{x})\,\mathbf{y}) \cdot \mathbf{y} \;\geq\; \alpha\,\|\mathbf{y}\|^2\,, \quad \text{a.e. } \mathbf{x} \in \Omega\,, \quad \forall\,\mathbf{y} \in \mathbb{R}^d\,, \tag{1}$$

for some $\alpha > 0$. Then, we also have that $\mathcal{K}^{-1} \in [L^\infty(\Omega)]^{d \times d}$ and

$$(\mathcal{K}^{-1}(\mathbf{x})\,\mathbf{y}) \cdot \mathbf{y} \geq \frac{\alpha}{\|\mathcal{K}\|^2_{\infty,\Omega}}\,\|\mathbf{y}\|^2\,, \quad \text{a.e. } \mathbf{x} \in \Omega\,, \quad \forall\,\mathbf{y} \in \mathbb{R}^d\,, \tag{2}$$

where we denote by $\|\cdot\|_{\infty,\Omega}$ the usual norm in $[L^\infty(\Omega)]^{d \times d}$.

Let $\mathbf{b} \in [L^\infty(\Omega)]^d$ be a solenoidal velocity field, that is

$$\mathrm{div}(\mathbf{b}) = 0 \quad \text{in } \Omega\,, \tag{3}$$

such that

$$\mathbf{b} \cdot \mathbf{n} \geq 0 \quad \text{on } \Gamma_N\,. \tag{4}$$

Then, given a source $f \in L^2(\Omega)$, $g \in H^{1/2}(\Gamma_D)$ and $z \in H^{-1/2}(\Gamma_N)$, we consider the following steady convection-diffusion problem: find the concentration $u : \Omega \to \mathbb{R}$ such that

$$\left\{ \begin{array}{rcll} -\,\mathrm{div}(\mathcal{K}\,\nabla u) \,+\, \mathbf{b} \cdot \nabla u & = & f & \text{in } \Omega\,, \\ u & = & g & \text{on } \Gamma_D\,, \\ \mathcal{K}\,\nabla u \cdot \mathbf{n} & = & z & \text{on } \Gamma_N\,, \end{array} \right. \tag{5}$$

where $\mathbf{n}$ is the unit outward normal vector to Γ.

We introduce the flux $\sigma := \mathcal{K}\nabla u$ in Ω as an additional unknown. Then, problem (5) can be stated equivalently as follows: find $\sigma : \Omega \to \mathbb{R}^d$ and $u : \Omega \to \mathbb{R}$ such that

$$\left\{ \begin{array}{rcll} -\mathrm{div}(\sigma) + \mathbf{b}\cdot\nabla u &=& f & \text{in } \Omega\,, \\ \mathcal{K}^{-1}\sigma - \nabla u &=& 0 & \text{in } \Omega\,, \\ u &=& g & \text{on } \Gamma_D\,, \\ \sigma\cdot\mathbf{n} &=& z & \text{on } \Gamma_N\,. \end{array} \right. \tag{6}$$

Given $s \in H^{-1/2}(\Gamma_N)$, we define the space $H_s := \{\tau \in H(\mathrm{div};\Omega) : \tau\cdot\mathbf{n} = s \quad \text{on } \Gamma_N\}$. Then, multiplying the two first equations in (6) by appropriate test functions and integrating by parts, we derive the following dual-mixed variational formulation of problem (6): find $\sigma \in H_z$ and $u \in H^1(\Omega)$ such that

$$\left\{ \begin{array}{l} \displaystyle\int_\Omega \mathcal{K}^{-1}\sigma\cdot\tau + \int_\Omega u\,\mathrm{div}(\tau) = \int_{\Gamma_D} g\,\tau\cdot\mathbf{n}\,, \quad \forall\,\tau\in H_0\,, \\ \displaystyle\int_\Omega \mathrm{div}(\sigma)\,v - \int_\Omega \mathbf{b}\cdot\nabla u\,v = -\int_\Omega f v\,, \quad \forall\, v\in H^1(\Omega)\,. \end{array} \right. \tag{7}$$

We remark that the Neumann boundary condition is essential whereas the Dirichlet boundary condition is natural in this formulation.

Let us consider the following decomposition: $\sigma = \sigma_0 + \sigma_z \in H_0 + H_z$, that is, $\sigma_0\cdot\mathbf{n} = 0$ on Γ_N and $\sigma_z\cdot\mathbf{n} = z$ on Γ_N. Then, problem (7) is equivalent to: find $(\sigma_0, u) \in H_0 \times H^1(\Omega)$ such that

$$\left\{ \begin{array}{l} \displaystyle\int_\Omega \mathcal{K}^{-1}\sigma_0\cdot\tau + \int_\Omega u\,\mathrm{div}(\tau) = \int_{\Gamma_D} g\,\tau\cdot\mathbf{n} - \int_\Omega \mathcal{K}^{-1}\sigma_z\cdot\tau\,, \quad \forall\,\tau\in H_0\,, \\ \displaystyle\int_\Omega \mathrm{div}(\sigma_0)\,v - \int_\Omega \mathbf{b}\cdot\nabla u\,v = -\int_\Omega f v - \int_\Omega \mathrm{div}(\sigma_z)\,v\,, \quad \forall\, v\in H^1(\Omega)\,. \end{array} \right. \tag{8}$$

Let us define the bilinear forms $a : H_0\times H_0 \to \mathbb{R}$, $b : H^1(\Omega)\times H_0 \to \mathbb{R}$ and $c : H^1(\Omega)\times H^1(\Omega) \to \mathbb{R}$ by

$$a(\zeta,\tau) := \int_\Omega \mathcal{K}^{-1}\zeta\cdot\tau\,, \quad b(w,\tau) := \int_\Omega w\,\mathrm{div}(\tau)\,, \quad c(w,v) := \int_\Omega \mathbf{b}\cdot\nabla w\,v$$

for all ζ, $\tau \in H_0$ and w, $v \in H^1(\Omega)$, and the linear functionals $m : H_0 \to \mathbb{R}$ and $l : H^1(\Omega) \to \mathbb{R}$ by

$$m(\tau) := \int_{\Gamma_D} g\,\tau\cdot\mathbf{n} - \int_\Omega \mathcal{K}^{-1}\sigma_z\cdot\tau\,, \quad l(v) := -\int_\Omega (f + \mathrm{div}(\sigma_z))\,v\,,$$

for all $\tau \in H_0$ and $v \in H^1(\Omega)$. Then, the variational formulation (8) can be rewritten as follows: find $(\sigma_0, u) \in H_0 \times H^1(\Omega)$ such that

$$\left\{ \begin{array}{rcll} a(\sigma_0, \tau) + b(u, \tau) & = & m(\tau)\,, & \forall\, \tau \in H_0\,, \\ b(v, \sigma_0) - c(u, v) & = & l(v)\,, & \forall\, v \in H^1(\Omega)\,. \end{array} \right.$$

Thus, problem (8) has a generalized saddle point structure. Since the bilinear form $c(\cdot,\cdot)$ is not symmetric, in order to ensure that problem (8) is well-posed using the generalized Babuška-Brezzi theory, we need the bilinear form $a(\cdot,\cdot)$ to be elliptic in H_0. However, we only have

$$a(\tau, \tau) \,\geq\, \frac{\alpha}{\|\mathcal{K}\|^2_{\infty,\Omega}}\, \|\tau\|^2_{[L^2(\Omega)]^d}\,, \quad \forall\, \tau \in H_0\,.$$

3 Augmented dual-mixed variational formulation

We follow the ideas in [25, 21, 24] and subtract the second equation in (8) from the first one and then add the following residual terms:

$$\kappa_1 \int_\Omega (\mathrm{div}(\sigma_0) - \mathbf{b}\cdot\nabla u)(\mathrm{div}(\tau) + \mathbf{b}\cdot\nabla v) \,=\, -\kappa_1 \int_\Omega (f \,+\, \mathrm{div}(\sigma_z))\,(\mathrm{div}(\tau) + \mathbf{b}\cdot\nabla v)$$

$$\kappa_2 \int_\Omega (\nabla u - \mathcal{K}^{-1}\sigma_0)\,\cdot\,(\nabla v + \mathcal{K}^{-1}\tau) \,=\, \kappa_2 \int_\Omega \mathcal{K}^{-1}\,\sigma_z\,\cdot\,(\nabla v + \mathcal{K}^{-1}\tau)$$

and

$$\kappa_3 \int_{\Gamma_D} u\, v = \kappa_3 \int_{\Gamma_D} g\, v\,,$$

where we assume that (σ_0, u) is a solution of (8) and $(\tau, v) \in \mathbf{H} := H_0 \times H^1(\Omega)$.

Proceeding in this way, we obtain the following augmented variational formulation: *find* $(\sigma_0, u) \in \mathbf{H}$ *such that*

$$A_s((\sigma_0, u), (\tau, v)) = F_s(\tau, v)\,, \quad \forall\, (\tau, v) \in \mathbf{H}, \tag{9}$$

where the bilinear form $A_s : \mathbf{H} \times \mathbf{H} \to \mathbb{R}$ and the linear functional $F_s : \mathbf{H} \to \mathbb{R}$ are defined by

$$A_s((\zeta, w), (\tau, v)) := \int_\Omega \mathcal{K}^{-1}\zeta\,\cdot\,\tau \,+\, \int_\Omega w\,\mathrm{div}(\tau) \,-\, \int_\Omega \mathrm{div}(\zeta)\,v \,+\, \int_\Omega \mathbf{b}\cdot\nabla w\, v$$
$$+\,\kappa_1 \int_\Omega (\mathrm{div}(\zeta) - \mathbf{b}\cdot\nabla w)\,(\mathrm{div}(\tau) + \mathbf{b}\cdot\nabla v) \,+\, \kappa_2 \int_\Omega (\nabla w - \mathcal{K}^{-1}\,\zeta)\,\cdot\,(\nabla v + \mathcal{K}^{-1}\,\tau)$$
$$+\,\kappa_3 \int_{\Gamma_D} w\, v$$

and

$$F_s(\tau, v) := \int_\Omega (f + \mathrm{div}(\sigma_z))\, v - \int_\Omega \mathcal{K}^{-1}\sigma_z \cdot \tau - \kappa_1 \int_\Omega (f + \mathrm{div}(\sigma_z))\,(\mathrm{div}(\tau) + \mathbf{b}\cdot\nabla v) + \kappa_2 \int_\Omega \mathcal{K}^{-1}\sigma_z(\nabla v + \mathcal{K}^{-1}\tau) + \int_{\Gamma_D} g\,\tau\cdot\mathbf{n} + \kappa_3 \int_{\Gamma_D} g\, v$$

for all $(\zeta, w), (\tau, v) \in \mathbf{H}$.

The stabilization parameters, κ_1, κ_2 and κ_3, are positive constants to be chosen so that the augmented bilinear form be coercive in the whole space $\mathbf{H}$. We endow $\mathbf{H}$ with the product norm

$$\|(\tau, v)\|_{\mathbf{H}} := (\|\tau\|^2_{H(\mathrm{div};\Omega)} + \|v\|^2_{H^1(\Omega)})^{1/2}, \quad \forall\, (\tau, v) \in \mathbf{H}.$$

Clearly, F_s is a linear continuous functional in $\mathbf{H}$. Indeed, using the Cauchy-Schwarz inequality, we have

$$\|F_s\| \leq C_F \left(||f||_{L^2(\Omega)} + ||g||_{H^{1/2}(\Gamma_D)} + ||\sigma_z||_{[L^2(\Omega)]^d} + ||\mathrm{div}(\sigma_z)||_{L^2(\Omega)} \right),$$

with $C_F := \max(1 + \kappa_3, \|\mathcal{K}^{-1}\|_{\infty,\Omega}(1 + \kappa_2(1 + \|\mathcal{K}^{-1}\|_{\infty,\Omega})), 1 + \kappa_1(1 + \sqrt{d}\|\mathbf{b}\|_{[L^\infty(\Omega)]^d}))$.

Similarly, it is easy to see that the bilinear form $A_s(\cdot,\cdot)$ is continuous in $\mathbf{H}$. In this case, using the Cauchy-Schwarz inequality, we obtain

$$|A_s((\zeta, w), (\tau, v))| \leq M\, ||(\zeta, w)||_{\mathbf{H}}\, ||(\tau, v)||_{\mathbf{H}}, \quad \forall\, (\zeta, w),\, (\tau, v) \in \mathbf{H},$$

with $M := 2 + \|\mathcal{K}^{-1}\|_{\infty,\Omega} + \sqrt{d}\|\mathbf{b}\|_{[L^\infty(\Omega)]^d} + \kappa_1(1 + \sqrt{d}\|\mathbf{b}\|_{[L^\infty(\Omega)]^d})^2 + \kappa_2(1 + \|\mathcal{K}^{-1}\|_{\infty,\Omega})^2 + \kappa_3$.

Now, let $(\tau, v) \in \mathbf{H}$. Then, from the definition of $A_s(\cdot,\cdot)$ and using (2) we have that

$$\begin{aligned} A_s((\tau, v), (\tau, v)) &\geq \frac{\alpha}{\|\mathcal{K}\|^2_{\infty,\Omega}} ||\tau||^2_{[L^2(\Omega)]^d} + \int_\Omega \mathbf{b}\cdot\nabla v\, v \\ &+ \kappa_1(||\mathrm{div}(\tau)||^2_{L^2(\Omega)} - ||\mathbf{b}\cdot\nabla v||^2_{L^2(\Omega)}) \\ &+ \kappa_2(||\nabla v||^2_{[L^2(\Omega)]^d} - ||\mathcal{K}^{-1}\tau||^2_{[L^2(\Omega)]^d}) + \kappa_3\|v\|^2_{L^2(\Gamma_D)}. \end{aligned}$$

Integrating by parts and using (3) and (4), we have for $v \in H^1(\Omega)$,

$$\int_\Omega \mathbf{b}\cdot\nabla v\, v = \frac{1}{2}\int_\Omega \mathbf{b}\cdot\nabla(v^2) = \frac{1}{2}\int_\Gamma \mathbf{b}\cdot\mathbf{n}\, v^2 \geq \frac{1}{2}\int_{\Gamma_D} \mathbf{b}\cdot\mathbf{n}\, v^2.$$

Then,

$$\begin{aligned} A_s((\tau, v), (\tau, v)) &\geq (\frac{\alpha}{\|\mathcal{K}\|^2_{\infty,\Omega}} - \kappa_2\,\|\mathcal{K}^{-1}\|^2_{\infty,\Omega})||\tau||^2_{[L^2(\Omega)]^d} + \kappa_1||\mathrm{div}(\tau)||^2_{L^2(\Omega)} \\ &+ (\kappa_2 - \kappa_1 d\|\mathbf{b}\|^2_{[L^\infty(\Omega)]^d})\|\nabla v\|^2_{[L^2(\Omega)]^d} + (\kappa_3 - \frac{1}{2}\|\mathbf{b}\cdot\mathbf{n}\|_{L^\infty(\Gamma_D)})\|v\|^2_{L^2(\Gamma_D)}. \end{aligned}$$

Therefore, using that

$$\|\nabla v\|^2_{[L^2(\Omega)]^d} + \|v\|^2_{L^2(\Gamma_D)} \geq c\,\|v\|^2_{H^1(\Omega)}\,,$$

for some positive constant c, the coercivity of $A_s(\cdot,\cdot)$ in $\mathbf{H}$ follows, with an ellipticity constant

$$C_{\mathtt{ell}} := \min\Big(\frac{\alpha}{\|\mathcal{K}\|^2_{\infty,\Omega}} - \kappa_2\,\|\mathcal{K}^{-1}\|^2_{\infty,\Omega}, \kappa_1, (\kappa_2 - \kappa_1 d\|\mathbf{b}\|^2_{[L^\infty(\Omega)]^d})c, (\kappa_3 - \frac{1}{2}\|\mathbf{b}\cdot\mathbf{n}\|_{L^\infty(\Gamma_D)})c\Big)\,,$$

provided that

$$0 < \kappa_1 < \frac{\kappa_2}{d\,||\mathbf{b}||^2_{[L^\infty(\Omega)]^d}}, \quad 0 < \kappa_2 < \frac{\alpha}{\|\mathcal{K}\|^2_{\infty,\Omega}\,\|\mathcal{K}^{-1}\|^2_{\infty,\Omega}} \quad \text{and} \quad \kappa_3 > \frac{1}{2}\|\mathbf{b}\cdot\mathbf{n}\|_{L^\infty(\Gamma_D)}\,. \tag{10}$$

We have the following result concerning the well-posedness of problem (9).

Theorem 1 *Under assumptions* (1), (3), (4) *and* (10), *problem* (9) *has a unique solution,* $(\sigma_0, u) \in \mathbf{H}$.

Proof. It follows from the previous considerations and Lax-Milgram Lemma. □

Remark. A feasible choice for the stabilization parameters would be

$$\kappa_1 = \frac{\kappa_2}{2\,d\,||\mathbf{b}||^2_{[L^\infty(\Omega)]^d}}, \qquad \kappa_2 = \frac{\alpha}{2\,\|\mathcal{K}\|^2_{\infty,\Omega}\,\|\mathcal{K}^{-1}\|^2_{\infty,\Omega}}, \qquad \kappa_3 = \|\mathbf{b}\cdot\mathbf{n}\|_{L^\infty(\Gamma_D)}.$$

For this choice, the ellipticity constant is

$$C_{\mathtt{ell}} = \min(\frac{\alpha}{2\,\|\mathcal{K}\|^2_{\infty,\Omega}}, \frac{\alpha}{4\,d\,||\mathbf{b}||^2_{[L^\infty(\Omega)]^d}\,\|\mathcal{K}\|^2_{\infty,\Omega}\,\|\mathcal{K}^{-1}\|^2_{\infty,\Omega}}, \frac{\alpha}{4\,\|\mathcal{K}\|^2_{\infty,\Omega}\,\|\mathcal{K}^{-1}\|^2_{\infty,\Omega}}\,c, \frac{1}{2}\|\mathbf{b}\cdot\mathbf{n}\|_{L^\infty(\Gamma_D)}c)\,.$$

In the particular case when $\mathcal{K} = \epsilon\,\mathcal{I}$, with $\epsilon > 0$, we have that

$$C_{\mathtt{ell}} = \min(\frac{1}{2\,\epsilon}, \frac{\epsilon}{4\,d\,||\mathbf{b}||^2_{[L^\infty(\Omega)]^d}}, \frac{\epsilon}{4}\,c, \frac{1}{2}\|\mathbf{b}\cdot\mathbf{n}\|_{L^\infty(\Gamma_D)}c)\,.$$

For $\epsilon << 1$ and $\|\mathbf{b}\| = \mathcal{O}(1)$, we have that $C_{\mathtt{ell}} = \mathcal{O}(\epsilon^{-1})$. □

4 Augmented mixed finite element method

Let $\{\mathcal{T}_h\}_{h>0}$ be a family of shape-regular meshes of $\bar{\Omega}$ made up of triangles if $d = 2$ or tetrahedra if $d = 3$. We denote by h_T the diameter of an element $T \in \mathcal{T}_h$ and define $h := \max_{T\in\mathcal{T}_h} h_T$. Let

H_h and V_h be any finite element subspaces of H_0 and $H^1(\Omega)$, respectively. Then, the Galerkin scheme associated to problem (9) reads: *find* $(\sigma_{0,h}, u_h) \in \mathbf{H}_h := H_h \times V_h$ *such that*

$$A_s((\sigma_{0,h}, u_h), (\tau_h, v_h)) = F_s(\tau_h, v_h)\,, \qquad \forall\, (\tau_h, v_h) \in \mathbf{H}_h\,. \tag{11}$$

Under the hypotheses of Theorem 1, problem (11) has a unique solution $(\sigma_{0,h}, u_h) \in H_h \times V_h$. Moreover, there exists a constant $C > 0$, independent of h, such that

$$||(\sigma_0 - \sigma_{0,h}, u - u_h)||_{\mathbf{H}} \leq C \inf_{(\tau_h, v_h) \in \mathbf{H}_h} ||(\sigma_0 - \tau_h, u - v_h)||_{\mathbf{H}}\,. \tag{12}$$

In order to establish a rate of convergence result, we consider specific finite element subspaces H_h and V_h. Hereafter, given $T \in \mathcal{T}_h$ and an integer $l \geq 0$, we denote by $\mathcal{P}_l(T)$ the space of polynomials of total degree at most l defined on T and, given an integer $r \geq 0$, we denote by $\mathcal{RT}_r(T)$ the local Raviart-Thomas space of order r (cf. [36]), that is,

$$\mathcal{RT}_r(T) := [\mathcal{P}_r(T)]^d \oplus [\mathbf{x}]\mathcal{P}_r(T) \subset [\mathcal{P}_{r+1}(T)]^d\,,$$

where $\mathbf{x}$ is a generic vector of $\mathbb{R}^d$.

Let $r \geq 0$ and $m \geq 1$. Then, we define (see [10, 36])

$$H_h := \mathcal{RT}_r = \Big\{\tau_h \in H_0 \,:\, \tau_h\big|_T \in \mathcal{RT}_r(T), \quad \forall\, T \in \mathcal{T}_h\Big\},$$

or

$$H_h := \mathcal{BDM}_{r+1} = \Big\{\tau_h \in H_0 \,:\, \tau_h\big|_T \in [\mathcal{P}_{r+1}(T)]^d, \quad \forall\, T \in \mathcal{T}_h\Big\},$$

and

$$V_h := \Big\{v_h \in \mathcal{C}(\overline{\Omega}) \,:\, v_h\big|_T \in \mathcal{P}_m(T), \quad \forall\, T \in \mathcal{T}_h\Big\}\,.$$

The corresponding rate of convergence is given in the next theorem.

Theorem 2 *Assume* $\sigma_0 \in [H^t(\Omega)]^d$, $\mathrm{div}(\sigma_0) \in H^t(\Omega)$ *and* $u \in H^{t+1}(\Omega)$. *Then, under the assumptions of Theorem 1, there exists* $C_{\mathbf{err}} > 0$, *independent of* h, *such that*

$$||(\sigma_0 - \sigma_{0,h}, u - u_h)||_{\mathbf{H}} \leq C_{\mathbf{err}}\, h^{\min\{t,m,r+1\}} \Big(||\sigma_0||_{[H^t(\Omega)]^d} + ||\mathrm{div}(\sigma_0)||_{H^t(\Omega)} + ||u||_{H^{t+1}(\Omega)}\Big)\,.$$

Proof. It follows straightforwardly from inequality (12) and the approximation properties of the corresponding finite element subspaces. □

5 A posteriori error analysis

In this section, we assume the hypotheses of Theorem 1 and develop a residual-based a posteriori error analysis of the augmented mixed finite element method (11). We derive a simple a posteriori error indicator and prove that it is reliable and locally efficient.

Let H_h and V_h be any finite element subspaces of H_0 and $H^1(\Omega)$, respectively, and let $(\sigma_0, u) \in \mathbf{H}$ and $(\sigma_{0,h}, u_h) \in \mathbf{H}_h := H_h \times V_h$ be the unique solutions to problems (9) and (11), respectively. Then, we consider the residual

$$R_h(\tau, v) := F_s(\tau, v) - A_s((\sigma_{0,h}, u_h), (\tau, v)) \,, \qquad \forall\, (\tau, v) \in \mathbf{H}. \tag{13}$$

Using the ellipticity of the bilinear form $A_s(\cdot, \cdot)$ and the definition of the residual (13), we deduce that

$$\|(\sigma_0 - \sigma_{0,h}, u - u_h)\|_{\mathbf{H}} \;\leq\; C_{\mathtt{ell}}^{-1} \sup_{\substack{(\tau,v)\in\mathbf{H} \\ (\tau,v)\neq(0,0)}} \frac{R_h(\tau, v)}{\|(\tau, v)\|_{\mathbf{H}}} \,. \tag{14}$$

Now, we remark that using the definitions of the linear functional F_s and the bilinear form $A_s(\cdot, \cdot)$, we can write

$$R_h(\tau, v) = R_1(\tau) + R_2(v) \,, \qquad \forall\, \tau \in H_0 \,, \quad \forall\, v \in H^1(\Omega) \,, \tag{15}$$

where $R_1 : H_0 \to \mathbb{R}$ and $R_2 : H^1(\Omega) \to \mathbb{R}$ are defined by

$$\begin{aligned} R_1(\tau) \;:=\; & -\int_\Omega (\mathcal{K}^{-1}(\sigma_z + \sigma_{0,h}) - \nabla u_h) \cdot \tau \;+\; \kappa_2 \int_\Omega (\mathcal{K}^{-1}(\sigma_z + \sigma_{0,h}) - \nabla u_h) \cdot \mathcal{K}^{-1}\tau \\ & -\; \kappa_1 \int_\Omega (f + \mathrm{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b} \cdot \nabla u_h)\, \mathrm{div}(\tau) \;+\; \int_{\Gamma_D} (g - u_h)\tau \cdot \mathbf{n} \end{aligned}$$

and

$$\begin{aligned} R_2(v) \;:=\; & \int_\Omega (f + \mathrm{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b} \cdot \nabla u_h)\, v \;-\; \kappa_1 \int_\Omega (f + \mathrm{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b} \cdot \nabla u_h)\, \mathbf{b} \cdot \nabla v \\ & +\; \kappa_2 \int_\Omega (\mathcal{K}^{-1}(\sigma_z + \sigma_{0,h}) - \nabla u_h) \cdot \nabla v \;+\; \kappa_3 \int_{\Gamma_D} (g - u_h) v \,. \end{aligned}$$

5.1 Notations and preliminary results

We let $\mathcal{T}_h$ be as in section 4 and assume that $\{\tau_h \in H_0 \,:\, (\tau_h)_{|T} \in (\mathcal{P}_0(T))^d \,, \forall\, T \in \mathcal{T}_h\} \subset H_h$ and that $X_h := \{v_h \in \mathcal{C}(\overline{\Omega}) \,:\, v_h|_T \in \mathcal{P}_1(T) \,, \quad \forall\, T \in \mathcal{T}_h\} \subset V_h$. Given an element $T \in \mathcal{T}_h$, we denote by $E(T)$ the set of the edges (if $d = 2$) or faces (if $d = 3$) of T and by E_h the set of all the edges ($d = 2$) / faces ($d = 3$) induced by the mesh $\mathcal{T}_h$. Then, we can write $E_h = E_i \cup E_{\Gamma_D} \cup E_{\Gamma_N}$,

where $E_i := \{e \in E_h \,:\, e \subset \Omega\}$, $E_{\Gamma_D} := \{e \in E_h \,:\, e \subset \overline{\Gamma}_D\}$ and $E_{\Gamma_N} := \{e \in E_h \,:\, e \subset \overline{\Gamma}_N\}$. Moreover, if $d = 2$, then for each edge $e \in E_h$, we fix a unit normal vector, $\mathbf{n}_e := (n_1, n_2)^{\mathtt{t}}$, and let $\mathbf{t}_e := (-n_2, n_1)^{\mathtt{t}}$ be the corresponding fixed unit tangential vector along e. If $d = 3$, for each face $e \in E_h$, we fix a unit normal vector $\mathbf{n}_e$ to e. Finally, given $v : \Omega \to \mathbb{R}$ and $\tau : \Omega \to \mathbb{R}^3$ sufficiently smooth, we denote $\mathrm{curl}(v) = \left(\frac{\partial v}{\partial x_2}, -\frac{\partial v}{\partial x_1}\right)^{\mathtt{t}}$ and $\mathbf{curl}(\tau) = \nabla \times \tau$.

We consider the Clément interpolation operator, $I_h : H^1(\Omega) \to X_h$. In the next lemma, we recall its local approximation properties.

Lemma 1 *There exist positive constants, c_1 and c_2, independent of h, such that for all $v \in H^1(\Omega)$ there holds*

$$\|v - I_h(v)\|_{H^m(T)} \leq c_1\, h_T^{1-m}\, \|v\|_{H^1(\omega(T))}, \quad \forall\, m \in \{0,1\}, \quad \forall\, T \in \mathcal{T}_h$$

$$\|v - I_h(v)\|_{L^2(e)} \leq c_2\, h_e^{1/2}\, \|v\|_{H^1(\omega(e))}, \quad \forall\, e \in E_h\,,$$

where $\omega(T) = \bigcup\{T' \in \mathcal{T}_h \,:\, T' \cap T \neq \emptyset\}$ and $\omega(e) = \bigcup\{T' \in \mathcal{T}_h \,:\, T' \cap e \neq \emptyset\}$.

Proof. See [12]. □

We also consider the Raviart-Thomas interpolation operator, $\Pi_h^k : [H^1(\Omega)]^d \to H_h$. We recall that given $\tau \in [H^1(\Omega)]^d$, $\Pi_h^k(\tau)$ is characterized by the following identities:

$$\int_e \Pi_h^k(\tau) \cdot \mathbf{n}_e\, q = \int_e \tau \cdot \mathbf{n}_e\, q\,, \quad \forall e \in E_h, \quad \forall q \in \mathcal{P}_k(e), \quad \forall\, k \geq 0 \tag{16}$$

$$\int_T \Pi_h^k(\tau) \cdot \rho = \int_T \tau \cdot \rho, \quad \forall \rho \in [\mathcal{P}_{k-1}(T)]^d, \quad \forall\, k \geq 1\,. \tag{17}$$

In the next lemma, we recall the approximation properties of the Raviart-Thomas interpolation operator.

Lemma 2 *There exist positive constants, c_3, c_4 and c_5, independent of h, such that*

$$\|\tau - \Pi_h^k(\tau)\|_{[L^2(T)]^d} \leq c_3\, h_T^m\, |\tau|_{[H^m(T)]^d}\,, \quad \forall\, T \in \mathcal{T}_h\,, \quad 1 \leq m \leq k+1\,,$$

and for all $\tau \in [H^1(\Omega)]^d$ with $\mathrm{div}(\tau) \in H^m(\Omega)$,

$$\|\mathrm{div}(\tau) - \mathrm{div}(\Pi_h^k(\tau))\|_{L^2(T)} \leq c_4 h_T^m |\mathrm{div}(\tau)|_{H^m(T)}\,, \quad 0 \leq m \leq k+1\,,$$

$$\|\tau \cdot \mathbf{n} - \Pi_h^k(\tau) \cdot \mathbf{n}\|_{L^2(e)} \leq c_5\, h_e^{1/2} \|\tau\|_{[H^1(T_e)]^d}, \quad \forall e \in E_h \quad \forall \tau \in [H^1(\Omega)]^d\,,$$

where T_e contains e on its boundary.

Proof. See [36]. □

Using (16) and (17) it is easy to show that

$$\mathrm{div}(\Pi_h^k(\tau)) = P_h^k \mathrm{div}(\tau) \tag{18}$$

where $P_h^k : L^2(\Omega) \to V_h$ is the L^2-orthogonal projector. It is well-known that $\forall\, v \in H^m(\Omega)$, $0 \leq m \leq k+1$, there holds

$$\|v - P_h^k v\|_{L^2(T)} \leq c_6\, h_T^m\, |v|_{H^m(T)}\,, \quad \forall\, T \in \mathcal{T}_h\,. \tag{19}$$

5.2 Two-dimensional case

In what follows, we assume that $d = 2$ and let $(\tau, v) \in H_0 \times H^1(\Omega)$. Then, there exists $\chi \in H^1(\Omega)$ and $z \in H^2(\Omega)$ such that

$$\tau = \mathrm{curl}(\chi) + \nabla z$$

and

$$\|\chi\|_{H^1(\Omega)} + \|z\|_{H^2(\Omega)} \leq C \, \|\tau\|_{H(\mathrm{div};\Omega)} \,, \tag{20}$$

for some positive constant C. In particular, $\mathrm{div}(\tau) = \Delta z$ in Ω.

Now, from the Galerkin orthogonality, we have that

$$R_h(\tau, v) = R_h(\tau - \tau_h, v - v_h) \,, \quad \forall \, \tau_h \in H_h \,, \quad \forall \, v_h \in V_h \,.$$

Let us consider a *discrete* Helmholtz decomposition of τ_h: we define $\chi_h = I_h(\chi)$ and

$$\tau_h = \mathrm{curl}(\chi_h) + \Pi_h^k \nabla z \in H_h \,.$$

Then,

$$\tau - \tau_h = \mathrm{curl}(\chi - \chi_h) + (id - \Pi_h^k)\nabla z \,,$$

and, from (18),

$$\mathrm{div}(\tau - \tau_h) = \mathrm{div}((id - \Pi_h^k)\nabla z) = (id - P_h^k)\Delta z = (id - P_h^k)\mathrm{div}(\tau) \,.$$

Therefore, we can decompose

$$R_1(\tau) = R_1(\tau - \tau_h) = \tilde{R}_1(\tau) + \hat{R}_1(z) + \bar{R}_1(\chi) \,, \tag{21}$$

where

$$\tilde{R}_1(\tau) \;=\; -\kappa_1 \int_\Omega (f + \mathrm{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b} \cdot \nabla u_h)(id - P_h^k)\mathrm{div}(\tau) \,,$$

$$\begin{aligned} \hat{R}_1(z) \;:=\; & -\int_\Omega (\mathcal{K}^{-1}(\sigma_z + \sigma_{0,h}) - \nabla u_h) \cdot (id - \Pi_h^k)\nabla z \\ & + \kappa_2 \int_\Omega (\mathcal{K}^{-1}(\sigma_z + \sigma_{0,h}) - \nabla u_h) \cdot \mathcal{K}^{-1}(id - \Pi_h^k)\nabla z + \int_{\Gamma_D} (g - u_h)(id - \Pi_h^k)\nabla z \cdot \mathbf{n} \,, \end{aligned}$$

and

$$\begin{aligned} \bar{R}_1(\chi) \;:=\; & -\int_\Omega (\mathcal{K}^{-1}(\sigma_z + \sigma_{0,h}) - \nabla u_h) \cdot \mathrm{curl}(\chi - \chi_h) \\ & + \kappa_2 \int_\Omega (\mathcal{K}^{-1}(\sigma_z + \sigma_{0,h}) - \nabla u_h) \cdot \mathcal{K}^{-1}\mathrm{curl}(\chi - \chi_h) + \int_{\Gamma_D} (g - u_h)\mathrm{curl}(\chi - \chi_h) \cdot \mathbf{n} \,. \end{aligned}$$

Our aim is to obtain upper bounds for each one of these three terms.

Lemma 3 *There holds*

$$|\tilde{R}_1(\tau)| \leq \kappa_1 \Big(\sum_{T\in\mathcal{T}_h} \|f + \mathrm{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b}\cdot\nabla u_h\|^2_{L^2(T)} \Big)^{1/2} \|\mathrm{div}(\tau)\|_{L^2(\Omega)} .$$

Proof. The proof follows by decomposing $\displaystyle\int_\Omega = \sum_{T\in\mathcal{T}_h} \int_T$, using the Cauchy-Schwarz inequality and (19). □

Lemma 4 *There exists a positive constant C, independent of h, such that*

$$|\hat{R}_1(z)| \leq C \Big(\sum_{T\in\mathcal{T}_h} h_T^2 \|\nabla u_h - \mathcal{K}^{-1}(\sigma_z + \sigma_{0,h})\|^2_{[L^2(T)]^2} + \sum_{e\in E_{\Gamma_D}} h_e \|g - u_h\|^2_{L^2(e)} \Big)^{1/2} \|\tau\|_{H(\mathrm{div};\Omega)} .$$

Proof. On the one hand, using the triangle inequality and the Cauchy-Schwarz inequality, we have

$$\Big|\int_\Omega (\nabla u_h - \mathcal{K}^{-1}(\sigma_z+\sigma_{0,h}))(id - \Pi_h^k)\nabla z\Big| \leq \sum_{T\in\mathcal{T}_h} \|\nabla u_h - \mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})\|_{[L^2(T)]^2} \|(id-\Pi_h^k)\nabla z\|_{[L^2(T)]^2}$$

Now, by Lemma 2,

$$\|(id - \Pi_h^k)\nabla z\|_{[L^2(T)]^2} \leq c_3\, h_T\, |\nabla z|_{[H^1(T)]^2} .$$

Then, using again the Cauchy-Schwarz inequality, we obtain

$$\Big|\int_\Omega (\nabla u_h - \mathcal{K}^{-1}(\sigma_z+\sigma_{0,h}))(id - \Pi_h^k)\nabla z\Big| \leq C \Big(\sum_{T\in\mathcal{T}_h} h_T^2 \|\nabla u_h - \mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})\|^2_{[L^2(T)]^2}\Big)^{1/2} \|\tau\|_{H(\mathrm{div};\Omega)}$$

where we used that $\Big(\displaystyle\sum_{T\in\mathcal{T}_h} |\nabla z|^2_{[H^1(T)]^2}\Big)^{1/2} \leq \|z\|_{H^2(\Omega)} \leq C\|\tau\|_{H(\mathrm{div};\Omega)}$.

Analogously, using the continuity of $\mathcal{K}^{-1}$, we have that

$$\Big|\kappa_2 \int_\Omega (\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h}) - \nabla u_h)\cdot \mathcal{K}^{-1}(id - \Pi_h^k)\nabla z\Big| \leq C \Big(\sum_{T\in\mathcal{T}_h} h_T^2 \|\nabla u_h - \mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})\|^2_{[L^2(T)]^2}\Big)^{1/2} \|\tau\|_{H(\mathrm{div};\Omega)} .$$

On the other hand, using the triangle inequality and the Cauchy-Schwarz inequality,

$$\Big|\int_{\Gamma_D} (g - u_h)(id - \Pi_h^k)\nabla z\cdot\mathbf{n}\Big| \leq \sum_{e\in E_{\Gamma_D}} \|g - u_h\|_{L^2(e)} \|(id - \Pi_h^k)\nabla z\cdot\mathbf{n}\|_{L^2(e)} .$$

Since $\nabla z \in [H^1(\Omega)]^2$, $\nabla z\cdot\mathbf{n} \in [H^{1/2}(\Gamma)]^2$. Then, by Lemma 2,

$$\|(id - \Pi_h^k)\nabla z\cdot\mathbf{n}\|_{L^2(e)} \leq c_5\, h_e^{1/2} \|\nabla z\|_{[H^1(T_e)]^2} .$$

Therefore,

$$|\int_{\Gamma_D}(g-u_h)(id-\Pi_h^k)\nabla z\cdot\mathbf{n}|\ \leq\ C\,(\sum_{e\in E_{\Gamma_D}} h_e\,\|g-u_h\|^2_{L^2(e)})^{1/2}\|\tau\|_{H(\mathrm{div};\Omega)}\,,$$

where we used that $(\sum_{e\in E_{\Gamma_D}}\|\nabla z\|^2_{[H^1(T_e)]^2})^{1/2}\ \leq\ \|z\|_{H^2(\Omega)}\ \leq\ C\,\|\tau\|_{H(\mathrm{div};\Omega)}$.

The proof follows by using the triangle inequality. □

Lemma 5 *Assume that $g\in H^1(\Gamma_D)$. Then, there exists $C>0$, independent of h, such that*

$$|\bar{R}_1(\chi)|\leq C\Big(\sum_{T\in\mathcal{T}_h}\|\nabla u_h-\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})\|^2_{[L^2(T)]^2}\ +\ \sum_{e\in E_{\Gamma_D}} h_e\,\|\frac{\partial}{\partial\mathbf{t}}(g-u_h)\|^2_{L^2(e)}\Big)^{1/2}\|\tau\|_{H(\mathrm{div};\Omega)}\,.$$

Proof. Proceeding similarly as in the proof of the previous Lemma and noting that, by virtue of Lemma 1,

$$\|\mathrm{curl}(\chi-\chi_h)\|_{[L^2(T)]^2}=\|\nabla(\chi-\chi_h)\|_{[L^2(T)]^2}\ \leq\ \|\chi-\chi_h\|_{H^1(T)}\ \leq\ c_1\,\|\chi\|_{H^1(\omega(T))},$$

we have

$$\begin{aligned}|-\int_\Omega(\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})-\nabla u_h)\ \cdot\ \mathrm{curl}(\chi-\chi_h)\ +\ \kappa_2\int_\Omega(\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})-\nabla u_h)\ \cdot\ \mathcal{K}^{-1}\mathrm{curl}(\chi-\chi_h)|\\ \leq\ C\sum_{T\in\mathcal{T}_h}\|\nabla u_h-\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})\|_{[L^2(T)]^2}\|\chi\|_{H^1(\omega(T))}\,.\end{aligned}$$

Now, since the number of triangles in $\omega(T)$ is bounded and, from (20), $\|\chi\|_{H^1(\Omega)}\ \leq\ C\,\|\tau\|_{H(\mathrm{div};\Omega)}$, it follows that

$$\begin{aligned}|-\int_\Omega(\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})-\nabla u_h)\ \cdot\ \mathrm{curl}(\chi-\chi_h)\ +\ \kappa_2\int_\Omega(\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})-\nabla u_h)\ \cdot\ \mathcal{K}^{-1}\mathrm{curl}(\chi-\chi_h)|\\ \leq\ C\,(\sum_{T\in\mathcal{T}_h}\|\nabla u_h-\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})\|^2_{[L^2(T)]^2})^{1/2}\|\tau\|_{H(\mathrm{div};\Omega)}\,.\end{aligned}\tag{22}$$

On the other hand, we remark that $\mathrm{curl}(\chi-\chi_h)\cdot\mathbf{n}\ =\ \frac{\partial}{\partial\mathbf{t}}(\chi-\chi_h)$ so,

$$|\int_{\Gamma_D}(g-u_h)\mathrm{curl}(\chi-\chi_h)\cdot\mathbf{n}|\ \leq\ \sum_{e\in E_{\Gamma_D}}|\int_e\frac{\partial}{\partial\mathbf{t}}(g-u_h)(\chi-\chi_h)|$$

Now, using the Cauchy-Schwarz inequality, Lemma 1 and noting that the number of elements in $\omega(e)$ is bounded, we deduce that

$$|\int_{\Gamma_D}(g-u_h)\mathrm{curl}(\chi-\chi_h)\cdot\mathbf{n}|\ \leq\ C\,(\sum_{e\in E_{\Gamma_D}} h_e\|\frac{\partial}{\partial\mathbf{t}_e}(g-u_h)\|^2_{L^2(e)})^{1/2}\|\tau\|_{H(\mathrm{div};\Omega)}\,.\tag{23}$$

The proof follows using the triangle inequality, (22) and (23). □

On the other hand, we recall that

$$R_2(v) = R_2(v - v_h), \quad \forall\, v_h \in V_h\,.$$

Let us take $v_h = I_h(v) \in V_h$. Then,

$$\begin{aligned} R_2(v) &= \int_\Omega (f + \operatorname{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b}\cdot\nabla u_h)(v - I_h(v)) \\ &- \kappa_1 \int_\Omega (f + \operatorname{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b}\cdot\nabla u_h)\mathbf{b}\cdot\nabla(v - I_h(v)) \\ &+ \kappa_2 \int_\Omega (\mathcal{K}^{-1}(\sigma_z + \sigma_{0,h}) - \nabla u_h)\cdot\nabla(v - I_h(v)) + \kappa_3 \int_{\Gamma_D} (g - u_h)(v - I_h(v)) \end{aligned}$$

Now, using the Cauchy-Schwarz inequality and Lemma 1 in each one of the terms above, we have that

$$\begin{aligned} |R_2(v)| &\leq C\Big(\|f + \operatorname{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b}\cdot\nabla u_h\|_{L^2(\Omega)} \\ &+ \|\nabla u_h - \mathcal{K}^{-1}(\sigma_z + \sigma_{0,h})\|_{[L^2(\Omega)]^2} + \Big(\sum_{e\in E_{\Gamma_D}} h_e \|g - u_h\|^2_{L^2(e)}\Big)^{1/2}\Big)\|v\|_{H^1(\Omega)}\,, \end{aligned} \tag{24}$$

where we used that $(\sum_{T\in\mathcal{T}_h} \|v\|^2_{H^1(\omega(T))})^{1/2} \leq C\,\|v\|_{H^1(\Omega)}$ and $(\sum_{e\in E_{\Gamma_D}} \|v\|^2_{H^1(\omega(e))})^{1/2} \leq C\,\|v\|_{H^1(\Omega)}$.

From the previous results, we have the following upper bound for the residual.

Proposition 1 *There exists a positive constant C, independent of h, such that*

$$\begin{aligned} \sup_{\substack{(\tau,v)\in\mathbf{H}\\(\tau,v)\neq(0,0)}} \frac{R_h(\tau,v)}{\|(\tau,v)\|_{\mathbf{H}}} &\leq C\Big(||f + \operatorname{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b}\cdot\nabla u_h||_{L^2(\Omega)} + ||\nabla u_h - \mathcal{K}^{-1}(\sigma_z + \sigma_{0,h})||_{[L^2(\Omega)]^2} \\ &+ \Big(\sum_{e\in E_{\Gamma_D}} h_e\,(\|g - u_h\|^2_{L^2(e)} + \|\frac{\partial}{\partial \mathbf{t}_e}(g - u_h)\|^2_{L^2(e)})\Big)^{1/2}\Big). \end{aligned}$$

Proof. It follows from (15), (21), the triangle inequality, Lemmas 3, 4 and 5, and (24). □

From (14) and Proposition 1, we have that

$$\begin{aligned} \|(\sigma_0 - \sigma_{0,h}, u - u_h)\|_{\mathbf{H}} &\leq C_{\mathtt{ell}}^{-1}\, C\Big(||f + \operatorname{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b}\cdot\nabla u_h||_{L^2(\Omega)} \\ &+ ||\nabla u_h - \mathcal{K}^{-1}(\sigma_z + \sigma_{0,h})||_{[L^2(\Omega)]^2} \\ &+ \Big(\sum_{e\in E_{\Gamma_D}} h_e\,(\|g - u_h\|^2_{L^2(e)} + \|\frac{\partial}{\partial \mathbf{t}_e}(g - u_h)\|^2_{L^2(e)})\Big)^{1/2}\Big). \end{aligned} \tag{25}$$

Motivated by this result, we define the global a posteriori error indicator

$$\theta := \Big(\sum_{T \in \mathcal{T}_h} \theta_T^2 \Big)^{1/2}, \tag{26}$$

where

$$\begin{aligned} \theta_T^2 &:= ||f + \mathrm{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b} \cdot \nabla u_h||^2_{L^2(T)} + ||\nabla u_h - \mathcal{K}^{-1}(\sigma_z + \sigma_{0,h})||^2_{[L^2(T)]^2} \\ &+ \sum_{e \in E_{\Gamma_D} \cap \partial T} h_e \left(\|g - u_h\|^2_{L^2(e)} + \|\frac{\partial}{\partial \mathbf{t}_e}(g - u_h)\|^2_{L^2(e)} \right). \end{aligned} \tag{27}$$

In the next theorem we establish the reliability of the error indicator θ.

Theorem 3 *Let $(\sigma_0, u) \in \mathbf{H}$ and $(\sigma_{0,h}, u_h) \in \mathbf{H}_h$ be the unique solutions to problems (9) and (11), respectively. Then, there exists a positive constant $C_{\mathtt{rel}}$, independent of h, such that*

$$\|(\sigma_0 - \sigma_{0,h}, u - u_h)\|_{\mathbf{H}} \leq C_{\mathtt{rel}}\, \theta\,.$$

Proof. It follows from (25) and the definition of θ. □

In order to prove the efficiency of θ, we first recall that $f = -\mathrm{div}(\sigma_0 + \sigma_z) + \mathbf{b} \cdot \nabla u$ and $\mathcal{K}^{-1}(\sigma_0 + \sigma_z) - \nabla u = \mathbf{0}$ in Ω. Let $T \in \mathcal{T}_h$. Then, using the triangle inequality

$$\begin{aligned} \|f + \mathrm{div}(\sigma_z + \sigma_{0,h}) - \mathbf{b} \cdot \nabla u_h\|^2_{L^2(T)} &= \|\mathrm{div}(\sigma_{0,h} - \sigma_0) - \mathbf{b} \cdot \nabla(u_h - u)\|^2_{L^2(T)} \\ &\leq 2\big(\|\mathrm{div}(\sigma_{0,h} - \sigma_0)\|^2_{L^2(T)} + \|\mathbf{b} \cdot \nabla(u_h - u)\|^2_{L^2(T)}\big) \\ &\leq 2\big(\|\mathrm{div}(\sigma_{0,h} - \sigma_0)\|^2_{L^2(T)} + 2\,\|\mathbf{b}\|^2_{[L^\infty(T)]^2} \|\nabla(u_h - u)\|^2_{[L^2(T)]^d}\big) \end{aligned} \tag{28}$$

and

$$\|\nabla u_h - \mathcal{K}^{-1}(\sigma_z + \sigma_{0,h})\|^2_{[L^2(T)]^2} \leq 2\left(\|\nabla(u_h - u)\|^2_{[L^2(T)]^2} + \|\mathcal{K}^{-1}\|^2_{\infty,T}\, \|\sigma_{0,h} - \sigma_0\|^2_{[L^2(T)]^2}\right). \tag{29}$$

Now, in order to bound the boundary terms on the right hand side of (27), we recall a discrete trace inequality from [2, Theorem 3.10] (see also [4]): there exists $c > 0$, depending only on the shape regularity of the triangulations, such that for each $T \in \mathcal{T}_h$ and $e \in E(T)$, there holds

$$\|v\|^2_{L^2(e)} \leq c_7 \left(h_e^{-1} \|v\|^2_{L^2(T)} + h_e |v|^2_{H^1(T)} \right), \quad \forall\, v \in H^1(T). \tag{30}$$

Lemma 6 *There exists $c_7 > 0$, independent of h, such that for each $e \in E_{\Gamma_D}$ there holds*

$$h_e \|g - u_h\|^2_{L^2(e)} \leq c_7 \left(\|u - u_h\|^2_{L^2(T_e)} + h^2_{T_e} |u - u_h|^2_{H^1(T_e)} \right),$$

where T_e is the triangle having e as an edge.

Proof. It follows straightforwardly by taking into account that $g = u$ on Γ_D and applying the discrete trace inequality (30). □

Now, in order to bound the last boundary term in (27), we use the localization technique [39]. Given $T \in \mathcal{T}_h$ and $e \in E(T)$, we let ψ_e be the usual edge-bubble function (see equation (1.6) in [39]), which satisfies $\psi_e|_T \in \mathcal{P}_2(T)$, $\mathrm{supp}(\psi_e) \subset \omega_e := \bigcup\{T' \in \mathcal{T}_h \,:\, e \in E(T')\}$, $\psi_e = 0$ on $\partial T \setminus e$, and $0 \leq \psi_e \leq 1$ in ω_e.

We also recall from [38] that, given a nonnegative integer k, there exists a linear operator $L : \mathcal{C}(e) \to \mathcal{C}(T)$, $T \in \omega_e$, that satisfies $L(p) \in \mathcal{P}_k(T)$ and $L(p)|_e = p$, $\forall\, p \in \mathcal{P}_k(e)$.

Lemma 7 *Given $k \in \mathbb{N}$, there exists positive constants c_8 and c_9, that only depend on k and the shape regularity of the triangulations (minimum angle condition), such that for each $T \in \mathcal{T}_h$ and $e \in E(T)$, there holds*

$$\|p\|_{L^2(e)} \,\leq\, c_8\, \|\psi_e^{1/2} p\|_{L^2(e)}\,, \qquad \forall\, p \in \mathcal{P}_k(e)\,, \tag{31}$$

$$\|\psi_e^{1/2} L(p)\|_{L^2(T)} \,\leq\, c_9\, h_e^{1/2}\, \|p\|_{L^2(e)}\,, \qquad \forall\, p \in \mathcal{P}_k(e)\,. \tag{32}$$

Proof. See Lemma 4.1 in [38]. □

We will also use the following inverse inequality (see Theorem 3.2.6 in [11]): for each $T \in \mathcal{T}_h$, there holds

$$|q|_{H^1(T)} \,\leq\, c_{10}\, h_T^{-1}\, ||q||_{L^2(T)}\,, \quad \forall\, q \in \mathcal{P}_k(T)\,, \tag{33}$$

where $c_{10} > 0$ depends only on k and the shape regularity of the triangulations.

Lemma 8 *Assume that $g \in H^1(\Gamma_D)$ is a piecewise polynomial on Γ_D. Then, there exists $C > 0$, independent of h, such that for each $e \in E_{\Gamma_D}$ there holds*

$$h_e\, \|\frac{\partial}{\partial \mathbf{t}_e}(g - u_h)\|^2_{L^2(e)} \,\leq\, C\, |u - u_h|^2_{H^1(T_e)}\,,$$

where T_e is the triangle having e as an edge.

Proof. Let $e \in E_{\Gamma_D}$, and define $\chi_e := \dfrac{\partial}{\partial \mathbf{t}_e}(g - u_h)$ on e, which is a polynomial on e. Then, thanks to (31) and the extension operator $L : \mathcal{C}(e) \to \mathcal{C}(T_e)$, we have that

$$\|\chi_e\|^2_{L^2(e)} \,\leq\, c_8^2\, \|\psi_e^{1/2}\, \chi_e\|^2_{L^2(e)} \,=\, c_8^2 \int_e \psi_e \chi_e \big(\frac{\partial}{\partial \mathbf{t}_e}(g - u_h)\big) \,=\, c_8^2 \int_{\partial T_e} \psi_e L(\chi_e) \nabla(u - u_h) \cdot \mathbf{t}\,.$$

Now, integrating by parts, we find that

$$\int_{\partial T_e} \psi_e L(\chi_e) \nabla(u - u_h) \cdot \mathbf{t} \,=\, -\int_{T_e} \mathrm{curl}(\psi_e L(\chi_e)) \cdot \nabla(u - u_h)\,.$$

Therefore, applying the Cauchy-Schwarz inequality, the inverse inequality (33), that $h_e \leq h_{T_e}$ and $0 \leq \psi_e \leq 1$, and (32) we have

$$\begin{aligned}
\|\chi_e\|^2_{L^2(e)} &\leq c_8^2 \, \|\mathrm{curl}(\psi_e L(\chi_e))\|_{[L^2(T_e)]^2} \|\nabla(u-u_h)\|_{[L^2(T_e)]^2} \\
&= c_8^2 \, |\psi_e L(\chi_e)|_{H^1(T_e)} |u-u_h|_{H^1(T_e)} \\
&\leq c_8^2 c_{10} h_e^{-1} \|\psi_e L(\chi_e)\|_{L^2(T_e)} |u-u_h|_{H^1(T_e)} \\
&\leq c_8^2 c_{10} h_e^{-1} \|\psi_e^{1/2} L(\chi_e)\|_{L^2(T_e)} |u-u_h|_{H^1(T_e)} \\
&\leq c_8^2 c_{10} c_9 h_e^{-1} h_e^{1/2} \|\chi_e\|_{L^2(e)} |u-u_h|_{H^1(T_e)}
\end{aligned}$$

and we end the proof. □

In summary we have proved the next theorem, which establishes the local efficiency of the a posteriori error indicator θ.

Theorem 4 *Assume $g \in H^1(\Gamma_D)$ is a piecewise polynomial on Γ_D. Then, there exists $C_{\mathtt{eff}} > 0$, independent of h, such that for all $T \in \mathcal{T}_h$ we have*

$$\theta_T^2 \leq C_{\mathtt{eff}}^{-1} \left(\|u-u_h\|^2_{H^1(T)} + \|\sigma_0 - \sigma_{0,h}\|^2_{H(div,T)} \right).$$

Proof. It follows from (26), (28), (29) and Lemmas 6 and 8. □

5.3 Three-dimensional case

In what follows, we let $d = 3$ and consider $(\tau, v) \in H_0 \times H^1(\Omega)$. Then, there exists $\chi \in [H^1(\Omega)]^3$ and $z \in H^2(\Omega)$ such that

$$\tau = \mathbf{curl}(\chi) + \nabla z$$

and

$$\|\chi\|_{[H^1(\Omega)]^3} + \|z\|_{H^2(\Omega)} \leq C \, \|\tau\|_{H(\mathrm{div};\Omega)},$$

for some positive constant C independent of τ (see Theorem 3.1 in [22]). In particular, $\mathrm{div}(\tau) = \Delta z$ in Ω.

We consider a *discrete* Helmholtz decomposition of τ_h: we define $\chi_h = \mathbf{I}_h(\chi)$, where $\mathbf{I}_h : [H^1(\Omega)]^3 \to X_h^3$ is the vector counterpart of the Clément interpolation operator, and

$$\tau_h = \mathbf{curl}(\chi_h) + \Pi_h^k \nabla z \in H_h.$$

Then, proceeding similarly as in the previous section, we have the decomposition

$$R_1(\tau) = R_1(\tau - \tau_h) = \tilde{R}_1(\tau) + \hat{R}_1(z) + \bar{R}_1(\chi), \tag{34}$$

where $\tilde{R}_1(\tau)$ and $\hat{R}_1(z)$ are formally as in the previous section and

$$\begin{aligned}\bar{R}_1(\chi) &:= -\int_\Omega (\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})-\nabla u_h)\cdot \mathrm{curl}(\chi-\chi_h) \\ &+ \kappa_2\int_\Omega (\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})-\nabla u_h)\cdot \mathcal{K}^{-1}\mathrm{curl}(\chi-\chi_h) + \int_{\Gamma_D}(g-u_h)\mathbf{curl}(\chi-\chi_h)\cdot\mathbf{n}\,.\end{aligned}$$

The analogous of Lemmas 3 and 4 follows straightforwardly. The analogous of Lemma 5 follows by noting that

$$\int_{\Gamma_D}(g-u_h)\mathbf{curl}(\chi-\chi_h)\cdot\mathbf{n} = \sum_{e\in\Gamma_D}\int_e(\chi-\chi_h)\cdot[\nabla(g-u_h)\times\mathbf{n}_e]\,.$$

Then, we define the global a posteriori error indicator

$$\eta := \Big(\sum_{T\in\mathcal{T}_h}\eta_T^2\Big)^{1/2}, \tag{35}$$

where

$$\begin{aligned}\eta_T^2 &:= ||f+\mathrm{div}(\sigma_z+\sigma_{0,h})-\mathbf{b}\cdot\nabla u_h||^2_{L^2(T)} + ||\nabla u_h-\mathcal{K}^{-1}(\sigma_z+\sigma_{0,h})||^2_{[L^2(T)]^2} \\ &+ \sum_{e\in E_{\Gamma_D}\cap\partial T} h_e\left(\|g-u_h\|^2_{L^2(e)}+\|\nabla(g-u_h)\times\mathbf{n}\|^2_{[L^2(e)]^3}\right).\end{aligned} \tag{36}$$

Following the same arguments as in the previous section, we have the analogous of Theorem 3.

In order to prove the efficiency of η, we proceed with the first three terms of (36) as before, and obtain the analogous to (28), (29) and Lemma 6. Finally, to bound the fourth term, we proceed as in Lemma 8.

6 Numerical experiments

In this section, we show some numerical experiments that illustrate the performance of the augmented scheme (11) and confirm the properties of the a posteriori error estimator θ defined in (26)-(27). The numerical experiments were performed with the finite element toolbox FENICS [32] using the Plaza and Carey refinement algorithm [35]. We present numerical results for the finite element pairs (H_h, V_h) given by $(\mathcal{RT}_0, \mathcal{L}_1)$, $(\mathcal{RT}_1, \mathcal{L}_2)$, $(\mathcal{BDM}_1, \mathcal{L}_1)$ and $(\mathcal{BDM}_2, \mathcal{L}_2)$ in $\mathbb{R}^2$.

We use the standard adaptive finite element method (AFEM) based on the loop:

$$\mathsf{SOLVE} \rightarrow \mathsf{ESTIMATE} \rightarrow \mathsf{MARK} \rightarrow \mathsf{REFINE}.$$

Hereafter, we replace the subscript h by k, where k is the counter of the adaptive loop. Then, given a mesh $\mathcal{T}_k$, the procedure SOLVE is a direct solver for computing the discrete solution (σ_k, u_k). ESTIMATE calculates the error indicators $\theta_k(T)$ for all $T \in \mathcal{T}_k$ depending on the computed solution and the data. Based on the values of $\{\theta_k(T)\}_{T\in\mathcal{T}_k}$, the procedure MARK generates a set of marked elements subject to refinement. For the elements selection, we rely on the *maximum* strategy: Given a threshold $\gamma \in (0, 1]$, any element $T' \in \mathcal{T}_k$ such that

$$\theta_k(T') > \gamma \max_{T\in\mathcal{T}_k} \theta_k(T), \tag{37}$$

is marked for refinement (in our experiments, we took $\gamma = 0.4$). Finally, the procedure REFINE creates a conforming refinement $\mathcal{T}_{k+1}$ of $\mathcal{T}_k$, using the Plaza and Carey refinement algorithm.

We will compare the performance of a finite element method based on uniform refinement with the adaptive method that we have described above.

In what follows, DOFs stands for the total number of degrees of freedom (unknowns) of (11) and we define the individual errors

$$e_k(\sigma_0) := \|\sigma_0 - \sigma_{0,k}\|_{H(\mathrm{div};\Omega)}\,, \quad e_k(u) := \|u - u_k\|_{H^1(\Omega)}\,,$$

and the total error

$$e_k(\sigma_0, u) := \big(e_k(\sigma_0)^2 + e_k(u)^2\big)^{1/2}.$$

The efficiency index with respect to the error estimator θ_k is defined as $\mathrm{eff}_k := \theta_k / e_k(\sigma_0, u)$.

We present five examples. The aim of the first one is to test the robustness of the discrete scheme (11) with respect to the stabilization parameters and to confirm the rate of convergence predicted by Theorem 2. In the second example, the solution presents an exponential boundary layer whereas in the third example the solution has two boundary layers. The fourth example has a singular solution and the fifth example explores the case of mixed boundary conditions. These four latter examples allow us to test the efficiency of the adaptive method.

6.1 A smooth example

Let $\Omega = (0, 2\,\pi) \times (0, 2\,\pi)$, $\Gamma_D = \Gamma$ and $\Gamma_N = \emptyset$. We consider a problem with anisotropic variable diffusion, with

$$\mathcal{K}(x, y) = \begin{pmatrix} 3 + \cos x & 0 \\ 0 & 3 + \cos y \end{pmatrix}, \quad \mathbf{b}(x, y) = (\cos x \sin y, -\sin x \cos y)^{\mathtt{t}}\,,$$

and choose f and g such that the exact solution of problem (5) is $u(x, y) = \cos(x + y)$.

We solve this problem with the finite element pairs $(\mathcal{RT}_0, \mathcal{L}_1)$, $(\mathcal{BDM}_1, \mathcal{L}_1)$, $(\mathcal{RT}_1, \mathcal{L}_2)$ and $(\mathcal{BDM}_2, \mathcal{L}_2)$ on a sequence of uniform meshes. In Figure 1 we show the decay of total errors

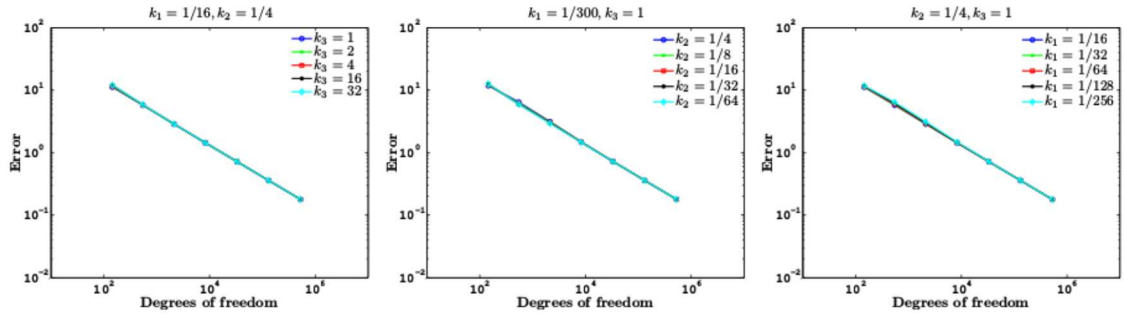

Figure 1: Example 1: Decay of total errors vs. number of degrees of freedom in the case of $(\mathcal{RT}_0, \mathcal{L}_1)$ and different values of κ_1, κ_2 and κ_3.

versus DOFs for the $(\mathcal{RT}_0, \mathcal{L}_1)$ finite element pair and different values of the stabilization parameters. We observe that optimal rates of convergence are attained in all cases and the results are independent of the values of the stabilization parameters, which shows the robustness of the method with respect to κ_1, κ_2 and κ_3.

Then, we choose

$$\kappa_1 = \frac{1}{16}\,, \qquad \kappa_2 = \frac{1}{4}\,, \qquad \kappa_3 = 1\,,$$

that satisfy conditions (10). In Figure 2 we show the decay of the error and estimator for both uniform and adaptive mesh refinement for the different finite elements. Convergence rates predicted by the theory for the total error (Theorem 2) are attained in all cases (we recall that $h^k \sim \text{DOFs}^{-k/d}$ on uniform meshes). Note that, since the exact solution is smooth, the error is uniformly distributed. Thus, when using the adaptive procedure we did not expect an improvement over the uniform one. Moreover, the errors $e_k(\sigma_0)$ and $e_k(u)$ decay with the same velocity as the total error (see Figure 3). In all cases the total error and the estimator have the same decay, which confirms that they are equivalent. In Figure 4 we plot the efficiency index for the different elements employed. We observe that efficiency indices are fairly stable.

Finally, the concentration and the flux obtained in the last iteration of the adaptive refinement algorithm when using $(\mathcal{RT}_0, \mathcal{L}_1)$ are plotted in Figure 5.

6.2 Sharp boundary layer

Now, we let $\Omega = (-1,1)\times(-1,1)$, $\Gamma_D = \Gamma$ and $\Gamma_N = \emptyset$. Let $\mathcal{K} = \epsilon\,\mathcal{I}$, with $\epsilon = 10^{-2}, 10^{-3}, 10^{-4}$, $\mathbf{b}(x,y) = (0,1)^{\mathrm{t}}$, $f = 0$, and

$$g(x,y) = x\,\frac{1 - e^{\frac{y-1}{\epsilon}}}{1 - e^{\frac{-2}{\epsilon}}}\,.$$

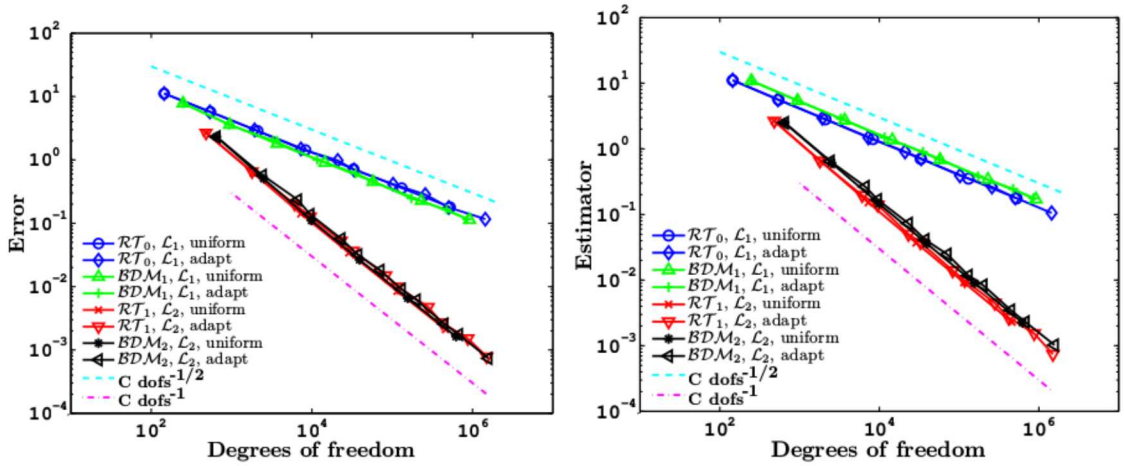

Figure 2: Example 1: Decay of total errors (left) and estimators (right) vs. number of degrees of freedom for different elements.

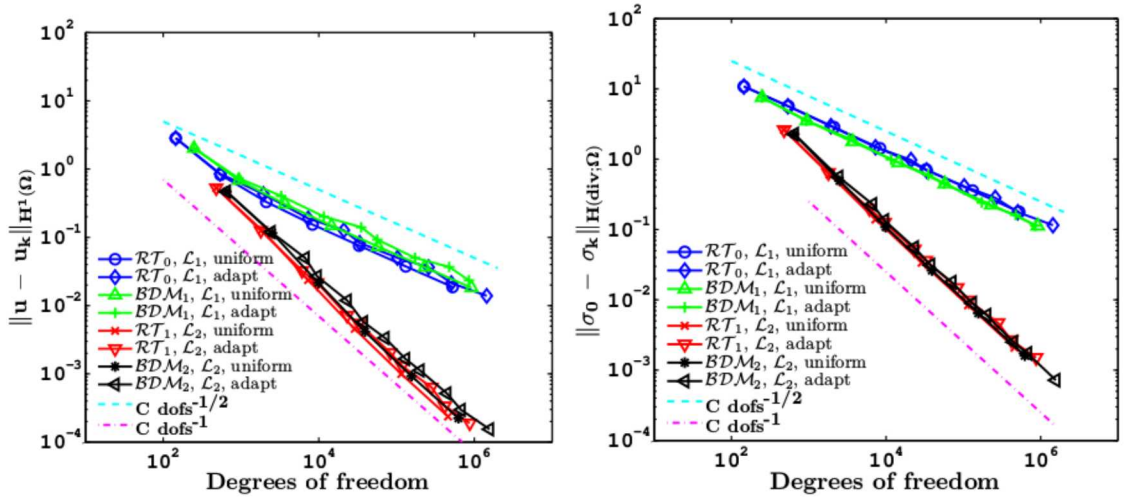

Figure 3: Example 1: Decay of $e_k(u)$ (left) and $e_k(\sigma)$ (right) vs. number of degrees of freedom for different elements.

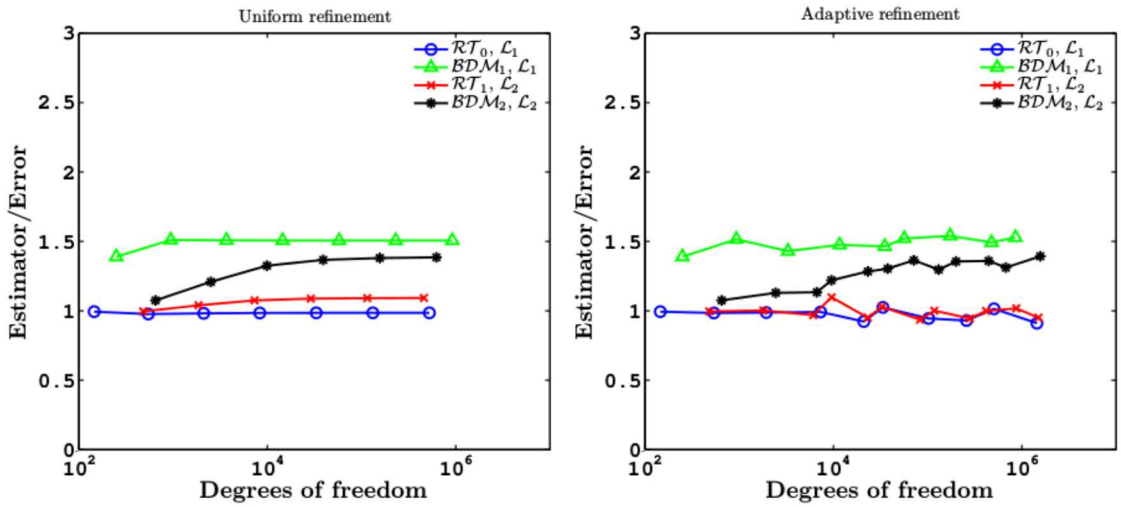

Figure 4: Example 1: Efficiency indices for different finite elements.

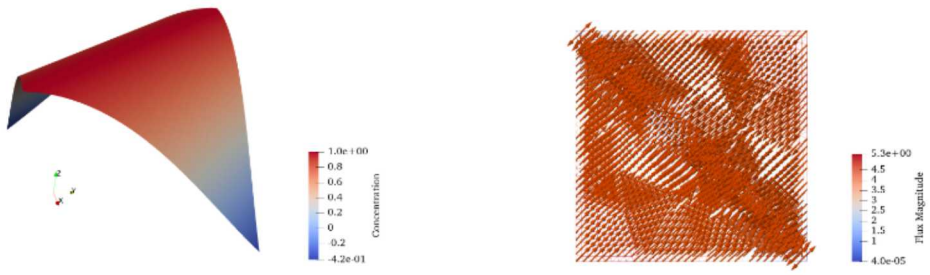

Figure 5: Example 1: Final concentration (left) and final flux (right) when using $(\mathcal{RT}_0, \mathcal{L}_1)$.

We apply the proposed augmented formulation to a boundary value problem inspired from [18], whose solution is given by $u(x,y) = x\,\frac{1-e^{\frac{y-1}{\epsilon}}}{1-e^{\frac{-2}{\epsilon}}}$. We remark that this solution has a sharp boundary layer at the boundary $y = 1$.

We performed experiments for the finite elements pairs $(\mathcal{RT}_0, \mathcal{L}_1)$, $(\mathcal{BDM}_1, \mathcal{L}_1)$, $(\mathcal{RT}_1, \mathcal{L}_2)$ and $(\mathcal{BDM}_2, \mathcal{L}_2)$. The results displayed in Figures 6 and 7 show that the individual and total errors, as well as the estimators proposed in this work decrease following the convergence rates predicted by the theory for the adaptive refinement algorithm, for the four finite elements implemented. Moreover, the meshes plotted in Figure 9, reveal that the algorithm based on the a posteriori error estimator θ defined in (26) is able to identify the exponential boundary layer and to refine the mesh around it. As a result, the adaptive procedure is more efficient than the uniform refinement, as confirmed by Figures 6 and 7.

In Figure 8 we plot the efficiency index for the four finite elements: for first order elements the efficiency index barely moves away from 1, while for second order elements we observe more oscillations with a tendency of stabilization around 2.5. Finally, Figure 10 shows the final concentration (left) and flux (right) obtained with the adaptive procedure.

6.3 Two sharp boundary layers

Next, we set $\Omega = (0,1) \times (0,1)$, $\Gamma_D = \Gamma$ and $\Gamma_N = \emptyset$. Let $\mathcal{K} = \epsilon\mathcal{I}$, with $\epsilon = 10^{-2}, 10^{-3}, 10^{-4}$, $\mathbf{b}(x,y) = (2,3)^{\mathsf{t}}$, and choose f and g such that the solution of the boundary value problem (5) is

$$u(x,y) = xy^2 - y^2 e^{\frac{2x-2}{\epsilon}} - xe^{\frac{3y-3}{\epsilon}} + e^{\frac{2x-5+ey}{\epsilon}}.$$

To our knowledge, this model problem was first proposed in [28], and then analyzed in [18].

In Figure 11, we can observe the total error and estimator for the uniform and adaptive refinements. From this graph, we conclude that the adaptive algorithm is more competitive than the uniform procedure. Figure 12 shows the decay of the individual errors, $e_k(u)$ and $e_k(\sigma)$, for the adaptive and uniform refinements. We observe a fast convergence of the adaptive procedure for the variables of interest (flux and concentration) as a consequence of the localization of the two boundary layers (see Figure 14).

Figure 13 shows the efficiency indices for the different finite elements and the different values of ϵ considered here. In Figure 14 we show the initial, an intermediary and final meshes for $(\mathcal{RT}_1, \mathcal{L}_2)$ and $\epsilon = 10^{-2}$. Finally, in Figure 15 we show the final flux and concentration.

6.4 An example with a singular solution

Let $\Omega = (0,1) \times (0,1)$ be the unit square, $\Gamma_D = \Gamma$ and $\Gamma_N = \emptyset$. Let $\mathcal{K} = \epsilon\mathcal{I}$, with $\epsilon = 10^{-4}$, $\mathbf{b} = (1,0)^{\mathsf{t}}$, and choose f and g such that the solution of the boundary value problem (5) is

$$u(x,y) = (2.1 - x - y)^{-2/3}.$$

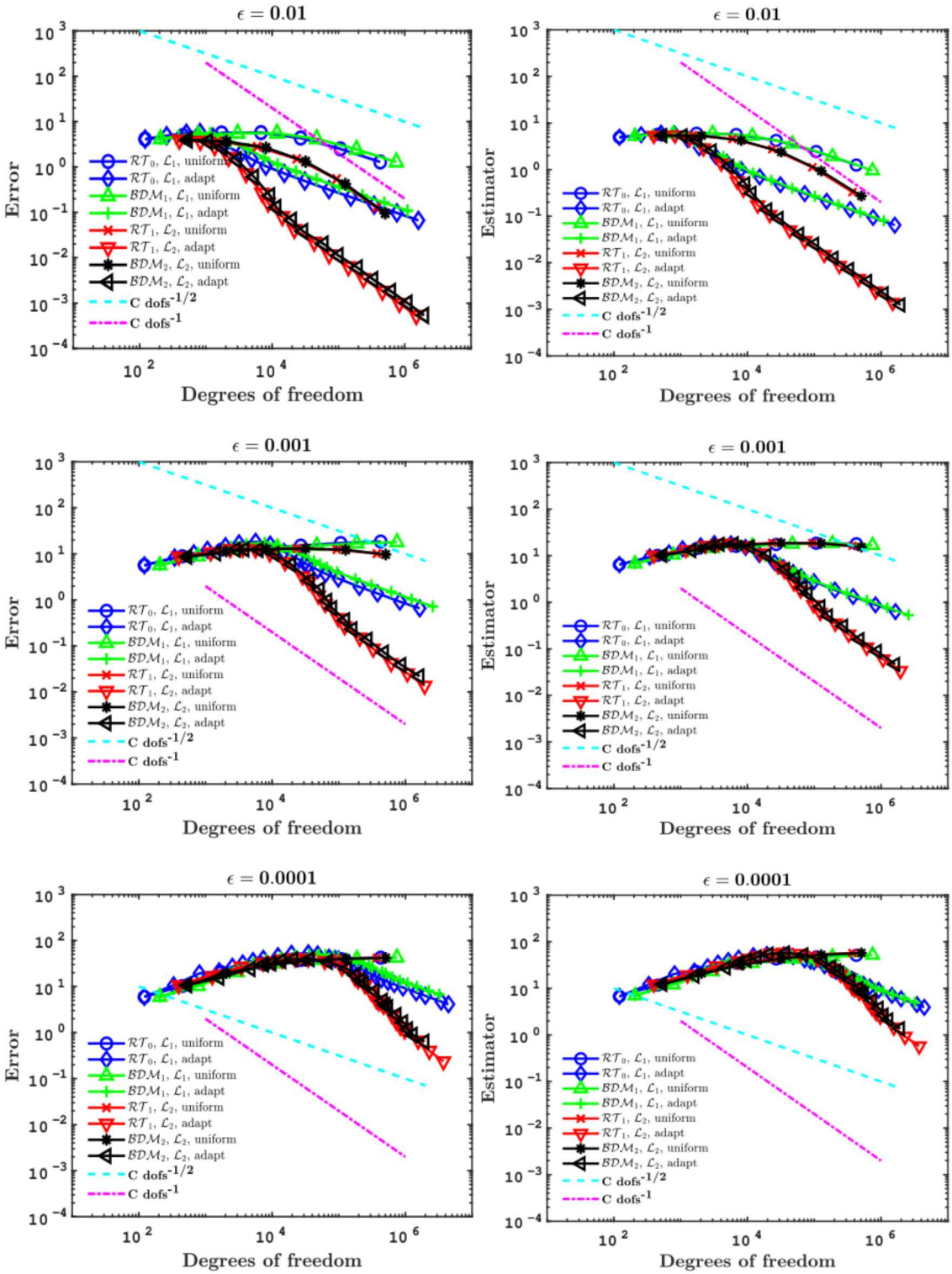

Figure 6: Example 2: Decay of total errors (left) and estimators (right) vs. number of degrees of freedom for different elements and values of ϵ.

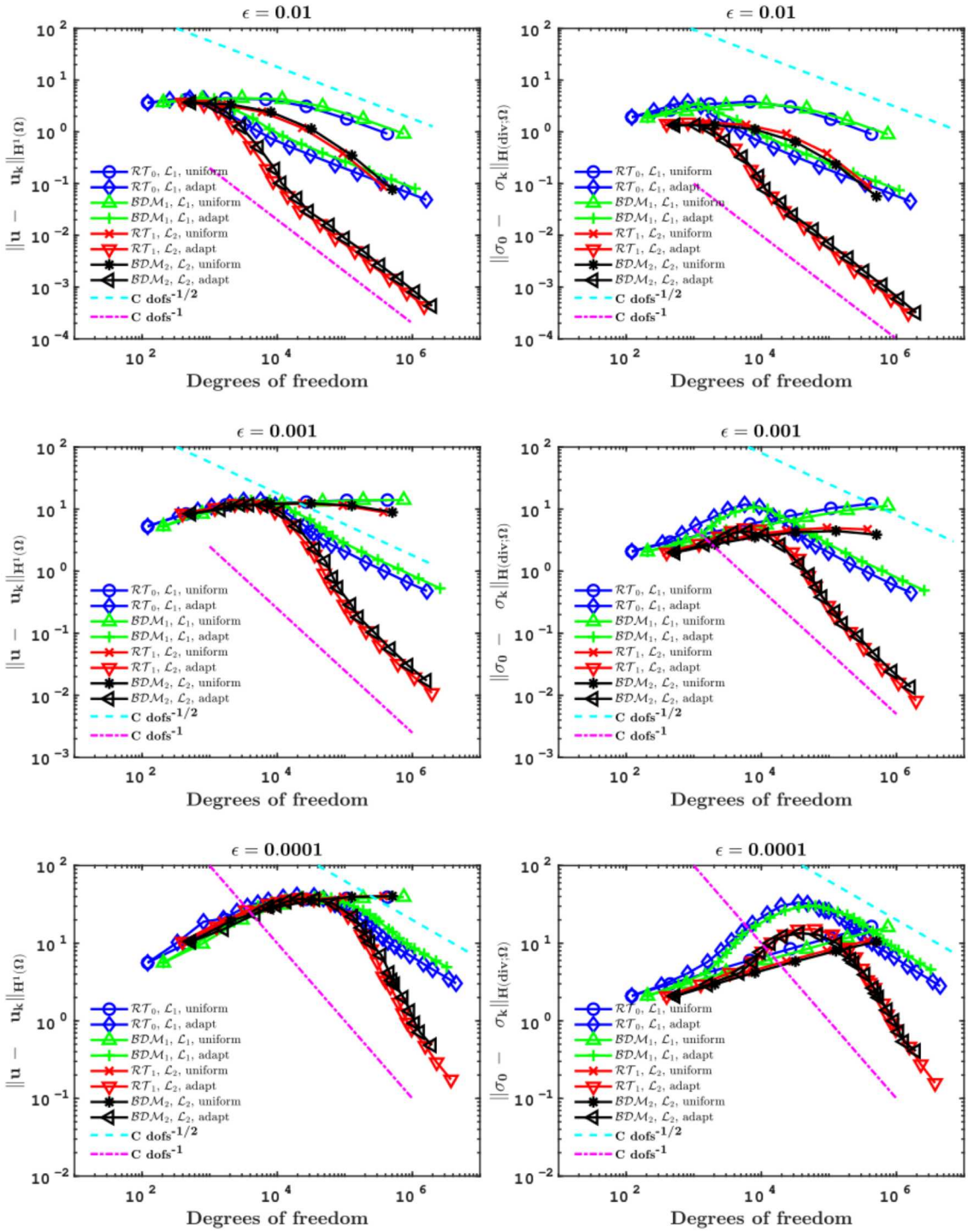

Figure 7: Example 2: Decay of $e_k(u)$ (left) and $e_k(\sigma)$ (right) vs. number of degrees of freedom for different elements and values of ϵ.

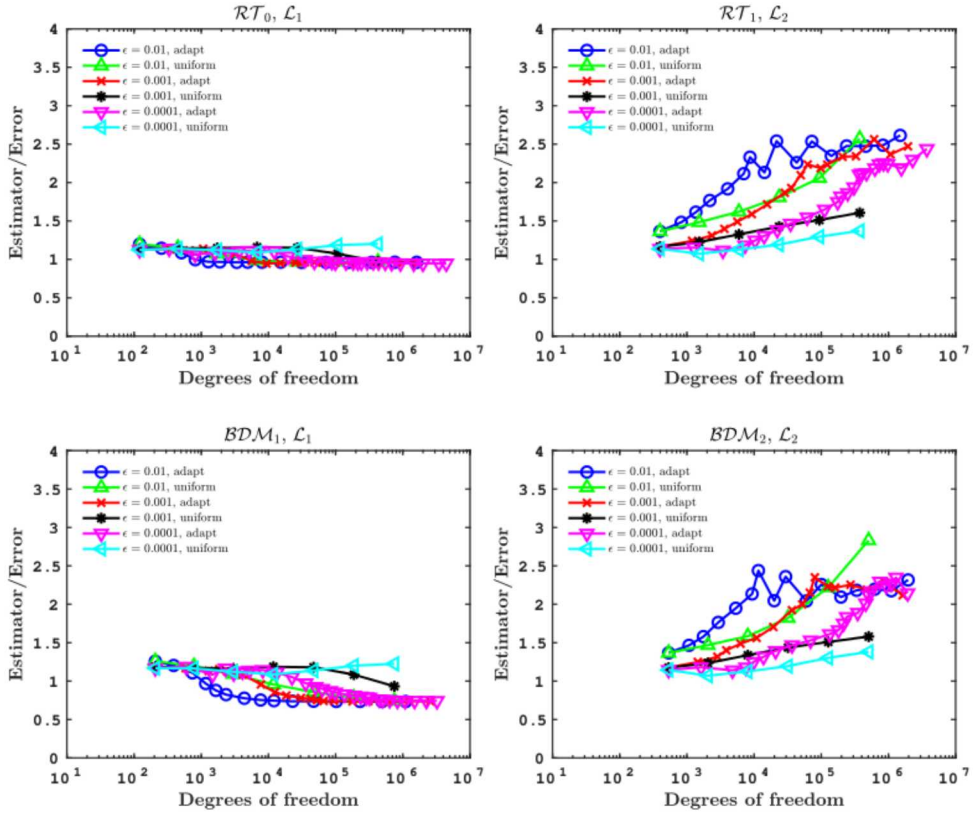

Figure 8: Example 2: Efficiency indices for different finite elements and values of ϵ.

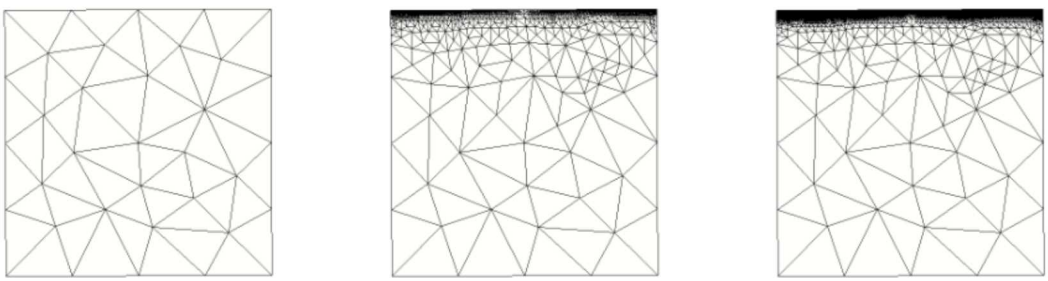

Figure 9: Example 2: Initial (397 DOFs), intermediate (8 702 DOFs) and final (1.5×10^6 DOFs) mesh in the case of $(\mathcal{RT}_1, \mathcal{L}_2)$ and for $\epsilon = 0.01$.

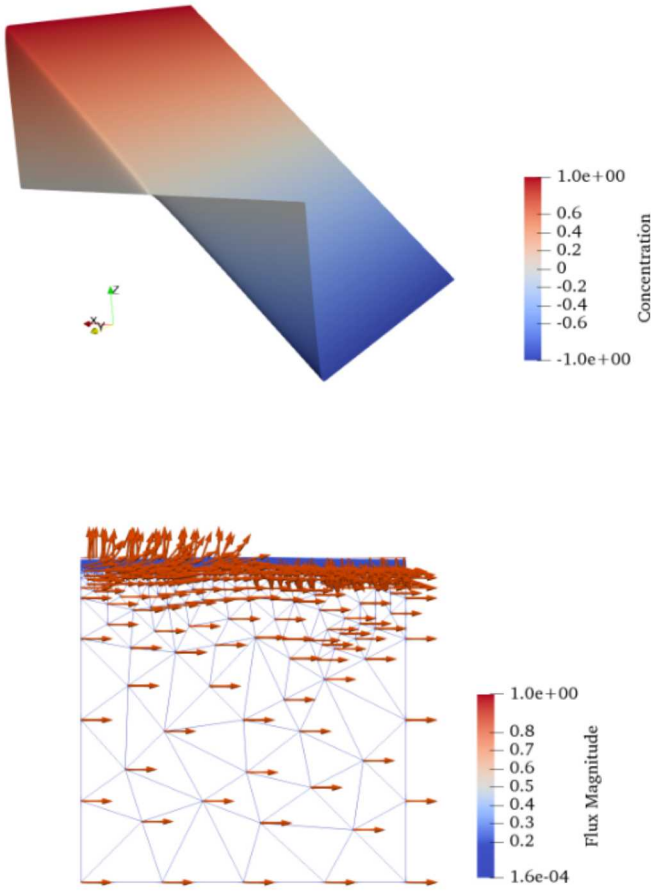

Figure 10: Example 2: Final concentration (top) and flux (bottom) in the case of $(\mathcal{RT}_1, \mathcal{L}_2)$ and for $\epsilon = 0.01$.

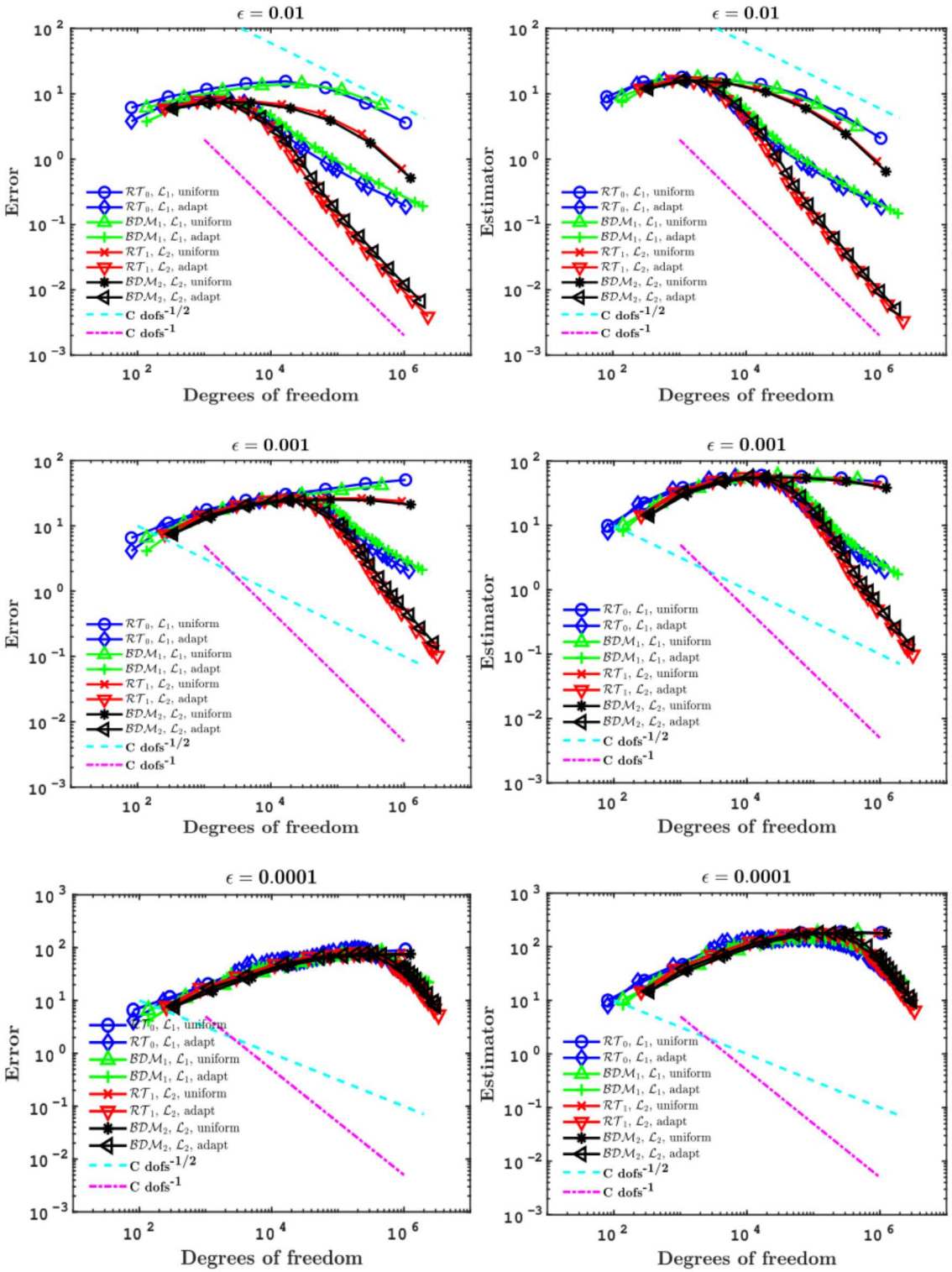

Figure 11: Example 3: Decay of total errors (left) and estimators (right) vs. number of degrees of freedom for different elements and different values of ϵ.

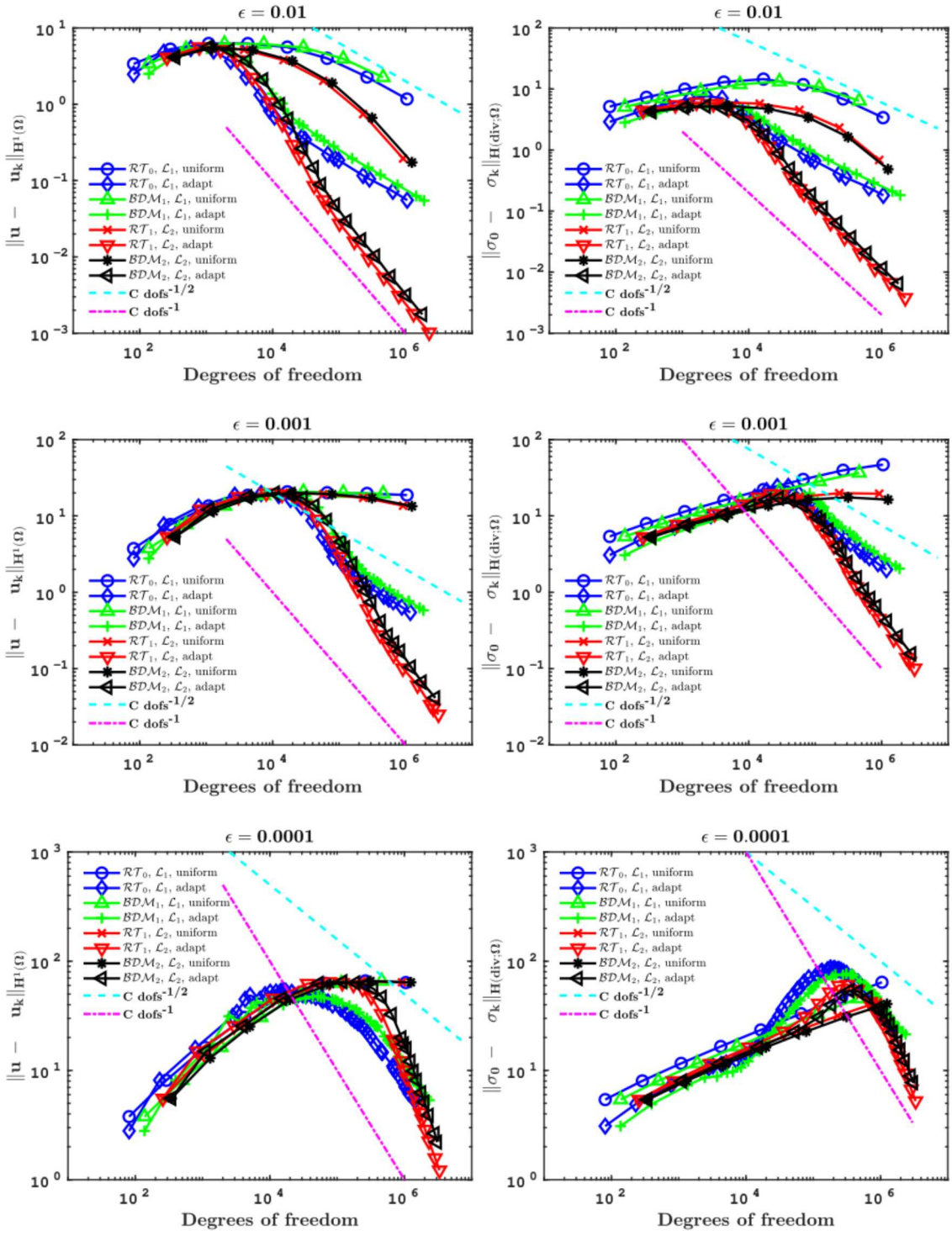

Figure 12: Example 3: Decay of $e_k(u)$ (left) and $e_k(\sigma)$ (right) vs. number of degrees of freedom for different elements and values of ϵ.

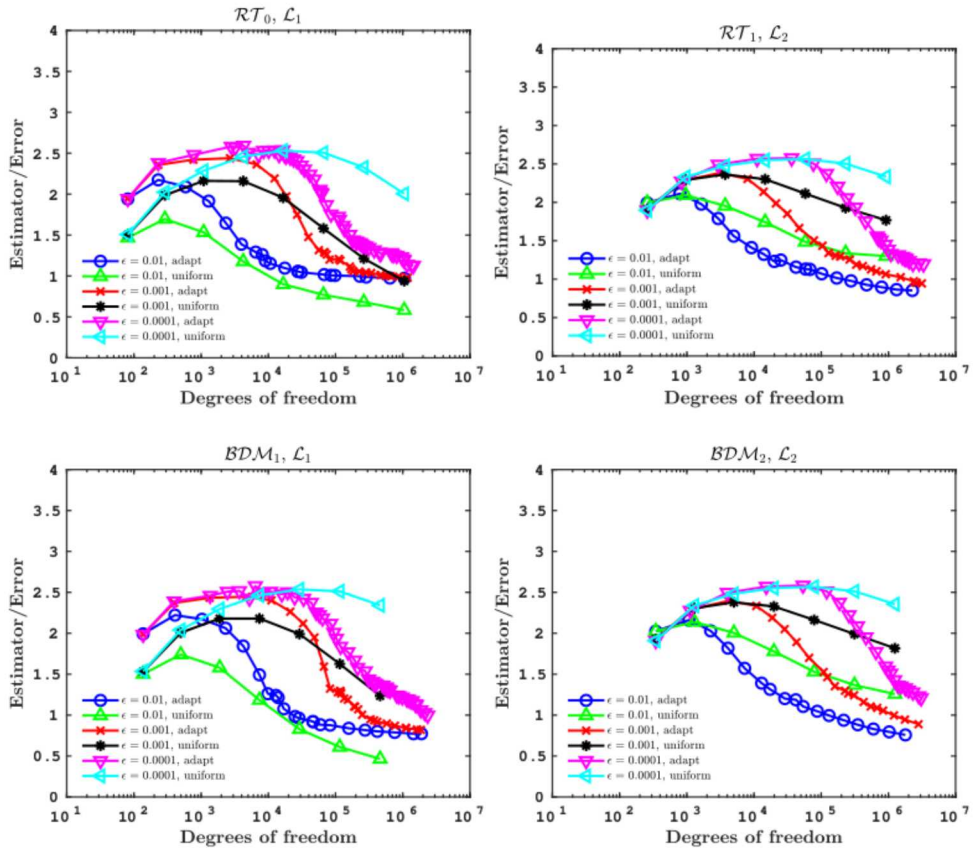

Figure 13: Example 3: Efficiency indices for different finite elements and values of ϵ.

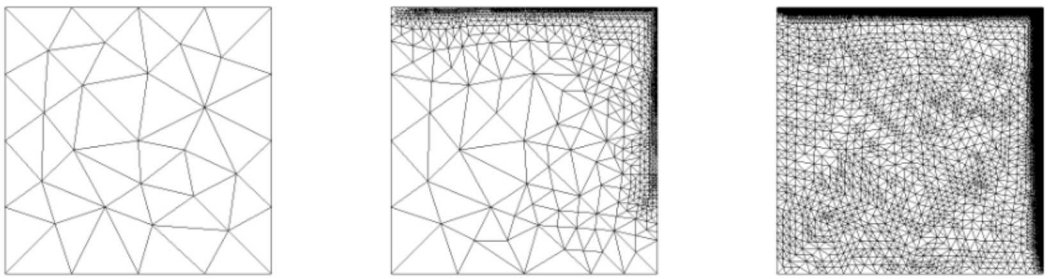

Figure 14: Example 3: Initial (397 DOFs), intermediate (42 082 DOFs) and final (2.2×10^6 DOFs) mesh in the case of $(\mathcal{RT}_1, \mathcal{L}_2)$ and for $\epsilon = 0.01$.

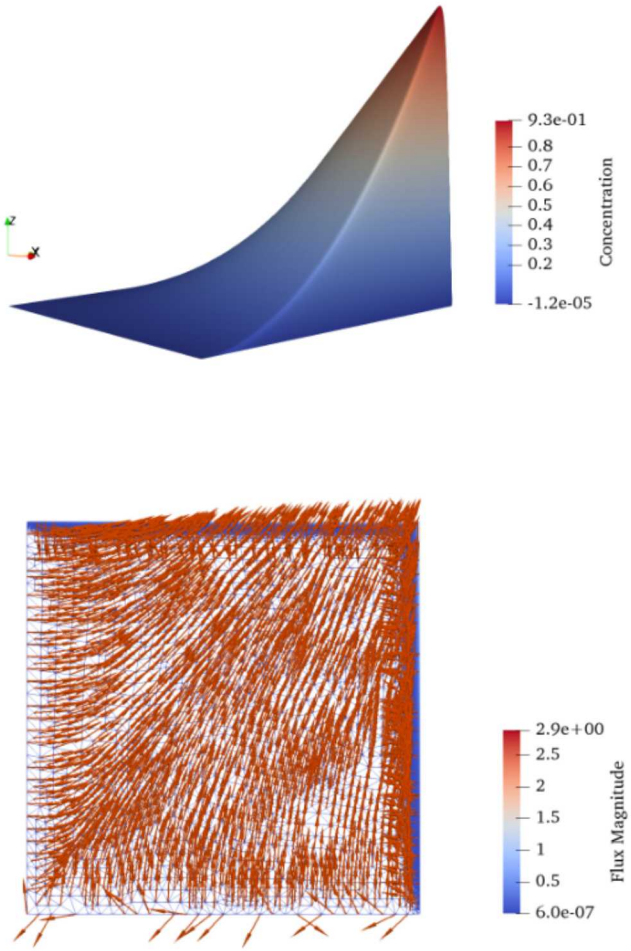

Figure 15: Example 3: Final concentration (top) and flux (bottom) in the case of $(\mathcal{RT}_1, \mathcal{L}_2)$ and for $\epsilon = 0.01$.

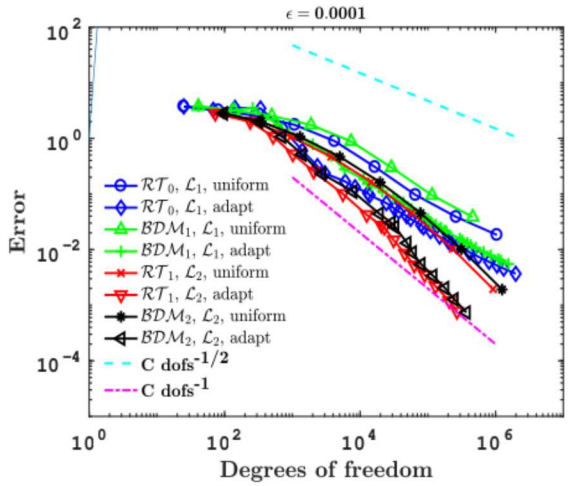

Figure 16: Example 4: Decay of total errors vs. number of degrees of freedom for different elements.

We remark that this solution presents a singularity in an exterior neighbourhood of the point $(1,1)$. In this example, we took $\kappa_1 = \frac{\epsilon}{8}$, $\kappa_2 = \frac{\epsilon}{2}$ and $\kappa_3 = 1$.

In Figure 16, we can observe the decay of the total error for the uniform and adaptive refinements for the 4 different finite elements implemented. From this graph, we conclude that in all cases the adaptive algorithm is more competitive than the uniform procedure. Figure 17 shows the efficiency indices for the different finite elements. In Figure 18 we show the initial mesh, and the meshes obtained after 6 and 12 iterations of the adaptive algorithm with the $(\mathcal{RT}_0, \mathcal{L}_1)$ element and $\epsilon = 10^{-4}$. We remark that the adapted meshes are highly refined around the singularity. Finally, in Figure 19 we show the final flux and concentration.

6.5 An example with mixed boundary conditions

Let $\Omega = (0,1) \times (0,1)$ be the unit square and define Γ_D to be the left and bottom boundaries of Ω and $\Gamma_N = \partial\Omega \setminus \Gamma_D$. Let $\mathcal{K} = \epsilon\mathcal{I}$, with $\epsilon = 10^{-3}$, $\mathbf{b} = (2,1)^{\mathbf{t}}$, and choose the data so that the solution of the boundary value problem is

$$u(x,y) = 1 + \frac{1}{2}\Big(\tanh(\frac{x}{a_\omega}) + \tanh(\frac{-y}{a_\omega})\Big)$$

with $a_\omega = 0.05$. We remark that this solution has two boundary layers around the Dirichlet boundaries. In this example, we took $\kappa_1 = \frac{\epsilon}{32}$, $\kappa_2 = \frac{\epsilon}{2}$ and $\kappa_3 = 2$.

In Figure 20, we observe the decay of the total error for the uniform and adaptive refinements for the 4 different finite elements implemented. From this graph, we conclude that in all cases

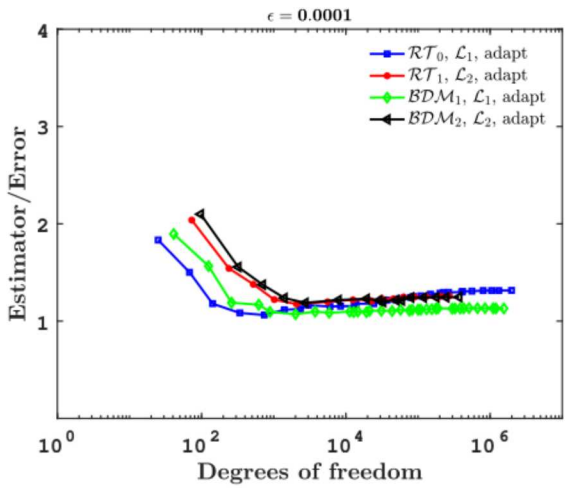

Figure 17: Example 4: Efficiency indices for different finite elements.

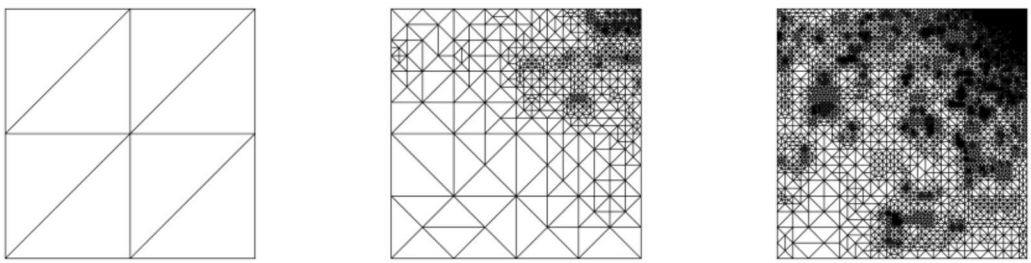

Figure 18: Example 4: Initial mesh (25 DOFs), mesh after 6 iterations (2 346 DOFs) and mesh after 12 iterations (24 415 DOFs) in the case of $(\mathcal{RT}_0, \mathcal{L}_1)$ and $\epsilon = 10^{-4}$.

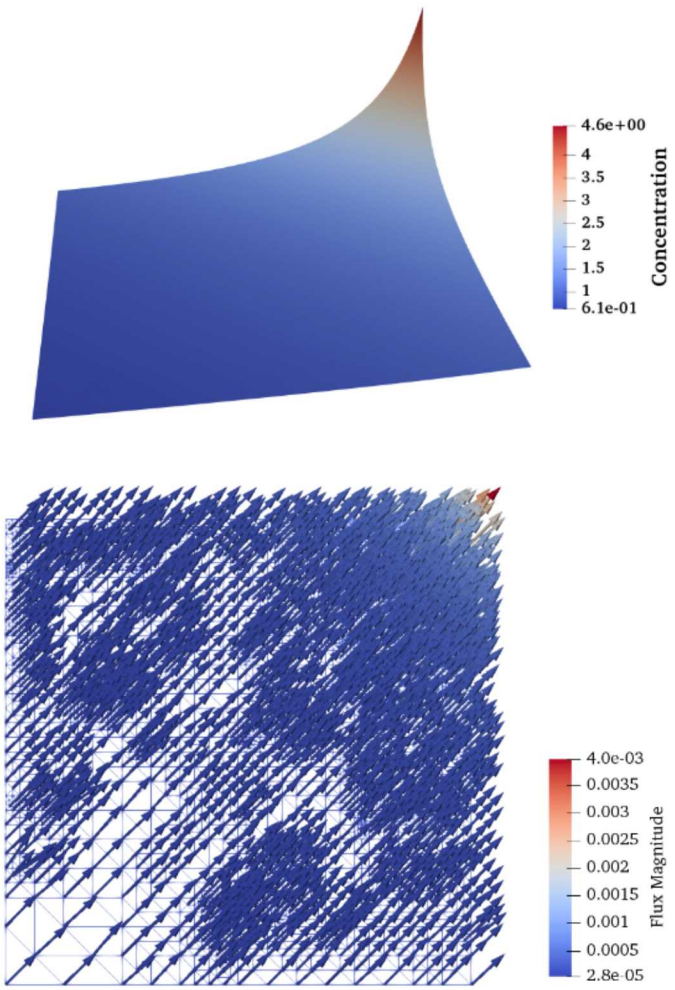

Figure 19: Example 4: Final concentration (top) and flux (bottom) for $(\mathcal{RT}_0, \mathcal{L}_1)$.

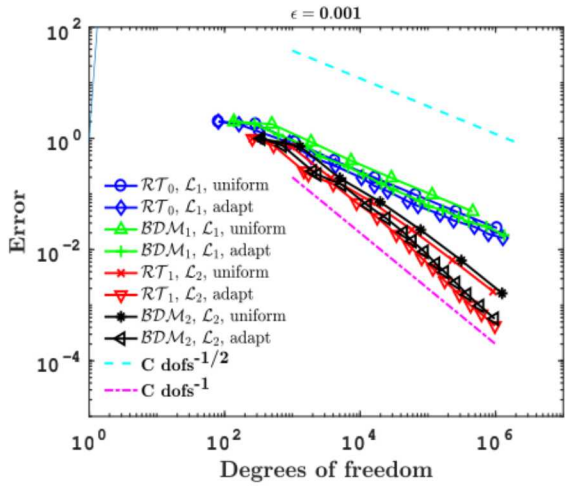

Figure 20: Example 5: Decay of total errors vs. number of degrees of freedom for different elements.

the adaptive algorithm is more competitive than the uniform procedure. Figure 21 shows the efficiency indices for the different finite elements. In Figure 22 we show the initial mesh, and the meshes obtained after 5 and 10 iterations of the adaptive algorithm with the $(\mathcal{RT}_0, \mathcal{L}_1)$ element and $\epsilon = 10^{-3}$. We remark that the adapted meshes are highly refined around the boundary layers. Finally, in Figure 23 we show the final flux and concentration.

7 Conclusions

We introduced new augmented mixed finite element methods for the linear convection-diffusion equation with mixed boundary conditions in two and three dimensions. We proved that for appropriate values of the stabilization parameters, the new variational formulation and the corresponding Galerkin schemes are well-posed and a Céa estimate holds. We derived the rate of convergence when the flux is approximated by Raviart-Thomas or Brezzi-Douglas-Marini elements, and the concentration is approximated by continuous piecewise polynomials. Moreover, we developed an a posteriori error analysis of residual type based on the use of a discrete Helmholtz decomposition, and propose a simple a posteriori error indicator that is reliable and locally efficient. This a posteriori error indicator consists of two terms in internal elements and elements with a side/face on the Neumann boundary, and it contains two additional terms on elements with a side/face on the Dirichlet boundary.

Finally, we presented some numerical examples that illustrate the performance of the augmented schemes and confirm the properties of the a posteriori error indicators when the finite element pairs $(\mathcal{RT}_0, \mathcal{L}_1)$, $(\mathcal{BDM}_1, \mathcal{L}_1)$, $(\mathcal{RT}_1, \mathcal{L}_2)$ and $(\mathcal{BDM}_2, \mathcal{L}_2)$ are used. The fist example

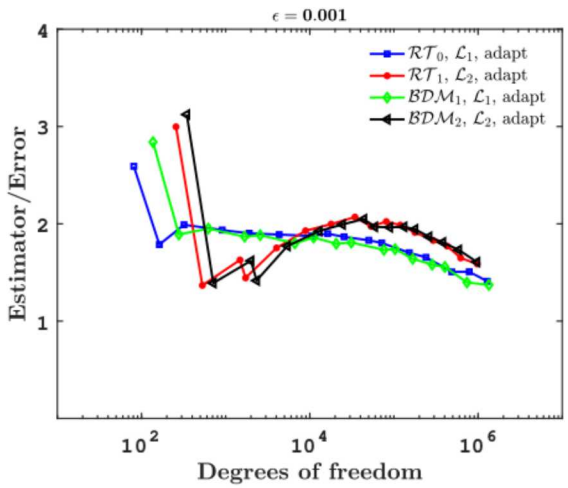

Figure 21: Example 5: Efficiency indices for different finite elements.

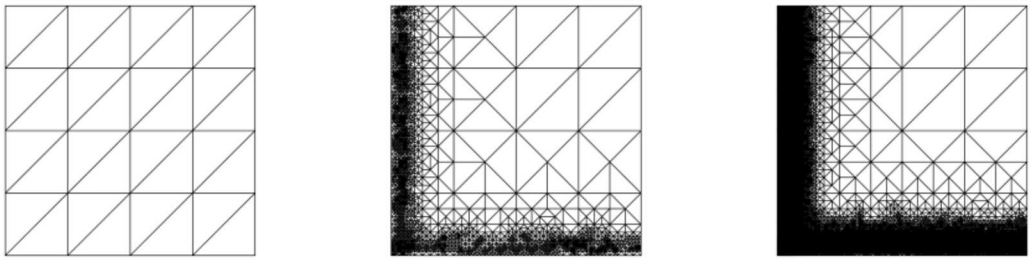

Figure 22: Example 5: Initial mesh (81 DOFs), mesh after 5 iterations (4 326 DOFs) and mesh after 10 iterations (71 626 DOFs) in the case of $(\mathcal{RT}_0, \mathcal{L}_1)$ and $\epsilon = 10^{-3}$.

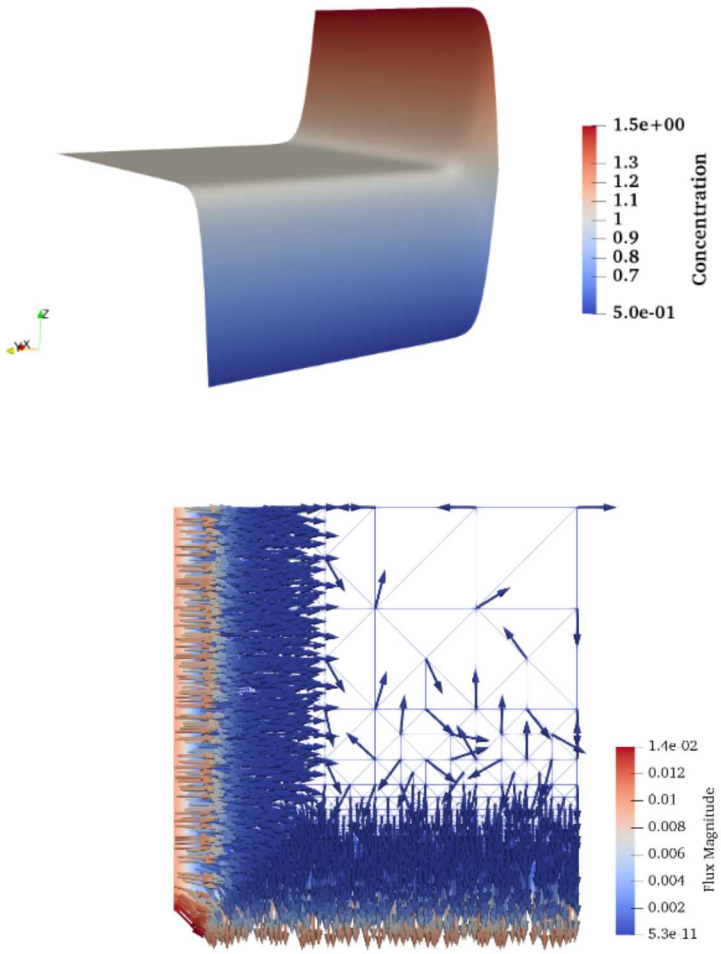

Figure 23: Example 5: Final concentration (top) and flux (bottom) for $(\mathcal{RT}_1, \mathcal{L}_2)$.

allowed us to test the robustness of the augmented mixed finite element schemes with respect to the stabilization parameters, and to confirm that optimal convergence rates predicted by the theory are attained in all cases when the solution is sufficiently smooth. Then, we tested the adaptive algorithm based on the a posteriori error indicator θ_T (see (27)) over an example with an exponential boundary layer, another one with two sharp boundary layers, an example with a singular solution and one example with mixed boundary conditions. We conclude that the adaptive algorithm is more competitive than the uniform refinement algorithm, and it is able to locate the boundary layers and singularities of the solutions.

We remark that we have taken as point of departure a centered mixed scheme. In order to tackle problems with a more important convection, we need to incorporate some kind of upwinding technique. This will be the subject of a forthcoming paper.

References

[1] B. Achchab, A. Agouzal, M. El Fatini and A. Souissi, *Robust hierarchical a posteriori error estimators for stabilized convection-diffusion problems*, Numerical Methods for Partial Differential Equations 28 (2012) 1717-1728.

[2] S. Agmon, Lectures on Elliptic Boundary Value Problems, Princeton, 1965.

[3] M. Ainsworth, A. Allendes, G.R. Barrenechea and R. Rankin, *Fully computable a posteriori error bounds for stabilised FEM approximations of convection-reaction-diffusion problems in three dimensions*, Int. J. Numer. Meth. Fluids 73 (2013) 765-790.

[4] D.N. Arnold, *An interior penalty finite element method with discontinuous elements*, SIAM J. Numer. Anal. 19 (4) (1982) 742-760.

[5] T.P. Barrios, R. Bustinza, G.C. García, E. Hernández, *On stabilized mixed methods for generalized Stokes problem based on the velocity-pseudostress formulation: A priori error estimates*, Comput. Methods Appl. Mech. Engrg. 237-240 (2012) 78-87.

[6] T. Barrios, R. Bustinza, G.C. García and M. González, *An a Posteriori Error Estimator for a New Stabilized Formulation of the Brinkman Problem*, Numerical Mathematics and Advanced Applications - ENUMATH 2013, (Springer, ed.). Lecture Notes in Computational Science and Engineering, Springer (2015), 253-261.

[7] T.P. Barrios, J.M. Cascón and M. González, *A posteriori error analysis of an augmented mixed finite element method for Darcy flow*, Comput. Methods Appl. Mech. Engrg. 283 (2015) 909-922.

[8] T.P. Barrios, J.M. Cascón and M. González, *A posteriori error estimation of a stabilized mixed finite element method for Darcy flow*, in Boundary and Interior Layers, Computational and Asymptotic Methods BAIL 2014, Petr Knobloch (editor), Lecture Notes in Computational Science and Engineering, 2015, 13-24.

[9] T.P. Barrios, J.M. Cascón and M. González, *Augmented mixed finite element method for the Oseen problem: A priori and a posteriori error analyses*, Comput. Methods Appl. Mech. Engrg. 313 (2017) 216-238.

[10] F. Brezzi and M. Fortin, Mixed and Hybrid Finite Element Methods, Springer-Verlag, 1991.

[11] P.G. Ciarlet, The Finite Element Method for Elliptic Problems, North-Holland, 1978.

[12] Ph. Clément, *Approximation by finite element functions using local regularization*, RAIRO Analyse numérique, vol. 9 (2), pp. 77-84 (1975).

[13] J. Douglas, Jr. and J.E. Roberts, *Mixed finite element methods for second order elliptic problems*, Mat. Apl. Comp. 1 (1982) 91-103.

[14] J. Douglas, Jr. and J.E. Roberts, *Global estimates for mixed methods for second order elliptic equations*, Math. Comp. 44(169) (1985) 39-52.

[15] S. Du, *A new residual a posteriori error estimates of mixed finite element methods for convection-diffusion-reaction equations*, Numerical Methods for Partial Differential Equations 30 (2014) 593-624.

[16] S. Du and Z. Zhang, *A robust residual-type a posteriori error estimator for convection-diffusion equations*, J. Sci. Computing 65 (2015) 138-170.

[17] M. Eigel and C. Merdon, *Equilibration a posteriori error estimation for convection-diffusion-reaction problems*, J. Sci. Computing 67 (2016) 747-768.

[18] A.H. Elsheikh, S. Smith and S.E. Chidiac, *Numerical investigation of the reliability of a posteriori error estimation for advection-diffusion equations*, Commun. Numer. Meth. Engng. 24 (2008) 711-726.

[19] M. Farhloul and A.S. Mounim, *A mixed-hybrid finite element method for convection-diffusion problems*, Appl. Math. Comput. 171 (2005) 1037-1047.

[20] G.N. Gatica, *Analysis of a new augmented mixed finite element method for linear elasticity allowing* $\mathbb{RT}_0$-$\mathbb{P}_1$-$\mathbb{P}_0$ *approximations*, ESAIM: M2AN Math. Model. Numer. Anal. 40 (2006) 1-28.

[21] G.N. Gatica, *An augmented mixed finite element method for linear elasticity with non-homogeneous Dirichlet conditions*, Electron. Trans. Numer. Anal. 26 (2007) 421-438.

[22] G.N. Gatica, *A note on a stable Helmholtz decomposition in 3D*, To appear in Applicable Analysis (2018), DOI: 10.1080/00036811.2018.1522627.

[23] G.N. Gatica, A. Márquez and S. Meddahi, *An augmented mixed finite element method for 3D linear elasticity problems*, J. Comput. Appl. Math. 231 (2009) 526-540.

[24] M. González, *Stabilized dual-mixed method for the problem of linear elasticity with mixed boundary conditions*, Applied Mathematics Letters 30 (2014) 1-5.

[25] M. González, S. Korotov and J. Jansson, *A posteriori error analysis of a stabilized mixed FEM for convection-diffusion problems*, Discrete Contin. Dyn. Syst. (2015), Dynamical systems, differential equations and applications. 10th AIMS Conference. Suppl., 525-532.

[26] T.J.R. Hughes and A.N. Brooks, *A multi-dimensional upwind scheme with no crosswind diffusion.* In T.J.R. Hughes (Ed.), FEMs for convection-dominated flows, ASME AMD 34 (1979) 19-35.

[27] J. Jaffre, *Decentrage et elements finis mixtes pour les equations de diffusion-convection*, Calcolo 21(2) (1984) 171-197.

[28] V. John, *A numerical study of a posteriori error estimators for convection-diffusion equations*, Comput. Methods Appl. Mech. Engrg. 190(5-7) (2000) 757-781.

[29] 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) 289-305.

[30] C. Johnson, U. Nävert and J. Pitkäranta, *Finite element methods for linear hyperbolic problems*, Comput. Methods Appl. Mech. Engrg. 45 (1984) 285-312.

[31] D. Kim and E.-J. Park, *A posteriori error estimators for the upstream weighting mixed methods for convection diffusion problems*, Comput. Methods Appl. Mech. Engrg. 197 (2008) 806-820.

[32] A. Logg, K.-A. Mardal and G.N. Wells Ed. *Automated Solution of Differential Equations by the Finite Element Method. The FEniCS Book*, Springer, 2012.

[33] A. Masud and T.J.R. Hughes, *A stabilized mixed finite element method for Darcy flow*, Comput. Methods Appl. Mech. Engrg. 191 (2002) 4341-4370.

[34] A. Papastavrou and R. Verfürth, *A posteriori error estimators for stationary convection-diffusion problems: a computational comparison*, Comput. Methods Appl. Mech. Engrg. 189(2) (2000) 449-462.

[35] A. Plaza and G. F. Carey, *Local refinement of simplicial grids based on the skeleton*, Appl. Numer. Math. 32 (2000) 195-218.

[36] J.E. Roberts and J.-M. Thomas, *Mixed and Hybrid Methods*, in Handbook of Numerical Analysis, edited by P.G. Ciarlet and J.L. Lions, vol. II, Finite Element Methods (Part 1). North-Holland, Amsterdam (1991).

[37] J.-M. Thomas, *Mixed finite elements methods for convection-diffusion problems*, Numerical Approximation of Partial Differential Equations, Elsevier (1987) 241-250.

[38] R. Verfürth, *A posteriori error estimation and adaptive mesh-refinement techniques*, J. Comput. Appl. Math. vol.50, pp. 67-83 (1994).

[39] R. Verfürth, A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques, Wiley-Teubner, 1996.

[40] R. Verfürth, *A posteriori error estimators for convection-diffusion equations*, Numer. Math. 80(4) (1998) 641-663.

[41] R. Verfürth, *Robust a posteriori error estimates for stationary convection-diffusion equations*, SIAM J. Numer. Anal. 43(4) (2005) 1766-1782.

[42] R. Verfürth, A Posteriori Error Estimation Techniques for finite Element Methods, Oxford University Press, 2013.

[43] M. Vohralík, *A posteriori error estimates for lowest-order mixed finite element discretizations of convection-diffusion-reaction equations*, SIAM J. Numer. Anal. 45(4) (2007) 1570-1599.