A posteriori error estimates including algebraic error: computable upper bounds and stopping criteria for iterative solvers

Pavel Jiranek, Zdenek Strakos, Martin Vohralík

A POSTERIORI ERROR ESTIMATES INCLUDING ALGEBRAIC ERROR AND STOPPING CRITERIA FOR ITERATIVE SOLVERS

PAVEL JIRÁNEK*, ZDENĚK STRAKOŠ†, AND MARTIN VOHRALÍK‡

Abstract. For the finite volume discretization of a second-order elliptic model problem, we derive a posteriori error estimates which take into account an inexact solution of the associated linear algebraic system. We show that the algebraic error can be bounded by constructing an equilibrated Raviart–Thomas–Nédélec discrete vector field whose divergence is given by a proper weighting of the residual vector. Next, claiming that the discretization error and the algebraic one should be in balance, we construct stopping criteria for iterative algebraic solvers. An attention is paid, in particular, to the conjugate gradient method which minimizes the energy norm of the algebraic error. Using this convenient balance, we also prove the efficiency of our a posteriori estimates, i.e., we show that they also represent a lower bound, up to a generic constant, for the overall energy error. A local version of this result is also stated. This makes our approach suitable for adaptive mesh refinement which also takes into account the algebraic error. Numerical experiments illustrate the proposed estimates and construction of efficient stopping criteria for algebraic iterative solvers.

Key words. Second-order elliptic partial differential equation, finite volume method, a posteriori error estimates, iterative methods for linear algebraic systems, conjugate gradient method, stopping criteria.

AMS subject classifications. 65N15, 65N30, 76M12, 65N22, 65F10

1. Introduction.

In numerical solution of partial differential equations, the computed result is an approximate solution found in some finite-dimensional space. A natural question is whether this solution is a sufficiently accurate approximation of the exact (weak) solution of the problem at hand. A posteriori error estimates aim at giving an answer to this question while providing upper bounds on the difference between the approximate and exact solutions that can be easily computed. Their mathematical theory for the finite element method was started by the pioneering paper by Babuška and Rheinboldt [7] and a vast amount of literature on this subject exists nowadays; we refer, e.g., to the books by Verfürth [48] or Ainsworth and Oden [2]. For the cell-centered finite volume method, Ohlberger [31] derives a posteriori estimates for the convection–diffusion–reaction case, whereas, for the pure diffusion case, Achdou et al. [1] use the equivalence of the discrete forms of the schemes with some finite element ones, Nicaise [30] gives estimates for Morley-type interpolants of the original piecewise constant finite volume approximation, and Kim [23] develops a framework applicable to any locally conservative method. Recently, general guaranteed a posteriori estimates for locally conservative methods have been derived in [50, 51, 52].

*Faculty of Mechatronics and Interdisciplinary Engineering Studies, Technical University of Liberec, Hálkova 6, 46117 Liberec, Czech Republic & CERFACS, 42 Avenue Gaspard Coriolis, 31100 Toulouse, France (pavel.jiranek@cerfacs.fr). The work of this author was supported by the grant No. 201/09/P464 of the GACR and by the project IAA100300802 of the GAAS.

†Institute of Computer Science, Academy of Sciences of the Czech Republic, Pod Vodárenskou věží 2, 18207 Prague, Czech Republic (strakos@cs.cas.cz). The work of this author was supported by the project IAA100300802 of the GAAS and by the Institutional Research Plan AV0Z10300504 "Computer Science for the Information Society: Models, Algorithms, Applications."

‡UPMC Univ. Paris 06, UMR 7598, Laboratoire Jacques-Louis Lions, 75005 Paris, France & CNRS, UMR 7598, Laboratoire Jacques-Louis Lions, 75005 Paris, France (vohralik@ann.jussieu.fr). The work of this author was supported by the GNR MoMaS project "Numerical Simulations and Mathematical Modeling of Underground Nuclear Waste Disposal", PACEN/CNRS, ANDRA, BRGM, CEA, EdF, IRSN, France.

Apart from few exceptions, existing a posteriori estimates rely on the assumption that the linear system resulting from discretization *is solved exactly*. This is not assumed, e.g., in the work by Wohlmuth and Hoppe [53], but the bounds are valid only for a sufficiently refined mesh, and/or contain various unspecified constants. Rüde [38, 39, 40] gives estimates of the energy norm of the error based on the norms of the residual functionals obtained from some particular stable splitting of the underlying Hilbert space. Repin [35] or Repin and Smolianski [36] do not use any information about the discretization method and the method for solving the resulting linear algebraic system. This makes the estimates very general but the price is that they may be rather costly and not sufficiently accurate.

A moderately sized system of linear algebraic equations can be solved by a direct method. For large systems, preconditioned iterative methods, see, e.g., Saad [41], become competitive, and with increasing size they represent the only viable alternative. It should, however, be emphasized that applications of direct and iterative methods are *principally different*. While in direct methods the whole solution process must be completed to the very end in order to get a meaningful numerical solution, iterative methods can produce an approximation of the solution *at each iteration step*. The amount of computational work depends on the number of iterations performed, and an efficient PDE solver should use this principal advantage by stopping the algebraic solver whenever the algebraic error drops to the level at which it does not significantly affect the whole error (cf. [6, 46]). The simplest, most often used, and mathematically most questionable stopping criterion is based on evaluation of the relative Euclidean norm of the residual vector, see, e.g., the discussion in [22, Section 17.5]. There is only a rough connection of the algebraic residual norm with the size of the whole error in approximation of the continuous problem (we discuss this point in detail in Section 7.1 below) and, usually, not even this connection is considered. Consequently, one either continues the algebraic iterations until the residual norm is not further reduced (i.e., one uses the iterative solver essentially as a direct solver, possibly wasting resources and computational time without getting any further improvement of the whole error), or stops earlier at a risk that the computed solution is not sufficiently accurate. For some enlightening comments we refer, e.g., to [32].

The question of stopping criteria has been addressed, e.g., by Becker et al. [10] with emphasize on the multigrid solver, see also [9] and the recent paper [25]. A remarkable early approach relating the algebraic and discretization errors is represented by the so-called cascadic conjugate gradient method of Deuflhard [16], which was further studied by several other authors, see, e.g., [42]. In [3], Arioli compares the bound on the discretization error with the error of the iterative method when solving self-adjoint second-order elliptic problems. He uses the relationship between the energy norm defined in the underlying Hilbert space for the weak formulation and its restriction onto the discrete space, in combination with the numerically stable algebraic error bounds [45], see also [46]. Arioli et al. [5] extend these results for non self-adjoint problems. Their approach is interesting and useful in some applications but relies on an *a priori* knowledge, not an a posteriori bound for the discretization error. It is also worth to point out the recent results on stopping criteria for Krylov subspace methods in the framework of mixed finite element methods applied to linear and nonlinear elliptic problems [4]. Stopping the algebraic iterative solver based on a priori information on the discretization error is also applied in the context of wavelet discretizations of elliptic partial differential equations by Burstedde and Kunoth [13]. Finally, the interesting technique of Patera and Rønquist [32], see also Maday and

Patera [24], gives computable lower and upper asymptotic bounds of a linear functional of an approximate linear system solution. If the asymptotics is attained for a reasonable number of iterations, this allows to construct a stopping criterion. Such criterion is, however, tailored to a fast converging preconditioned primal-dual conjugate gradient Lanczos method, and, at least in the presented form, it does not relate the discretization and algebraic parts of the error.

In this paper we consider a second-order elliptic pure diffusion model problem: find a real-valued function p defined on Ω such that

$$-\nabla\cdot(\mathbf{S}\nabla p) = f \quad \text{in} \quad \Omega, \qquad p = g \quad \text{on} \quad \Gamma := \partial\Omega, \tag{1.1}$$

where Ω is a polygonal/polyhedral domain (open, bounded, and connected set) in $\mathbb{R}^d$, $d = 2, 3$, $\mathbf{S}$ is a diffusion tensor, f is a source term, and g prescribes the Dirichlet boundary condition. Details are given in Section 2. For the discretization of problem (1.1) on simplicial meshes, we consider in Section 3 a general locally conservative cell-centered finite volume scheme, cf. Eymard et al. [17].

The first goal of this paper is to derive a posteriori error estimates which take into account an *inexact solution of the associated linear algebraic system*. Section 5 extends for this purpose the a posteriori error estimates from [50, 51]. The derived upper bound consists of three estimators: an estimator measuring the nonconformity of the approximate solution, which essentially reflects the discretization error; an estimator corresponding to the interpolation error in the approximation of the source term f which in general turns out to be a higher-order term; and an abstract algebraic error estimator corresponding to the inexact solution of the discrete linear algebraic problem, based on equilibrated vector fields $\mathbf{r}_h$ from the lowest-order Raviart–Thomas–Nédélec space whose divergences are given by a proper weighting of the algebraic residual vector.

The second goal of this paper is to construct, in the context of solving problem (1.1), efficient *stopping criteria for iterative algebraic solvers*. Our approach is based on comparison of the discretization and algebraic error estimates, see Section 6. Under the assumption of a convenient balance between the two estimates we also prove the (local) efficiency of our estimates. They can thus correctly predict the overall error size and distribution and are suitable for adaptive mesh refinement which takes into account the inaccuracy of the algebraic computations.

Section 7 gives fully computable upper bounds or estimates for the abstract algebraic error estimator of Section 5. The first upper bound is given directly by the components of the algebraic residual vector. The second approach is based on the estimates for the algebraic error measured in the energy norm, for which there exist efficient estimates, namely in the conjugate gradient method. The last approach is based on a factual construction of a vector field $\mathbf{r}_h$ and on the use of its complementary energy $\|\mathbf{S}^{-\frac{1}{2}}\mathbf{r}_h\|$ as the algebraic error estimator. All three approaches are numerically illustrated in Section 8 on several examples.

2. Notation, assumptions, and the continuous problem. Our notation is standard, see [15, 12, 17], and it is included here for completeness. It can be skipped and used as a reference, if needed, while reading the rest of the paper.

Recall that Ω is a polygonal domain in $\mathbb{R}^2$ or a polyhedral domain in $\mathbb{R}^3$ with the boundary Γ. Let $\mathcal{T}_h$ be a partition of Ω into closed simplices, i.e., triangles if $d = 2$ and tetrahedra if $d = 3$, such that $\overline{\Omega} = \cup_{K\in\mathcal{T}_h} K$. Moreover, we assume that the partition is conforming in the sense that if $K, L \in \mathcal{T}_h$, $K \neq L$, then $K \cap L$ is either an empty set, a common face, edge, or vertex of K and L. For $K \in \mathcal{T}_h$, we denote

by $\mathcal{E}_K$ the set of sides (edges if $d = 2$, faces if $d = 3$) of K, by $\mathcal{E}_h = \cup_{K\in\mathcal{T}_h}\mathcal{E}_K$ the set of all sides of $\mathcal{T}_h$, and by $\mathcal{E}_h^{\rm int}$ and $\mathcal{E}_h^{\rm ext}$, respectively, the interior and exterior sides. We also use the notation $\mathfrak{E}_K$ for the set of all $\sigma \in \mathcal{E}_h^{\rm int}$ which share at least a vertex with a $K \in \mathcal{T}_h$. For interior sides such that $\sigma = \sigma_{K,L} := \partial K \cap \partial L$, i.e., $\sigma_{K,L}$ is a part of the boundary ∂K and, at the same time, a part of the boundary ∂L, we shall call K and L neighbors. We denote the set of neighbors of a given element $K \in \mathcal{T}_h$ by $\mathcal{T}_K$; $\mathfrak{T}_K$ stands for all triangles sharing at least a vertex with $K \in \mathcal{T}_h$. For $K \in \mathcal{T}_h$, $\mathbf{n}$ will always denote its exterior normal vector; we shall also employ the notation $\mathbf{n}_\sigma$ for a normal vector of a side $\sigma \in \mathcal{E}_h$, whose orientation is chosen arbitrarily but fixed for interior sides and coinciding with the exterior normal of Ω for exterior sides. For $\sigma_{K,L} \in \mathcal{E}_h^{\rm int}$ such that $\mathbf{n}_\sigma$ points from K to L and a sufficiently smooth function φ we also define the jump operator $[\![\cdot]\!]$ by $[\![\varphi]\!] := (\varphi|_K)|_\sigma - (\varphi|_L)|_\sigma$. Finally, a family of meshes $\mathcal{T} := \{\mathcal{T}_h;\, h > 0\}$ is parameterized by $h := \max_{K\in\mathcal{T}_h} h_K$, where h_K is the diameter of K (we also denote by h_σ the diameter of $\sigma \in \mathcal{E}_h$).

For a given domain $S \subset \mathbb{R}^d$, let $L^2(S)$ be the space of square-integrable (in the Lebesgue sense) functions over S, $(\cdot,\cdot)_S$ the $L^2(S)$ inner product, and $\|\cdot\|_S$ the associated norm (we omit the index S when $S = \Omega$). By $|S|$ we denote the Lebesgue measure of S and by $|\sigma|$ the $(d-1)$-dimensional Lebesgue measure of a $(d-1)$-dimensional surface σ in $\mathbb{R}^d$. Let $\mathcal{H}(S)$ be a set of real-valued functions defined on S. By $[\mathcal{H}(S)]^d$ we denote the set of vector functions with d components each belonging to $\mathcal{H}(S)$. Let next $H^1(S)$ be the Sobolev space with square-integrable weak derivatives up to order one, $H^1_0(S) \subset H^1(S)$ its subspace of functions with traces vanishing on Γ, $H^{1/2}(S)$ the trace space, $\mathbf{H}(\mathrm{div}, S) := \{\mathbf{v} \in [L^2(S)]^d;\, \nabla\cdot\mathbf{v} \in L^2(S)\}$ the space of functions with square-integrable weak divergences, and let finally $\langle\cdot,\cdot\rangle_{\partial S}$ stand for $(d-1)$-dimensional $L^2(\partial S)$-inner product on ∂S. We also let $H^1_\Gamma(\Omega) := \{\varphi \in H^1(\Omega);\, \varphi|_\Gamma = g\}$ be the set of functions satisfying the Dirichlet boundary condition on Γ in the sense of traces. For a given partition $\mathcal{T}_h$ of Ω, let $H^1(\mathcal{T}_h) := \{\varphi \in L^2(\Omega);\, \varphi|_K \in H^1(K)\ \forall K \in \mathcal{T}_h\}$ be the broken Sobolev space. Finally, we let $W(\mathcal{T}_h)$ be the space of functions with mean values of the traces continuous across interior sides, i.e., $W(\mathcal{T}_h) := \{\varphi \in H^1(\mathcal{T}_h);\, \langle[\![\varphi]\!], 1\rangle_\sigma = 0\ \forall\sigma \in \mathcal{E}_h^{\rm int}\}$, and $W_0(\mathcal{T}_h)$ its subspace with mean values of traces over boundary sides equal to zero, $W_0(\mathcal{T}_h) := \{\varphi \in W(\mathcal{T}_h);\, \langle\varphi, 1\rangle_\sigma = 0\ \forall\sigma \in \mathcal{E}_h^{\rm ext}\}$.

We next denote by $\mathbb{P}_k(S)$ the space of polynomials on S of total degree less than or equal to k and by $\mathbb{P}_k(\mathcal{T}_h) := \{\varphi_h \in L^2(\Omega);\, \varphi_h|_K \in \mathbb{P}_k(K)\ \forall K \in \mathcal{T}_h\}$ the space of piecewise k-degree polynomials on $\mathcal{T}_h$. We define $\mathbf{RTN}(K) := [\mathbb{P}_0(K)]^d + \mathbf{x}\mathbb{P}_0(K)$ for an element $K \in \mathcal{T}_h$ the local and $\mathbf{RTN}(\mathcal{T}_h) := \{\mathbf{v}_h \in [L^2(\Omega)]^d;\, \mathbf{v}_h|_K \in \mathbf{RTN}(K)\ \forall K \in \mathcal{T}_h\} \cap \mathbf{H}(\mathrm{div}, \Omega)$ the global lowest-order Raviart–Thomas–Nédélec space of specific piecewise linear vector functions. Recall that the normal components of $\mathbf{v}_h \in \mathbf{RTN}(K)$, $\mathbf{v}_h \cdot \mathbf{n}$, are constant on each $\sigma \in \mathcal{E}_K$ and that they represent the degrees of freedom of $\mathbf{RTN}(K)$. By consequence, the constraint $\mathbf{v}_h \in \mathbf{H}(\mathrm{div}, \Omega)$ imposing the normal continuity of the traces is expressed as $\mathbf{v}_h|_K \cdot \mathbf{n} + \mathbf{v}_h|_L \cdot \mathbf{n} = 0$ for all $\sigma_{K,L} \in \mathcal{E}_h^{\rm int}$ and there is one degree of freedom per side in $\mathbf{RTN}(\mathcal{T}_h)$. Recall also that $\nabla\cdot\mathbf{v}_h$ is constant for $\mathbf{v}_h \in \mathbf{RTN}(K)$. For more details, we refer to Brezzi and Fortin [12] or Quarteroni and Valli [33].

ASSUMPTION 2.1 (Data). *Let $\mathbf{S}$ be a symmetric, bounded, and uniformly positive definite tensor, piecewise constant on $\mathcal{T}_h$. Let in particular $c_{\mathbf{S},K} > 0$ and $C_{\mathbf{S},K} > 0$ denote its smallest and largest eigenvalues on each $K \in \mathcal{T}_h$. In addition, let $f \in \mathbb{P}_l(\mathcal{T}_h)$ be an elementwise l-degree polynomial function and $g \in H^{1/2}(\Gamma)$.*

The assumptions on $\mathbf{S}$ and f are made for the sake of simplicity and are usually

satisfied in practice. Otherwise, interpolation can be used in order to get the desired properties. In the sequel, we will employ the notation $\mathbf{S}_K := \mathbf{S}|_K$, and, in general, $\varphi_K := \varphi_h|_K$ for $\varphi_h \in \mathbb{P}_0(\mathcal{T}_h)$.

We define a bilinear form $\mathcal{B}$ by

$$\mathcal{B}(p,\varphi) := \sum_{K\in\mathcal{T}_h} (\mathbf{S}\nabla p, \nabla\varphi)_K, \qquad p,\varphi \in H^1(\mathcal{T}_h)$$

and the corresponding energy (semi-)norm by

$$|||\varphi|||^2 := \mathcal{B}(\varphi,\varphi). \tag{2.1}$$

Note that $|||\cdot|||$ becomes a norm on the space $W_0(\mathcal{T}_h)$, cf. [49]. Also note that $\mathcal{B}$ is well-defined for functions from the space $H^1(\Omega)$ as well as from the broken space $H^1(\mathcal{T}_h)$. The weak formulation of problem (1.1) is then to find $p \in H^1_\Gamma(\Omega)$ such that

$$\mathcal{B}(p,\varphi) = (f,\varphi) \qquad \forall\varphi \in H^1_0(\Omega). \tag{2.2}$$

Assumption 2.1 implies that problem (2.2) admits a unique solution [15].

3. Finite volume methods and postprocessing. We start with description of the finite volume methods for problem (1.1). In these methods, the approximation p_h of the solution p in (1.1) is only piecewise constant and it is not appropriate for an energy a posteriori error estimate, as $\nabla p_h = 0$. We therefore construct a locally postprocessed approximation using information about the known fluxes. Finally, we will in the a posteriori error estimates need an $H^1(\Omega)$-conforming approximation using the so-called Oswald interpolate.

3.1. Finite volume methods. A general cell-centered finite volume method for problem (1.1) (cf., e.g., [17]) can be written as: find $p_h \in \mathbb{P}_0(\mathcal{T}_h)$ such that

$$\sum_{\sigma\in\mathcal{E}_K} U_{K,\sigma} = f_K|K| \qquad \forall K \in \mathcal{T}_h, \tag{3.1}$$

where $f_K := (f,1)_K/|K|$ and $U_{K,\sigma}$ is the diffusive flux through the side σ of an element K, depending linearly on the values of p_h, so that (3.1) represents a system of linear algebraic equations of the form

$$\mathbb{S}P = H, \tag{3.2}$$

where $\mathbb{S} \in \mathbb{R}^{N\times N}$ and $P, H \in \mathbb{R}^N$ with N being the number of elements in the partition $\mathcal{T}_h$. Here we only assume the continuity of the fluxes, i.e., $U_{K,\sigma} = -U_{L,\sigma}$ for all $\sigma = \sigma_{K,L} \in \mathcal{E}_h^{\rm int}$, so that practically all finite volume schemes can be included. We next give an example which clarifies the ideas.

Let there be a point $\mathbf{x}_K \in K$ for each $K \in \mathcal{T}_h$ such that if $\sigma_{K,L} \in \mathcal{E}_h^{\rm int}$, then $\mathbf{x}_K \neq \mathbf{x}_L$ and the straight line connecting $\mathbf{x}_K$ and $\mathbf{x}_L$ is orthogonal to $\sigma_{K,L}$. Let an analogous orthogonality condition hold also on the boundary. Then $\mathcal{T}_h$ is admissible in the sense of [17, Definition 9.1]. Under the additional assumption $\mathbf{S}_K = s_K\mathbb{I}$ ($\mathbb{I}$ denotes the identity matrix) on each $K \in \mathcal{T}_h$, the following choice is possible:

$$\begin{aligned} U_{K,\sigma} &= -s_{K,L}\frac{|\sigma_{K,L}|}{d_{K,L}}(p_L - p_K) \quad \text{for } \sigma = \sigma_{K,L} \in \mathcal{E}_h^{\rm int}, \\ U_{K,\sigma} &= -s_K\frac{|\sigma|}{d_{K,\sigma}}(g_\sigma - p_K) \qquad \text{for } \sigma \in \mathcal{E}_K \cap \mathcal{E}_h^{\rm ext}. \end{aligned} \tag{3.3}$$

Here p_K are the cell values of p_h ($p_K := p_h|_K$ for all $K \in \mathcal{T}_h$) and the value of $s_{K,L}$ on a side $\sigma = \sigma_{K,L} \in \mathcal{E}_h^{\mathrm{int}}$ is given by $s_{K,L} = \omega_{\sigma,K} s_K + \omega_{\sigma,L} s_L$, where $\omega_{\sigma,K} = \omega_{\sigma,L} = \frac{1}{2}$ in the case of the arithmetic averaging and $\omega_{\sigma,K} = s_L/(s_K+s_L)$ and $\omega_{\sigma,L} = s_K/(s_K+s_L)$ in the case of the harmonic averaging of the diffusion coefficients s_K. The symbol $d_{K,L}$ stands for the Euclidean distance between the points $\mathbf{x}_K$ and $\mathbf{x}_L$ and $d_{K,\sigma}$ for the distance between $\mathbf{x}_K$ and $\sigma \in \mathcal{E}_K \cap \mathcal{E}_h^{\mathrm{ext}}$. Finally, $g_\sigma := \langle g, 1\rangle_\sigma/|\sigma|$ is the mean value of g on a side $\sigma \in \mathcal{E}_h^{\mathrm{ext}}$. To express (3.1), (3.3) in the matrix form (3.2), let the elements of $\mathcal{T}_h$ be enumerated using a bijection $\ell : \mathcal{T}_h \to \{1, \dots, N\}$. With the corresponding ordering of the unknown values p_K of p_h defined by $(P)_{\ell(K)} = p_K$ for each $K \in \mathcal{T}_h$, and denoting respectively by $(\cdot)_{kl}$ and $(\cdot)_k$ the matrix and vector components, the system matrix $\mathbb{S}$ and the right-hand side vector H are all zero except the elements defined by

$$
\begin{aligned}
(\mathbb{S})_{\ell(K),\ell(K)} &= \sum_{L \in \mathcal{T}_K} s_{K,L} \frac{|\sigma_{K,L}|}{d_{K,L}} \quad + \sum_{\sigma \in \mathcal{E}_K \cap \mathcal{E}_h^{\mathrm{ext}}} s_K \frac{|\sigma|}{d_{K,\sigma}}, \\
(\mathbb{S})_{\ell(K),\ell(L)} &= -s_{K,L} \frac{|\sigma_{K,L}|}{d_{K,L}}, \qquad L \in \mathcal{T}_K, \\
(H)_{\ell(K)} &= f_K |K| \quad + \sum_{\sigma \in \mathcal{E}_K \cap \mathcal{E}_h^{\mathrm{ext}}} s_K \frac{|\sigma|}{d_{K,\sigma}} g_\sigma.
\end{aligned}
$$

The system matrix $\mathbb{S}$ is therefore symmetric and positive definite and, moreover, irreducibly diagonally dominant (for the definition of this term, see, e.g., [47]).

3.2. Postprocessing. Let $\mathbf{u}_h \in \mathbf{RTN}(\mathcal{T}_h)$ be prescribed by the fluxes $U_{K,\sigma}$, i.e., for each $K \in \mathcal{T}_h$ and $\sigma \in \mathcal{E}_K$, let $\mathbf{u}_h$ be such that

$$
(\mathbf{u}_h|_K \cdot \mathbf{n})|_\sigma := U_{K,\sigma}/|\sigma|. \tag{3.4}
$$

We define a postprocessed approximation $\tilde{p}_h \in \mathbb{P}_2(\mathcal{T}_h)$ in the following way:

$$
-\mathbf{S}_K \nabla \tilde{p}_h|_K = \mathbf{u}_h|_K, \qquad \forall K \in \mathcal{T}_h, \tag{3.5a}
$$

$$
(1-\mu_K)(\tilde{p}_h, 1)_K/|K| + \mu_K \tilde{p}_h(\mathbf{x}_K) = p_K, \qquad \forall K \in \mathcal{T}_h. \tag{3.5b}
$$

Here $\mu_K = 0$ or 1, depending on whether in the particular finite volume scheme (3.1) p_K represents the approximate mean value of p_h on $K \in \mathcal{T}_h$ or the approximate point value in $\mathbf{x}_K$, respectively. It is not difficult to show that such $\tilde{p}_h$ exists, is unique, but nonconforming (does not belong to $H^1(\Omega)$), see [50, Section 4.1] and [51, Section 3.2.1]. For the finite volume scheme (3.1), (3.3), $\tilde{p}_h \in W(\mathcal{T}_h)$ if $f = 0$, but if $f \neq 0$ then $\tilde{p}_h \notin W(\mathcal{T}_h)$ in general. Under the condition that the finite volume scheme at hand satisfies some convergence properties it is shown in [51] that $\nabla \tilde{p}_h \to \nabla p$ and $\tilde{p}_h \to p$ in the $L^2(\Omega)$-norm for $h \to 0$ and that optimal a priori error estimates hold. Note finally that the described postprocessing is local on each element and its cost is negligible.

3.3. Oswald interpolation operator. For a given function $\varphi_h \in \mathbb{P}_k(\mathcal{T}_h)$, the Oswald interpolation operator $\mathcal{I}_{\mathrm{Os}}$ from $\mathbb{P}_k(\mathcal{T}_h)$ to $\mathbb{P}_k(\mathcal{T}_h) \cap H^1(\Omega)$ is defined as follows (cf., e.g., [1]): let $\mathbf{x}$ be a Lagrangian node, i.e., a point where the Lagrangian degree of freedom for $\mathbb{P}_k(\mathcal{T}_h) \cap H^1(\Omega)$ is prescribed, see [15, Section 2.2]. If $\mathbf{x}$ lies in the interior of some $K \in \mathcal{T}_h$ or in the interior of some boundary side, $\mathcal{I}_{\mathrm{Os}}(\varphi_h)(\mathbf{x}) = \varphi_h(\mathbf{x})$.

Otherwise, the value of $\mathcal{I}_{\mathrm{Os}}(\varphi_h)$ at $\mathbf{x}$ is defined by the average of the values of φ_h at this node from the neighboring elements, i.e.,

$$\mathcal{I}_{\mathrm{Os}}(\varphi_h)(\mathbf{x}) = \frac{1}{N_{\mathbf{x}}} \sum_{K \in \mathcal{T}_{\mathbf{x}}} \varphi_h|_K(\mathbf{x}),$$

where $\mathcal{T}_{\mathbf{x}} := \{K \in \mathcal{T}_h; \mathbf{x} \in K\}$ is the set of elements of $\mathcal{T}_h$ containing the node $\mathbf{x}$ and $N_{\mathbf{x}}$ denotes the number of elements contained in this set. Finally, let $\mathcal{I}^{\Gamma}_{\mathrm{Os}}(\varphi_h)$ be a modified Oswald interpolate (cf. [51]) differing from $\mathcal{I}_{\mathrm{Os}}(\varphi_h)$ only on such $K \in \mathcal{T}_h$ that contain a boundary side and such that

$$\mathcal{I}^{\Gamma}_{\mathrm{Os}}(\varphi_h)|_{\Gamma} = g \qquad \text{in the sense of traces.}$$

4. Inexact solution of systems of linear algebraic equations. Let P^{a} be an approximate solution of (3.2), i.e., $\mathbb{S}P^{\mathrm{a}} \approx H$. We then have the equation

$$\mathbb{S}P^{\mathrm{a}} = H - R, \tag{4.1}$$

where $R := H - \mathbb{S}P^{\mathrm{a}}$ is the algebraic residual vector associated with the approximation P^{a}. This means that an approximate solution P^{a} of problem (3.2) is the exact solution of the same problem with a perturbed right-hand side $H^{\mathrm{a}} := H - R$. Defining $p^{\mathrm{a}}_h \in \mathbb{P}_0(\mathcal{T}_h)$ by $p^{\mathrm{a}}_K := (P^{\mathrm{a}})_{\ell(K)}$ and a residual function $\rho_h \in \mathbb{P}_0(\mathcal{T}_h)$ associated with the algebraic residual vector R by

$$\rho_K := (R)_{\ell(K)}/|K|, \qquad K \in \mathcal{T}_h, \tag{4.2}$$

equation (4.1) is equivalent to the set of conservation equations

$$\sum_{\sigma \in \mathcal{E}_K} U^{\mathrm{a}}_{K,\sigma} = f_K|K| - \rho_K|K| \qquad \forall K \in \mathcal{T}_h. \tag{4.3}$$

The fluxes $U^{\mathrm{a}}_{K,\sigma}$ are of the same form as $U_{K,\sigma}$, with the values of p_h replaced by p^{a}_h.

Compared to (3.1), equation (4.3) contains an additional term on the right-hand side representing the error from the inexact solution of the algebraic system. We can now define $\mathbf{u}^{\mathrm{a}}_h \in \mathbf{RTN}(\mathcal{T}_h)$ by $(\mathbf{u}^{\mathrm{a}}_h|_K \cdot \mathbf{n})|_\sigma := U^{\mathrm{a}}_{K,\sigma}/|\sigma|$, so that (4.3) implies

$$\langle \mathbf{u}^{\mathrm{a}}_h \cdot \mathbf{n}, 1\rangle_{\partial K} = f_K|K| - \rho_K|K| \qquad \forall K \in \mathcal{T}_h. \tag{4.4}$$

We can consequently, as in Section 3.2, build a postprocessed approximation $\tilde{p}^{\mathrm{a}}_h \in \mathbb{P}_2(\mathcal{T}_h)$ by

$$-\mathbf{S}_K \nabla \tilde{p}^{\mathrm{a}}_h|_K = \mathbf{u}^{\mathrm{a}}_h|_K, \qquad \forall K \in \mathcal{T}_h, \tag{4.5a}$$

$$(1-\mu_K)(\tilde{p}^{\mathrm{a}}_h, 1)_K/|K| + \mu_K \tilde{p}^{\mathrm{a}}_h(\mathbf{x}_K) = p^{\mathrm{a}}_K, \qquad \forall K \in \mathcal{T}_h. \tag{4.5b}$$

The backward error idea of incorporating the algebraic error into (4.1), together with the construction (4.4) and (4.5), will form a basis for our a posteriori error estimates presented next.

5. A posteriori error estimates including the algebraic error. We first recall the following result proved as a part of [50, Lemma 7.1] (here $|||\cdot|||$ is the energy (semi-)norm defined by (2.1)):

LEMMA 5.1 (Abstract a posteriori error estimate). *Consider arbitrary* $p \in H^1_\Gamma(\Omega)$ *and* $\tilde{p} \in H^1(\mathcal{T}_h)$. *Then*

$$|||p-\tilde{p}||| \leq \inf_{s\in H^1_\Gamma(\Omega)} |||\tilde{p}-s||| + \sup_{\substack{\varphi\in H^1_0(\Omega)\\ |||\varphi|||=1}} \mathcal{B}(p-\tilde{p},\varphi).$$

Before formulating the a posteriori error estimate, we recall the Poincaré inequality. It states that for a convex polygon/polyhedron K and $\varphi \in H^1(K)$,

$$\|\varphi-\varphi_K\|_K \leq \frac{1}{\pi} h_K \|\nabla\varphi\|_K, \tag{5.1}$$

where $\varphi_K := (\varphi,1)_K/|K|$ is the mean of φ over K. Our a posteriori error estimates are based on the following theorem:

THEOREM 5.2 (A posteriori error estimate including the algebraic error). *Let* p *be the weak solution of* (1.1) *given by* (2.2) *with the data satisfying Assumption 2.1. Let a couple* $p_h^{\mathrm{a}} \in \mathbb{P}_0(\mathcal{T}_h)$, $\mathbf{u}_h^{\mathrm{a}} \in \mathbf{RTN}(\mathcal{T}_h)$ *be given, where* $\mathbf{u}_h^{\mathrm{a}}$ *satisfies* (4.4) *for some given function* $\rho_h \in \mathbb{P}_0(\mathcal{T}_h)$. *Finally, let* $\tilde{p}_h^{\mathrm{a}} \in \mathbb{P}_2(\mathcal{T}_h)$ *be the postprocessed approximation given by* (4.5a)–(4.5b). *Then*

$$|||p-\tilde{p}_h^{\mathrm{a}}||| \leq \eta_{\mathrm{NC}} + \eta_{\mathrm{O}} + \eta_{\mathrm{AE}}, \tag{5.2}$$

where the global nonconformity *and* oscillation estimators *are given by*

$$\eta_{\mathrm{NC}} := \left\{ \sum_{K\in\mathcal{T}_h} \eta_{\mathrm{NC},K}^2 \right\}^{\frac{1}{2}} \qquad \text{and} \qquad \eta_{\mathrm{O}} := \left\{ \sum_{K\in\mathcal{T}_h} \eta_{\mathrm{O},K}^2 \right\}^{\frac{1}{2}},$$

respectively, and η_{AE} *stands for the* algebraic error estimator *defined by*

$$\eta_{\mathrm{AE}} := \inf_{\substack{\mathbf{r}_h\in\mathbf{RTN}(\mathcal{T}_h)\\ \nabla\cdot\mathbf{r}_h=\rho_h}} \sup_{\substack{\varphi\in H^1_0(\Omega)\\ |||\varphi|||=1}} (\mathbf{r}_h, \nabla\varphi). \tag{5.3}$$

The local nonconformity *and* oscillation estimators *are respectively given by*

$$\eta_{\mathrm{NC},K} := |||\tilde{p}_h^{\mathrm{a}} - \mathcal{I}_{\mathrm{Os}}^{\Gamma}(\tilde{p}_h^{\mathrm{a}})|||_K, \quad \eta_{\mathrm{O},K} := \frac{1}{\pi\sqrt{c_{\mathbf{S},K}}}\, h_K \|f-f_K\|_K,$$

and $\mathcal{I}_{\mathrm{Os}}^{\Gamma}(\tilde{p}_h^{\mathrm{a}})$ *is the modified Oswald interpolant of* $\tilde{p}_h^{\mathrm{a}}$ *described in Section 3.3.*

Proof. For any $s \in H^1_\Gamma(\Omega)$ we have from Lemma 5.1

$$\begin{aligned} |||p-\tilde{p}_h^{\mathrm{a}}||| &\leq |||\tilde{p}_h^{\mathrm{a}}-s||| + \sup_{\substack{\varphi\in H^1_0(\Omega)\\ |||\varphi|||=1}} \mathcal{B}(p-\tilde{p}_h^{\mathrm{a}},\varphi) \\ &= |||\tilde{p}_h^{\mathrm{a}}-s||| + \sup_{\substack{\varphi\in H^1_0(\Omega)\\ |||\varphi|||=1}} [T_{\mathrm{O}}(\varphi) + T_{\mathrm{AE}}(\varphi)] \\ &\leq |||\tilde{p}_h^{\mathrm{a}}-s||| + \sup_{\substack{\varphi\in H^1_0(\Omega)\\ |||\varphi|||=1}} T_{\mathrm{O}}(\varphi) + \sup_{\substack{\varphi\in H^1_0(\Omega)\\ |||\varphi|||=1}} T_{\mathrm{AE}}(\varphi), \end{aligned} \tag{5.4}$$

where $T_{\mathrm{O}}(\varphi) := \sum_{K\in\mathcal{T}_h} (\mathbf{S}\nabla(p-\tilde{p}_h^{\mathrm{a}}) + \mathbf{r}_h, \nabla\varphi)_K$ and $T_{\mathrm{AE}}(\varphi) := -(\mathbf{r}_h, \nabla\varphi)$ for an arbitrary $\mathbf{r}_h \in \mathbf{RTN}(\mathcal{T}_h)$ such that $\nabla\cdot\mathbf{r}_h = \rho_h$.

The term $T_{\mathrm{O}}(\varphi)$ can be expressed using the definition of the weak solution (2.2), (4.5a), and the Green theorem as (recall that $\mathbf{r}_h$, $\mathbf{u}_h^{\mathrm{a}} \in \mathbf{H}(\mathrm{div},\Omega)$ and $\varphi \in H_0^1(\Omega)$)

$$\begin{aligned} T_{\mathrm{O}}(\varphi) &= (f,\varphi) - \sum_{K\in\mathcal{T}_h} (\mathbf{S}\nabla\tilde{p}_h^{\mathrm{a}} - \mathbf{r}_h, \nabla\varphi)_K \\ &= (f,\varphi) + (\mathbf{r}_h + \mathbf{u}_h^{\mathrm{a}}, \nabla\varphi) = (f - \nabla\cdot(\mathbf{r}_h + \mathbf{u}_h^{\mathrm{a}}), \varphi). \end{aligned} \tag{5.5}$$

Since the divergence is piecewise constant for functions in $\mathbf{RTN}(\mathcal{T}_h)$, the Green theorem with (4.4) gives for any $K \in \mathcal{T}_h$

$$(\nabla\cdot\mathbf{u}_h^{\mathrm{a}})|_K = (\nabla\cdot\mathbf{u}_h^{\mathrm{a}}, 1)_K/|K| = \langle \mathbf{u}_h^{\mathrm{a}}\cdot\mathbf{n}, 1\rangle_{\partial K}/|K| = f_K - \rho_K. \tag{5.6}$$

Thus, employing $\nabla\cdot\mathbf{r}_h|_K = \rho_K$,

$$f - \nabla\cdot(\mathbf{r}_h + \mathbf{u}_h^{\mathrm{a}}) = f - \rho_K - f_K + \rho_K = f - f_K \qquad \forall K \in \mathcal{T}_h.$$

Now let $\varphi_K := (\varphi,1)_K/|K|$ be the mean value of φ over K. Using the above identities, we can rewrite (5.5) in the form

$$T_{\mathrm{O}}(\varphi) = \sum_{K\in\mathcal{T}_h} (f - f_K, \varphi - \varphi_K)_K$$

and from the Cauchy–Schwarz inequality, the Poincaré inequality (5.1), and using $|||\varphi||| = 1$, we obtain the estimate

$$T_{\mathrm{O}}(\varphi) \le \sum_{K\in\mathcal{T}_h} \|f - f_K\|_K \|\varphi - \varphi_K\|_K \le \sum_{K\in\mathcal{T}_h} \eta_{\mathrm{O},K} |||\varphi|||_K \le \left\{ \sum_{K\in\mathcal{T}_h} \eta_{\mathrm{O},K}^2 \right\}^{\frac{1}{2}}.$$

With (5.4), putting $s = \mathcal{I}_{\mathrm{Os}}^{\Gamma}(\tilde{p}_h^{\mathrm{a}})$ and noticing that $\mathbf{r}_h \in \mathbf{RTN}(\mathcal{T}_h)$ such that $\nabla\cdot\mathbf{r}_h = \rho_h$ was chosen arbitrarily, the proof is finished. □

Remark 5.3 (Form of the a posteriori error estimate). *By* (4.5a) *and by definition* (2.1) *of the energy (semi-)norm, posing* $\mathbf{u} := -\mathbf{S}\nabla p$,

$$|||p - \tilde{p}_h^{\mathrm{a}}||| = \left\|\mathbf{S}^{-\frac{1}{2}}(\mathbf{u} - \mathbf{u}_h^{\mathrm{a}})\right\|,$$

so that the a posteriori error estimate of Theorem 5.2 equivalently controls the energy norm of the error in the flux.

The a posteriori error estimate given in Theorem 5.2 consists of three parts: the nonconformity estimator η_{NC} indicating the departure of the approximate solution $\tilde{p}_h^{\mathrm{a}}$ from the space $H^1(\Omega)$, the oscillation estimator η_{O} which measures the interpolation error in the right-hand side of problem (1.1), and the algebraic error estimator η_{AE} which accounts for the error from the inexact solution of the algebraic system. Note that the nonconformity estimator depends on the actual approximation $\tilde{p}_h^{\mathrm{a}}$ of $\tilde{p}_h$ and thus implicitly also on ρ_h and *not only on the discretization error*, whereas the algebraic error estimator depends *only on the residual function* ρ_h associated with the algebraic residual vector R, see (4.2). We discuss computable upper bounds on η_{AE} in Section 7 below. Finally, whenever $f \in H^1(\mathcal{T}_h)$, the oscillation estimator η_{O} is of higher order by the Poincaré inequality (5.1) (it converges as $O(h^2)$ for $h \to 0$) and its value is significant only on coarse grids or for highly varying $\mathbf{S}$. We shall give some more details in the next section.

The following remark follows from the freedom of choice of s and $\mathbf{r}_h$ in the proof of Theorem 5.2:

REMARK 5.4 (Abstract form of Theorem 5.2). *With the assumptions of Theorem 5.2,*

$$|||p - \tilde{p}_h^{\mathrm{a}}||| \leq \eta_{\mathrm{NC}}^{\mathrm{A}} + \eta_{\mathrm{O}} + \eta_{\mathrm{AE}}^{\mathrm{A}}$$

with

$$\eta_{\mathrm{NC}}^{\mathrm{A}} := \inf_{s \in H^1_\Gamma(\Omega)} |||\tilde{p}_h^{\mathrm{a}} - s|||, \quad \eta_{\mathrm{AE}}^{\mathrm{A}} := \inf_{\substack{\mathbf{r} \in \mathbf{H}(\mathrm{div},\Omega) \\ \nabla\cdot\mathbf{r} = \rho_h}} \sup_{\substack{\varphi \in H^1_0(\Omega) \\ |||\varphi|||=1}} (\mathbf{r}, \nabla\varphi), \tag{5.7}$$

and η_{O} as in Theorem 5.2. Please note that

$$\eta_{\mathrm{NC}}^{\mathrm{A}} \leq \eta_{\mathrm{NC}} \quad \textit{and} \quad \eta_{\mathrm{AE}}^{\mathrm{A}} \leq \eta_{\mathrm{AE}}.$$

We now show that the abstract algebraic error estimator $\eta_{\mathrm{AE}}^{\mathrm{A}}$ given above is equal to the complementary energy of the flux of the solution of the original problem (1.1) with homogeneous Dirichlet boundary condition and the right-hand side replaced by the residual function ρ_h.

THEOREM 5.5 (Equivalence of the abstract algebraic error estimator and of the minimal complementary energy). *Consider an arbitrary $\rho_h \in \mathbb{P}_0(\mathcal{T}_h)$ and $\eta_{\mathrm{AE}}^{\mathrm{A}}$ given by* (5.7). *Then*

$$\eta_{\mathrm{AE}}^{\mathrm{A}} = \|\mathbf{S}^{-\frac{1}{2}}\mathbf{q}\|,$$

where $\mathbf{q} \in \mathbf{H}(\mathrm{div},\Omega)$, $\nabla\cdot\mathbf{q} = \rho_h$, is the unique minimizer of the complementary energy characterized by

$$\|\mathbf{S}^{-\frac{1}{2}}\mathbf{q}\| = \min_{\substack{\mathbf{r} \in \mathbf{H}(\mathrm{div},\Omega) \\ \nabla\cdot\mathbf{r} = \rho_h}} \|\mathbf{S}^{-\frac{1}{2}}\mathbf{r}\|, \tag{5.8}$$

or, equivalently, by $\mathbf{q} = -\mathbf{S}\nabla e$, where $e \in H^1_0(\Omega)$ is the unique weak solution of

$$-\nabla\cdot(\mathbf{S}\nabla e) = \rho_h \quad \textit{in} \quad \Omega, \qquad e = 0 \quad \textit{on} \quad \Gamma. \tag{5.9}$$

Proof. Using the Cauchy–Schwarz inequality,

$$\begin{aligned}
\eta_{\mathrm{AE}}^{\mathrm{A}} &= \inf_{\substack{\mathbf{r} \in \mathbf{H}(\mathrm{div},\Omega) \\ \nabla\cdot\mathbf{r} = \rho_h}} \sup_{\substack{\varphi \in H^1_0(\Omega) \\ |||\varphi|||=1}} (\mathbf{r}, \nabla\varphi) \leq \sup_{\substack{\varphi \in H^1_0(\Omega) \\ |||\varphi|||=1}} (\mathbf{q}, \nabla\varphi) \\
&= \sup_{\substack{\varphi \in H^1_0(\Omega) \\ |||\varphi|||=1}} (\mathbf{S}^{-\frac{1}{2}}\mathbf{q}, \mathbf{S}^{\frac{1}{2}}\nabla\varphi) \leq \sup_{\substack{\varphi \in H^1_0(\Omega) \\ |||\varphi|||=1}} \big(\|\mathbf{S}^{-\frac{1}{2}}\mathbf{q}\| |||\varphi|||\big) = \|\mathbf{S}^{-\frac{1}{2}}\mathbf{q}\|.
\end{aligned}$$

Before proceeding to the converse, let us recall that the problem of finding $\mathbf{q}$ as the minimizer of the complementary energy is equivalent to the problem of finding $\mathbf{q} \in \mathbf{H}(\mathrm{div},\Omega)$, $\nabla\cdot\mathbf{q} = \rho_h$, such that

$$(\mathbf{S}^{-1}\mathbf{q}, \mathbf{v}) = 0 \qquad \forall \mathbf{v} \in \mathbf{H}(\mathrm{div},\Omega);\ \nabla\cdot\mathbf{v} = 0, \tag{5.10}$$

see, e.g., [33, Theorem 7.1.1]. Let now $\mathbf{r} \in \mathbf{H}(\mathrm{div}, \Omega)$ such that $\nabla \cdot \mathbf{r} = \rho_h$ be arbitrary. Then, by (5.10), it holds $(\mathbf{S}^{-1}\mathbf{q}, \mathbf{q} - \mathbf{r}) = 0$, and using that $\mathbf{q} = -\mathbf{S}\nabla e$, we get

$$\|\mathbf{S}^{-\frac{1}{2}}\mathbf{q}\|^2 = (\mathbf{S}^{-1}\mathbf{q}, \mathbf{q}) = (\mathbf{S}^{-1}\mathbf{q}, \mathbf{q} - \mathbf{r}) + (\mathbf{S}^{-1}\mathbf{q}, \mathbf{r}) = (-\nabla e, \mathbf{r}).$$

Hence

$$\|\mathbf{S}^{-\frac{1}{2}}\mathbf{q}\| = |||e||| = \left(\mathbf{r}, \frac{-\nabla e}{|||e|||}\right) \leq \sup_{\substack{\varphi \in H^1_0(\Omega) \\ |||\varphi||| = 1}} (\mathbf{r}, \nabla\varphi),$$

which concludes the proof by virtue of the fact that $\mathbf{r} \in \mathbf{H}(\mathrm{div}, \Omega)$ such that $\nabla \cdot \mathbf{r} = \rho_h$ was chosen arbitrarily. □

6. Stopping criterion for iterative solvers and efficiency of the a posteriori error estimate. In PDE solvers, the discretization and algebraic errors should be in balance. This requirement leads to a stopping criterion for iterative algebraic solvers applied to discretized linear algebraic systems. Using this approach, we also prove in this section global and local efficiency of our a posteriori error estimates in the sense that the estimators also represent global and local lower bounds (up to a generic constant) for the error in the energy (semi-)norm. Please note that all the results presented below still hold when η_{AE} is replaced by one of its computable upper bounds presented in Section 7 below.

6.1. Stopping criterion. A stopping criterion that we propose requires the value of the algebraic error estimator to be related to the nonconformity one via

$$\eta_{\mathrm{AE}} \leq \gamma\, \eta_{\mathrm{NC}}, \qquad 0 < \gamma \leq 1, \tag{6.1}$$

where γ is typically close to 1. This leads to the bound

$$|||p - \tilde{p}_h^{\mathrm{a}}||| \leq (1 + \gamma)\eta_{\mathrm{NC}} + \eta_{\mathrm{O}}.$$

Let η_{AE} can be constructed using local contributions $\eta_{\mathrm{AE},K}$ corresponding to individual elements $K \in \mathcal{T}_h$ so that

$$\eta_{\mathrm{AE}} = \left\{ \sum_{K \in \mathcal{T}_h} \eta^2_{\mathrm{AE},K} \right\}^{\frac{1}{2}}. \tag{6.2}$$

We will use such construction below. Then one can consider also a *local* stopping criterion of the form

$$\eta_{\mathrm{AE},K} \leq \gamma_K\, \eta_{\mathrm{NC},K}, \qquad 0 < \gamma_K \leq 1 \qquad \forall K \in \mathcal{T}_h, \tag{6.3}$$

where γ_K are typically close to 1.

In the rest of this section we will employ the notation $c_{\mathbf{S},\mathfrak{T}_K} := \min_{L \in \mathfrak{T}_K} c_{\mathbf{S},L}$, which is the lower bound on the eigenvalues of the diffusion tensor $\mathbf{S}$ on the patch of elements $\mathfrak{T}_K$ (see Section 2). The notation C, $\widetilde{C}$, and $\overline{C}$ will stand for generic constants dependent on the quantities specified below, possibly different at different occurrences. We will also make use of the following assumption:

ASSUMPTION 6.1 (Shape regularity of $\mathcal{T}$). *There exists a constant $\theta_{\mathcal{T}} > 0$ such that $\min_{K \in \mathcal{T}_h} h_K/\varrho_K \leq \theta_{\mathcal{T}}$ for all $\mathcal{T}_h \in \mathcal{T}$, where ϱ_K is the diameter of the largest ball inscribed in K.*

6.2. Efficiency of the estimates. Let us first introduce some additional notation. For $\varphi \in H^1(\mathcal{T}_h)$ and $\sigma = \sigma_{L,M} \in \mathcal{E}_h^{\mathrm{int}}$, put $\|\varphi\|_{\#,\sigma} := h_\sigma^{-\frac{1}{2}} |\langle [\![\varphi]\!], 1\rangle_\sigma|$. Note that $\langle [\![\varphi]\!], 1\rangle_\sigma = 0$ for all $\sigma \in \mathcal{E}_h^{\mathrm{int}}$ for $\varphi \in W(\mathcal{T}_h)$, as $\|\varphi\|_{\#,\sigma}$ only measures the jump in the mean values of φ. Then, for a given set of sides $\mathcal{E}$, let

$$|||\varphi|||_{\#,\mathcal{E}}^2 := c_{\mathbf{S},\mathcal{E}} \sum_{\sigma\in\mathcal{E}} \|\varphi\|_{\#,\sigma}^2 ,$$

where $c_{\mathbf{S},\mathcal{E}}$ is the minimum of the values $c_{\mathbf{S},K}$ over all $K \in \mathcal{T}_h$ which have at least one side in the set $\mathcal{E}$. The following theorem shows that using the *local* stopping criterion (6.3), the derived estimates also represent *local* lower bounds for the error. Consequently, they are suitable for adaptive mesh refinement.

THEOREM 6.2 (Local efficiency of the a posteriori error estimate). *Let the assumptions of Theorem 5.2 and Assumption 6.1 be satisfied. Let* (6.2) *hold together with* (6.3)*. Then, for each $K \in \mathcal{T}_h$,*

$$\begin{aligned}\eta_{\mathrm{NC},K} + \eta_{\mathrm{AE},K} \le (1+\gamma_K)\big(&C C_{\mathbf{S},K}^{\frac{1}{2}} c_{\mathbf{S},\mathfrak{T}_K}^{-\frac{1}{2}} \left(|||p - \tilde{p}_h^{\mathrm{a}}|||_{\mathfrak{T}_K} + |||p - \tilde{p}_h^{\mathrm{a}}|||_{\#,\mathfrak{E}_K}\right)\\ &+ |||\mathcal{I}_{\mathrm{Os}}(\tilde{p}_h^{\mathrm{a}}) - \mathcal{I}_{\mathrm{Os}}^{\Gamma}(\tilde{p}_h^{\mathrm{a}})|||_K\big).\end{aligned}$$

If, moreover, the local algebraic error estimators are given by $\eta_{\mathrm{AE},K} = \|\mathbf{S}^{-\frac{1}{2}}\mathbf{r}_h\|_K$ for some $\mathbf{r}_h$ such that $\mathbf{r}_h \in \mathbf{RTN}(\mathcal{T}_h)$, $\nabla\cdot\mathbf{r}_h = \rho_h$, then

$$\begin{aligned}\eta_{\mathrm{NC},K} + \eta_{\mathrm{O},K} + \eta_{\mathrm{AE},K} \le \widetilde{C}(1+\gamma_K)\big(&|||p - \tilde{p}_h^{\mathrm{a}}|||_{\mathfrak{T}_K} + |||p - \tilde{p}_h^{\mathrm{a}}|||_{\#,\mathfrak{E}_K}\\ &+ |||\mathcal{I}_{\mathrm{Os}}(\tilde{p}_h^{\mathrm{a}}) - \mathcal{I}_{\mathrm{Os}}^{\Gamma}(\tilde{p}_h^{\mathrm{a}})|||_K\big).\end{aligned} \tag{6.4}$$

Here the constant C depends only on the space dimension d and on the shape regularity parameter $\theta_{\mathcal{T}}$ and $\widetilde{C}$ depends in addition on the polynomial degree l of f (see Assumption 2.1) and on the ratio $C_{\mathbf{S},K}/c_{\mathbf{S},\mathfrak{T}_K}$.

Proof. It has been proved in [50, Theorem 4.4], [51, Theorem 4.2], and [52, Theorem 6.15], using the tools from [48] and [1], that for any piecewise polynomial function $\tilde{p}_h^{\mathrm{a}} \in \mathbb{P}_m(\mathcal{T}_h)$,

$$\begin{aligned}\eta_{\mathrm{NC},K} \le C\Bigg(&C_{\mathbf{S},K}^{\frac{1}{2}} c_{\mathbf{S},\mathfrak{T}_K}^{-\frac{1}{2}} |||p - \tilde{p}_h^{\mathrm{a}}|||_{\mathfrak{T}_K} + C_{\mathbf{S},K}^{\frac{1}{2}} \sum_{\sigma\in\mathfrak{E}_K} \|p - \tilde{p}_h^{\mathrm{a}}\|_{\#,\sigma}\Bigg) \qquad (6.5a)\\ &+ |||\mathcal{I}_{\mathrm{Os}}(\tilde{p}_h^{\mathrm{a}}) - \mathcal{I}_{\mathrm{Os}}^{\Gamma}(\tilde{p}_h^{\mathrm{a}})|||_K,\\ \pi^{-1} c_{\mathbf{S},K}^{-\frac{1}{2}} h_K \|f + \nabla\cdot(\mathbf{S}\nabla\tilde{p}_h^{\mathrm{a}})\|_K &\le \overline{C} C_{\mathbf{S},K}^{\frac{1}{2}} c_{\mathbf{S},K}^{-\frac{1}{2}} |||p - \tilde{p}_h^{\mathrm{a}}|||_K, \qquad (6.5b)\end{aligned}$$

where the constant C depends only on d, $\theta_{\mathcal{T}}$, and the polynomial degree m of $\tilde{p}_h^{\mathrm{a}}$, and $\overline{C}$ depends in addition on the polynomial degree l of f.

The first assertion of the theorem is thus an immediate consequence of (6.5a) and of (6.3). For the second one, we have to bound $\eta_{\mathrm{O},K}$. Using $f_K = (\nabla\cdot\mathbf{u}_h^{\mathrm{a}})|_K + \rho_K$ from (5.6), $\mathbf{u}_h^{\mathrm{a}}|_K = -\mathbf{S}_K\nabla\tilde{p}_h^{\mathrm{a}}|_K$ from (4.5a), the triangle inequality, and $\nabla\cdot\mathbf{r}_h = \rho_h$, we have

$$\eta_{\mathrm{O},K} = \pi^{-1} c_{\mathbf{S},K}^{-\frac{1}{2}} h_K \|f - f_K\|_K \le \pi^{-1} c_{\mathbf{S},K}^{-\frac{1}{2}} h_K (\|f + \nabla\cdot(\mathbf{S}\nabla\tilde{p}_h^{\mathrm{a}})\|_K + \|\nabla\cdot\mathbf{r}_h\|_K).$$

The first term on the right-hand side of this inequality is bounded by (6.5b). Using the inverse inequality (cf. [33, Proposition 6.3.2]) and Assumption 2.1,

$$\|\nabla\cdot\mathbf{r}_h\|_K \le C h_K^{-1} \|\mathbf{r}_h\|_K \le C h_K^{-1} C_{\mathbf{S},K}^{\frac{1}{2}} \|\mathbf{S}^{-\frac{1}{2}}\mathbf{r}_h\|_K$$

for some constant C only depending on d and $\theta_{\mathcal{T}}$. Thus, using (6.3),

$$\eta_{\mathrm{O},K} \le \overline{C} C_{\mathbf{S},K}^{\frac{1}{2}} c_{\mathbf{S},K}^{-\frac{1}{2}} |||p - \tilde{p}_h^{\mathrm{a}}|||_K + C C_{\mathbf{S},K}^{\frac{1}{2}} c_{\mathbf{S},K}^{-\frac{1}{2}} \gamma_K \eta_{\mathrm{NC},K}.$$

The assertion (6.4) thus follows by combining the above estimate with the previous ones. ∎

Using the *global* stopping criterion (6.1) without (6.2) and (6.3), we obtain the following *global* lower bound (note that the result for estimators η_{NC} and η_{AE} is standard and sufficient, as the estimator η_{O} represents only data oscillations and is generally of higher order; it can also be included as shown in Theorem 6.2):

THEOREM 6.3 (Global efficiency of the a posteriori error estimate). *Let the assumptions of Theorem 5.2 and Assumption 6.1 be satisfied and let* (6.1) *hold. Then*

$$\eta_{\mathrm{NC}} + \eta_{\mathrm{AE}} \le \widetilde{C}(1+\gamma)(|||p - \tilde{p}_h^{\mathrm{a}}||| + |||p - \tilde{p}_h^{\mathrm{a}}|||_{\#,\mathcal{E}_h^{\mathrm{int}}} + |||\mathcal{I}_{\mathrm{Os}}(\tilde{p}_h^{\mathrm{a}}) - \mathcal{I}_{\mathrm{Os}}^{\Gamma}(\tilde{p}_h^{\mathrm{a}})|||),$$

where the constant $\widetilde{C}$ *depends only on* d, $\theta_{\mathcal{T}}$, *and* $\max_{K\in\mathcal{T}_h} C_{\mathbf{S},K}/c_{\mathbf{S},\mathfrak{T}_K}$.

Proof. From (6.1), $\eta_{\mathrm{NC}} + \eta_{\mathrm{AE}} \le (1+\gamma)\eta_{\mathrm{NC}}$. Using the definition of η_{NC}, employing (6.5a) and the inequality $(a+b)^2 \le 2a^2 + 2b^2$, we have

$$\eta_{\mathrm{NC}} + \eta_{\mathrm{AE}} \le (1+\gamma)\sqrt{2}\Bigg\{ \sum_{K\in\mathcal{T}_h} \big(C C_{\mathbf{S},K} c_{\mathbf{S},\mathfrak{T}_K}^{-1} (|||p - \tilde{p}_h^{\mathrm{a}}|||_{\mathfrak{T}_K}^2 + |||p - \tilde{p}_h^{\mathrm{a}}|||_{\#,\mathfrak{E}_K}^2) + |||\mathcal{I}_{\mathrm{Os}}(\tilde{p}_h^{\mathrm{a}}) - \mathcal{I}_{\mathrm{Os}}^{\Gamma}(\tilde{p}_h^{\mathrm{a}})|||_K^2 \big) \Bigg\}^{\frac{1}{2}},$$

from where the assertion of the theorem follows. ∎

We remark that the terms $|||\mathcal{I}_{\mathrm{Os}}(\tilde{p}_h^{\mathrm{a}}) - \mathcal{I}_{\mathrm{Os}}^{\Gamma}(\tilde{p}_h^{\mathrm{a}})|||_K$ in the above theorems penalize the possible violation of the Dirichlet boundary condition and they can be nonzero only for boundary simplices. The term $|||p - \tilde{p}_h^{\mathrm{a}}|||_{\#,\mathcal{E}_h^{\mathrm{int}}} = |||\tilde{p}_h^{\mathrm{a}}|||_{\#,\mathcal{E}_h^{\mathrm{int}}}$ then accounts for the discontinuity of the means of traces of the postprocessed approximation $\tilde{p}_h^{\mathrm{a}}$ and for a part of the algebraic error. In our numerical experiments it was negligible. Bound (6.4) is in particular relevant to the cases investigated in Section 7.3 below, where the algebraic error estimator admits the desired form.

7. Computable upper bounds and estimates for the algebraic error estimator. The algebraic error estimator η_{AE} of Section 5 was defined in a general way without specification of the techniques for computing it. In this section we discuss three different approaches giving computable upper bounds on η_{AE} or its efficient estimates.

7.1. Simple bound using the algebraic residual vector. A guaranteed upper bound on the algebraic error η_{AE} can be obtained using a *weighted* Euclidean norm of the algebraic residual vector R defined in (4.1). This worst case-like scenario approach can lead to large overestimation, cf. Section 8 below; for a supportive algebraic reasoning see, e.g., [22, Section 17.5].

LEMMA 7.1 (Algebraic error estimator using the algebraic residual vector). *The algebraic error estimator* η_{AE} *from Theorem 5.2 can be bounded as*

$$\eta_{\mathrm{AE}} \le \eta_{\mathrm{AE}}^{(1)} := \sqrt{\frac{C_{\mathrm{F},\Omega}}{c_{\mathbf{S},\Omega}}}\, h_\Omega \left\{ \sum_{K\in\mathcal{T}_h} \rho_K^2 |K| \right\}^{\frac{1}{2}}, \tag{7.1}$$

where $c_{\mathbf{S},\Omega} := \min_{K\in\mathcal{T}_h} c_{\mathbf{S},K}$ *and* h_Ω *is the diameter of the domain* Ω.

Proof. Using the Green theorem and the Cauchy–Schwarz inequality, we have

$$(\mathbf{r}_h, \nabla\varphi) = -(\nabla\cdot\mathbf{r}_h, \varphi) = -\sum_{K\in\mathcal{T}_h} (\rho_K, \varphi)_K \le \left\{ \sum_{K\in\mathcal{T}_h} \rho_K^2 |K| \right\}^{\frac{1}{2}} \|\varphi\|. \tag{7.2}$$

As $\varphi \in H_0^1(\Omega)$, we can now relate $\|\varphi\|$ to $|||\varphi|||$ using the Friedrichs inequality by

$$\|\varphi\| \le \sqrt{C_{\mathrm{F},\Omega}}\, h_\Omega \|\nabla\varphi\| \le \sqrt{\frac{C_{\mathrm{F},\Omega}}{c_{\mathbf{S},\Omega}}}\, h_\Omega |||\varphi|||\,. \tag{7.3}$$

Considering $|||\varphi||| = 1$ and combining (7.2) and (7.3) proves the statement. As for the value of $C_{\mathrm{F},\Omega}$, we refer to, e.g., Nečas [29, Section 1.2] or Rektorys [34, Chapter 30]; it ranges between $1/\pi^2$ and 1. Note that h_Ω may be replaced by the infimum over the thicknesses of Ω in the given direction, cf., e.g., [49]. ∎

We point out that (7.1) can be rewritten in the algebraic form as

$$\eta_{\mathrm{AE}}^{(1)} = \sqrt{\frac{C_{\mathrm{F},\Omega}}{c_{\mathbf{S},\Omega}}}\, h_\Omega \sqrt{R^t \mathbb{D}^{-1} R} = \sqrt{\frac{C_{\mathrm{F},\Omega}}{c_{\mathbf{S},\Omega}}} h_\Omega \, \|R\|_{\mathbb{D}^{-1}}, \tag{7.4}$$

where $\mathbb{D} := \mathrm{diag}(|\ell^{-1}(k)|)_{k=1}^N$ is a finite volume-type mass matrix and ℓ represents the enumeration of elements in $\mathcal{T}_h$ defined in Section 3.1.

7.2. Estimate based on the energy norm of the algebraic error. Inspired by Theorem 5.5, consider the approximation of (5.9) by the finite volume scheme given in Section 3.1. It consists in finding $e_h \in \mathbb{P}_0(\mathcal{T}_h)$ such that

$$\sum_{\sigma\in\mathcal{E}_K} U_{K,\sigma} = \rho_K |K| \qquad \forall K \in \mathcal{T}_h, \tag{7.5}$$

where $U_{K,\sigma}$ are the prescribed fluxes, which depend linearly on the values of e_h. In matrix form, this leads to

$$\mathbb{S}E = R, \tag{7.6}$$

where $\mathbb{S}$ is the matrix from (3.2). The matrix $\mathbb{S}$ is symmetric and positive definite (SPD), see Section 3.1, so that it induces an algebraic energy norm $\|\cdot\|_{\mathbb{S}}$ by $\|X\|_{\mathbb{S}}^2 := X^t\mathbb{S}X$ for a vector $X \in \mathbb{R}^N$. We now shed some light on the relationship between η_{AE} and $\|E\|_{\mathbb{S}}$.

Let us construct a postprocessed error $\tilde{e}_h \in \mathbb{P}_2(\mathcal{T}_h)$ from e_h and $U_{K,\sigma}$ given by (7.5) as described in Section 3.2 and put $\mathbf{q}_h := -\mathbf{S}\nabla\tilde{e}_h$. Then $\mathbf{q}_h \in \mathbf{RTN}(\mathcal{T}_h)$ and $\nabla\cdot\mathbf{q}_h = \rho_h$ by (7.5), so that

$$\eta_{\mathrm{AE}} \le \eta_{\mathrm{AE}}^{(2)} := \|\mathbf{S}^{-\frac{1}{2}}\mathbf{q}_h\| \tag{7.7}$$

follows directly from definition (5.3) of η_{AE} and the Cauchy–Schwarz inequality, see the proof of Theorem 5.5. Suppose for the moment that $\tilde{e}_h \in W_0(\mathcal{T}_h)$ and that $\mu_K = 0$ in (3.5b) for all $K \in \mathcal{T}_h$. Under these conditions and using the Green theorem,

$$\begin{aligned}
\big(\eta_{\mathrm{AE}}^{(2)}\big)^2 = \|\mathbf{S}^{-\frac{1}{2}}\mathbf{q}_h\|^2 &= \sum_{K\in\mathcal{T}_h} (\mathbf{S}\nabla\tilde{e}_h, \nabla\tilde{e}_h)_K \\
&= \sum_{K\in\mathcal{T}_h} \big\{ (-\nabla\cdot(\mathbf{S}\nabla\tilde{e}_h), \tilde{e}_h)_K + \langle \mathbf{S}\nabla\tilde{e}_h|_K\cdot\mathbf{n}, \tilde{e}_h\rangle_{\partial K} \big\} \\
&= \sum_{K\in\mathcal{T}_h} (-\nabla\cdot(\mathbf{S}\nabla\tilde{e}_h), \tilde{e}_h)_K = \sum_{K\in\mathcal{T}_h} e_K \rho_K |K| = \|E\|_{\mathbb{S}}^2.
\end{aligned}$$

Here the term $\sum_{K\in\mathcal{T}_h}\langle \mathbf{S}\nabla\tilde{e}_h|_K\cdot\mathbf{n},\tilde{e}_h\rangle_{\partial K}$ vanishes due to the fact that $\mathbf{S}\nabla\tilde{e}_h\cdot\mathbf{n}$ is sidewise constant as $\mathbf{S}\nabla\tilde{e}_h\in\mathbf{RTN}(\mathcal{T}_h)$ and the assumption $\tilde{e}_h\in W_0(\mathcal{T}_h)$. Unfortunately, as discussed in Section 3.2, $\tilde{e}_h$ in the finite volume method does not in general belong to the space $W_0(\mathcal{T}_h)$. Numerical experiments however show that the violations of the means of traces continuity are typically very slight. Therefore

$$\eta_{\mathrm{AE}}\le\eta_{\mathrm{AE}}^{(2)}=\|\mathbf{S}^{-\frac{1}{2}}\mathbf{q}_h\|\approx\|E\|_{\mathbb{S}},\tag{7.8}$$

and $\|E\|_{\mathbb{S}}$ suggest itself as an a posteriori algebraic error estimate. We now switch to linear algebra considerations of estimating $\|E\|_{\mathbb{S}}$.

Let a system of the form (3.2) be given and let $\mathbb{S}$ be SPD, so the conjugate gradient method (CG) [21] can be used. Let P_n^{CG} be the CG approximation to the solution P computed at the iteration step n, $P^{\mathrm{a}}=P_n^{\mathrm{CG}}$, $\mathbb{S}P_n^{\mathrm{CG}}=H-R_n^{\mathrm{CG}}$, see (4.1), $E_n^{\mathrm{CG}}:=P-P_n^{\mathrm{CG}}$, $\mathbb{S}E_n^{\mathrm{CG}}=R_n^{\mathrm{CG}}$, see (7.6). Since the original paper [21] it is known that the Euclidean norm of the residual $\|R_n^{\mathrm{CG}}\|$ does not represent a reliable measure of the quality of the CG approximation P_n^{CG}. CG minimizes the algebraic energy norm of the error $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}$ over the Krylov subspaces

$$\mathcal{K}_n(\mathbb{S},R_0^{\mathrm{CG}}):=\mathrm{span}\{R_0^{\mathrm{CG}},\mathbb{S}R_0^{\mathrm{CG}},\dots,\mathbb{S}^{n-1}R_0^{\mathrm{CG}}\}=\mathrm{span}\{R_0^{\mathrm{CG}},R_1^{\mathrm{CG}},\dots,R_{n-1}^{\mathrm{CG}}\},$$

$R_0^{\mathrm{CG}}:=H-\mathbb{S}P_0^{\mathrm{CG}}$. Therefore $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}$ is the appropriate convergence measure which should be used for evaluation of the algebraic error. It can unfortunately not be computed and its efficient estimation is nontrivial. Using the inequalities

$$\frac{1}{\sigma_{\max}(\mathbb{S})}\,\|R_n^{\mathrm{CG}}\|^2\le\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}^2=\|R_n^{\mathrm{CG}}\|_{\mathbb{S}^{-1}}^2\le\frac{1}{\sigma_{\min}(\mathbb{S})}\,\|R_n^{\mathrm{CG}}\|^2,\tag{7.9}$$

where $\sigma_{\max}$ and $\sigma_{\min}$ denote respectively the largest and the smallest singular values (eigenvalues) of the matrix $\mathbb{S}$, the algebraic energy norm of the error $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}$ can be approximated for *well-conditioned* $\mathbb{S}$ by the Euclidean norm of the CG residual, cf. [27, Section 4]. In many practical cases $\mathbb{S}$ is, however, ill-conditioned, and this approach can give misleading information. In practice, preconditioning is used to accelerate convergence. In theory, preconditioned CG (PCG) can be viewed as CG applied to the preconditioned system, and therefore (7.9) holds for the quantities relevant to PCG, cf. [46]. However, the energy norm of the error in PCG is identical to the energy norm of the error in CG applied to the unpreconditioned system (i.e. to the original data), see [46, Section 3, pp. 794–795]. Consequently, if the condition number of the preconditioned system is small, then the Euclidean norm of the preconditioned residual provides a good information on the size of the energy norm of the error with respect to the original data. Upper bounds can be in theory constructed using the Gauss-Radau quadrature, which uses the *a-priori knowledge* of $\sigma_{\min}(\mathbb{S})$ or using techniques based on the anti-Gauss quadrature, cf. [18, 19, 20, 27, 14]. Due to rounding errors, the upper bounds can not be guaranteed in practice, see [45, 46]. Despite some open questions and intricate implementation issues, which are out of the scope of this paper, estimates for $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}$ can be computed at a very low cost.

In the sequel, we restrict ourselves to presenting *a lower bound* for $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}$, following [21, 45, 46, 28]. Its justification is based on the matching moments idea which can be considered a basic principle behind CG and other Krylov subspace methods, see [43]. In CG, the approximate solution is updated using the formula

$$P_{n+1}^{\mathrm{CG}}=P_n^{\mathrm{CG}}+\mu_n^{\mathrm{CG}}D_n^{\mathrm{CG}},$$

where μ_n^{CG} is the scalar coefficient giving the minimum of the energy norm of the error along the line defined by the previous approximation P_n^{CG} and the search direction D_n^{CG}, see [21]. Considering ν additional conjugate gradients iterations, we obtain, see [45, 27] and a detailed survey in [28, Sections 3.3 and 5.3],

$$\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}^2 = \sum_{j=n}^{n+\nu} \mu_j^{\mathrm{CG}} \|R_j^{\mathrm{CG}}\|^2 + \|E_{n+\nu}^{\mathrm{CG}}\|_{\mathbb{S}}^2. \tag{7.10}$$

The squared algebraic energy norm of the error is at step n approximated from below by

$$\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}^2 \approx \left(\hat{\eta}_{\mathrm{AE}}^{(2)}\right)^2 := \sum_{j=n}^{n+\nu} \mu_j^{\mathrm{CG}} \|R_j^{\mathrm{CG}}\|^2, \tag{7.11}$$

with the inaccuracy given by the squared size of the algebraic energy error at the $(n+\nu)$-th step. If $\|E_{n+\nu}^{\mathrm{CG}}\|_{\mathbb{S}}^2$ is significantly smaller than $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}^2$, then $\hat{\eta}_{\mathrm{AE}}^{(2)}$ represents an accurate approximation of $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}$. The choice of ν depends on the problem to be solved, and an efficient algorithm for an adaptive choice of ν is still under investigation.

7.3. Guaranteed upper bound using a particular construction of the vector function $\mathbf{r}_h$. The following corollary is an immediate consequence of the definition of η_{AE} in Theorem 5.2, cf. the proof of Theorem 5.5:

COROLLARY 7.2 (Algebraic error estimator based on an explicitly constructed $\mathbf{r}_h$). *Consider an arbitrary $\mathbf{r}_h \in \mathbf{RTN}(\mathcal{T}_h)$ such that $\nabla \cdot \mathbf{r}_h = \rho_h$. Then the algebraic error estimator η_{AE} from Theorem 5.2 can be bounded from above by*

$$\eta_{\mathrm{AE}} \le \eta_{\mathrm{AE}}^{(3)}(\mathbf{r}_h) := \|\mathbf{S}^{-\frac{1}{2}} \mathbf{r}_h\|. \tag{7.12}$$

Proof. Let $\mathbf{r}_h \in \mathbf{RTN}(\mathcal{T}_h)$ such that $\nabla \cdot \mathbf{r}_h = \rho_h$. Then

$$\eta_{\mathrm{AE}} \le \sup_{\substack{\varphi \in H_0^1(\Omega) \\ |||\varphi|||=1}} (\mathbf{r}_h, \nabla\varphi) = \sup_{\substack{\varphi \in H_0^1(\Omega) \\ |||\varphi|||=1}} (\mathbf{S}^{-\frac{1}{2}} \mathbf{r}_h, \mathbf{S}^{\frac{1}{2}} \nabla\varphi) \le \|\mathbf{S}^{-\frac{1}{2}} \mathbf{r}_h\| = \eta_{\mathrm{AE}}^{(3)}(\mathbf{r_h})$$

using the definition of η_{AE} in Theorem 5.2 and the Cauchy-Schwarz inequality. ∎

We now present a simple algorithm with a linear complexity in the number of mesh elements which finds an appropriate function $\mathbf{r}_h$ without a need of solving any global problem. The first step is to find an enumeration of the elements of $\mathcal{T}_h$ such that for each K_i, there is a side $\sigma \in \mathcal{E}_{K_i}$ which does not lie on the boundary of $\cup_{j=1}^{i-1} K_j$. Such an enumeration of the elements of $\mathcal{T}_h$ can be always found for meshes consisting of simplices using, e.g., the standard depth-first search in the graph associated with the partition $\mathcal{T}_h$. The algorithm is described as follows: set $\mathcal{T} := \mathcal{T}_h$, $i := N$, and while $i \ge 2$:

1. find $K \in \mathcal{T}$ such that there is a side $\sigma \in K$ which lies on the boundary of $\mathcal{T}$;
2. set $K_i := K$, $\mathcal{T} := \mathcal{T} \setminus K$, $i := i - 1$.

Finally denote as K_1 the last element.

With such an enumeration, we construct $\mathbf{r}_h$ locally on each element of $\mathcal{T}_h$ while proceeding sequentially for $i = 1, 2, \ldots, N$:

1. find $\mathbf{r}_i \in \mathbf{RTN}(K_i)$ such that

$$\mathbf{r}_i = \arg \min_{\widetilde{\mathbf{r}} \in \widetilde{\mathbf{RTN}}(K_i)} \|\mathbf{S}^{-\frac{1}{2}} \widetilde{\mathbf{r}}\|_{K_i},$$

where $\widetilde{\mathbf{RTN}}(K_i)$ are functions of $\mathbf{RTN}(K_i)$ such that

$$\nabla \cdot \widetilde{\mathbf{r}}_i = \rho_{K_i}, \quad \widetilde{\mathbf{r}}_i \cdot \mathbf{n}_\sigma = \mathbf{r}_h \cdot \mathbf{n}_\sigma \text{ on all } \sigma \in \mathcal{E}_{K_i} \cap \mathcal{E}_{K_j},\ j < i;$$

2. set $\mathbf{r}_h|_{K_i} := \mathbf{r}_i$.

The vector function $\mathbf{r}_h$ constructed in this way is not optimal but, as shown in the experiments, it represents a good candidate for giving a useful estimate.

8. Numerical experiments. In this section we illustrate the proposed estimates and stopping criteria on model problems with both homogeneous and inhomogeneous diffusion tensors. We will consider two examples.

EXAMPLE 8.1 (Laplace equation). *We consider the Laplace equation* $-\Delta p = 0$, *i.e.,* $\mathbf{S} = \mathbb{I}$ *and* $f = 0$ *in* (1.1), $\Omega = (-1,1) \times (-1,1)$. *Let*

$$p(x,y) = \exp\left(\frac{x}{10}\right) \cos\left(\frac{y}{10}\right)$$

and let g *in* (1.1) *be defined by the values of this* p *on the boundary* Γ *of* Ω. *Then* p *is the (weak as well as classical) solution of problem* (1.1).

EXAMPLE 8.2 (Problem with an inhomogeneous diffusion tensor). *We consider the diffusion equation* $-\nabla \cdot (\mathbf{S}\nabla p) = 0$ *and suppose that* $\Omega = (-1,1) \times (-1,1)$ *is divided into four subdomains* Ω_i *corresponding to the axis quadrants numbered counterclockwise. Let* $\mathbf{S}$ *be piecewise constant and equal to* $s_i\mathbb{I}$ *in* Ω_i. *Then with the two choices of* s_i *presented in Table 8.1, the analytical solution in each subdomain* Ω_i *has in polar coordinates* (ϱ, ϑ) *the form*

$$p(\varrho,\vartheta)|_{\Omega_i} = \varrho^\alpha (a_i \sin(\alpha\vartheta) + b_i \cos(\alpha\vartheta)) \tag{8.1}$$

with the Dirichlet boundary condition imposed accordingly to (8.1), *where the coefficients* α, a_i, *and* b_i *are also given in Table 8.1, see [37]. Note that* p *belongs only to* $H^{1+\alpha}(\Omega)$ *and it exhibits a singularity at the origin. It is continuous, but only the normal component of its flux* $-\mathbf{S}\nabla p$ *is continuous across the interfaces.*

$s_1 = s_3 = 5$, $s_2 = s_4 = 1$				$s_1 = s_3 = 100$, $s_2 = s_4 = 1$			
$\alpha =$	0.53544095			$\alpha =$	0.12690207		
$a_1 =$	0.44721360	$b_1 =$	1.00000000	$a_1 =$	0.10000000	$b_1 =$	1.00000000
$a_2 =$	−0.74535599	$b_2 =$	2.33333333	$a_2 =$	−9.60396040	$b_2 =$	2.96039604
$a_3 =$	−0.94411759	$b_3 =$	0.55555556	$a_3 =$	−0.48035487	$b_3 =$	−0.88275659
$a_4 =$	−2.40170264	$b_4 =$	−0.48148148	$a_4 =$	7.70156488	$b_4 =$	−6.45646175

TABLE 8.1
The values of the coefficients in (8.1) *for the two choices of the diffusion tensor* $\mathbf{S}$.

In our experiments we use the finite volume scheme (3.1), (3.3), which we extend from triangular grids admissible in the sense of [17, Definition 9.1] to strictly Delaunay triangular meshes, cf. [17, Example 9.1]. For the diffusion tensor the harmonic averaging is employed and modified by taking into account the distances of the circumcenters $\mathbf{x}_K$, $K \in \mathcal{T}_h$, from the sides of K; for details, we refer to [51].

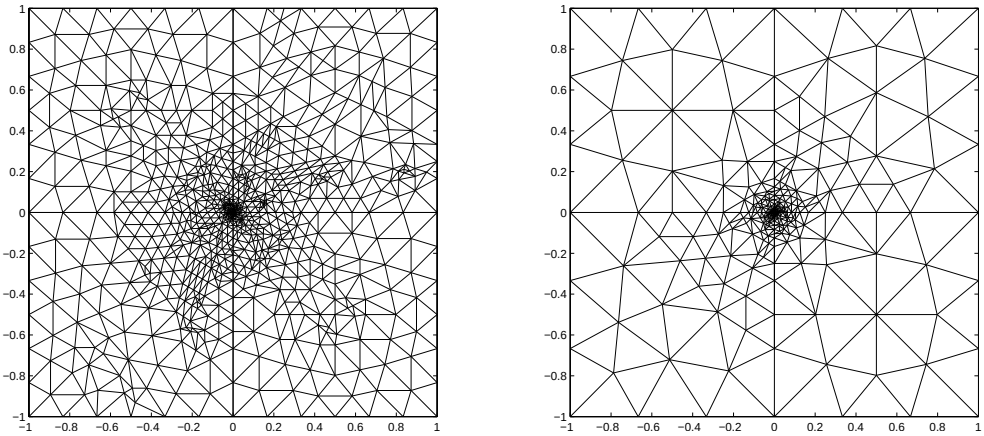

FIG. 8.1. *Adaptively refined mesh with 1812 elements for Example 8.2 with* $s_1 = s_3 = 5$, $s_2 = s_4 = 1$ *(left) and with 1736 elements for the problem with* $s_1 = s_3 = 100$, $s_2 = s_4 = 1$ *(right).*

We start our computations with an unstructured mesh $\mathcal{T}_h$ of Ω consisting of 112 elements. In Example 8.1 the mesh is refined uniformly, i.e., each triangular element in $\mathcal{T}_h$ is subdivided into four elements. In Example 8.2 it is refined adaptively. The adaptive mesh refinement strategy is described in detail in [51]; the essential point is in equilibration of the estimated local discretization errors while keeping the mesh strictly Delaunay. The refinement process is stopped when the number of elements in $\mathcal{T}_h$ exceeds 1700, which results in all cases in algebraic systems of similar size. This relatively small number of elements was chosen because of the second choice of coefficients in Example 8.2. Due to significant singularity, for around 2000 triangles, the diameter of the smallest triangles near the origin is about 10^{-15}. The final mesh in Example 8.1 consists of 1792 elements, in the first case of Example 8.2 of 1812 elements, and in the second case of Example 8.2 of 1736 elements. The last two meshes are shown in Figure 8.1. Recall that the matrix size is equal to the number of mesh elements.

The arising algebraic systems (3.2) are solved approximately by CG preconditioned by the incomplete Cholesky factorization with no fill-in (IC(0)), see [26]. For illustrative purposes, we use for all meshes the zero initial guess. In practical computations, the approximate solution from the previous refinement level should be interpolated onto the current mesh and used as a starting vector, together with the possible scaling, see [28, p. 530]. In our experiments, for each approximate solution $P^{\mathrm{a}} = P_n^{\mathrm{CG}}$ of (3.2), we evaluate the estimator η_{NC} defined in Theorem 5.2 as $|||\tilde{p}_h^{\mathrm{a}} - \mathcal{I}_{\mathrm{Os}}(\tilde{p}_h^{\mathrm{a}})|||$ (we consider the additional error from the inhomogeneous boundary condition negligible). Then we compute the algebraic error estimators described in Section 7. Note that η_{O} is zero since $f = 0$ in both examples. CG is stopped when the *local stopping criterion* (6.3) based on the estimator $\eta_{\mathrm{AE}}^{(3)}(\mathbf{r}_h)$ is satisfied, i.e., when

$$\eta_{\mathrm{AE},K}^{(3)}(\mathbf{r}_h) := \|\mathbf{S}^{-\frac{1}{2}}\mathbf{r}_h\|_K \leq \gamma\, \eta_{\mathrm{NC},K} \qquad \forall K \in \mathcal{T}_h.$$

In order to illustrate the behavior of the nonconforming and algebraic error estimators, we have chosen $\gamma = 10^{-3}$. In practical computations, it is advisable to use a value of γ much closer to one, in dependence on the given problem, and $\eta_{\mathrm{NC},K}$ should not be evaluated at every CG step, see the comment on efficiency in Section 9 below.

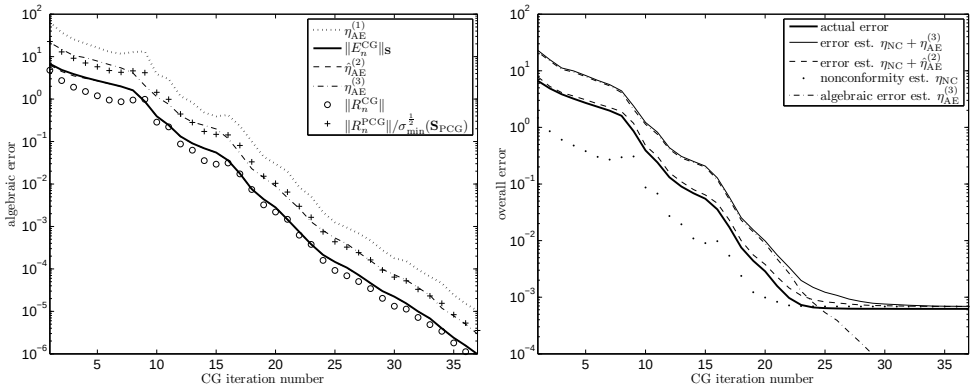

FIG. 8.2. *Different errors and estimators for Example 8.1, uniformly refined mesh with* 1792 *elements. Left: algebraic error only; right: overall error.*

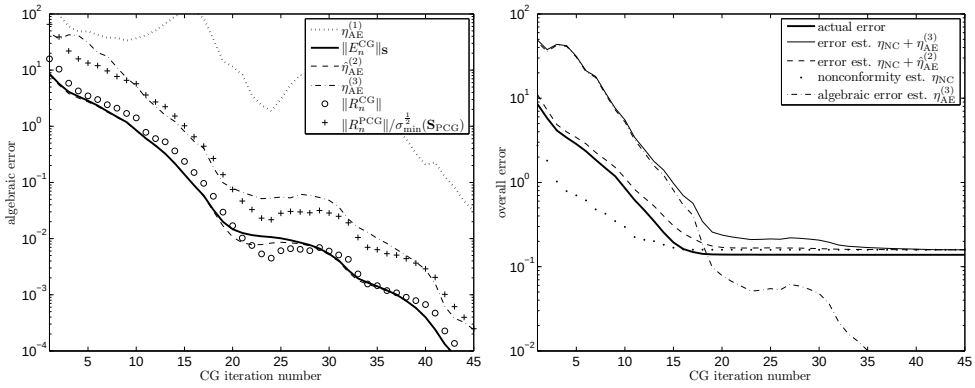

FIG. 8.3. *Different errors and estimators for Example 8.2 with* $s_1 = s_3 = 5$, $s_2 = s_4 = 1$, *adaptively refined mesh with* 1812 *elements. Left: algebraic error only; right: overall error.*

Results for meshes obtained at the last stage of the uniform or adaptive mesh refinement process are illustrated in Figures 8.2–8.4. The results for the Laplace equation in Example 8.1 are plotted in Figure 8.2. The results for Example 8.2 with the inhomogeneous $\mathbf{S}$ with s_i given in the left and right part of Table 8.1 are plotted in Figure 8.3 and Figure 8.4, respectively.

Left parts of Figures 8.2–8.4 show the values of the algebraic error estimators described in Section 7, together with the true algebraic energy norm of the error $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}$ (solid lines), the Euclidean norm of the algebraic residual $\|R_n^{\mathrm{CG}}\|$ (circles), and the upper bound $\|R_n^{\mathrm{PCG}}\|/\sigma_{\min}^{1/2}(\mathbb{S}_{\mathrm{PCG}})$ (crosses) for $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}$ constructed from the preconditioned residual, see (7.9). Please note that $\sigma_{\min}(\mathbb{S}_{\mathrm{PCG}})$ is not available and must be approximated. The true algebraic energy error $\|E_n^{\mathrm{CG}}\|_{\mathbb{S}}$ is evaluated by solving $\mathbb{S}E_n^{\mathrm{CG}} = R_n^{\mathrm{CG}}$ using a direct solver. The estimator $\eta_{\mathrm{AE}}^{(1)}$ based on the weighted norm of the algebraic residual vector, see Lemma 7.1, is plotted by dotted lines. The estimate $\hat{\eta}_{\mathrm{AE}}^{(2)}$ evaluated for $\nu = 5$ is plotted by dashed lines, and the estimator $\eta_{\mathrm{AE}}^{(3)}$ of Section 7.3 by dash-dotted lines.

The estimate $\hat{\eta}_{\mathrm{AE}}^{(2)}$ is close to $\|E\|_{\mathbb{S}}$, with some visible but insignificant underestimations (due to the rather slow convergence of CG, cf. [45, 46]) in Figures 8.3 and 8.4.

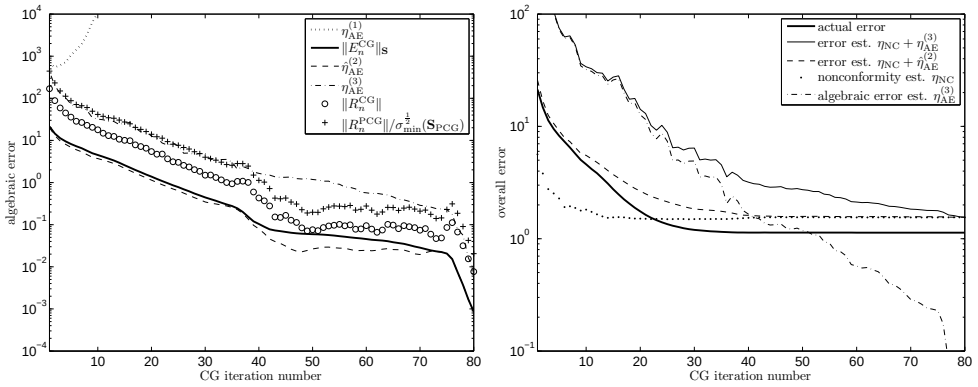

FIG. 8.4. *Different errors and estimators for Example 8.2 with* $s_1 = s_3 = 100$, $s_2 = s_4 = 1$, *adaptively refined mesh with* 1736 *elements. Left: algebraic error only; right: overall error.*

The estimator $\eta^{(3)}_{\rm AE}$ represents a *guaranteed upper bound* for the algebraic error. The estimator $\eta^{(1)}_{\rm AE}$, as expected, provides the worst information among all considered measures of the algebraic error. This is in particular evident in Example 8.2 where the adaptive mesh refinement is employed, see Figures 8.3 and 8.4 (in Figure 8.4 it is out of scale for almost all iterations). For both examples, $\|R^{\rm CG}_n\|$ is remarkably close to $\|E^{\rm CG}_n\|_{\mathbb{S}}$. For examples of a different behavior see [46]. The upper bound constructed from the preconditioned residual is here quite tight.

On right parts of Figures 8.2–8.4 we present the actual energy (semi-)norm of the overall error $|||p-\tilde p^{\rm a}_h|||$ (bold solid lines). We compute it in each triangle by the 7-point quadrature formula, see, e.g., [54, Section 9.10] (we consider the associated additional error negligible). The guaranteed upper bound $\eta_{\rm NC}+\eta^{(3)}_{\rm AE}$ on $|||p-\tilde p^{\rm a}_h|||$ is represented by solid lines, while its components, the nonconformity estimator $\eta_{\rm NC}$ and the algebraic error estimator $\eta^{(3)}_{\rm AE}$, are plotted by dots and dash-dotted lines, respectively. For comparison, we also include the estimate $\eta_{\rm NC}+\hat\eta^{(2)}_{\rm AE}$ plotted by dashed lines.

Figures 8.2–8.4 show that for small number of iterations the algebraic part of the error dominates. As the number of iterations of the conjugate gradient method grows, the algebraic part of the error drops to the level of the nonconformity error, which is reflected by the fact that the curves of $\eta_{\rm NC}$ and $\eta^{(3)}_{\rm AE}$ intersect. While $\eta_{\rm NC}$ almost stagnates, the estimate on the algebraic error $\eta^{(3)}_{\rm AE}$ further decreases and it ultimately gets negligible in comparison with the nonconformity error. Our stopping criteria for iterative solvers (6.1) and (6.3) essentially state that it is meaningless to continue the algebraic computation after $\eta^{(3)}_{{\rm AE},K}(\mathbf{r}_h) \approx \gamma\,\eta_{{\rm NC},K}$ is reached.

The quality of our estimates, i.e., the effectivity indices $(\eta_{\rm NC}+\eta^{(1)}_{\rm AE})/|||p-\tilde p^{\rm a}_h|||$ (dotted line), $(\eta_{\rm NC}+\hat\eta^{(2)}_{\rm AE})/|||p-\tilde p^{\rm a}_h|||$ (dashed line), and $(\eta_{\rm NC}+\eta^{(3)}_{\rm AE})/|||p-\tilde p^{\rm a}_h|||$ (solid line), is illustrated in the left part of Figure 8.5 and in Figure 8.6. Estimate $\eta^{(1)}_{\rm AE}$ overestimates largely the actual algebraic error and the corresponding effectivity index is very poor (in the right part of Figure 8.6 it is completely out of scale). Recall that the estimate $\eta_{\rm NC}+\eta^{(3)}_{\rm AE}$ gives a guaranteed upper bound. Its effectivity index is very reasonable even in the first PCG iterations in the second case of Example 8.2. Finally, even though $\hat\eta^{(2)}_{\rm AE}$ does not represent a guaranteed upper bound for $\eta^{(2)}_{\rm AE}$, the estimate

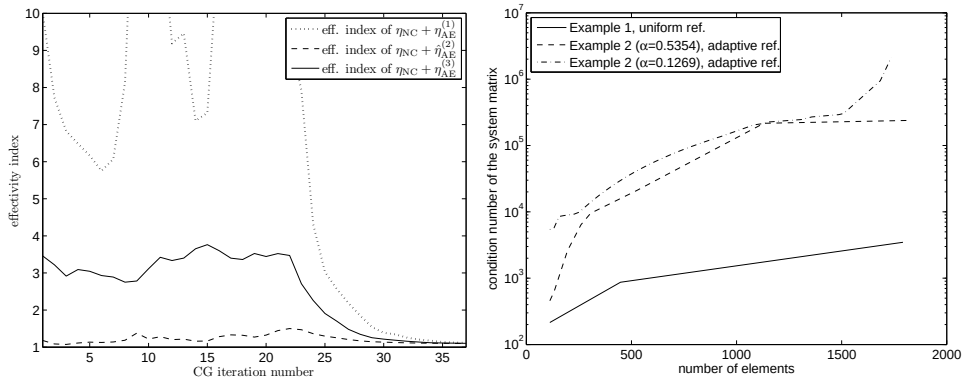

FIG. 8.5. *Effectivity indices for Example 8.1 (left), and condition number of system matrix* $\mathbb{S}$ *for Examples 8.1 and 8.2 (right).*

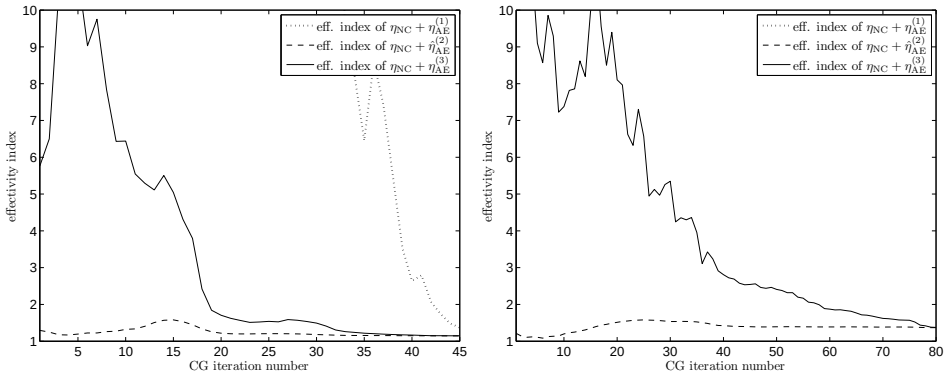

FIG. 8.6. *Effectivity indices for Example 8.2 with* $s_1 = s_3 = 5$, $s_2 = s_4 = 1$ *(left) and* $s_1 = s_3 = 100$, $s_2 = s_4 = 1$ *(right). The dotted line is essentially out of the scale of the figure.*

$\eta_{\mathrm{NC}} + \hat{\eta}_{\mathrm{AE}}^{(2)}$ gives in our experiments very tight estimates for the overall error. The effectivity index is here in all cases remarkably close to one.

Without taking into consideration the algebraic part of the error, it is sometimes claimed in the literature that adaptive mesh refinement can provide *an arbitrary accurate numerical solution*. Similar claims should be in some cases examined and revisited. Adaptive discretization in the presence of singularity can lead to highly ill-conditioned systems of linear algebraic equations. This can have two main effects:

- the iterative solvers can become slow and the computation of the numerical solution can become expensive;
- the maximum attainable accuracy of the (direct as well as iterative) linear algebraic solvers can for highly ill-conditioned systems become very poor, which can prevent reaching the desired accuracy of the numerical solution of the original problem regardless how small the discretization error becomes.

Right part of Figure 8.5 shows for our examples the dependence of the spectral condition number of the system matrix $\mathbb{S}$ on the number of elements in the mesh. In the case of the homogeneous diffusion tensor and the uniform mesh refinement of Example 8.1, the condition number of $\mathbb{S}$ is growing according to the well-known theoretical result as $O(N)$. In Example 8.2 with inhomogeneous diffusion coefficients, adaptive mesh

refinement compensates for the effect of the singularity. This results in the growth of the condition number of the system matrix $\mathbb{S}$, see the right part of Figure 8.5. If we proceed with the refinement, the condition number of $\mathbb{S}$ will soon reach reach the value of the inverse of machine precision, which will make algebraic computations practically meaningless. Though a more detailed discussion of this phenomenon is beyond the scope of this paper, we believe that its role can be substantial and it will have to be systematically investigated in a near future. If the conditioning of $\mathbb{S}$ is reasonably bounded independently of the mesh, see, e.g., [11, Section 9.6], then the matter is resolved.

9. Concluding remarks. Deriving tight a posteriori estimates under the assumption that the associated systems of linear algebraic equations are solved exactly is much easier than without this assumption. It however precludes the efficient use of such estimates in practical large scale computations, where the linear systems, solved by iterative algebraic solvers, are never solved exactly, and should even be solved inexactly on purpose.

Efficient usage of iterative algebraic solvers requires balancing the algebraic and discretization errors. It is useless to make a large number of algebraic solver iterations after the algebraic error drops significantly below the discretization error. A stopping criterion must be cheap to compute. This may seem in contradiction with evaluation of the η_{NC} estimator presented above, with the cost proportional to the number of mesh elements. But η_{NC} does not need to be evaluated at each iteration of CG. A viable strategy is to monitor the algebraic convergence at a negligible cost using the algebraic error estimator $\hat{\eta}_{\mathrm{AE}}^{(2)}$ (in addition to monitoring the CG and PCG residuals), and to evaluate any other estimators only after $\hat{\eta}_{\mathrm{AE}}^{(2)}$ drops below a certain level. The strategy of evaluating error estimators can be tailored for a given problem in order to minimize the overall extra cost in comparison with the cost of actual computations.

If an adaptive mesh refinement leads in the presence of singularity to pathologically ill-conditioned linear algebraic systems, this can eventually prevent obtaining a numerical solution with a single digit of accuracy. Modeling, discretization, and computation form interconnected stages of a *single solution process*. As stated in [8, p. 273], "The purpose of computation is not to produce a solution with least error but to produce reliably, robustly and *affordably* a solution which is within a user-specified tolerance." Therefore the *errors on the different stages should be in balance*, see, e.g., [44]. Considering the numerical analysis and the discretization stages separately from computations is philosophically wrong. Similar approaches will lead in solving difficult problems to dead ends.

Acknowledgments. This work was initiated during the summer school CEMRACS organized by the Jacques-Louis Lions laboratory (LJLL) in summer 2007 in Luminy, France and the authors gratefully acknowledge all the support. The second author thanks for the support during his visit of the LJLL in September 2008. We would also like to thank the anonymous referees for their valuable comments and suggestions.

REFERENCES

[1] Y. Achdou, C. Bernardi, and F. Coquel, *A priori and a posteriori analysis of finite volume discretizations of Darcy's equations*, Numer. Math., 96 (2003), pp. 17–42.

[2] M. Ainsworth and J. T. Oden, *A posteriori error estimation in finite element analysis*, Pure and Applied Mathematics (New York), Wiley-Interscience [John Wiley & Sons], New York, 2000.
[3] M. Arioli, *A stopping criterion for the conjugate gradient algorithm in a finite element method framework*, Numer. Math., 97 (2004), pp. 1–24.
[4] M. Arioli and D. Loghin, *Stopping criteria for mixed finite element problems*, Electron. Trans. Numer. Anal., 29 (2007/08), pp. 178–192.
[5] M. Arioli, D. Loghin, and A. J. Wathen, *Stopping criteria for iterations in finite element methods*, Numer. Math., 99 (2005), pp. 381–410.
[6] I. Babuška, *Numerical stability in problems of linear algebra*, SIAM J. Numer. Anal., 9 (1972), pp. 53–77.
[7] I. Babuška and W. C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal., 15 (1978), pp. 736–754.
[8] B. J. C. Baxter and A. Iserles, *On the foundations of computational mathematics*, in Handbook of numerical analysis, Vol. XI, Handb. Numer. Anal., XI, North-Holland, Amsterdam, 2003, pp. 3–34.
[9] R. Becker, *An adaptive finite element method for the Stokes equations including control of the iteration error*, in ENUMATH 97 (Heidelberg), World Sci. Publ., River Edge, NJ, 1998, pp. 609–620.
[10] R. Becker, C. Johnson, and R. Rannacher, *Adaptive error control for multigrid finite element methods*, Computing, 55 (1995), pp. 271–288.
[11] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Texts in Applied Mathematics, Springer, 3rd ed., 2007.
[12] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods*, vol. 15 of Springer Series in Computational Mathematics, Springer-Verlag, New York, 1991.
[13] C. Burstedde and A. Kunoth, *Fast iterative solution of elliptic control problems in wavelet discretization*, J. Comput. Appl. Math., 196 (2006), pp. 299–319.
[14] D. Calvetti, S. Morigi, L. Reichel, and F. Sgallari, *Computable error bounds and estimates for the conjugate gradient method*, Numer. Algorithms, 25 (2000), pp. 75–88. Mathematical journey through analysis, matrix theory and scientific computation (Kent, OH, 1999).
[15] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, vol. 4 of Studies in Mathematics and its Applications, North-Holland, Amsterdam, 1978.
[16] P. Deuflhard, *Cascadic conjugate gradient methods for elliptic partial differential equations: algorithm and numerical results*, in Domain decomposition methods in scientific and engineering computing (University Park, PA, 1993), vol. 180 of Contemp. Math., Amer. Math. Soc., Providence, RI, 1994, pp. 29–42.
[17] R. Eymard, T. Gallouët, and R. Herbin, *Finite volume methods*, in Handbook of Numerical Analysis, Vol. VII, North-Holland, Amsterdam, 2000, pp. 713–1020.
[18] G. H. Golub and G. Meurant, *Matrices, moments and quadrature*, in Numerical analysis 1993 (Dundee, 1993), vol. 303 of Pitman Res. Notes Math. Ser., Longman Sci. Tech., Harlow, 1994, pp. 105–156.
[19] G. H. Golub and G. Meurant, *Matrices, moments and quadrature II: how to compute the norm of the error in iterative methods*, BIT, 37 (1997), pp. 687–705.
[20] G. H. Golub and Z. Strakoš, *Estimates in quadratic formulas*, Numer. Algorithms, 8 (1994), pp. 241–268.
[21] M. R. Hestenes and E. Stiefel, *Methods of conjugate gradients for solving linear systems*, J. Res. Natl. Bur. Stand., 49 (1952), pp. 409–436.
[22] N. J. Higham, *Accuracy and Stability of Numerical Algorithms*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 2002.
[23] K. Y. Kim, *A posteriori error analysis for locally conservative mixed methods*, Math. Comp., 76 (2007), pp. 43–66.
[24] Y. Maday and A. T. Patera, *Numerical analysis of a posteriori finite element bounds for linear functional outputs*, Math. Models Methods Appl. Sci., 10 (2000), pp. 785–799.
[25] D. Meidner, R. Rannacher, and J. Vihharev, *Goal-oriented error control of the iterative solution of finite element equations*, J. Numer. Math., 17 (2009), pp. 143–172.
[26] J. A. Meijerink and H. A. van der Vorst, *An iterative solution method for linear systems of which the coefficient matrix is a symmetric M-matrix*, Math. Comp., 31 (1977), pp. 148–162.
[27] G. Meurant, *The computation of bounds for the norm of the error in the conjugate gradient algorithm*, Numer. Algorithms, 16 (1997), pp. 77–87.
[28] G. Meurant and Z. Strakoš, *The Lanczos and conjugate gradient algorithms in finite pre-*

cision arithmetic, Acta Numer., 15 (2006), pp. 471–542.
[29] J. Nečas, *Les méthodes directes en théorie des équations elliptiques*, Masson, Paris, 1967.
[30] S. Nicaise, *A posteriori error estimations of some cell-centered finite volume methods*, SIAM J. Numer. Anal., 43 (2005), pp. 1481–1503.
[31] M. Ohlberger, *A posteriori error estimate for finite volume approximations to singularly perturbed nonlinear convection–diffusion equations*, Numer. Math., 87 (2001), pp. 737–761.
[32] A. T. Patera and E. M. Rønquist, *A general output bound result: application to discretization and iteration error estimation and control*, Math. Models Methods Appl. Sci., 11 (2001), pp. 685–712.
[33] A. Quarteroni and A. Valli, *Numerical approximation of partial differential equations*, vol. 23 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1994.
[34] K. Rektorys, *Variational Methods in Mathematics, Science, and Engineering*, Kluwer, Dordrecht, 1982.
[35] S. Repin, *A posteriori error estimation for nonlinear variational problems by duality theory*, Zapiski Nauchnykh Seminarov, 243 (1997), pp. 201–214.
[36] S. I. Repin and A. Smolianski, *Functional-type a posteriori error estimates for mixed finite element methods*, Russian J. Numer. Anal. Math. Modelling, 20 (2005), pp. 365–382.
[37] B. Rivière, M. F. Wheeler, and K. Banas, *Part II. Discontinuous Galerkin method applied to single phase flow in porous media*, Comput. Geosci., 4 (2000), pp. 337–349.
[38] U. Rüde, *Fully adaptive multigrid methods*, SIAM J. Numer. Anal., 30 (1993), pp. 230–248.
[39] U. Rüde, *Error estimators based on stable splittings*, in Proceedings of the 7th International Conference on Domain Decomposition in Science and Engineering Computing, Pennsylvania State University, D. Keyes, ed., vol. 180, Providence: American Mathematical Society, 1994, pp. 111–118.
[40] U. Rüde, *On the multilevel adaptive iterative method*, SIAM J. Sci. Comput., 15 (1994), pp. 577–586. Iterative methods in numerical linear algebra (Copper Mountain Resort, CO, 1992).
[41] Y. Saad, *Iterative Methods for Sparse Linear Systems*, SIAM, 2nd ed., 2003.
[42] V. Shaidurov and L. Tobiska, *The convergence of the cascadic conjugate-gradient method applied to elliptic problems in domains with re-entrant corners*, Math. Comp., 69 (2000), pp. 501–520.
[43] Z. Strakoš, *Model reduction using the Vorobyev moment problem*, Numer. Algorithms, 51 (2009), pp. 363–379.
[44] Z. Strakoš and J. Liesen, *On numerical stability in large scale linear algebraic computations*, ZAMM Z. Angew. Math. Mech., 85 (2005), pp. 307–325.
[45] Z. Strakoš and P. Tichý, *On error estimation in the conjugate gradient method and why it works in finite precision computations*, Electron. Trans. Numer. Anal., 13 (2002), pp. 56–80.
[46] Z. Strakoš and P. Tichý, *Error estimation in preconditioned conjugate gradients*, BIT, 45 (2005), pp. 789–817.
[47] R. S. Varga, *Matrix Iterative Analysis*, Springer, Berlin, 2 ed., 1992.
[48] R. Verfürth, *A review of a posteriori error estimation and adaptive mesh-refinement techniques*, Teubner-Wiley, Stuttgart, 1996.
[49] M. Vohralík, *On the discrete Poincaré–Friedrichs inequalities for nonconforming approximations of the Sobolev space H^1*, Numer. Funct. Anal. Optim., 26 (2005), pp. 925–952.
[50] ———, *A posteriori error estimates for lowest-order mixed finite element discretizations of convection-diffusion-reaction equations*, SIAM J. Numer. Anal., 45 (2007), pp. 1570–1599.
[51] ———, *Residual flux-based a posteriori error estimates for finite volume and related locally conservative methods*, Numer. Math., 111 (2008), pp. 121–158.
[52] ———, *Unified primal formulation-based a priori and a posteriori error analysis of mixed finite element methods*, Math. Comp., (2009). Accepted for publication.
[53] B. I. Wohlmuth and R. H. W. Hoppe, *A comparison of a posteriori error estimators for mixed finite element discretizations by Raviart–Thomas elements*, Math. Comp., 68 (1999), pp. 1347–1378.
[54] O. C. Zienkiewicz and R. L. Taylor, *The Finite Element Method. Volume I: The Basis*, Butterworth-Heinemann, Oxford, 5th ed., 2000.