Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 339 (2018) 320–340
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

A hybrid a posteriori error estimator for conforming finite element approximations☆

Difeng Cai, Zhiqiang Cai*

Department of Mathematics, Purdue University, West Lafayette, IN 47907-2067, USA

Received 7 March 2018; accepted 30 April 2018
Available online 16 May 2018

Highlights

- The hybrid estimator resolves unreliability of ZZ type estimators on coarse meshes.
- The hybrid estimator extends the improved ZZ estimator to higher order elements.
- The hybrid estimator is explicit and is more accurate than the residual estimator.

Abstract

This paper introduces a hybrid a posteriori error estimator for the conforming finite element method, which may be regarded as a combination of the explicit residual and the improved ZZ error estimators. With comparable cost, the hybrid estimator is more accurate than the residual estimator. It is shown that the hybrid estimator is reliable on all meshes, unlike estimators of the ZZ type. Moreover, the reliability constant is independent of the jump of the diffusion coefficients for elliptic interface problems under the monotonicity assumption of the coefficients. Finally, numerical examples confirm the robustness of the estimator with respect to coefficient jumps and also better effectivity index compared to the residual estimator.

Keywords: Finite element method; A posteriori error estimation; Adaptive mesh refinement; Diffusion problem

1. Introduction

Adaptive mesh refinement is necessary in the discretization of partial differential equations (PDEs) in order to handle computational challenges [1]. A posteriori error estimates play a crucial role in adaptive mesh refinement, where one tries to estimate the error by computing quantities (called error estimators) based on numerical solution as well as data from the underlying PDE. It is well known that the explicit residual error estimators (see, e.g., [2–6]) are computationally inexpensive with applications to a large class of problems. Moreover, for computationally challenging problems such as interface problems, proper weighted residual estimators (see, e.g., [5,7]) generate efficient meshes.

☆ This work was supported in part by the National Science Foundation under grant DMS-1522707.
* Corresponding author.
E-mail addresses: cai92@purdue.edu (D. Cai), caiz@purdue.edu (Z. Cai).

https://doi.org/10.1016/j.cma.2018.04.050
0045-7825/

However, it is also known that residual estimators usually overestimate the true error by a large margin compared to estimators of the Zienkiewicz–Zhu (ZZ) type (cf. [8]). In this paper we introduce a hybrid a posteriori error estimator for the conforming finite element method, which is more accurate than the residual error estimator and is reliable on all meshes unlike the ZZ estimators.

By first recovering a gradient in the conforming C^0 linear vector finite element space from the numerical gradient, the Zienkiewicz–Zhu (ZZ) estimator [9] is defined as the L^2 norm of the difference between the recovered and the numerical gradients. Due to its simplicity, universality, and asymptotic exactness for smooth problems, the ZZ estimator enjoys a high popularity in the engineering community (see, e.g., [6,10–12]).

Despite its popularity, it is also well known that estimators of the ZZ type have several major drawbacks. First, adaptive mesh refinement (AMR) algorithms using the ZZ estimator are not efficient to reduce global error for non-smooth problems, e.g., interface problems (see, e.g., [13]). By exploring the mathematical structure of the underlying problem and the characteristics of finite element approximations, [14] identified the reason for this failure, and [15] introduced an improved ZZ estimator for the lowest order conforming elements which is explicit and efficient for non-smooth problems. Second, estimators of the ZZ type are not reliable on coarse meshes relative to the underlying problem. A simple one-dimensional example in [4] shows that the ZZ estimator equals zero but the true error is arbitrarily large. For a two-dimensional example, see Section 5.4. Moreover, estimators of the ZZ type work only for elements of the lowest order and research for higher order elements is still in its infancy (see, e.g., [16,17]). Bank, Xu, and Zheng in [16] recovered higher order derivatives, and Naga and Zhang in [17] approximated the numerical solution by a higher order polynomial. Both the approaches demonstrate some appealing features like super-convergence and asymptotic exactness under certain smoothness assumptions of the exact solution.

Comparing with the residual estimator, we realized that the scaled element residual is no longer higher order when the recovered flux is the L^2 projection of the numerical flux in an $H(\mathrm{div})$-conforming space. By simply adding an appropriately weighted element residual to the improved ZZ estimator, it was shown in [18] that the resulting estimator for higher order elements is reliable on all meshes and more accurate than the residual estimator. Computing the L^2 projection of the numerical flux in an $H(\mathrm{div})$-conforming space requires solving a global problem and, hence, the estimator in [18] is more expensive than the residual estimator.

The purpose of this paper is to introduce an explicit flux recovery in an $H(\mathrm{div})$-conforming space so that the resulting hybrid error estimator is more accurate than the residual estimator with similar computational cost and applicability. To do so, we first specify the desired normal component of the recovered flux on each face as a weighted average of the normal components of the numerical fluxes. Then the recovered flux is chosen to satisfy a compatible divergence equation on each element. In particular, we are able to derive an explicit formula for a recovered flux in an $H(\mathrm{div})$-conforming space and the formula is automatically valid for higher order finite element approximations. Unlike existing ZZ-type estimators, which are not reliable on coarse meshes, we incorporate the divergence error in the estimator and the resulting error estimator of hybrid type is proved to be reliable on all meshes.

This hybrid estimator displays a strong connection to the explicit residual estimator as we can prove that the proposed estimator is actually equivalent to the residual estimator [5] with constants independent of the diffusion coefficients (see Section 4.2). As a result, the robustness of the residual estimator with respect to coefficient jumps carries over to the hybrid estimator. Despite the theoretical equivalence, numerical results show that the hybrid estimator is more accurate than the residual estimator. Hence the hybrid estimator can be viewed as a substitute of the residual estimator with an improved accuracy. The innate link to the residual estimator lends comparable generality to the hybrid estimator and future work includes applying the technique to convection–diffusion problems.

The rest of the paper is organized as follows. In Section 2, we introduce the model problem with a conforming finite element discretization and some notation. In Section 3, we present the explicit flux recovery. In Section 4, after defining the element indicator and the resulting global error estimator, we prove the robust local efficiency and global reliability. The equivalence between the proposed local indicator and the standard residual-based indicator is established in Section 4.2. Numerical results are presented in Section 5 to demonstrate the performance of the proposed estimator and a counter example is included in the end to illustrate the unreliability of ZZ-type estimators on a coarse mesh.

2. Problem and finite element approximation

Let Ω be a bounded polygonal domain in $\mathbb{R}^d$ $(d = 2,\ 3)$ with Lipschitz boundary $\partial\Omega$, where $\partial\Omega = \bar{\Gamma}_D \cup \bar{\Gamma}_N$ and $\Gamma_D \cap \Gamma_N = \emptyset$. For simplicity, assume that $\mathrm{meas}(\Gamma_D) > 0$. Consider the diffusion equation

$$-\operatorname{div}(A\nabla u) = f \quad \text{in } \Omega \tag{2.1}$$

with boundary conditions

$$u|_{\Gamma_D} = 0 \text{ and } -A\nabla u \cdot \boldsymbol{n}|_{\Gamma_N} = g_N. \tag{2.2}$$

Here, $f \in L^2(\Omega)$, $g_N \in L^2(\Gamma_N)$, $\boldsymbol{n}$ denotes the unit outward vector normal to Γ_N; A is piecewise constant in Ω and for almost all $x \in \Omega$, $A(x)$ is a symmetric positive definite matrix. If A is a scalar multiple of the identity matrix in a region, we would simply say A is a scalar in that region. Let $\bigcup_{i=1}^{L} \Omega_i$ be a disjoint partition of the domain Ω such that A is constant on each subdomain Ω_i.

The corresponding weak formulation for the problem in (2.1)–(2.2) is to find $u \in H^1_D(\Omega) := \{v \in H^1(\Omega) : v|_{\Gamma_D} = 0\}$ such that

$$a(u, v) := \int_\Omega A\nabla u \cdot \nabla v dx = \int_\Omega f v dx - \int_{\Gamma_N} g_N v ds, \quad \forall\, v \in H^1_D(\Omega). \tag{2.3}$$

The well-posedness of the weak formulation in (2.3) follows from the Riesz Representation Theorem.

Let $\mathcal{T}$ be a regular triangulation of Ω (see, e.g., [19]). For simplicity, assume that $K \in \mathcal{T}$ is either a triangle ($d = 2$) or a tetrahedron ($d = 3$). Define the following sets associated with the triangulation $\mathcal{T}$:

$\mathcal{N}$: the set of all vertices,

$\mathcal{E}$: the set of all edges ($d = 2$)/faces ($d = 3$),

$\mathcal{E}_I$: the set of all interior edges ($d = 2$)/faces ($d = 3$),

$\mathcal{E}_D$: the set of edges ($d = 2$)/faces ($d = 3$) on Γ_D,

$\mathcal{E}_N$: the set of edges ($d = 2$)/faces ($d = 3$) on Γ_N,

$\mathcal{E}_K$: the set of edges ($d = 2$)/faces ($d = 3$) in an element $K \in \mathcal{T}$.

Let h_K and h_e denote the diameters of $K \in \mathcal{T}$ and $e \in \mathcal{E}$, respectively. Similar to [5], assume that for each $K \in \mathcal{T}$, $A|_K$ is a symmetric, positive definite constant matrix that is nearly scalar, i.e., there exists a moderate constant $\kappa > 0$ such that

$$\rho_{\max}(A|_K) \le \kappa \rho_{\min}(A|_K), \quad \forall\, K \in \mathcal{T}, \tag{2.4}$$

where $\rho_{\max}$ and $\rho_{\min}$ denote maximal and minimal eigenvalues of a matrix, respectively. Obviously, if A is piecewise scalar with respect to $\mathcal{T}$, then (2.4) holds with $\kappa \equiv 1$. Let $(\cdot, \cdot)_S$ and $\|\cdot\|_S$ denote the L^2 inner product and norm on set S, respectively, and the subscript S is omitted when $S = \Omega$. Set

$$\alpha_{\max} := \max_{K \in \mathcal{T}} \rho_{\max}(A|_K), \quad \alpha_{\min} := \min_{K \in \mathcal{T}} \rho_{\min}(A|_K),$$

and define the piecewise constant function $\alpha \in L^2(\Omega)$ such that

$$\alpha = \alpha_K := \rho_{\max}(A|_K) \quad \text{in } K \in \mathcal{T}. \tag{2.5}$$

Furthermore, as in [5], for each $e \in \mathcal{E}$, define

$$\alpha_e := \max_{K \subseteq \omega_e} \alpha_K,$$

where ω_e denotes the union of elements adjacent to e.

Let $P_k(K)$ and $P_k(e)$ denote the sets of polynomials of degree less than or equal to $k \ge 0$ on $K \in \mathcal{T}$ and $e \in \mathcal{E}$, respectively. The conforming finite element space of order k is defined by

$$V_{\mathcal{T}} = \{v \in H^1_D(\Omega) : v|_K \in P_k(K), \quad \forall\, K \in \mathcal{T}\}.$$

The finite element solution $u_{\mathcal{T}} \in V_{\mathcal{T}}$ satisfies

$$a(u_{\mathcal{T}}, v) = \int_\Omega f v dx - \int_{\Gamma_N} g_N v ds, \quad \forall\, v \in V_{\mathcal{T}}. \tag{2.6}$$

Similarly, the well-posedness of (2.6) follows from the Riesz Representation Theorem.

Let $\bar{f}$ be the L^2 projection of f onto the piecewise polynomial space of degree $k-1$ with respect to $\mathcal{T}$ and denote by $\bar{g}_N$ the L^2 projection of g_N onto the piecewise polynomial space of degree $k-1$ with respect to $\mathcal{E}_N$. That is,

$$\bar{f}|_K := \Pi_K^{k-1} f, \quad \forall\, K \in \mathcal{T} \quad \text{and} \quad \bar{g}_N|_e := \Pi_e^{k-1} g_N, \quad \forall\, e \in \mathcal{E}_N,$$

where Π_K^{k-1} and Π_e^{k-1} denote the L^2 projection from $L^2(K)$ to $P_{k-1}(K)$ and from $L^2(e)$ to $P_{k-1}(e)$, respectively. The so-called *data oscillations* in f and g_N are defined by

$$\operatorname{osc}(f, K) := \frac{h_K}{\alpha_K^{1/2}}\|f - \bar{f}\|_K, \quad \operatorname{osc}(f, \mathcal{T}) := \left(\sum_{K\in\mathcal{T}} \operatorname{osc}(f, K)^2\right)^{1/2},$$

and

$$\operatorname{osc}(g_N, e) := \frac{h_e^{1/2}}{\alpha_e^{1/2}}\|g_N - \bar{g}_N\|_e, \quad \operatorname{osc}(g_N, \mathcal{E}_N) := \left(\sum_{e\in\mathcal{E}_N} \operatorname{osc}(g_N, e)^2\right)^{1/2},$$

respectively.

3. Element-based flux recovery

Denote the true and the numerical fluxes by

$$\boldsymbol{\sigma} = -A\nabla u \ \text{ and } \ \boldsymbol{\sigma}_{\mathcal{T}} = -A\nabla u_{\mathcal{T}},$$

respectively. It is well-known that $\boldsymbol{\sigma} \in H(\operatorname{div}; \Omega) = \{\boldsymbol{\tau} \in L^2(\Omega)^d : \operatorname{div}\boldsymbol{\tau} \in L^2(\Omega)\}$ and that $\boldsymbol{\sigma}_{\mathcal{T}} \notin H(\operatorname{div}; \Omega)$ in general. Hence as considered in [14], it is reasonable to find a flux $\hat{\boldsymbol{\sigma}} \in H(\operatorname{div}; \Omega)$ such that $\hat{\boldsymbol{\sigma}}$ is the projection of the numerical flux $\boldsymbol{\sigma}_{\mathcal{T}}$ with respect to the weighted L^2 inner product $(A^{-1}\cdot, \cdot)$. Equivalently, with

$$H_{N,g}(\operatorname{div}; \Omega) := \left\{\boldsymbol{\tau} \in H(\operatorname{div}; \Omega) : \boldsymbol{\tau}\cdot\boldsymbol{n} = g_N \text{ on } \Gamma_N\right\},$$

the projection $\hat{\boldsymbol{\sigma}} \in H_{N,g}(\operatorname{div}; \Omega)$ satisfies

$$\|A^{-1/2}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_{\mathcal{T}})\| = \min_{\boldsymbol{\tau}\in H_{N,g}(\operatorname{div};\Omega)} \|A^{-1/2}(\boldsymbol{\tau} - \boldsymbol{\sigma}_{\mathcal{T}})\|. \tag{3.1}$$

Since $H_{N,g}(\operatorname{div}; \Omega)$ is a nonempty closed convex subset of the Hilbert space $H(\operatorname{div}; \Omega)$, there exists a unique solution $\hat{\boldsymbol{\sigma}} \in H_{N,g}(\operatorname{div}; \Omega)$ for (3.1).

Let $\mathrm{RT}_{k-1}(K)$ and $\mathrm{BDM}_k(K)$ be the Raviart–Thomas space of index $k-1$ ($k \geq 1$) and the Brezzi–Douglas–Marini space of index k on element $K \in \mathcal{T}$, respectively. Namely,

$$\mathrm{RT}_{k-1}(K) := P_{k-1}(K)^d + \boldsymbol{x}P_{k-1}(K) \quad \text{and} \quad \mathrm{BDM}_k(K) := P_k(K)^d.$$

Let $\mathcal{W}(K) := \mathrm{RT}_{k-1}(K)$ or $\mathrm{BDM}_k(K)$ and define

$$\mathcal{W}_g := \{\boldsymbol{\tau} \in H(\operatorname{div}; \Omega) : \boldsymbol{\tau}|_K \in \mathcal{W}(K),\ \forall\, K \in \mathcal{T} \text{ and } \boldsymbol{\tau}\cdot\boldsymbol{n}|_e = \bar{g}_N,\ \forall\, e \in \mathcal{E}_N\}.$$

Corresponding to the finite element space $V_{\mathcal{T}}$ of order k ($k = 1, 2, \dots$), the approximation of the desired flux will be based on $\mathcal{W}_g$.

The finite element approximation of problem (3.1) is to find $\hat{\boldsymbol{\sigma}} \in \mathcal{W}_g$ such that

$$\|A^{-1/2}(\hat{\boldsymbol{\sigma}} - \boldsymbol{\sigma}_{\mathcal{T}})\| = \min_{\boldsymbol{\tau}\in\mathcal{W}_g} \|A^{-1/2}(\boldsymbol{\tau} - \boldsymbol{\sigma}_{\mathcal{T}})\|. \tag{3.2}$$

The global minimization in (3.2) was employed in [18] to perform flux recovery for higher order finite element approximations. However, a global recovery procedure is computationally expensive, while a local procedure, especially an element-based one, is more appealing for engineering practice.

We show in the rest of this section how to construct a flux in $\mathcal{W}_g$ that solves a boundary value problem imposed in each element $K \in \mathcal{T}$. A desired flux can be obtained via either a minimization on K (Section 3.3.1) or an explicit formula with degrees of freedom properly assigned (Section 3.3.2).

For each $e \in \mathcal{E}$, associate a unit normal vector $\boldsymbol{n}_e$, where $\boldsymbol{n}_e$ is chosen as the unit outward normal if $e \in \mathcal{E}_N$. For an interior edge ($d = 2$) /face ($d = 3$) $e \in \mathcal{E}_I$, denote by K_e^+ and K_e^- the two elements with common edge/face e such that the unit outward normal of K_e^+ on e coincides with $\boldsymbol{n}_e$. Moreover, $|K|$ and $|e|$ denote the area ($d = 2$) / volume ($d = 3$) of an element K and length of an edge e ($d = 2$) / area of a face e ($d = 3$), respectively.

For each element $K \in \mathcal{T}$ and $e \in \mathcal{E}_K$, in order to determine if the unit normal $\boldsymbol{n}_e$ is an *outward* normal with respect to K, we define below a sign function in $L^2(\partial K)$:

$$s_K : \partial K \to \{-1, 1\}, \quad s_K|_e = \begin{cases} 1, & \text{if } \boldsymbol{n}_e \text{ is an outward normal on } \partial K, \\ -1, & \text{if } \boldsymbol{n}_e \text{ is an inward normal on } \partial K. \end{cases}$$

That is to say, if $\boldsymbol{n}$ is the unit outward normal on $e \in \mathcal{E}_K$ in terms of K, then we have the expression $\boldsymbol{n} = s_K|_e \boldsymbol{n}_e$.

The flux recovery procedure is element-based. To be more specific, we impose a simple boundary value problem for the desired flux $\hat{\boldsymbol{\sigma}}$ in each element $K \in \mathcal{T}$ and then choose a solution that satisfies certain stability estimate as our desired flux.

We are looking for a flux $\hat{\boldsymbol{\sigma}} \in \mathcal{W}_g$ that satisfies in each element $K \in \mathcal{T}$:

$$\begin{cases} \operatorname{div} \hat{\boldsymbol{\sigma}} = \hat{f}_K & \text{in } K, \\ \hat{\boldsymbol{\sigma}} \cdot \boldsymbol{n}_e = \hat{g}_e & \text{on } e \in \mathcal{E}_K, \end{cases} \tag{3.3}$$

where $\hat{g}_e \in P_{k-1}(e)$ and $\hat{f}_K \in P_{k-1}(K)$ will be given in Section 3.1 and Section 3.2, respectively. The choice of a particular recovered flux among all solutions to (3.3) is presented in Section 3.3.

3.1. Choice of $\hat{g}_e$

The normal component of the recovered flux on each $e \in \mathcal{E}$ is defined as a weighted average of the normal components of the numerical fluxes, i.e.,

$$\hat{g}_e := \begin{cases} (1-\lambda_e)\boldsymbol{\sigma}_{\mathcal{T}}|_{K_e^+} \cdot \boldsymbol{n}_e + \lambda_e \boldsymbol{\sigma}_{\mathcal{T}}|_{K_e^-} \cdot \boldsymbol{n}_e, & \text{if } e \in \mathcal{E}_I, \\ \bar{g}_N, & \text{if } e \in \mathcal{E}_N, \\ \boldsymbol{\sigma}_{\mathcal{T}} \cdot \boldsymbol{n}_e, & \text{if } e \in \mathcal{E}_D, \end{cases} \tag{3.4}$$

where, as in [15], the weight λ_e is defined by

$$\lambda_e := \frac{\gamma_e^-}{\gamma_e^+ + \gamma_e^-}, \quad \forall e \in \mathcal{E}_I, \quad \text{with} \quad \gamma_e^{\pm} := (A^{-1}\boldsymbol{\phi}_e, \boldsymbol{\phi}_e)_{K_e^{\pm}}. \tag{3.5}$$

Here, $\boldsymbol{\phi}_e$ is the nodal basis function in RT_0 associated with e, i.e.,

$$\int_{e'} \boldsymbol{\phi}_e \cdot \boldsymbol{n}_{e'} ds = \delta_{ee'}, \quad \forall e' \in \mathcal{E}.$$

Apart from the weight defined in (3.5), a similar but easier-to-compute weight is given in Remark 3.2.

3.2. Choice of $\hat{f}_K$

Having determined the boundary data $\hat{g}_e$ in boundary value problem (3.3), we see that the right-hand side $\hat{f}_K$ must satisfy the compatibility condition:

$$\int_K \hat{f}_K dx = \int_{\partial K} \hat{\boldsymbol{\sigma}} \cdot \boldsymbol{n} ds = \sum_{e \in \mathcal{E}_K} \int_e s_K \hat{g}_e ds, \quad \forall K \in \mathcal{T}. \tag{3.6}$$

We choose $\hat{f}_K$ to be the L^2-projection of $\bar{f}$ onto the following admissible set

$$\mathcal{F}_K := \left\{ p \in P_{k-1}(K) : \int_K p dx = \sum_{e \in \mathcal{E}_K} \int_e s_K \hat{g}_e ds \right\}.$$

That is, $\hat{f}_K$ is the solution of

$$\|\hat{f}_K - \bar{f}\|_K = \min_{p \in \mathcal{F}_K} \|p - \bar{f}\|_K. \tag{3.7}$$

We next show that the solution of (3.7) is given by

$$\hat{f}_K := \bar{f} + J_K, \quad \text{with } J_K := |K|^{-1} \left(\sum_{e \in \mathcal{E}_K} \int_e s_K \hat{g}_e ds - \int_K \bar{f} dx \right). \tag{3.8}$$

Proposition 3.1. *With J_K defined in (3.8), we have $\bar{f} + J_K \in \mathcal{F}_K$ and*

$$\min_{p\in\mathcal{F}_K} \|p - \bar{f}\|_K = \|J_K\|_K.$$

Proof. For any $p \in \mathcal{F}_K$, it follows immediately from (3.8) and the Cauchy–Schwarz inequality that

$$|K||J_K| = \left| \sum_{e\in\mathcal{E}_K} \int_e s_K \hat{g}_e ds - \int_K \bar{f} dx \right| = \left| \int_K p - \bar{f} dx \right| \le |K|^{1/2} \|p - \bar{f}\|_K,$$

which is equivalent to

$$\min_{p\in\mathcal{F}_K} \|p - \bar{f}\|_K \ge \|J_K\|_K.$$

A simple calculation shows that $\bar{f} + J_K \in \mathcal{F}_K$. Hence the lower bound $\|J_K\|_K$ can be achieved at $\bar{f} + J_K$ and the minimizer is unique as $\mathcal{F}_K$ is a closed convex subset of $P_{k-1}(K)$. □

3.3. Choice of recovered flux

3.3.1. Choosing $\hat{\boldsymbol{\sigma}}_{\mathcal{T}}$ as the solution of a local minimization problem

With $\hat{g}_e$ and $\hat{f}_K$ defined in (3.4) and (3.8), respectively, for each element $K \in \mathcal{T}$, we define the set of discrete solutions to (3.3) by

$$\mathcal{S}(K) := \{\boldsymbol{\tau} \in \mathcal{W}(K) : \operatorname{div} \boldsymbol{\tau} = \hat{f}_K \text{ in } K \text{ and } \boldsymbol{\tau}\cdot\boldsymbol{n}_e = \hat{g}_e \text{ on } e \in \mathcal{E}_K\}. \tag{3.9}$$

Since the solution to the local problem in (3.3) may not be unique, we can choose a solution $\hat{\boldsymbol{\sigma}}_{\mathcal{T}} \in \mathcal{W}_g$ such that it also solves the local minimization problem in each element $K \in \mathcal{T}$:

$$\|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{\mathcal{T}} - \boldsymbol{\sigma}_{\mathcal{T}})\|_K = \min_{\boldsymbol{\tau}\in\mathcal{S}(K)} \|A^{-1/2}(\boldsymbol{\tau} - \boldsymbol{\sigma}_{\mathcal{T}})\|_K, \quad \forall\, K \in \mathcal{T}. \tag{3.10}$$

3.3.2. Explicit construction of $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}$

Instead of looking for a local minimizer as in (3.10), below we present an explicit construction of a flux $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}$ that solves (3.3) by assigning degrees of freedom in $\mathcal{W}(K) = \mathrm{RT}_{k-1}(K)$ or $\mathrm{BDM}_k(K)$ for each $K \in \mathcal{T}$. Though $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}$ may not necessarily minimize the quantity in (3.10), it can be seen later in Section 4.1 that the quantity $\|A^{-1/2}(\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} - \boldsymbol{\sigma}_{\mathcal{T}})\|_K$ is also small enough to guarantee the local efficiency.

To this end, we introduce the following spaces :

$$H_k(K) := \{\boldsymbol{q} \in P_k(K)^d : \operatorname{div} \boldsymbol{q} = 0 \text{ in } K \text{ and } \boldsymbol{q}\cdot\boldsymbol{n}|_{\partial K} = 0\} \quad (k \ge 1) \tag{3.11}$$

and

$$Q_{k-2}(K) := \{\boldsymbol{q} \in P_{k-2}(K)^d : (\boldsymbol{q}, \nabla p)_K = 0,\ \forall\, p \in P_{k-1}(K)\} \quad (k \ge 2). \tag{3.12}$$

The space $H_k(K)$ is defined in [20] to fix the degrees of freedom in $\mathrm{BDM}_k(K)$, and the space $Q_{k-2}(K)$ gives the following orthogonal decomposition of $P_{k-2}(K)^d$ with respect to the L^2 inner product:

$$P_{k-2}(K)^d = \{\nabla p : p \in P_{k-1}(K)\} \oplus Q_{k-2}(K). \tag{3.13}$$

The recovered flux $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} \in \mathcal{W}_g$ is defined as follows in each $K \in \mathcal{T}$ $(k \ge 1)$:

$$\begin{cases} \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}\cdot\boldsymbol{n}_e = \hat{g}_e, \quad \forall\, e \in \mathcal{E}_K, \\ \displaystyle\int_K \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} \cdot \nabla p dx = \sum_{e\in\mathcal{E}_K} (s_K \hat{g}_e, p)_e - (\hat{f}_K, p)_K, \quad \forall\, p \in P_{k-1}(K), \\ \displaystyle\int_K \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} \cdot \boldsymbol{q} dx = \int_K \boldsymbol{\sigma}_{\mathcal{T}} \cdot \boldsymbol{q} dx, \quad \forall\, \boldsymbol{q} \in \mathcal{Q}(K), \end{cases} \tag{3.14}$$

where

$$\mathcal{Q}(K) = \begin{cases} Q_{k-2}(K), & \text{if } \mathcal{W}(K) = \mathrm{RT}_{k-1}(K),\ k \geq 2, \\ H_k(K), & \text{if } \mathcal{W}(K) = \mathrm{BDM}_k(K). \end{cases} \tag{3.15}$$

This choice of recovered flux ensures that the quantity $\|\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} - \boldsymbol{\sigma}_{\mathcal{T}}\|_K$ is small enough to fulfill certain stability estimate explained in Section 4.1. Besides, it is easily seen that $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}$ does solve (3.3), i.e., $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}|_K \in \mathcal{S}(K)$.

Remark 3.1. For the lowest order $(k = 1)$ conforming finite element discretization of interface problem (2.1), the explicit recovery procedure in [14] can be viewed as a special case of the procedure in Section 3 when $\mathcal{W}(K) = \mathrm{RT}_0(K)$ except that the weight λ_e was chosen differently from the one in (3.5). In fact, the degrees of freedom in $\mathrm{RT}_0(K)$ imply that $\hat{\boldsymbol{\sigma}} \in \mathrm{RT}_0(K)$ is uniquely determined by its normal components on ∂K, hence expressing $\hat{\boldsymbol{\sigma}}$ in terms of basis functions in $\mathrm{RT}_0(K)$ [14] is equivalent to imposing normal components on ∂K as in (3.3) and (3.4). In this case, $\hat{f}_K$ in (3.3) is a constant uniquely determined by $\hat{g}_e$ according to (3.6).

Remark 3.2. In addition to (3.5), another choice of the weight λ_e in (3.4) can be:

$$\lambda_e := \frac{\alpha_{K_e^-}^{-1} h_{K_e^-}}{\alpha_{K_e^+}^{-1} h_{K_e^+} + \alpha_{K_e^-}^{-1} h_{K_e^-}}. \tag{3.16}$$

With this choice, it can be verified that the resulting $\hat{g}_e$ in (3.4) satisfies the following minimization property for any $e \in \mathcal{E}_K \cap \mathcal{E}_I$:

$$\hat{g}_e = \underset{\tilde{g} \in L^2(e)}{\operatorname{argmin}} \left\{ \alpha_{K_e^+}^{-1} h_{K_e^+} \|\tilde{g} - \boldsymbol{\sigma}_{\mathcal{T}}|_{K_e^+} \cdot \boldsymbol{n}_e\|_e^2 + \alpha_{K_e^-}^{-1} h_{K_e^-} \|\tilde{g} - \boldsymbol{\sigma}_{\mathcal{T}}|_{K_e^-} \cdot \boldsymbol{n}_e\|_e^2 \right\}.$$

Remark 3.3. In addition to (3.8), we can also choose

$$\hat{f}_K := \operatorname{div} \boldsymbol{\sigma}_{\mathcal{T}} + J_K,$$

where J_K is a constant such that (3.6) is satisfied. It can be seen later from the proof of Theorem 4.1 that this choice of $\hat{f}_K$ also guarantees the robust efficiency of the proposed local indicator in next section. However, as we shall see in Theorem 4.3, the quantity $\|\hat{f}_K - \bar{f}\|_K$ appears in the upper bound of the exact error, so it is more desirable to minimize this quantity as in (3.7) when choosing $\hat{f}_K$. Hence we believe that $\hat{f}_K$ defined in (3.8) is a better choice.

4. Error estimator

Let $\hat{\boldsymbol{\sigma}}_{\mathcal{T}}$ and $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}$ be the fluxes derived in (3.10) and (3.14), respectively. Define the local indicators corresponding to $\hat{\boldsymbol{\sigma}}_{\mathcal{T}}$ and $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}$ by

$$\xi_K := \left(h_K^2 \alpha_K^{-1} \|\bar{f} - \hat{f}_K\|_K^2 + \|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{\mathcal{T}} - \boldsymbol{\sigma}_{\mathcal{T}})\|_K^2 \right)^{1/2} \tag{4.1}$$

and

$$\tilde{\xi}_K := \left(h_K^2 \alpha_K^{-1} \|\bar{f} - \hat{f}_K\|_K^2 + \|A^{-1/2}(\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} - \boldsymbol{\sigma}_{\mathcal{T}})\|_K^2 \right)^{1/2}, \tag{4.2}$$

respectively. The corresponding global estimators are then defined by

$$\xi := \left(\sum_{K \in \mathcal{T}} \xi_K^2 \right)^{1/2} \quad \text{and} \quad \tilde{\xi} := \left(\sum_{K \in \mathcal{T}} \tilde{\xi}_K^2 \right)^{1/2}. \tag{4.3}$$

Due to the minimization property of $\hat{\boldsymbol{\sigma}}_{\mathcal{T}}$ in (3.10), it immediately follows that

$$\xi_K \leq \tilde{\xi}_K \quad \text{and} \quad \xi \leq \tilde{\xi}. \tag{4.4}$$

We prove the robust local efficiency of the local indicators and the global reliability of the global estimators in Section 4.1 and Section 4.3, respectively.

4.1. Efficiency

Let $\eta_{R,K}$ denote the residual-based error indicator in K (cf. [5–7]). Namely,

$$\eta_{R,K}^2 := \frac{h_K^2}{\alpha_K}\| \bar{f} - \operatorname{div} \boldsymbol{\sigma}_{\mathcal{T}} \|_K^2 + \frac{1}{2}\sum_{e\in\mathcal{E}_K\cap\mathcal{E}_I}\frac{h_K}{\alpha_e}\| j_e \|_e^2 + \sum_{e\in\mathcal{E}_K\cap\mathcal{E}_N}\frac{h_K}{\alpha_e}\| j_e \|_e^2, \tag{4.5}$$

where j_e is the jump of the normal components of the numerical fluxes defined by

$$j_e := \begin{cases} (\boldsymbol{\sigma}_{\mathcal{T}}|_{K_e^+} - \boldsymbol{\sigma}_{\mathcal{T}}|_{K_e^-})\cdot \boldsymbol{n}_e, & \text{if } e \in \mathcal{E}_I, \\ \boldsymbol{\sigma}_{\mathcal{T}}\cdot \boldsymbol{n}_e - \bar{g}_N, & \text{if } e \in \mathcal{E}_N, \\ 0, & \text{if } e \in \mathcal{E}_D. \end{cases}$$

We first bound $\tilde{\xi}_K$ from above by inter-element jumps and element residuals, then the local efficiency of $\tilde{\xi}_K$ follows immediately from that of $\eta_{R,K}$.

4.1.1. A lemma on a stability estimate

The following lemma plays a crucial role in the proof of the efficiency of $\tilde{\xi}_K$. We will use c (or C) with or without subscripts in this paper to denote a generic positive constant, possibly different at different occurrences, that is independent of $\alpha_{\max}/\alpha_{\min}$, but may depend on shape parameter of the mesh $\mathcal{T}$, the polynomial degree k, and κ in (2.4).

Lemma 4.1. *For $k \geq 1$, let $f \in P_{k-1}(K)$ and $g|_e \in P_{k-1}(e)$ for all $e \in \mathcal{E}_K$ satisfy the following compatibility condition*

$$\int_K f\,dx = \int_{\partial K} g\,ds.$$

With $\mathcal{W}(K) = \mathrm{RT}_{k-1}(K)$ or $\mathrm{BDM}_k(K)$, if the vector field $\boldsymbol{\tau} \in \mathcal{W}(K)$ is defined by

$$\begin{cases} \boldsymbol{\tau}\cdot\boldsymbol{n} = g, \quad \text{on } \partial K, \\ \displaystyle\int_K \boldsymbol{\tau}\cdot\nabla p\,dx = (g,p)_{\partial K} - (f,p)_K, \quad \forall\, p \in P_{k-1}(K), \\ \displaystyle\int_K \boldsymbol{\tau}\cdot\boldsymbol{q}\,dx = 0, \quad \forall\, \boldsymbol{q} \in \mathcal{Q}(K), \end{cases} \tag{4.6}$$

where $\mathcal{Q}(K)$ is defined in (3.15), then $\boldsymbol{\tau}$ satisfies the following divergence equation

$$\begin{cases} div\,\boldsymbol{\tau} = f & \text{in } K, \\ \boldsymbol{\tau}\cdot\boldsymbol{n} = g & \text{on } \partial K \end{cases} \tag{4.7}$$

and the following stability estimate

$$\|\boldsymbol{\tau}\|_K \leq c\left(h_K\|f\|_K + h_K^{1/2}\|g\|_{\partial K}\right). \tag{4.8}$$

To prove Lemma 4.1, we make use of a norm equivalence result of $\mathcal{W}(K)$ in [21]. To this end, for $k \geq 1$ and $\boldsymbol{\tau} \in \mathcal{W}(K)$, if $\mathcal{W}(K) = \mathrm{RT}_{k-1}(K)$, let

$$G(\boldsymbol{\tau}) = h_K^{1/2}\sum_{e\in\mathcal{E}_K}\sup_{\substack{p\in P_{k-1}(e)\\ \|p\|_e=1}}\left|\int_e(\boldsymbol{\tau}\cdot\boldsymbol{n})\,p\,ds\right| + \sup_{\substack{\boldsymbol{q}\in P_{k-2}(K)^d\\ \|\boldsymbol{q}\|_K=1}}\left|\int_K \boldsymbol{\tau}\cdot\boldsymbol{q}\,dx\right|,$$

where if $k = 1$, the second term in $G(\boldsymbol{\tau})$ above vanishes because $P_{-1}(K)$ is considered empty (the degrees of freedom for $\mathrm{RT}_0(K)$ are on the boundary ∂K). If $\mathcal{W}(K) = \mathrm{BDM}_k(K)$, let

$$G(\boldsymbol{\tau}) = h_K^{1/2}\sum_{e\in\mathcal{E}_K}\sup_{\substack{p\in P_k(e)\\ \|p\|_e=1}}\left|\int_e(\boldsymbol{\tau}\cdot\boldsymbol{n})p\,ds\right| + \sup_{\substack{p\in P_{k-1}(K)\\ \|\nabla p\|_K=1}}\left|\int_K \boldsymbol{\tau}\cdot\nabla p\,dx\right| + \sup_{\substack{\boldsymbol{q}\in H_k(K)\\ \|\boldsymbol{q}\|_K=1}}\left|\int_K \boldsymbol{\tau}\cdot\boldsymbol{q}\,dx\right|.$$

Then a standard scaling argument (cf. [21, Lemma 3.5]) implies that $G(\boldsymbol{\tau})$ is equivalent to the L^2-norm in $\mathcal{W}(K)$ with constants independent of h_K, i.e.,

$$\frac{1}{c}\|\boldsymbol{\tau}\|_K \leq G(\boldsymbol{\tau}) \leq c\|\boldsymbol{\tau}\|_K. \tag{4.9}$$

Proof of Lemma 4.1. Let $\boldsymbol{\tau} \in \mathcal{W}(K)$ be given in (4.6). Obviously, $\boldsymbol{\tau}$ satisfies the boundary condition in (4.7). To show the validity of the first equation in (4.7), it follows from integration by parts and the first two equations in (4.6) that for any $p \in P_{k-1}(K)$,

$$(\operatorname{div}\boldsymbol{\tau}, p)_K = (\boldsymbol{\tau}\cdot\boldsymbol{n}, p)_{\partial K} - (\boldsymbol{\tau}, \nabla p)_K = (f, p)_K.$$

Next we prove that $\boldsymbol{\tau}$ satisfies the stability estimate in (4.8) for the case $\mathcal{W}(K) = \mathrm{RT}_{k-1}(K)$. (For $\mathcal{W}(K) = \mathrm{BDM}_k(K)$, (4.8) can be established in a similar fashion.) According to (4.9), the degrees of freedom defined in (4.6), and the orthogonal decomposition of $P_{k-2}(K)^d$ in (3.13), we have

$$\begin{aligned}\|\boldsymbol{\tau}\|_K \le c\, G(\boldsymbol{\tau}) &\le c\left(h_K^{1/2}\sum_{e\in\mathcal{E}_K}\|g\|_e + \sup_{\substack{p\in P_{k-1}(K)\\ \|\nabla p\|_K=1}}\left|\int_K \boldsymbol{\tau}\cdot\nabla p\,dx\right|\right)\\ &\le c\left(h_K^{1/2}\|g\|_{\partial K} + \sup_{\substack{p\in P_{k-1}(K)\\ \|\nabla p\|_K=1\\ \int_K p dx=0}}\left|\int_{\partial K} gp\,ds - \int_K fp\,dx\right|\right).\end{aligned} \tag{4.10}$$

For $p \in P_{k-1}(K)$ with $\int_K p dx = 0$ and $\|\nabla p\|_K = 1$, it follows from the Poincaré and the trace inequalities (cf. [22]) that

$$\|p\|_{\partial K} \le c h_K^{1/2}\|\nabla p\|_K = c h_K^{1/2} \quad \text{and} \quad \|p\|_K \le \frac{1}{\pi}h_K\|\nabla p\|_K = \frac{1}{\pi}h_K,$$

which, together with the Cauchy–Schwarz inequality, implies

$$\left|\int_{\partial K} gp ds - \int_K fp dx\right| \le c\left(h_K^{1/2}\|g\|_{\partial K} + h_K\|f\|_K\right).$$

Now (4.8) is a direct consequence of (4.10). This completes the proof of the lemma. □

4.1.2. Proof of local efficiency

We first simplify the expression of J_K to reveal its relation to the jumps of normal components of $\boldsymbol{\sigma}_{\mathcal{T}}$ across $e \in \mathcal{E}_K$ as well as element residual, and then derive estimates for J_K.

By the divergence theorem, J_K in (3.8) can be expressed as

$$J_K = |K|^{-1}\left(\int_K (\operatorname{div}\boldsymbol{\sigma}_{\mathcal{T}} - \bar{f})dx + \sum_{e\in\mathcal{E}_K}\int_e s_K(\hat{g}_e - \boldsymbol{\sigma}_{\mathcal{T}}|_K\cdot\boldsymbol{n}_e)ds\right). \tag{4.11}$$

A simple calculation implies that

$$\hat{g}_e - \boldsymbol{\sigma}_{\mathcal{T}}|_K\cdot\boldsymbol{n}_e = \begin{cases}\dfrac{-\gamma_e^-}{\gamma_e^+ + \gamma_e^-}j_e, & \text{if } K = K_e^+,\\ \dfrac{\gamma_e^+}{\gamma_e^+ + \gamma_e^-}j_e, & \text{if } K = K_e^-\end{cases} \tag{4.12}$$

for $e \in \mathcal{E}_I$ and $\hat{g}_e - \boldsymbol{\sigma}_{\mathcal{T}}\cdot\boldsymbol{n}_e = -j_e$ for $e \in \mathcal{E}_N \cup \mathcal{E}_D$.

To this end, for $e \in \mathcal{E}_K$, define

$$\mu_{K,e} := \begin{cases}\dfrac{(A^{-1}\boldsymbol{\phi}_e, \boldsymbol{\phi}_e)_{K_e'}}{(A^{-1}\boldsymbol{\phi}_e, \boldsymbol{\phi}_e)_K + (A^{-1}\boldsymbol{\phi}_e, \boldsymbol{\phi}_e)_{K_e'}}, & \text{if } e \in \mathcal{E}_I,\\ 1, & \text{if } e \in \mathcal{E}_N \cup \mathcal{E}_D,\end{cases}$$

where K_e' denotes the element adjacent to K that shares e with K. Since $\|\boldsymbol{\phi}_e\|_K$ (or $\|\boldsymbol{\phi}_e\|_{K_e'}$) only depends on K (or K_e'), it can be computed that

$$\mu_{K,e} \le c\,\frac{\alpha_{K_e'}^{-1}}{\alpha_K^{-1} + \alpha_{K_e'}^{-1}} \le c\left(\frac{\alpha_K}{\alpha_e}\right)^{1/2}, \quad \forall\, e \in \mathcal{E}_K \cap \mathcal{E}_I, \tag{4.13}$$

where c may depend on κ, but is independent of α.

Thus it follows immediately from (4.12) and (4.13) that

$$\|\hat{g}_e - \boldsymbol{\sigma}_{\mathcal{T}}|_K \cdot \boldsymbol{n}_e\|_e \le \mu_{K,e}\|j_e\|_e \le c\left(\frac{\alpha_K}{\alpha_e}\right)^{1/2}\|j_e\|_e, \quad \forall e \in \mathcal{E}_K, \tag{4.14}$$

which, together with (4.11) and the triangle and the Cauchy–Schwarz inequalities, implies that

$$\|J_K\|_K \le \|\operatorname{div}\boldsymbol{\sigma}_{\mathcal{T}} - \bar{f}\|_K + ch_K^{-1/2}\sum_{e\in\mathcal{E}_K}\left(\frac{\alpha_K}{\alpha_e}\right)^{1/2}\|j_e\|_e. \tag{4.15}$$

Now we are able to prove the local efficiency of $\tilde{\xi}_K$ and of ξ_K.

Theorem 4.1. *For the local indicators ξ_K and $\tilde{\xi}_K$ defined in (4.1) and (4.2), respectively, there exists a constant c independent of $\alpha_{\max}/\alpha_{\min}$ such that*

$$\xi_K^2 \le \tilde{\xi}_K^2 \le c\left(\|A^{1/2}\nabla(u-u_{\mathcal{T}})\|_{\omega_K}^2 + \sum_{K'\subset\omega_K} osc(f,K')^2 + \sum_{e\in\mathcal{E}_K\cap\mathcal{E}_N} osc(g_N,e)^2\right),$$

where ω_K denotes the union of elements that share at least one edge ($d=2$) or one face ($d=3$) with K.

Proof. Since $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}$ solves (3.3), the error flux $\widetilde{\boldsymbol{\sigma}}^{\Delta} := \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} - \boldsymbol{\sigma}_{\mathcal{T}}$ satisfies (4.7) with

$$f = \bar{f} - \operatorname{div}\boldsymbol{\sigma}_{\mathcal{T}} + J_K \quad \text{and} \quad g = \hat{g}_e - \boldsymbol{\sigma}_{\mathcal{T}}|_K \cdot \boldsymbol{n}_e \text{ on } e \in \mathcal{E}_K.$$

Moreover, by the definition of $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}$ in (3.14), the error flux satisfies the third equation in (4.6). Hence, it follows from Lemma 4.1, the triangle inequality, (4.15), and (4.14) that

$$\begin{aligned}\|A^{-1/2}(\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} - \boldsymbol{\sigma}_{\mathcal{T}})\|_K &\le c\alpha_K^{-1/2}\left(h_K\|\bar{f} - \operatorname{div}\boldsymbol{\sigma}_{\mathcal{T}} + J_K\|_K + h_K^{1/2}\sum_{e\in\mathcal{E}_K}\|\hat{g}_e - \boldsymbol{\sigma}_{\mathcal{T}}|_K\cdot\boldsymbol{n}_e\|_e\right)\\ &\le c\left(h_K\alpha_K^{-1/2}\|\operatorname{div}\boldsymbol{\sigma}_{\mathcal{T}} - \bar{f}\|_K + \sum_{e\in\mathcal{E}_K}\frac{h_K^{1/2}}{\sqrt{\alpha_e}}\|j_e\|_e\right),\end{aligned} \tag{4.16}$$

where c may depend on κ, but is independent of α.

Analogously, by recalling in (3.8) that $\hat{f}_K - \bar{f} = J_K$ and applying (4.15), we have

$$h_K\alpha_K^{-1/2}\|\bar{f} - \hat{f}_K\|_K \le c\left(h_K\alpha_K^{-1/2}\|\operatorname{div}\boldsymbol{\sigma}_{\mathcal{T}} - \bar{f}\|_K + \sum_{e\in\mathcal{E}_K}\frac{h_K^{1/2}}{\sqrt{\alpha_e}}\|j_e\|_e\right). \tag{4.17}$$

It is easily seen from the definition of $\tilde{\xi}_K$ in (4.2), the property in (4.4) and the estimates in (4.16) and (4.17) that

$$\xi_K^2 \le \tilde{\xi}_K^2 \le c\,\eta_{R,K}^2, \tag{4.18}$$

with c independent of α. The local efficiency of ξ_K or $\tilde{\xi}_K$ now follows immediately from that of $\eta_{R,K}$ (cf. [5–7]). This completes the proof of the theorem. □

Remark 4.1. In terms of Lemma 4.1, different from the explicit construction in (4.6), the existence of a vector field $\boldsymbol{\tau} \in \mathcal{W}(K)$ that satisfies the divergence equation (4.7) and the stability estimate (4.8) was shown in [22, Lemma 3.1] via an indirect argument, which requires solving a Poisson equation with Neumann boundary conditions. Hence no explicit formula for such a vector field $\boldsymbol{\tau}$ was given in [22, Lemma 3.1].

4.2. Equivalence between residual estimator $\eta_{R,K}$ and hybrid estimators $\xi_K, \tilde{\xi}_K$

It is shown in (4.18) that the local error indicators, ξ_K and $\tilde{\xi}_K$, based on flux recovery can be bounded from above by the residual-based indicator $\eta_{R,K}$ with a constant independent of α. In this section, we prove the opposite direction, thus establishing the equivalence between $\xi_K, \tilde{\xi}_K$ and $\eta_{R,K}$ with constants independent of α.

Theorem 4.2. *Let $\eta_{R,K}$ be the residual indicator in (4.5) and ξ_K and $\tilde{\xi}_K$ be the hybrid indicators in (4.1) and (4.2), respectively. Then there exist positive constants C_1 and C_2 independent of α such that*

$$C_1\xi_K \le \eta_{R,K} \le C_2 \sum_{K'\subset\omega_K} \xi_{K'} \quad \text{and} \quad C_1\tilde{\xi}_K \le \eta_{R,K} \le C_2 \sum_{K'\subset\omega_K} \tilde{\xi}_{K'}, \tag{4.19}$$

where ω_K is same as in Theorem 4.1.

Proof. The lower bounds in (4.19) were established in (4.18). To prove the validity of the upper bounds in (4.19), we adopt the idea on the proof of local efficiency of $\eta_{R,K}$ in [5]. To this end, let ϕ_e denote the edge/face bubble function associated with edge ($d = 2$)/face ($d = 3$) e and ϕ_K the element bubble function associated with element K. For each element K contained in ω_e, by extending j_e to be a polynomial in K with stability estimate (cf. [6,23])

$$\|j_e\|_K \le Ch_e^{1/2}\|j_e\|_e, \quad K \subseteq \omega_e,$$

we deduce from integration by parts, the Cauchy–Schwarz and the inverse inequalities that

$$\begin{aligned}\|j_e\|_e^2 &\le C\int_e \phi_e j_e^2 ds = C\sum_{K\subseteq\omega_e}\left(\int_K(\boldsymbol{\sigma}_{\mathcal{T}}-\hat{\boldsymbol{\sigma}}_{\mathcal{T}})\cdot\nabla(\phi_e j_e)dx + \int_K \phi_e j_e \mathrm{div}(\boldsymbol{\sigma}_{\mathcal{T}}-\hat{\boldsymbol{\sigma}}_{\mathcal{T}})dx\right)\\ &\le C\sum_{K\subseteq\omega_e}\left(\alpha_K^{1/2}h_e^{-1/2}\|A^{-1/2}(\boldsymbol{\sigma}_{\mathcal{T}}-\hat{\boldsymbol{\sigma}}_{\mathcal{T}})\|_K + h_e^{1/2}\|\mathrm{div}\,\boldsymbol{\sigma}_{\mathcal{T}}-\hat{f}_K\|_K\right)\|j_e\|_e\end{aligned}$$

and that

$$\begin{aligned}\|\hat{f}_K - \mathrm{div}\,\boldsymbol{\sigma}_{\mathcal{T}}\|_K^2 &\le C\int_K \phi_K(\mathrm{div}\hat{\boldsymbol{\sigma}}_{\mathcal{T}} - \mathrm{div}\boldsymbol{\sigma}_{\mathcal{T}})^2 dx\\ &= -C\int_K \nabla(\phi_K(\mathrm{div}\hat{\boldsymbol{\sigma}}_{\mathcal{T}} - \mathrm{div}\boldsymbol{\sigma}_{\mathcal{T}}))\cdot(\hat{\boldsymbol{\sigma}}_{\mathcal{T}}-\boldsymbol{\sigma}_{\mathcal{T}})dx\\ &\le C\alpha_K^{1/2}h_K^{-1}\|\hat{f}_K - \mathrm{div}\,\boldsymbol{\sigma}_{\mathcal{T}}\|_K\|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{\mathcal{T}}-\boldsymbol{\sigma}_{\mathcal{T}})\|_K,\end{aligned}$$

where C may depend on κ, but is independent of α. Hence we conclude that

$$h_K\alpha_K^{-1/2}\|\hat{f}_K - \mathrm{div}\,\boldsymbol{\sigma}_{\mathcal{T}}\|_K \le C\|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{\mathcal{T}}-\boldsymbol{\sigma}_{\mathcal{T}})\|_K,$$

and

$$h_e^{1/2}\alpha_e^{-1/2}\|j_e\|_e \le C\|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{\mathcal{T}}-\boldsymbol{\sigma}_{\mathcal{T}})\|_{\omega_e}$$

with C independent α. Now the upper bounds in (4.19) are a direct consequence of the definition of $\eta_{R,K}$ and ξ_K in the respective (4.1) and (4.5), the triangle inequality, and (4.4). This completes the proof of the theorem. □

4.3. Reliability

Due to the equivalence result in Theorem 4.2, the robust global reliability of ξ or $\tilde{\xi}$ follows immediately from that of the residual-based estimator η_R [6] under the monotonicity assumption (Hypothesis 2.7 in [5]) of α:

> for any two different subdomains $\overline{\Omega}_i$ and $\overline{\Omega}_j$ that share at least one point, there exists a connected path passing from $\overline{\Omega}_i$ to $\overline{\Omega}_j$ through adjacent subdomains such that α is monotone along this path.

In fact, the assumption can be further weakened to the quasi-monotonicity condition [7].

Theorem 4.3. *Assume that α defined in (2.5) satisfies the monotonicity assumption. Then the estimators ξ and $\tilde{\xi}$ defined in (4.3) satisfy the following reliability bound:*

$$\begin{aligned}\|A^{1/2}\nabla(u-u_{\mathcal{T}})\| &\le c\left(\xi + osc(f,\mathcal{T}) + osc(g_N,\mathcal{E}_N)\right)\\ &\le c\left(\tilde{\xi} + osc(f,\mathcal{T}) + osc(g_N,\mathcal{E}_N)\right)\end{aligned}$$

with c independent of $\alpha_{\max}/\alpha_{\min}$.

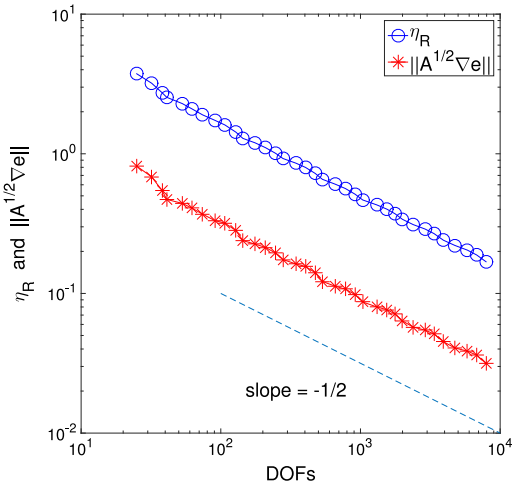

Fig. 1. Example 1: P_1 - error - η_R.

5. Numerical experiments

We consider adaptive finite element for solving three examples in Section For all examples, $\Omega = (-1, 1)^2$ and the initial mesh consists of 4×4 congruent squares, each of which is partitioned into two triangles connecting bottom-left and top-right corners. Dörfler's marking strategy [24] is used with $\theta_D = 0.5$ as in [24,25]. Namely, in the refinement of $\mathcal{T}$, a minimal subset $\hat{\mathcal{T}}$ of $\mathcal{T}$ is constructed such that

$$\left(\sum_{K\in\hat{\mathcal{T}}}\eta_K^2\right)^{1/2} \geq \theta_D\left(\sum_{K\in\mathcal{T}}\eta_K^2\right)^{1/2}, \tag{5.1}$$

where η_K denotes an error indicator on element K. The newest-vertex bisection [26] is used in the refinement. The following notation will be used:

- exact error $e := u - u_{\mathcal{T}}$;
- effectivity index is denoted by: eff-ind;
- degrees of freedom: DOFs;
- stopping criterion: $\|A^{1/2}\nabla e\| \leq \epsilon_{\text{rel}}\|A^{1/2}\nabla u\|$ where ϵ_{rel} is a prescribed tolerance.

Examples in Sections 5.1 and 5.2 are designed to illustrate the well-known fact that residual estimator usually overestimates the true error by a large margin [8] and also to show the improved accuracy by using the proposed hybrid estimator. In Section 5.3 we solve the numerical benchmark known as Kellogg's example [25,27] to justify the robustness and generality of the proposed estimator.

5.1. Example 1

To illustrate the effectivity of residual estimator, we first consider the Poisson equation with Dirichlet boundary condition and the data is chosen such that the exact solution is a quadratic polynomial

$$u(x, y) = -x^2 - y^2.$$

The tolerance in the stopping criterion is chosen as $\epsilon_{\text{rel}} = 0.01$. With P_1 finite element approximation, the numerical results are shown in Table 1 and Figs. 1–2. It can be seen from Table 1 that, for such a smooth problem, the residual estimator η_R is much less accurate compared to the proposed hybrid estimator ξ.

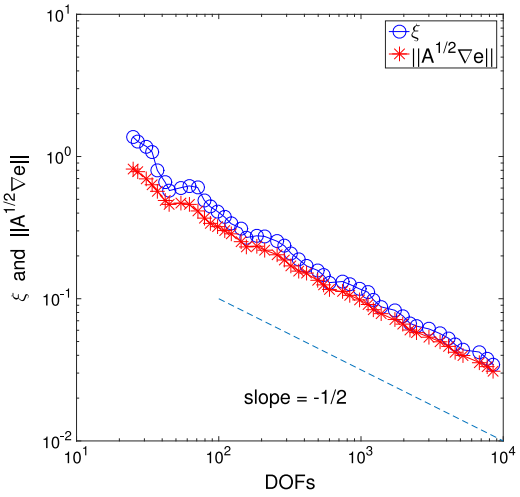

Fig. 2. Example 1: P_1 - error - ξ.

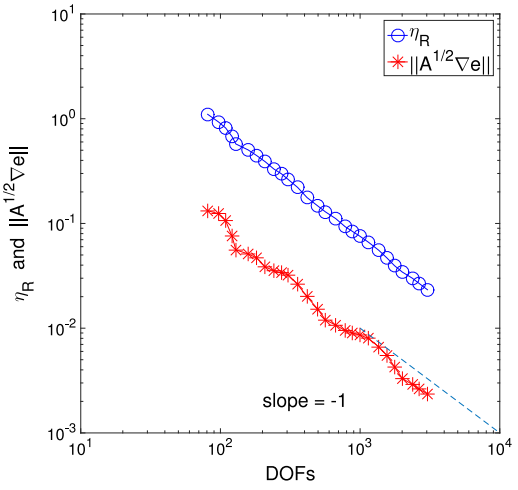

Fig. 3. Example 2: P_2 - error - η_R.

Table 1
Example 1 — quadratic solution and P_1 discretization ($\epsilon_{\text{rel}} = 0.01$).

Estimator	DOFs	$\|A^{1/2}\nabla e\|/\|A^{1/2}\nabla u\|$	eff-ind
η_R	7926	9.6E−3	5.35
ξ	8520	9.4E−3	1.11

5.2. *Example 2*

We consider again the Poisson equation as in Section 5.1, where the exact solution is now a quartic polynomial

$$u(x, y) = -(x^2 - 1)(y^2 - 1).$$

The tolerance is chosen as $\epsilon_{\text{rel}} = 0.001$ and P_2 discretization is used. The numerical results are presented in Table 2 and Figs. 3–4, from which we see that the residual estimator η_R overestimates the true error by a factor of 9.87, while the proposed estimator ξ remains accurate.

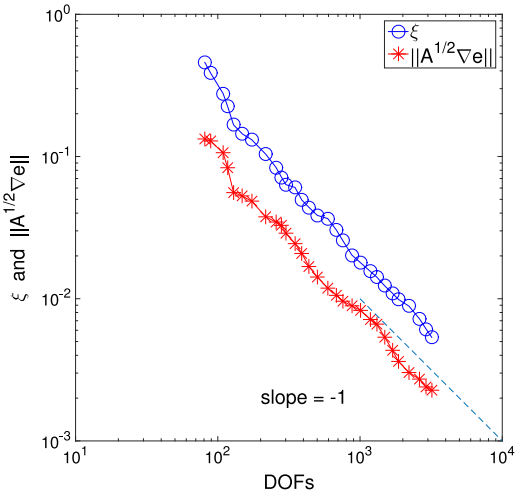

Fig. 4. Example 2: P_2 - erro - ξ.

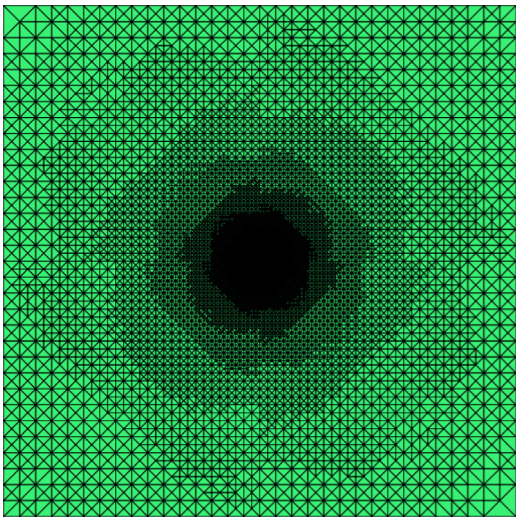

Fig. 5. Example 3: P_1 - mesh - η_R.

Table 2
Example 2 — quartic solution and P_2 discretization ($\epsilon_{\text{rel}} = 0.001$).

Estimator	DOFs	$\|A^{1/2}\nabla e\|/\|A^{1/2}\nabla u\|$	eff-ind
η_R	3037	9.9E−4	9.87
ξ	3193	9.5E−4	2.36

5.3. *Example 3—Kellogg's example*

We consider solving the Kellogg's example [27]. The parameters are same as in [25] and are collected here for completeness. The domain is $\Omega = (-1, 1)^2$ and the diffusion coefficient is chosen as $A = \alpha_1 I$ in the first and third quadrants, and $A = \alpha_2 I$ in the second and fourth quadrants, where

$$\alpha_1 \approx 161.4476387975881 \quad \text{and} \quad \alpha_2 = 1.$$

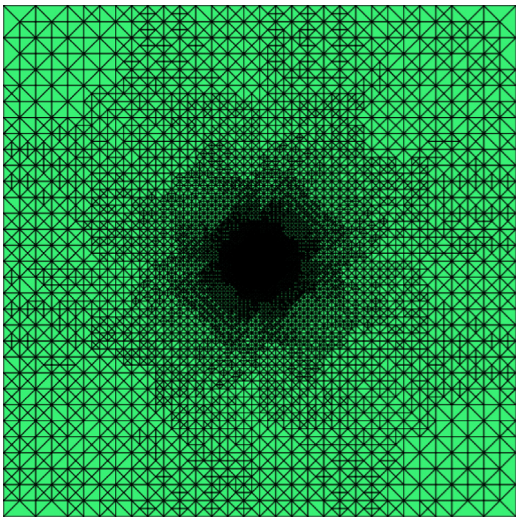

Fig. 6. Example 3: P_1 - mesh - ξ.

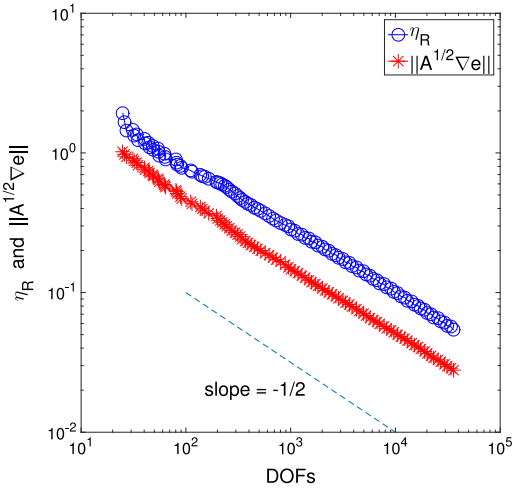

Fig. 7. Example 3: P_1 - error - η_R.

For $f = 0$, an exact solution in polar coordinate is given by $u(r, \theta) = r^\beta \psi(\theta)$ with

$$\psi(\theta) := \begin{cases} \cos((\pi/2 - \tau)\beta)\cos((\theta - \pi/4)\beta), & \text{if } 0 \le \theta \le \pi/2, \\ \cos(\pi\beta/4)\cos((\theta - \pi + \tau)\beta), & \text{if } \pi/2 \le \theta \le \pi, \\ \cos(\tau\beta)\cos((\theta - 5\pi/4)\beta), & \text{if } \pi \le \theta \le 3\pi/2, \\ \cos((\pi/2 - \tau)\beta)\cos((\theta - 3\pi/2 - \tau)\beta), & \text{if } 0 \le \theta \le \pi/2, \end{cases}$$

$$\beta = 0.1 \quad \text{and} \quad \tau \approx 14.92256510455152.$$

The regularity of u is quite low as $u \notin H^{1.1}(\Omega)$.

We perform numerical tests by using both P_1 and P_2 conforming elements. For $P_k (k = 1, 2)$ element and RT_{k-1} flux recovery, the recovered fluxes in Section 3.3 are identical, i.e., $\hat{\boldsymbol{\sigma}}_{\mathcal{T}} = \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}$, consequently, $\xi_K = \tilde{\xi}_K$. The relative error tolerance is chosen as $\epsilon_{\text{rel}} = 0.05$.

The numerical results are collected in Table 3 and Figs. 5–12.

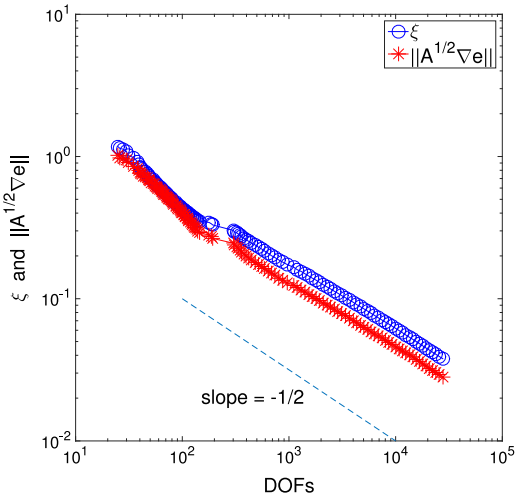

Fig. 8. Example 3: P_1 - error - ξ.

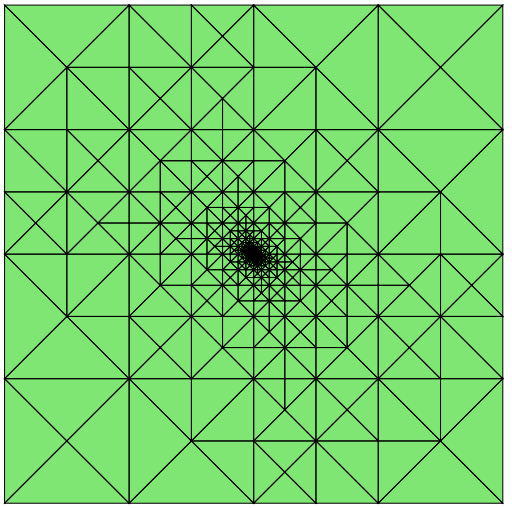

Fig. 9. Example 3: P_2 - mesh - η_R.

Table 3
Example 3 — Kellogg's Example with P_1 and P_2 discretizations ($\epsilon_{\text{rel}} = 0.05$).

u_τ	Estimator	DOFs	$\|A^{1/2}\nabla e\|/\|A^{1/2}\nabla u\|$	eff-ind
P_1	η_R	35707	4.9E−2	1.96
	ξ	29072	4.8E−2	1.35
P_2	η_R	5133	4.9E−2	2.48
	ξ	4429	4.9E−2	1.50

From the meshes in Figs. 5–6 and Figs. 9–10, we see that both estimators are robust with respect to the large jump of the diffusion coefficient. Moreover, optimal convergence rates are observed from Figs. 7–8 for P_1 element and from Figs. 11–12 for P_2 element. From Table 3, Figs. 7–8 and Figs. 11–12, we observe that, although the two estimators are proved to be equivalent in Section 4.2, ξ (or $\tilde{\xi}$) is more accurate than η_R.

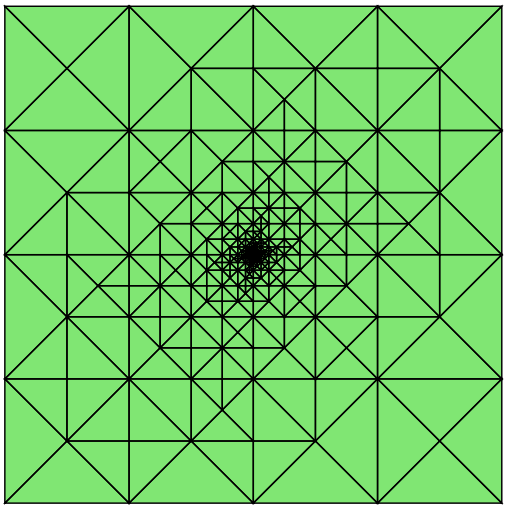

Fig. 10. Example 3: P_2 - mesh - ξ.

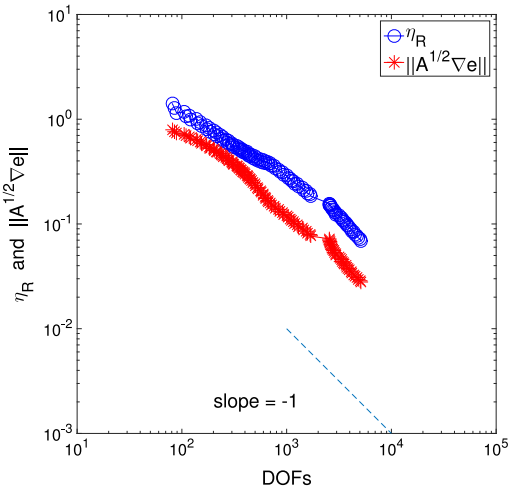

Fig. 11. Example 3: P_2 - error - η_R.

Next we present numerical results to demonstrate that, for the estimator ξ(as well as $\tilde{\xi}$), the term ξ_f defined below, which measures the divergence error between recovered flux and exact flux (actually J_K in (3.8)), is not a higher order term.

$$\xi_f := \left(\sum_{K\in\mathcal{T}} h_K^2\alpha_K^{-1}\|\bar{f}-\hat{f}_K\|_K^2\right)^{1/2}$$
$$= \left(\sum_{K\in\mathcal{T}} h_K^2\alpha_K^{-1}\|J_K\|_K^2\right)^{1/2} = \left(\sum_{K\in\mathcal{T}} h_K^2\alpha_K^{-1}|K|J_K|^2\right)^{1/2}.$$

It can be seen from Figs. 13 and 14 that ξ_f is of the same order as the true error.

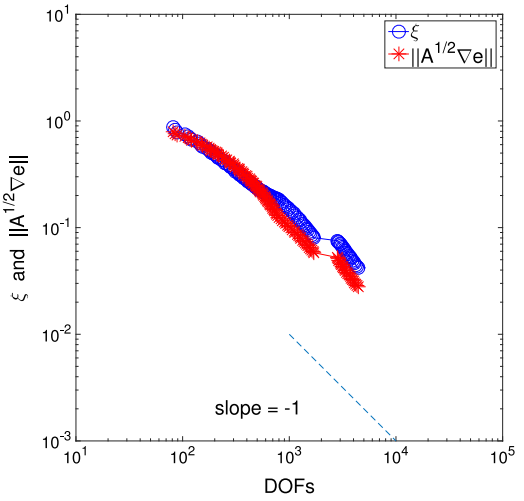

Fig. 12. Example 3: P_2 - error - ξ.

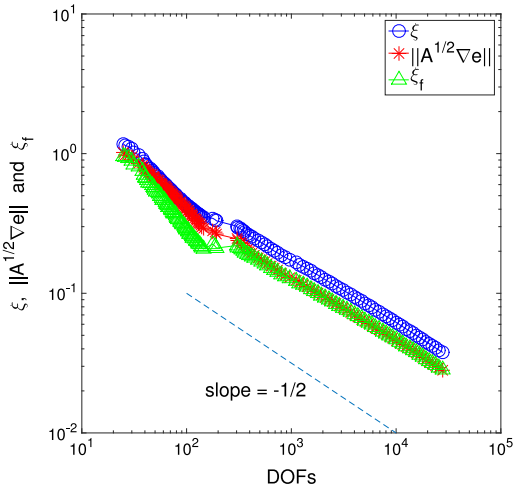

Fig. 13. Example 3: P_1 - error, ξ and ξ_f.

Remark 5.1. When computing $\hat{g}_e$ in (3.4), we use the weight λ_e defined in (3.16) for the ease of implementation.

5.4. Loss of reliability for ZZ-type estimators

This section aims to show that all ZZ-type estimators, without incorporating f, are in general not reliable on coarse meshes. A counter example is constructed below, where all estimators solely computed from the finite element solution vanish everywhere while the true error is not zero.

Let Ω be the unit square. Consider a triangulation $\mathcal{T}$ of Ω as illustrated in Fig. 15, where the top-left element is denoted by K_1.

Consider the P_1 conforming finite element discretization associated with mesh $\mathcal{T}$ of the following homogeneous Dirichlet problem:

$$\begin{cases} \Delta u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

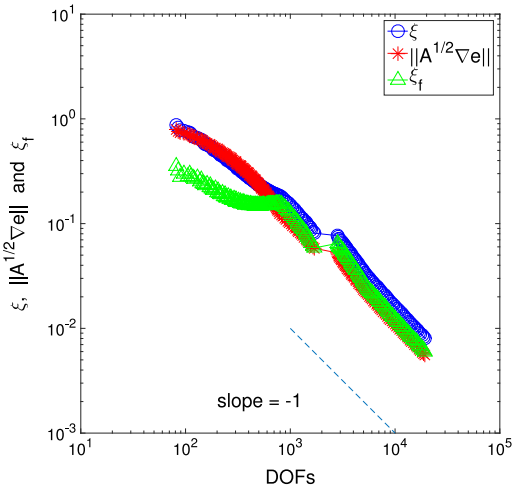

Fig. 14. Example 3: P_2 - error, ξ and ξ_f.

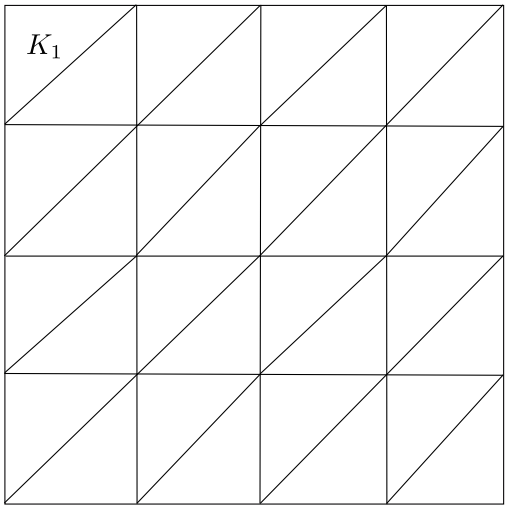

Fig. 15. Initial mesh $\mathcal{T}$ and element K_1.

where $f = 2018$ in K_1 and $f = 0$ in $\Omega \setminus K_1$. Note that for any $v \in V_{\mathcal{T}}$, $v|_{K_1} = 0$, which implies that

$$(f, v) = 0, \quad \forall\, v \in V_{\mathcal{T}}.$$

Therefore, the finite element solution is $u_{\mathcal{T}} = 0$ in Ω. All ZZ-type estimators computed solely from $u_{\mathcal{T}}$ are equal to zero. On the other hand, the true error in energy norm is $\|\nabla u\| > 0$. Hence those estimators are not reliable and consequently can not lead to a convergent adaptive finite element algorithm. In fact, based on Dörfler's marking strategy as before, since those estimators vanish everywhere, it suffices to choose an arbitrary element to form $\hat{\mathcal{T}}$ and (5.1) automatically holds true. With the initial mesh $\mathcal{T}$ in Fig. 15, if the element in $\hat{\mathcal{T}}$ is chosen as the bottom-left element on the horizontal boundary, then the mesh after 100 refinement steps based on unreliable estimators is shown in Fig. 16, where the numerical solution (as well as the resulting estimator) is always zero and convergence can never be achieved.

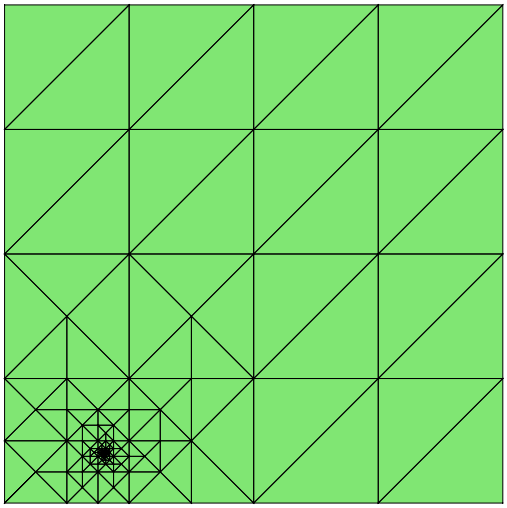

Fig. 16. The mesh after 100 refinement steps with unreliable estimators.

Remark 5.2. Note that there is no oscillation error in f in the counter example above, so the MNS marking strategy proposed in [25] coincides with Dörfler's marking strategy. Also, it is easy to see that such a counter example can always be constructed by choosing f to be orthogonal to $V_{\mathcal{T}}$ no matter the mesh $\mathcal{T}$ is considered coarse or fine.

References

[1] U.S.N.C.T.A. Mechanics, C.E.T. Systems, B.M.E. Design, D.E.P. Sciences, N.R. Council, Research directions in computational mechanics, A Series, National Academies Press, 1991.

[2] I. Babuška, W.C. Rheinboldt, Error estimates for adaptive finite element computations, SIAM J. Numer. Anal. 15 (4) (1978) 736–754.

[3] I. Babuška, A. Miller, A feedback element method with a posteriori error estimation: Part I. The finite element method and some basic properties of the a posteriori error estimator, Comput. Methods Appl. Mech. Engrg. 61 (1) (1987) 1–40.

[4] M. Ainsworth, J.T. Oden, A Posteriori Error Estimation in Finite Element Analysis, Wiley, 2000 A Wiley-Interscience publication.

[5] C. Bernardi, R. Verfürth, Adaptive finite element methods for elliptic equations with non-smooth coefficients, Numer. Math. 85 (4) (2000) 579–608.

[6] R. Verfürth, A posteriori error estimation techniques for finite element methods, Numerical Mathematics and Scientific Computation, OUP, Oxford, 2013.

[7] M. Petzoldt, A posteriori error estimators for elliptic equations with discontinuous coefficients, Adv. Comput. Math. 16 (1) (2002) 47–75.

[8] C. Carstensen, C. Merdon, Estimator competition for Poisson problems, J. Comput. Math. (2010) 309–330.

[9] O.C. Zienkiewicz, J.Z. Zhu, A simple error estimator and adaptive procedure for practical engineerng analysis, Internat. J. Numer. Methods Engrg. 24 (2) (1987) 337–357.

[10] O.C. Zienkiewicz, J.Z. Zhu, The superconvergent patch recovery and a posteriori error estimates. Part 1: The recovery technique, Internat. J. Numer. Methods Engrg. 33 (7) (1992) 1331–1364.

[11] C. Carstensen, 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 (239) (2002) 945–969.

[12] R. Bank, J. Xu, Asymptotically exact a posteriori error estimators, part I: Grids with superconvergence, SIAM J. Numer. Anal. 41 (6) (2003) 2294–2312.

[13] J.S. Ovall, Two dangers to avoid when using gradient recovery methods for finite element error estimation and adaptivity, Technical report 6, 2006, Max-Planck-Institute fur Mathematick in den Naturwissenschaften, Bonn, Germany.

[14] Z. Cai, S. Zhang, Recovery-based error estimator for interface problems: Conforming linear elements, SIAM J. Numer. Anal. 47 (3) (2009) 2132–2156.

[15] Z. Cai, C. He, S. Zhang, Improved ZZ a posteriori error estimators for diffusion problems: Conforming linear elements, Comput. Methods Appl. Mech. Engrg. 313 (2017) 433–449.

[16] R. Bank, J. Xu, B. Zheng, Superconvergent derivative recovery for Lagrange triangular elements of degree p on unstructured grids, SIAM J. Numer. Anal. 45 (5) (2007) 2032–2046.

[17] A. Naga, Z. Zhang, The polynomial-preserving recovery for higher order finite element methods in 2D and 3D, Discrete Contin. Dyn. Syst. Ser. B 5 (3) (2005) 769.

[18] Z. Cai, S. Zhang, Flux recovery and a posteriori error estimators: Conforming elements for scalar elliptic equations, SIAM J. Numer. Anal. 48 (2) (2010) 578–602.

[19] P. Ciarlet, The Finite Element Method for Elliptic Problems, Society for Industrial and Applied Mathematics, 2002.
[20] D. Boffi, M. Fortin, F. Brezzi, Mixed finite element methods and applications, Springer Series in Computational Mathematics, Springer, Berlin, Heidelberg, 2013.
[21] S. Cochez-Dhondt, S. Nicaise, Equilibrated error estimators for discontinuous Galerkin methods, Numer. Methods Partial Differential Equations 24 (5) (2008) 1236–1252.
[22] R. Verfürth, A note on constant-free a posteriori error estimates, SIAM J. Numer. Anal. 47 (4) (2009) 3180–3194.
[23] R. Verfürth, A posteriori error estimation and adaptive mesh-refinement techniques, J. Comput. Appl. Math. 50 (1) (1994) 67–83.
[24] W. Dörfler, A convergent adaptive algorithm for Poissons equation, SIAM J. Numer. Anal. 33 (3) (1996) 1106–1124.
[25] P. Morin, R.H. Nochetto, K.G. Siebert, Convergence of adaptive finite element methods, SIAM Rev. 44 (4) (2002) 631–658.
[26] E.G. Sewell, Automatic generation of triangulations for piecewise polynomial approximation, Purdue University, West Lafayette, IN, 1972.
[27] R.B. Kellogg, On the Poisson equation with intersecting interfaces, Appl. Anal. 4 (2) (1974) 101–129.