EXPLICIT AND AVERAGING A POSTERIORI ERROR ESTIMATES FOR ADAPTIVE FINITE VOLUME METHODS

C. CARSTENSEN, R. LAZAROV, AND S. TOMOV

ABSTRACT. Local mesh-refining algorithms known from adaptive finite element methods are adopted for locally conservative and monotone finite volume discretizations of boundary value problems for steady-state convection-diffusion-reaction equations. The paper establishes residual-type explicit error estimators and averaging techniques for a posteriori finite volume error control with and without upwind in global H^1 and L^2 norms. Reliability and efficiency is verified theoretically and confirmed empirically with experimental support for the superiority of the suggested adaptive mesh-refining algorithms over uniform mesh-refining. A discussion on an adaptive real-life simulation of the concentration in a non-homogeneous aquifer concludes the paper.

1. INTRODUCTION

We consider the following convection-diffusion-reaction problem: Find $u = u(x)$ such that

$$\left\{\begin{array}{rll} Lu \equiv \nabla\cdot(-A\nabla u + \underline{b}u) + \gamma u &= f & \text{in } \Omega,\\ u &= 0 & \text{on } \Gamma_D,\\ (-A\nabla u + \underline{b}u)\cdot\underline{n} &= g & \text{on } \Gamma_N^{in},\\ -(A\nabla u)\cdot\underline{n} &= 0 & \text{on } \Gamma_N^{out}. \end{array}\right. \tag{1.1}$$

Here Ω is a bounded polygonal domain in R^d, $d = 2, 3$, $A = A(x)$ is $d \times d$ symmetric, bounded and uniformly positive definite matrix in Ω, $\underline{b}$ is a given vector function, $\underline{n}$ is the unit outer vector normal to $\partial\Omega$, and f is a given source function. We have also used the notation ∇u for the gradient of a scalar function u and $\nabla\cdot\underline{b}$ for the divergence of a vector function $\underline{b}$ in R^d. The boundary of Ω, $\partial\Omega$, is split into Dirichlet, Γ_D, and Neumann, Γ_N, parts. Further, the Neumann boundary is divided into two parts: $\Gamma_N = \Gamma_N^{in} \cup \Gamma_N^{out}$, where $\Gamma_N^{in} = \{x \in \Gamma_N : \underline{n}(x)\cdot\underline{b}(x) < 0\}$ and $\Gamma_N^{out} = \{x \in \Gamma_N : \underline{n}(x)\cdot\underline{b}(x) \geq 0\}$. We assume that Γ_D has positive surface measure.

This problem is a prototype for flow and transport in porous media. For example, $u(x)$ can represent the pressure head in an aquifer or the concentration of a chemical that is dissolved and distributed in the ground-water due to the processes of diffusion, dispersion, and absorption. In many cases $A = \epsilon I$, where I is the identity matrix in R^d and $\epsilon > 0$

Date: April 9, 2003.

1991 *Mathematics Subject Classification.* 65N30, 65M35.

Key words and phrases. convection-diffusion-reaction equations, 3-D problems, finite volume approximation, a posteriori error estimators, residual estimators, averaging estimators, ZZ refinement indicator.

This work has been partially supported by NSF under grant DMS–9973328. It was finalized when the first author was guest at the Isaac Newton Institute for Mathematical Science, Cambridge, UK.

is a small parameter. This corresponds to the important and difficult class of singularly perturbed convection-diffusion problems (see, e.g. the monograph of Ross, Stynes, and Tobiska [29]). Further, $u(x)$ can be viewed as a limit for $t = \infty$ of the solution $u = u(x,t)$ of the corresponding time-dependent problem:

$$u_t + Lu = f,\ t > 0,\ x \in \Omega \tag{1.2}$$

with boundary conditions as above and an initial condition $u(x,0) = u_0(x)$, where u_0 is a given function in Ω. Various generalizations, mostly considering nonlinear terms, are possible and widely used in the applications. For example γu is replaced by nonlinear reaction term $\gamma(u)$ or the linear convective term $\underline{b}u$ is replaced by a nonlinear flux $\underline{b}(u)$. In this work we stay in the framework of the model problem (1.1) and focus on its 3-dimensional setting.

The development of efficient solution methods featuring error control is important for various applications. Our study has been motivated by the research in ground-water modeling and petroleum reservoir simulations (see, e.g. [17, 20]). The solutions of problems in that area exhibit steep gradients and rapid changes due to localized boundary data, discontinuities in the coefficients of the differential equation, and/or other local phenomena (for example extraction/injection wells, faults etc.). In order to accurately resolve such local behavior the numerical method should be able to detect the regions of significant changes of the solution and to refine the grid locally in a balanced manner, so that the overall accuracy is uniform in the whole domain.

Equation (1.1) expresses conservation of the properly scaled quantity u over any subdomain contained in Ω. In the context of ground-water fluid flows $u(x)$ is in general either the water mass or the mass of the chemical dissolved in the water. Numerical methods that have this property over a number of non-overlapping subdomains that cover the whole domain are called locally conservative. Finite volumes (control volumes, box schemes), mixed finite elements, and discontinuous Galerkin methods have this highly desirable property. The simplicity of the finite volume approximations combined with their local conservation property and flexibility motivated our study.

There are few works related to a posteriori error estimates for finite volume methods. In [1] L. Angermann has studied a balanced a posteriori error estimate for finite volume discretizations for convection-diffusion equations in 2-D on Voronoi meshes. The derivation of the error estimator is based on the idea of his previous work [2] on finite element method. The estimator for the finite volume method contains two new terms, which have been studied. In our paper we use similar approach, namely, the error estimates for the finite volume method are derived by using the relation between the finite volume and finite element methods (see, e.g. [6]). We note that despite the recent progress (see, e.g. the monographs [21, 24]) the theory of the finite volume methods is still under development. This in turn raises certain difficulties in establishing an independent and sharp a posteriori error analysis for the finite volume approximations.

A posteriori error indicators and estimators for the finite element method have been used and studied in the past two decades. Since the pioneering paper of Babuska and Rheinboldt [4] the research in this field has expanded in various directions that include *Residual Based* method (see the survey paper of Verfürth [33]), *Hierarchical Based* error estimators [7], estimators based on post-processing of the approximate solution gradient [34, 35], error estimators that control the error or its gradient in the maximum norm, etc. One popular

approach is to evaluate certain local residuals and obtain the a posteriori error indicator by solving local Dirichlet or Neumann problems by taking the local residuals as data [4, 7]. Another variation of the method that controls the global L^2- and H^1-norms of the error uses the Galerkin orthogonality, a priori interpolation estimates, and global stability (see, for example, [18]). Furthermore, solving appropriate dual problems, instead of using the a priori interpolation estimates, leads to error estimators controlling various kinds of error functional [9]. Solving finite element problems in an enriched by hierarchical bases functions space gives rise to *Hierarchical Based* error estimators [7]. There are error estimators based on optimal a priori estimates in maximum norm [19]. Another type of error estimators/indicators, widely (and in most cases heuristicly) used in many adaptive finite element codes, is based on post-processing (averaging) of the approximate solution gradient (see [34, 35]). In the context of the finite element method for elliptic partial differential equations, averaging or recovery techniques are justified in [8, 12, 28]. Finally, for an extensive study of the efficiency and the reliability of the local estimators and indicators for finite element approximations we refer to the recent monograph of Babuska and Strouboulis [5].

In this paper we adapt the finite element local error estimation techniques to the case of finite volume approximations. We consider mainly the *Residual Based* a-posteriori error estimators and analyze the one that uses Galerkin orthogonality, a priori interpolation estimates, and global stability in L^2- and H^1-norms. Our theoretical and experimental findings are similar to those in [1] and could be summarized as follows. The a posteriori error estimates in the finite volume element method are quite close to those in the finite element method and the mathematical tools from the finite element theory can be successfully applied for their analysis. Our computational experiments with various model problems confirm this conclusion. For more computational examples we refer to [23].

The paper is organized as follows. We start with the finite volume element formulation in Section 2. The section defines the used notations, approximations, and gives some general results from the finite volume approximations. Section 3 studies the *Residual Based* error estimator, followed by a short description of the used adaptive refinement strategy (in Section 4). Finally, in Section 5, we present numerous computational results for 2-D and 3-D test problems which illustrate the adaptive strategy and support our theoretical findings.

2. Finite volume element approximation

Subsection 2.1 introduces the notations used in the paper. In Subsection 2.2 we define the finite volume element approximations and give a priori estimate for the error.

2.1. **Notations.** We denote by $L^2(K)$ the square-integrable real-valued functions over $K \subset \Omega$, by $(\cdot,\cdot)_{L^2(K)}$ the inner product in $L^2(K)$, by $|\cdot|_{H^1(K)}$ and $||\cdot||_{H^1(K)}$ respectively the seminorm and norm of the Sobolev space $H^1(K)$, namely

$$||u||_{L^2(K)} := (u,u)^{1/2}_{L^2(K)}, \quad |u|_{H^1(K)} := (\nabla u, \nabla u)^{1/2}_{L^2(K)}, \quad ||u||^2_{H^1(K)} := ||u||^2_{L^2(K)} + |u|^2_{H^1(K)}.$$

In addition, if $K = \Omega$ we suppress the index K and also write $(\cdot,\cdot)_{L^2(\Omega)} := (\cdot,\cdot)$ and $||\cdot||_{L^2} := ||\cdot||$. Further, we use the Hilbert space $H^1_D(\Omega) = \{v \in H^1(\Omega) : \ v|_{\Gamma_D} = 0\}$. Finally, we denote by $H^{1/2}(\partial K)$ the space of the traces of functions in $H^1(K)$ on the boundary ∂K.

To avoid writing unknown constants we use the notation $a \lesssim b$ instead of the inequality $a \le Cb$ where the constant C is independent of the mesh-size h.

We use Ilin's inequality (see, e.g. [26]) in our analysis: let Ω_δ be a strip along $\partial\Omega$ of width δ, then for any function $u \in H^1(\Omega)$

$$||u||_{L^2(\Omega_\delta)} \lesssim \delta^{1/2}||u||_{H^1(\Omega)}. \tag{2.3}$$

We use the weak formulation of problem (1.1). To formulate it we first introduce the bilinear form $a(\cdot,\cdot)$ defined on $H^1_D(\Omega) \times H^1_D(\Omega)$ as:

$$a(u,v) := (A\nabla u - \underline{b}u, \nabla v) + (\gamma u, v) + \int_{\Gamma_N^{out}} \underline{b}\cdot\underline{n}\ u\ v\, ds. \tag{2.4}$$

We assume that the coefficients of problem (1.1) are such that:
(a) the form is $H^1_D(\Omega)$-elliptic, i.e., there is a constant $c_0 > 0$ such that

$$c_0||u||_{H^1} \le a(u,u) \quad \text{for all } u \in H^1_D(\Omega);$$

(b) the form is bounded on $H^1_D(\Omega)$, i.e., there is a constant $c_1 > 0$ such that

$$a(u,v) \le c_1||u||_{H^1}||v||_{H^1} \quad \text{for all } u,\ v \in H^1_D(\Omega).$$

The above two conditions guarantee that the expression $a(u,u)$ is equivalent to the norm in $H^1_D(\Omega)$. Further, we shall use the notation $||u||_a^2 = a(u,u)$ and call this expression "energy" norm.

A sufficient condition for the ellipticity of the bilinear form is $\gamma(x) + 0.5\nabla\cdot\underline{b}(x) \ge 0$ for all $x \in \Omega$, while a sufficient condition for the continuity is boundedness of the coefficients $A(x)$, $\underline{b}(x)$ and $\gamma(x)$ in Ω. Further in the paper we assume that these conditions are satisfied. Then (1.1) has the following weak form: Find $u \in H^1_D(\Omega)$ such that

$$a(u,v) = F(v) := (f,v) - \int_{\Gamma_N^{in}} gv\, ds \quad \text{for all } v \in H^1_D(\Omega). \tag{2.5}$$

2.2. **Approximation method.** We assume that Ω is a convex polygonal domain. Further, Ω is partitioned into triangular (for the 2-D case) or tetrahedral (for the 3-D case) finite elements denoted by K. The elements are considered to be closed sets and the splitting is denoted by $\mathcal{T}$ and often called triangulation of Ω. Our analysis will be valid also for domains with smooth boundary. In this case we have to modify the triangulation so that the methods does not loose accuracy due to approximation of the domain. Such schemes have been discussed in [16].

We introduce the set $N_h = \{x_i : x_i \text{ is a vertex of element } K \in \mathcal{T}\}$ and denote by N_h^0 the set of all vertices in N_h except those on Γ_D. For a given vertex x_i we denote by $\Pi(i)$ the index set of all neighbors of x_i in N_h, i.e., all vertices that are connected to x_i by an edge.

For a given finite element triangulation $\mathcal{T}$, we construct a dual mesh $\mathcal{T}^*$ (based upon $\mathcal{T}$), whose elements are called control volumes (boxes, finite volumes, etc.). There are various ways to introduce the control volumes. Almost all approaches can be described in the following general scheme. In each element $K \in \mathcal{T}$ a point q is selected. For the 3-D case, on each of the four faces $\overline{x_ix_jx_k}$ of K a point x_{ijk} is selected and on each of the six edges $\overline{x_ix_j}$ a point x_{ij} is selected. Then q is connected to the points x_{ijk}, and in the corresponding faces, the points x_{ijk} are connected to the points x_{ij} by straight lines (see Figure 1). Control volume associated with a vertex x_i is denoted by V_i and defined as the union of the "quarter"

elements $K \in \mathcal{T}$, which have x_i as a vertex (see Figure 1). The interface between two control volumes, V_i and V_j, is denoted by γ_{ij}, i.e. $\overline{V}_i \cap \overline{V}_j = \gamma_{ij}$.

We assume that $\mathcal{T}$ is locally quasi-uniform, that is for $K \in \mathcal{T}$, $|K| \lesssim \rho(K)^d$, where $\rho(K)$ is the radius of the largest ball contained in K and $|K|$ denotes the area or volume of K. In the context of locally refined grids, this means that the smallest interior angle is bounded away from zero and any two neighboring finite elements are of approximately the same size whereas elements that are far away may have quite different sizes.

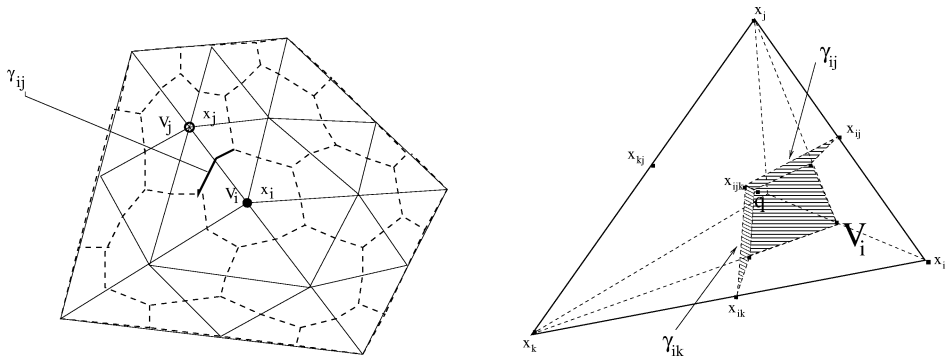

FIGURE 1. Left: Finite element and finite volume partitions in 2-D; Right: Contribution from one element to control volume V_i, γ_{ij} and γ_{ik} in 3-D; Point q is the element's medicenter and internal points for the faces are the medicenters of the faces.

In our 3-D computations q is the center of gravity of the element K, x_{ijk} are the centers of gravity of the corresponding faces, and x_{ij} are the mid-point (center of gravity) of the corresponding edges (as shown on Figure 1).

In 2-D, another possibility is to choose q to be the center of the circumscribed circle of K. This type of control volumes form Voronoi or PEBI meshes (see, e.g. [21], pp. 764, 825). Then obviously, γ_{ij} are the perpendicular bisectors of the three edges of K (see Figure 2). This construction requires that all finite elements are triangles of acute type, which we shall assume whenever such triangulation is used.

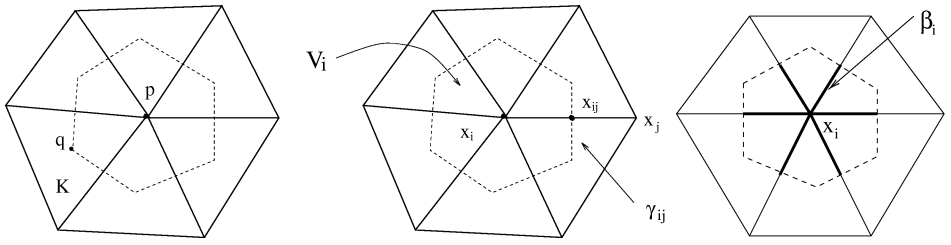

FIGURE 2. Control volumes with circumcenters as internal points (Voronoi meshes) and interface γ_{ij} of V_i and V_j. The rightmost picture shows the segments β_i in bold.

We define the linear finite element space S_h as

$$S_h = \{v \in C(\Omega) : \quad v|_K \text{ is affine for all } \quad K \in \mathcal{T} \text{ and } v|_{\Gamma_D} = 0\}$$

and its dual volume element space S_h^* by

$$S_h^* = \{v \in L^2(\Omega) : \ v|_V \text{ is constant for all } V \in \mathcal{T}^* \text{ and } v|_{\Gamma_D} = 0\}.$$

Obviously, $S_h = \text{span}\{\phi_i : \ x_i \in N_h^0\}$ and $S_h^* = \text{span}\{\chi_i : \ x_i \in N_h^0\}$, where ϕ_i denotes the standard nodal linear basis function associated with the node x_i and χ_i the characteristic function of the volume V_i. Let $I_h : C(\Omega) \cap H_D^1(\Omega) \to S_h$ be the interpolation operator and $I_h^* : C(\Omega) \cap H_D^1(\Omega) \to S_h^*$ and $P_h^* : C(\Omega) \cap H_D^1(\Omega) \to S_h^*$ be the piecewise constant interpolation and projection operators:

$$I_h u = \sum_{x_i \in N_h} u(x_i)\phi_i(x), \ I_h^* u = \sum_{x_i \in N_h} u(x_i)\chi_i(x), \text{ and } P_h^* u = \sum_{x_i \in N_h} \bar{u}_i \chi_i(x).$$

Here $\bar{u}_i$ is the averaged value of u over the volume V_i for $x_i \in N_h^0$, i.e., $\bar{u}_i = \int_{V_i} u\, dx/|V_i|$, and $\bar{u}_i = 0$ for $x_i \in \Gamma_D$. In fact, I_h makes also sense as an interpolation operator from S_h^* to S_h. Namely, if $v^* \in S_h^*$, then $I_h v^* \in S_h$ and $I_h v^*(x_i) = v^*(x_i)$.

Further, for $v^* \in S_h^*$, we use the notation $v_i^* = v^*(x_i)$. We also define the "total flux" and its approximation by

$$\underline{\sigma} := -A\nabla u + \underline{b}u, \qquad \underline{\sigma}_h := -A\nabla_h u_h + \underline{b}u_h$$

and assume that the coefficients $A(x)$ and $\underline{b}(x)$ are elementwise smooth. Also, we denote by $\nabla_h \cdot$ the $\mathcal{T}$-piecewise divergence and by ∇_h the $\mathcal{T}$-piecewise gradient. Integrals involving piecewise quantities are considered as sums over the pieces where the quantities are defined.

The finite volume element approximation u_h of (1.1) is the solution to the problem: Find $u_h \in S_h$ such that

$$a_h(u_h, v_h^*) := A(u_h, v_h^*) + C(u_h, v_h^*) = F(v_h^*) \quad \text{for all } v_h^* \in S_h^*. \tag{2.6}$$

Here the bilinear forms $A(u_h, v^*)$ and $C(u_h, v^*)$ are defined on $S_h \times S_h^*$ and the linear form $F(v^*)$ is defined on S_h^*. They are given by

$$A(u_h, v^*) = \sum_{x_i \in N_h^0} v_i^* \left(- \int_{\partial V_i \setminus \Gamma_N} (A\nabla_h u_h) \cdot \underline{n} ds + \int_{V_i} \gamma u_h dx \right), \tag{2.7}$$

$$C(u_h, v^*) = \sum_{x_i \in N_h^0} v_i^* \int_{\partial V_i \setminus \Gamma_N^{in}} (\underline{b} \cdot \underline{n})\ u_h\, ds, \tag{2.8}$$

$$F(v^*) = \sum_{x_i \in N_h^0} v_i^* \left\{ \int_{V_i} f dx - \int_{\partial V_i \cap \Gamma_N^{in}} g\, ds \right\}. \tag{2.9}$$

Obviously, $\nabla \cdot \underline{\sigma}_h$ is well defined over $V_i \cap K$ for all $V_i \in \mathcal{T}^*$ and $K \in \mathcal{T}$. This ensures, in particular, that the surface integrals in (2.7) and (2.8) exist.

In addition to $C(u_h, v^*)$ we introduce form $C^{up}(u_h, v^*)$ that uses upwind approximation. Approximation (2.7) – (2.9) can be used for moderate convection fields and dominating diffusion. For small diffusion, for example when $A = \epsilon I$ with ϵ small, approximation (2.7) – (2.9) gives oscillating numerical results which we would like to avoid. We are interested

in approximation methods that produce solutions satisfying the maximum principle and are locally conservative. Such schemes are also known as monotone schemes (see, e.g. [22, 29]). A well-known sufficient condition for a scheme to be monotone is that the corresponding stiffness matrix is an M–matrix (see [31] p. 182, p. 260 and [29] p. 202).

The upwind approximation that we use for problems with large convection (or small diffusion) is locally mass conservative and gives the desired stabilization. We split the integral over ∂V_i on integrals over $\gamma_{ij} = \partial V_i \cap \partial V_j$, (see Figure 1) and introduce out-flow and in-flow parts of the boundary of the volume V_i. This splitting can be characterized by the quantities $(\underline{b} \cdot \underline{n})_+ = \max(0,\ \underline{b} \cdot \underline{n})$ and $(\underline{b} \cdot \underline{n})_- = \min(0,\ \underline{b} \cdot \underline{n})$, where $\underline{n}$ is the outer unit vector normal to ∂V_i. Then we introduce

$$C^{up}(u_h, v^*) = \sum_{x_i \in N_h^0} v_i^* \left\{ \sum_{j \in \Pi(i)} \int_{\gamma_{ij}} \left((\underline{b} \cdot \underline{n})_+ u_h(x_i) + (\underline{b} \cdot \underline{n})_- u_h(x_j) \right) ds + \int_{\Gamma_N^{out} \cap \partial V_i} (\underline{b} \cdot \underline{n})\, u_h(x_i)\, ds \right\}. \tag{2.10}$$

This approximation is well defined for any $\underline{b}$. In order to avoid technicalities in our analysis we assume that the vector field $\underline{b}$ is piecewise smooth and has small variation over each finite element. Thus, the quantity $\underline{b} \cdot \underline{n}$ does not change sign over γ_{ij}.

The upwind finite volume element approximation u_h of (1.1) becomes: Find $u_h \in S_h$ such that

$$a_h^{up}(u_h, v^*) := A(u_h, v^*) + C^{up}(u_h, v^*) = F(v^*) \text{ for all } v^* \in S_h^*. \tag{2.11}$$

This is an extension of the classical upwind approximation of the convection term and is closely related to the discontinuous Galerkin approximation (see, e.g. [19]) or to the Tabata scheme for Galerkin finite element method [32]. It is also related to the scheme on Voronoi meshes derived by Mishev [25]. A different type of weighted upwind approximation on Voronoi meshes in 2-D has been studied by Angermann [1].

3. A posteriori error analysis

This section is devoted to the mathematical derivation of computable error bounds in the energy norm. Throughout the section, $u \in H_D^1$ denotes the exact solution of (2.5) and $u_h \in S_h$ denotes the discrete solution of either (2.6) or (2.11). Then, $e := u - u_h \in H_D^1(\Omega)$ is the (unknown) error and $\overline{e} := P_h^* e \in S_h^*$ is its $\mathcal{T}^*$-piecewise integral mean. We denote by $\mathcal{E}$ the set of all interior edges/faces in $\mathcal{T}$ respectively in two/three dimensions. Also, for a vertex $x_i \in N_h^0$ let $\beta_i := V_i \cap \mathcal{E}$ (see Figure 2). For any $E \in \mathcal{E}$ let $[\underline{\sigma}_h] \cdot \underline{n}$ denote the jump of $\underline{\sigma}_h$ across E in normal to E direction $\underline{n}$. The orientation of $\underline{n}$ is not important as long as the jump is in the same direction. In general, if $\underline{n}$ is present in a boundary integral, it will denote the outward unit vector normal to the boundary. With every element $K \in \mathcal{T}$, edge/face $E \in \mathcal{E}$, and volume $V_i \in \mathcal{T}^*$ we associate local mesh size denoted correspondingly by h_K, h_E, and h_i. Since the mesh is locally quasi-uniform the introduced mesh sizes are locally equivalent, i.e., bound each other from above and below with constants independent of the mesh size. Then, we introduce a global discontinuous mesh size function $h(x)$, $x \in \Omega$

that assumes value h_K, h_E, and h_i depending on $x \in K \setminus \partial K$, $x \in E$, or $x = x_i$, respectively. Finally, we use the following short-hand notation for integration over all faces E in $\mathcal{E}$:

$$\int_{\mathcal{E}} v ds := \sum_{E \in \mathcal{E}} \int_E v ds.$$

3.1. **Energy-norm a posteriori error estimate of the scheme without upwind.** We consider problem (2.6) and begin our analysis with the case when the form $C(\cdot,\cdot)$ is evaluated by (2.8). We first give a representation of the error and introduce some locally computable quantities. In Theorem 3.1 we show that these quantities give a reliable estimate for the error. Further, we introduce the error estimator, based on local "averaging" of the "total flux" $\underline{\sigma}$ over the control volumes and show that this estimator is reliable up to higher order terms.

The following lemma gives a representation of the error:

Lemma 3.1. *Let the coefficients of the convection-diffusion-reaction problem (1.1) be such that the assumptions from Subsection 2.1 be satisfied. Then for the error $e = u - u_h$, where u is the solution of (2.5) and u_h is the solution of (2.6) we have*

$$\begin{aligned} \|e\|_a^2 = (f - \nabla_h \cdot \underline{\sigma}_h - \gamma u_h,\ e - \overline{e}) - \int_{\mathcal{E}} [\underline{\sigma}_h] \cdot \underline{n}\, (e - \overline{e})\, ds \\ - \int_{\Gamma_N^{in}} (g - \underline{\sigma}_h \cdot \underline{n})\, (e - \overline{e})\, ds - \int_{\Gamma_N^{out}} (A \nabla_h u_h) \cdot \underline{n}\, (e - \overline{e})\, ds. \end{aligned} \tag{3.12}$$

Proof. We take $v = e \in H_D^1(\Omega)$ in (2.5) and use the definition of $a(\cdot,\cdot)$ by (2.4) to get

$$\begin{aligned} a(e,e) &= a(u,e) - a(u_h,e) \\ &= (f - \gamma u_h, e) + (\underline{\sigma}_h, \nabla e) - \int_{\Gamma_N^{in}} g e\, ds - \int_{\Gamma_N^{out}} (\underline{b} \cdot \underline{n})\, u_h\, e\, ds. \end{aligned}$$

We integrate the second term on the right-hand side by parts on each element $K \in \mathcal{T}$

$$\int_K \underline{\sigma}_h \cdot \nabla e\, ds = \int_{\partial K} (\underline{\sigma}_h \cdot \underline{n}) e\, ds - \int_K e \nabla \cdot \underline{\sigma}_h\, dx.$$

The sum over all elements yields the jump contributions $[\underline{\sigma}_h] \cdot \underline{n}$ along $\mathcal{E}$ and eventually proves

$$\begin{aligned} a(e,e) &= (f - \nabla_h \cdot \underline{\sigma}_h - \gamma u_h, e) - \int_{\mathcal{E}} [\underline{\sigma}_h] \cdot \underline{n}\, e\, ds \\ &\quad - \int_{\Gamma_N^{in}} (g - \underline{\sigma}_h \cdot \underline{n}) e\, ds - \int_{\Gamma_N^{out}} (A \nabla_h u_h) \cdot \underline{n}\, e\, ds. \end{aligned} \tag{3.13}$$

It remains to see that the preceding right-hand side vanishes if e is replaced by $\overline{e}$. For each control volume V_i we have from (2.6)–(2.8) that

$$\int_{\partial V_i \setminus \Gamma_N} \underline{\sigma}_h \cdot \underline{n}\, ds = \int_{V_i} (f - \gamma u_h)\, dx - \int_{\partial V_i \cap \Gamma_N^{out}} (\underline{b} \cdot \underline{n})\, u_h\, ds - \int_{\partial V_i \cap \Gamma_N^{in}} g\, ds.$$

The Gauß divergence theorem is applied to each non-void $K\cap V_i$, $K\in\mathcal{T}$ so that the left-hand side of the above inequality becomes

$$\int_{\partial V_i\setminus\Gamma_N}\underline{\sigma}_h\cdot\underline{n}\,ds=\int_{V_i}\nabla_h\cdot\underline{\sigma}_h\,dx+\int_{\beta_i}[\underline{\sigma}_h]\cdot\underline{n}\,ds-\int_{\partial V_i\cap\Gamma_N}\underline{\sigma}_h\cdot\underline{n}\,ds.$$

The difference of the preceding two identities is multiplied by $\overline{e}(x_i)$ and summed over all control volumes. This results in

$$0=(f-\nabla_h\cdot\underline{\sigma}_h-\gamma u_h,\overline{e})-\int_{\mathcal{E}}[\underline{\sigma}_h]\cdot\underline{n}\,\overline{e}\,ds-\int_{\Gamma_N^{in}}(g-\underline{\sigma}_h\cdot\underline{n})\,\overline{e}\,ds-\int_{\Gamma_N^{out}}A\nabla_h u_h\cdot\underline{n}\,\overline{e}\,ds.$$

Subtracting this identity from (3.13) concludes the proof of (3.12). □

Motivated by the above considerations we introduce the following locally computable quantities that play main role in the design of adaptive algorithms and their a posteriori error analysis.

Definition 3.1. *Set*

$$R_K(x):=(f-\nabla\cdot\underline{\sigma}_h-\gamma u_h)(x),\ x\in K,$$

$$R_E(x):=([\underline{\sigma}_h]\cdot\underline{n})(x),\ x\in E,\ \textit{for } E\cap\Gamma_N=\emptyset,$$

$$R_E^{in}(x):=(g-\underline{\sigma}_h\cdot\underline{n})(x),\ x\in E,\ \textit{for } E\subset\Gamma_N^{in},$$

$$R_E^{out}(x):=(A\nabla u_h\cdot\underline{n})(x),\ x\in E,\ \textit{for } E\subset\Gamma_N^{out}$$

and define

$$\eta_R:=\|h\,R_K\|_{L^2(\Omega)},\ \eta_E:=\|h^{1/2}\,R_E\|_{L^2(\mathcal{E})},$$

$$\eta_N:=\|h^{1/2}\,R_E^{in}\|_{L^2(\Gamma_N^{in})}+\|h^{1/2}\,R_E^{out}\|_{L^2(\Gamma_N^{out})}.$$

Lemma 3.2. *There holds*

$$\int_{\mathcal{E}}[\underline{\sigma}_h]\cdot\underline{n}(e-\overline{e})\,ds\lesssim\eta_E\,\|\nabla e\|.$$

Proof. A well-established trace inequality (cf., e.g., [10, Theorem 1.6.6] or [13, Theorem 1.4]) and scaling argument leads to

$$h_E^{1/2}\|v\|_{L^2(E)}\lesssim\|v\|_{L^2(K)}+h_E\|\nabla v\|_{L^2(K)}\tag{3.14}$$

for all $v\in H^1(K)$ and edges E of an element $K\in\mathcal{T}$. An application to $v:=e-\overline{e}$ on each $K\cap V_i$, where $K\in\mathcal{T}$ and $x_i\in N_h$, leads to

$$\begin{aligned}\int_{\beta_i}[\underline{\sigma}_h]\cdot\underline{n}(e-\overline{e})\,ds&\leq\|[\underline{\sigma}_h]\cdot\underline{n}\|_{L^2(\beta_i)}\|e-\overline{e}\|_{L^2(\beta_i)}\\&\lesssim h_i^{1/2}\|[\underline{\sigma}_h]\cdot\underline{n}\|_{L^2(\beta_i)}\left(h_i^{-1}\|e-\overline{e}\|_{L^2(V_i)}+\|\nabla e\|_{L^2(V_i)}\right).\end{aligned}$$

Further, Poincaré's inequality for $x_i\in N_h^0$ (in which case $\int_{V_i}(e-\overline{e})\,dx=0$) or Friedrichs' inequality for $x_i\in N_h\setminus N_h^0$ (in which case $\overline{e}=0$ on V_i and $e=0$ on $\partial V_i\cap\Gamma_D$) shows that

$$h_i^{-1}\|e-\overline{e}\|_{L^2(V_i)}\lesssim\|\nabla e\|_{L^2(V_i)}.\tag{3.15}$$

Substituting the last result into the preceding inequality yields

$$\int_{\beta_i} [\underline{\sigma}_h] \cdot \underline{n}(e - \overline{e})\, ds \lesssim \|h^{1/2}\, [\underline{\sigma}_h] \cdot \underline{n}\|_{L^2(\beta_i)} \|\nabla e\|_{L^2(V_i)}$$

for all $x_i \in N_h$. A summation over all vertices yields the assertion. □

Below we establish that the sum of the quantities η_R, η_E, and η_N gives a reliable estimate for the error in the global energy norm.

Theorem 3.1. *There holds*

$$\|e\|_a \lesssim \eta_R + \eta_E + \eta_N.$$

Proof. The identity (3.12) of Lemma 3.1 represents $\|e\|_a^2$ as a sum of four terms. We bound the first term using Cauchy's inequality, the second one using Lemma 3.2, and the remaining two terms using again the Cauchy's inequality:

$$\|e\|_a^2 \lesssim \eta_R \|h^{-1}(e - \overline{e})\| + \eta_E \|\nabla e\| + \eta_N \|h^{-1/2}(e - \overline{e})\|_{L^2(\Gamma_N)}.$$

Inequality (3.15) is combined with the trace inequality (3.14) to obtain

$$\|h^{-1/2}(e - \overline{e})\|^2_{L^2(\Gamma_N)} + \|h^{-1}(e - \overline{e})\|^2 \lesssim \sum_{x_i \in N_h} \left(h_i^{-2} \|e - \overline{e}\|^2_{L^2(V_i)} + \|\nabla e\|^2_{L^2(V_i)} \right) \lesssim \|\nabla e\|^2.$$

Condition (a) of Subsection 2.1 yields $\|\nabla e\| \lesssim \|e\|_a$. This and the preceding two inequalities conclude the proof of the theorem. □

Now we introduce an error estimator that is based on local averaging (post-processing) of the "total flux" $\underline{\sigma}_h$. For finite element approximations this estimator, often called ZZ-estimator, has been justified by Carstensen and Bartels [8, 12] and Rodriguez [28].

Definition 3.2. *Let P_i be the L^2-projection onto the affine functions on V_i. We define the error indicator η_Z for $A(x)$ and $\underline{b}(x)$ smooth over the volumes $V_i \in \mathcal{T}^*$ as*

$$\eta_Z := \left(\sum_{x_i \in N_h} \|\underline{\sigma}_h - P_i \underline{\sigma}_h\|^2_{L^2(V_i)} \right)^{1/2}.$$

Remark 3.1. *In our numerical experiments we have allowed $A(x)$ to have jumps that are aligned with the partition $\mathcal{T}$. In such cases we have changed the projection P_i. For example, if $V_i = V_i^1 \cup V_i^2$ and $A(x)$ is smooth on V_i^1 and V_i^2 but has jump across their interface, then P_i is defined in a piecewise way*

$$||\underline{\sigma}_h - P_i \underline{\sigma}_h||^2_{L^2(V_i)} = ||\underline{\sigma}_h - P_i^1 \underline{\sigma}_h||^2_{L^2(V_i^1)} + ||\underline{\sigma}_h - P_i^2 \underline{\sigma}_h||^2_{L^2(V_i^2)},$$

where P_i^1 and P_i^2 are the L^2-projections on the affine functions on V_i^1 and V_i^2, respectively.

Lemma 3.3. *There holds*

$$h_i^{1/2} \|[\underline{\sigma}_h] \cdot \underline{n}\|_{L^2(\beta_i)} \lesssim \|\underline{\sigma}_h - P_i \underline{\sigma}_h\|_{L^2(V_i)} + h.o.t. \quad \forall\, V_i \in \mathcal{T}^*. \tag{3.16}$$

As a corollary we have that $\eta_E \lesssim \eta_Z + h.o.t.$ The multiplicative constants in the notation $\lesssim$ depend on the shape of the elements in $\mathcal{T}$ and the shape of the control volumes in $\mathcal{T}^*$, while the h.o.t. depends on the smoothness of A and $\underline{b}$.

Proof. If A and $\underline{b}$ are polynomials (or more general, belong to a finite dimensional space) then $\underline{\sigma}_h|_K$ is in a finite dimensional space for any $K \in \mathcal{T}$. In this case we easily prove (3.16) without *h.o.t.* by an equivalence-of-norm argument on finite dimensional spaces. Namely, both sides of (3.16) define semi-norms for finite dimensional $\underline{\sigma}_h$. If $\|\underline{\sigma}_h - P_i\underline{\sigma}_h\|_{L^2(V_i)} = 0$ for some $\underline{\sigma}_h$, then $\underline{\sigma}_h = P_i\underline{\sigma}_h$ on V_i. Since $P_i\underline{\sigma}_h$ is linear on V_i, this shows that $\underline{\sigma}_h$ is also linear. Therefore, the jump $[\underline{\sigma}_h]$ is zero on β_i, i.e., the left-hand side of (3.16) vanishes as well. This proves that the semi-norm on the right-hand side is stronger than the semi-norm on the left-hand side and so proves (3.16). A scaling argument shows that the multiplicative constant behind $\lesssim$ is independent of h_i.

The case when A and $\underline{b}$ are smooth functions but $\underline{\sigma}_h|_K$ is not finite dimensional over $K \in \mathcal{T}$ is treated using approximation. Namely, we introduce finite dimensional approximations $\overline{\underline{\sigma}}_h$ of $\underline{\sigma}_h$ for any $K \in \mathcal{T}$ based on approximations of A and $\underline{b}$, take into account that

$$||\underline{\sigma}_h - \overline{\underline{\sigma}}_h||_{L^2(V_i)} = h.o.t. \quad \text{and} \quad ||[\underline{\sigma}_h - \overline{\underline{\sigma}}_h] \cdot \underline{n}||_{L^2(V_i)} = h.o.t.,$$

and use the result for the finite dimensional case to get (3.16).

By squaring (3.16) and summing over all $x_i \in N_h$ we get $\eta_E \lesssim \eta_Z + h.o.t.$ □

Theorem 3.2. *Assume that $f \in H^1(\Omega)$. Then, there holds*

$$\|e\|_a \lesssim \eta_Z + \eta_N + h.o.t. \tag{3.17}$$

Remark 3.2. *The finite volume scheme at hand is of first order. Therefore, if the considered differential problem has sufficiently regular solution then $||e||_a \lesssim h$. Terms of order higher than first are denoted by h.o.t. Let $\Omega_D := \cup\{V_i : x_i \in N_h \cap \Gamma_D\}$ be a strip of width h around Γ_D. Then for $f \in H^1(\Omega)$ we apply Ilin's inequality (2.3), $||f||_{L^2(\Omega_D)} \lesssim h^{1/2}||f||_{H^1(\Omega)}$, so that*

$$\|h^2\ \nabla(\gamma u_h)\| + \|h\ f\|_{L^2(\Omega_D)} + \|h^2\ \nabla f\| = h.o.t..$$

Proof of Theorem 3.2. To prove the theorem we use again the error representation from Lemma 3.1. In Theorem 3.1 we have bounded the third and fourth sum from the error representation by $\eta_N||\nabla e||$ and the second sum by $\eta_E||\nabla e||$. Further, η_E was bounded in Lemma (3.3) by $\eta_Z + h.o.t.$, so it remains to establish the bound

$$(f - \nabla \cdot \underline{\sigma}_h - cu_h,\ e - \overline{e}) \lesssim (\eta_Z + h.o.t.)||\nabla e||.$$

For $x_i \in N_h^0$ denote by $\overline{f}$ and $\overline{cu_h}$ the integral means over V_i of f and cu_h, respectively. Then we have

$$\begin{aligned}\int_{V_i} (f - \nabla_h \cdot \underline{\sigma}_h - \gamma u_h)(e - \overline{e})\,dx &= \int_{V_i} (f - \overline{f})(e - \overline{e})\,dx \\ &\quad - \int_{V_i} \nabla_h \cdot (\underline{\sigma}_h - P_i\underline{\sigma}_h)(e - \overline{e})\,dx - \int_{V_i} (\gamma u_h - \overline{\gamma u_h})(e - \overline{e})\,dx \\ &\le \|e - \overline{e}\|_{L^2(V_i)} \Big(\|f - \overline{f}\|_{L^2(V_i)} + \|\nabla_h \cdot (\underline{\sigma}_h - P_i\underline{\sigma}_h)\|_{L^2(V_i)} + \|\gamma u_h - \overline{\gamma u_h}\|_{L^2(V_i)} \Big).\end{aligned} \tag{3.18}$$

Poincaré's inequality gives

$$\begin{aligned}\|e - \overline{e}\|_{L^2(V_i)} &\lesssim h_i \|\nabla e\|_{L^2(V_i)}, \\ \|f - \overline{f}\|_{L^2(V_i)} &\lesssim h_i \|\nabla f\|_{L^2(V_i)}, \\ \|\gamma u_h - \overline{\gamma u_h}\|_{L^2(V_i)} &\lesssim h_i \|\nabla(\gamma u_h)\|_{L^2(V_i)}.\end{aligned} \tag{3.19}$$

The term $\|\nabla_h \cdot (\underline{\sigma}_h - P_i\underline{\sigma}_h)\|_{L^2(V_i)}$ is treated by the inverse estimate

$$\|\nabla_h \cdot (\underline{\sigma}_h - P_i\underline{\sigma}_h)\|_{L^2(V_i)} \lesssim h_i^{-1}\|\underline{\sigma}_h - P_i\underline{\sigma}_h\|_{L^2(V_i)} + h.o.t. \tag{3.20}$$

As in the proof of Lemma 3.3, we first prove (3.20) when $\underline{\sigma}_h$ is finite dimensional by equivalence-of-norms followed by a scaling argument, and then for the general case by a perturbation analysis. The combination of (3.18)–(3.20) shows

$$\begin{aligned}\int_{V_i} (f - \nabla_h \cdot \underline{\sigma}_h - \gamma u_h)(e - \overline{e})\, dx \\ \lesssim \|\nabla e\|_{L^2(V_i)} \left(\|h^2\, \nabla f\|_{L^2(V_i)} + \|h^2\, \nabla(\gamma u_h)\|_{L^2(V_i)} + \|\underline{\sigma}_h - P_i\underline{\sigma}_h\|_{L^2(V_i)} + h.o.t.\right).\end{aligned} \tag{3.21}$$

So far (3.21) holds for $x_i \in N_h^0$. For $x_i \in N_h \cap \Gamma_D$ we replace $\overline{e}$, $\overline{f}$, and $\overline{\gamma u_h}$ by zero and deduce the first and third inequalities of (3.19) from Friedrichs' inequality (notice that e and γu_h vanish on $\Gamma_D \cap V_i$). The inverse estimate (3.20) holds for $x_i \in N_h \cap \Gamma_D$ as well. The aforementioned arguments prove (3.21) with $\|h^2\, \nabla f\|_{L^2(V_i)}$ replaced by $\|h\, f\|_{L^2(V_i)}$. This shows

$$(f - \nabla_h \cdot \underline{\sigma}_h - \gamma u_h, e - \overline{e}) \lesssim (\eta_Z + \|h^2 \nabla(\gamma u_h)\| + \|h^2 \nabla f\| + \|hf\|_{L^2(\Omega_D)} + h.o.t.)||\nabla e||.$$

The last result, the discussion at the beginning of the theorem, Remark 3.2, and the ellipticity assumption conclude the proof of the theorem. □

The aforementioned error estimates are sharp. The converse estimates hold even in a more local form than displayed.

Theorem 3.3. *There holds*

$$\eta_Z + \eta_R + \eta_E + \eta_N \lesssim \|e\|_a + h.o.t.$$

Proof. We will prove that the quantities η_R, η_E, η_N, and η_Z are bounded by $C\ \|e\|_a + h.o.t.$ The *h.o.t.* appears by applying averaging techniques as in the proof of Lemma 3.3 and therefore we will consider only the case when $\underline{\sigma}_h$ is finite dimensional. First, we will bound the contributions to η_N due to Γ_N^{in}, namely, we will prove

$$||h^{1/2}\, (g - \underline{\sigma}_h \cdot \underline{n})||_{L^2(\Gamma_N^{in})} \lesssim ||e||_a + h.o.t. \tag{3.22}$$

We consider an element $K \in \mathcal{T}$ that has an edge/face $E \subset \Gamma_N^{in}$. We will use the pair $(K,\, E)$ in the rest of the proof (see Figure 3.1).

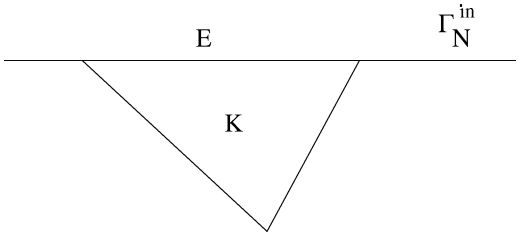

FIGURE 3. The pair $(K,\, E)$ of edge $E \subset \Gamma_N^{in}$ and element K used in the proof of inequality (3.22).

First, we note that

$$h_E^{1/2}\|g-\overline{g}\|_{L^2(E)} = h.o.t. \quad \text{for} \quad \overline{g} := \int_E g\,ds/|E|.$$

Then

$$\|g-\underline{\sigma}_h\cdot\underline{n}\|_{L^2(E)} \le \|g-\overline{g}\|_{L^2(E)} + \|\overline{g}-\underline{\sigma}_h\cdot\underline{n}\|_{L^2(E)} \lesssim \|\overline{g}-\underline{\sigma}_h\cdot\underline{n}\|_{L^2(E)} + \ h.o.t.$$

We prove below that

$$\|\overline{g}-\underline{\sigma}_h\cdot\underline{n}\|_{L^2(E)} \lesssim h_E^{-1/2}\|\underline{\sigma}-\underline{\sigma}_h\|_{L^2(K)} + h.o.t.,$$

so that summation over all $E\subset\Gamma_N^{in}$ yields (3.22).

Consider an edge-bubble function $b_E\in H^1(\Omega)$, $b_E\ge 0$, $b_E(x)=0$ on $\Omega\setminus K$ and $\partial K\setminus E$, with properties

$$\int_E b_E\,ds = \int_E ds,\ \|b_E\|_{L^\infty(K)}\lesssim 1,\ \|\nabla b_E\|_{L^\infty(K)}\lesssim 1/h_E. \tag{3.23}$$

A 2-D example of such bubble is $b_E = 6\phi_1\phi_2$, where ϕ_1 and ϕ_2 are the standard linear nodal basis functions associated with the end points of the edge E. Let $z\in H^1(K)$ be the harmonic extension of $(\overline{g}-\underline{\sigma}_h\cdot\underline{n})b_E$ from ∂K to K. The extension is bounded in H^1 [27, Theorem 4.1.1] on a reference element $\hat{K}$ by the $H^{1/2}(\hat{E})$ norm of the extended quantity, and since all norms are equivalent on a finite dimensional space, by its $L^2(\hat{E})$ norm. Therefore, a scaling argument gives

$$h_E^{1/2}\|\nabla z\|_{L^2(K)} + h_E^{-1/2}\|z\|_{L^2(K)} \lesssim \|b_E(\overline{g}-\underline{\sigma}_h\cdot\underline{n})\|_{L^2(E)}. \tag{3.24}$$

We define the linear operator P_K into the space of polynomials of degree 2 on an element $K\in\mathcal{T}$ as

$$(b_K\ P_K z, p_h)_{L^2(K)} = (z,p_h)_{L^2(K)}$$

for all polynomials p_h of degree 2. Here $b_K\in H^1(\Omega)$, $b_K\ge 0$ is an element-bubble function with properties

$$\text{supp } b_K\subset K,\ \int_K b_K\,ds = \int_K ds,\ \|b_K\|_{L^\infty(K)}\lesssim 1,\ \|\nabla b_K\|_{L^\infty(K)}\lesssim 1/h_K.$$

A 2-D example of such bubble is $b_K = 60\phi_1\phi_2\phi_3$, where ϕ_1, ϕ_2, and ϕ_3 are the standard linear nodal basis functions associated with the vertices of the element K. Then $\tilde{z} := z - b_K\,P_K z$ by construction has the properties

$$\begin{aligned}&\tilde{z} = (\overline{g}-\underline{\sigma}_h\cdot\underline{n})b_E \quad \text{on } E,\ \tilde{z}=0 \quad \text{on } \partial K\setminus E,\\ &(\tilde{z},\ p_h)_{L^2(K)} = 0 \quad \text{for all polynomials } p_h \text{ of degree 2.}\end{aligned}$$

Inequality (3.24) remains valid for z replaced by $\tilde{z}$, because of the following. Choosing $p_h = P_K z$ in the definition of P_K yields

$$\|b_K^{1/2}P_K z\|^2_{L^2(K)} = (z,P_K z)_{L^2(K)} \lesssim \|z\|_{L^2(K)}\|P_K z\|_{L^2(K)}.$$

We use norm equivalence on finite dimensional spaces on a reference element and scaling to K to get that the quantities $||b_K P_K z||_{L^2(K)}$, $\|b_K^{1/2}P_K z\|_{L^2(K)}$, and $\|P_K z\|_{L^2(K)}$ are equivalent

up to constants independent of h, and therefore $\|b_K P_K z\|_{L^2(K)} \lesssim \|z\|_{L^2(K)}$. We use again the equivalence of norms argument, inverse inequality, and the properties of z to get that

$$\begin{aligned}\|\nabla(b_K P_K z)\|_{L^2(K)} &\lesssim \|\nabla b_K\|_{L^2(K)}\|P_K z\|_{L^2(K)} + \|b_K \nabla(P_K z)\|_{L^2(K)}\\ &\lesssim h_E^{-1}\|P_K z\|_{L^2(K)} + h_E^{-1}\|z\|_{L^2(K)}\\ &\lesssim h_E^{-1/2}\|b_E(\overline{g} - \underline{\sigma}_h\cdot\underline{n})\|_{L^2(E)}.\end{aligned}$$

Combined with the bound for $\|b_K P_K z\|_{L^2(K)}$, this completes the proof of (3.22) for $z = \tilde{z}$.

Given a polynomial p_h of degree 2, using the Gauß divergence theorem and the properties of $\tilde{z}$, we deduce

$$\begin{aligned}\int_E b_E(\overline{g} - \underline{\sigma}_h\cdot\underline{n})(\underline{\sigma} - \underline{\sigma}_h)\cdot\underline{n}\,ds &= \int_{\partial K} \tilde{z}(\underline{\sigma} - \underline{\sigma}_h)\cdot\underline{n}\,ds\\ &= \int_K (\underline{\sigma} - \underline{\sigma}_h)\cdot\nabla\tilde{z}\,dx + \int_K \tilde{z}(\nabla\cdot(\underline{\sigma} - \underline{\sigma}_h) - p_h)\,dx\\ &\lesssim \Big(\|\underline{\sigma} - \underline{\sigma}_h\|_{L^2(K)} + h_E\|\nabla\cdot(\underline{\sigma} - \underline{\sigma}_h) - p_h\|_{L^2(K)}\Big) h_E^{-1/2}\|b_E(\overline{g} - \underline{\sigma}_h\cdot\underline{n})\|_{L^2(E)}.\end{aligned}$$

Choosing proper p_h in the second term of the last inequality makes that term *h.o.t.* Indeed, write down first the equality (see the basic problem (1.1))

$$\nabla\cdot(\underline{\sigma} - \underline{\sigma}_h) - p_h = cu - f - (\nabla\cdot A)\cdot\nabla u_h) + u_h\,\nabla\cdot\underline{b} + \underline{b}\cdot\nabla u_h - p_h. \tag{3.25}$$

Here $\nabla\cdot A$ is understood as a vector with components divergence of the rows of $A(x)$. Let $\tilde{f}$, $\widetilde{cu}$, $\widetilde{\nabla\cdot\underline{b}}$, $\tilde{\underline{b}}$, and $\widetilde{\nabla\cdot A}$ are the linear approximations on K of f, cu, $\nabla\cdot\underline{b}$, $\underline{b}$, and $\nabla\cdot A$, respectively.

Now, we choose p_h to be the following polynomial of degree two on K

$$p_h = \widetilde{cu} - \tilde{f} - (\widetilde{\nabla\cdot A})\cdot\nabla u_h + u_h\widetilde{\nabla\cdot\underline{b}} + \tilde{\underline{b}}\cdot\nabla u_h,$$

take the $L^2(K)$ norm of (3.25), and use triangle's inequality to get

$$\begin{aligned}\|\nabla\cdot(\underline{\sigma} - \underline{\sigma}_h) - p_h\|_{L^2(K)} &\le \|f - \tilde{f}\|_{L^2(K)} + \|cu - \widetilde{cu}\|_{L^2(K)} + \|u_h(\nabla\cdot\underline{b} - \widetilde{\nabla\cdot\underline{b}})\|_{L^2(K)}\\ &\quad + \|(\underline{b} - \tilde{\underline{b}})\cdot\nabla u_h\|_{L^2(K)} + \|\nabla u_h\cdot(\nabla\cdot A - \widetilde{\nabla\cdot A})\|_{L^2(K)}\\ &\lesssim \Big(\|u\|_{H^2(K)} + \|u_h\|_{H^1(K)}\Big)\ h.o.t. + \|h_K\nabla f\|_{L^2(K)}.\end{aligned}$$

Therefore, (note that $g = \underline{\sigma}\cdot\underline{n}$ on Γ_N^{in})

$$\begin{aligned}\|b_E^{1/2}(\overline{g} - \underline{\sigma}_h\cdot\underline{n})\|^2_{L^2(E)} &= \int_E \tilde{z}(\overline{g} - g)\,ds + \int_E \tilde{z}(\underline{\sigma} - \underline{\sigma}_h)\cdot\underline{n}\,ds\\ &\lesssim h_E^{-1/2}\Big(\|\underline{\sigma} - \underline{\sigma}_h\|_{L^2(K)} + h.o.t.\Big)\ \|b_E^{1/2}(\overline{g} - \underline{\sigma}_h\cdot\underline{n})\|_{L^2(E)}\end{aligned}$$

and so

$$\|b_E^{1/2}(\overline{g} - \underline{\sigma}_h \cdot \underline{n})\|_{L^2(E)} \lesssim h_E^{-1/2}\|\underline{\sigma} - \underline{\sigma}_h\|_{L^2(K)} + h.o.t.$$

Using again the equivalence-of-norms estimate (equivalence of norms on finite dimensional spaces on reference element and scaling)

$$\|\overline{g} - \underline{\sigma}_h \cdot \underline{n}\|_{L^2(E)} \lesssim \|b_E^{1/2}(\overline{g} - \underline{\sigma}_h \cdot \underline{n})\|_{L^2(E)}$$

we finally prove that

$$\begin{aligned} \|g - \underline{\sigma}_h \cdot \underline{n}\|_{L^2(E)} &\leq \|g - \overline{g}\|_{L^2(E)} + \|\overline{g} - \underline{\sigma}_h \cdot \underline{n}\|_{L^2(E)} \\ &\lesssim \|b_E^{1/2}(\overline{g} - \underline{\sigma}_h \cdot \underline{n})\|_{L^2(E)} + \ h.o.t. \\ &\lesssim h_E^{-1/2}\|\underline{\sigma} - \underline{\sigma}_h\|_{L^2(K)} + \ h.o.t. \end{aligned}$$

Similarly, $||A\nabla u_h \cdot \underline{n}||_{L^2(E)} \lesssim h_E^{-1/2}\|\underline{\sigma} - \underline{\sigma}_h\|_{L^2(K)} + \ h.o.t.$ for $E \subset \Gamma_N^{out}$, which, combined with the result for $E \subset \Gamma_N^{in}$, proves that $\eta_N \lesssim ||e||_a + \ h.o.t.$

A similar technique shows that $\eta_E \lesssim ||e||_a + \ h.o.t.$

The inequality $\eta_R \lesssim ||e||_a + h.o.t.$ can be proved in the following way. We use the average $\bar{R}_K$ of the residual $R_K := f - \nabla \cdot \underline{\sigma}_h - \gamma u_h$ over an element K to derive

$$||\bar{R}_K||_{L^2(K)} \leq ||R_K - \bar{R}_K||_{L^2(K)} + ||R_K||_{L^2(K)} = \ h.o.t. + ||R_K||_{L^2(K)}.$$

We use the technique from Lemma 3.1 to deduce the equality $(R_K, b_K\bar{R}_K)_{L^2(K)} = a(e, b_K\bar{R}_K)$ and therefore

$$(R_K, b_K\bar{R}_K)_{L^2(K)} = ||b_K^{1/2}R_K||^2_{L^2(K)} - (R_K, b_K(R_K - \bar{R}_K))_{L^2(K)} = a(e, b_K\bar{R}_K)$$

$$\lesssim ||e||_{H^1(K)}||b_K\bar{R}_K||_{H^1(K)} \lesssim ||e||_{H^1(K)}h_K^{-1}||\bar{R}_K||_{L^2(K)} \lesssim h_K^{-1}||e||_{H^1(K)}||R_K||_{L^2(K)} + h.o.t.$$

Here we used the inverse inequality and the boundedness of the coefficients of the differential equation (1.1). Then we take the term $(R_K, b_K(R_K - \bar{R}_K))_{L^2(K)}$ to the right-hand side and consider it as h.o.t. We finally use that $||b_K^{1/2}R_K||_{L^2(K)} \approx ||R_K||_{L^2(K)}$ to obtain

$$||R_K||_{L^2(K)} \lesssim h_K^{-1}||e||_{H^1(K)} + h.o.t.$$

A summation over all $K \in \mathcal{T}$ yields $\eta_R \lesssim ||e||_a + h.o.t.$

Now we prove the remaining inequality $\eta_Z \lesssim ||e||_a + h.o.t.$ Since P_i is a linear $L^2(V_i)$ projector, we have that

$$\|\underline{\sigma}_h - P_i\underline{\sigma}_h\|_{L^2(V_i)} \leq \|\underline{\sigma}_h - P_i\underline{\sigma}\|_{L^2(V_i)}.$$

Adding and subtracting $\underline{\sigma}$ in the right-hand side and applying triangle's inequality we get

$$\|\underline{\sigma}_h - P_i\underline{\sigma}\|_{L^2(V_i)} \leq \|\underline{\sigma}_h - \underline{\sigma}\|_{L^2(V_i)} + \|\underline{\sigma} - P_i\underline{\sigma}\|_{L^2(V_i)} = \|\underline{\sigma}_h - \underline{\sigma}\|_{L^2(V_i)} + h.o.t.$$

since $\|\underline{\sigma} - P_i\underline{\sigma}\|_{L^2(V_i)} = h.o.t.$ for $\underline{\sigma}$ smooth. The summation over all x_i concludes the proof of the theorem. □

3.2. **Analysis of the upwind scheme in H^1 norm.** This section is devoted to the case when an upwind approximation is applied to the the convection term, namely we consider problem (2.11).

Definition 3.3. *For an element $K \in \mathcal{T}$ we denote by $\gamma_K := \cup_{\gamma_{ij}}(K \cap \gamma_{ij})$ and set*

$$\eta_E^{up} := \left(\sum_{K\in\mathcal{T}} \sum_{\gamma_{ij}\subset\gamma_K} \|h^{1/2}\underline{b}\cdot\underline{n}\,(u_h(x_i)-u_h)\|^2_{L^2(\gamma_{ij})} \right)^{1/2},$$

$$\eta_N^{up} := \|h^{1/2}\underline{b}\cdot\underline{n}\,\nabla u_h\|_{L^2(\Gamma_N^{out})}.$$

Theorem 3.4. *Assume that $f \in H^1(\Omega)$. Then*

$$\|e\|_a \lesssim \eta_Z + \eta_N + \eta_E^{up} + \eta_N^{up} + h.o.t. \tag{3.26}$$

Proof. Since $a_h^{up}(u_h, v^*) = F(v^*)$ and $a_h(u, v^*) = F(v^*)$ for $v^* \in S_h^*$ we have the orthogonality condition: $a_h(u, v^*) - a_h^{up}(u_h, v^*) = 0$. Choosing $v^* = \bar{e}$ we get the following representation for the energy norm of the error

$$\begin{aligned}
\|e\|_a^2 &= a(e,e) - a_h(u,\bar{e}) + a_h^{up}(u_h,\bar{e}) \\
&= \{a(e,e) - a_h(e,\bar{e})\} + \{a_h^{up}(u_h,\bar{e}) - a_h(u_h,\bar{e})\} \\
&= \{a(e,e) - a_h(e,\bar{e})\} + \{C_h^{up}(u_h,\bar{e}) - C_h(u_h,\bar{e})\}.
\end{aligned}$$

For the first term, $a(e,e)-a_h(e,\bar{e})$, we use the same approach as in the analysis of the scheme without upwind (see Lemma 3.1) and show that

$$\begin{aligned}
a(e,e) - a_h(u,\bar{e}) = (f - \nabla_h\cdot\underline{\sigma}_h - \gamma u_h,\ e-\bar{e}) - \int_{\mathcal{E}} [\underline{\sigma}_h]\cdot\underline{n}\,(e-\bar{e})\,ds \\
- \int_{\Gamma_N^{in}} (g - \underline{\sigma}_h\cdot\underline{n})\,(e-\bar{e})\,ds - \int_{\Gamma_N^{out}} (A\nabla_h u_h)\cdot\underline{n}\,(e-\bar{e})\,ds.
\end{aligned} \tag{3.27}$$

This presentation allows us to use estimate (3.17) of Theorem 3.2.

For the second term, $C_h^{up}(u_h,\bar{e}) - C_h(u_h,\bar{e})$, we get

$$\begin{aligned}
C_h^{up}(u_h,\bar{e}) - C_h(u_h,\bar{e}) = \sum_{x_i\in N_h^0} \bar{e}_i \Bigg\{ \sum_{j\in\Pi(i)} \int_{\gamma_{ij}} ((\underline{b}\cdot\underline{n})_+ u_h(x_i) + (\underline{b}\cdot\underline{n})_- u_h(x_j) - \underline{b}\cdot\underline{n} u_h)\,ds \\
+ \int_{\partial V_i\cap\Gamma_N^{out}} (\underline{b}\cdot\underline{n}\,u_h(x_i) - \underline{b}\cdot\underline{n}\,u_h)\,ds \Bigg\}.
\end{aligned}$$

Here the unit normal vector $\underline{n}$ on γ_{ij} is oriented in such a way that $\underline{b}\cdot\underline{n} \geq 0$. We want to express the above sum as sum over the elements. To do so we specify that the indexes (ij)

are oriented so that $(x_i - x_j)\cdot \underline{n} \leq 0$. We get that

$$C_h^{up}(u_h,\bar{e}) - C_h(u_h,\bar{e}) = \sum_{K\in\mathcal{T}_h} \left\{ \sum_{\gamma_{ij}\subset K} (\bar{e}_i - \bar{e}_j) \int_{\gamma_{ij}} \underline{b}\cdot\underline{n}\,(u_h(x_i) - u_h)\,ds + \sum_{V_i\cap K} \bar{e}_i \int_{\partial V_i\cap\Gamma_N^{out}} \underline{b}\cdot\underline{n}\,(u_h(x_i) - u_h)\,ds \right\}.$$

We denote by $[\bar{e}] := e_i - \bar{e}_j$ the jump of $\bar{e}$ across γ_{ij} and take into account that $[\bar{e} - e] = [\bar{e}]$. Then, by Schwartz inequality, the term involving integral over γ_{ij} is bounded by $C||[e - \bar{e}]||_{L^2(\gamma_{ij})}\ ||\underline{b}\cdot\underline{n}\,(u_h(x_i) - u_h)||_{L^2(\gamma_{ij})}$. As before, using trace, Poincaré's, and/or Friedrichs' inequalities we get

$$||[e - \bar{e}]||_{L^2(\gamma_{ij})} \lesssim h_i^{1/2}\|\nabla e\|_{L^2(V_i)},$$

which bounds the integrals over γ_{ij} in the error representation with η_E^{up}.

For the terms involving integration over Γ_N^{out} we have

$$|u_h(x_i) - u_h(x)| \leq |\nabla u_h\cdot\underline{t}(x)|.|x_i - x|.$$

Here $\underline{t}(x)$ is a unit vector along $\partial V_i\cap K$, an edge in 2-D or a face in 3-D. Then in 2-D $\underline{t}$ is simply a unit vector perpendicular to $\underline{n}$, while in 3-D $\underline{t}(x)$ depends on the position of x on the face and is again perpendicular to $\underline{n}$. In both cases $|u_h(x_i) - u_h(x)| \leq |h_K\nabla u_h|$. Using Schwarz inequality we bound the term involving integration over Γ_N^{out} in the following way

$$\sum_{V_i\cap K} \bar{e}_i \int_{\partial V_i\cap\Gamma_N^{out}} \underline{b}\cdot\underline{n}\,(u_h(x_i) - u_h)\,ds \leq C||\nabla e||.||h^{1/2}\underline{b}\cdot\underline{n}\,\nabla u_h||_{L^2(\Gamma_N^{out})},$$

which eventually gives the term η_N^{up} in (3.26) and completes the proof. □

3.3. **Error estimates in L^2.** We use dual techniques to get error estimators for different quantities of the error. In this subsection we will show how to use the duality technique in order to derive an error estimator in the global $L^2(\Omega)$-norm for the scheme without upwind. The main assumption in this section is that the solution of problem (1.1) is H^2 regular.

Definition 3.4. *We define the residual L^2 a posteriori error estimator $\tilde{\rho}$ as*

$$\tilde{\rho} := (\tilde{\eta}_R^2 + \tilde{\eta}_E^2 + \tilde{\eta}_N^2)^{1/2}, \tag{3.28}$$

where

$$\begin{aligned}
\tilde{\eta}_R^2 &:= ||h\,(R_K - \bar{R}_K)||^2 + ||h^2\,R_K||^2,\\
\tilde{\eta}_E^2 &:= ||h^{1/2}\,(R_E - \bar{R}_E)||^2_{L^2(\mathcal{E})} + ||h^{3/2}\,R_E||^2_{L^2(\mathcal{E})},\\
\tilde{\eta}_N^2 &:= ||h^{1/2}\,(R_E^{in} - \bar{R}_E^{in})||^2_{L^2(\Gamma_N^{in})} + ||h^{3/2}\,R_E^{in}||^2_{L^2(\Gamma_N^{in})}\\
&\quad + ||h^{1/2}\,(R_E^{out} - \bar{R}_E^{out})||^2_{L^2(\Gamma_N^{out})} + ||h^{3/2}\,R_E^{out}||^2_{L^2(\Gamma_N^{out})},
\end{aligned}$$

and $\bar{R}_K$, $\bar{R}_E$, $\bar{R}_E^{in}$, *and* $\bar{R}_E^{out}$ *are the* $K \in \mathcal{T}$, $E \in \mathcal{E}$, $E \in \Gamma_N^{in}$, *and* $E \in \Gamma_N^{out}$ *piecewise mean values of, correspondingly,* R_K, R_E, R_E^{in}, *and* R_E^{out} *introduced in Definition 3.1.*

Our aim is to show that the estimator $\tilde{\rho}$ is reliable in the $L^2(\Omega)$ norm. The a posteriori $L^2(\Omega)$ error analysis involves the continuous dual problem: Find $\tilde{e} \in H_D^1(\Omega)$ such that

$$a(v, \tilde{e}) = (e, v) \text{ for any } \ v \in H_D^1(\Omega), \tag{3.29}$$

where e is the exact error, defined as before.

Theorem 3.5. *Let the solution* $\tilde{e}$ *of the dual problem (3.29) be* $H^2(\Omega)$*-regular. If the coefficients of our basic problem (1.1) are sufficiently regular, namely* R_K, R_E, R_E^{in}, *and* R_E^{out} *are correspondingly in* $H^1(K)$, $H^{1/2}(E)$, $H^{1/2}(\Gamma_N^{in})$, *and* $H^{1/2}(\Gamma_N^{out})$, *then the residual* L^2 *a posteriori error estimator (3.28) from Definition 3.4 is reliable, i.e.,* $||e|| \lesssim \tilde{\rho}$.

Proof. Let $v = e$ in (3.29) and argue as in the proof of Lemma 3.1 to show

$$\begin{aligned} ||e||^2 = a(e, \tilde{e}) \ &= \ (R_K, \tilde{e} - e^*) - (R_E, \tilde{e} - e^*)_{L^2(\mathcal{E})} \\ &\quad - (R_E^{in}, \tilde{e} - e^*)_{L^2(\Gamma_N^{in})} - (R_E^{out}, \tilde{e} - e^*)_{L^2(\Gamma_N^{out})}, \end{aligned} \tag{3.30}$$

for an arbitrary $e^* \in S_h^*$. To evaluate the right hand side of this identity we use the nodal interpolation operator I_h and its properties. If $\tilde{e} \in H^2(\Omega)$ the Sobolev inequalities [10, Theorem 4.3.4] guarantee that $I_h\tilde{e}$ is well defined. The properties of the interpolant are well established in the finite element literature (see, for example [10]), namely,

$$h_K^{-2}||\tilde{e} - I_h\tilde{e}||_{L^2(K)} + h_K^{-1}|\tilde{e} - I_h\tilde{e}|_{H^1(K)} + h_K^{-3/2}||\tilde{e} - I_h\tilde{e}||_{L^2(\partial K)} \le C_{I,K}|\tilde{e}|_{H^2(K)}. \tag{3.31}$$

Now, in equation (3.30), we choose $e^* = I_h^* I_h \tilde{e}$ so that $\tilde{e} - e^* = (\tilde{e} - I_h\tilde{e}) + (I_h\tilde{e} - I_h^* I_h\tilde{e})$. Further, we apply Schwarz inequality on the integrals involving $\tilde{e} - I_h\tilde{e}$ and use (3.31) to get the bound

$$\begin{aligned} &(R_K, \tilde{e} - I_h\tilde{e}) - (R_E, \tilde{e} - I_h\tilde{e})_{L^2(\mathcal{E})} - (R_E^{in}, \tilde{e} - I_h\tilde{e})_{L^2(\Gamma_N^{in})} - (R_E^{out}, \tilde{e} - I_h\tilde{e})_{L^2(\Gamma_N^{out})} \\ &\qquad \lesssim \left(||h^2 \ R_K|| + ||h^{3/2} \ R_E||_{L^2(\mathcal{E})} + ||h^{3/2} \ R_E^{in}||_{L^2(\Gamma_N^{in})} + ||h^{3/2} \ R_E^{out}||_{L^2(\Gamma_N^{out})} \right) |\tilde{e}|_{H^2(\Omega)}. \end{aligned}$$

For the integrals involving $I_h\tilde{e} - I_h^* I_h\tilde{e}$ we first note that if K is a fixed element in $\mathcal{T}$, then for every vertex x_i of K, the quantities $|K \cap V_i|$ (volume in 3-D and area in 2-D) are equal. Also, for vertices x_i on the face/edge E we have that the boundary quantities $|E \cap V_i|$ (area in 3-D and length in 2-D) are also equal. Therefore,

$$\int_K (I_h\tilde{e} - I_h^* I_h\tilde{e}) dx = 0, \qquad \int_E (I_h\tilde{e} - I_h^* I_h\tilde{e}) ds = 0.$$

We apply the last fact to the integrals involving $I_h\tilde{e} - I_h^* I_h\tilde{e}$ in order to subtract from R_K, R_E, R_E^{in}, and R_E^{out} their mean values $\bar{R}_K$, $\bar{R}_E$, $\bar{R}_E^{in}$, and $\bar{R}_E^{out}$. Then, using Schwarz and

Poincaré's inequalities we bound the term involving $I_h\tilde{e} - I_h^* I_h\tilde{e}$, namely,

$$\begin{aligned}
&|(R_K, I_h^* I_h\tilde{e} - I_h\tilde{e}) - (R_E, I_h^* I_h\tilde{e} - I_h\tilde{e})_{L^2(\mathcal{E})} \\
&\qquad -(R_E^{in}, I_h^* I_h\tilde{e} - I_h\tilde{e})_{L^2(\Gamma_N^{in})} - (R_E^{out}, I_h^* I_h\tilde{e} - I_h\tilde{e})_{L^2(\Gamma_N^{out})}| \\
&\lesssim \Big(||h\ (R_K - \bar{R}_K)|| + ||h^{1/2}(R_E - \bar{R}_E)||_{L^2(\mathcal{E})} \\
&\qquad + ||h^{1/2}\ (R_E^{in} - \bar{R}_E^{in})||_{L^2(\Gamma_N^{in})} + ||h^{1/2}\ (R_E^{out} - \bar{R}_E^{out})||_{L^2(\Gamma_N^{out})} \Big) ||\tilde{e}||_{H^2(K)},
\end{aligned}$$

where we have used the inequality

$$\begin{aligned}
||I_h\tilde{e} - I_h^* I_h\tilde{e}||_{L^2(K)} &\lesssim h_K |I_h\tilde{e}|_{H^1(K)} \lesssim h_K |\tilde{e} - I_h\tilde{e}|_{H^1(K)} + h_K |\tilde{e}|_{H^1(K)} \\
&\lesssim h_K^2 |\tilde{e}|_{H^2(K)} + h_K |\tilde{e}|_{H^1(K)} \lesssim h_K |\tilde{e}|_{H^2(K)}.
\end{aligned}$$

Applying the above estimates, the stability of the dual problem with respect to the right hand side, $||\tilde{e}||_{H^2(\Omega)} \leq C||e||$, and obvious manipulations, we get that the L^2 a posteriori error estimator $\tilde{\rho}$ is reliable. Moreover, since the coefficients of (1.1) are sufficiently regular we can apply Poincaré's inequality to the terms $||R_K - \bar{R}_K||_{L^2(K)}$, $||R_E - \bar{R}_E||_{L^2(E)}$, $||R_E^{in} - \bar{R}_E^{in}||_{L^2(E)}$, and $||R_E^{out} - \bar{R}_E^{out}||_{L^2(E)}$ to get one additional power of h that will make the error estimator of second order.

Note that we did not explicitly apply the Poincaré's inequality in the definition of the error estimator in order to make it well defined for problems with less that the stated in the theorem regularity. □

4. Adaptive grid refinement and solution strategy

In this section we present the adaptive mesh refinement strategy that we use. For a given finite element partitioning $\mathcal{T}$, desired error tolerance ρ, and norm in which the tolerance to be achieved, say $|||\cdot|||$, do the following :

- compute the finite volume approximation $u_h \in S_h$, as given in Subsection 2.2;
- using the a posteriori error analysis, compute the errors ρ_K for all $K \in \mathcal{T}$;
- mark those finite elements K for which $\rho_K \geq \rho/\sqrt{N}$; here N is the number of elements in $\mathcal{T}$;
- if $\sum_{K\in\mathcal{T}} \rho_K^2 > \rho^2$, then refine the marked elements;
- additionally refine until a conforming mesh is reached;
- repeat the above process until no elements have been refined.

For the 2-D case we refine marked elements by uniformly splitting the marked triangles into four. The refinement to conformity is done by bisection through the longest edge. For the 3-D version of the code the elements (tetrahedrons) are refined using the algorithm described by D. Arnold, et.al. in [3].

The described procedure yields error control and optimal mesh (heuristics), which are the goals in the adaptive algorithm. The obtained in the process nested meshes are used to define multilevel preconditioners. The initial guess for every new level is taken to be the interpolation of u_h from the previous level.

5. Numerical examples

Here we present two sets of numerical examples to test the our theoretical results. The first two examples are simple 2-dimensional elliptic problems while the remaining tests illustrate our approach on 3-dimensional problems of flow and transport in porous media.

5.1. **2-dimensional test problems.** In Example 1 we consider problems with known solutions and compare the behavior of the error estimators with the exact errors. Example 2 is for discontinuous matrix $A(x)$ with unknown solution.

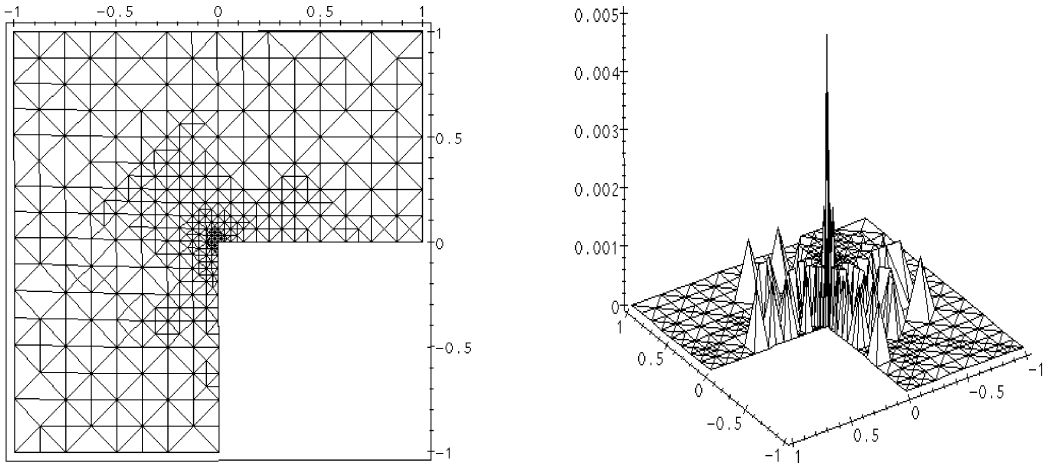

Figure 4. Locally refined mesh and the corresponding error after 4 levels of refinement

Example 1. We consider three Dirichlet problems for the Poisson equation on an L-shaped domain with known exact solutions $u = r^{4/3} sin\frac{4\theta}{3}$ (Problem 1), $u = r^{2/3} sin\frac{2\theta}{3}$ (Problem 2), and $u = r^{1/2} sin\frac{\theta}{2}$ (Problem 3). These functions belong to $H^{1+s}(\Omega)$ with s almost $4/3$, $2/3$ and $1/2$, respectively. On Figure 4 we show the mesh and the error for Problem 2 after 4 level of local refinement.

The theory gives that the a posteriori error estimators η_E and η_Z are equivalent to the H^1-norm of the error. This theoretical result is confirmed by our the computations which are summarized on Figure 5. The left picture gives the exact error (solid line) and the a posteriori error estimators η_Z (dashed line) and η_E (dash-dotted line) for the three problems over the different levels of the mesh. The levels are obtained by uniform refinement (splitting every triangle into 4) and have correspondingly 65, 255, 833, 3201, 12545, 49665, and 197633 nodes. The errors are printed in logarithmic scale in order to demonstrate the linear over the levels error reduction. For exact solutions in $H^{1+1/2-\epsilon}$, $H^{1+2/3-\epsilon}$, and $H^{1+4/3-\epsilon}$ ($\epsilon > 0$) one can see the theoretically expected rate of error reduction over the levels of $1/2$, $2/3$, and 1, correspondingly. One can observe that both η_Z and η_E are equivalent to the exact error, as proved in the theoretical section. The same is true when the local refinement method from Section 4 is applied (see Figure 5 (Center)). The error tolerances, supplied to the refinement procedures for the three problems, are the exact errors on level 7 of the uniformly refined mesh (i.e., the one with $197,633$ nodes). Refinement based on η_Z leads to final meshes of 169618, 8365, and 2508 for Problems 1, 2, and 3, respectively (see Figure 5

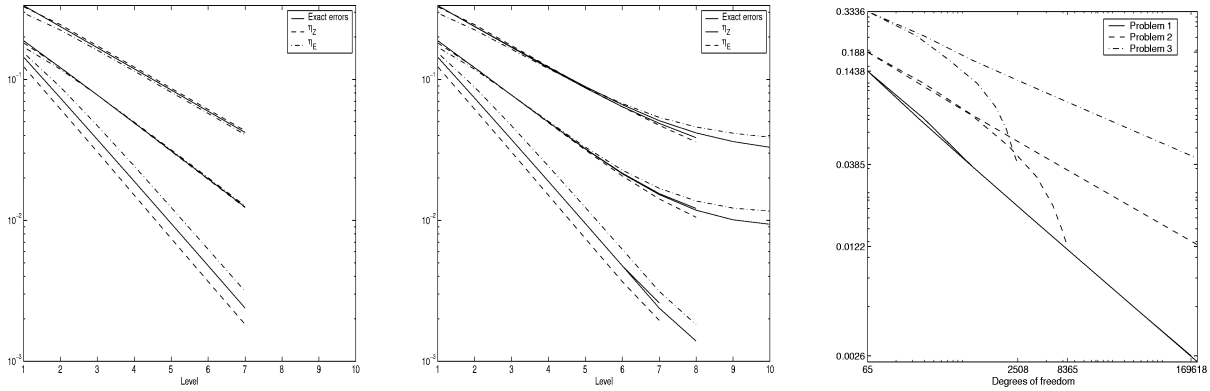

FIGURE 5. Comparison of the H^1-norm of the error for solutions in $H^{1+1/2-\epsilon}$ (Problem 3), $H^{1+2/3-\epsilon}$, (Problem 2), and $H^{1+4/3-\epsilon}$, (Problem 1) on a sequence of globally refined grids and for grids refined locally by using the a posteriori error estimates: Left: Exact error, η_Z, and η_E for uniformly refined grids; Center: Locally refined meshes over different mesh levels; Right: Accuracy versus the degrees of freedom for globally and locally refined meshes.

(Right)). Refinement based on η_E leads to final meshes of 1986, 10097, and 581852 nodes for Problems 3, 2, and 1, correspondingly. Note the difference in the order of the mesh sizes for uniform refinement and local refinement for the first two problems. Note also that the error reduction for locally refined grids over different levels is the same as for uniformly refined grids. For the third problem we have full elliptic regularity and η_Z/η_E are supposed to lead to uniform refinement, which is confirmed by the numerical experiment.

Example 2. We consider problem (1.1) with Ω shown on Figure 6. In this problem Γ_D is the upper boundary, $\underline{b} = (1, -0.5)$, and $f = 0$. The domain is taken to have three layers (seen on Figure 6) with $A(x) = 0.01\ I$ in the top layer, $0.05\ I$ in the internal, and $0.001\ I$ in the bottom. The Dirichlet boundary value is 1 for $x < 0.2$ and 0 otherwise. On the Neumann boundary we take $g = 0$. Figure 6 shows the mesh on level 4 (left) with $3,032$ nodes and $5,910$ triangles. On the right are the solution level curves.

5.2. **3-dimensional problems of flow and transport in porous media.** A steady state flow, with Darcy velocity $\underline{v}$ measured in ft/yr, has been established in a parallelepiped shaped reservoir $\Omega = [0, 1000] \times [-500, 500] \times [0, 500]$ (see Figure 7, right). First, we determine the pressure $p(x)$ in Ω as the solution $u(x)$ of problem (1.1) with $\underline{b} = 0$, $\gamma = 0$, and $A(x) = D(x)$, where $D(x)$ is the permeability tensor. The pressure at faces $x_1 = 0$ and $x_1 = 1000$ is constant, correspondingly 2000 and 0. The rest of the boundary is subject to no-flow condition. We take the permeability $D(x)$ to be $32\ I$ everywhere in Ω except in the layer (seen on Figure 7, middle) where $D(x)$ is taken to be ten times smaller than in the rest of the domain, i.e., in the layer $D(x) = 3.2\ I$.

Also, we have six production wells. For all of them x_3 is in the range $0 \ldots 400$. Their (x_1, x_2) coordinates are correspondingly (200, -250), (400, -250), (200, 0), (400, 0), (200, 250), and (400, 250). We treat a well simply as a line-delta function (sink) along the well axis. Production rates $Q = 16,000\ l/yr$ for wells in plane $x_2 = 0$ and $Q = 8,000\ l/yr$ for the rest, are the intensities of the sink. Figure 7 shows half of the mesh and the contour curves

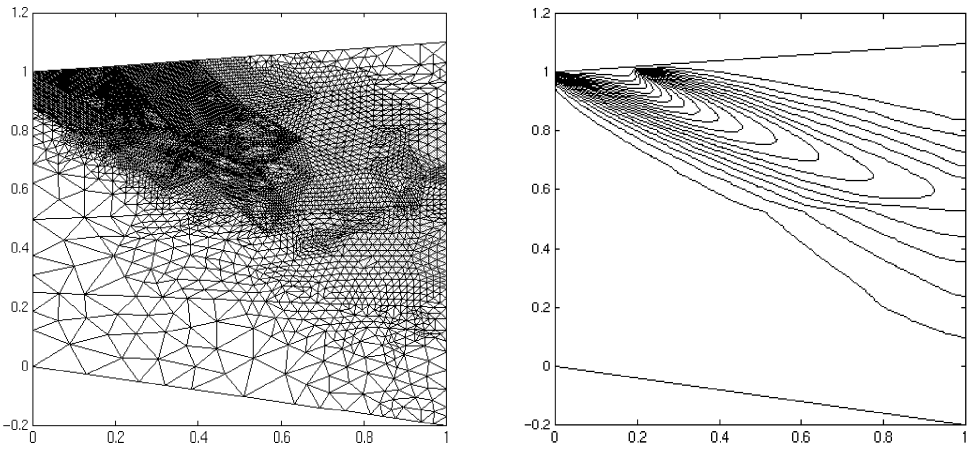

FIGURE 6. Convection-diffusion problem; the inhomogeneities are represented by three layers; Left: the locally refined mesh after 4 level of adaptive refinement (3,032 nodes and 5,910 triangles); Right: the level curves of the solution

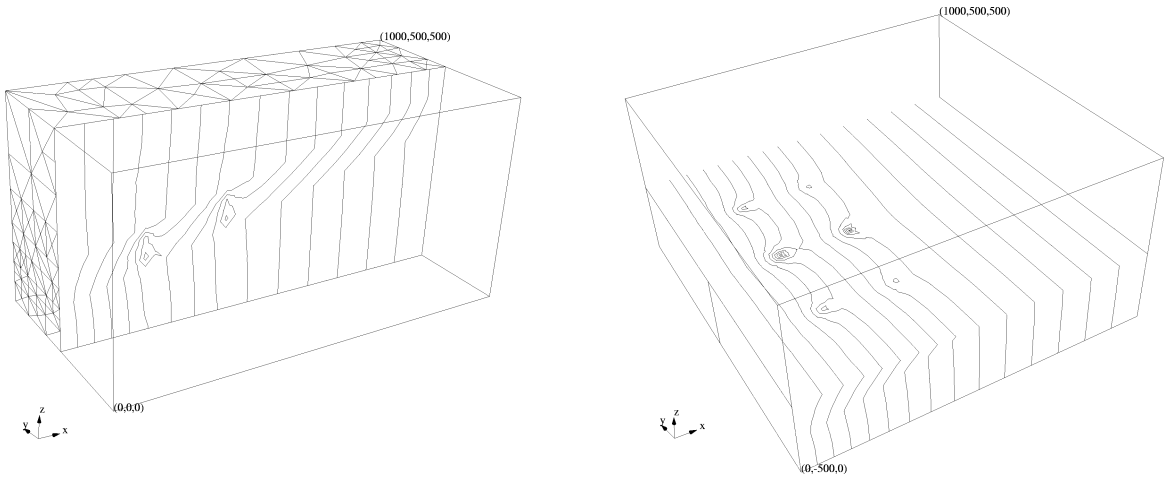

FIGURE 7. Pressure computations for a non-homogeneous reservoir: (left) contour curves of the pressure for the cross-section $x_2 = 250$; (right) contour curves of the pressure for the cross-section $x_3 = 200$.

of the pressure for the cross-section $x_2 = 250$ (left) after 5 levels of local refinement. It has 19,850 tetrahedrons and 3,905 nodes. The right picture shows the contour curves for the cross-section $x_3 = 200$.

The weighted pressure gradient $-D\nabla p$ forces the ground water to flow. The transport of a contaminant dissolved in the water, in our case benzene, is described by the convection-diffusion-reaction equation (1.1), where $u(x)$ represents the benzene concentration, $\underline{b}$ is the Darcy velocity $\underline{v} = -D\nabla p$, γ is the biodegradation rate, and $A(x)$ is the diffusion-dispersion tensor:

$$A(x) = k_{diff} I + k_t \underline{v}^T \underline{v}/|\underline{v}| + k_l(|\underline{v}|^2 I - \underline{v}^T \underline{v})/|\underline{v}|.$$

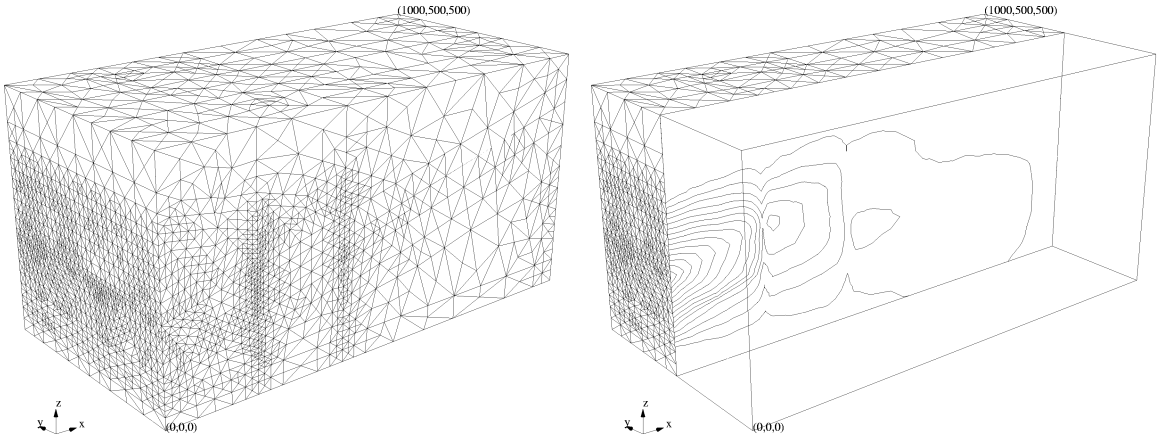

FIGURE 8. Concentration computations for a non-homogeneous reservoir. Left: the 3-D mesh on refinement level 11 with 219,789 tetrahedrons and 39,752 nodes. Right: concentration contour curves for cross-section $x_2 = 250$.

Here $k_{diff} = 0.0001$, $k_t = 21$, and $k_l = 2.1$ are the coefficients of diffusion, transverse, and longitudinal dispersions, respectively. A steady piecewise linear in x_3 and constant in x_2 leakage of benzene of maximum 30 mg/l is applied on the boundary strip $x_1 = 0$ and $50 \leq x_3 \leq 350$. The leakage is 30 mg/l at $x_3 = 200$ and drops linearly to 0 at $x_3 = 50$ and 350. The rest of the boundary is subject to homogeneous Neumann boundary condition. The dispersion/convection process causes the dissolved benzene to disperse in the reservoir. The biodegradation is transforming it into a solid substance which is absorbed by the soil. This leads to a decrease in the benzene. The computations are for the case of low absorption rate $\gamma = 0.05$. Figure 8 shows the obtained half mesh (left) on refinement level 11. The mesh has 219,789 tetrahedrons and 39,752 nodes. The first 5 level of refinement are for the pressure equation, the rest for the concentration. Figure 8 (right) shows the level curves for the concentration in the reservoir cross-section $x_2 = 250$ on the same refinement level. Figure 10 shows half of the domain Ω with the 4 wells and two iso-surfaces for the concentration, correspondingly equal to 3.5 and 8.

REFERENCES

[1] L. ANGERMANN, *Balanced a posteriori error estimates for finite-volume type discretizations of convection-dominated elliptic problems*, Computing, 55(4) (1995), pp. 305–324.

[2] L. ANGERMANN, *An a-posteriori estimation for the solution of an elliptic singularly perturbed problem*, IMA J. Numer. Anal., 12 (1992), pp. 201–215.

[3] D.N. ARNOLD, A. MUKHERJEE, AND L. POULY, *Locally adapted tetrahedral meshes using bisection*, SIAM J. Sci. Computing, 22(2) (2000), pp. 431–448.

[4] I. BABUSKA AND W.C. RHEINBOLDT, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal., 15 (1978), pp. 736–754.

[5] I. BABUSKA AND T. STROUBOULIS, *The Finite Element Method and Its Reliability*, Oxford University Press, London, 2001.

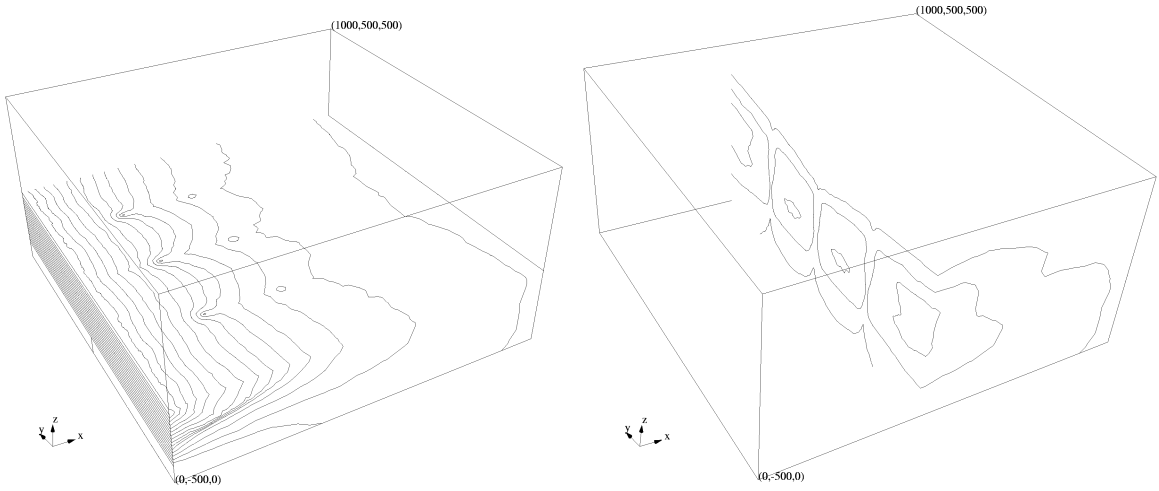

FIGURE 9. Concentration level curves at cross-sections $x_3 = 200$ (left) and $x_1 = 400$ (right).

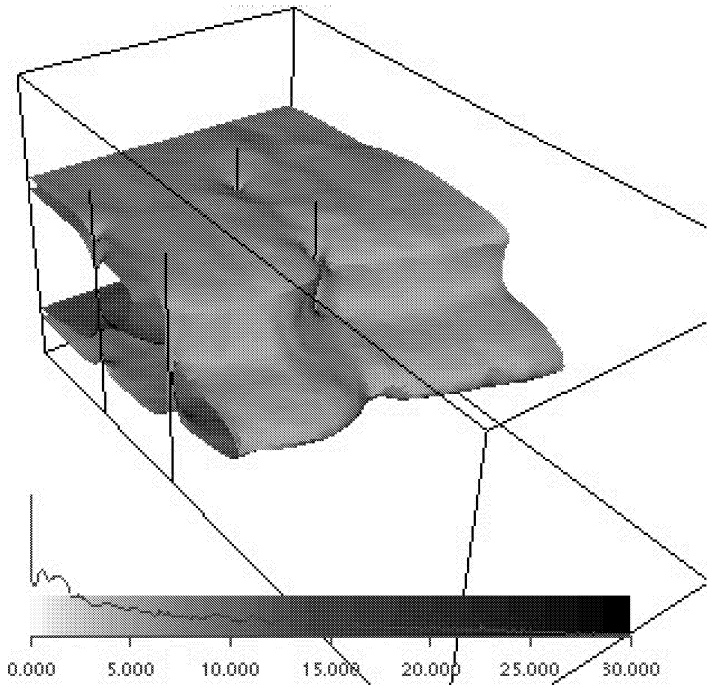

FIGURE 10. Half of the domain Ω with 4 wells and two iso-surfaces for concentration correspondingly equal to 3.5 and 8.

[6] R.E. BANK AND D.J. ROSE, *Some error estimates for the box method*, SIAM J. Numer. Anal., 24 (1987), pp. 777–787.

[7] R.E. BANK AND R.K. SMITH, *A posteriori error estimates based on hierarchical bases*, SIAM J. Numer. Anal., 30 (1993), pp. 921–932.

[8] S. BARTELS AND C. CARSTENSEN, *Each Averaging Technique Yields Reliable A Posteriori Error Control in FEM on Unstructured Grids. Part II: Higher Order FEM*, Math. Comp. 71 (2002), pp. 971-994.

[9] R. Becker and R. Rannacher, *A feed-back approach to error control in finite element methods: Basic analysis and examples*, East-West J. Numer. Math. 4 (1996), pp. 237–264.
[10] S.C. Brenner and L.R. Scott, *The Mathematical Theory of Finite Element Methods*, Springer, 1996.
[11] Z. Cai, *On the finite volume element method,* Numer. Math., 58 (1991), pp. 713–735.
[12] C. Carstensen and S. Bartels, *Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids Part I: Low order conforming, nonconforming, and mixed FEM*, Math. Comp. 71 (2002), pp. 945-969.
[13] C. Carstensen and S.A. Funken, *Constants in Clément-interpolation error and residual-based a posteriori estimates in finite element methods*, East-West Journal of Numerical Analysis, 8 (3) (2000), pp. 153-175.
[14] C. Carstensen and S.A. Funken, *Fully reliable localized error control in the FEM*, SIAM J. Sci. Comp., 21 (4) (2000), pp. 1465-1484.
[15] C. Carstensen and S.A. Funken, *A posteriori error control in low-order finite element discretizations of incompressible stationary flow problems*, Math. Comp. 70 (2001), pp. 1353–1381.
[16] S.H. Chou and Q. Li, *Error estimates in L^2, H^1, and L^∞ in covolume methods for elliptic and parabolic problems: A unified approach*, Math. Comp. 69 (229) (2000), pp. 103–120.
[17] G. Dagan, *Flow and Transport in Porous Formations*, Springer-Verlag, Berlin-Heidelberg, 1989.
[18] K. Eriksson, D. Estep, P. Hansbo, C. Johnson *Computational Differential Equations*, Cambridge University Press, 1996.
[19] K. Eriksson and C. Johnson, *An Adaptive Finite Element Method for Linear Elliptic Problems*, Math. Comp., 50 (1988), pp. 361–382.
[20] R.E. Ewing, *The need for multidisciplinary involvement in ground water contaminant simulations*, Proceedings of Next Generation Environmental Models and Computational Methods (G. Delic and M. Wheeler, eds.), SIAM, Philadelphia, PA, (1997), pp. 227–245.
[21] R. Eymard, T. Gallouët, and R. Herbin, *Finite Volume Methods*, in Handbook of Numerical Analysis, VII, pp. 713–1020, North-Holland, Amsterdam, 2000.
[22] T. Ikeda, *Maximum Principle in Finite Element Models for Convection-Diffusion Phenomena*, Lecture Notes in Numer. Appl. Analysis, Vol. 4, North-Holland, Amsterdam New York Oxford, 1983.
[23] R.D. Lazarov and S.Z. Tomov, *A posteriori error estimates for finite volume element approximations of convection-diffusion-reaction equations*, Computational Geosciences, 6 (2002), pp. 483–503. (http://www.isc.tamu.edu/iscpubs/0107.ps).
[24] R.H. Li, Z.Y. Chen, and W. Wu, *Generalized Difference Method for Differential Equations. Numerical Analysis of Finite Volume Methods*, Marcel Dekker, Inc., New York-Basel, 2000.
[25] I.D. Mishev, *Finite volume methods on Voronoi meshes*, Numer. Methods for Partial Differential Equations, 14 (1998), pp. 193–212.
[26] L. Oganesian and V.L. Ruhovetz, *Variational-Difference Methods for Solving Elliptic Equations*, Erevan, Publ. House of Armenian Academy of Sciences, 1979.
[27] A. Quarteroni and A. Valli, *Domain Decomposition Methods for Partial Differential Equations*, Clarendon Press, Oxford, 1999.
[28] R. Rodriguez, *Some remarks on Zienkiewicz-Zhu estimator*, Numerical Methods for Partial Differential equations, 10, (1994), pp. 625–635.
[29] H.-O. Ross, M. Stynes, and L. Tobiska, *Numerical Methods for Singularly Perturbed Differential Equations*, Springer, 1996.
[30] R. Rannacher and R. Scott, *Some optimal error estimates for piecewise linear element approximations*, Math. Comp., 38 (1982), pp. 437–445.
[31] A.A. Samarskii, *The Theory of Difference Schemes*, Marcel Dekker, Inc., 2001.
[32] M. Tabata, *A finite element approximation corresponding to the upwind finite differencing*, Mem. Numer. Math., 4 (1977), pp. 47–63.
[33] R. Verfürth, *A posteriori error estimation and adaptive mesh-refinement techniques*, J. Comp. Appl. Math., 50 (1994), pp. 67–83.

[34] O.C. Zienkiewicz and J.Z. Zhu, *A simple error estimator and adaptive procedure for practical engineering analysis*, Int. J. Numer. Meth. Engng., 24 (1987), pp. 337–357.
[35] O.C. Zienkiewicz and J.Z. Zhu, *Adaptivity and mesh generation*, Int. J. Numer. Methods Engrg. 32 (1991), pp. 783–810.

Institute for Applied Mathematics and Numerical Analysis, Vienna University of Technology, Wiedner Hauptstrasse 8-10/115, A-1040 Wien, Austria
E-mail address: Carsten.Carstensen@tuwien.ac.at

Department of Mathematics, Texas A&M University, College Station, TX 77843-3368
E-mail address: lazarov@isc.tamu.edu

Brookhaven National Laboratory, Information Technology Division, Bldg. 515, Upton, NY 11973
E-mail address: tomov@bnl.gov