Numer. Math. (2011) 117:679–716
DOI 10.1007/s00211-010-0324-5

Numerische
Mathematik

Decay rates of adaptive finite elements with Dörfler marking

Christian Kreuzer · Kunibert G. Siebert

Received: 11 November 2009 / Revised: 28 April 2010 / Published online: 17 July 2010

Abstract We investigate the decay rate for an adaptive finite element discretization of a second order linear, symmetric, elliptic PDE. We allow for any kind of estimator that is locally equivalent to the standard residual estimator. This includes in particular hierarchical estimators, estimators based on the solution of local problems, estimators based on local averaging, equilibrated residual estimators, the ZZ-estimator, etc. The adaptive method selects elements for refinement with Dörfler marking and performs a minimal refinement in that no interior node property is needed. Based on the local equivalence to the residual estimator we prove an error reduction property. In combination with minimal Dörfler marking this yields an optimal decay rate in terms of degrees of freedom.

Mathematics Subject Classification (2000) Primary 65N30 · 65N12 · 65N50 · 65N15

1 Introduction

We consider the approximation to second order, linear symmetric elliptic partial differential equations by adaptive finite elements with the standard adaptive loop

$$\mathsf{SOLVE} \quad \rightarrow \quad \mathsf{ESTIMATE} \quad \rightarrow \quad \mathsf{MARK} \quad \rightarrow \quad \mathsf{REFINE}. \tag{1.1}$$

C. Kreuzer (✉) · K. G. Siebert
Fakultät für Mathematik, Universität Duisburg-Essen,
Forsthausweg 2, 47057 Duisburg, Germany
e-mail: christian.kreuzer@uni-due.de
URL: http://www.numa.uni-due.de

K. G. Siebert
e-mail: kg.siebert@uni-due.de

This is: compute the Ritz-approximation in a H^1 conforming finite element space over the current triangulation, calculate an a posteriori error estimator, use Dörfler marking for the selection of elements to be refined, and refine the current grid into a new one. For sake of clarity we restrict ourselves to the Poisson problem, conforming refinement of simplicial grids by bisection, and continuous, piecewise affine finite elements.

After fixing the modules SOLVE, MARK, and REFINE we vary in the module ESTIMATE by taking different kinds of estimators into account. In particular we use the standard residual estimator as used in [14], variations of it [1,39,41], hierarchical estimators [5,35,38,39], an estimator based on local problems on stars [25], an equilibrated residual estimator [8,6], and the ZZ-estimator [42,12]. We prove that the standard iteration (1.1) with any of these estimators produces a sequence of discrete solutions that converges with an optimal rate in terms of degrees of freedom (DOFs) to the true solution.

Before embarking on the details we want to refer to the books [1,9,7,39] for the basic theory of adaptive finite elements and additional references. Plain convergence of the adaptive iteration (1.1) is by now completely understood, even for a larger problem class, different kind of estimators and marking strategies, and more general type of grids; compare with the results by Morin, Siebert, and Veeser [26] and Siebert [31]. We also would like to mention the overview article by Nochetto, Siebert, and Veeser summarizing the main aspects in the convergence and optimality analysis for adaptive finite elements [27]. Below we restrict ourselves to references intimately connected with this article.

One key ingredient in constructive approximation of some given function u is a discrete function U that is completely determined from local values. This means, changing U within a single element does not affect U in any other element of the grid. Another important ingredient is an upper bound for the local error on an element that reduces by a constant factor upon refining the element; compare with [4] and [27].

A fundamental problem in the optimality analysis of adaptive finite element methods is the fact that the Ritz approximation U is a *global* projection. Consequently, refinement of a single element affects the Ritz approximation everywhere. This in turn inhibits a completely local upper bound on single elements.

Despite this problem it turns out that a global error quantity that is strictly reduced by (1.1) can be used to prove an optimal decay rate. Based on the fundamental paper by Dörfler [18], Morin et al. showed that the energy error is such a contracting quantity when using Dörfler marking for both estimator and oscillation and asking for a sufficient refinement of marked elements and its direct neighbors [23,24]. This result was generalized in several directions, for instance in [13,22,25]. In particular important in the course of this article is the paper by Diening and Kreuzer dealing with the p-Laplacian [16]. Observing that oscillation is dominated by the estimator they showed that the sum of error and oscillation is strictly reduced by (1.1) *without* marking for oscillation.

Binev et al. were the first to prove an optimal decay rate for an adaptive finite element method [3]. Roughly speaking, they utilized the algorithm from [23] to improve the discrete solution and then added a coarsening step to regain the optimal rate. Stevenson realized that *minimal* Dörfler marking makes it possible to relate an optimal approximation to u with the current Ritz approximation [33]. Thanks to this important observation he was able to remove the coarsening step of [3]. In respect thereof, Dörfler marking is needed for ensuring an improvement of the discrete solution, and minimal Dörfler marking then in turn yields an optimal decay rate for the error.

Stevenson altered the standard loop (1.1) by separating marking for the estimator and oscillation using an inner loop for improving oscillation. In light of the discussion in [14, Sect. 6] it is clear that marking for two independent quantities simultaneously might lead to sub-optimal convergence rates.

Based on [16], Cascón et al. showed that marking for oscillation is not mandatory for deriving a contractive error quantity for (1.1) [14]. To be more precise: considering the standard residual estimator $\hat{\mathcal{E}}_{\mathcal{T}}$, they showed that Dörfler marking implies that a scaled sum of energy error and estimator is strictly reduced by (1.1). Up to oscillation this error notion is equivalent to the true error. Asking for minimal Dörfler marking they also proved an optimal decay rate for the standard iteration (1.1) by adapting the ideas of Stevenson to this error notion.

The contribution of this article is an optimality result for the standard adaptive loop (1.1) for different kind of estimators, in particular the ones mentioned above that are of a quite different nature. Following the ideas in [14], an optimal decay rate would be an immediate consequence from a contraction result for a scaled sum of energy error and estimator. However, it is not clear that for a general kind of estimator such a result holds true. Indeed, even monotonicity is not clear.

Philosophically spoken, the contraction result for the residual estimator $\hat{\mathcal{E}}_{\mathcal{T}}$ is a combination of the following two extreme cases. If oscillation is zero (or relatively small) the energy error strictly reduces as already used in [23,24,33]. The strict reduction of the energy error does even compensate for a possible non-monotone behavior of $\hat{\mathcal{E}}_{\mathcal{T}}$. Assume next that the discrete solution does not change and therefore the energy error stalls. In this situation oscillation is large and $\hat{\mathcal{E}}_{\mathcal{T}}$ strictly reduces thanks to a strict reduction of the weights of the indicators on refined elements. A suitable combination of these extreme cases gives the aforementioned contraction for a scaled sum of both quantities.

The behavior of the residual estimator in the latter case is related to the fact that it is a scaled L^2 norm of the residual, which is stronger than the H^{-1} norm. Therefore, $\hat{\mathcal{E}}_{\mathcal{T}}$ has a tendency to *overestimate* the true error, in particular when oscillation is large. Other estimators compute for instance an approximation to the H^{-1} norm of the residual by only evaluating it on an enlarged but still finite dimensional space. Consequently, theses estimators have the tendency to *underestimate* the error. In fact, such estimators only become reliable, if one adds oscillation to the upper bound.

The dilemma that the contraction result in [14] is a consequence of the potential overestimation of $\hat{\mathcal{E}}_{\mathcal{T}}$ and we are now dealing with estimators that potentially underestimate the error makes it evident that we should not aim at a reduction property in one *single* iteration.

We master this problem by utilizing a local equivalence of an estimator $\mathcal{E}_{\mathcal{T}}$ to the standard residual estimator $\hat{\mathcal{E}}_{\mathcal{T}}$. This equivalence has the following important consequence: The Dörfler marking for $\mathcal{E}_{\mathcal{T}}$ in (1.1) implies a Döfler property for $\hat{\mathcal{E}}_{\mathcal{T}}$. Consequently, [14] implies a contraction of a scaled sum of energy error and $\hat{\mathcal{E}}_{\mathcal{T}}$. Although $\hat{\mathcal{E}}_{\mathcal{T}}$ is not explicitly used in (1.1), this contracting quantity is then the key to start the optimality analysis.

Another approach would be to split the indicators into two parts, where one part is reducing upon refining an element and the other one can be handled by a discrete local lower bound; compare for instance with [16]. Following this approach one can show that a scaled sum of energy error and oscillation is strictly reduced after a *fixed* number of iterations. This approach needs a modification of the module MARK and is considered currently by Cascón and Nochetto. Utilizing the local equivalence of a given estimator $\mathcal{E}_{\mathcal{T}}$ to $\hat{\mathcal{E}}_{\mathcal{T}}$ we do not need to modify any of the modules in (1.1).

For the outline of the article and summary of the main result we next state the problem under consideration. Given a bounded polyhedral domain $\Omega \subset \mathbb{R}^d$ and a source term $f \in L^2(\Omega)$ we look for a weak solution u such that

$$-\Delta u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega. \tag{1.2}$$

In Sect. 2 we introduce the variational formulation of (1.2) and state the discrete problem. We then consider several estimators and list their basic properties. In Sect. 3 we state and prove the main results. We start with the basic assumptions on the adaptive iteration (1.1) and then define the appropriate approximation class $\mathbb{A}_s$, which involves the true solution u as well as data f. We then prove the following results for the sequence $\{\mathcal{T}_k, U_k, \mathcal{E}_k\}_{k\geq 0}$ of grids, discrete solutions, and estimators produced by (1.1).

Main Result 1 (Reduction Property) *There exist* $0 < \alpha < 1$ *and* $\Lambda_1 > 0$ *such that*

$$\left(|||U_k - u|||^2 + \mathcal{E}_k^2\right)^{1/2} \leq \Lambda_1\, \alpha^{k-\ell} \left(|||U_\ell - u|||^2 + \mathcal{E}_\ell^2\right)^{1/2} \quad \forall\, 0 \leq \ell \leq k.$$

Main Result 2 (Optimal Decay Rate) *There exists a constant* Λ_2 *such that if* $(u, f) \in \mathbb{A}_s$ *there holds*

$$\left(|||U_k - u|||^2 + \mathcal{E}_k^2\right)^{1/2} \leq \Lambda_2\, |u, f|_s\, (\#\mathcal{T}_k - \#\mathcal{T}_0)^{-s}.$$

Concerning the constants Λ_1 and Λ_2 it is important that the constants of the local equivalence to the standard residual estimator are of moderate size. This is true for the Poisson problem. In fact, the reliability and efficiency proofs for some estimators rely on this local equivalence. We conclude this article in Sect. 4 with an extension of the theory to singularly perturbed problems relying on robust estimators.

We finally would like to point out that the presented framework is neither restricted to the Poisson problem nor to linear finite elements. This restriction solely accounts for the clarity of the used arguments. For a generalization of the results we shortly list the important ingredients for the used arguments.

The results transfer to other elliptic problems that stem from a minimization of some energy and where the standard residual estimator $\hat{\mathcal{E}}_{\mathcal{T}}$ provides an upper and lower bound for the true error; compare for instance with [14] for a more general symmetric and linear problem and [16] for a nonlinear problem. In case of linear problems, the energy minimization implies a crucial orthogonal error decomposition that is heavily used. The indicators of $\hat{\mathcal{E}}_{\mathcal{T}}$ strictly reduce under refinement, if the discrete solution does not change. This is one key to construct a suitable contracting quantity for (1.1).

The proof of Main Result 1 additionally utilizes the local equivalence of some estimator $\mathcal{E}_\mathcal{T}$ to the standard residual estimator $\hat{\mathcal{E}}_\mathcal{T}$. Such an equivalence can be shown for a larger problem class and other discretizations. Dörfler marking then ensures the refinement of sufficiently many elements for the reduction property of Main Result 1.

The proof of Main Result 2 follows the ideas of [33] and relies besides the reduction property on a localized upper bound for the difference of two Ritz-projections. Such a bound can be deduced as consequence of the local equivalence of $\mathcal{E}_\mathcal{T}$ and $\hat{\mathcal{E}}_\mathcal{T}$ or may be derived directly for $\mathcal{E}_\mathcal{T}$. In combination with minimal Dörfler marking this implies that not too many elements are refined, which in turn yields the optimal decay rate of Main Result 2.

2 Problem, discretization, and error estimation

In this section we introduce the continuous and discrete problem. In order to avoid technical difficulties we consider the Poisson problem discretized by linear finite elements. Subsequently we introduce diverse kinds of estimators.

2.1 Continuous and discrete problem

For any measurable subset $\omega \subset \mathbb{R}^d$ with non-empty interior let $L^2(\omega)$ be the space of real square integrable Lebesgue functions over ω with scalar product $\langle \cdot, \cdot \rangle_\omega$. We denote by $H^1(\omega)$ the usual Sobolev space of functions in $L^2(\omega)$ whose first derivatives are also in $L^2(\omega)$, endowed with the norm

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

Finally, we let $\mathbb{V} := H^1_0(\Omega)$ be the space of functions in $H^1(\Omega)$ with vanishing trace on $\partial\Omega$.

Continuous Problem. The weak solution u of (1.2) is the unique solution of the variational problem

$$u \in \mathbb{V}: \qquad \mathcal{B}[u, v] = \langle f, v \rangle_\Omega \qquad \forall\, v \in \mathbb{V}, \tag{2.1}$$

where the bilinear $\mathcal{B}\colon \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ form is defined to be $\mathcal{B}[w, v] := \int_\Omega \nabla v \cdot \nabla w \, dx$.

The semi-norm $|\cdot|_{H^1(\Omega)} := \|\nabla \cdot\|_{L^2(\Omega)}$ is a norm on $\mathbb{V} = H^1_0(\Omega)$ that is equivalent to $\|\cdot\|_{H^1(\Omega)}$, thanks to the Poincaré–Friedrichs inequality [19]. Therefore, $\mathcal{B}$ is a scalar product on $\mathbb{V}$ inducing the energy norm $|\!|\!|\cdot|\!|\!| := \mathcal{B}[\cdot, \cdot]^{1/2} = |\cdot|_{H^1(\Omega)}$. The restriction of the energy norm to a subset $\omega \subset \Omega$ is denoted by $|\!|\!|v|\!|\!|_\omega := (\int_\omega |\nabla v|^2 \, dx)^{1/2}$. Existence and uniqueness of a weak solution $u \in \mathbb{V}$ to (2.1) is therefore a direct consequence of the Riesz Representation Theorem [19, Theorem 5.7].

Discrete Problem. We use linear finite elements for the discretization. To be more precise: Given a conforming triangulation $\mathcal{T}$ of Ω, which is built from closed simplexes, we let $\mathbb{V}(\mathcal{T})$ be the space of continuous, piecewise affine functions over $\mathcal{T}$ with vanishing trace on $\partial\Omega$. The discrete solution is the *Ritz-approximation* $U \in \mathbb{V}(\mathcal{T})$ of u, i.e., U is the unique solution of the discrete problem

$$U \in \mathbb{V}(\mathcal{T}): \qquad \mathcal{B}[U,\, V] = \langle f,\, V\rangle_\Omega \qquad \forall\, V \in \mathbb{V}(\mathcal{T}). \tag{2.2}$$

Note, that we hereby assume exact integration and exact linear algebra.

2.2 Error estimation

In general, a posteriori error estimation aims at deriving a computable bound for a negative norm of the residual $\mathrm{Res}(U) \in H^{-1}(\Omega)$ defined as

$$\langle \mathrm{Res}(U),\, v\rangle := \mathcal{B}[U,\, v] - \langle f,\, v\rangle_\Omega = \mathcal{B}[U-u,\, v] \qquad \forall\, v \in H_0^1(\Omega).$$

In our setting, the norm of $\mathrm{Res}(U)$ is induced by the energy norm

$$|||\mathrm{Res}(U)|||_* := \sup_{v\in\mathbb{V}} \frac{\langle \mathrm{Res}(U),\, v\rangle}{|||v|||}.$$

In the derivation of such bounds, the normal flux of ∇U across inter-element sides plays an important role. Denoting by Σ the skeleton of $\mathcal{T}$, i.e., the union of all sides in $\mathcal{T}$, we define the jump residual $J(U) \in L^2(\Sigma)$ as follows. For an interior side $\sigma = T_1 \cap T_2$ we let $J(U)$ be the normal flux of ∇U, i.e.,

$$J(U)_{|\sigma} := [\![\nabla U]\!]_{|\sigma} = \left(\nabla U_{|T_1} \cdot \boldsymbol{n}_{T_1} + \nabla U_{|T_2} \cdot \boldsymbol{n}_{T_2}\right)_{|\sigma},$$

where $\boldsymbol{n}_T$ is the outer normal of T. On a boundary side $\sigma \subset \partial\Omega$ we set $J(U) = 0$. Note, that $J(U)$ is piecewise constant over Σ. The definition of the jump residual and Green's formula then imply

$$\langle \mathrm{Res}(U),\, v\rangle = \int_\Sigma J(U) v \, do - \langle f,\, v\rangle_\Omega.$$

We next introduce the standard residual estimator $\hat{\mathcal{E}}_\mathcal{T}$ and then consider several error estimators $\mathcal{E}_\mathcal{T}$ that are equivalent to $\hat{\mathcal{E}}_\mathcal{T}$. In the context of convergence and, especially, optimality analysis of adaptive methods, the specific choice of the weights in the definition of $\hat{\mathcal{E}}_\mathcal{T}$ is one of the key ingredients. The equivalence of another estimator $\mathcal{E}_\mathcal{T}$ to $\hat{\mathcal{E}}_\mathcal{T}$ substitutes for this as we shall see in Sect. 3.

Any explicit constant C or implicit constant hidden in '$\lesssim$','$\gtrsim$', or '$\approx$' only depends on the shape regularity of the grid $\mathcal{T}$, Ω, and the dimension d.

2.2.1 Standard residual estimator

A paramount role in this paper plays the following definition of the residual estimator that we denote by $\hat{\mathcal{E}}_\mathcal{T}$. Denoting by $h_\mathcal{T}\colon \Omega \to \mathbb{R}^+$ the piecewise constant mesh size function defined as

$$h_{\mathcal{T}}|_T = |T|^{1/d} \qquad \forall\, T \in \mathcal{T},$$

the element indicators of the standard residual estimator are given by

$$\hat{\mathcal{E}}^2_{\mathcal{T}}(U,T) := \|h_{\mathcal{T}} f\|^2_{L^2(T)} + \|h_{\mathcal{T}}^{1/2} J(U)\|^2_{L^2(\partial T)} \qquad \forall\, T \in \mathcal{T}. \tag{2.3}$$

The set $\{\hat{\mathcal{E}}^2_{\mathcal{T}}(U,T)\}_{T\in\mathcal{T}}$ of indicators builds up the estimator $\hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T})$, which is reliable and efficient, i.e.,

$$|\!|\!| U-u |\!|\!|^2 \le \hat{C}_1 \hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}) := \hat{C}_1 \sum_{T\in\mathcal{T}} \hat{\mathcal{E}}^2_{\mathcal{T}}(U,T), \tag{2.4a}$$

$$\hat{C}_2 \hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}) \le |\!|\!| U-u |\!|\!|^2 + \widehat{\operatorname{osc}}^2_{\mathcal{T}}(f,\mathcal{T}) := |\!|\!| U-u |\!|\!|^2 + \sum_{T\in\mathcal{T}} \widehat{\operatorname{osc}}^2_{\mathcal{T}}(f,T). \tag{2.4b}$$

Local *data oscillation* on T is defined with the mean value $f_T := \frac{1}{|T|}\int_T f\,dx$ of f as

$$\widehat{\operatorname{osc}}_{\mathcal{T}}(f,T) := \|h_{\mathcal{T}}(f - f_T)\|_{L^2(T)}.$$

Generically, data oscillation is of higher order.

2.2.2 *Variants of the residual estimator*

In this section we briefly introduce two variants of the residual estimator introduced above. These variants are not directly included in [14].

A common definition of the indicators for the residual estimator is

$$\mathcal{E}^2_{\mathcal{T}}(U,T) := h_T^2 \|f\|^2_{L^2(T)} + \sum_{\sigma\subset\partial T} h_\sigma \|J(U)\|^2_{L^2(\sigma)} \qquad \forall\, T \in \mathcal{T}, \tag{2.5}$$

where we have some scope of choosing the weights h_T and h_σ. A typical choice is $h_T := \operatorname{diam}(T)$ and $h_\sigma := \operatorname{diam}(\sigma)$, see for instance [23,39]. Any choice of h_T and h_σ such that

$$h_{\mathcal{T}|T} \lesssim h_T \lesssim h_{\mathcal{T}|T} \quad \text{and} \quad h_{\mathcal{T}|T} \lesssim h_\sigma \lesssim h_{\mathcal{T}|T} \qquad \forall\, T \in \mathcal{T}$$

results in the obvious element by element equivalence

$$\mathcal{E}^2_{\mathcal{T}}(U,T) \lesssim \hat{\mathcal{E}}^2_{\mathcal{T}}(U,T) \lesssim \mathcal{E}^2_{\mathcal{T}}(U,T) \qquad \forall\, T \in \mathcal{T}. \tag{2.6}$$

The equivalence of the indicators readily implies

$$|\!|\!| U-u |\!|\!|^2 \lesssim \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{T}) := \sum_{T\in\mathcal{T}} \mathcal{E}^2_{\mathcal{T}}(U,T), \tag{2.7a}$$

$$\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{T}) \lesssim |\!|\!| U-u |\!|\!|^2 + \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{T}), \tag{2.7b}$$

where oscillation is defined as above, i.e., $\operatorname{osc}_{\mathcal{T}} = \widehat{\operatorname{osc}}_{\mathcal{T}}$.

Veeser and Verfürth have given a definition of the residual estimator with explicit constants that is organized by stars [41]. Denote by $\mathcal{N} = \mathcal{N}(\mathcal{T})$ the set of all vertices of $\mathcal{T}$. The finite element star ω_z at $z \in \mathcal{N}$ is the support of the hat function ϕ_z at z, which is

$$\omega_z := \omega(\mathcal{T}, z) = \bigcup\{T \in \mathcal{T} : z \in T\}.$$

Note that $\sum_{z\in\mathcal{N}} \phi_z \equiv 1$ is a partition of unity and that the set of hat functions at interior nodes $z \in \mathring{\mathcal{N}} := \mathcal{N} \cap \Omega$ is a basis of $\mathbb{V}(\mathcal{T})$.

The indicators are not given element-wise but are indexed by the vertices and defined as follows:

$$\mathcal{E}^2_{\mathcal{T}}(U, z) := c(\omega_z)\, h_z^2 \|f\|^2_{L^2(\omega_z)} + c(\Sigma_z) \sum_{\sigma\subset\Sigma_z} \frac{h_z^2}{h_\sigma^\perp} \|J(U)\|^2_{L^2(\sigma)} \qquad \forall\, z \in \mathcal{N}, \tag{2.8}$$

where Σ_z is the union of all sides that have z as a common vertex. The derivation of the estimator relies on a precise tracking of constants from local Poincaré and trace inequalities. These constants scale as stated in (2.8) with the local mesh sizes

$$h_z := \operatorname{diam}(\omega_z) \quad \text{and} \quad h_\sigma^\perp := \frac{|\omega_\sigma|}{|\sigma|},$$

where for $\sigma \in \mathcal{S}$ with the two adjacent elements $T_1, T_2 \in \mathcal{T}$ such that $\sigma = T_1 \cap T_2$ we set $\omega_\sigma = \omega_\sigma(\mathcal{T}, \sigma) = T_1 \cup T_2$. Shape regularity of $\mathcal{T}$ readily implies

$$h_{\mathcal{T}|T} \le h_z \lesssim h_{\mathcal{T}|T} \quad \text{and} \quad h_{\mathcal{T}|T} \lesssim h_z^2/h_\sigma^\perp \lesssim h_{\mathcal{T}|T} \qquad \forall\, T \subset \omega_z. \tag{2.9}$$

The additional constants $c(\omega_z)$ and $c(\Sigma_z)$ are related to the shape of the star and are uniformly bounded [41, Sect. 5].

The estimator satisfies the upper

$$|\!|\!| U - u |\!|\!|^2 \lesssim \mathcal{E}^2_{\mathcal{T}}(U, \mathcal{N}) := \sum_{z\in\mathcal{N}} \mathcal{E}^2_{\mathcal{T}}(U, z). \tag{2.10a}$$

Standard techniques, in combination with the uniform estimates for $c(\omega_z)$ and $c(\Sigma_z)$ give the lower bound

$$\mathcal{E}^2_{\mathcal{T}}(U, \mathcal{N}) \lesssim |\!|\!| U - u |\!|\!|^2 + \operatorname{osc}^2_{\mathcal{T}}(f, \mathcal{N}) := |\!|\!| U - u |\!|\!|^2 + \sum_{z\in\mathcal{N}} \operatorname{osc}^2_{\mathcal{T}}(f, z). \tag{2.10b}$$

Hereby, local data oscillation on the star ω_z is defined as

$$\operatorname{osc}^2_{\mathcal{T}}(f, z) := c(\omega_z)\, h_z^2 \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)},$$

where $f_{z|T} = \frac{1}{|T|}\int_T f\phi_z\, dx$ is the weighted mean value of f in $T \in \mathcal{T}$.

We next turn to the local equivalence to $\hat{\mathcal{E}}_{\mathcal{T}}$. For $\mathcal{N}' \subset \mathcal{N}$ we have

$$\mathcal{E}_{\mathcal{T}}^2(U, \mathcal{N}') \lesssim \sum_{T \subset \mathcal{T}(\mathcal{N}')} \hat{\mathcal{E}}_{\mathcal{T}}^2(U, T) =: \hat{\mathcal{E}}_{\mathcal{T}}^2\left(U, \mathcal{T}(\mathcal{N}')\right) \tag{2.11a}$$

with $\mathcal{T}(\mathcal{N}') := \{T \in \mathcal{T} \mid T \subset \omega_z \text{ for some } z \in \mathcal{N}'\}$. On the other hand, any indicator $\hat{\mathcal{E}}_{\mathcal{T}}(U, T)$ is controlled by the sum of the indicators $\mathcal{E}_{\mathcal{T}}(U, z)$ with $z \in T$. Moreover, for $\mathcal{T}' \subset \mathcal{T}$ we have

$$\hat{\mathcal{E}}_{\mathcal{T}}^2(U, \mathcal{T}') \lesssim \sum_{z \in \mathcal{N}(\mathcal{T}')} \mathcal{E}_{\mathcal{T}}^2(U, z) =: \mathcal{E}_{\mathcal{T}}^2\left(U, \mathcal{N}(\mathcal{T}')\right), \tag{2.11b}$$

where $\mathcal{N}(\mathcal{T}') := \{z \in \mathcal{N} \mid z \in T \text{ for some } T \in \mathcal{T}'\}$. In summary, both variants of the residual estimator are locally equivalent to $\hat{\mathcal{E}}_{\mathcal{T}}$.

2.2.3 Hierarchical estimators

The idea of hierarchical estimators is based upon evaluating the residual with sufficiently many discrete functions that do not belong to $\mathbb{V}(\mathcal{T})$. Suitable functions are side and element bubble functions that are either higher order finite elements on the same grid or linear finite elements on a refined mesh. Most results of this section can, e.g., be found in [35,38,39]; compare also with [1,5,17].

We use a variant of the hierarchical estimator indexed by the *interior* sides of $\mathcal{T}$, which we denote by $\mathcal{S}$. For the precise definition of the side and element bubble functions we introduce the following spaces. We let $\mathbb{V}_p(\mathcal{T}) \subset \mathbb{V}$ be continuous piecewise polynomials of degree $p \in \mathbb{N}$ over $\mathcal{T}$ with vanishing trace on $\partial\Omega$. Furthermore, we denote by $\mathcal{T}_+$ the conforming refinement of $\mathcal{T}$ such that for any $\sigma \in \mathcal{S}$ and $T \in \mathcal{T}$ there exist $z_\sigma, z_T \in \mathcal{N}(\mathcal{T}_+) \setminus \mathcal{N}(\mathcal{T})$ that belong to the interior of σ respectively T. Recall the notation ω_σ, $\sigma \in \mathcal{S}$ for the union of adjacent elements given in Sect. 2.2.2. We then let either $\phi_\sigma \in \mathbb{V}_d(\mathcal{T})$ or $\phi_\sigma \in \mathbb{V}(\mathcal{T}_+)$ be the unique function satisfying

$$\phi_\sigma(z_\sigma) > 0, \qquad \operatorname{supp}(\phi_\sigma) \subset \omega_\sigma, \quad \text{and} \quad |||\phi_\sigma||| = 1.$$

For $T \in \mathcal{T}$ we either select $\phi_T \in \mathbb{V}_{d+1}(\mathcal{T})$ or $\phi_T \in \mathbb{V}(\mathcal{T}_+)$ uniquely determined by

$$\phi_T(z_T) > 0, \qquad \operatorname{supp}(\phi_T) \subset T, \quad \text{and} \quad |||\phi_T||| = 1.$$

For $\sigma \in \mathcal{S}$ the associated indicator is then given as

$$\mathcal{E}_{\mathcal{T}}^2(U, \sigma) := \langle \operatorname{Res}(U), \phi_\sigma \rangle^2 + \sum_{T \subset \omega_\sigma} \left(\langle \operatorname{Res}(U), \phi_T \rangle^2 + \|h_{\mathcal{T}}(f - f_T)\|_{L^2(T)}^2 \right).$$

Assuming $\mathcal{S} \neq \emptyset$, the estimator is reliable and efficient, i.e.,

$$|||U-u|||^2 \lesssim \sum_{\sigma\in\mathcal{S}} \mathcal{E}^2_{\mathcal{T}}(U,\sigma) =: \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{S}) \tag{2.12a}$$

$$\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{S}) \lesssim |||U-u|||^2 + \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{S}) =: |||U-u|||^2 + \sum_{\sigma\in\mathcal{S}} \operatorname{osc}^2_{\mathcal{T}}(f,\sigma), \tag{2.12b}$$

where local oscillation on ω_σ is defined as

$$\operatorname{osc}^2_{\mathcal{T}}(f,\sigma) := \sum_{T\subset\omega_\sigma} \widehat{\operatorname{osc}}^2_{\mathcal{T}}(f,T),$$

compare with [38,39]. Note, that the assumption $\mathcal{S} \neq \emptyset$ just rules out trivial cases like $\#\mathcal{T} = 1$.

Recalling $|||\phi_T||| = 1$, the local Friedrich's inequality $\|\phi_T\|_{L^2(T)} \le h_T := \operatorname{diam}(T)$ directly implies

$$|\langle \operatorname{Res}(U),\, \phi_T\rangle| = \left|\int_T f\,\phi_T\,dx\right| \le \|h_T f\|_{L^2(T)}.$$

In this vein, $\|\phi_\sigma\|_{L^2(\omega_\sigma)} \le h_\sigma := \operatorname{diam}(\omega_\sigma)$ and in combination with a scaled trace inequality we bound

$$|\langle \operatorname{Res}(U),\, \phi_\sigma\rangle| = \left|\int_\sigma J(U)\,\phi_\sigma\,do - \int_{\omega_\sigma} f\,\phi_\sigma\,dx\right| \lesssim \|h_\sigma^{1/2} J(U)\|_{L^2(\sigma)} + \|h_\sigma f\|_{L^2(\omega_\sigma)}.$$

The local mesh sizes h_T and h_σ are locally equivalent to $h_{\mathcal{T}}$ in ω_σ. Consequently, thanks to the finite overlap of patches ω_σ, $\sigma \in \mathcal{S}$, we have

$$\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{S}') \lesssim \sum_{T\in\mathcal{T}(\mathcal{S}')} \hat{\mathcal{E}}^2_{\mathcal{T}}(U,T) = \hat{\mathcal{E}}^2_{\mathcal{T}}\left(U,\mathcal{T}(\mathcal{S}')\right), \tag{2.13a}$$

for all $\mathcal{S}' \subset \mathcal{S}$ with $\mathcal{T}(\mathcal{S}') := \{T \in \mathcal{T} \mid T \subset \omega_\sigma \text{ for some } \sigma \in \mathcal{S}'\}$. Note, that we also have used the fact that oscillation is dominated by the residual estimator, i.e., $\operatorname{osc}_{\mathcal{T}}(f,\sigma) \le \hat{\mathcal{E}}_{\mathcal{T}}(U,\mathcal{T}(\{\sigma\}))$.

Verfürth has shown in [39, (1.52) & (1.57)] the estimates

$$\begin{aligned}
\|h_\sigma^{1/2} J(U)\|^2_{L^2(\sigma)} &\lesssim \langle \operatorname{Res}(U),\, \phi_\sigma\rangle^2 + \|h_{\mathcal{T}} f\|^2_{L^2(\omega_\sigma)},\\
\|h_{\mathcal{T}} f\|^2_{L^2(T)} &\lesssim \langle \operatorname{Res}(U),\, \phi_T\rangle^2 + \|h_{\mathcal{T}}(f - f_T)\|^2_{L^2(T)}.
\end{aligned}$$

Therefore, the indicator $\hat{\mathcal{E}}_{\mathcal{T}}(U, T)$ on T is controlled by the hierarchical indicators at its interior sides. Hence, for $\mathcal{S}(\mathcal{T}') := \{\sigma \in \mathcal{S} \mid \sigma \subset T \text{ for some } T \in \mathcal{T}'\}$, $\mathcal{T}' \subset \mathcal{T}$:

$$\hat{\mathcal{E}}^2_{\mathcal{T}}(U, \mathcal{T}') \lesssim \sum_{\sigma \subset T \cap \Omega} \mathcal{E}^2_{\mathcal{T}}(U, \sigma) := \mathcal{E}^2_{\mathcal{T}}\left(U, \mathcal{S}(\mathcal{T}')\right), \tag{2.13b}$$

which proves local equivalence of the residual estimator $\hat{\mathcal{E}}_{\mathcal{T}}$ and the hierarchical estimator.

2.2.4 Estimators based on local problems

Morin, Nochetto, and Siebert have introduced for two space dimension an estimator that is based on solving local problems on stars [25]. For $z \in \mathcal{N}$, the finite element star ω_z, its skeleton Σ_z, and the hat-function ϕ_z are already introduced in Sect. 2.2.2.

The local problems are solved in a local function space $\mathbb{W}_z$ consisting of continuous piecewise quadratic finite elements inside the star ω_z that have vanishing trace on $\partial\omega_z$. For an interior node $z \in \mathring{\mathcal{N}} = \mathcal{N} \cap \Omega$ the elements $\psi \in \mathbb{W}_z$ are additionally required to satisfy $\int_{\omega_z} \psi\, \phi_z\, dx = 0$.

The star estimator is then defined as follows. For each vertex $z \in \mathcal{N}$ solve the linear problem

$$\eta_z \in \mathbb{W}_z : \quad \int_{\omega_z} \nabla\eta_z \cdot \nabla\psi\, \phi_z\, dx = \int_{\Sigma_z} J(U)\psi\phi_z\, do - \int_{\omega_z} f\, \psi\phi_z\, dx \quad \forall\, \psi \in \mathbb{W}_z$$

and set

$$\mathcal{E}^2_{\mathcal{T}}(U, z) := \|\nabla\eta_z\phi_z^{1/2}\|^2_{L^2(\omega_z)} + h_z^2\|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)}.$$

Hereafter, $f_z = \int_{\omega_z} f\, \phi_z\, dx / \int_{\omega_z} \phi_z\, dx$ is for $z \in \mathring{\mathcal{N}}$ a weighted mean value of f over ω_z and $f_z = 0$ for a boundary vertex $z \in \mathcal{N} \cap \partial\Omega$. As above, the local mesh size is defined to be $h_z = \operatorname{diam}(\omega_z)$.

In [25] it has been shown that this estimator is reliable and efficient, i.e.,

$$|\!|\!|U - u|\!|\!|^2 \lesssim \sum_{z \in \mathcal{N}} \mathcal{E}^2_{\mathcal{T}}(U, z) =: \mathcal{E}^2_{\mathcal{T}}(U, \mathcal{N}) \tag{2.14a}$$

$$\mathcal{E}^2_{\mathcal{T}}(U, \mathcal{N}) \lesssim |\!|\!|U - u|\!|\!|^2 + \operatorname{osc}^2_{\mathcal{T}}(f, \mathcal{N}) := |\!|\!|U - u|\!|\!|^2 + \sum_{z \in \mathcal{N}} \operatorname{osc}^2_{\mathcal{T}}(f, z), \tag{2.14b}$$

where $\operatorname{osc}^2_{\mathcal{T}}(f, z) := \|h_z(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)}$.

Similar arguments as for the hierarchical estimator yield

$$\mathcal{E}^2_{\mathcal{T}}(U, z) \lesssim h_z^2\|f\phi_z^{1/2}\|^2_{L^2(\omega_z)} + h_z\|J(U)\phi_z^{1/2}\|^2_{L^2(\Sigma_z)},$$

whence, recalling that $\{\phi_z\}_{z\in\mathcal{N}}$ is a partition of unity, we arrive at

$$\mathcal{E}_{\mathcal{T}}^2(U,\mathcal{N}') \lesssim \sum_{T\in\mathcal{T}(\mathcal{N}')} \hat{\mathcal{E}}_{\mathcal{T}}^2(U,T) := \hat{\mathcal{E}}_{\mathcal{T}}^2\left(U,\mathcal{T}(\mathcal{N}')\right), \quad \mathcal{N}'\subset\mathcal{N}, \qquad (2.15a)$$

with $\mathcal{T}(\mathcal{N}')$ as in Sect. 2.2.2.

To bound the residual estimator in terms of the star estimator we first construct suitable test functions. For $\sigma\in\mathcal{S}$ let $\psi_\sigma\in\mathbb{V}_2(\mathcal{T})$ be the piecewise quadratic edge bubble function with $\operatorname{supp}(\psi_\sigma)=\omega_\sigma$; compare with Sect. 2.2.3. For $z\in\mathcal{N}$ let $\psi_z\in\mathbb{V}_2(\mathcal{T})$ be the piecewise quadratic Lagrange basis function associated with z. Therefore, $\operatorname{supp}(\psi_z)=\omega_z$, $\psi_z(z)=1$, and ψ_z equals zero at the midpoints of all edges $\sigma\subset\Sigma_z$.

Let $\sigma\in\mathcal{S}$ be arbitrarily chosen and fix any of its vertices $z\in\mathcal{N}\cap\sigma$. Following ideas for the construction of an interpolation operator into $\mathbb{W}_z$ in the proof of [25, Lemma 3.7] we set

$$\alpha_z=\tfrac{4}{d}\tfrac{|\omega_\sigma|}{|\omega_z|},\qquad \alpha_\sigma=\tfrac{3}{d}-\tfrac{\alpha_z}{2},\qquad \text{and}\quad \alpha_{\sigma'}=-\tfrac{\alpha_z}{2}\quad \forall\sigma'\subset\Sigma_z,\ \sigma'\neq\sigma$$

and define

$$\psi_\sigma^z := \begin{cases} \displaystyle\sum_{\sigma'\subset\Sigma_z}\alpha_{\sigma'}\psi_{\sigma'}+\alpha_z\psi_z, & z\in\mathring{\mathcal{N}},\\ \displaystyle\psi_\sigma\Big/\int_\sigma\psi_\sigma\phi_z\,do, & \text{otherwise.}\end{cases}$$

The construction of the test function ψ_σ^z implies the following properties:

$$\psi_\sigma^z\in\mathbb{W}_z,\qquad \int_\sigma\psi_\sigma^z\phi_z\,do=\tfrac{1}{d}|\sigma|,\qquad \int_{\sigma'}\psi_\sigma^z\phi_z\,do=0,\quad \forall\sigma'\subset\Sigma_z,\ \sigma'\neq\sigma,$$

and

$$\|\nabla\psi_\sigma^z\phi_z^{1/2}\|_{L^2(\omega_z)}\lesssim 1;$$

see the proof of [25, Lemma 3.7]. Using the short form $J_\sigma=J(U)_{|\sigma}\in\mathbb{R}$ we obtain by definition of the solution η_z of the local problem and the definition of ψ_σ^z

$$\begin{aligned}\int_\sigma J^2(U)\phi_z\,do &= J_\sigma^2\int_\sigma\phi_z\,do = J_\sigma^2\int_\sigma\psi_\sigma^z\phi_z\,do=\int_{\Sigma_z}J(U)\left(J_\sigma\psi_\sigma^z\right)\phi_z\,do\\ &=\int_{\omega_z}\nabla\eta_z\cdot\left(J_\sigma\nabla\psi_\sigma^z\right)\phi_z\,dx-\int_{\omega_z}f\left(J_\sigma\psi_\sigma^z\right)\phi_z\,dx,\end{aligned}$$

since $J_\sigma\psi_\sigma^z\in\mathbb{W}_z$. Recalling the definition of f_z we conclude from $J_\sigma\psi_\sigma^z\in\mathbb{W}_z$ the identity $\int_{\omega_z}f\left(J_\sigma\psi_\sigma^z\right)\phi_z\,dx=\int_{\omega_z}(f-f_z)\left(J_\sigma\psi_\sigma^z\right)\phi_z\,dx$. The Cauchy–Schwartz inequality therefore implies

$$\int\limits_{\sigma} J_\sigma^2 \phi_z \, do \le \left(\|\nabla \eta_z \phi_z^{1/2}\|^2_{L^2(\omega_z)} + h_z^2 \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)} \right)^{1/2} \\ \times \left(\|J_\sigma \nabla \psi_\sigma^z \, \phi_z^{1/2}\|^2_{L^2(\omega_z)} + h_z^{-2} \|J_\sigma \psi_\sigma^z \, \phi_z^{1/2}\|^2_{L^2(\omega_z)} \right)^{1/2}.$$

The equivalence of norms on $\mathbb{W}_z$ together with standard scaling arguments yields

$$\|J_\sigma \nabla \psi_\sigma^z \, \phi_z^{1/2}\|_{L^2(\omega_z)} \lesssim h_z^{-1/2} \|J_\sigma \phi_z^{1/2}\|_{L^2(\sigma)}, \\ \|J_\sigma \psi_\sigma^z \, \phi_z^{1/2}\|_{L^2(\omega_z)} \lesssim h_z^{1/2} \|J_\sigma \phi_z^{1/2}\|_{L^2(\sigma)}.$$

Summarizing, for any $\mathcal{T}' \subset \mathcal{T}$ we have deduced

$$\begin{aligned} \sum_{T \in \mathcal{T}'} \|h_{\mathcal{T}}^{1/2} J(U)\|^2_{L^2(\partial T)} &\lesssim \sum_{z \in \mathcal{N}(\mathcal{T}')} \|h_z^{1/2} J(U) \phi_z^{1/2}\|^2_{L^2(\Sigma_z)} \\ &\lesssim \sum_{z \in \mathcal{N}(\mathcal{T}')} \|\nabla \eta_z \phi_z^{1/2}\|^2_{L^2(\omega_z)} + h_z^2 \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)} \\ &= \mathcal{E}^2_{\mathcal{T}}(U, \mathcal{N}(\mathcal{T}')), \end{aligned}$$

where we used the definition of $\mathcal{N}(\mathcal{T}')$ from Sect. 2.2.2.

We next turn to $\|h_{\mathcal{T}} f\|_{L^2(T)}$ where we use the well-known fact, that the element residual is dominated by the jump residual plus oscillation on the star. To be more precise: for any interior node $z \in \mathring{\mathcal{N}}$ we deduce from the discrete problem (2.2) for the hat-function $\phi_z \in \mathbb{V}(\mathcal{T})$

$$\int\limits_{\omega_z} f \phi_z \, dx = \mathcal{B}[U, \, \phi_z] = \int\limits_{\Sigma_z} J(U) \phi_z \, do.$$

Since $f_z \in \mathbb{R}$ we therefore infer

$$\begin{aligned} \int\limits_{\omega_z} |f_z|^2 \phi_z \, dx &= \int\limits_{\omega_z} f f_z \phi_z \, dx + \int\limits_{\omega_z} (f_z - f) f_z \phi_z \, dx \\ &= \int\limits_{\Sigma_z} J(U) f_z \phi_z \, do + \int\limits_{\omega_z} (f_z - f) f_z \phi_z \, dx \\ &\lesssim \left(h_z^{-1} \|J(U) \phi_z^{1/2}\|^2_{L^2(\Sigma_z)} + \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)^2} \right)^{1/2} \|f_z\|_{L^2(\omega_z)}, \end{aligned}$$

using $h_z \| f_z \|^2_{L^2(\Sigma_z)} \lesssim \| f_z \|^2_{L^2(\omega_z)}$. Recalling that $f_z = 0$ for boundary nodes, which implies

$$\begin{aligned}\sum_{T\in\mathcal{T}'} \|h_{\mathcal{T}} f\|^2_{L^2(T)} &\le \sum_{z\in\mathcal{N}(\mathcal{T}')} \|h_{\mathcal{T}} f \phi_z^{1/2}\|^2_{L^2(T)} \\ &\le \sum_{z\in\mathcal{N}(\mathcal{T}')} 2h_z^2 \| f_z \phi_z^{1/2}\|^2_{L^2(\omega_z)} + 2h_z^2 \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)} \\ &\lesssim h_z \| J(U)\phi_z^{1/2}\|^2_{\Sigma_z} + h_z^2 \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)}\end{aligned}$$

for all $\mathcal{T}' \subset \mathcal{T}$. Combining this with the bound for the jump residual we see

$$\begin{aligned}\hat{\mathcal{E}}^2_{\mathcal{T}}(U, \mathcal{T}') &\lesssim \sum_{z\in\mathcal{N}(\mathcal{T}')} h_z \| J(U)\|^2_{\Sigma_z} + h_z^2 \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)} \\ &\lesssim \sum_{z\in\mathcal{N}(\mathcal{T}')} \mathcal{E}^2_{\mathcal{T}}(U, z) =: \mathcal{E}^2_{\mathcal{T}}(U, \mathcal{N}(\mathcal{T}')), \qquad \mathcal{T}' \subset \mathcal{T}. \qquad (2.15b)\end{aligned}$$

In summary, we have shown that the residual estimator $\hat{\mathcal{E}}_{\mathcal{T}}$ and the star estimator are locally equivalent.

2.2.5 Equilibrated residual estimators

In this section we analyze error estimators motivated by the fundamental theorem of Prager and Synge: For any $\xi \in H(\mathrm{div}; \Omega)$ with $\mathrm{div}\, \xi + f = 0$ it holds

$$\|\nabla(v - u)\|^2_{L^2(\Omega)} + \|\xi - \nabla u\|^2_{L^2(\Omega)} = \|\nabla v - \xi\|^2_{L^2(\Omega)} \qquad \forall v \in \mathbb{V};$$

compare with [8,30]. Assuming that for the Ritz projection $v = U \in \mathbb{V}(\mathcal{T})$ we can *compute* a suitable function ξ, we obtain the constant free a posteriori error bound

$$\|\nabla(U - u)\|^2_{L^2(\Omega)} \le \|\nabla U - \xi\|^2_{L^2(\Omega)}.$$

Braess and Schöberl construct a suitable ξ based on a local flux equilibration using broken Raviart–Thomas spaces; see [10]. For ease of exposition we restrict ourselves here to the case $d = 2$ and consider the broken Raviart–Thomas space

$$\mathbb{RT}^{-1}(\mathcal{T}) := \left\{ \boldsymbol{g} \in L^2(\Omega; \mathbb{R}^2) \mid \boldsymbol{g}_{|T}(x) = \boldsymbol{a} + bx, \boldsymbol{a} \in \mathbb{R}^2, b \in \mathbb{R}\ \forall T \in \mathcal{T} \right\}.$$

The space $\mathbb{RT}^{-1}(\mathcal{T})$ does not require any continuity across inter-element sides.

The construction of the estimator is based on the solution of local divergence equations in the local spaces $\mathbb{RT}^{-1}(\mathcal{T}; z) := \{\boldsymbol{g}_{|\omega_z} \mid \boldsymbol{g} \in \mathbb{RT}^{-1}(\mathcal{T})\}$, $z \in \mathcal{N}$. To be more precise: Given $z \in \mathcal{N}$ find $\xi_z \in \mathbb{RT}^{-1}(\mathcal{T}; z)$ with minimal L^2-norm such that

$$\begin{aligned} \operatorname{div} \xi_{z|T} &= -\frac{1}{|T|} \int_T f\phi_z \, dx =: f_{z|T}, && \text{in each } T \subset \omega_z, \\ [\![\xi_z]\!]_{|\sigma} &= \int_\sigma J(U)\phi_z \, do = \frac{1}{2} J(U)_{|\sigma}, && \text{on each } \sigma \subset \Sigma_z, \\ \xi_z \cdot \boldsymbol{n} &= 0, && \text{on } \partial\omega_z, \end{aligned} \tag{2.16}$$

where $\boldsymbol{n}$ denotes the outer unit normal on $\partial\omega_z$. Note, that a necessary condition for (2.16) being well-posed is the Galerkin orthogonality $\int_{\Sigma_z} J(U)\phi_z \, do = \int_{\omega_z} f\phi_z \, dx$. The solution ξ_z of (2.16) depends linearly on discrete data f_z and $\nabla U_{|\omega_z}$. Since ξ_z has minimal L^2-norm we conclude $\xi_z \equiv 0$ if and only if $f_z \equiv 0$ in ω_z and ∇U is constant on ω_z. Therefore, applying equivalence of norms on finite dimensional spaces in combination with standard scaling arguments we arrive at

$$\|\xi_z\|^2_{L^2(\omega_z)} \approx h_z^2 \|f_z \phi_z^{1/2}\|^2_{L^2(\omega_z)} + h_z \|J(U)\phi_z^{1/2}\|^2_{L^2(\Sigma_z)}. \tag{2.17}$$

This means, that up to oscillation $\|\xi_z\|_{L^2(\omega_z)}$ is locally equivalent to the residual estimator organized by stars; compare with Sect. 2.2.2.

We next construct $\xi \in H(\mathrm{div}; \Omega)$ by using the partition of unity $\{\phi_z\}_{z\in\mathcal{N}}$:

$$\operatorname{div} \xi_{\mathcal{T}}(U) := \operatorname{div} \sum_{z\in\mathcal{N}} \xi_z = \operatorname{Res}(U) + f - f_{\mathcal{T}}$$

in distributional sense, where we used that $f_{\mathcal{T}} = \sum_{z\in\mathcal{N}} f_z\phi_z$ for the piecewise constant L^2 best approximation $f_{\mathcal{T}}$ to f of Sect. 2.2.1. Taking $\xi := \nabla U + \xi_{\mathcal{T}}(U)$ the theorem of Prager and Synge in combination with a perturbation argument yields

$$\|\nabla u - \nabla U\|^2_{L^2(\Omega)} \le \|\xi_{\mathcal{T}}(U)\|^2_{L^2(\Omega)} + C \sum_{z\in\mathcal{N}} h_z^2 \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)}. \tag{2.18a}$$

see [8,10]. This is the upper bound for the indicators

$$\mathcal{E}^2_{\mathcal{T}}(U, z) := \|\xi_z\|^2_{L^2(\omega_z)} + h_z^2 \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)}.$$

The equivalence (2.17), in combination with (2.10b) yields the lower bound

$$\mathcal{E}^2_{\mathcal{T}}(U, \mathcal{N}) := \sum_{z\in\mathcal{N}} \mathcal{E}^2_{\mathcal{T}}(U, z) \lesssim \|\nabla u - \nabla U\|^2_{L^2(\Omega)} + \operatorname{osc}^2_{\mathcal{T}}(f, \mathcal{N}) \tag{2.18b}$$

with

$$\operatorname{osc}^2_{\mathcal{T}}(f, z) := h_z^2 \|(f - f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)} \quad \text{and} \quad \operatorname{osc}^2_{\mathcal{T}}(f, \mathcal{N}) := \sum_{z\in\mathcal{N}} \operatorname{osc}^2_{\mathcal{T}}(f, z).$$

The local equivalence to the standard residual estimator follows by (2.17) similar to Sect. 2.2.2. In addition we have to use the fact that the residual estimator dominates the oscillation to obtain

$$\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{N}') \lesssim \hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}(\mathcal{N}')), \qquad \forall \mathcal{N}' \subset \mathcal{N}, \tag{2.19a}$$

$$\hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}') \lesssim \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{N}(\mathcal{T}')), \qquad \forall \mathcal{T}' \subset \mathcal{T}, \tag{2.19b}$$

where $\mathcal{T}(\mathcal{N}')$ and $\mathcal{N}(\mathcal{T}')$ are defined in Sect. 2.2.2.

Remark 2.1 Assuming that f is piecewise constant over an initial triangulation $\mathcal{T}_0$, we could also include the estimator by Braess et al. [6]. However, it is unclear how to adopt the result for general f.

2.2.6 Recovery based estimators

Zienkiewicz and Zhu introduced an estimator based upon gradient recovery [42]. Philosophically, the ZZ-estimator is an estimate for $\|\nabla(U-u)\|_{L^2(\Omega)}$ rather than for $|\!|\!|\mathrm{Res}(U)|\!|\!|_*$. For the particular problem at hand it holds $|\!|\!|\cdot|\!|\!| = \|\nabla\cdot\|_{L^2(\Omega)}$, whence the ZZ-estimator is also an estimator for the energy norm that fits into our framework.

Denoting by $\mathbb{V}(\mathcal{T};\mathbb{R}^d)$ the space of continuous, piecewise affine vector fields over $\mathcal{T}$, the averaging operator $\mathcal{G}_{\mathcal{T}}:\mathbb{V}\to\mathbb{V}(\mathcal{T};\mathbb{R}^d)$ is defined from the nodal values $(\mathcal{G}_{\mathcal{T}}V)(z) = \frac{1}{|\omega_z|}\int_{\omega_z}\nabla V\,dx$ for $z\in\mathcal{N}$. Based on this operator we define the local error indicators by

$$\mathcal{E}^2_{\mathcal{T}}(U,z) := \left\{\|(\nabla U - \mathcal{G}_{\mathcal{T}}U)\phi_z^{1/2}\|^2_{L^2(\omega_z)} + h_z^2\|(f-f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)}\right\}, \quad z\in\mathcal{N}.$$

Here, we use $h_z = \operatorname{diam}(\omega_z)$, and $f_z = \int_{\omega_z} f\phi_z\,dx / \int_{\omega_z}\phi_z\,dx$, if $z\in\mathring{\mathcal{N}} := \mathcal{N}\cap\Omega$, and $f_z = 0$, otherwise. The resulting estimator indexed by vertices is reliable and efficient, i.e.,

$$\|\nabla(U-u)\|^2_{L^2(\Omega)} \lesssim \sum_{z\in\mathcal{N}} \mathcal{E}^2_{\mathcal{T}}(U,z) =: \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{N}) \tag{2.20a}$$

$$\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{N}) \lesssim \|\nabla(U-u)\|^2_{L^2(\Omega)} + \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{N}). \tag{2.20b}$$

Here, oscillation for a vertex z is defined by

$$\operatorname{osc}^2_{\mathcal{T}}(f,z) := h_z^2\|(f-f_z)\phi_z^{1/2}\|^2_{L^2(\omega_z)} \quad\text{and}\quad \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{N}) := \sum_{z\in\mathcal{N}}\operatorname{osc}^2_{\mathcal{T}}(f,z);$$

compare for instance with [12].

Using equivalence of norms on finite dimensional spaces in combination with scaling arguments we obtain

$$h_z^{1/2}\|J(U)\|_{L^2(\Sigma_z)} \lesssim \|\nabla U - \mathcal{G}_{\mathcal{T}}U\|_{L^2(\omega_z)} \lesssim h_z^{1/2}\|J(U)\|_{L^2(\Sigma_z)} \qquad \forall z\in\mathcal{N} \tag{2.21}$$

recalling that Σ_z is the union of sides emanating from $z \in \mathcal{N}$. The oscillation part of the indicator $\mathcal{E}_{\mathcal{T}}(U, z)$ is dominated by the element residuals of $\hat{\mathcal{E}}_{\mathcal{T}}$ which in summary implies

$$\mathcal{E}_{\mathcal{T}}^2(U, \mathcal{N}') \lesssim \sum_{T \in \mathcal{T}(\mathcal{N}')} \hat{\mathcal{E}}_{\mathcal{T}}^2(U, T) =: \hat{\mathcal{E}}_{\mathcal{T}}^2(U, \mathcal{T}(\mathcal{N}')), \quad \mathcal{N}' \subset \mathcal{N}, \tag{2.22a}$$

with $\mathcal{T}(\mathcal{N}')$ defined as in Sect. 2.2.2.

For estimating the residual estimator by the ZZ-estimator we proceed as in Sect. 2.2.4 to obtain

$$h_z^2 \| f \phi_z^{1/2} \|_{L^2(\omega_z)}^2 \lesssim h_z \| J(U) \phi_z^{1/2} \|_{L^2(\Sigma_z)}^2 + h_z^2 \| (f - f_z) \phi_z^{1/2} \|_{L^2(\omega_z)}^2 \qquad \forall z \in \mathcal{N}.$$

Combining this with (2.21) and recalling that $\{\phi_z\}_{z \in \mathcal{N}}$ is a partition of unity we deduce

$$\hat{\mathcal{E}}_{\mathcal{T}}^2(U, \mathcal{T}') \lesssim \sum_{z \in \mathcal{N}(\mathcal{T}')} \mathcal{E}_{\mathcal{T}}^2(U, z) = \mathcal{E}_{\mathcal{T}}^2(U, \mathcal{N}(\mathcal{T}')), \quad \mathcal{T}' \subset \mathcal{T}, \tag{2.22b}$$

where we use the definition of $\mathcal{N}(\mathcal{T}')$ from Sect. 2.2.2. This shows that the residual estimator $\hat{\mathcal{E}}_{\mathcal{T}}$ and the ZZ-estimator are locally equivalent.

3 An optimal adaptive finite element method

In this section we analyze the standard adaptive loop (1.1) with main focus on different estimators. We first state the precise assumptions on the adaptive algorithm and then prove the main results.

3.1 Adaptive discretization

Before embarking on the adaptive algorithm we describe the framework for refinement and error estimation.

3.1.1 Framework for refinement

We restrict ourselves to refinement by bisection; compare with [2,20,21,37] as well as [27,32] and the references therein. To be more precise: Refinement is based on an initial conforming triangulation $\mathcal{T}_0$ of Ω and a procedure REFINE with the following properties. Given a conforming triangulation $\mathcal{T}$ and a subset $\mathcal{M} \subset \mathcal{T}$ of *marked elements*

$$\mathcal{T}_* = \mathsf{REFINE}(\mathcal{T}, \mathcal{M})$$

is a conforming refinement of $\mathcal{T}$ such that all elements in $\mathcal{M}$ are bisected. In general, additional elements are refined in order to ensure conformity. The input $\mathcal{T}$ can either be $\mathcal{T}_0$ or the output of a previous application of REFINE. The class of all conforming triangulations that can be produced from $\mathcal{T}_0$ by REFINE we denote by $\mathbb{T}$. For $\mathcal{T} \in \mathbb{T}$ we call $\mathcal{T}_* \in \mathbb{T}$ a refinement of $\mathcal{T}$ if $\mathcal{T}_*$ is produced from $\mathcal{T}$ by a finite number of applications of REFINE and we denote this by $\mathcal{T} \le \mathcal{T}_*$ or $\mathcal{T}_* \ge \mathcal{T}$.

One key property of the refinement rule is uniform shape regularity for any $\mathcal{T} \in \mathbb{T}$. This means that constants depending on the shape regularity are uniformly bounded by a constant depending on $\mathcal{T}_0$. In particular, all constants of Sect. 2.2 are uniform for $\mathcal{T} \in \mathbb{T}$. Hereafter, all explicit constants C, C_i, or implicit constants hidden in '$\lesssim$', '$\gtrsim$', and '$\approx$' do only depend on the class $\mathbb{T}$, the domain Ω, and the dimension d. Furthermore, we only deal with conforming grids, this means, whenever we refer to some triangulation $\mathcal{T}$ or $\mathcal{T}_*$ we implicitly assume $\mathcal{T}, \mathcal{T}_* \in \mathbb{T}$.

3.1.2 Framework for error estimation

We next set up a framework for error estimation that includes all estimators presented in Sect. 2.2 and that might also be suitable to treat other estimators. As we have seen, the various estimators are indexed differently, namely by elements, sides, or vertices, and the local equivalence of Sect. 2.2 involves local element patches depending on the index set of $\mathcal{E}_\mathcal{T}$. We introduce some notation that allows for a unified presentation. We denote by $\mathcal{I}$ the index set that is used for indexing the estimator, i.e., $\mathcal{I}$ is either $\mathcal{T}$, $\mathcal{S}$, or $\mathcal{N}$. For any subset $\mathcal{I}' \subset \mathcal{I}$ we use the notation

$$\mathcal{E}^2_\mathcal{T}(U,\mathcal{I}') := \sum_{I\in\mathcal{I}'} \mathcal{E}^2_\mathcal{T}(U,I) \quad \text{and} \quad \operatorname{osc}^2_\mathcal{T}(U,\mathcal{I}') := \sum_{I\in\mathcal{I}'} \operatorname{osc}^2_\mathcal{T}(U,I).$$

The local equivalences of Sect. 2.2 involve the patches ω_I that are defined as follows. If $\mathcal{I} = \mathcal{T}$ then $\omega_T := \omega(\mathcal{T}, T) := T$, else if $\mathcal{I} = \mathcal{S}$ we use $\omega_\sigma := \omega(\mathcal{T}, \sigma) = T_1 \cup T_2$ for an interior side $\sigma = T_1 \cap T_2$, and if $\mathcal{I} = \mathcal{N}$ we set $\omega_z := \omega(\mathcal{T}, z) = \operatorname{supp}(\phi_z)$. Note, that the definition of the patches depends on the underlying triangulation. A patch $\omega(\mathcal{T}, I)$ may change during refinement although the index I does not change.

We use the following relation between an index set and the corresponding triangulation

$$\mathcal{T}(\mathcal{I}') := \{T \in \mathcal{T} \mid T \subset \omega_I \quad \text{for some } I \in \mathcal{I}'\} \qquad \forall \mathcal{I}' \subset \mathcal{I}$$

and the converse relation

$$\mathcal{I}(\mathcal{T}') := \{I \in \mathcal{I} \mid I \subset T \quad \text{for some } T \in \mathcal{T}'\} \qquad \forall \mathcal{T}' \subset \mathcal{T}.$$

The definition of $\mathcal{T}(\mathcal{I}')$ and $\mathcal{I}(\mathcal{T}')$ for subsets $\mathcal{I}' \subset \mathcal{I}$ and $\mathcal{T}' \subset \mathcal{T}$ allows us to sum indicators over the corresponding subsets.

For a given index $I \in \mathcal{I}$, the set $\mathcal{T}(\{I\})$ contains exactly those elements $T \in \mathcal{T}$ that enter in the definition of $\mathcal{E}_\mathcal{T}(U, I)$. In this vein, the set of *active indices* on $T \in \mathcal{T}$ is $\mathcal{I}(\{T\})$, i.e., T enters in the definition of $\mathcal{E}_\mathcal{T}(U, I)$ for all $I \in \mathcal{I}(\{T\})$. Using this notion, the local equivalences (2.6), (2.11), (2.13), (2.15), and (2.22) for the different estimators then read

$$\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{I}') \lesssim \hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}(\mathcal{I}')) \quad \text{and} \quad \hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}') \lesssim \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{I}(\mathcal{T}')) \tag{3.1}$$

for all $\mathcal{I}' \subset \mathcal{I}$ and $\mathcal{T}' \subset \mathcal{T}$.

When dealing with a refinement $\mathcal{T} \leq \mathcal{T}_*$ we need the notion of *refined index set*. Denoting by $\mathcal{I} = \mathcal{I}(\mathcal{T})$ and $\mathcal{I}_* = \mathcal{I}(\mathcal{T}_*)$ the respective index sets, the set of refined indices is

$$\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I}) := \mathcal{I} \setminus \mathcal{I}_* \cup \{I \in \mathcal{I} \cap \mathcal{I}_* \mid \omega(\mathcal{T}, I) \neq \omega(\mathcal{T}_*, I)\},$$

i.e., $\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I})$ contains all those indices $I \in \mathcal{I}$ such that at least one element inside the patch $\omega(\mathcal{T}, I)$ is refined when going from $\mathcal{T}$ to $\mathcal{T}_*$. Note, that $\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{T})$ is the set of *refined elements*, i.e., $\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{T}) = \mathcal{T} \setminus \mathcal{T}_*$. The notion of refined index set is more involved than the set of refined elements, which can be seen from the following lemma.

Lemma 3.1 (Relation of Refined Indices and Elements) *For given triangulation $\mathcal{T}$ and refinement $\mathcal{T}_* \geq \mathcal{T}$ holds*

$$\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I}) = \mathcal{I}(\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{T})) = \mathcal{I}(\mathcal{T}_* \backslash \mathcal{T}).$$

Proof $\boxed{1}$ We first show $\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I}) \subseteq \mathcal{I}(\mathcal{T}_* \backslash \mathcal{T})$. Considering the case $I \in \mathcal{I} \backslash \mathcal{I}_* = \mathcal{I}(\mathcal{T}) \setminus \mathcal{I}(\mathcal{T}_*)$ we realize that there is a $T \in \mathcal{T} \setminus \mathcal{T}_*$ such that $I \subset T$. Therefore, $I \in \mathcal{I}(\mathcal{T} \setminus \mathcal{T}_*)$. For $I \in \{I \in \mathcal{I} \cap \mathcal{I}_* \mid \omega(\mathcal{T}, I) \neq \omega(\mathcal{T}_*, I)\}$ there exists $T \in \mathcal{T} \setminus \mathcal{T}_*$ such that $T \subset \omega(\mathcal{T}, I)$. This implies $I \subset T$ and consequently $I \in \mathcal{I}(\mathcal{T} \setminus \mathcal{T}_*)$.

$\boxed{2}$ We next show $\mathcal{I}(\mathcal{T}_* \setminus \mathcal{T}) \subseteq \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I})$. For $I \in \mathcal{I}(\mathcal{T} \setminus \mathcal{T}_*)$ we distinguish two cases: $I \in \mathcal{I} \setminus \mathcal{I}_*$ and $I \in \mathcal{I} \cap \mathcal{I}_*$. In the first case we obviously have $I \in \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I})$. In the later one there is $T \in \mathcal{T}$ and $T_* \in \mathcal{T}_*$ such that $I \subset T$ and $I \subset T_*$. This in turn implies $\omega(\mathcal{T}, I) \neq \omega(\mathcal{T}_*, I)$ and therefore $I \in \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I})$. □

3.1.3 The adaptive algorithm: assumptions and basic properties

We are now in the position to formulate the adaptive method and to state the assumptions on its modules. Starting with $\mathcal{T}_0$ the adaptive loop is an iteration of the following main steps:

$$\begin{aligned}
&(1)\; U_k := \mathsf{SOLVE}\,(\mathbb{V}(\mathcal{T}_k)).\\
&(2)\; \{\mathcal{E}_k(U_k, I)\}_{I\in\mathcal{I}_k} := \mathsf{ESTIMATE}\,(U_k, \mathcal{I}_k, \mathcal{T}_k).\\
&(3)\; \mathcal{M}_k := \mathsf{MARK}\left(\{\mathcal{E}_k(U_k, I)\}_{I\in\mathcal{I}_k},\, \mathcal{I}_k\right).\\
&(4)\; \mathcal{T}_{k+1} := \mathsf{REFINE}\,(\mathcal{T}_k,\; \mathcal{T}(\mathcal{M}_k)),\ \text{increment}\, k.
\end{aligned} \tag{3.2}$$

We next state the precise assumptions on the modules.

SOLVE. Given a grid $\mathcal{T} \in \mathbb{T}$, the output

$$U = \mathsf{SOLVE}(\mathcal{T})$$

is the exact Ritz-approximation $U \in \mathbb{V}(\mathcal{T})$ to u, i.e., U is the solution to (2.2).

The presented results crucially rely on the following properties of the finite element spaces. Whenever $\mathcal{T} \leq \mathcal{T}_*$ holds, the finite element spaces are nested, i.e., $\mathbb{V}(\mathcal{T}) \subset \mathbb{V}(\mathcal{T}_*)$. Nesting of spaces in combination with the fact that $\mathcal{B}$ is a scalar product yields the following orthogonal error decomposition

$$|\!|\!| U - u |\!|\!|^2 = |\!|\!| U_* - u |\!|\!|^2 + |\!|\!| U_* - U |\!|\!|^2 , \tag{3.3}$$

where $U \in \mathbb{V}(\mathcal{T})$ and $U_* \in \mathbb{V}(\mathcal{T}_*)$ are the Ritz-approximations to u in $\mathbb{V}(\mathcal{T})$ respectively $\mathbb{V}(\mathcal{T}^*)$.

ESTIMATE. Given a grid $\mathcal{T} \in \mathbb{T}$, the index set $\mathcal{I}$, and the discrete solution $U \in \mathbb{V}(\mathcal{T})$ the output

$$\{\mathcal{E}_{\mathcal{T}}(U, I)\}_{I \in \mathcal{I}} = \mathsf{ESTIMATE}(U, \mathcal{I}, \mathcal{T})$$

is a set of *local error indicators* $\mathcal{E}_{\mathcal{T}}(U, I)$ that build up an *error estimator* $\mathcal{E}_{\mathcal{T}}(U, \mathcal{I})$ with the following properties:

(1) The estimator is *reliable and efficient*, i.e., there exist constants C_1 and C_2 such that

$$|\!|\!| U - u |\!|\!|^2 \leq C_1 \mathcal{E}^2_{\mathcal{T}}(U, \mathcal{I}) \tag{3.4a}$$

and

$$C_2 \mathcal{E}^2_{\mathcal{T}}(U, \mathcal{I}) \leq |\!|\!| U - u |\!|\!|^2 + \operatorname{osc}^2_{\mathcal{T}}(f, \mathcal{I}). \tag{3.4b}$$

(2) The estimator provides a *localized upper bound*, i.e., there exists a constant $\bar{C}_1$ such that for any refinement $\mathcal{T}_* \geq \mathcal{T}$ and the Ritz-approximation $U_* \in \mathbb{V}(\mathcal{T}_*)$ there holds

$$|\!|\!| U - U_* |\!|\!|^2 \leq \bar{C}_1 \mathcal{E}^2_{\mathcal{T}} \left(U, \mathcal{R}_{\mathcal{T} \to \mathcal{T}_*}(\mathcal{I}) \right). \tag{3.5}$$

(3) We require that $\mathcal{E}_{\mathcal{T}}$ is *locally equivalent* to the *residual estimator* $\hat{\mathcal{E}}_{\mathcal{T}}$, i.e., there exist constants $0 < C_3 \leq C_4$ with

$$C_3 \mathcal{E}^2_{\mathcal{T}}(U, \mathcal{I}') \leq \hat{\mathcal{E}}^2_{\mathcal{T}}(U, \mathcal{T}(\mathcal{I}')) \qquad \forall \mathcal{I}' \subset \mathcal{I} \tag{3.6a}$$

and

$$\hat{\mathcal{E}}^2_{\mathcal{T}}(U, \mathcal{T}') \leq C_4 \mathcal{E}^2_{\mathcal{T}}(U, \mathcal{I}(\mathcal{T}')) \qquad \forall \mathcal{T}' \subset \mathcal{T}; \tag{3.6b}$$

(4) We suppose that oscillation $\operatorname{osc}_{\mathcal{T}}$ has the following properties. The estimator dominates oscillation, i.e.,

$$\operatorname{osc}_{\mathcal{T}}(f, I) \leq \mathcal{E}_{\mathcal{T}}(U, I) \qquad \forall\, I \in \mathcal{I}, \tag{3.7a}$$

Let $\mathcal{T}_* \leq \mathcal{T}$ be a refinement of $\mathcal{T}$. Oscillation is quasi-monotone in the sense that

$$\operatorname{osc}_{\mathcal{T}_*}(f, \mathcal{I}_*) \leq C_5 \operatorname{osc}_{\mathcal{T}}(f, \mathcal{I}), \tag{3.7b}$$

and oscillation does not change in non-refined regions, i.e., there holds

$$\operatorname{osc}_{\mathcal{T}_*}(f, I) = \operatorname{osc}_{\mathcal{T}}(f, I) \qquad \forall\, I \in \mathcal{I} \setminus \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I}). \tag{3.7c}$$

All the estimators of Sect. 2.2 satisfy these assumptions. In particular, estimates (2.4), (2.10), (2.12), (2.14), (2.20), and (2.18) are the upper and lower bound (3.4). The local equivalences of the estimators to $\hat{\mathcal{E}}_{\mathcal{T}}$ (2.6), (2.11), (2.13), (2.15), (2.19), and (2.22) are summarized in (3.6). We shall prove the localized upper bound (3.5) in Sect. 3.2 separately.

The required properties (3.7) of oscillation are rather standard. That is why we did not label them explicitly in Sect. 2.2. In fact, the residual type estimators of Sects. 2.2.1 and 2.2.2 dominate the oscillation since oscillation is defined as an L^2 projection of the element residual to the piecewise constant functions. In Sects. 2.2.3–2.2.5 oscillation is part of the estimator, which obviously implies (3.7a). Assumptions (3.7b) and (3.7c) are a direct consequence of the local mesh-size reduction; compare with [23] and [25]. Moreover, in most cases $C_5 = 1$.

MARK. The marking procedure utilizes *Dörfler Marking* [18], i.e., for fixed parameter $\theta \in (0, 1]$ the output

$$\mathcal{M} = \mathsf{MARK}(\{\mathcal{E}_{\mathcal{T}}(U, I)\}_{I\in\mathcal{I}}, \mathcal{I})$$

is a subset $\mathcal{M} \subset \mathcal{I}$ of selected indices satisfying the Dörfler property

$$\theta \mathcal{E}_{\mathcal{T}}(U, \mathcal{I}) \leq \mathcal{E}_{\mathcal{T}}(U, \mathcal{M}). \tag{3.8}$$

We additionally suppose that the output $\mathcal{M}$ has *minimal cardinality* and that the marking parameter θ satisfies $\theta \in (0, \theta_*)$ with

$$\theta_*^2 = C_2/(\bar{C}_1 + 1), \tag{3.9}$$

where C_2 is the constant of the lower bound (3.4b) and $\bar{C}_1$ the constant of the localized upper bound (3.5).

Dörfler marking for $\mathcal{E}_{\mathcal{T}}$ in combination with the local equivalence (3.6) of $\mathcal{E}_{\mathcal{T}}$ and $\hat{\mathcal{E}}_{\mathcal{T}}$ implies a Dörfler property of $\hat{\mathcal{E}}_{\mathcal{T}}$ with parameter $\hat{\theta} \in (0, \theta]$. This observation is the key to derive a strictly monotone error quantity; compare with Theorem 3.6 and Corollary 3.6 below.

Proposition 3.2 (Dörfler Property) *For given $\mathcal{T}$ let $U \in \mathbb{V}(\mathcal{T})$ be the Ritz projection and $\mathcal{E}_{\mathcal{T}}(U, \mathcal{I})$ the corresponding estimator. Assume that $\mathcal{E}_{\mathcal{T}}$ is locally equivalent to $\hat{\mathcal{E}}_{\mathcal{T}}$ in the sense of* (3.6) *and that $\mathcal{E}_{\mathcal{T}}$ satisfies the Dörfler property* (3.8) *on $\mathcal{M} \subset \mathcal{I}$.*

Then the residual estimator $\hat{\mathcal{E}}_{\mathcal{T}}$ satisfies the Dörfler property on $\mathcal{T}(\mathcal{M})$ with parameter $\hat{\theta} = (C_3/C_4)^{1/2}\,\theta \in (0,\theta]$, i.e.,

$$\hat{\theta}\hat{\mathcal{E}}_{\mathcal{T}}(U,\mathcal{T}) \le \hat{\mathcal{E}}_{\mathcal{T}}(U,\mathcal{T}(\mathcal{M})).$$

Proof The local equivalence of $\mathcal{E}_{\mathcal{T}}$ and $\hat{\mathcal{E}}_{\mathcal{T}}$ readily imply

$$\hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}(\mathcal{M})) \ge C_3\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{M}) \ge C_3\theta^2\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{I}) \ge \frac{C_3}{C_4}\theta^2\hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}) = \hat{\theta}^2\hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}).$$

This is the assertion. □

REFINE. From the output $\mathcal{M} \subset \mathcal{I}$ from MARK of *marked indices* we have to define a set of *marked elements* as input for REFINE. We recall that for any $I \in \mathcal{M}$ exactly the elements in $\mathcal{T}(\{I\})$ enter the definition of $\mathcal{E}_{\mathcal{T}}(U,I)$, whence a natural choice as input for refine is $\mathcal{T}(\mathcal{M})$. Note, that thanks to uniform shape-regularity there holds

$$\#\mathcal{T}(\mathcal{M}) \lesssim \#\mathcal{M}.$$

We therefore suppose that $\mathsf{REFINE}(\mathcal{T},\mathcal{T}(\mathcal{M}))$ outputs the *smallest* conforming refinement $\mathcal{T}_*$ of $\mathcal{T}$ such that all elements in $\mathcal{T}(\mathcal{M})$ are bisected b times, where $b \in \mathbb{N}$ is fixed. In addition, we pose restrictions on the initial grid ensuring that any uniform refinement $\mathcal{T}_g$ of generation $g \in \mathbb{N}_0$ of $\mathcal{T}_0$ is conforming. We thereby call $\mathcal{T}_g$ a uniform refinement of generation g, if $\mathcal{T}_g$ is generated by bisecting recurrently all elements in $\mathcal{T}_0$ exactly g times. Note, that in general $\mathcal{T}_g$ is non-conforming but a proper distribution of refinement edges on the initial triangulation $\mathcal{T}_0$ guarantees that any uniform refinement of generation $g \in \mathbb{N}_0$ is conforming. Conditions how to assign the refinement edges for the elements on $\mathcal{T}_0$ are given in Sect. 4 in [34]; compare also with [27, Assumption 1]. The restriction on the initial grid $\mathcal{T}_0$ has the following consequence.

Lemma 3.3 (Complexity of REFINE) *Assume that any uniform refinement $\mathcal{T}_g$ of generation $g \in \mathbb{N}_0$ of $\mathcal{T}_0$ is conforming. For $k \ge 0$ let $\{\mathcal{T}_k\}_{k\ge 0}$ be any sequence of refinements of $\mathcal{T}_0$ where $\mathcal{T}_{k+1}$ is generated from $\mathcal{T}_k$ by $\mathcal{T}_{k+1} = \mathsf{REFINE}(\mathcal{T}_k,\ \mathcal{T}_k(\mathcal{M}_k))$ with a subset $\mathcal{M}_k \subset \mathcal{I}_k$.*

Then, there exists a constant C_0 solely depending on $\mathcal{T}_0$, the number b of bisections, and the dimension d such that

$$\#\mathcal{T}_k - \#\mathcal{T}_0 \le C_0 \sum_{\ell=0}^{k-1} \#\mathcal{M}_\ell \qquad \forall\, k \ge 1.$$

Proof Binev et al. have shown for $d = 2$ [3, Theorem 2.4] and Stevenson for $d \ge 2$ [34, Theorem 6.1] the estimate

$$\#\mathcal{T}_k - \#\mathcal{T}_0 \lesssim \sum_{\ell=0}^{k-1} \#\mathcal{T}_\ell(\mathcal{M}_\ell) \qquad \forall\, k \ge 1,$$

when REFINE bisects all marked elements once. The claim follows readily from $\#\mathcal{T}_\ell(\mathcal{M}_\ell) \lesssim \#\mathcal{M}_\ell$. For the generalization to $b > 1$ bisections compare with the remark after Lemma 2.3 in [14]. □

3.2 Localized upper bound

One key ingredient in the optimality proof is the localized upper bound (3.5) for the difference of two Ritz projections. Such a bound is a consequence of Galerkin orthogonality in combination with a suitable interpolation operator that locally preserves discrete functions. It was first proved by Stevenson [33, Theorem 4.1] with a Clément interpolant [15]. Cascón et al. completely localized the bound to the set of refined elements by using a Scott–Zhang interpolant [36] with the following properties [14, Lemma 3.6].

Lemma 3.4 *For given $\mathcal{T} \le \mathcal{T}_*$ let $\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*} := \mathcal{T} \setminus \mathcal{T}_*$ be the set of refined elements and set $\Omega_{\mathcal{R}} := \bigcup\{T : T \in \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}\}$.*

There exists interpolation operator $P_{\mathcal{R}} : \mathbb{V} \to \mathbb{V}(\mathcal{T})$ such that

$$V_* - P_{\mathcal{R}} V_* \equiv 0 \quad \text{in } \Omega \setminus \Omega_{\mathcal{R}} \qquad \forall\, V_* \in \mathbb{V}(\mathcal{T}_*). \tag{3.10a}$$

The operator is H^1 stable, i.e.,

$$\|\nabla(P_{\mathcal{R}} v - v)\|^2_{L^2(\Omega)} \le C_{P_{\mathcal{R}}} \|\nabla v\|^2_{L^2(\Omega)} \qquad \forall\, v \in \mathbb{V}, \tag{3.10b}$$

and satisfies the interpolation estimate

$$\sum_{T\in\mathcal{T}} \left\{ \|h_{\mathcal{T}}(v - P_{\mathcal{R}} v)\|^2_{L^2(T)} + \|h_{\mathcal{T}}^{-1/2}(v - P_{\mathcal{R}} v)\|^2_{L^2(\partial T)} \right\} \lesssim \|\nabla v\|_{L^2(\Omega)}. \tag{3.10c}$$

for all $v \in \mathbb{V}$. The operator $P_{\mathcal{R}}$ depends on $\mathcal{T}$ and $\mathcal{T}_*$ whereas $C_{P_{\mathcal{R}}}$ and the constant hidden in '$\lesssim$' do only depend on $\mathbb{T}$ and not on the particular choice of $\mathcal{T}$ and $\mathcal{T}_*$.

Using the interpolation operator $P_{\mathcal{R}}$ the localized upper bound (3.5) for the residual estimator of Sect. 2.2.1 and its first variant in Sect. 2.2.2 can be shown exactly with the same arguments as the upper bound (3.4a). The localization to the set of refined elements in the localized upper bound is a direct consequence of (3.10a). Moreover, using $P_{\mathcal{R}}$ in both upper bound and localized upper bound results in the same constant $C_1 = \bar{C}_1$. Note, that this is an important aspect, since the constant $\bar{C}_1$ of the localized upper bound enters in the restriction of the marking parameter θ_* in (3.9).

Unfortunately, the upper bounds for all other estimators but the equilibrated residual estimator from Sect. 2.2.5 are derived with different kind of interpolation operators that are not locally preserving discrete functions. The derivation of the upper bound for the equilibrated residual estimator does not use any interpolant. Galerkin orthogonality is only used implicitly in that the local problems (2.16) are well-posed. It is therefore not obvious how to localize the upper bound for the difference of two discrete solutions. Before addressing the localized upper bounds for these estimators we want to investigate the relation of the constants C_1 and $\bar{C}_1$.

Assume that C_1 and $\bar{C}_1$ are the optimal constants, which are in general not known. Since the localized upper bound has to be valid for any refinement $\mathcal{T}_*$ of $\mathcal{T}$ we can conclude $C_1 \le \bar{C}_1$. We may wonder if we can expect $C_1 = \bar{C}_1$. In order to explore this matter we assume the 'best' estimator, namely the exact error $|||U - U_*|||$, which satisfies (3.4a) with $C_1 = 1$. In general, $U_* \neq U$ in $\Omega \setminus \Omega_{\mathcal{R}}$, using the notion of Lemma 3.4. This implies

$$|||U - U_*|||_{\Omega_{\mathcal{R}}} < |||U - U_*|||_{\Omega} .$$

Consequently, we cannot expect the same constant $C_1 = \bar{C}_1$. Moreover, an estimate

$$|||U - U_*|||_{\Omega} \le \bar{C}_1 \; |||U - U_*|||_{\Omega_{\mathcal{R}}} \tag{3.11}$$

cannot be valid for generic functions $U \in \mathbb{V}(\mathcal{T})$ and $U_* \in \mathbb{V}(\mathcal{T}_*)$ but relies on U being the Ritz approximation to U_*. Galerkin orthogonality implies for the energy error of $E_* = U - U_* \in \mathbb{V}(\mathcal{T}_*)$

$$\begin{aligned}|||U - U_*|||_{\Omega}^2 &= \mathcal{B}[U - U_*,\ E_*] = \mathcal{B}[U - U_*,\ E_* - P_{\mathcal{R}} E_*]\\ &\le |||U - U_*|||_{\Omega_{\mathcal{R}}}\ |||P_{\mathcal{R}} E_* - E_*|||_{\Omega_{\mathcal{R}}} \le |||U - U_*|||_{\Omega_{\mathcal{R}}}\ C_{P_{\mathcal{R}}}\ |||U - U_*|||_{\Omega} ,\end{aligned}$$

thanks to (3.10a). This shows (3.11) with $\bar{C}_1 \le C_{P_{\mathcal{R}}}$.

This motivates to prove the localized upper bound for any estimator where the upper bound involves an interpolation operator as follows. In a first step localize a test function $V_* \in \mathbb{V}(\mathcal{T}_*)$ by employing Galerkin orthogonality with $P_{\mathcal{R}} V_*$:

$$\sup_{|||V_*|||=1} \langle \mathrm{Res}(U),\ V_* \rangle = \sup_{|||V_*|||=1} \langle \mathrm{Res}(U),\ V_* - P_{\mathcal{R}} V_* \rangle \le C_{P_{\mathcal{R}}} \sup_{\substack{|||W_*|||=1\\ \mathrm{supp}\, W_* \subset \Omega_{\mathcal{R}}}} \langle \mathrm{Res}(U),\ W_* \rangle.$$

Proceed then further as in the proof of the upper bound using the interpolation operator applied to W_* with support in $\Omega_{\mathcal{R}}$. For all the estimators but the one from Sect. 2.2.5 the respective interpolation operators are defined via local projections $\Pi_I W_*$ on ω_I, $I \in \mathcal{I}$. By definition of the refined index set $\mathcal{R}_{\mathcal{I} \to \mathcal{I}_*}$ we see $W_* \equiv 0$ in ω_I and therefore $\Pi_I W_* \equiv 0$ for all $I \in \mathcal{I} \setminus \mathcal{R}_{\mathcal{I} \to \mathcal{I}_*}$. From this we conclude the localized upper bound (3.5) with $\bar{C}_1 \le C_{P_{\mathcal{R}}} C_1$, where C_1 is the constant from the upper bound (3.4a). Recalling the discussion above using the true error as estimator, it is apparent that we cannot expect a better bound for $\bar{C}_1$ in general.

It remains to derive the localized upper bound for the equilibrated residual estimator from Sect. 2.2.5. Since Galerkin orthogonality only enters implicitly the derivation of the upper bound (2.18a) it is not clear that a localized upper bound can directly be shown. We master this difficulty by employing the equivalence to the standard residual estimator. The following result can be used for any estimator $\mathcal{E}_{\mathcal{T}}$ that is locally equivalent to $\hat{\mathcal{E}}_{\mathcal{T}}$ in the sense of (3.6).

Lemma 3.5 (Localized Upper Bound) *For given grids $\mathcal{T} \leq \mathcal{T}_*$ let $U \in \mathbb{V}(\mathcal{T})$ and $U_* \in \mathbb{V}(\mathcal{T}_*)$ be the corresponding Ritz projections. Assume that the estimator $\mathcal{E}_\mathcal{T}$ is locally equivalent to $\hat{\mathcal{E}}_\mathcal{T}$ in the sense of* (3.6).

Then there holds with $\bar{C}_1 = \hat{C}_1 C_4$ the localized upper bound

$$|||U - U_*|||^2 \leq \bar{C}_1\, \mathcal{E}^2_\mathcal{T}\left(U, \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I})\right). \tag{3.12}$$

Proof From [14, Lemma 3.6] we know

$$|||U - U_*|||^2 \leq \hat{C}_1 \hat{\mathcal{E}}^2_\mathcal{T}(U, \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{T})), \tag{3.13}$$

where $\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{T}) = \mathcal{T}_* \setminus \mathcal{T}$ is the set of all refined elements and $\hat{C}_1$ is the same constant as in the upper bound (2.4a). The local equivalence of $\mathcal{E}_\mathcal{T}$ to $\hat{\mathcal{E}}_\mathcal{T}$ (3.6) therefore implies

$$|||U - U_*|||^2 \leq \hat{C}_1 \hat{\mathcal{E}}^2_\mathcal{T}\left(U, \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{T})\right) \leq \hat{C}_1 C_4 \mathcal{E}^2_\mathcal{T}\left(U, \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I})\right),$$

since $\mathcal{I}(\mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{T})) = \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I})$; compare with Lemma 3.1. □

3.3 Error reduction and optimal decay rate

In this section we prove the main results of this article. In doing this, we denote by $\{\mathcal{T}_k, \mathbb{V}_k, U_k, \mathcal{E}_k, \mathcal{I}_k, \mathcal{M}_k\}_{k\geq 0}$ the sequence of refinements $\mathcal{T}_k$ of $\mathcal{T}_0$, conforming discrete spaces $\mathbb{V}_k = \mathbb{V}(\mathcal{T}_k) \subset \mathbb{V}$, discrete solutions $U_k \in \mathbb{V}_k$, a posteriori error estimators $\mathcal{E}_k$, index sets $\mathcal{I}_k$, and marked indices $\mathcal{M}_k \subset \mathcal{I}_k$ generated by iteration (3.2). For sake of convenience we replace the subscript $\mathcal{T}_k$ by the iteration counter k, for instance $\mathcal{E}_k = \mathcal{E}_{\mathcal{T}_k}$.

3.3.1 Error reduction property

We next proof under the assumptions of Sect. 3.1.3 Main Result 1. The Dörfler property of the residual estimator $\hat{\mathcal{E}}_\mathcal{T}$, which is a consequence of Dörfler marking for $\mathcal{E}_\mathcal{T}$, compare with Proposition 3.2, allows us to prove the following contraction property.

Theorem 3.6 (Contraction Property) *Iteration* (3.2) *is a contraction for a scaled sum of energy error and* residual *estimator, i.e., there exist $\alpha \in (0, 1)$ and $\gamma > 0$ such that for all $k \geq 0$*

$$\left(|||U_{k+1} - u|||^2 + \gamma\hat{\mathcal{E}}^2_{k+1}(U_{k+1}, \mathcal{T}_{k+1})\right)^{1/2} \leq \alpha\left(|||U_k - u|||^2 + \gamma\hat{\mathcal{E}}^2_k(U_k, \mathcal{T}_k)\right)^{1/2}.$$

Proof We observe that the argument of REFINE is $\mathcal{T}(\mathcal{M})$. On $\mathcal{T}(\mathcal{M})$ the residual estimator $\hat{\mathcal{E}}_\mathcal{T}$ satisfies the Dörfler property

$$\hat{\theta}\hat{\mathcal{E}}_\mathcal{T}(U, \mathcal{T}) \leq \hat{\mathcal{E}}_\mathcal{T}(U, \mathcal{T}(\mathcal{M}))$$

with $\hat{\theta} = \sqrt{\frac{C_3}{C_4}}\theta > 0$, thanks to Proposition 3.2. Therefore, the assertion is an immediate consequence of [14, Theorem 4.1]. □

We want to remark that this contraction property does not require any restriction $\theta < \theta_*$ of the marking parameter nor minimality of the set $\mathcal{M}$ of marked indices. Moreover, the restrictive assumption on $\mathcal{T}_0$ that any uniform refinement is conforming is not used. The Dörfler property for $\hat{\mathcal{E}}_{\mathcal{T}}$ hinges on the definition of marked elements $\mathcal{T}(\mathcal{M})$ in Sect. 3.1.3.

Main Result 1 is now a direct consequence of the estimator equivalence and the contraction property.

Corollary 3.7 (Reduction Property) *Let $\alpha \in (0,1)$ and $\gamma > 0$ be the contraction and scaling constants from Theorem 3.6.*

Then there holds with $\Lambda_1 = \max\{1, \gamma C_4\}^{1/2} / \min\{1, \gamma C_3\}^{1/2}$ for all $0 \le \ell \le k$

$$\left(|\!|\!| U_k - u |\!|\!|^2 + \mathcal{E}_k^2(U_k, \mathcal{I}_k)\right)^{1/2} \le \Lambda_1 \, \alpha^{k-\ell} \left(|\!|\!| U_\ell - u |\!|\!|^2 + \mathcal{E}_\ell^2(U_\ell, \mathcal{I}_\ell)\right)^{1/2}.$$

Proof Set $\mathcal{E}_k = \mathcal{E}_k(U_k, \mathcal{I}_k)$ and $\hat{\mathcal{E}}_k = \hat{\mathcal{E}}_k(U_k, \mathcal{I}_k)$. Recalling the equivalence (3.6) of the estimators and the contraction property of the residual estimator from Theorem 3.6 we estimate

$$\begin{aligned}
\min\{1, \gamma C_3\} \left(|\!|\!| U_k - u |\!|\!|^2 + \mathcal{E}_k^2\right) &\le |\!|\!| U_k - u |\!|\!|^2 + \gamma C_3 \mathcal{E}_k^2 \\
&\le |\!|\!| U_k - u |\!|\!|^2 + \gamma \hat{\mathcal{E}}_k^2 \le \alpha^{2(k-\ell)} \left(|\!|\!| U_\ell - u |\!|\!|^2 + \gamma \hat{\mathcal{E}}_\ell^2\right) \\
&\le \alpha^{2(k-\ell)} \left(|\!|\!| U_\ell - u |\!|\!|^2 + \gamma C_4 \mathcal{E}_\ell^2\right) \\
&\le \alpha^{2(k-\ell)} \max\{1, \gamma C_4\} \left(|\!|\!| U_\ell - u |\!|\!|^2 + \mathcal{E}_\ell^2\right).
\end{aligned}$$

This is the desired error reduction property and finishes the proof. □

As already alluded in the introduction, relying on the local equivalence of $\mathcal{E}_{\mathcal{T}}$ to $\hat{\mathcal{E}}_{\mathcal{T}}$ it is of importance that the constants C_3 and C_4 are of moderate size, which is true for the Poisson problem. In Sect. 4 we consider robust estimators for perturbed Poisson problems and discuss how this affects the constant Λ_1.

3.3.2 Approximation class

We next introduce an appropriate error notion for the adaptive method together with the corresponding approximation class. Decisions of AFEM are driven only by virtue of the local error indicators. This indicates that the estimator is the quantity we can expect optimal rates for. On the one hand, oscillation is dominated by the estimator (3.7a), hence by the upper bound (3.4a) we have

$$|\!|\!| u - U |\!|\!|^2 + \operatorname{osc}_{\mathcal{T}}^2(f, \mathcal{T}) \le (1 + C_1)\, \mathcal{E}_{\mathcal{T}}^2(f, \mathcal{T}).$$

On the other hand we have by the lower bound (3.4b) that

$$C_2 \mathcal{E}^2_{\mathcal{T}}(f,\mathcal{T}) \le \left(\vert\!\vert\!\vert u-U \vert\!\vert\!\vert^2 + \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{T}) \right),$$

i.e., the right-hand side is equivalent to the estimator and its square root is called the *total error*.

This motivates the following definition of $\mathbb{A}_s$. Let $\mathbb{T}_N \subset \mathbb{T}$ be the set of all possible conforming triangulations generated from $\mathcal{T}_0$ with at most N elements more than $\mathcal{T}_0$:

$$\mathbb{T}_N := \{\mathcal{T} \in \mathbb{T} \mid \#\mathcal{T} - \#\mathcal{T}_0 \le N\}.$$

The quality of the best approximation to the total error in the set $\mathbb{T}_N$ is given by

$$\sigma(N; u, f) := \inf_{\mathcal{T}\in\mathbb{T}_N} \inf_{V\in\mathbb{V}(\mathcal{T})} \left(\vert\!\vert\!\vert u-V \vert\!\vert\!\vert^2 + \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{T}) \right)^{1/2}.$$

We now define the nonlinear approximation class $\mathbb{A}_s$ to be

$$\mathbb{A}_s := \left\{ (u,f) \mid |u,f|_s := \sup_{N>0} \left(N^s\, \sigma(N;u,f) \right) < \infty \right\}.$$

Remark 3.8 (Equivalent definitions of $\mathbb{A}_s$) The definition of $\mathbb{A}_s$ seems to depend on the particular error estimator, respectively its oscillation. This is not the case, which can be seen as follows. Let

$$\hat{\sigma}(N; u, f) := \inf_{\mathcal{T}\in\mathbb{T}_N} \inf_{V\in\mathbb{V}(\mathcal{T})} \left(\vert\!\vert\!\vert u-V \vert\!\vert\!\vert^2 + \widehat{\operatorname{osc}}^2_{\mathcal{T}}(f,\mathcal{T}) \right)^{1/2}$$

and

$$\hat{\mathbb{A}}_s := \left\{ (u,f) \mid \widehat{|u,f|}_s := \sup_{N>0} \left(N^s\, \hat{\sigma}(N;u,f) \right) < \infty \right\}$$

be the approximation class according to the standard residual estimator of Sect. 2.2.1. Recalling the equivalence of estimators (3.6) and observing the optimality of the Ritz approximation $U \in \mathbb{V}(\mathcal{T})$

$$\vert\!\vert\!\vert u-U \vert\!\vert\!\vert^2 + \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{T}) = \inf_{V\in\mathbb{V}(\mathcal{T})} \vert\!\vert\!\vert u-V \vert\!\vert\!\vert^2 + \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{T})$$

we obtain equivalence of the total errors, namely

$$\vert\!\vert\!\vert u-U \vert\!\vert\!\vert^2 + \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{T}) \approx \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{I}) \approx \hat{\mathcal{E}}^2_{\mathcal{T}}(U,\mathcal{T}) \approx \vert\!\vert\!\vert u-U \vert\!\vert\!\vert^2 + \widehat{\operatorname{osc}}^2_{\mathcal{T}}(f,\mathcal{T}),$$

and vice versa. Therefore, we conclude for all N

$$\hat{\sigma}(N; u, f) \approx \sigma(N; u, f),$$

which yields $\hat{\mathbb{A}}_s = \mathbb{A}_s$.

Remark 3.9 The definition of the approximation class $\mathbb{A}_s$ follows [14]. It is not a standard approximation class as used in approximation theory in contrast to

$$\mathcal{A}_s := \left\{ v \in \mathbb{V} \mid \ |v|_{\mathcal{A}_s} := \sup_{N>0} \left(N^s \inf_{\mathcal{T}\in\mathbb{T}_N} \inf_{V\in\mathbb{V}(\mathcal{T})} |\!|\!| v - V |\!|\!| \right) < \infty \right\},$$
$$\bar{\mathcal{A}}_s := \left\{ g \in L^2(\Omega) \mid \ |g|_{\bar{\mathcal{A}}_s} := \sup_{N>0} \left(N^s \inf_{\mathcal{T}\in\mathbb{T}_N} \|h_{\mathcal{T}}(g - P_{\mathcal{T}}\, g)\|_{L^2(\Omega)} \right) < \infty \right\},$$

where $P_{\mathcal{T}}$ is the L^2 projection onto the space of piecewise constants over $\mathcal{T}$; compare with [3,4,33]. However we have the following equivalence

$$(u, f) \in \mathbb{A}_s \qquad \Longleftrightarrow \qquad u \in \mathcal{A}_s \text{ and } f \in \bar{\mathcal{A}}_s;$$

compare for instance with [14, Lemma 5.3]. Moreover, from the reduction of the mesh-size $h_{\mathcal{T}}$ using uniform refinement and $f \in L^2(\Omega)$ it can easily be shown that $f \in \bar{\mathcal{A}}_{1/d}$ [14, Lemma 5.4]. The highest attainable order for linear elements is $s = 1/d$. Therefore, for all $s \le 1/d$ we have

$$(u, f) \in \mathbb{A}_s \qquad \Longleftrightarrow \qquad u \in \mathcal{A}_s.$$

The membership of (u, f) in $\mathbb{A}_s$ implies the following property: For any $\varepsilon > 0$ there exists $\mathcal{T}_\varepsilon \in \mathbb{T}$ and $U_\varepsilon \in \mathbb{V}(\mathcal{T}_\varepsilon)$ such that

$$|\!|\!| U_\varepsilon - u |\!|\!|^2 + \operatorname{osc}^2_{\mathcal{T}_\varepsilon}(f, \mathcal{I}_\varepsilon) \le \varepsilon^2 \qquad \text{and} \qquad \#\mathcal{T}_\varepsilon - \#\mathcal{T}_0 \lesssim |u, f|_s^{1/s}\, \varepsilon^{-1/s}, \quad (3.14)$$

where the constant hidden in '$\lesssim$' is close to 1.

3.3.3 Minimal Dörfler marking

We next derive properties that are related to the minimal Dörfler marking. The proofs mainly follow the presentations in [14,27]. However there are some important modifications that cannot be shortly summarized.

Lemma 3.10 (Optimal Marking) *Let $\mathcal{T} \le \mathcal{T}_*$ two grids with corresponding index sets $\mathcal{I} = \mathcal{I}(\mathcal{T})$, $\mathcal{I}_* = \mathcal{I}(\mathcal{T}_*)$ and discrete solutions $U \in \mathbb{V}(\mathcal{T})$ and $U_* \in \mathbb{V}(\mathcal{T}_*)$ such that*

$$|\!|\!| U_* - u |\!|\!|^2 + \operatorname{osc}^2_{\mathcal{T}_*}(U_*, \mathcal{I}_*) \le \mu \left(|\!|\!| U - u |\!|\!|^2 + \operatorname{osc}^2_{\mathcal{T}}(U, \mathcal{I}) \right)$$

with $\mu := \left(1 - \theta^2/\theta_*^2\right) > 0$.

Then the set $\mathcal{R} = \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{I})$ of refined indices satisfies the Dörfler property

$$\theta\, \mathcal{E}_{\mathcal{T}}(U, \mathcal{I}) \le \mathcal{E}_{\mathcal{T}}(U, \mathcal{R}).$$

Proof $\boxed{1}$ Combining the global lower bound with the assumption of the lemma we can write

$$\begin{aligned}(1-\mu)\,C_2\,\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{I}) &\le (1-\mu)\left(|\!|\!|U-u|\!|\!|^2+\operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{I})\right)\\ &\le \left(|\!|\!|U-u|\!|\!|^2-|\!|\!|u-U_*|\!|\!|^2\right)\\ &\quad+\left(\operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{I})-\operatorname{osc}^2_{\mathcal{T}_*}(f,\mathcal{I}_*)\right). \qquad (3.15)\end{aligned}$$

$\boxed{2}$ The orthogonality relation (3.3) in combination with the localized upper bound Lemma 3.5 yields

$$|\!|\!|U-u|\!|\!|^2-|\!|\!|u-U_*|\!|\!|^2=|\!|\!|U-U_*|\!|\!|^2\le \bar{C}_1\,\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{R}). \qquad (3.16)$$

To deal with oscillation we consider two subsets of $\mathcal{I}$. In $\mathcal{R}$ we use that oscillation is dominated by the estimator (3.7a) to deduce

$$\operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{R})-\operatorname{osc}^2_{\mathcal{T}_*}(f,\mathcal{R})\le \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{R})\le \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{R}).$$

In the complement $\mathcal{I}\setminus\mathcal{R}$ oscillation does not change, compare with (3.7c). Therefore,

$$\operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{I}\setminus\mathcal{R})-\operatorname{osc}^2_{\mathcal{T}_*}(f,\mathcal{I}\setminus\mathcal{R})=0.$$

In summary we have deduced

$$\operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{I})-\operatorname{osc}^2_{\mathcal{T}_*}(f,\mathcal{I}_*)\le \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{R}). \qquad (3.17)$$

$\boxed{3}$ Bounding the right-hand side of (3.15) by (3.16) and (3.17) we end up with

$$(1-\mu)\,C_2\,\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{T})\le(\bar{C}_1+1)\,\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{R}).$$

The definition of μ implies $\theta^2=(1-\mu)\theta_*^2=(1-\mu)C_2/(\bar{C}_1+1)$ which yields

$$\theta^2\mathcal{E}^2_{\mathcal{T}}(U,\mathcal{I})\le \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{R}).$$

□

Combining this result with properties of the approximation class $\mathbb{A}_s$ we are in the position to bound the cardinality of $\mathcal{M}_k$ in terms of $|u,f|_s^{1/s}$ and the total error.

Proposition 3.11 (Cardinality of $\mathcal{M}_k$) *Let θ be the marking parameter of AFEM and let $\mu=1-\theta^2/\theta_*^2>0$ as in Lemma 3.10. If $(u,f)\in\mathbb{A}_s$ then*

$$\#\mathcal{M}_k\lesssim \mu^{-1/(2s)}|u,f|_s^{1/s}\left(|\!|\!|u-U_k|\!|\!|^2+\operatorname{osc}^2_k(U_k,\mathcal{I}_k)\right)^{-1/(2s)}\qquad\forall k\ge 0.$$

Proof $\boxed{1}$ We set

$$\varepsilon^2 := \frac{\mu}{C_5}\Big(|||u - U_k|||^2 + \operatorname{osc}_k^2(U_k, \mathcal{T}_k)\Big).$$

Since $(u, f) \in \mathbb{A}_s$, in view of (3.14) there exists $\mathcal{T}_\varepsilon \in \mathbb{T}$ and $U_\varepsilon \in \mathbb{V}(\mathcal{T}_\varepsilon)$ such that

$$|||u - U_\varepsilon|||^2 + \operatorname{osc}^2_{\mathcal{T}_\varepsilon}(U_\varepsilon, \mathcal{I}_\varepsilon) \le \varepsilon^2 \quad \text{and} \quad \#\mathcal{T}_\varepsilon - \#\mathcal{T}_0 \lesssim |u, f|_s^{1/s}\, \varepsilon^{-1/s}. \tag{3.18}$$

To relate the total error with respect to U and U_ε we introduce the overlay $\mathcal{T}_* = \mathcal{T}_k \oplus \mathcal{T}_\varepsilon$, which is is the smallest common refinement of $\mathcal{T}_k$ and $\mathcal{T}_\varepsilon$, i.e., $\mathcal{T}_* \in \mathbb{T}$ with minimal cardinality such that $\mathcal{T}_k, \mathcal{T}_\varepsilon \le \mathcal{T}_*$. The cardinality of $\mathcal{T}_*$ can be estimated by

$$\#\mathcal{T}_* \le \#\mathcal{T}_k + \#\mathcal{T}_\varepsilon - \#\mathcal{T}_0; \tag{3.19}$$

compare for instance [14, Lemma 3.7].

$\boxed{2}$ Let U_* be the Ritz-projection in $\mathbb{V}(\mathcal{T}_*)$. Since $\mathcal{T}_\varepsilon \le \mathcal{T}_*$ nesting of spaces implies $\mathbb{V}(\mathcal{T}_\varepsilon) \subset \mathbb{V}(\mathcal{T}_*)$. Therefore, the best approximation property of U_* with respect to $|||\cdot|||$ and (3.7b) yield

$$\begin{aligned} |||u - U_*|||^2 + \operatorname{osc}^2_{\mathcal{T}_*}(f, \mathcal{T}_*) &\le C_5 \left(|||u - U_\varepsilon|||^2 + \operatorname{osc}^2_{\mathcal{T}_\varepsilon}(f, \mathcal{T}_\varepsilon)\right) \\ &\le C_5\, \varepsilon^2 = \mu \left(|||u - U_k|||^2 + \operatorname{osc}^2_k(U_k, \mathcal{T}_k)\right), \end{aligned}$$

i.e., the total error over $\mathcal{T}_*$ reduces by a factor μ relative to that one over $\mathcal{T}_k$. Upon applying Lemma 3.10 we therefore conclude that the set $\mathcal{R}_k = \mathcal{R}_{\mathcal{T}_k \to \mathcal{T}_*}(\mathcal{I}_k)$ of refined indices satisfies a Dörfler property (3.8) with parameter $\theta < \theta_*$.

$\boxed{3}$ Since MARK selects a minimal set $\mathcal{M}_k \subset \mathcal{I}_k$ satisfying this property we deduce

$$\#\mathcal{M}_k \le \#\mathcal{R}_k \lesssim \#\mathcal{T}_* - \#\mathcal{T}_k \le \#\mathcal{T}_\varepsilon - \#\mathcal{T}_0 \lesssim |u, f|_s^{1/s}\, \varepsilon^{-1/s},$$

where we have employed (3.19). Recalling the definition of ε we have proved the assertion. □

3.3.4 Optimal decay rate

We are ready to prove Main Result 2. It is an immediate consequence of the following theorem, which combines the strict error reduction of Corollary 3.7 with Proposition 3.11.

Theorem 3.12 (Optimal Decay Rate) *If $(u, f) \in \mathbb{A}_s$ then iteration (1.1) generates a sequence $\{\mathcal{T}_k, \mathbb{V}_k, U_k\}_{k \ge 0}$ with an optimal decay rate for the total error and the* *estimator, this is*

$$\left(\|\!|u - U_k\|\!|^2 + \mathrm{osc}_k^2(U_k, \mathcal{T}_k)\right)^{1/2} \lesssim \mu^{-1/2}|u, f|_s\,(\#\mathcal{T}_k - \#\mathcal{T}_0)^{-s} \qquad \forall\, k \geq 0$$

and

$$\mathcal{E}_k(U_k, \mathcal{T}_k) \lesssim \mu^{-1/2}|u, f|_s\,(\#\mathcal{T}_k - \#\mathcal{T}_0)^{-s} \qquad \forall\, k \geq 0.$$

Proof $\boxed{1}$ In this proof we use the notation $\mathcal{E}_k = \mathcal{E}_k(U_k, \mathcal{T}_k)$ and $\mathrm{osc}_k = \mathrm{osc}_k(f, \mathcal{T}_k)$. Since oscillation is dominated by the estimator (3.7a) we conclude in combination with the lower bound (3.4b)

$$\|\!|U_\ell - u\|\!|^2 + \mathrm{osc}_\ell^2 \leq \|\!|U_\ell - u\|\!|^2 + \mathcal{E}_\ell^2 \leq \left(1 + C_2^{-1}\right)\left(\|\!|U_\ell - u\|\!|^2 + \mathrm{osc}_\ell^2\right).$$

We utilize error reduction property of Corollary 3.7 to proceed for $0 \leq \ell < k$ by

$$\begin{aligned}\|\!|U_k - u\|\!|^2 + \mathrm{osc}_k^2 \leq \|\!|U_k - u\|\!|^2 + \mathcal{E}_k^2 &\leq \Lambda_1^2\alpha^{2(k-\ell)}\left(\|\!|U_\ell - u\|\!|^2 + \mathcal{E}_\ell^2\right)\\ &\leq \Lambda_1^2(1 + C_2^{-1})\,\alpha^{2(k-\ell)}\left(\|\!|U_\ell - u\|\!|^2 + \mathrm{osc}_\ell^2\right).\end{aligned}$$

$\boxed{2}$ From of Lemma 3.3 and Proposition 3.11 we therefore conclude

$$\begin{aligned}\#\mathcal{T}_k - \#\mathcal{T}_0 &\lesssim \sum_{\ell=0}^{k-1}\#\mathcal{M}_\ell \lesssim \mu^{-1/(2s)}|u, f|_s^{1/s}\sum_{\ell=0}^{k-1}\left(\|\!|U_\ell - u\|\!|^2 + \mathrm{osc}_\ell^2\right)^{-1/(2s)}\\ &\lesssim \mu^{-1/(2s)}|u, f|_s^{1/s}\left(\|\!|U_k - u\|\!|^2 + \mathrm{osc}_k^2\right)^{-1/(2s)}\sum_{\ell=0}^{k-1}\alpha^{(k-\ell)/s}\\ &\leq C(\alpha)\mu^{-1/(2s)}|u, f|_s^{1/s}\left(\|\!|U_k - u\|\!|^2 + \mathrm{osc}_k^2\right)^{-1/(2s)},\end{aligned}$$

where we have used boundedness of the geometric series

$$\sum_{\ell=0}^{k-1}\alpha^{(k-\ell)/s} = \sum_{j=1}^{k}\alpha^{j/s} < \sum_{j=1}^{\infty}\alpha^{j/s} =: C(\alpha) < \infty,$$

since $\alpha < 1$. This shows the first assertion. The second assertion follows immediately once again using the lower bound (3.4b). □

4 Extensions: robust estimators

In this section we apply our theory to two variants of the Poisson problem, namely a diffusion problem with jumping diffusion parameter and a reaction dominated diffusion-reaction problem. The problems are included in the theory presented in [14] relying on the standard residual estimator, which is not robust with respect to the conditions of the problems. The result in [14] implies asymptotically an optimal decay rate. However, the non-robustness of the estimator leads to a degeneration of the maximal marking parameter θ_* in (3.9) as well as the constant in the complexity result with increasing condition number.

For both problem robust estimators are known [29,40]. Robustness means that the constants C_1 and C_2 of the upper and lower bound are independent of the condition. Relying on such robust estimators we analyze the adaptive iteration with main focus on robustness in the maximal marking parameter θ_*.

4.1 Discontinuous coefficients

We consider the diffusion problem

$$-\operatorname{div} a(x)\nabla u = f \quad \text{in } \Omega, \;\; u = 0 \quad \text{on } \partial\Omega. \tag{4.1}$$

with a diffusion coefficient $a\colon \Omega \to \mathbb{R}$ that is strictly positive and bounded, i.e., there exists $0 < a_* \le a^* < \infty$ with

$$a_* \le a \le a^* \qquad \text{in } \Omega.$$

In addition, we assume that a is piecewise constant over some initial triangulation $\mathcal{T}_0$ and that the jumps of a are quasi monotone with respect to the triangulation $\mathcal{T}_0$. This last condition is preserved during refinement if $d = 2$, whereas in $3d$ we suppose that this condition is preserved during refinement. For more details on this topic see [11,29,28] and the references therein.

This problem gives rise to the following intrinsic scalar product with corresponding energy norm on $\mathbb{V}$:

$$\mathcal{B}[v,\, w] := \int_\Omega a\, \nabla v \cdot \nabla w \, dx, \qquad |||v||| := \mathcal{B}[v,\, v]^{1/2}.$$

We next recall the robust estimator derived by Petzold [29]. For $T \in \mathcal{T}$ and $\sigma \in \mathcal{S}$ define the weights

$$\beta_T := h_T a_{|T}^{-1/2} \quad \text{and} \quad \beta_\sigma := h_\sigma \max\{a_{|T} \mid T \subset \omega_\sigma\}^{-1}$$

with $h_T := \operatorname{diam}(T)$ and $h_\sigma := \operatorname{diam}(\sigma)$. The local indicators are then given by

$$\mathcal{E}_{\mathcal{T}}^2(U, T) := \beta_T^2 \|f\|_{L^2(T)}^2 + \sum_{\sigma \subset \partial T} \beta_\sigma \|J(U)\|_{L^2(\sigma)}^2.$$

In this case $J(U)_{|\sigma} = [\![a\, \nabla U]\!]_{|\sigma}$ denotes the normal flux of $a\, \nabla U$ across interior sides $\sigma \in \mathcal{S}$ and 0 if σ is a boundary side.

Petzold has shown in [29], that there are constants C_1, C_2 that are independent of the global condition number $\kappa = a^*/a_*$ of the problem such that

$$|||u-U|||^2 \le C_1 \sum_{T\in\mathcal{T}} \mathcal{E}^2_{\mathcal{T}}(U,T) = C_1 \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{T}) \tag{4.2a}$$

$$C_2 \mathcal{E}^2_{\mathcal{T}}(U,\mathcal{T}) \le |||u-U|||^2 + \operatorname{osc}^2_{\mathcal{T}}(f,\mathcal{T}) := |||u-U|||^2 + \sum_{T\in\mathcal{T}} \operatorname{osc}^2_{\mathcal{T}}(f,T). \tag{4.2b}$$

Local data oscillation on T is defined by the mean value $f_T := \frac{1}{|T|}\int_T f\,dx$ of f as

$$\operatorname{osc}^2_{\mathcal{T}}(f,T) := \beta^2_T \|f-f_T\|^2_{L^2(T)}.$$

The robustness of the estimator with respect to κ relies on suitable interpolation estimates utilizing the energy norm $|||\cdot|||$. Hereby, the quasi-monotonicity of the jumps of a with respect to the $\mathcal{T}$ is important; compare with [29, Lemma 4.9].

Problem (4.1) is included in the theory in [14] using the *non-robust* residual estimator $\hat{\mathcal{E}}_{\mathcal{T}}$ that is built in this case from the indicators

$$\hat{\mathcal{E}}^2_{\mathcal{T}}(U,T) := \|h_{\mathcal{T}} f\|^2_{L^2(T)} + \|h^{1/2}_{\mathcal{T}} J(U)\|^2_{L^2(\partial T)}, \qquad T\in\mathcal{T}.$$

Using the bounds for a we therefore have the element by element equivalence

$$a_*\,\mathcal{E}^2_{\mathcal{T}}(U,T) \lesssim \hat{\mathcal{E}}^2_{\mathcal{T}}(U,T) \qquad \text{and} \qquad \hat{\mathcal{E}}^2_{\mathcal{T}}(U,T) \lesssim a^*\,\mathcal{E}^2_{\mathcal{T}}(U,T) \tag{4.3}$$

where we explicitly traced the dependence on a^* and a_*. This implies the error reduction property for the sum of error and estimator according to Corollary 3.7.

Moreover, the local equivalence (4.3) allows us to employ Lemma 3.5, i.e., the robust estimator $\mathcal{E}_{\mathcal{T}}$ of Petzold satisfies the localized upper bound (3.5). Consequently, Theorem 3.12 implies an optimal decay rate for true error and estimator.

However, establishing the localized upper bound for a robust estimator via the equivalence to a non-robust estimator results in a maximal Dörfler marking parameter θ_* that degenerates with the condition of the problem. To be more precise, the explicit constants in (4.3) gives $\bar{C}_1 = \hat{C}_1\, C_4 \approx \kappa$ since $\hat{C}_1 \approx a_*^{-1}$. This in turn implies $\theta^2_* = C_2/(\bar{C}_1+1) \approx C_2/(\kappa+1)$, which gets small for large condition numbers κ. Such a restriction for the choice of the marking parameter θ when using a robust estimator is not acceptable.

It is therefore of utter importance to derive a localized upper bound with a constant $\bar{C}_1$ that is independent of the condition κ. We sketch the proof of such a bound in Sect. 4.3. Consequently the restriction of the marking parameter θ_* is robust with respect to κ. Utilizing the equivalence of $\mathcal{E}_{\mathcal{T}}$ and $\hat{\mathcal{E}}_{\mathcal{T}}$ the constants Λ_1 and Λ_2 of the main results depend on κ. We next comment on this.

Remark 4.1 The constant $\Lambda_1 = \max\{1,\gamma C_4\}^{1/2}/\min\{1,\gamma C_3\}^{1/2}$ in Corollary 3.7 obviously depends on $C_4/C_3 \approx a^*/a_* = \kappa$. It additionally depends on κ via γ; compare with [14, Theorem 4.1]. Moreover, the contraction constant α from Theorem 3.6

degenerates with κ getting large. Both constants Λ_1 and α enter in the constant Λ_2 of Theorem 3.12, the first one linearly, the latter one via the geometric series $\sum_{j=1}^{\infty} \alpha^{j/s}$.

Replacing on T the local mesh-sizes $h_T = \operatorname{diam}(T)$ and $h_\sigma = \operatorname{diam}(\sigma)$ in the definition of the weights β_T and β_σ by $|T|^{1/d} = h_{\mathcal{T}|T}$, it is easy to see that the resulting estimator fits into the framework considered in [14] and we deduce Theorems 3.6 and 3.12 directly. However, the constant Λ_2 still depends on κ due to the techniques used in [14] that employ both the H^1 norm and the energy norm.

For the particular problem at hand, we can avoid the H^1 norm by replacing [14, Propositon 3.3] utilizing the following robust estimate for two discrete functions $V, W \in \mathbb{V}(\mathcal{T})$ that only involves the energy norm:

$$\mathcal{E}^2_{\mathcal{T}}(V, T) \leq \mathcal{E}^2_{\mathcal{T}}(W, T) + C \sum_{\sigma \in \partial T} |||V - W|||^2_{\omega_\sigma} .$$

This modification in turn allows for a proof of an optimal decay rate with a constant Λ_2 that is robust with respect to the condition κ.

4.2 A singularly perturbed reaction-diffusion problem

We investigate the singularly perturbed reaction-diffusion equation

$$-\Delta u + \kappa\, u = f \quad \text{in } \Omega, \qquad u = 0 \quad \text{on } \partial\Omega, \tag{4.4}$$

where $\kappa \gg 1$ is the condition of the problem. The variational formulation of the problem leads to an intrinsic scalar product on $\mathbb{V}$ with corresponding energy norm:

$$\mathcal{B}[v,\ w] := \int_\Omega \nabla v \cdot \nabla w + \kappa\, v w \, dx, \qquad |||v||| := \mathcal{B}[v,\ v]^{1/2}.$$

Verfürth has derived for this problem an estimator that is robust in κ [40]. For $T \in \mathcal{T}$ and $\sigma \in \mathcal{S}$ we define the weights

$$\beta_T := \min\{h_T, \kappa^{-1/2}\} \qquad \text{and} \qquad \beta_\sigma := \min\{h_\sigma, \kappa^{-1/2}\},$$

where the local mesh-sizes h_T and h_σ are defined as in Sect. 2.2.2. The indicators are then defined as

$$\mathcal{E}^2_{\mathcal{T}}(U, T) := \beta_T^2 \, \| f - \kappa\, U \|^2_{L^2(T)} + \sum_{\sigma \subset \partial T} \beta_\sigma \, \| J(U) \|^2_{L^2(\sigma)}.$$

Thereby, $J(U) = [\![\nabla U]\!]$ is the normal flux of ∇U over interior sides and $J(U) = 0$ for boundary sides. Thanks to the choice of the weights β_T, β_σ and a suitable interpolation operator [40, Proposition 4.1] there exist constants C_1 and C_2 independent of κ such that

$$|\!|\!| u - U |\!|\!|^2 \le C_1 \sum_{T\in\mathcal{T}} \mathcal{E}^2_{\mathcal{T}}(U, T) =: \mathcal{E}^2_{\mathcal{T}}(U, \mathcal{T}) \tag{4.5a}$$

$$C_2\mathcal{E}^2_{\mathcal{T}}(U, \mathcal{T}) \le |\!|\!| u - U |\!|\!|^2 + \operatorname{osc}^2_{\mathcal{T}}(f, \mathcal{T}) := |\!|\!| u - U |\!|\!|^2 + \sum_{T\in\mathcal{T}} \operatorname{osc}^2_{\mathcal{T}}(f, T). \tag{4.5b}$$

For this problem we let $f_T \in \mathbb{P}_1$ be the L^2-projection of f onto the affine functions over T which results in the *data* oscillation term

$$\operatorname{osc}^2_{\mathcal{T}}(f, T) := \beta_T^2 \|f - f_T\|^2_{L^2(T)};$$

compare with [40, Proposition 4.1].

The reaction-diffusion problem (4.4) is also included in the theory in [14] relying on the *non-robust* standard residual estimator defined from the indicators

$$\hat{\mathcal{E}}^2_{\mathcal{T}}(U, T) := \|h_{\mathcal{T}}(f - \kappa\, U)\|^2_{L^2(T)} + \|h^{1/2}_{\mathcal{T}} J(U)\|^2_{L^2(\partial T)} \qquad T \in \mathcal{T}.$$

Obviously, the estimators $\hat{\mathcal{E}}_{\mathcal{T}}$ and $\mathcal{E}_{\mathcal{T}}$ are locally equivalent with a constant depending on the condition κ. To be more precise: In case $h_T \le \kappa^{-1/2}$ it holds $\beta_T = h_T \approx h_{\mathcal{T}|T}$ and for $h_T > \kappa^{-1/2}$ we estimate $\kappa^{-1/2} = \beta_T \le h_T \lesssim \beta_T(\kappa^{1/2} h_{\mathcal{T}|T})$. The same estimates apply to β_σ with $\sigma \subset T$, whence

$$\mathcal{E}^2_{\mathcal{T}}(U, T) \lesssim \hat{\mathcal{E}}^2_{\mathcal{T}}(U, T) \le C_4 \mathcal{E}^2_{\mathcal{T}}(U, T) \tag{4.6}$$

with $C_4 \approx \|h_{\mathcal{T}_0}\|^2_\infty \kappa$. This is the local equivalence of indicators (3.6), which in turn yields the reduction property as stated in Corollary 3.7.

In Sect. 4.3 we derive the localized upper bound (3.5) for $\mathcal{E}_{\mathcal{T}}$ with a constant $\bar{C}_1$ independent of κ. This again implies an optimal decay rate in terms of DOFs for both true error and estimator with a maximal marking parameter θ_* in (3.9), which is robust in κ.

The constants Λ_1 and Λ_2 of the main results do depend on κ; compare with Remark 4.1. For this problem the techniques from [14] do not apply to the estimator $\mathcal{E}_{\mathcal{T}}$ since the weights β_T and β_σ are not *strictly* monotone during refinement. In view of this, robustness of the constants Λ_1 and Λ_2 with respect to κ is an open problem and deserves future investigation.

4.3 Localized upper bound

Given $\mathcal{T} \le \mathcal{T}_*$ we aim for a robust estimate of the error $E_* = U_* - U$ between the two Ritz approximations $U \in \mathbb{V}(\mathcal{T})$ and $U_* \in \mathbb{V}(\mathcal{T}_*)$ involving only the error indicators of the refined elements $\mathcal{R} = \mathcal{R}_{\mathcal{T}\to\mathcal{T}_*}(\mathcal{T}) = \mathcal{T} \setminus \mathcal{T}_*$. The key ingredient in the proof of the localized upper bound is an interpolation operator $\Pi_{\mathcal{T}}$ that localizes $E_* - \Pi_{\mathcal{T}} E_*$ to the set of refined elements; see Sect. 3.2.

We observe that the estimators of Sects. 4.1 and 4.2 have a similar structure, namely

$$\mathcal{E}^2_{\mathcal{T}}(U, T) = \beta_T^2 \| R(U) \|^2_{L^2(T)} + \sum_{\sigma \subset \partial T} \beta_\sigma \| J(U) \|^2_{L^2(\sigma)} \qquad T \in \mathcal{T},$$

with the interior residual $R(U) \in L^2(T)$, the jump residual $J(U) \in L^2(\sigma)$, and the respective weights β_T, β_σ introduced in Sects. 4.1 and 4.2. We call an interpolation operator $\Pi_{\mathcal{T}} : \mathbb{V} \to \mathbb{V}(\mathcal{T})$ robust with respect to the condition κ of the problem if there exists a constant $\bar{C}$ independent of κ such that

$$\sum_{T \in \mathcal{T}} \left\{ \beta_T^{-2} \| v - \Pi_{\mathcal{T}} v \|^2_{L^2(T)} + \sum_{\sigma \subset T} \beta_\sigma^{-1} \| v - \Pi_{\mathcal{T}} v \|^2_{L^2(\sigma)} \right\} \le \bar{C} \, |\!|\!| v |\!|\!|^2 . \tag{4.7}$$

Utilizing modifications of the Clément interpolant, such interpolation estimates are proved in [29, Lemma 4.9] for the discontinuous coefficient problem from Sect. 4.1 and in [40, Proposition 4.1] for the reaction-diffusion problem of Sect. 4.2. The modified Clément interpolants preserve locally discrete functions but do not strictly localize to the set of refined elements $\mathcal{R}$. However, it holds for any $V_* \in \mathbb{V}(\mathcal{T}_*)$

$$V_* - \Pi_{\mathcal{T}} V_* \equiv 0 \quad \text{in } T \qquad \text{if } T \cap T' = \emptyset \quad \text{for all } T' \in \mathcal{R}.$$

Therefore, enlarging $\mathcal{R}$ by one ring of additional elements around $\mathcal{R}$, namely

$$\overline{\mathcal{R}} = \overline{\mathcal{R}}_{\mathcal{T} \to \mathcal{T}_*} := \{ T \in \mathcal{T} \mid T \cap T' \neq \emptyset \text{ for some } T' \in \mathcal{R} \},$$

we see $V_* - \Pi_{\mathcal{T}} V_* \equiv 0$ for all $T \in \mathcal{T} \setminus \overline{\mathcal{R}}_{\mathcal{T} \to \mathcal{T}_*}$. Utilizing Galerkin orthogonality and standard arguments we therefore conclude

$$\begin{aligned}
|\!|\!| E_* |\!|\!|^2 &= \mathcal{B}[E_*,\ E_* - \Pi_{\mathcal{T}} E_*] \\
&= \sum_{T \in \overline{\mathcal{R}}} \langle r_T,\ E_* - \Pi_{\mathcal{T}} E_* \rangle_T - \tfrac{1}{2} \langle J(U),\ E_* - \Pi_{\mathcal{T}} E_* \rangle_{\partial T} \\
&\le \sum_{T \in \overline{\mathcal{R}}} \Bigg\{ \beta_T \| r_T \|_{L^2(T)} \, \beta_T^{-1} \| E_* - \Pi_{\mathcal{T}} E_* \|_{L^2(T)} \\
&\qquad + \sum_{\sigma \subset T} \beta_\sigma^{1/2} \| J_k(U) \|_{L^2(\sigma)} \, \beta_\sigma^{-1/2} \| E_* - \Pi_{\mathcal{T}} E_* \|_{L^2(\sigma)} \Bigg\} \\
&\le \bar{C} \mathcal{E}_{\mathcal{T}}(U, \overline{\mathcal{R}}) \, |\!|\!| E_* |\!|\!| ,
\end{aligned} \tag{4.8}$$

using (4.7) in the last step. Upon replacing $\mathcal{R}$ by $\overline{\mathcal{R}}$ in (3.5) we have shown the localized upper bound with $\bar{C}_1$ bounded by the interpolation constant $\bar{C}$ of $\Pi_{\mathcal{T}}$ that is robust in κ. The constant is the same as in the upper bound (4.2a) respectively (4.5a).

We would like to remark that a complete localization to $\mathcal{R}$ as discussed in Sect. 3.2 by first using the interpolation operator $P_{\mathcal{R}}$ of Lemma 3.4 is out of question. This operator is only stable with respect to the H^1 norm and not with respect to the energy norm. Consequently, any use of $P_{\mathcal{R}}$ would result in a κ dependent constant.

The replacement of $\mathcal{R}$ by $\overline{\mathcal{R}}$ in (3.5) does not prevent us from proving Theorem 3.12. This can be seen as follows. Lemma 3.10 still holds true if we also substitute $\mathcal{R}$ by $\overline{\mathcal{R}}$. This Lemma comes into play in Step $\boxed{3}$ of the proof to Proposition 3.11. In particular, adopting the notation of the proof, we first use shape-regularity to bound $\#\overline{\mathcal{R}} \lesssim \#\mathcal{R}$ and then proceed by

$$\#\mathcal{M} \le \#\mathcal{R}_k \le \#\overline{\mathcal{R}}_k \lesssim \#\mathcal{R}_k \le |u, f|_s^{1/s}\, \varepsilon^{-1/s},$$

which proves the proposition. This in turn yields Theorem 3.12.

References

1. Ainsworth, M., Oden, J.T.: A posteriori error estimation in finite element analysis. Wiley-Interscience, New York (2000)
2. Bänsch, E.: Local mesh refinement in 2 and 3 dimensions. IMPACT Comput. Sci. Eng. **3**, 181–191 (1991)
3. Binev, P., Dahmen, W., DeVore, R.: Adaptive finite element methods with convergence rates. Numer. Math. **97**, 219–268 (2004)
4. Binev, P., Dahmen, W., DeVore, R., Petrushev, P.: Approximation classes for adaptive methods. Serdica Math. J. **28**(4), 391–416 (2002). Dedicated to the memory of Vassil Popov on the occasion of his 60th birthday. MR MR1965238 (2004b:65176)
5. Bornemann, F.A., Erdmann, B., Kornhuber, R.: A posteriori error estimates for elliptic problems in two and three space dimensions. SIAM J. Numer. Anal. **33**(3), 1188–1204 (1996)
6. Braess, D., Pillwein, V., Schöberl, J.: Equilibrated residual error estimates are p-robust. Comput. Methods Appl. Mech. Eng. **198**(13–14), 1189–1197 (2009)
7. Bangerth, W., Rannacher, R.: Adaptive finite element methods for differential equations. In: Lectures in Mathematics, ETH Zürich. Birkhäuser, Basel (2003)
8. Braess, D.: Finite Elements, 3rd edn. Cambridge University Press, Cambridge (2007). Theory, fast solvers, and applications in elasticity theory. Translated from German by Larry L. Schumaker
9. Babuška, I., Strouboulis, T.: The finite element method and its reliability. In: Numerical Mathematics and Scientific Computation. The Clarendon Press, New York (2001)
10. Braess, D., Schöberl, J.: Equilibrated residual error estimator for edge elements. Math. Comput. **77**(262), 651–672 (2008)
11. Bernardi, C., Verfürth, R.: Adaptive finite element methods for elliptic equations with non-smooth coefficients. Numer. Math. **85**(4), 579–608 (2000)
12. Carstensen, C., Bartels, S.: Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. Part I: low order conforming, nonconforming, and mixed FEM. Math. Comput. **71**, 945–969 (2002)
13. Chen, Z., Feng, J.: An adaptive finite element algorithm with reliable and efficient error control for linear parabolic problems. Math. Comput. **73**, 1167–1193 (2004)
14. Cascón, J.M., Kreuzer, C., Nochetto, R.H., Siebert, K.G.: Quasi-optimal convergence rate for an adaptive finite element method. SIAM J. Numer. Anal. **46**(5), 2524–2550 (2008)
15. Clément, P.: Approximation by finite element functions using local regularization. RAIRO **9**, 77–84 (1975)
16. Diening, L., Kreuzer, Ch.: Convergence of an adaptive finite element method for the p-Laplacian equation. SIAM J. Numer. Anal. **46**(2), 614–638 (2008)
17. Dörfler, W., Nochetto, R.H.: Small data oscillation implies the saturation assumption. Numer. Math. **91**(1), 1–12 (2002)

18. Dörfler, W.: A convergent adaptive algorithm for Poisson's equation. SIAM J. Numer. Anal. **33**, 1106–1124 (1996)
19. Gilbarg, D., Trudinger, N.S.: Elliptic partial differential equations of second order. In: Classics in Mathematics. Springer, New York (2001)
20. Kossaczký, I.: A recursive approach to local mesh refinement in two and three dimensions. J. Comput. Appl. Math. **55**, 275–288 (1994)
21. Maubach, J.M.: Local bisection refinement for n-simplicial grids generated by reflection. SIAM J. Sci. Comput. **16**, 210–227 (1995)
22. Mekchay, K., Nochetto, R.H.: Convergence of adaptive finite element methods for general second order linear elliptic PDEs. SIAM J. Numer. Anal. **43**(5), 1803–1827 (2005)
23. Morin, P., Nochetto, R.H., Siebert, K.G.: Data oscillation and convergence of adaptive FEM. SIAM J. Numer. Anal. **38**, 466–488 (2000)
24. Morin, P., Nochetto, R.H., Siebert, K.G.: Convergence of adaptive finite element methods. SIAM Rev. **44**, 631–658 (2002)
25. Morin, P., Nochetto, R.H., Siebert, K.G.: Local problems on stars: a posteriori error estimators, convergence, and performance. Math. Comput. **72**, 1067–1097 (2003)
26. Morin, P., Siebert, K.G., Veeser, A.: A basic convergence result for conforming adaptive finite elements. Math. Models Methods Appl. **18**, 707–737 (2008)
27. Nochetto, R.H., Siebert, K.G., Veeser, A.: Theory of adaptive finite element methods: an introduction. In: DeVore, R.A., Kunoth, A. (eds.) Multiscale, Nonlinear and Adaptive Approximation. Springer, New York, pp. 409–542 (2009)
28. Petzoldt, M.: Regularity and error estimators for elliptic equations with discontinuous coefficients. Ph.D. thesis, FU Berlin (2001)
29. Petzoldt, M.: A posteriori error estimators for elliptic equations with discontinuous coefficients. Adv. Comput. Math. **16**, 47–75 (2002)
30. Prager, W., Synge, J.L.: Approximations in elasticity based on the concept of function space. Quart. Appl. Math. **5**, 241–269 (1947)
31. Siebert, K.G.: A convergence proof for adaptive finite elements without lower bound. IMA J. Numer. Anal. doi:10.1093/imanum/drq001 (published online) (2010, May)
32. Schmidt, A., Siebert, K.G.: Design of adaptive finite element software. The finite element toolbox ALBERTA. In: Lecture Notes in Computational Science and Engineering, vol. 42. Springer, New York (2005)
33. Stevenson, R.: Optimality of a standard adaptive finite element method. Found. Comput. Math. **7**(2), 245–269 (2007)
34. Stevenson, R.: The completion of locally refined simplicial partitions created by bisection. Math. Comput. **77**(261), 227–241 (2008)
35. Siebert, K.G., Veeser, A.: A unilaterally constrained quadratic minimization with adaptive finite elements. SIAM J. Optim. **18**(1), 260–289 (2007)
36. Scott, L.R., Zhang, S.: Finite element interpolation of nonsmooth functions satisfying boundary conditions. Math. Comput. **54**(190), 483–493 (1990)
37. Traxler, C.T.: An algorithm for adaptive mesh refinement in n dimensions. Computing **59**, 115–137 (1997)
38. Veeser, A.: Convergent adaptive finite elements for the nonlinear Laplacian. Numer. Math. **92**(4), 743–770 (2002)
39. Verfürth, R.: A review of a posteriori error estimation and adaptive mesh-refinement techniques. Adv. Numer. Math. John Wiley, Chichester (1996)
40. Verfürth, R.: Robust a posteriori error estimators for a singularly perturbed reaction-diffusion equation. Numer. Math. **78**(3), 479–493 (1998)
41. Veeser, A., Verfürth, R.: Explicit upper bounds for dual norms of residuals. SIAM J. Numer. Anal. (2009, to appear)
42. Zienkiewicz, O.C., Zhu, J.Z.: A simple error estimator and adaptive procedure for practical engineering analysis. Int. J. Numer. Methods Eng. **24**, 337–357 (1987)