Numerische Mathematik (2020) 144:1–21
https://doi.org/10.1007/s00211-019-01075-1

Numerische Mathematik

Robust equilibrated a posteriori error estimator for higher order finite element approximations to diffusion problems

Difeng Cai[1] · Zhiqiang Cai[2] · Shun Zhang[3]

Received: 22 January 2018 / Revised: 24 July 2019 / Published online: 10 October 2019

Abstract
We present a patch-based equilibrated flux recovery procedure for the conforming finite element approximation to diffusion problems. The recovered flux is computed as the solution to a local constraint-free minimization problem on each patch. The approach is valid for higher order conforming elements in both two and three dimensions. The resulting estimator admits guaranteed reliability and the robust local efficiency is proved under the quasi-monotonicity condition of the diffusion coefficient. Numerical experiments are given to confirm the theoretical results.

Mathematics Subject Classification 65N15 · 65N30 · 65N50

1 Introduction

A posteriori error estimators based on equilibrated flux recovery have been popular recently, since they usually yield guaranteed upper bounds of the true error as a result of the Prager–Synge Theorem [5,22]. Estimators of this type are perfect for discretization

Zhiqiang Cai: This work was supported in part by the National Science Foundation under Grant DMS-1522707. Shun Zhang: This work was supported in part by Hong Kong Research Grants Council under the GRF Grant Project No. 11305319, CityU 9042090.

✉ Zhiqiang Cai
caiz@purdue.edu

Difeng Cai
difeng.cai@emory.edu

Shun Zhang
shun.zhang@cityu.edu.hk

[1] Department of Mathematics, Emory University, 201 Dowman Drive, Atlanta, GA 30322, USA

[2] Department of Mathematics, Purdue University, 150 N. University Street, West Lafayette, IN 47907-2067, USA

[3] Department of Mathematics, City University of Hong Kong, Kowloon, Hong Kong SAR, China

error control on both fine and coarse meshes, and error control on pre-asymptotic meshes is important in practice but very difficult. The intial work in this direction by Ladeveze and Leguillon [17], used a partition of unity to reduce the construction of an equilibrated flux to vertex patch based local computations. Hence this approach is computationally more efficient compared to global flux recovery procedures.

Driven by the advantage of such a local procedure, a vast number of approaches on equilibrated flux recovery for diffusion problems have been proposed (cf. [1,5], etc.). In particular, for the conforming linear finite element approximation to the Poisson equation in two dimensions, an equilibrated flux in the lowest order Raviart–Thomas space was explicitly constructed by Braess and Schöberl [7]. Their procedure starts with a decomposition of the error flux into local error fluxes by a partition of unity, local error fluxes are then approximated by vertex patch problems, and finally each vertex patch problem is solved explicitly by computing the normal components of the recovered local error flux on each edge through circling elements around the vertex at the center.

Extensions of this simple procedure to three dimensions and to higher order elements are non-trivial. Nevertheless, there are many efforts in this regards recently. An attempt was made in [12] on extension to higher order elements, but the resulting admissible flux (constructed in [12, p. 157]) is actually not equilibrated in general. The newest result on extension to three dimensions for the linear elements is reported in an unpublished manuscript by Ern and Vohralik [15] using techniques from the polytopes [26]. Their method is based on a specific enumeration of all faces in a vertex patch. However, it is not easy to obtain such an enumeration in practice even though it exists theoretically. One aim of this paper is to introduce a new approach to efficiently compute an admissible equilibrated flux for higher order finite elements in both two and three dimensions.

For singularly-perturbed diffusion-reaction and elliptic interface problems, the resulting equilibrated indicator in [7] is not robust with respect to parameters of the underlying problem (see [12,24]). To guarantee the robustness, Cai and Zhang [12] followed the ideas in [7,24] and introduced an additional minimization procedure on each vertex patch for higher order finite element approximations to the elliptic interface problem. Moreover, it was shown that the equilibrated estimator introduced in [12] is robust under the quasi-monotonicity condition [21] of the diffusion coefficient. Apart from the equilibrated flux recovery, other types of recovery-based estimators are available, including the Zienkiewicz–Zhu (ZZ) estimator [27,28], the derivative recovery [2], the polynomial-preserving recovery [20], the global projection [11], the hybrid estimator [9], etc. Those methods do not impose equilibrium condition on the recovered flux and hence the corresponding estimator can not provide a *guaranteed* upper bound of the true error on coarse meshes.

In this paper, we focus on the computation of the equilibrated flux directly (instead of the error flux as in [5,7,12]) in $H(\mathrm{div})$-conforming Raviart–Thomas spaces. A simple procedure is presented to compute an admissible equilibrated flux. In order to obtain a robust estimator, similar to [12], a local minimization is imposed to generate the desired equilibrated flux. Thanks to the construction of an admissible equilibrated flux, the local minimization problem can be solved easily by a simple projection. It should be emphasized that the proposed approach is valid for higher order conforming

finite elements in both two and three dimensions, while the procedures presented in [5,7] based on circling elements around the center vertex in a vertex patch only work for the lowest order discretizations in two dimensions. A complete algorithm is given in Sect. 4.

Theoretically, the proposed estimator is shown to be efficient, where the efficiency bound is independent of the coefficient jump under the quasi-monotonicity condition [21]. Similar to [12], the proof relies on the mixed formulation of the local minimization problem. However, due to the direct recovery of the exact flux in $H(\mathrm{div};\Omega)$ instead of the error flux in a broken space (not in $H(\mathrm{div};\Omega)$), the constraint of normal component across inter-element faces is removed and the proof (presented in Sect. 6.3) is much simpler than that in [12].

The rest of the paper is organized as follows. In Sect. 2, we introduce the model problem and its finite element approximation. In Sect. 3, the equilibrated flux recovery based on minimization on vertex patches is formulated, and Sect. 4 presents a complete procedure on solving the patch-based constrained minimization problem efficiently. The resulting error estimator is defined in Sect. 5. The robust local efficiency bound is proved in Sect. 6. Numerical experiments are presented in Sect. 7 to illustrate the performance of the proposed estimator for both P_1 and P_2 elements.

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 $\partial\Omega = \overline{\Gamma}_D \cup \overline{\Gamma}_N$ and $\Gamma_D \cap \Gamma_N = \emptyset$. For simplicity, assume that $\mathrm{meas}(\Gamma_D) > 0$. Consider the following diffusion problem:

$$\begin{cases} -\mathrm{div}(A\nabla u) = f, & \text{in } \Omega, \\ u = 0, & \text{on } \Gamma_D, \\ (-A\nabla u)\cdot \boldsymbol{n} = g_N, & \text{on } \Gamma_N, \end{cases} \tag{2.1}$$

where for almost all $x \in \Omega$, $A(x)$ is a symmetric, positive definite matrix whose smallest eigenvalue is no less than a positive constant independent of x, $f \in L^2(\Omega)$, $g_N \in L^2(\Gamma_N)$, and $\boldsymbol{n}$ is the unit outward vector normal to Γ_N.

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

$$a(u,v) := \int_\Omega A\nabla u\cdot\nabla v dx = \int_\Omega fv dx - \int_{\Gamma_N} g_N v ds, \quad \forall\, v \in H^1_D(\Omega). \tag{2.2}$$

It can be verified that the bilinear form $a(\cdot,\cdot)$ defines an inner product in $H^1_D(\Omega)$. Thus the well-posedness of (2.2) follows from the Riesz Representation Theorem.

Let $\mathcal{T} = \{K\}$ be a regular triangular $(d = 2)$/tetrahedral $(d = 3)$ partition of Ω. We define the following sets associated with the mesh $\mathcal{T}$:

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

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

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

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

$\mathcal{E}_N$: set of edges ($d = 2$)/faces ($d = 3$) on Γ_N,
$\mathcal{E}_S$: set of edges ($d = 2$)/faces ($d = 3$) contained in the closure of S.

Denote by h_K and h_e the diameters of $K \in \mathcal{T}$ and $e \in \mathcal{E}$, respectively. We use $\boldsymbol{n}$ to denote a unit outward vector normal to the boundary of an element or to the boundary of the domain Ω. For each $e \in \mathcal{E}$, we associate a unit vector normal to e, denoted by $\boldsymbol{n}_e$, where $\boldsymbol{n}_e$ is an outward vector if $e \subset \partial\Omega$. For $e \in \mathcal{E}_I$, let K_e^+ and K_e^- denote the two elements adjacent to e such that the unit outward normal for K_e^+ on e coincides with $\boldsymbol{n}_e$. For any $K \in \mathcal{T}$ and $e \in \mathcal{E}_K$, define the following sign function in $L^2(\partial K)$ associated with K:

$$s_K : \partial K \to \{-1, 1\}, \quad s_K|_e = \begin{cases} 1, & \text{if } \boldsymbol{n}_e \text{ is outward,} \\ -1, & \text{if } \boldsymbol{n}_e \text{ is inward.} \end{cases} \tag{2.3}$$

Then $s_K|_e \boldsymbol{n}_e$ is outward. For $e \in \mathcal{E}_I$ and a piecewise continuous function ϕ, we define $[\![\phi]\!]_e := \phi_+ - \phi_-$, where ϕ_+ and ϕ_- denote the limits of ϕ on e from K_e^+ and K_e^-, respectively.

Assume that for each $K \in \mathcal{T}$, $A|_K$ is a symmetric, positive definite, constant matrix. Let $P_k(K)$ and $P_k(e)$ denote the sets of polynomials of degree less than or equal to $k \geq 0$ on K and e, respectively. Correspondingly, let Π_K^k and Π_e^k denote the L^2 projections from $L^2(K)$ to $P_k(K)$ and from $L^2(e)$ to $P_k(e)$, respectively. Let $(\cdot, \cdot)_S$ and $\|\cdot\|_S$ denote the L^2 inner product and the L^2 norm over a set S, respectively. The subscript is omitted when $S = \Omega$.

We define the continuous finite element space of order k ($k \geq 1$) by

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

To simplify the presentation, as in [6,12], we assume that $f \in L^2(\Omega)$ and $g_N \in L^2(\Gamma_N)$ are piecewise polynomials such that

$$f|_K \in P_{k-1}(K), \quad \forall\, K \in \mathcal{T} \quad \text{and} \quad g_N|_e \in P_{k-1}(e), \quad \forall\, e \in \mathcal{E}_N.$$

Otherwise, the piecewise polynomial approximation of the data is used and the data oscillation is regarded as a higher order term (cf. [3,7,9]). The finite element solution $u_{\mathcal{T}} \in V_{\mathcal{T}}^k$ satisfies

$$a(u_{\mathcal{T}}, v_{\mathcal{T}}) = \int_\Omega f v_{\mathcal{T}} dx - \int_{\Gamma_N} g_N v_{\mathcal{T}} ds, \quad \forall\, v_{\mathcal{T}} \in V_{\mathcal{T}}^k. \tag{2.4}$$

The well-posedness of problem (2.4) follows from the Riesz Representation Theorem.

3 Equilibrated flux recovery

The idea of an equilibrated flux recovery is to construct a flux with certain properties that are satisfied by the true flux, $\boldsymbol{\sigma} = -A\nabla u$. Specifically, we notice that the true flux

enjoys the following properties:

$$\boldsymbol{\sigma} \in H(\mathrm{div};\Omega), \quad \mathrm{div}\,\boldsymbol{\sigma} = f \text{ in } \Omega, \quad \text{and } \boldsymbol{\sigma}\cdot\boldsymbol{n} = g_N \text{ on } \Gamma_N, \tag{3.1}$$

where $H(\mathrm{div};\Omega)$ is the space of all square-integrable vector fields whose divergence is also square-integrable over Ω. The Prager–Synge Theorem (cf. [22] or [5, p. 148]) states that, for any $\boldsymbol{\tau}$ satisfing the conditions in (3.1), the following identity holds

$$\|A^{1/2}\nabla(u_{\mathcal{T}} - u)\| + \|A^{1/2}\nabla u + A^{-1/2}\boldsymbol{\tau}\| = \|A^{1/2}\nabla u_{\mathcal{T}} + A^{-1/2}\boldsymbol{\tau}\|,$$

which immediately implies

$$\|A^{1/2}\nabla(u_{\mathcal{T}} - u)\| \le \|A^{1/2}\nabla u_{\mathcal{T}} + A^{-1/2}\boldsymbol{\tau}\|. \tag{3.2}$$

Thus we aim to find a flux that fulfills (3.1) and that minimizes the right-hand side of (3.2). However, it would be expensive to solve a global constraint minimization problem, so a local procedure is usually preferred (see e.g., [7,12]).

3.1 Localization

Let ϕ_z be the nodal basis function associated with vertex z, i.e., ϕ_z is continuous and piecewise linear over mesh $\mathcal{T}$ with $\phi_z(z) = 1$ and $\phi_z(z') = 0, \forall\, z' \in \mathcal{N}\backslash\{z\}$. Consider a partition of unity via ϕ_z:

$$\sum_{z\in\mathcal{N}} \phi_z(x) = 1, \quad \forall\, x \in \overline{\Omega}. \tag{3.3}$$

The vertex patch associated to a vertex z is given by

$$\omega_z := \mathrm{supp}\,\phi_z.$$

The partition of unity in (3.3) produces a decomposition of $\boldsymbol{\sigma}$:

$$\boldsymbol{\sigma} = \sum_{z\in\mathcal{N}} \phi_z\boldsymbol{\sigma} = \sum_{z\in\mathcal{N}} \boldsymbol{\sigma}_z,$$

where $\boldsymbol{\sigma}_z$ vanishes outside ω_z. Thus we try to construct an approximation of $\boldsymbol{\sigma}_z$ such that its sum over all vertices satisfies (3.1).

3.2 Conditions at continuous level

To derive conditions on approximations to $\boldsymbol{\sigma}_z$, let us look at necessary conditions for the true flux. Note that $\boldsymbol{\sigma}_z = -\phi_z A\nabla u$ and that

$$\mathrm{div}\,\boldsymbol{\sigma}_z = \nabla\phi_z\cdot\boldsymbol{\sigma} + \phi_z f.$$

However, the true $\boldsymbol{\sigma}$ is unknown, so we replace it by its approximation, i.e., the numerical flux: $\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} = -A\nabla u_{\mathcal{T}}$, to obtain the first condition

$$\operatorname{div} \hat{\boldsymbol{\sigma}}_z = \nabla \phi_z \cdot \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} + \phi_z f. \tag{3.4}$$

The boundary conditions for $\hat{\boldsymbol{\sigma}}_z \in H(\operatorname{div}; \omega_z)$ can be immediately seen as below:

$$\begin{cases} \hat{\boldsymbol{\sigma}}_z|_e \cdot \boldsymbol{n}_e = \phi_z g_N, & \text{if } e \in \mathcal{E}_N, \\ \hat{\boldsymbol{\sigma}}_z|_e \cdot \boldsymbol{n}_e = 0, & \text{if } e \in \mathcal{E}_z. \end{cases} \tag{3.5}$$

Here $\mathcal{E}_z$ is a subset of boundary edges $(d = 2)$/faces $(d = 3)$ associated with vertex patch ω_z where ϕ_z vanishes:

$$\mathcal{E}_z := \begin{cases} \{e \in \mathcal{E} : e \subset \partial\omega_z\}, & \text{if } z \notin \partial\Omega; \\ \{e \in \mathcal{E} : e \subset \partial\omega_z \backslash \partial\Omega\}, & \text{if } z \in \partial\Omega. \end{cases} \tag{3.6}$$

Moreover, $\hat{\boldsymbol{\sigma}}_z$ vanishes outside ω_z. It can be verified that the two conditions in (3.5) are compatible.

Define

$$\hat{\boldsymbol{\sigma}} = \sum_{z \in \mathcal{N}} \hat{\boldsymbol{\sigma}}_z. \tag{3.7}$$

Then we we have the following proposition.

Proposition 3.1 *For each vertex z, assume that $\hat{\boldsymbol{\sigma}}_z$ satisfies the conditions in (3.4) and (3.5). Then the flux $\hat{\boldsymbol{\sigma}}$ defined in (3.7) satisfies (3.1).*

Proof According to (3.4) and (3.3), we have

$$\operatorname{div} \hat{\boldsymbol{\sigma}} = \sum_{z \in \mathcal{N}} \operatorname{div} \hat{\boldsymbol{\sigma}}_z = \sum_{z \in \mathcal{N}} \left(\nabla \phi_z \cdot \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} + \phi_z f \right) = f.$$

From (3.5), it follows immediately that $\hat{\boldsymbol{\sigma}} \in H(\operatorname{div}; \Omega)$ and that $\hat{\boldsymbol{\sigma}} \cdot \boldsymbol{n} = g_N$ on Γ_N. Therefore, we conclude that $\hat{\boldsymbol{\sigma}}$ satisfies (3.1). □

We next show that the conditions on $\hat{\boldsymbol{\sigma}}_z$ are well-posed. In other words, we prove the existence of $\hat{\boldsymbol{\sigma}}_z$ with conditions (3.4) and (3.5) at the continuous level.

Theorem 3.1 *Suppose that $u_{\mathcal{T}}$ is the finite element solution defined in (2.4). For each vertex $z \in \mathcal{N}$, there exists a flux $\hat{\boldsymbol{\sigma}}_z \in H(div; \Omega)$ satisfying (3.4) and (3.5).*

Proof Consider the following Neumann problem:

$$\begin{cases} \operatorname{div}(\nabla v) = \nabla \phi_z \cdot \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} + \phi_z f, & \text{in } \omega_z, \\ \nabla v \cdot \boldsymbol{n} = 0, & \text{on } \partial\omega_z \backslash \partial\Omega, \\ \nabla v \cdot \boldsymbol{n} = \phi_z g_N, & \text{on } \partial\omega_z \cap \Gamma_N, \\ \nabla v \cdot \boldsymbol{n} = C, & \text{on } \partial\omega_z \cap \Gamma_D, \end{cases} \tag{3.8}$$

where C is a constant defined by

$$C := \text{meas}(\partial\omega_z \cap \Gamma_D)^{-1} \left(\int_{\omega_z} \nabla\phi_z \cdot \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} + \phi_z f dx - \int_{\partial\omega_z \cap \Gamma_N} \phi_z g_N ds \right)$$

if $z \in \partial\Omega$ and $\text{meas}(\partial\omega_z \cap \Gamma_D) > 0$; otherwise, $C = 0$. We show that the compatibility condition below for the Neumann problem in (3.8) always holds true:

$$\int_{\omega_z} \nabla\phi_z \cdot \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} + \phi_z f dx = \int_{\partial\omega_z \cap \Gamma_D} C ds + \int_{\partial\omega_z \cap \Gamma_N} \phi_z g_N ds. \tag{3.9}$$

According to the definition of C, it is obvious that (3.9) is true when $z \in \partial\Omega$ and $\text{meas}(\partial\omega_z \cap \Gamma_D) > 0$. Otherwise, we have $\phi_z \in V_{\mathcal{T}}^1$, and (3.9) follows immediately from the weak formulation for $u_{\mathcal{T}}$ in (2.4) by choosing $v_{\mathcal{T}} = \phi_z \in V_{\mathcal{T}}^1$. Therefore, we conclude that the Neumann problem in (3.8) is solvable.

Let $v \in H^1(\omega_z)$ be a solution to (3.8). By setting $\nabla v = 0$ outside ω_z, we see that $\nabla v \cdot \boldsymbol{n}$ is continuous across $\partial\omega_z \cap \Omega$, so $\nabla v \in H(\text{div}; \Omega)$. Therefore, $\hat{\boldsymbol{\sigma}}_z := \nabla v \in H(\text{div}; \Omega)$ satisfies (3.4) and (3.5). □

3.3 Local equilibrated flux recovery in Raviart–Thomas space via minimization

In this section, we consider conditions (3.4) and (3.5) at the discrete level. To approximate the flux, we consider the Raviart–Thomas space associated with the triangulation $\mathcal{T} = \{K\}$. For each $K \in \mathcal{T}$, the Raviart–Thomas space of index $k-1$ on element K is defined by

$$\text{RT}_{k-1}(K) := \left\{ \boldsymbol{\tau} \in L^2(K)^d : \boldsymbol{\tau} = \boldsymbol{p} + \boldsymbol{x}q,\ \boldsymbol{p} \in P_{k-1}(K)^d,\ q \in P_{k-1}(K) \right\}.$$

The $H(\text{div}; \Omega)$-conforming and the broken Raviart–Thomas spaces of index $k-1$ are then given by

$$\text{RT}_{k-1} := \{\boldsymbol{\tau} \in H(\text{div}; \Omega) : \boldsymbol{\tau}|_K \in \text{RT}_{k-1}(K),\ \forall K \in \mathcal{T}\}$$

and

$$\text{RT}_{k-1}^{-1} := \{\boldsymbol{\tau} \in L^2(\Omega)^d : \boldsymbol{\tau}|_K \in \text{RT}_{k-1}(K),\ \forall K \in \mathcal{T}\},$$

respectively. Corresponding to the conforming finite element space $V_{\mathcal{T}}^k$, approximation $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}$ to $\hat{\boldsymbol{\sigma}}_z$ is required to satisfy

$$\text{div}\, \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}|_K \in P_{k-1}(K), \quad \forall K \in \mathcal{T} \quad \text{and} \quad \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \cdot \boldsymbol{n}_e \in P_{k-1}(e), \quad \forall e \in \mathcal{E}.$$

Therefore, at the discrete level, the conditions are the discrete equilibrium equation:

$$\text{div}\, \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}|_K = \bar{f}_z|_K := \Pi_K^{k-1} \left(\nabla\phi_z \cdot \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} + \phi_z f \right), \quad \forall K \in \mathcal{T} \tag{3.10}$$

and the boundary conditions

$$\begin{cases} \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \cdot \boldsymbol{n}_e = \Pi_e^{k-1}(\phi_z g_N), & \text{if } e \in \mathcal{E}_N, \\ \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \cdot \boldsymbol{n}_e = 0, & \text{if } e \in \mathcal{E}_z. \end{cases} \tag{3.11}$$

By incorporating the constraints in (3.11), we define the following subset of RT_{k-1} associated with vertex z:

$$\mathrm{RT}_{z,g} := \{\boldsymbol{\tau} \in \mathrm{RT}_{k-1} : \operatorname{supp} \boldsymbol{\tau} \subseteq \omega_z,\ \boldsymbol{\tau} \text{ satisfies (3.11)}\}. \tag{3.12}$$

Moreover, we define below the set Σ_z as the collection of desired flux satisfying (3.10) and (3.11):

$$\Sigma_z := \{\boldsymbol{\tau} \in \mathrm{RT}_{z,g} : \operatorname{div} \boldsymbol{\tau} = \bar{f}_z\}. \tag{3.13}$$

For each vertex $z \in \mathcal{N}$, we look for a flux $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \in \Sigma_z$ such that $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}$ solves the following constrained minimization problem on ω_z:

$$\|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\|_{\omega_z} = \min_{\boldsymbol{\tau} \in \Sigma_z} \|A^{-1/2}(\boldsymbol{\tau} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\|_{\omega_z}, \tag{3.14}$$

where

$$\widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}|_K = \Pi_{\mathrm{RT}_{k-1}(K)}\left(\phi_z \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}\right), \quad \forall\, K \in \mathcal{T} \tag{3.15}$$

and $\Pi_{\mathrm{RT}_{k-1}(K)}$ denotes the interpolation operator from $H(\mathrm{div}; K)$ to $\mathrm{RT}_{k-1}(K)$ (cf. [4, Ch.2.5] or [5, Ch.III.5]).

Since Σ_z is a closed convex subset of RT_{k-1}, the minimization problem in (3.14) is uniquely solvable whenever Σ_z is non-empty. As we shall see later in Sect. 5, (3.14) will be chosen as the local indicator.

We prove the existence of an equilibrated local flux $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \in \Sigma_z$ first in Theorem 3.2, then we construct one $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \in \Sigma_z$ in Sect. 4.

As a discrete version of Theorem 3.1, the existence of an equilibrated local flux $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \in \Sigma_z$ can be easily proved.

Theorem 3.2 *Suppose that $u_{\mathcal{T}} \in V_{\mathcal{T}}^k$ $(k \geq 1)$ is the finite element solution defined in (2.4). For each vertex $z \in \mathcal{N}$, there exists a $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \in \mathrm{RT}_{k-1}$ satisfying (3.10) and (3.11). Hence Σ_z is non-empty.*

Proof According to Theorem 3.1, there exists a $\boldsymbol{\tau}_z \in H(\mathrm{div}; \Omega)$ satisfying the conditions in (3.4) and (3.5) at the continuous level. Define $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}$ by setting $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}|_K := \Pi_{\mathrm{RT}_{k-1}(K)}\boldsymbol{\tau}_z$ for all $K \in \mathcal{T}$. Then it is easy to see that $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \in \mathrm{RT}_{k-1}$ and $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}$ satisfies (3.10) and (3.11). □

Remark 3.1 It can be easily verified that there is a relation between $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} \in \Sigma_z$ in Sect. 3.3 and the recovered error flux $\boldsymbol{\sigma}_z^{\Delta} \in \mathrm{RT}_{k-1}^{-1}$ in [12] (see also [5,7] without imposing the minimization in (3.14)) given by

$$\boldsymbol{\sigma}_z^{\Delta} = \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}},$$

where $\widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}$ is defined in (3.15).

4 Solution of the constrained minimization problem

To compute the recovered local flux on each vertex patch, Braess [5] used an explicit procedure in two dimensions to compute the recovered flux, which requires an enumeration of elements in a specific direction around the center vertex. This simple approach is only valid for the RT_0 recovery in the lowest order finite element discretization and the extension to three dimensions is not straightforward due to the enumeration issues. Moreover, without imposing any minimization in the recovery, the computed estimator in [5,7] is not robust with respect to the coefficient jump. To ensure the robustness of the estimator, a constrained minimization as in (3.14), valid for higher order elements, was introduced in [12]. Since the nonhomogeneous constraint in (3.14) is more difficult to solve than the homogeneous constraint, following the procedure in [5], an attempt to construct an admissible equilibrated flux was made in [12], but again the computed flux was equilibrated only for the lowest order discretization in two dimensions. Recently, for Poisson equations, Ern and Vohralik [15] extended the idea in [5] to three dimensions based on a specific enumeration of all faces in a vertex patch, but it is not straightforward to obtain such an enumeration in practice even though it exists theoretically. Hence, for interface problems, no simple procedure was presented so far regarding the robust equilibrated flux recovery for higher order conforming elements in both two and three dimensions.

In this section, we first present a simple algorithm to construct an admissible equilibrated flux $\hat{\boldsymbol{\sigma}}^{f}_{z,\mathcal{T}} \in \Sigma_z$, valid for higher order elements in d $(d = 2, 3)$ dimensions. Then it suffices to solve the following minimization problem over the divergence free subspace of $\mathrm{RT}_{z,0}$:

$$\hat{\boldsymbol{\sigma}}^{0}_{z,\mathcal{T}} = \underset{\substack{\boldsymbol{\tau}\in \mathrm{RT}_{z,0} \\ \operatorname{div}\boldsymbol{\tau}=0}}{\arg\min} \|A^{-1/2}(\boldsymbol{\tau} + \hat{\boldsymbol{\sigma}}^{f}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\|_{\omega_z} \tag{4.1}$$

Note that the basis functions of $\mathrm{RT}_{z,0}$ are known explicitly [see (4.4)]. Setting

$$\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} = \hat{\boldsymbol{\sigma}}^{f}_{z,\mathcal{T}} + \hat{\boldsymbol{\sigma}}^{0}_{z,\mathcal{T}}$$

gives the minimizer in (3.14).

We define

$$\tau_e = 0 \quad \text{for } e \in \mathcal{E}_z \quad \text{and} \quad \tau_e = \int_e \phi_z g_N ds \quad \text{for } e \in \mathcal{E}_N,$$

and τ_e $(e \in \mathcal{E}_{\omega_z} \backslash (\mathcal{E}_z \cup \mathcal{E}_N))$ to be a solution of the following linear system

$$\sum_{e\in\mathcal{E}_K\backslash\mathcal{E}_z} s_K \tau_e = \int_K \bar{f}_z dx, \quad \forall\, K \in \mathcal{T}_z, \tag{4.2}$$

which is indeed solvable according to Proposition 4.1. In practice, it can be solved via singular value decomposition (SVD).

Once all τ_e ($e \in \mathcal{E}_{\omega_z}$) are known, an equilibrated local flux $\hat{\boldsymbol{\sigma}}^f_{z,\mathcal{T}} \in \Sigma_z$ can be constructed below by assigning degrees of freedom in $\mathrm{RT}_{k-1}(K)$ on each $K \in \mathcal{T}_z$:

$$\begin{cases} \hat{\boldsymbol{\sigma}}^f_{z,\mathcal{T}} \cdot \boldsymbol{n}_e = \tau_e/|e|, & \forall\, e \in \mathcal{E}_K \backslash \mathcal{E}_N, \\ \hat{\boldsymbol{\sigma}}^f_{z,\mathcal{T}} \cdot \boldsymbol{n}_e = \Pi_e^{k-1}(\phi_z g_N), & \forall\, e \in \mathcal{E}_K \cap \mathcal{E}_N, \\ \int_K \hat{\boldsymbol{\sigma}}^f_{z,\mathcal{T}} \cdot \nabla p \, dx = \sum_{e\in\mathcal{E}_K} (s_K \hat{\boldsymbol{\sigma}}^f_{z,\mathcal{T}} \cdot \boldsymbol{n}_e, p)_e - (\bar{f}_z, p)_K, & \forall\, p \in P_{k-1}(K), \\ \int_K \hat{\boldsymbol{\sigma}}^f_{z,\mathcal{T}} \cdot \boldsymbol{q} \, dx = 0, & \forall\, \boldsymbol{q} \in Q_{k-2}(K), \end{cases} \tag{4.3}$$

where

$$Q_{k-2}(K) := \left\{ \boldsymbol{q} \in P_{k-2}(K)^d : (\boldsymbol{q}, \nabla p) = 0,\ \forall\, p \in P_{k-1}(K) \right\}.$$

Now in view of the minimization in (4.1) with the homogeneous constraint, the solution of $\hat{\boldsymbol{\sigma}}^0_{z,\mathcal{T}}$ in (4.1) has been considered in [12] and we briefly review it here. According to [12], we know that

$$N_z := \left\{ \boldsymbol{\tau} \in \mathrm{RT}_{z,0} : \operatorname{div} \boldsymbol{\tau} = 0 \right\} = \begin{cases} \nabla^\perp S^k_{z,0}, & \text{if } d = 2, \\ \nabla \times N d^k_{z,0}, & \text{if } d = 3, \end{cases} \tag{4.4}$$

where

$$\begin{aligned} &\nabla^\perp v = \left(-\frac{\partial v}{\partial y}, \frac{\partial v}{\partial x} \right), \\ &S^k_{z,0} := \left\{ v \in H^1(\omega_z) : v|_K \in P_k(K)\ \forall K \in \mathcal{T}_z \text{ and } v|_e = 0 \text{ on } e \in \mathcal{E}_z \cup \mathcal{E}_N \right\}, \\ &N d^k_{z,0} := \left\{ \boldsymbol{\tau} \in H(\mathbf{curl}; \omega_z) : \boldsymbol{\tau}|_K \in N d^k(K)\ \forall K \in \mathcal{T}_z \text{ and } \boldsymbol{\tau} \times \boldsymbol{n}_e = 0 \text{ on } e \in \mathcal{E}_z \cup \mathcal{E}_N \right\}, \\ &N d^k(K) := \left\{ \boldsymbol{\tau} \in L^2(K)^d : \boldsymbol{\tau} = \boldsymbol{a} + \boldsymbol{b}, \boldsymbol{a} \in P_k(K)^d, \boldsymbol{b} \in P^h_{k+1}(K)^d \text{ and } \boldsymbol{b} \cdot \boldsymbol{x} = 0 \right\} \end{aligned}$$

with $P^h_k(K)^d$ the space of homogeneous polynomials of order k on element K. The minimizer $\hat{\boldsymbol{\sigma}}^0_{z,\mathcal{T}} \in N_z$ of (4.1) is then characterized by

$$\left(A^{-1} \hat{\boldsymbol{\sigma}}^0_{z,\mathcal{T}}, \boldsymbol{\tau} \right)_{\omega_z} = \left(A^{-1} (\widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \hat{\boldsymbol{\sigma}}^f_{z,\mathcal{T}}), \boldsymbol{\tau} \right)_{\omega_z}, \quad \forall\, \boldsymbol{\tau} \in N_z. \tag{4.5}$$

The algorithm for computing the desired flux $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}$ is summarized below.

1. Solve the linear system of τ_e in (4.2) for $e \in \mathcal{E}_{\omega_z} \backslash (\mathcal{E}_z \cup \mathcal{E}_N)$ using SVD.
2. Define $\hat{\boldsymbol{\sigma}}^f_{z,\mathcal{T}}$ as in (4.3).
3. Solve the linear system associated with (4.5) and then obtain $\hat{\boldsymbol{\sigma}}^0_{z,\mathcal{T}}$.
4. Set $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} = \hat{\boldsymbol{\sigma}}^f_{z,\mathcal{T}} + \hat{\boldsymbol{\sigma}}^0_{z,\mathcal{T}}$.

Proposition 4.1 *The linear system of τ_e for $e \in \mathcal{E}_{\omega_z} \backslash (\mathcal{E}_z \cup \mathcal{E}_N)$ in (4.2) is solvable.*

Proof According to Theorem 3.2, there exists a $\hat{\boldsymbol{\sigma}}^{*}_{z,\mathcal{T}} \in \Sigma_z$. By applying the divergence theorem to $\operatorname{div} \hat{\boldsymbol{\sigma}}^{*}_{z,\mathcal{T}} = \bar{f}_z$ in each element $K \in \mathcal{T}_z$, we obtain the linear system in (4.2) with

$$\tau_e = \int_e \hat{\boldsymbol{\sigma}}^{*}_{z,\mathcal{T}} \cdot \boldsymbol{n}_e ds.$$

This implies that (4.2) is solvable. □

Remark 4.1 As long as the first three equations in (4.3) are satisfied, $\hat{\boldsymbol{\sigma}}^{f}_{z,\mathcal{T}} \in \Sigma_z$. Hence the last equation in (4.3) can actually be arbitrary.

5 A posteriori error estimate

5.1 Global error estimator and guaranteed reliability

Consider the global error estimator

$$\xi := \|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}})\|, \tag{5.1}$$

where $\hat{\boldsymbol{\sigma}}_{\mathcal{T}} = \sum_{z\in\mathcal{N}} \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}$ is the recovered global flux. According to Proposition 3.1 and inequality (3.2), the estimator above has guaranteed reliability, i.e.,

$$\|A^{1/2}\nabla(u_{\mathcal{T}} - u)\| \le \|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}})\|.$$

5.2 Local error indicator

The local error indicator is given by

$$\xi_K := \|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}})\|_K. \tag{5.2}$$

Its local efficiency is established through the triangle inequality and the local efficiency of the following indicator

$$\xi_z := \|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\|. \tag{5.3}$$

Note that ξ_z measures the error inside the vertex patch ω_z and is needed in the proof only and that ξ_K is used in the adaptive mesh refinement.

Remark 5.1 It is tempting for one to choose the local error indicator as

$$\|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \phi_z\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}})\|_{\omega_z},$$

where $u_{\mathcal{T}}$ is the finite element solution in $V^k_{\mathcal{T}}$ $(k \ge 1)$ and $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}$ is the recovered flux in RT_{k-1}. However, this is not correct as it can not guarantee the local efficiency in

general. For example, suppose that A is the identity matrix and the linear conforming finite element solution coincides with the exact solution such that $\nabla u_{\mathcal{T}} = \nabla u \neq 0$ in $K \subset \omega_z$, then the true error is 0, but $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} + \phi_z \nabla u_{\mathcal{T}} \neq 0$ because $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}|_K \in \mathrm{RT}_0(K)$ while $\phi_z \nabla u_{\mathcal{T}}|_K \notin \mathrm{RT}_0(K)$.

Remark 5.2 In existing literature, corresponding to the P_k element, both RT_{k-1} and RT_k spaces are considered in the recovery of $\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}}$. One of the earlier work in this direction, i.e., [7] (see also [5]) performed flux recovery in RT_0 space for P_1 element, which is a natural choice as used in most literature on recovery-based estimators (cf. [1,10,11,13,18]). A generalization in [12] handles P_k element with RT_{k-1} flux recovery. In those cases, as pointed out in Remark 5.1, the mapping of $\phi_z \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}|_K$ to $\mathrm{RT}_{k-1}(K)$ see (3.15)] is necessary in order to guarantee the local efficiency of the resulting error indicator. On the other hand, the work in [6] chose to use RT_k flux recovery for the P_k element. In that case, the mapping in (3.15) is not necessary, but due to the increased degrees of freedom, it would be computationally more expensive.

6 Local efficiency

We prove local efficiency of the local indicator ξ_z as well as ξ_K for the interface problem. For simplicity, assume that $A = \alpha I$, where $\alpha > 0$ is piecewise constant with respect to $\mathcal{T}$, i.e., $\alpha|_K = \alpha_K > 0,\ \forall\, K \in \mathcal{T}$. Define

$$\alpha_{\max} := \max_{K \in \mathcal{T}} \alpha_K \quad \text{and} \quad \alpha_{\min} := \min_{K \in \mathcal{T}} \alpha_K .$$

Furthermore, for each $e \in \mathcal{E}$, define

$$\alpha_{e,M} := \max_{K \subseteq \omega_e} \alpha_K \quad \text{and} \quad \alpha_{e,m} := \min_{K \subseteq \omega_e} \alpha_K , \tag{6.1}$$

where ω_e denotes the union of elements adjacent to e.

For each vertex z, let $\mathcal{T}_z$ be a subset of $\mathcal{T}$ such that $\mathcal{T}_z$ is the collection of all elements contained in the vertex patch ω_z. Let K_z be an element in $\mathcal{T}_z$ such that

$$\alpha_{K_z} = \max_{K \in \mathcal{T}_z} \alpha_K . \tag{6.2}$$

We define

$$P_z := \left\{ v \in L^2(\omega_z) : v|_K \in P_{k-1}(K),\ \forall\, K \in \mathcal{T}_z \right\}$$

and

$$\overline{P}_z := \begin{cases} \{v \in P_z : \int_{K_z} v\, dx = 0\}, & \text{if } z \in \mathcal{N} \backslash \Gamma_D, \\ \{v \in P_z : v|_{\Gamma_D} = 0\}, & \text{if } z \in \mathcal{N} \cap \Gamma_D. \end{cases} \tag{6.3}$$

We justify the local efficiency of ξ_z via proving the stability of mixed formulations corresponding to the constrained minimization problem in (3.14). The mixed formulation will be based on the Hilbert spaces: $\mathrm{RT}_{z,0}$ and $\overline{P}_z$, where $\mathrm{RT}_{z,0}$ is equipped with the $H(\mathrm{div};\Omega)$ inner product and $\overline{P}_z$ is equipped with the usual L^2 inner product.

6.1 Mixed formulation associated with the constrained minimization problem

By choosing a $\hat{\boldsymbol{\sigma}}_{z,g} \in \mathrm{RT}_{z,g}$, any $\boldsymbol{\tau} \in \mathrm{RT}_{z,g}$ can be written as

$$\boldsymbol{\tau} = \hat{\boldsymbol{\sigma}}_{z,g} + \boldsymbol{\tau}_0, \quad \boldsymbol{\tau}_0 \in \mathrm{RT}_{z,0}.$$

Hence the minimization problem in (3.14) is equivalent to: find $\hat{\boldsymbol{\sigma}}_{z,0} \in \mathrm{RT}_{z,0}$ such that

$$\|A^{-1/2}(\hat{\boldsymbol{\sigma}}_{z,0} + \hat{\boldsymbol{\sigma}}_{z,g} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\| = \min_{\substack{\boldsymbol{\tau}_0 \in \mathrm{RT}_{z,0} \\ \operatorname{div} \boldsymbol{\tau}_0 = \bar{f}_z - \operatorname{div} \hat{\boldsymbol{\sigma}}_{z,g}}} \|A^{-1/2}(\boldsymbol{\tau}_0 + \hat{\boldsymbol{\sigma}}_{z,g} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\|. \quad (6.4)$$

The corresponding mixed formulation (cf. [5]) is to find $(\hat{\boldsymbol{\sigma}}_{z,0}, u_z) \in \mathrm{RT}_{z,0} \times \overline{P}_z$ such that

$$\begin{cases} \left(A^{-1}\hat{\boldsymbol{\sigma}}_{z,0}, \boldsymbol{\tau}\right) + (\operatorname{div} \boldsymbol{\tau}, u_z) = \left(A^{-1}(\widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \hat{\boldsymbol{\sigma}}_{z,g}), \boldsymbol{\tau}\right), & \forall\, \boldsymbol{\tau} \in \mathrm{RT}_{z,0}, \\ (\operatorname{div} \hat{\boldsymbol{\sigma}}_{z,0}, v) = (\bar{f}_z - \operatorname{div} \hat{\boldsymbol{\sigma}}_{z,g}, v), & \forall\, v \in \overline{P}_z. \end{cases} \quad (6.5)$$

$\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} = \hat{\boldsymbol{\sigma}}_{z,0} + \hat{\boldsymbol{\sigma}}_{z,g} \in \mathrm{RT}_{z,g}$ is then the desired flux that solves (3.14).

6.2 *Inf-sup* conditions

Given a vertex $z \in \mathcal{N}$, we define a mesh-dependent norm on $\overline{P}_z$:

$$\|v\|_{\mathcal{T}_z}^2 := \sum_{K \in \mathcal{T}_z} \|\alpha^{1/2} \nabla v\|_K^2 + \sum_{e \in \mathcal{E}_{\omega_z} \setminus \mathcal{E}_{\partial\omega_z}} \alpha_{e,m} h_e^{-1} \|[\![v]\!]\|_e^2, \quad \forall\, v \in \overline{P}_z. \quad (6.6)$$

It is easy to verify that $\|\cdot\|_{\mathcal{T}_z}$ is a well-defined norm on $\overline{P}_z$.

Here and thereafter, we will use C with or without subscripts to denote a generic positive constant, possibly different at different occurrences, that is independent of $\alpha_{\max}/\alpha_{\min}$, but may depend on the shape parameter of the mesh $\mathcal{T}$ and the polynomial degree k. The result below is proved in a similar fashion as [8, Lemma 2.3].

Lemma 6.1 *There exists a positive constant C, depending only on polynomial degree k and the shape parameter of $\mathcal{T}$, such that the following inf-sup condition holds:*

$$\|v\|_{\mathcal{T}_z} \le C \sup_{\boldsymbol{\tau} \in \mathrm{RT}_{z,0}} \frac{(div\, \boldsymbol{\tau}, v)}{\|\alpha^{-1/2}\boldsymbol{\tau}\|}, \quad \forall\, v \in \overline{P}_z. \quad (6.7)$$

Proof For any given $v \in \overline{P}_z$, to establish the *inf-sup* condition in (6.7), it suffices to construct a $\boldsymbol{\tau} \in \mathrm{RT}_{z,0}$ such that

$$\|\alpha^{-1/2}\boldsymbol{\tau}\| \le C\|v\|_{\mathcal{T}_z} \quad \text{and} \quad (\operatorname{div} \boldsymbol{\tau}, v) = \|v\|_{\mathcal{T}_z}^2. \quad (6.8)$$

To this end, according to the degrees of freedom for $\mathrm{RT}_{k-1}(K)$, there is a unique vector field $\boldsymbol{\tau} \in \mathrm{RT}_{z,0}$ such that

$$\begin{cases} \boldsymbol{\tau} \cdot \boldsymbol{n}_e = \alpha_{e,m} h_e^{-1} [\![v]\!]_e, & \text{on } e \in \mathcal{E}_{\omega_z} \backslash \mathcal{E}_{\partial\omega_z}, \\ \boldsymbol{\tau} \cdot \boldsymbol{n}_e = 0, & \text{on } e \in \mathcal{E}_{\partial\omega_z}, \\ (\boldsymbol{\tau}, \boldsymbol{q})_K = (-\alpha \nabla v, \boldsymbol{q})_K, & \forall\, \boldsymbol{q} \in P_{k-2}(K)^d, \quad \forall\, K \in \mathcal{T}_z. \end{cases} \tag{6.9}$$

A standard scaling argument implies that (cf. [8])

$$\|\boldsymbol{\tau}\|_K^2 \le C \left(\|\alpha \nabla v\|_K^2 + h_K \|\boldsymbol{\tau} \cdot \boldsymbol{n}\|_{\partial K}^2 \right), \tag{6.10}$$

which, together with (6.1), yields that

$$\|\alpha^{-1/2} \boldsymbol{\tau}\|_K^2 \le C \left(\|\alpha^{1/2} \nabla v\|_K^2 + \sum_{e \in \mathcal{E}_K \backslash \mathcal{E}_{\partial\omega_z}} \alpha_{e,m} h_e^{-1} \|[\![v]\!]\|_e^2 \right).$$

Summing over $K \in \mathcal{T}_z$ implies the inequality in (6.8). The equality in (6.8) is a direct consequence of integration by parts element-wisely. This proves that the $\boldsymbol{\tau}$ defined in (6.9) satisfies (6.8) and, hence, the lemma. □

6.3 Proof of local efficiency via mixed formulation

We prove the local efficiency of ξ_z via bounding ξ_z from above by the residual-based local indicator defined below, which is known to have local efficiency (cf. [3,21,25]):

$$\eta_K := \left(\frac{h_K^2}{\alpha_K} \|f - f_K\|_K^2 + \frac{1}{2} \sum_{e \in \mathcal{E}_K \cap \mathcal{E}_I} \frac{h_K}{\alpha_{e,M}} \|j_e\|_e^2 + \sum_{e \in \mathcal{E}_K \cap \mathcal{E}_N} \frac{h_K}{\alpha_{e,M}} \|j_e\|_e^2 \right)^{1/2}, \tag{6.11}$$

where

$$f_K := \operatorname{div} \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}|_K \quad \text{and} \quad j_e := \begin{cases} [\![\widetilde{\boldsymbol{\sigma}}_{\mathcal{T}} \cdot \boldsymbol{n}_e]\!]_e, & \text{if } e \in \mathcal{E}_I, \\ \widetilde{\boldsymbol{\sigma}}_{\mathcal{T}}|_e \cdot \boldsymbol{n}_e - g_N, & \text{if } e \in \mathcal{E}_N, \\ 0, & \text{if } e \in \mathcal{E}_D. \end{cases} \tag{6.12}$$

It is well-known [3,25] that the local residual indicator has the following robust efficiency bound:

$$\eta_K \le C \|\alpha^{1/2} \nabla (u - u_{\mathcal{T}})\|_{\omega_K}, \tag{6.13}$$

where ω_K denotes the union of all elements that share an edge $(d = 2)$/face $(d = 3)$ with K.

To bound ξ_z from above by η_K, we need the quasi-monotonicity condition on the distribution of α [21], which is weaker than the Hypothesis 2.7 in [3]. The quasi-monotonicity condition [21] is cited below.

Definition 6.1 With K_z given in (6.2) at the beginning of Sect. 6, α is called *quasi-monotone with respect to the vertex patch* ω_z if the following conditions are satisfied:

for each $K \in \mathcal{T}_z$, there exists a Lipschitz set $\tilde{\omega}_{K,z}$ containing only elements from $\mathcal{T}_z$ such that

- if $z \notin \Gamma_D$, then $K \cup K_z \subseteq \tilde{\omega}_{K,z}$ and $\alpha_K \leq \alpha_{K'}$, $\forall\, K' \subseteq \tilde{\omega}_{K,z}$;
- if $z \in \Gamma_D$, then $K \subseteq \tilde{\omega}_{K,z}$, $\mathrm{meas}(\partial\tilde{\omega}_{K,z} \cap \Gamma_D) > 0$ and $\alpha_K \leq \alpha_{K'}$, $\forall\, K' \subseteq \tilde{\omega}_{K,z}$.

Under the quasi-monotonicity condition, the following result was proved in [12, Corollary 5.10].

Lemma 6.2 *Assume that α is quasi-monotone. For each $v \in \overline{P}_z$, there exists a constant $C > 0$ such that*

$$\sum_{K\in\mathcal{T}_z} h_K^{-2}\|\alpha^{1/2}v\|_K^2 \leq C\,\|v\|_{\mathcal{T}_z}^2.$$

To show the local efficiency of ξ_z, the following result is needed.

Proposition 6.1 *For each vertex $z \in \mathcal{N}$, there exists a vector field $\boldsymbol{\tau}_z \in \mathrm{RT}_{z,g}$ such that for each element $K \in \mathcal{T}_z$,*

$$div\,\boldsymbol{\tau}_z = div\,\widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}} + J_z, \quad in\ K \quad and \quad \|\alpha^{-1/2}(\boldsymbol{\tau}_z - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\|_K \leq C\eta_K, \tag{6.14}$$

where

$$J_z|_K = \sum_{e\in\mathcal{E}_K} |K|^{-1}\int_e -\nu_{z,K,e}\,\Pi_e^{k-1}(\phi_z j_e)ds \tag{6.15}$$

and

$$\nu_{z,K,e} := \begin{cases} \dfrac{\sqrt{\alpha_K}}{\sqrt{\alpha_{K_e^+}}+\sqrt{\alpha_{K_e^-}}}, & \text{if } e \in \mathcal{E}_{\omega_z}\backslash\mathcal{E}_{\partial\omega_z}, \\ 1, & \text{otherwise.} \end{cases} \tag{6.16}$$

The construction of a vector field $\boldsymbol{\tau}_z \in \mathrm{RT}_{z,g}$ satisfying the estimates (6.20) and (6.14) in Proposition 6.1 is similar to the flux recovery in [9]. Namely, we pose a boundary value problem for $\boldsymbol{\tau}_z$ in each element in vertex patch ω_z and choose a solution that fulfills the stability estimate.

Let $|K|$ 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.

Proof of Proposition 6.1 For each vertex $z \in \mathcal{N}$, consider the following boundary value problem for $\boldsymbol{\tau}_z \in \mathrm{RT}_{z,g}$ on each element $K \in \mathcal{T}_z$:

$$\begin{cases} \mathrm{div}\,(\boldsymbol{\tau}_z - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}) = J_z, & \text{in } K, \\ (\boldsymbol{\tau}_z - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}|_K)\cdot\boldsymbol{n}_e = -s_K|_e\nu_{z,K,e}\,\Pi_e^{k-1}(\phi_z j_e), & \text{on } e \in \mathcal{E}_K. \end{cases} \tag{6.17}$$

It can be verified that the choice of J_z in (6.15) guarantees the solvability of each local problem in (6.17) and $\boldsymbol{\tau}_z \in \mathrm{RT}_{z,g}$.

According to [24, Lemma 3.1], there exists a $\boldsymbol{\tau}_z|_K \in \mathrm{RT}_{k-1}(K)$ such that the following stability estimate holds

$$\|\boldsymbol{\tau}_z - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}\|_K \leq C \left(h_K \|J_z\|_K + h_K^{1/2} \sum_{e \in \mathcal{E}_K} \nu_{z,K,e} \|\Pi_e^{k-1}(\phi_z j_e)\|_e \right). \tag{6.18}$$

Due to the facts that

$$\|\Pi_e^{k-1}(\phi_z j_e)\|_e \leq \|j_e\|_e \quad \text{and} \quad \nu_{z,K,e} \leq \frac{\sqrt{\alpha_K}}{\sqrt{\alpha_{e,M}}}, \tag{6.19}$$

by the Cauchy–Schwarz inequality, we have

$$h_K \alpha^{-1/2} \|J_z\|_K \leq \sum_{e \in \mathcal{E}_K} \frac{h_K^{1/2}}{\sqrt{\alpha_{e,M}}} \|j_e\|_e \leq C\, \eta_K. \tag{6.20}$$

Now, the inequality in (6.14) is a direct consequence of (6.18) and (6.19). This completes the proof of the proposition. □

Now we are in a position to state the local efficiency of ξ_z as well as ξ_K.

Theorem 6.1 *Assume that α is quasi-monotone. With η_K in (6.11), the following estimates hold true:*

$$\xi_z \leq C_1 \left(\sum_{K \in \mathcal{T}_z} \eta_K^2 \right)^{1/2} \leq C_2 \, \|\alpha^{1/2} \nabla (u - u_{\mathcal{T}})\|_{\hat{\omega}_z}, \tag{6.21}$$

$$\xi_K \leq \sum_{z \in \mathcal{N} \cap \partial K} \xi_z \leq \sum_{z \in \mathcal{N} \cap \partial K} C \|\alpha^{1/2} \nabla (u - u_{\mathcal{T}})\|_{\hat{\omega}_z}. \tag{6.22}$$

where $\hat{\omega}_z$ denotes the union of elements that share at least one edge ($d = 2$) or one face ($d = 3$) with an element in ω_z.

Proof Note that (6.22) is an immediate result of the triangle inequality and (6.21), so it suffices to show (6.21).

The second inequality in (6.21) is a direct consequence of the local efficiency bound of η_K in (6.13). To prove the first inequality in (6.21), let $\boldsymbol{\tau}_z \in \mathrm{RT}_{z,g}$ be the vector field in Proposition 6.1. The Cauchy–Schwarz inequality and (6.14) imply that

$$\begin{aligned}
\xi_z^2 &= \left(\alpha^{-1} (\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}), \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}} \right) \\
&= \left(\alpha^{-1} (\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}), \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \boldsymbol{\tau}_z \right) + \left(\alpha^{-1} (\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}), \boldsymbol{\tau}_z - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}} \right) \\
&\leq \left(\alpha^{-1} (\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}), \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \boldsymbol{\tau}_z \right) + C \|\alpha^{-1/2} (\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\| \left(\sum_{K \in \mathcal{T}_z} \eta_K^2 \right)^{1/2}.
\end{aligned}$$

Now it suffices to show that

$$b \equiv \left(\alpha^{-1}(\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}), \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \boldsymbol{\tau}_z\right) \le C\|\alpha^{-1/2}(\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\| \left(\sum_{K\in\mathcal{T}_z} \eta_K^2\right)^{1/2}. \tag{6.23}$$

To this end, it follows from the first equation in (6.5) with $\boldsymbol{\tau} = \hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \boldsymbol{\tau}_z \in \mathrm{RT}_{z,0}$, (6.14), the Cauchy–Schwarz and the triangle inequalities, (6.20), and Lemma 6.2 that

$$\begin{aligned} b &= \left(-\mathrm{div}\,(\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \boldsymbol{\tau}_z), u_z\right) = \sum_{K\in\mathcal{T}_z} \left(J_z - \mathrm{div}\,(\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}), u_z\right)_K \\ &= \sum_{K\in\mathcal{T}_z} (J_z - \phi_z(f - f_K), u_z)_K \le \sum_{K\in\mathcal{T}_z} \alpha_K^{-1/2} \left(\|J_z\|_K + \|f - f_K\|_K\right) \|\alpha^{1/2} u_z\|_K \\ &\le C \left(\sum_{K\in\mathcal{T}_z} \eta_K^2\right)^{1/2} \left(\sum_{K\in\mathcal{T}_z} h_K^{-2} \|\alpha^{1/2} u_z\|_K^2\right)^{1/2} \le C\|u_z\|_{\mathcal{T}_z} \left(\sum_{K\in\mathcal{T}_z} \eta_K^2\right)^{1/2}. \end{aligned}$$

By Lemma 6.1, the first equation in (6.5), and the Cauchy–Schwarz inequality, we have

$$\begin{aligned} \|u_z\|_{\mathcal{T}_z} &\le C \sup_{\boldsymbol{\tau}\in\mathrm{RT}_{z,0}} \frac{(\mathrm{div}\,\boldsymbol{\tau}, u_z)}{\|\alpha^{-1/2}\boldsymbol{\tau}\|} = C \sup_{\boldsymbol{\tau}\in\mathrm{RT}_{z,0}} \frac{(-\alpha^{-1}(\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}}), \boldsymbol{\tau})}{\|\alpha^{-1/2}\boldsymbol{\tau}\|} \\ &\le C\|\alpha^{-1/2}(\hat{\boldsymbol{\sigma}}_{z,\mathcal{T}} - \widetilde{\boldsymbol{\sigma}}_{z,\mathcal{T}})\|. \end{aligned}$$

Combining the above two inequalities gives (6.23). This proves the first inequality in (6.21) and, hence, the theorem. □

7 Numerical experiments

We consider solving the Kellogg's example [16] with parameters in [19]. Namely, $\Omega = (-1, 1)^2$ and the diffusion coefficient is $\alpha = 161.4476387975881$ in the first and third quadrants, $\alpha = 1$ in the second and fourth quadrants. For $f = 0$, a solution in the polar coordinates 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 with conforming P_1 and P_2 elements. RT_{k-1} flux recovery is used for P_k element for $k = 1$ and 2. The initial mesh consists of 4×4

Table 1 Kellogg's example: P_1 and P_2 discretizations

$u_{\mathcal{T}}$	ϵ_{rel}	DOFs	$\|A^{1/2}\nabla e\|/\|A^{1/2}\nabla u\|$	eff-ind
P_1	0.05	12410	4.9E-2	1.47
P_2	0.01	10237	9.9E-3	1.62

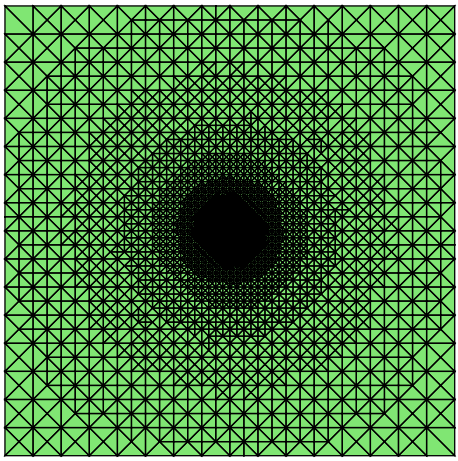


Fig. 1 P_1 element: mesh

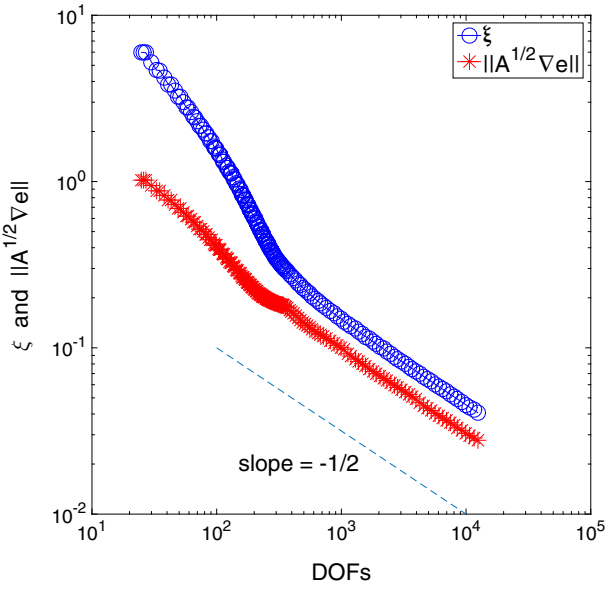
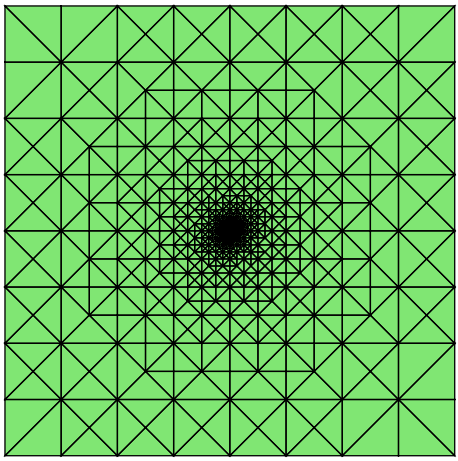


Fig. 2 P_1 element: error

congruent squares, each of which is partitioned into two triangles connecting bottom-left and top-right corners. We use Dörfler's marking strategy [14] with $\theta_D = 0.5$ as in [14,19]. That is, in the refinement of $\mathcal{T}$, a minimal subset $\hat{\mathcal{T}}$ of $\mathcal{T}$ is constructed such that

Fig. 3 P_2 element: mesh



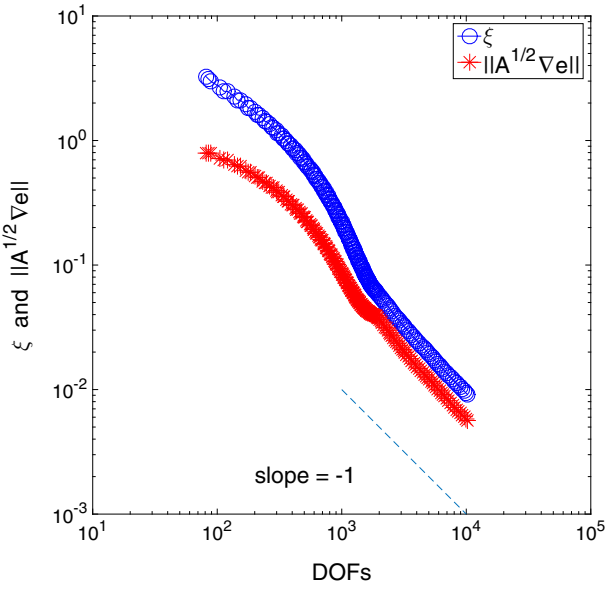


Fig. 4 P_2 element: error

$$\left(\sum_{K\in\hat{\mathcal{T}}}\xi_K^2\right)^{1/2} \geq \theta_D\left(\sum_{K\in\mathcal{T}}\xi_K^2\right)^{1/2}. \tag{7.1}$$

The newest-vertex bisection [23] is used in the refinement.

The following notation will be used:

- exact error $e := u - u_{\mathcal{T}}$;
- effectivity index: eff-ind;
- degrees of freedom: DOFs;

- stopping criterion: $\|A^{1/2}\nabla e\| \leq \epsilon_{\text{rel}}\|A^{1/2}\nabla u\|$ with

$$\epsilon_{\text{rel}} = \begin{cases} 0.05, & \text{for } P_1 \text{ element,} \\ 0.01, & \text{for } P_2 \text{ element.} \end{cases}$$

Numerical results are shown in Table 1 and in Figs. 1, 2, 3 and 4. First we notice from the plots in Figs. 2 and 4 that the estimator is always larger than the true error, confirming the guaranteed error control. It can be seen from Figs. 1 and 3 that the mesh refinement is homogeneous with respect to the singularity regardless of different scales of the diffusion coefficient in different quadrants, which implies the robustness of the estimator with respect to the coefficient jump. Optimal convergence rates are observed for both P_1 element in Fig. 2 and P_2 element in Fig. 4. Table 1 shows that the effectivity index is close to 1, so the estimator is considered accurate.

References

1. Ainsworth, M., Oden, J.T.: A Posteriori Error Estimation in Finite Element Analysis. Wiley, Wiley (2000)
2. Bank, R., Xu, J., Zheng, B.: Superconvergent derivative recovery for Lagrange triangular elements of degree p on unstructured grids. SIAM J. Numer. Anal. **45**(5), 2032–2046 (2007)
3. Bernardi, C., Verfürth, R.: Adaptive finite element methods for elliptic equations with non-smooth coefficients. Numer. Math. **85**(4), 579–608 (2000)
4. Boffi, D., Fortin, M., Brezzi, F.: Mixed Finite Element Methods and Applications. Springer Series in Computational Mathematics. Springer, Berlin (2013)
5. Braess, D.: Finite Elements: Theory, Fast Solvers, and Applications in Solid Mechanics. Cambridge University Press, Cambridge (2007)
6. Braess, D., Pillwein, V., Schöberl, J.: Equilibrated residual error estimates are p-robust. Comput. Methods Appl. Mech. Eng. **198**(13), 1189–1197 (2009)
7. Braess, D., Schöberl, J.: Equilibrated residual error estimator for edge elements. Math. Comput. **77**(262), 651–672 (2008)
8. Braess, D., Verfürth, R.: A posteriori error estimators for the Raviart–Thomas element. SIAM J. Numer. Anal. **33**(6), 2431–2444 (1996)
9. Cai, D., Cai, Z.: A hybrid a posteriori error estimator for conforming finite element approximations. Comput. Methods Appl. Mech. Eng. **339**, 320–340 (2018)
10. Cai, Z., Zhang, S.: Recovery-based error estimator for interface problems: conforming linear elements. SIAM J. Numer. Anal. **47**(3), 2132–2156 (2009)
11. Cai, Z., Zhang, S.: Flux recovery and a posteriori error estimators: conforming elements for scalar elliptic equations. SIAM J. Numer. Anal. **48**(2), 578–602 (2010)
12. Cai, Z., Zhang, S.: Robust equilibrated residual error estimator for diffusion problems: conforming elements. SIAM J. Numer. Anal. **50**(1), 151–170 (2012)
13. Cochez-Dhondt, S., Nicaise, S.: Equilibrated error estimators for discontinuous Galerkin methods. Numer. Methods Partial Differ. Equ. **24**(5), 1236–1252 (2008)
14. Dörfler, W.: A convergent adaptive algorithm for Poisson's equation. SIAM J. Numer. Anal. **33**(3), 1106–1124 (1996)
15. Ern, A., Vohralík, M.: Stable broken H^1 and $\mathbf{H}(\text{div})$ polynomial extensions for polynomial-degree-robust potential and flux reconstruction in three space dimensions. ArXiv e-prints (2017)
16. Kellogg, R.B.: On the Poisson equation with intersecting interfaces. Appl. Anal. **4**(2), 101–129 (1974)
17. Ladeveze, P., Leguillon, D.: Error estimate procedure in the finite element method and applications. SIAM J. Numer. Anal. **20**(3), 485–509 (1983)
18. Luce, R., Wohlmuth, B.: A local a posteriori error estimator based on equilibrated fluxes. SIAM J. Numer. Anal. **42**(4), 1394–1414 (2004)

19. Morin, P., Nochetto, R.H., Siebert, K.G.: Convergence of adaptive finite element methods. SIAM Rev. **44**(4), 631–658 (2002)
20. Naga, A., Zhang, Z.: The polynomial-preserving recovery for higher order finite element methods in 2D and 3D. Discrete Contin. Dyn. Syst. Ser. B **5**(3), 769 (2005)
21. Petzoldt, M.: A posteriori error estimators for elliptic equations with discontinuous coefficients. Adv. Comput. Math. **16**(1), 47–75 (2002)
22. Prager, W., Synge, J.L.: Approximations in elasticity based on the concept of function space. Q. Appl. Math. **5**, 241–269 (1947)
23. Sewell, E.G.: Automatic generation of triangulations for piecewise polynomial approximation. Ph.D. thesis, Purdue University, West Lafayette, IN (1972)
24. Verfürth, R.: A note on constant-free a posteriori error estimates. SIAM J. Numer. Anal. **47**(4), 3180–3194 (2009)
25. Verfürth, R.: A Posteriori Error Estimation Techniques for Finite Element Methods. Numerical Mathematics and Scientific Computation. OUP, Oxford (2013)
26. Ziegler, G.M.: Lectures on Polytopes, vol. 152. Springer, New York (2012)
27. Zienkiewicz, O.C., Zhu, J.Z.: A simple error estimator and adaptive procedure for practical engineerng analysis. Int. J. Numer. Methods Eng. **24**(2), 337–357 (1987)
28. Zienkiewicz, O.C., Zhu, J.Z.: The superconvergent patch recovery and a posteriori error estimates. Part 1. The recovery technique. Int. J. Numer. Methods Eng. **33**(7), 1331–1364 (1992)