SIAM J. NUMER. ANAL.
Vol. 55, No. 6, pp. 2644–2665

AXIOMS OF ADAPTIVITY WITH SEPARATE MARKING FOR DATA RESOLUTION*

C. CARSTENSEN† AND H. RABUS†

Abstract. Mixed finite element methods with flux errors in $H(\mathrm{div})$-norms and div-least-squares finite element methods require a separate marking strategy in obligatory adaptive mesh-refining. The refinement indicator $\sigma^2(\mathcal{T},K) = \eta^2(\mathcal{T},K) + \mu^2(K)$ of a finite element domain K in an admissible triangulation $\mathcal{T}$ consists of some residual-based error estimator $\eta(\mathcal{T},K)$ with some reduction property under local mesh-refining and some data approximation error $\mu(K)$. Separate marking means either Dörfler marking if $\mu^2(\mathcal{T}) \leq \kappa\eta^2(\mathcal{T})$ or otherwise an optimal data approximation algorithm with controlled accuracy. The axioms are sufficient conditions on the estimators $\eta(\mathcal{T},K)$ and data approximation errors $\mu(K)$ for optimal asymptotic convergence rates. The enfolded set of axioms of this paper simplifies [C. Carstensen, M. Feischl, M. Page, and D. Praetorius, *Comput. Math. Appl.*, 67 (2014), pp. 1195–1253] for collective marking, treats separate marking established for the first time in an abstract framework, generalizes [C. Carstensen and E.-J. Park, *SIAM J. Numer. Anal.*, 53 (2015), pp. 43–62] for least-squares schemes, and extends [C. Carstensen and H. Rabus, *Math. Comp.*, 80 (2011), pp. 649–667] to the mixed finite element method with flux error control in $H(\mathrm{div})$. The paper gives an outline of the mathematical analysis for optimal convergence rates but also serves as a reference so that future contributions merely verify a few axioms in a new application in order to ensure optimal mesh-refinement of the adaptive algorithm.

Key words. adaptivity, finite element method, nonstandard finite element method, mixed finite element method, optimal convergence, least-squares finite element method

AMS subject classifications. 65N12, 65N15, 65N30, 65N50, 65Y20

DOI. 10.1137/16M1068050

1. Introduction.

The understanding of adaptive mesh-refinement for finite element methods (FEMs) has been one of the most important fields in the computational sciences with PDEs over the last two decades following early theoretical contributions due to Ivo Babuška et al. in the 1980s and starting in 2D with Willy Dörfler in 1997 before the first proofs of optimal rates appeared in [Ste07, CKNS08]; more detailed historic remarks and references can be found in [CFPP14] with a first abstract approach and four axioms for the collective marking strategy CAFEM. Those four axioms describe elementary properties of the total error estimator that are sufficient for optimal convergence rates and do not include local efficiency as this is not available for boundary element methods. Standard adaptive schemes are based on a total error estimator and collective marking on each level outlined in pseudocode as follows:

CAFEM$(\theta, \mathcal{T}_0)$
  for $\ell = 0, 1, \ldots$ **do**
    COMPUTE $\sigma_\ell(K)$ for all $K \in \mathcal{T}_\ell$
    $\mathcal{T}_{\ell+1} :=$ `Dörfler_marking`$(\theta, \sigma_\ell(K) : K \in \mathcal{T}_\ell)$
  end for

*Received by the editors March 30, 2016; accepted for publication (in revised form) August 10, 2017; published electronically November 2, 2017.
http://www.siam.org/journals/sinum/55-6/M106805.html
Funding: The work of the authors was supported by the Deutsche Forschungsgemeinschaft in the Priority Program 1748 *Reliable simulation techniques in solid mechanics. Development of non-standard discretization methods, mechanical and mathematical analysis* under the project *Foundation and application of generalized mixed FEM towards nonlinear problems in solid mechanics* (CA 151/22-1).

†Department of Mathematics, Humboldt-Universität zu Berlin, 10099 Berlin, Germany (cc@math.hu-berlin.de, rabus@math.hu-berlin.de).

In earlier contributions [Ste07, CKNS08], the preceding step COMPUTE σ_ℓ is realized by the call of SOLVE and ESTIMATE with respect to a triangulation $\mathcal{T}_\ell$ and for a given application at hand. This paper generalizes CAFEM and includes certain data approximation terms, which do not allow for a forthcoming reduction property (A2) in general. Mixed finite element schemes when the fluxes are measured in the $H(\mathrm{div})$ norm or least-squares FEMs [CDR16, BC17, BCS17] are examples where CAFEM is not successful and a separate marking is obligatory.

This paper simplifies the axioms from [CFPP14], also works without the concept of nonlinear approximation classes [BDdV04, Ste07, CKNS08], and so avoids any notion of efficiency. The focus of this paper is on adaptive finite element algorithms with separate marking (SAFEMs) for the separate data approximation. The proposed algorithm is a modification of the standard AFEM: Dörfler marking is applied if the estimated error dominates the data approximation error, while an optimal data approximation is performed otherwise; this is outlined in pseudocode as follows:

SAFEM($\theta_A, \kappa, \rho_B, \mathcal{T}_0$)
  for $\ell = 0, 1, \ldots$ **do**
    COMPUTE $\eta_\ell(K)$, $\mu(K)$ for all $K \in \mathcal{T}_\ell$
    if $\mu_\ell^2 := \mu^2(\mathcal{T}_\ell) \le \kappa\eta_\ell^2 = \kappa\eta_\ell^2(\mathcal{T}_\ell, \mathcal{T}_\ell)$ **then** // CASE (A)
      $\mathcal{T}_{\ell+1} := \texttt{Dörfler\_marking}(\theta_A, \eta_\ell(K) : K \in \mathcal{T}_\ell)$
    else // CASE (B)
      $\mathcal{T}_{\ell+1} := \mathcal{T}_\ell \oplus \texttt{appx}(\rho_B\mu_\ell^2, \mu(K) : K \in \mathcal{T}_0)$
    end if
  end for

AFEMs based on separate marking have been designed and established by several authors. The algorithm in [Ste07], for example, follows a similar idea with a routine RHS applied on each level ℓ for a prescribed tolerance $\mu_0/2^\ell$ in the controlled data approximation. The algorithm in [BM08] runs Dörfler marking also for the data approximation reduction, while [CDN12] suggests separate marking for data in $H^{-1}(\Omega)$.

The algorithm SAFEM of this paper combines ideas from [Ste07, BM08, CR11, Rab15] and distinguishes two cases, Cases (A) and (B), where the refinement is with respect to the dominant refinement indication η_ℓ^2 or μ_ℓ^2. The refinement in Case (B) depends on the data approximation error and is independent of the discrete solution. This allows for any optimal algorithm for data approximation: The output $\mathcal{T}_{\mathrm{Tol}} = \texttt{appx}(\mathrm{Tol}, \mu(K) : K \in \mathcal{T}_0)$ is expected to satisfy

$$\mu^2(\mathcal{T}_{\mathrm{Tol}}) \le \mathrm{Tol} \quad \text{and} \quad |\mathcal{T}_{\mathrm{Tol}}| - |\mathcal{T}_0| \le \Lambda_5 \, \mathrm{Tol}^{-1/s}\,.$$

Optimal convergence rates for the estimators follow from axioms (A1)–(A4) generalized from [CFPP14] and (B1)–(B2) for optimal data approximation with quasimonotonicity (QM). The results of this paper not only provide a guideline for nonexperts in asymptotic convergence theory but also allow them a direct reference to the axioms and reduce the work required to prove properties (A1)–(A4), (B1)–(B2), and (QM) in an example at hand. If the global efficiency of the estimator is known, rate-optimality of the corresponding error follows immediately.

The analysis for AFEMs based on collective marking as in [CFPP14] is included, when $\sigma^2(\mathcal{T}, \bullet) = \eta^2(\mathcal{T}, \bullet) + \mu^2(\mathcal{T}, \bullet)$ replaces $\eta^2(\mathcal{T}, \bullet)$ in Case (A) and the refinement indicator in Case (B) is set to zero. Thus, only Case (A) of SAFEM applies for collective marking, the axioms (A1)–(A4) are equivalent to those in [CFPP14], and (B1) and (B2) are automatically satisfied in this setting.

The subroutine `appx` in SAFEM can be realized by some Dörfler marking (similar to the algorithm in [BM08]) or by the algorithm APPROX from [BDdV04, BdV04] (applied in [CR11, Rab15]). However, the flexibility in the data reduction allows applications of SAFEM to problems with data approximation terms that do *not* satisfy an estimator reduction property but do satisfy *quasimonotonicity*. In those cases, Dörfler marking is not guaranteed to yield optimality. Two applications motivate the present paper, where the data approximation in the L^2 term cannot be avoided: mixed FEM with flux error estimation in $H(\mathrm{div})$ (rather than solely in $L^2(\Omega)$ from [CHX09, CR11]) and least-squares FEM. There already exist contributions based on the axioms of this paper [CP15, BC17, BCS17] (it is straightforward to re-interpret the first two; the third is already based on the precise axioms of this paper) for least-squares FEM. But it is also important for the application of mixed FEM to well-posed general linear second-order problems [CDNP16], for the low-order terms enforce the consideration of the divergence of the flux in L^2. The recent comprehensive a posteriori error analysis in [CPS16] provides an efficient and reliable control in natural norms: the error in the flux in $H(\mathrm{div},\Omega)$ and the error in the displacements in $L^2(\Omega)$.

The separate marking algorithms [Ste07, CDN12] may also be analyzed within the abstract framework of this paper.

The remaining parts of this paper are organized as follows. Section 2 presents more details on SAFEM and guides the reader through the conditions in (A1)–(A4) and (B1)–(B2) for the refinement indicators η and μ and asserts the optimal convergence rate of SAFEM in Theorem 2.1 with respect to σ. Together with efficiency, this automatically leads to rate-optimality of the error. A collection of remarks follows in section 3 before section 4 presents the proofs. Section 5 contains an application to mixed FEMs in $H(\mathrm{div})$, where separate marking is obligatory, with a discrete version (A3) of [CPS16].

The notation $A \lesssim B$ abbreviates $A \leq CB$ for some positive generic constant C, which depends only on the initial triangulation $\mathcal{T}_0$ and on the universal constants in the axioms; while $A \approx B$ abbreviates $A \lesssim B \lesssim A$. Throughout this paper standard notation of Lebesgue and Sobolev spaces and their norms applies. The modulus sign $|\bullet|$ denotes the Euclidean length, the area or volume of a domain, as well as the counting measure; e.g., $|\mathcal{M}|$ is the cardinality of $\mathcal{M}$ and equals the number of elements in a subset $\mathcal{M}$ of a triangulation.

2. Axioms and results. The axioms concern general conditions of the estimators η and μ, which play different roles in the adaptive algorithm and are based on the set $\mathbb{T}$ of admissible triangulations.

2.1. Partitions and admissible triangulations. Let $\mathcal{T}_0$ be a regular triangulation of the domain Ω into (tagged) n-simplices in $\mathbb{R}^n$. Any refinement $\mathcal{P}$ from $\mathcal{T}_0$ by the newest vertex bisection (NVB) is called a partition and is written $\mathcal{P} \in \mathbb{P}(\mathcal{T}_0) =: \mathbb{P}$. A partition $\mathcal{P} \in \mathbb{P}$, which is a regular triangulation in the sense of Ciarlet (sometimes called conforming triangulation), is called admissible and is written $\mathcal{P} \in \mathbb{T}(\mathcal{T}_0) =: \mathbb{T}$.

The input of the underlying refinement procedure $\mathcal{T}_{\mathrm{out}} := \mathrm{REFINE}(\mathcal{T}_{\mathrm{in}}, \mathcal{M})$ is an admissible triangulation $\mathcal{T}_{\mathrm{in}} \in \mathbb{T}$ and some subset $\mathcal{M} \subseteq \mathcal{T}_{\mathrm{in}}$ thereof; the output $\mathcal{T}_{\mathrm{out}} \in \mathbb{T}(\mathcal{T}_{\mathrm{in}})$ is the smallest admissible refinement of $\mathcal{T}_{\mathrm{in}}$ in which all $T \in \mathcal{M} \subseteq \mathcal{T}_{\mathrm{in}}$ are at least bisected, i.e., $\mathcal{M} \subseteq \mathcal{T}_{\mathrm{in}} \setminus \mathcal{T}_{\mathrm{out}}$. The procedure REFINE specifies the NVB with completion (to avoid hanging nodes, etc.), and more details may be found in [Ste08]. Given $\mathcal{T}, \mathcal{T}' \in \mathbb{T}$ its overlay $\mathcal{T} \oplus \mathcal{T}' \in \mathbb{T}(\mathcal{T}) \cap \mathbb{T}(\mathcal{T}')$ is the smallest common refinement of $\mathcal{T}$ and $\mathcal{T}'$.

2.2. Estimators and distance. The axioms are defined in terms of refinement indicators η and μ plus a global distance δ. For any admissible triangulation $\mathcal{T} \in \mathbb{T}$ and any element domain $K \in \mathcal{T}$ let $\eta(\mathcal{T}, K)$ and $\mu(K)$ be a nonnegative real number with squares $\eta^2(\mathcal{T}, K)$ and $\mu^2(K)$ and their sums

$$\eta^2(\mathcal{T}, \mathcal{M}) := \sum_{K\in\mathcal{M}} \eta^2(\mathcal{T}, K), \qquad \mu^2(\mathcal{M}) := \sum_{K\in\mathcal{M}} \mu^2(K) \qquad \text{for any } \mathcal{M} \subseteq \mathcal{T}. \tag{2.1}$$

The distance $\delta(\mathcal{T}, \hat{\mathcal{T}})$ of $\mathcal{T} \in \mathbb{T}$ and its refinement $\hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ is a nonnegative real and for simplicity can be regarded as the squared energy norm of the difference of the discrete solutions on $\mathcal{T}$ and $\hat{\mathcal{T}}$. The estimators are utilized in the adaptive algorithm and are linked with the distance function in the axioms below. The output of the adaptive algorithm is a sequence $\mathcal{T}_0, \mathcal{T}_1, \mathcal{T}_2, \dots$ of successive refinements that start with $\mathcal{T}_0$ and give rise to the abbreviations (with a subindex ℓ to refer to the triangulation as part of the output of SAFEM)

$$\eta_\ell(K) := \eta(\mathcal{T}_\ell, K) \quad \text{for } K \in \mathcal{T}_\ell \quad \text{and} \quad \eta_\ell := \eta(\mathcal{T}_\ell, \mathcal{T}_\ell) := \Bigg(\sum_{K\in\mathcal{T}_\ell} \eta_\ell^2(K)\Bigg)^{1/2}.$$

The sum $\sigma^2 := \eta^2 + \mu^2$ and their local variants are frequently utilized throughout this paper with $\sigma_\ell^2 := \eta_\ell^2 + \mu_\ell^2$ for $\mu_\ell^2 := \mu^2(\mathcal{T}_\ell) := \sum_{K\in\mathcal{T}_\ell} \mu^2(K)$.

2.3. Adaptive algorithm. In more detail, SAFEM calls SELECT and REFINE to realize the Dörfler marking in Case (A) from the introduction; more details on `appx` in Case (B) follow in subsection 3.3.

SAFEM($\theta_A, \kappa, \rho_B, \mathcal{T}_0$)
  Input: Initial coarse triangulation $\mathcal{T}_0$, $0 < \theta_A < 1$, $0 < \rho_B < 1$, $0 < \kappa$
  for $\ell = 0, 1, \dots$ **do**
    COMPUTE refinement indicators $\eta_\ell^2(K)$ and $\mu^2(K)$ for all $K \in \mathcal{T}_\ell$
    if $\mu_\ell^2 \le \kappa\eta_\ell^2$ **then** // CASE (A)
      SELECT a subset $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ of element domains of (almost) minimal cardinality with

$$\theta_A \eta_\ell^2 \le \eta_\ell^2(\mathcal{M}_\ell) := \sum_{K\in\mathcal{M}_\ell} \eta_\ell^2(K) \tag{2.2}$$

      COMPUTE $\mathcal{T}_{\ell+1} := \text{REFINE}(\mathcal{T}_\ell, \mathcal{M}_\ell)$
    else // CASE (B)
      RUN $\mathcal{T} = \texttt{appx}(\text{Tol}, \mu(K) : K \in \mathcal{T}_0)$ with $\text{Tol} = \rho_B \mu_\ell^2$
      COMPUTE $\mathcal{T}_{\ell+1} := \mathcal{T}_\ell \oplus \mathcal{T}$
    end if
  end for
  Output: $\mathcal{T}_k$, η_k, μ_k, $\sigma_k := \sqrt{\eta_k^2 + \mu_k^2}$ for $k = 0, 1, \dots$

The selection of $\mathcal{M}_\ell$ with almost minimal cardinality means that $|\mathcal{M}_\ell| \lesssim |\mathcal{M}_\ell^\star|$, where $\mathcal{M}_\ell^\star$ denotes some set of minimal cardinality with (2.2). The point is that this can be realized in linear CPU time [Ste07].

2.4. Axioms. The universal positive constants Λ_{ref}, Λ_1, Λ_2, Λ_3, Λ_4, Λ_6, and $\widehat{\Lambda}_3 \ge 0$ as well as $0 < \rho_2 < 1$ in the axioms (A1)–(A4), (B2), and (QM) below solely depend on $\mathbb{T}$ (whence merely on $\mathcal{T}_0$); the parameters $s > 0$ and Λ_5 in (B1) also depend on the algorithm `appx` and the optimal data approximation rate.

The axioms (A1)–(A3) and (B2) concern an arbitrary triangulation $\mathcal{T} \in \mathbb{T}$ and any refinement $\hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ of it, while (A4) solely concerns the outcome of SAFEM. Recall the sum conventions for $\eta(\mathcal{T}, \mathcal{M})$ and $\mu(\mathcal{T})$ in subsection 2.2.

(A1) Stability. $\forall \mathcal{T} \in \mathbb{T}$, $\forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$

$$\left| \eta(\hat{\mathcal{T}}, \mathcal{T} \cap \hat{\mathcal{T}}) - \eta(\mathcal{T}, \mathcal{T} \cap \hat{\mathcal{T}}) \right| \leq \Lambda_1 \delta(\mathcal{T}, \hat{\mathcal{T}}). \tag{A1}$$

(A2) Reduction. $\forall \mathcal{T} \in \mathbb{T}$, $\forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$

$$\eta(\hat{\mathcal{T}}, \hat{\mathcal{T}} \setminus \mathcal{T}) \leq \rho_2 \eta(\mathcal{T}, \mathcal{T} \setminus \hat{\mathcal{T}}) + \Lambda_2 \delta(\mathcal{T}, \hat{\mathcal{T}}). \tag{A2}$$

(A3) Discrete reliability. $\forall \mathcal{T} \in \mathbb{T}$, $\forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ $\exists \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}}) \subseteq \mathcal{T}$ with $\mathcal{T} \setminus \hat{\mathcal{T}} \subseteq \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}})$,

$$\begin{aligned} &\left| \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}}) \right| \leq \Lambda_{\mathrm{ref}} \left| \mathcal{T} \setminus \hat{\mathcal{T}} \right| \quad \text{and} \\ &\delta^2(\mathcal{T}, \hat{\mathcal{T}}) \leq \Lambda_3 \left(\eta^2(\mathcal{T}, \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}})) + \mu^2(\mathcal{T}) \right) + \widehat{\Lambda}_3 \eta^2(\hat{\mathcal{T}}). \end{aligned} \tag{A3}$$

(A4) Quasiorthogonality of discrete solutions. $\forall \ell \in \mathbb{N}_0$

$$\sum_{k=\ell}^{\infty} \delta^2(\mathcal{T}_k, \mathcal{T}_{k+1}) \leq \Lambda_4 \sigma_\ell^2. \tag{A4}$$

(B1) Rate s data approximation. $\forall \, \mathrm{Tol} > 0$, $\mathcal{T}_{\mathrm{Tol}} := \mathtt{appx}(\mathrm{Tol}, \mu(K) : K \in \mathcal{T}_0) \in \mathbb{T}$ satisfies

$$|\mathcal{T}_{\mathrm{Tol}}| - |\mathcal{T}_0| \leq \Lambda_5 \, \mathrm{Tol}^{-1/(2s)} \quad \text{and} \quad \mu^2(\mathcal{T}_{\mathrm{Tol}}) \leq \mathrm{Tol}\,. \tag{B1}$$

(B2) Quasimonotonicity of μ. $\forall \mathcal{T} \in \mathbb{T}$, $\forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$, $\mu(\hat{\mathcal{T}}) \leq \Lambda_6 \mu(\mathcal{T})$.

Theorem 3.2 below asserts that the aforementioned axioms imply quasimonotonicity of σ for small values of $\widehat{\Lambda}_3$.

(QM) Quasimonotonicity of σ. $\forall \mathcal{T} \in \mathbb{T}$, $\forall \hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$, $\sigma(\hat{\mathcal{T}}) \leq \Lambda_7 \sigma(\mathcal{T})$.

2.5. Optimal convergence rates. The axioms (A1)–(A4), (B1)–(B2), and (QM) ensure quasioptimality of SAFEM for sufficiently small parameters θ_A and κ as stated in Theorem 2.1 below. Recall that $\sigma^2 := \eta^2 + \mu^2$ and set

$$\sigma^2(\mathcal{T}) \equiv \sigma(\mathcal{T})^2 := \sigma^2(\mathcal{T}, \mathcal{T}) := \sum_{K \in \mathcal{T}} \sigma^2(\mathcal{T}, K) \text{ for } \mathcal{T} \in \mathbb{T} \text{ and } \sigma_\ell := \sigma(\mathcal{T}_\ell).$$

For any $N \in \mathbb{N}_0$, the comparison with the optimal rates concerns the optimal value

$$\min \sigma(\mathbb{T}(N)) := \min\{\sigma(\mathcal{T}) : \mathcal{T} \in \mathbb{T}(N)\}$$

of all admissible triangulations

$$\mathbb{T}(N) := \{\mathcal{T} \in \mathbb{T} : |\mathcal{T}| \leq |\mathcal{T}_0| + N\}$$

of cardinality $|\mathcal{T}| \leq |\mathcal{T}_0| + N$ with at most N extra cells.

THEOREM 2.1 (quasioptimality). *Suppose* (A1)–(A4), (B1)–(B2) *and* (QM). *This leads to the existence of some $\kappa_0 > 0$, which is $+\infty$ if $\Lambda_6 = 1$, such that any choice of κ, θ_A, and ρ_B with*

$$0 < \kappa < \kappa_1 := \min\left\{\kappa_0, \Lambda_1^{-2} \Lambda_3^{-1}\right\}, \quad 0 < \theta_A < \theta_0 := (1 - \kappa \Lambda_1^2 \Lambda_3)/(1 + \Lambda_1^2 \Lambda_3),$$

and $0 < \rho_B < 1$ *implies the following. The outputs* $(\mathcal{T}_\ell)_{\ell\in\mathbb{N}_0}$ *and* $(\sigma_\ell)_{\ell\in\mathbb{N}_0}$ *of* SAFEM *satisfy the equivalence*

$$\Lambda_5^s + \sup_{\ell\in\mathbb{N}_0} (1 + |\mathcal{T}_\ell| - |\mathcal{T}_0|)^s\, \sigma_\ell \approx \Lambda_5^s + \sup_{N\in\mathbb{N}_0} (1+N)^s \min \sigma(\mathbb{T}(N)). \tag{2.3}$$

The theorem guarantees optimal rates in that the left-hand side of the equivalence (2.3) is smaller than infinity if the right-hand side is, and vice versa. The quotient is bounded below and from above by the equivalence constants, which continuously depend on Λ_{ref}, Λ_1, Λ_2, Λ_3, $\widehat{\Lambda}_3$, Λ_4, Λ_6, ρ_B, ρ_2, θ_A, κ, and $s > 0$ but not on Λ_5.

The (possibly unknown) parameter s is not utilized in SAFEM. The axioms (B1)–(B2) specify *sufficient* conditions for optimal convergence, where the parameter $s > 0$ is arbitrary and may refer to a related nonlinear approximation class.

3. Remarks.

3.1. Weak form of (A4).

The axiom (A4) can be weakened with some parameter $\varepsilon > 0$, which vanishes in (A4)≡(A4$_0$).

(A4$_{\boldsymbol\varepsilon}$) Quasiorthogonality with $\varepsilon > 0$. $\exists \varepsilon > 0\ \exists 0 < \Lambda_{4(\varepsilon)} < \infty\ \forall \ell, m \in \mathbb{N}_0$,

$$\sum_{k=\ell}^{\ell+m} \delta_{k,k+1}^2 \le \Lambda_{4(\varepsilon)} \sigma_\ell^2 + \varepsilon \sum_{k=\ell}^{\ell+m} \sigma_k^2. \tag{A4$_\varepsilon$}$$

The axiom (A4$_\varepsilon$) implies (A4$_{\varepsilon'}$) for all $0 \le \varepsilon < \varepsilon'$ with the same constant $\Lambda_{4(\varepsilon)} = \Lambda_{4(\varepsilon')}$, and (A4) is (A4$_0$), i.e., (A4$_\varepsilon$) for $\varepsilon = 0$. Conversely, as $\varepsilon \searrow 0$ it may be expected that $\Lambda_{4(\varepsilon)} \to \infty$. In the presence of (A1)–(A2), this is not the case. In fact, (A1)–(A2) and (A4$_\varepsilon$) imply (A4) for sufficiently small $\varepsilon > 0$.

THEOREM 3.1 ((A4$_\varepsilon$)⇒(A4)). *Let* θ_A *be the parameter of* SAFEM, *let* $0 < \rho_{12} < 1$ *be the reduction factor for the total error estimator with constant* $0 < \Lambda_{12} < \infty$ *in Theorem* 4.1 *below, and let* $0 \le \varepsilon < (1-\rho_{12})/\Lambda_{12}$. *Then* (A1)–(A2) *and* (A4$_\varepsilon$) *imply* (A4) *with* $\Lambda_4 := \Lambda_{4(\varepsilon)} + \varepsilon(1 + \Lambda_{12}\Lambda_{4(\varepsilon)})/(1 - \rho_{12} - \varepsilon\Lambda_{12})$.

This was first observed in [CFPP14] for CAFEM and is proved in subsection 4.2 for completeness and applied below in Theorem 5.1.

3.2. Quasimonotonicity.

The axiom (B2) explicitly ensures the quasimonotonicity of μ. The strict inequality $\widehat{M} := (\Lambda_1^2 + \Lambda_2^2)\widehat{\Lambda}_3 < 1$ implies (QM) with $\Lambda_7 := \sqrt{\Lambda_6^2 + \Lambda_8^2}$, $M := (\Lambda_1^2 + \Lambda_2^2)\Lambda_3$, and

$$\Lambda_8 := \frac{1 + M(1-\widehat{M}) + \widehat{M} + 2\sqrt{M(1-\widehat{M}) + \widehat{M}}}{(1-\widehat{M})^2}.$$

THEOREM 3.2 (quasimonotonicity). *Suppose* (A1)–(A3) *and* $\widehat{M} < 1$. *Then, any* $\mathcal{T} \in \mathbb{T}$ *and* $\hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ *satisfy* $\eta(\hat{\mathcal{T}}) \le \Lambda_8 \sigma(\mathcal{T})$.

Proof. Given $\lambda := (\sqrt{M + \widehat{M} - M\widehat{M}} - \widehat{M})/(M + \widehat{M}) < 1/\widehat{M} - 1$, recall the following implication of the axioms (A1)–(A3): namely,

$$\begin{aligned}
\eta^2(\hat{\mathcal{T}}, \hat{\mathcal{T}} \cap \mathcal{T}) &\le (1 + 1/\lambda)\eta^2(\mathcal{T}, \hat{\mathcal{T}} \cap \mathcal{T}) + (1+\lambda)\Lambda_1^2 \delta^2(\mathcal{T}, \hat{\mathcal{T}}),\\
\eta^2(\hat{\mathcal{T}}, \hat{\mathcal{T}} \setminus \mathcal{T}) &\le (1 + 1/\lambda)\rho_2^2\eta^2(\mathcal{T}, \mathcal{T} \setminus \hat{\mathcal{T}}) + (1+\lambda)\Lambda_2^2 \delta^2(\mathcal{T}, \hat{\mathcal{T}}),\\
\delta^2(\mathcal{T}, \hat{\mathcal{T}}) &\le \Lambda_3 \sigma^2(\mathcal{T}) + \widehat{\Lambda}_3 \eta^2(\hat{\mathcal{T}}).
\end{aligned}$$

Those inequalities plus the split $\eta^2(\hat{\mathcal{T}}) = \eta^2(\hat{\mathcal{T}}, \hat{\mathcal{T}} \cap \mathcal{T}) + \eta^2(\hat{\mathcal{T}}, \hat{\mathcal{T}} \setminus \mathcal{T})$ verify

$$\eta^2(\hat{\mathcal{T}}) \leq (1 + 1/\lambda)\eta^2(\mathcal{T}) + (1+\lambda)(\Lambda_1^2 + \Lambda_2^2)\left(\Lambda_3\sigma^2(\mathcal{T}) + \widehat{\Lambda}_3\eta^2(\hat{\mathcal{T}})\right). \qquad \square$$

3.3. Optimal data approximation with Approx. Case (B) of SAFEM runs a data approximation algorithm $\texttt{appx}(\mathrm{Tol}, \mu(K) : K \in \mathcal{T}_0)$ with output in $\mathbb{T}$. The data approximation algorithm Approx [BDdV04, BdV04] is based on the refinement of partitions and has been incorporated into separate marking algorithms of the type SAFEM of this paper for data approximation reduction in [CR11, Rab15]. Approx is one possible realization of appx in SAFEM.

Let $\hat{\mathcal{P}}$ be some NVB refinement of $\mathcal{P} \in \mathbb{P}$. Let $K \in \mathcal{P}$ and $\hat{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$; then the refinement of K in $\hat{\mathcal{P}}$ is the set $\hat{\mathcal{P}}(K) := \{T \in \hat{\mathcal{P}} \,|\, T \subseteq K\}$ in the following.

(SA) Subadditivity. $\exists \Lambda_6 < \infty$, $\forall \mathcal{P} \in \mathbb{P}$, $\forall \hat{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$, $\forall \mathcal{M} \subseteq \mathcal{P}$

$$\text{(SA)} \qquad \mu^2(\hat{\mathcal{P}}(\mathcal{M})) := \sum_{K \in \mathcal{M}} \sum_{T \in \hat{\mathcal{P}}(K)} \mu^2(T) \leq \Lambda_6 \mu^2(\mathcal{M}).$$

Note that the notation of the data approximation term μ is a straightforward extension of its definition in (2.1) for admissible triangulations to partitions.

The algorithm Approx is introduced and analyzed in [BDdV04, BdV04] with input tolerance $\mathrm{Tol}' := \mathrm{Tol}/\Lambda_6 = \rho_B \mu_\ell / \Lambda_6$ and the values $\mu(K)$ on the coarse triangulation $\mathcal{T}_0$.

Theorem 3.3 (see [BdV04, BDdV04]). (SA) *in* Approx *implies* (B1)–(B2) *with rate-s-optimality in the sense that*

$$M(s, \mu) := \sup_{N \in \mathbb{N}_0} (1+N)^s \min \mu(\mathbb{T}(N)) \approx \Lambda_5^s \tag{3.1}$$

holds for all $s > 0$ (and $M(s,\mu) < \infty$ if and only if $\Lambda_5 < \infty$) with equivalence constants which may depend on s.

Proof. This follows from near optimality proven in [BdV04, Theorem 6.1] and [BDdV04, Lemma 4.4]. $\square$

Note that Approx given in [BdV04, BDdV04] is even instance optimal.

3.4. Collective Dörfler marking is optimal for $\|hf\|_{L^2(\Omega)}$. The optimality of CAFEM also suggests an algorithm for the reduction of the L^2-norm of a mesh-size weight h times a given data like $h\,f$.

Given $f \in L^2(\Omega)$ in the polyhedral domain $\Omega \subseteq \mathbb{R}^n$ partitioned into the regular triangulation $\mathcal{T}_0$, set $\eta(\mathcal{T}_\ell, K) := |K|^{2/n} |f|_{L^2(K)}$ for all $K \in \mathcal{T}_\ell$. Let $\eta_\ell = \eta(\mathcal{T}_\ell, \mathcal{T}_\ell)$. Then, (A1)–(A4) are satisfied with appropriate weight functions $h_\mathcal{T}$ of mesh-sizes in $\mathcal{T}$ (resp., the mesh-sizes $h_{\hat{\mathcal{T}}} =: \hat{h}_T$ with respect to the finer triangulation $\hat{\mathcal{T}}$) and

$$\delta(\mathcal{T}, \hat{\mathcal{T}}) := \left\| (h_\mathcal{T} - \hat{h}_T) f \right\|_{L^2(\Omega)}.$$

Hence CAFEM with collective Dörfler marking implies optimal data approximation for this particular data error term with a mesh-size weight $h_\mathcal{T}$. This is in agreement with the well-established fact that first-order conforming and nonconforming FEMs do not need a data reduction with SAFEM.

4. Proofs. This section is devoted to the proof of Theorem 2.1. The abbreviation $\delta_{\ell,\ell+1} := \delta(\mathcal{T}_\ell, \mathcal{T}_{\ell+1})$ applies in this section.

4.1. Estimator reduction. The constant $\Lambda_6 \geq 1$ in the following theorem leads to κ_0 set to $+\infty$ for $\Lambda_6 = 1$; this holds in all the examples of this paper.

THEOREM 4.1 ((A12) reduction). *Suppose* (A1)–(A2) *and parameters* $0 < \theta_A \leq 1$, $0 < \kappa$, *and* $0 < \rho_B < 1/\Lambda_6$ *from* SAFEM. *Any choice of* γ *and* λ *with*

$$0 < \gamma < \rho_2^{-2} - 1 \text{ and } 0 < \lambda < \min\left\{\left(1-(1+\gamma)\rho_2^2\right)\frac{\theta_A}{1-\theta_A}, \kappa(1-\rho_B)\right\} \tag{4.1}$$

leads to the constants

$$0 < \Lambda_{12} := (1+1/\lambda)\Lambda_1^2 + (1+1/\gamma)\Lambda_2^2 < \infty, \tag{4.2}$$

$$0 < \rho_A := (1+\lambda)(1-\theta_A) + (1+\gamma)\rho_2^2\theta_A < 1, \tag{4.3}$$

$$0 < \kappa_0 := (1-\rho_A)/(\Lambda_6 - 1) \text{ (with } \kappa_0 := +\infty \text{ if } \Lambda_6 = 1\text{)}, \tag{4.4}$$

$$0 < \rho_{12} := \max\{\rho_A + \kappa\Lambda_6, 1 + \lambda + \kappa\rho_B\}/(1+\kappa) \leq 1. \tag{4.5}$$

Then $0 < \kappa < \kappa_0$ *implies* $\rho_{12} < 1$ *and*

$$\sigma_{\ell+1}^2 \leq \rho_{12}\sigma_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2 \quad \text{for all } \ell \in \mathbb{N}_0 \tag{A12}$$

for the output σ_ℓ^2 *of* SAFEM.

Proof. For γ and λ as in (4.1), the axioms (A1)–(A2) imply

$$\eta_{\ell+1}^2(\mathcal{T}_{\ell+1} \cap \mathcal{T}_\ell) \leq (1+\lambda)\eta_\ell^2(\mathcal{T}_{\ell+1} \cap \mathcal{T}_\ell) + (1+1/\lambda)\Lambda_1^2\delta_{\ell,\ell+1}^2,$$
$$\eta_{\ell+1}^2(\mathcal{T}_{\ell+1} \setminus \mathcal{T}_\ell) \leq (1+\gamma)\rho_2^2\eta_\ell^2(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+1}) + (1+1/\gamma)\Lambda_2^2\delta_{\ell,\ell+1}^2.$$

The sum of those two inequalities leads to

$$\eta_{\ell+1}^2 \leq (1+\lambda)\eta_\ell^2 + ((1+\gamma)\rho_2^2 - (1+\lambda))\eta_\ell^2(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+1}) + \Lambda_{12}\delta_{\ell,\ell+1}^2. \tag{4.6}$$

The restrictions on λ and γ ensure $(1+\gamma)\rho_2^2 < 1 < 1+\lambda$. Thus, in general,

$$\eta_{\ell+1}^2 \leq (1+\lambda)\eta_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2.$$

In Case (A) on the level ℓ, when Dörfler's marking ensures $\theta_A\eta_\ell^2 \leq \eta_\ell^2(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+1})$, this and (4.6) lead to an improvement of the last estimate, namely

$$\eta_{\ell+1}^2 \leq \left((1+\lambda)(1-\theta_A) + (1+\gamma)\rho_2^2\theta_A\right)\eta_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2 = \rho_A\eta_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2.$$

The restrictions on λ and γ reveal $\rho_A < 1$. Altogether, let

$$R_\ell := \begin{cases} \rho_A & \text{in Case (A) on level } \ell, \\ 1+\lambda & \text{in Case (B) on level } \ell. \end{cases} \tag{4.7}$$

Then, the output of SAFEM satisfies

$$\eta_{\ell+1}^2 \leq R_\ell\eta_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2 \quad \text{for all } \ell \in \mathbb{N}_0. \tag{4.8}$$

In Case (A) on any level ℓ with $R_\ell = \rho_A$ from (4.3) and Λ_{12} from (4.2), it also holds that $\mu_{\ell+1}^2 \leq \Lambda_6 \mu_\ell^2$ and $\mu_\ell^2 \leq \kappa \eta_\ell^2$. Since $\alpha := (\Lambda_6 - \rho_A)/(\kappa+1) > 0$, this and (4.8) lead to

$$\sigma_{\ell+1}^2 \leq (\rho_A + \alpha\kappa)\eta_\ell^2 + (\Lambda_6 - \alpha)\mu_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2 = \frac{\rho_A + \kappa\Lambda_6}{1+\kappa}\sigma_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2.$$

In Case (B) on the level ℓ with $R_\ell = 1+\lambda$, it holds that $\mu_{\ell+1}^2 \leq \rho_B \mu_\ell^2$ and $\kappa\eta_\ell^2 < \mu_\ell^2$. Since $\beta := \kappa(1+\lambda-\rho_B)/(1+\kappa) > 0$, this and (4.8) lead to

$$\sigma_{\ell+1}^2 < (1+\lambda-\beta)\eta_\ell^2 + (\rho_B + \beta/\kappa)\mu_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2 = \frac{1+\kappa\rho_B+\lambda}{1+\kappa}\sigma_\ell^2 + \Lambda_{12}\delta_{\ell,\ell+1}^2.$$

This proves the total error estimator reduction (A12) with ρ_{12} from (4.5). $\square$

4.2. Convergence. The plain convergence follows from the estimator reduction (A12) plus quasiorthogonality (A4). The elementary proofs are adopted from [CFPP14] and given for completeness.

Theorem 4.2. *Suppose* $0 < \theta_A \leq 1$, $0 < \kappa$, $0 < \rho_B < 1$, *and suppose* (A4) *and* (A12) *with constants* $0 < \rho_{12} < 1$ *and* $0 < \Lambda_{12} < \infty$. *Then* $\Lambda := (1+\Lambda_{12}\Lambda_4)/(1-\rho_{12})$, $q := \Lambda/(1+\Lambda) < 1$, *and the output of* SAFEM *satisfy the following assertions.*

(a) *(Plain convergence)* $\forall \ell, m \in \mathbb{N}_0$, $\sum_{k=\ell}^{\ell+m} \sigma_k^2 \leq \Lambda\sigma_\ell^2$.
(b) *(R-linear convergence on each level)* $\forall \ell, m \in \mathbb{N}_0$, $\sigma_{\ell+m}^2 \leq \frac{q^m}{1-q}\sigma_\ell^2$.
(c) *(Reciprocal sum)* $\forall s > 0$, $\forall \ell \in \mathbb{N}$, $\sum_{k=0}^{\ell-1} \sigma_k^{-1/s} \leq \frac{q^{1/(2s)}\sigma_\ell^{-1/s}}{(1-q)^{1/(2s)}(1-q^{1/(2s)})}$.

Proof of Theorem 4.2(a). For all ℓ, $m \in \mathbb{N}_0$, (A12) implies

$$\sum_{k=\ell}^{\ell+m} \sigma_k^2 = \sigma_\ell^2 + \sum_{k=\ell+1}^{\ell+m} \sigma_k^2 \leq \sigma_\ell^2 + \rho_{12}\sum_{k=\ell}^{\ell+m}\sigma_k^2 + \Lambda_{12}\sum_{k=\ell}^{\ell+m}\delta_{k,k+1}^2. \tag{4.9}$$

This plus (A4) verifies

$$(1-\rho_{12})\sum_{k=\ell}^{\ell+m}\sigma_k^2 \leq \sigma_\ell^2 + \Lambda_{12}\Lambda_4\sigma_\ell^2.$$

This proves (a) with the asserted constant Λ. $\square$

Proof of Theorem 3.1. The same argument as in the proof of Theorem 4.2(a) shows that (A12) and (A4$_\varepsilon$) imply (A4) for small ε. In fact, (4.9) and (A4$_\varepsilon$) show

$$(1-\rho_{12})\sum_{k=\ell}^{\ell+m}\sigma_k \leq \sigma_\ell^2 + \Lambda_{12}\left(\Lambda_{4(\varepsilon)}\sigma_\ell^2 + \varepsilon\sum_{k=\ell}^{\ell+m}\sigma_k\right).$$

In other words

$$(1-\rho_{12}-\varepsilon\Lambda_{12})\sum_{k=\ell}^{\ell+m}\sigma_k^2 \leq \left(1+\Lambda_{12}\Lambda_{4(\varepsilon)}\right)\sigma_\ell^2.$$

This plus (A4$_\varepsilon$) leads to (A4) with $\Lambda_4 := \Lambda_{4(\varepsilon)} + \varepsilon(1+\Lambda_{12}\Lambda_{4(\varepsilon)})/(1-\rho_{12}-\varepsilon\Lambda_{12})$,

$$\sum_{k=\ell}^{\ell+m}\delta_{k,k+1}^2 \leq \Lambda_{4(\varepsilon)}\sigma_\ell^2 + \varepsilon\sum_{k=\ell}^{\ell+m}\sigma_k^2 \leq \Lambda_4\sigma_\ell^2. \qquad \square$$

Proof of Theorem 4.2(b). The assertion (a) implies the convergence of the series

$$\xi_{\ell+1}^2 := \sum_{k=\ell+1}^{\infty} \sigma_k^2 \leq \Lambda \sigma_\ell^2 < \infty.$$

The addition of $\Lambda \xi_{\ell+1}^2$ to the previous inequality results in

$$(\Lambda+1)\xi_{\ell+1}^2 \leq \Lambda \xi_\ell^2; \text{ hence } \xi_{\ell+1}^2 \leq q\xi_\ell^2. \tag{4.10}$$

The successive application of the previous contraction (4.10) shows

$$\sigma_{\ell+m}^2 \leq \xi_{\ell+m}^2 \leq q^m \xi_\ell^2 = q^m \left(\sigma_\ell^2 + \xi_{\ell+1}^2\right) \leq q^m(1+\Lambda)\sigma_\ell^2. \qquad \square$$

Proof of Theorem 4.2(c). The R-linear convergence of (b) leads to

$$\sigma_k^{-1/s} \leq \frac{q^{(\ell-k)/(2s)}}{(1-q)^{1/(2s)}} \sigma_\ell^{-1/s} \qquad \text{for all } 0 \leq k < \ell.$$

This proves

$$\sum_{k=0}^{\ell-1} \sigma_k^{-1/s} \leq \frac{\sigma_\ell^{-1/s}}{(1-q)^{1/(2s)}} \sum_{k=0}^{\ell-1} \left(q^{1/(2s)}\right)^{\ell-k} \leq \frac{\sigma_\ell^{-1/s} q^{1/(2s)}}{(1-q)^{1/(2s)}(1-q^{1/(2s)})}. \qquad \square$$

LEMMA 4.3 (comparison). *Suppose* (A1)–(A4), (B1)–(B2) *with* $0 < s < \infty$, (QM), $0 < q < 1$ *from Theorem* 4.2(b), *and let* $0 < \xi < 1$ *and* $0 < \nu < \infty$*; let*

$$M := M(s,\sigma) := \sup_{N \in \mathbb{N}_0} (N+1)^s \min \sigma(\mathbb{T}(N)) < \infty, \tag{4.11}$$

similar to the definition of $M(s,\mu)$ *in* (3.1). *Then for any level* $\ell \in \mathbb{N}_0$ *of* SAFEM *with a triangulation* $\mathcal{T}_\ell$, *there exists a refinement* $\hat{\mathcal{T}}_\ell \in \mathbb{T}(\mathcal{T}_\ell)$ *with* (a)–(c).

(a) $\sigma(\hat{\mathcal{T}}_\ell) \leq \xi \sigma_\ell$;

(b) $\sqrt{1-q}\,\xi\, \sigma_\ell \left|\mathcal{T}_\ell \setminus \hat{\mathcal{T}}_\ell\right|^s \leq \Lambda_7 M$;

(c) $$(1-\xi^2(1+\nu+(1+1/\nu)\Lambda_1^2\widehat{\Lambda}_3))\eta_\ell^2 \leq \left(1+(1+1/\nu)\Lambda_1^2\Lambda_3\right)\eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell)) + ((1+\nu)\xi^2+(1+1/\nu)\Lambda_1^2(\Lambda_3+\widehat{\Lambda}_3\xi^2))\mu_\ell^2.$$

Proof. Two pathological situations are excluded in the beginning of the proof. First, if $\sigma_\ell = 0$, then $\hat{\mathcal{T}}_\ell = \mathcal{T}_\ell$ satisfies the assumptions (a)–(c). Second, Theorem 4.2 guarantees convergence of some sequence of triangulations, and (QM) implies that this holds for uniform refinements as well. Hence there exists a refinement $\hat{\mathcal{T}}_\ell$ of $\mathcal{T}_\ell$ with (a) and $\hat{\mathcal{T}}_\ell \cap \mathcal{T}_\ell = \emptyset$. The latter implies (c) even in case $M \equiv M(s,\sigma) = \infty$ when (b) is obvious.

Throughout the remaining parts of the proof, it is therefore assumed that $M < \infty$ and $\sigma_\ell > 0$. Then (QM) implies $0 < \sigma_0 \leq M < \infty$.

1. *Setup.* Let $N_\ell \in \mathbb{N}_0$ be minimal with

$$(N_\ell+1)^{-s} \leq \frac{\xi\sqrt{1-q}}{\Lambda_7 M}\, \sigma_\ell. \tag{4.12}$$

The quasimonotonicity (QM) followed by the definition of $M := M(s,\sigma) < \infty$ in (4.11) and $0 < q < 1, 0 < \xi < 1$ lead to

$$\frac{\xi\sqrt{1-q}}{\Lambda_7}\sigma_\ell \leq \xi\sqrt{1-q}\,\sigma_0 \leq \xi\sqrt{1-q}M < M.$$

Hence, $(N_\ell+1)^{-s}<1$ and so $N_\ell\geq 1$. Since $N_\ell\in\mathbb{N}$ is minimal with (4.12),

$$0<(N_\ell+1)^{-s}\leq\frac{\xi\sqrt{1-q}}{\Lambda_7 M}\sigma_\ell<N_\ell^{-s}.$$

This implies

$$N_\ell^s<\frac{\Lambda_7 M}{\xi\sqrt{1-q}}\sigma_\ell^{-1}. \tag{4.13}$$

2. *Design of* $\hat{\mathcal{T}}_\ell$. The definition of $M<\infty$ yields the existence of some optimal $\tilde{\mathcal{T}}_\ell\in\mathbb{T}(N_\ell)$ with

$$(N_\ell+1)^s\,\sigma(\tilde{\mathcal{T}}_\ell)\leq M. \tag{4.14}$$

The overlay triangulation $\hat{\mathcal{T}}_\ell:=\mathcal{T}_\ell\oplus\tilde{\mathcal{T}}_\ell$ [CKNS08, Ste07] satisfies

$$\left|\hat{\mathcal{T}}_\ell\right|+|\mathcal{T}_0|\leq|\mathcal{T}_\ell|+\left|\tilde{\mathcal{T}}_\ell\right|. \tag{4.15}$$

3. *Proof of* (a). The quasimonotonicity (QM) followed by (4.14) and (4.12) shows

$$\sigma(\hat{\mathcal{T}}_\ell)\leq\Lambda_7\sigma(\tilde{\mathcal{T}}_\ell)\leq\Lambda_7 M(N_\ell+1)^{-s}\leq\xi\sigma_\ell\sqrt{1-q}<\xi\sigma_\ell.$$ □

4. *Proof of* (b). The definition of $\tilde{\mathcal{T}}_\ell$, the overlay estimate in (4.15), and the upper bound for N_ℓ in (4.13) lead to

$$\left|\mathcal{T}_\ell\setminus\hat{\mathcal{T}}_\ell\right|\leq\left|\hat{\mathcal{T}}_\ell\right|-|\mathcal{T}_\ell|\leq\left|\tilde{\mathcal{T}}_\ell\right|-|\mathcal{T}_0|\leq N_\ell\leq\left(\frac{\Lambda_7 M}{\xi\sigma_\ell\sqrt{1-q}}\right)^{1/s}.$$ □

5. *Proof of* (c). For any $0<\nu<\infty, 0<\xi<1$, (A1) and (A3) result in

$$\begin{aligned}\eta_\ell^2(\mathcal{T}_\ell\cap\hat{\mathcal{T}}_\ell)&\leq(1+\nu)\eta^2(\hat{\mathcal{T}}_\ell,\hat{\mathcal{T}}_\ell\cap\mathcal{T}_\ell)+(1+1/\nu)\Lambda_1^2\delta^2(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell)\\&\leq\left(1+\nu+(1+1/\nu)\Lambda_1^2\widehat{\Lambda}_3\right)\eta^2(\hat{\mathcal{T}}_\ell)\\&\quad+(1+1/\nu)\Lambda_1^2\Lambda_3\left(\eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell))+\mu_\ell^2\right).\end{aligned}$$

This, (a), and $\mathcal{T}_\ell\setminus\hat{\mathcal{T}}_\ell\subseteq\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell)$ result in

$$\begin{aligned}\eta_\ell^2&=\eta_\ell^2(\mathcal{T}_\ell\cap\hat{\mathcal{T}}_\ell)+\eta_\ell^2(\mathcal{T}_\ell\setminus\hat{\mathcal{T}}_\ell)\\&\leq\left(1+\nu+(1+1/\nu)\Lambda_1^2\widehat{\Lambda}_3\right)\xi^2\sigma_\ell^2+\left(1+(1+1/\nu)\Lambda_1^2\Lambda_3\right)\eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell))\\&\quad+(1+1/\nu)\Lambda_1^2\Lambda_3\mu_\ell^2.\end{aligned}$$

Some rearrangements with $\sigma_\ell^2=\eta_\ell^2+\mu_\ell^2$ prove (c). □

4.3. Proof of Theorem 2.1.

Proof of "$\lesssim$" in (2.3) *of Theorem* 2.1. Since $\theta_A<\theta_0$ and the function

$$f(\xi,\nu):=\frac{1-\xi^2\left((1+\kappa)(1+\nu)+(1+\kappa)(1+1/\nu)\Lambda_1^2\widehat{\Lambda}_3\right)-\kappa(1+1/\nu)\Lambda_1^2\Lambda_3}{1+(1+1/\nu)\Lambda_1^2\Lambda_3}$$

is strictly smaller than $\theta_0 = \lim_{\nu\to\infty} f(0,\nu)$, there exist ν, ξ such that

$$\theta_A < f(\xi,\nu) < \theta_0.$$

Given κ_0 from Theorem 4.1, assume $\kappa < \kappa_1 := \min\left\{\kappa_0, \Lambda_1^{-2}\Lambda_3^{-1}\right\}$.

Case (A). Lemma 4.3(c) and $\mu_\ell^2 \le \kappa\eta_\ell^2$ prove that $\mathcal{R}(\mathcal{T}_\ell, \hat{\mathcal{T}}_\ell)$ satisfies

$$\begin{aligned}&\Big(1-(1+\kappa)\xi^2(1+\nu)-(1+\kappa)\xi^2(1+1/\nu)\Lambda_1^2\widehat{\Lambda}_3-\kappa(1+1/\nu)\Lambda_1^2\Lambda_3\Big)\,\eta_\ell^2\\&\qquad\le \left(1+(1+1/\nu)\Lambda_1^2\Lambda_3\right)\eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell)).\end{aligned}$$

This reads $\theta_A\eta_\ell^2 \le f(\xi,\nu)\eta_\ell^2 \le \eta_\ell^2(\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell))$ and implies that $\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell)$ satisfies Dörfler marking in Case (A).

Let $\mathcal{M}_\ell =: \mathcal{M}_\ell^{(0)}$ be the set of marked elements in the Dörfler marking on level ℓ, while $\mathcal{M}_\ell^\star$ is the optimal set of marked elements. Hence, there exists $0 < \Lambda_{\mathrm{opt}} < \infty$ such that

$$|\mathcal{M}_\ell| \le \Lambda_{\mathrm{opt}}\,|\mathcal{M}_\ell^\star| \le \Lambda_{\mathrm{opt}}\left|\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell)\right|.$$

The control over $\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell)$ of (A3) in Lemma 4.3(b) results in

$$\left|\mathcal{R}(\mathcal{T}_\ell,\hat{\mathcal{T}}_\ell)\right| \le \Lambda_{\mathrm{ref}}\left|\mathcal{T}_\ell\setminus\hat{\mathcal{T}}_\ell\right| \le \Lambda_{\mathrm{ref}}\left(\frac{\Lambda_7 M}{\sqrt{1-q}\,\xi\sigma_\ell}\right)^{1/s}.$$

Hence, $\Lambda_9 := \Lambda_{\mathrm{opt}}\Lambda_{\mathrm{ref}}\Lambda_7^{1/s}(\sqrt{1-q}\,\xi)^{-1/s}$ satisfies

$$\left|\mathcal{M}_\ell^{(0)}\right| = |\mathcal{M}_\ell| \le \Lambda_9 M^{1/s}\sigma_\ell^{-1/s}. \tag{4.16}$$

Case (B). The output of `appx` with input triangulation $\mathcal{T}_0$ and input tolerance $\mathrm{Tol} := \rho_B\mu_\ell^2$ on the level ℓ satisfies (B1). Since $\sigma_\ell^2 = \eta_\ell^2 + \mu_\ell^2 \le (1+1/\kappa)\mu_\ell^2$ in Case (B), this leads to

$$|\mathcal{T}_{\mathrm{Tol}}| - |\mathcal{T}_0| \le \Lambda_5(1+1/\kappa)\rho_B^{-1/(2s)}\sigma_\ell^{-1/s}.$$

Let $\mathcal{T}_\ell^{(0)} := \mathcal{T}_\ell$. According to [CR11, Rab15] for $\mathcal{T}_{\ell+1} = \mathcal{T}_\ell \oplus \mathcal{T}_{\mathrm{Tol}}$ there exists a finite sequence $(\mathcal{M}_\ell^{(k)})_{k=0,\dots,K(\ell)}$ of sets of marked element domains that satisfies

$$\mathcal{T}_\ell^{(k+1)} = \textsc{Refine}(\mathcal{T}_\ell^{(k)}, \mathcal{M}_\ell^{(k)}) \quad \text{for all } k = 0,\dots,K(\ell)-1,$$

which finally leads to $\mathcal{T}_{\ell+1} = \mathcal{T}_\ell^{(K(\ell))}$. This observation and the estimate for the overlay with the sequence $(\mathcal{M}_\ell^{(k)})_{k=0,\dots,K(\ell)}$ [CR11, Theorem 3.3] show

$$\sum_{k=0}^{K(\ell)} |\mathcal{M}_\ell^{(k)}| \le |\mathcal{T}_{\mathrm{Tol}}| - |\mathcal{T}_0| \lesssim \Lambda_5(1+1/\kappa)\rho_B^{-1/(2s)}\sigma_\ell^{-1/s}. \tag{4.17}$$

The estimate from [CR11, Theorem 3.3] is for 2D only; however, it is expected to hold in general.

End of the proof of "$\lesssim$". The overhead control of [BDdV04, Ste08] guarantees

$$|\mathcal{T}_\ell| - |\mathcal{T}_0| \le \Lambda_{\mathrm{BDdV}}\sum_{j=0}^{\ell-1}\sum_{k=0}^{K(j)} |\mathcal{M}_j^{(k)}|. \tag{4.18}$$

With (4.16)–(4.17) and Theorem 4.2(c), this proves

$$|\mathcal{T}_\ell| - |\mathcal{T}_0| \lesssim (\Lambda_5 + M^{1/s})\sigma_\ell^{-1/s}. \tag{4.19}$$

Finally, $1 \le |\mathcal{T}_\ell| - |\mathcal{T}_0|$ implies $1 + |\mathcal{T}_\ell| - |\mathcal{T}_0| \le 2(|\mathcal{T}_\ell| - |\mathcal{T}_0|)$ while $|\mathcal{T}_\ell| = |\mathcal{T}_0|$ implies $1 \le \sigma_\ell^{-1/s}(\Lambda_5 + M^{1/s})$. Hence (4.19) proves $\sigma_\ell(1 + |\mathcal{T}_\ell| - |\mathcal{T}_0|)^s \lesssim \Lambda_5^s + M$ and so "$\lesssim$" in the assertion of Theorem 2.1. ∎

Proof of "$\gtrsim$" in (2.3) *of Theorem* 2.1. Given $N \in \mathbb{N}_0$, suppose that $\min \sigma(\mathbb{T}(N))$ is positive and so $\sigma_\ell > 0$ for all $\ell \in \mathbb{N}_0$ with $N_\ell := |\mathcal{T}_\ell| - |\mathcal{T}_0| \le N$. For level ℓ in SAFEM this leads to $N_{\ell+1} > N_\ell$, for it only stops with $\mathcal{T}_\ell = \mathcal{T}_{\ell+1} = \mathcal{T}_{\ell+2} = \dots$ when $\sigma_\ell = 0$. Hence there exists some level ℓ with $N_\ell < N \le N_{\ell+1}$. This implies

$$(N+1)^s \min \sigma(\mathbb{T}(N)) \le (N_{\ell+1}+1)^s \sigma_\ell, \tag{4.20}$$

which is evident in case $\min \sigma(\mathbb{T}(N)) = 0$.

In Case (A) on the level ℓ of SAFEM, there is a one-level refinement to create $\mathcal{T}_{\ell+1}$, where each simplex in $\mathcal{T}_\ell$ creates a finite number $\le K(n)$ of children in a completion step. The constant $K(n) \ge 2$ depends only on the spatial dimension n [GSS14]. This leads to the bound $|\mathcal{T}_{\ell+1}| \le K(n)\,|\mathcal{T}_\ell|$ and then to

$$(N_{\ell+1}+1)/(N_\ell+1) \le K(n) + (K(n)-1)(|\mathcal{T}_0|-1) \lesssim 1.$$

In Case (B) on the level ℓ of SAFEM, the refinement $\mathcal{T}_{\ell+1} := \mathcal{T}_\ell \oplus \mathcal{T}_{\mathrm{Tol}}$ is controlled by $|\mathcal{T}_{\mathrm{Tol}}| - |\mathcal{T}_0| \le \Lambda_5 \,\mathrm{Tol}^{-1/(2s)} \le \Lambda_5 \rho_B^{-1/(2s)} \mu_\ell^{-1/s}$. Since $\sigma_\ell^2 \le (1+1/\kappa)\mu_\ell^2$ in Case (B), the overlay estimate of [CKNS08, Ste07] proves

$$N_{\ell+1} - N_\ell \le |\mathcal{T}_{\mathrm{Tol}}| - |\mathcal{T}_0| \le \Lambda_5 \rho_B^{-1/(2s)} (1+1/\kappa)^{1/(2s)} \sigma_\ell^{-1/s}.$$

This leads to the bound

$$2^{-s}(N_{\ell+1}+1)^s \le (N_\ell+1)^s + \rho_B^{-1/2}(1+1/\kappa)^{1/2}\Lambda_5.$$

Consequently, in both Cases (A) and (B), it follows that

$$(N_{\ell+1}+1)^s\sigma_\ell \le \left(K(n) + (K(n)-1)(|\mathcal{T}_0|-1)\right)^s (N_\ell+1)^s\sigma_\ell + 2^s\rho_B^{-s/2}(1+1/\kappa)^{s/2}\Lambda_5^s.$$

With $S := \sup_{\ell\in\mathbb{N}_0} (N_\ell+1)^s\,\sigma_\ell$, this and (4.20) imply

$$(N+1)^s \min \sigma(\mathbb{T}(N)) \le \left(K(n) + (K(n)-1)(|\mathcal{T}_0|-1)\right)^s S + 2^s\rho_B^{-s/2}(1+1/\kappa)^{s/2}\Lambda_5^s.$$

Since this holds for any $N \in \mathbb{N}_0$, the previous N-independent upper bound is greater than or equal to the supremum M. This concludes the proof of "$\gtrsim$" in (2.3). ∎

5. Application to mixed FEM.

The a posteriori error analysis of mixed finite element schemes [Car97, Alo96] was completed in [CPS16] with a reliable and efficient error control in the natural functional analytical framework $H(\mathrm{div},\Omega) \times L^2(\Omega)$ for the dual formulation of a Poisson model problem.

5.1. Mixed formulation of a Poisson model problem.

Given the right-hand side $f \in L^2(\Omega)$ in a bounded simply connected polyhedral Lipschitz domain $\Omega \subset \mathbb{R}^n$ for $n = 2, 3$, the dual formulation of the Laplace equation seeks $p \in H(\mathrm{div},\Omega)$ and $u \in L^2(\Omega)$ with

$$\begin{aligned} a(p,q) + b(q,u) &= 0 && \text{for all } q \in H(\mathrm{div},\Omega),\\ b(p,v) = -F(v) &:= -\int_\Omega fv \,\mathrm{dx} && \text{for all } v \in L^2(\Omega). \end{aligned}$$

Therein, the bilinear forms model the L^2 scalar product and the divergence term,

$$a(p,q) := \int_\Omega p\cdot q \,\mathrm{dx} \quad \text{and} \quad b(q,v) := \int_\Omega v \operatorname{div} q \,\mathrm{dx}\,.$$

It is well established that the weak solution $u \in V := H_0^1(\Omega)$ to $-\Delta u = f$ in Ω specifies the flux $p := \nabla u$ and that the two formulations are equivalent and allow for a unique solution [BBF13].

Given an admissible triangulation $\mathcal{T} \in \mathbb{T}$, let $\Sigma(\mathcal{T}) \times U(\mathcal{T})$ be a stable pair of discrete subspaces with the Raviart–Thomas or Brezzi–Douglas–Marini finite element space $\Sigma(\mathcal{T}) \subset H(\operatorname{div}, \Omega)$ of order $k \in \mathbb{N}_0$ (note the index shift of k in other references) and $U(\mathcal{T}) = \mathcal{P}_k(\mathcal{T})$ based on triangles for $n = 2$ or tetrahedra for $n = 3$; cf., e.g., [BBF13] for the precise definition and stability and commuting diagram properties of those finite element spaces. In particular, there exists a unique solution $(p_h, u_h) \in \Sigma(\mathcal{T}) \times U(\mathcal{T})$ to the discrete problem

$$a(p_h, q_h) + b(q_h, u_h) = 0 \qquad \text{for all } q_h \in \Sigma(\mathcal{T}), \tag{5.1}$$

$$b(p_h, v_h) = -F(v_h) \qquad \text{for all } v_h \in P_k(\mathcal{T}). \tag{5.2}$$

5.2. Error estimators and main result. Given the unique discrete solution (p_h, u_h) (resp., $(\widehat{p_h}, \widehat{u_h})$) with respect to the triangulation $\mathcal{T} \in \mathbb{T}$ (resp., its refinement $\hat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$), the error estimators of [Car97, CPS16] and the distance function in natural norms read

$$\begin{aligned}
\eta^2(\mathcal{T}, K) &:= |K|^{2/n}\, \|p_h - \nabla_{NC} u_h\|^2_{L^2(K)} + |K|^{2/n}\, \|\operatorname{curl} p_h\|^2_{L^2(K)}\\
&\quad + |K|^{1/n} \sum_{E \in \mathcal{E}(K)} \|[p_h]_E \times \nu_E\|^2_{L^2(E)},\\
\mu^2(K) &:= \|f - \Pi_k f\|^2_{L^2(K)} \qquad \text{for any } K \in \mathcal{T},\\
\delta^2(\mathcal{T}, \hat{\mathcal{T}}) &:= \|\widehat{p_h} - p_h\|^2_{\mathrm{H}(\mathrm{div},\Omega)} + \|\widehat{u_h} - u_h\|^2_{L^2(\Omega)}.
\end{aligned}$$

The standard notation applies to the simplex K of area or volume $|K|$ and its set $\mathcal{E}(K)$ of the three edges or four faces and the L^2 projection Π_k onto $P_k(K)$ (also denoting the L^2 projection onto $P_k(\mathcal{T})$). The jump $[\bullet]_E$ across an interior edge or face E with normal unit vector ν_E is the difference of the respective traces $[q]_E := q|_{T_+} - q|_{T_-}$ on E from the two neighboring triangles $T_\pm$. Given the homogeneous Dirichlet conditions on the boundary, the jump partner is zero; $[q]_E := q|_{\omega_E}$ for the boundary side $E \in \mathcal{E}(\partial\Omega)$ of the simplex $\overline{\omega_E}$. The above notation is 3D and curl reduces in 2D to the scalar function $\operatorname{curl} p_h := \partial p_h(1)/\partial x_2 - \partial p_h(2)/\partial x_1$ for the (piecewise) smooth vector function $p_h = (p_h(1), p_h(2))$, and $[p_h]_E \times \nu_E$ denotes the tangential component of $[p_h]_E$.

In the lowest-order case $k = 0$, the Lagrange multiplier u_h does *not* enter the estimators ($\nabla_{NC} P_0(\mathcal{T}) = 0$), and hence the distance function may be reduced to the flux approximations only. This simplification allows for a coarse initial triangulation for $k = 0$, while for $k \geq 1$, the initial triangulation $\mathcal{T}_0$ has to be sufficiently fine. The subsequent theorem asserts the axioms of this paper and so allows for optimal convergence rates of the adaptive algorithm.

THEOREM 5.1 ((A1)–(A4)). *In the lowest-order case $k = 0$, the aforementioned estimators and distance functions satisfy* (A1)–(A4) *and* (B2) *with $\Lambda_{\mathrm{ref}} = 1 = \Lambda_6$ and $\widehat{\Lambda}_3 = 0$ for $\mathcal{R}(\mathcal{T}, \hat{\mathcal{T}}) := \mathcal{T} \setminus \hat{\mathcal{T}}$ for $n = 2$; for $n = 3$, $\mathcal{R}(\mathcal{T}, \hat{\mathcal{T}})$ is $\mathcal{T} \setminus \hat{\mathcal{T}}$ plus one extra layer of tetrahedra around it in $\mathcal{T}$.*

In the remaining cases $k \geq 1$, this holds under the additional condition that the mesh-size $h_{\max}$ of the initial mesh $\mathcal{T}_0$ is sufficiently small.

The estimator is reliable and efficient [Car97, CPS16] in that the exact (resp., discrete) solution (p,u) (resp., (p_h,u_h) with respect to $\mathcal{T}\in\mathbb{T}$) satisfies

$$\sigma(\mathcal{T}) \approx \|p-p_h\|_{H(\mathrm{div},\Omega)} + \|u-u_h\|_{L^2(\Omega)}.$$

The divergence contribution in the $H(\mathrm{div},\Omega)$-norm of $p-p_h$ on the right-hand side is $\|f-\Pi_k f\|_{L^2(\Omega)}$, and so the right-hand side includes the data approximation part $\mu(\mathcal{T})$. Consequently, the optimal rates of the estimators are equivalent to the optimal rates of the errors in terms of nonlinear approximation classes with respect to the natural norms in $H(\mathrm{div})\times L^2$ of the mixed FEM.

5.3. Proof of (A1)–(A2). The proof of (A1)–(A2) follows [CKNS08] and relies on all kinds of elementary (reverse) triangle inequalities and Cauchy inequalities plus on the following lemma.

LEMMA 5.2 (discrete jump control). *There exists a universal constant C_{jc}, which depends on the shape regularity in $\mathbb{T}$ and the degree $k\in\mathbb{N}_0$, such that any $g\in P_k(\mathcal{T})$ with jumps*

$$[g]_E = \begin{cases} (g|_{T_+})|_E - (g|_{T_-})|_E & \text{for } E\in\mathcal{E}(\Omega) \text{ with } E=\partial T_+\cap\partial T_-,\\ g|_E & \text{for } E\in\mathcal{E}(\partial\Omega)\cap\mathcal{E}(K)\end{cases}$$

across any side $E\in\mathcal{E}$ satisfies

$$\sum_{K\in\mathcal{T}} |K|^{1/n} \sum_{E\in\mathcal{E}(K)} \|[g]_E\|^2_{L^2(E)} \leq C^2_{\mathrm{jc}} \|g\|^2_{L^2(\Omega)}.$$

Proof. Recall the discrete trace inequality [DPE12, p. 27] on a side $E\in\mathcal{E}(K)$ of a simplex $K\in\mathcal{T}$ for the polynomial $g|_K$ of degree at most k in the form

$$|K|^{1/(2n)}\,\|g|_K\|_{L^2(E)} \leq C_{\mathrm{dtr}}\,\|g\|_{L^2(K)}$$

with a constant C_{dtr}, which depends on the shape regularity in $\mathbb{T}$ and on $k\in\mathbb{N}_0$. The contributions to the left-hand side of the asserted inequality for an interior side $E=\partial T_+\cap\partial T_-$ with patch $\omega_E := \mathrm{int}(T_+\cup T_-)$ sum up to

$$\left(|T_+|^{1/n} + |T_-|^{1/n}\right)\|[g]_E\|^2_{L^2(E)}.$$

The triangle inequality, the discrete trace inequality for $K=T_\pm$, and the Cauchy inequality in $\mathbb{R}^2$ lead to

$$\begin{aligned}\|[g]_E\|^2_{L^2(E)} &\leq \left(\|g|_{T_+}\|_{L^2(E)} + \|g|_{T_-}\|_{L^2(E)}\right)^2\\ &\leq C^2_{\mathrm{dtr}}\left(|T_+|^{-1/(2n)}\|g\|_{L^2(T_+)} + |T_-|^{-1/(2n)}\|g\|_{L^2(T_-)}\right)^2\\ &\leq C^2_{\mathrm{dtr}}\left(|T_+|^{-1/n} + |T_-|^{-1/n}\right)\,\|g\|^2_{L^2(\omega_E)}.\end{aligned}$$

Let $h_\pm := |T_\pm|/(n\,|E|)$ be the heights of the two neighboring simplices $T_\pm$ which share the side E of length or area $|E|$. Then, $\left(|T_+|^{1/n}+|T_-|^{1/n}\right)\left(|T_+|^{-1/n}+|T_-|^{-1/n}\right)$ can be written as $(h_+^{1/n} + h_-^{1/n})(h_+^{-1/n} + h_-^{-1/n})$. The shape regularity in $\mathbb{T}$ bounds the quotient of the heights $h_\pm$ from above by some universal constant C_{sr}^2. Consequently,

$$\left(|T_+|^{1/n} + |T_-|^{1/n}\right) \|[g]_E\|_{L^2(E)}^2 \le C_{\mathrm{dtr}}^2 C_{\mathrm{sr}}^2 \|g\|_{L^2(\omega_E)}^2 .$$

With the same constant, this estimate holds for a boundary side $E = \partial T_+ \cap \partial\Omega$ with $\omega_E := \mathrm{int}(T_+)$ (and without $T_- := \emptyset$). The sum of all those contributions reads

$$\sum_{K\in\mathcal{T}} |K|^{1/n} \sum_{E\in\mathcal{E}(K)} \|[g]_E\|_{L^2(E)}^2 \le C_{\mathrm{dtr}}^2 C_{\mathrm{sr}}^2 \sum_{E\in\mathcal{E}} \|g\|_{L^2(\omega_E)}^2 .$$

Since at most $n+1$ of the side-patches $(\omega_E : E \in \mathcal{E})$ overlap, this proves the assertion with $C_{\mathrm{jc}} := \sqrt{n+1}\, C_{\mathrm{dtr}} C_{\mathrm{sr}}$. ∎

Proof of stability (A1) *in Theorem* 5.1. A reverse triangle inequality in $\mathbb{R}^m$ (for the number $m := |\mathcal{T} \cap \hat{\mathcal{T}}|$ of simplices in $\mathcal{T} \cap \hat{\mathcal{T}}$) implies

$$|\eta(\hat{\mathcal{T}}, \mathcal{T} \cap \hat{\mathcal{T}}) - \eta(\mathcal{T}, \mathcal{T} \cap \hat{\mathcal{T}})|^2 \le \sum_{T\in\mathcal{T}\cap\hat{\mathcal{T}}} \left(\eta(\hat{\mathcal{T}}, T) - \eta(\mathcal{T}, T)\right)^2 .$$

Each of the terms $\eta(\hat{\mathcal{T}}, T)$ and $\eta(\mathcal{T}, T)$ is a Euclidean norm in $\mathbb{R}^{3+n}$ of terms, which are Lebesgue norms and so allow for a reverse triangle inequality. With the abbreviations $g := \widehat{p_h} - p_h$ and $e := \widehat{u_h} - u_h$, this leads to

$$\begin{aligned}\left(\eta(\hat{\mathcal{T}}, T) - \eta(\mathcal{T}, T)\right)^2 \le{}& |T|^{2/n} \|g - \nabla e\|_{L^2(T)}^2 + |T|^{2/n} \|\mathrm{curl}\, g\|_{L^2(T)}^2 \\ &+ |T|^{1/n} \sum_{E\in\mathcal{E}(T)} \|[g]_E \times \nu_E\|_{L^2(E)}^2 .\end{aligned}$$

The sum over all $T \in \mathcal{T} \cap \hat{\mathcal{T}}$ involves volume and jump terms. The latter terms are bounded via Lemma 5.2. The volume terms are estimated by inverse estimates (with universal constant C_{inv}) for the polynomials and their derivatives. This and $|T| \le h_{\max}^n/n!$ result in

$$|\eta(\hat{\mathcal{T}}, \mathcal{T} \cap \hat{\mathcal{T}}) - \eta(\mathcal{T}, \mathcal{T} \cap \hat{\mathcal{T}})|^2 \le (C_{\mathrm{jc}}^2 + C_{\mathrm{inv}}^2 + 2h_{\max}^n/n!)\|g\|_{L^2(\Omega)}^2 + 2C_{\mathrm{inv}}^2 \|e\|_{L^2(\Omega)}^2 .$$

This is (A1) with $\Lambda_1 = \max\{\sqrt{C_{\mathrm{jc}}^2 + C_{\mathrm{inv}}^2 + 2h_{\max}^n/n!}, \sqrt{2}C_{\mathrm{inv}}\}$. ∎

Proof of reduction (A2) *in Theorem* 5.1. The error estimator for the m refined simplices $T \in \hat{\mathcal{T}}(K) := \{T \in \hat{\mathcal{T}} | T \subset K\}$ of $K \in \mathcal{T} \setminus \hat{\mathcal{T}}$ reads

$$\begin{aligned}\eta^2(\hat{\mathcal{T}}, \hat{\mathcal{T}} \setminus \mathcal{T}) = \sum_{T\in\hat{\mathcal{T}}\setminus\mathcal{T}} \Big(&|T|^{2/n} \|\widehat{p_h} - \nabla\widehat{u_h}\|_{L^2(T)}^2 + |T|^{2/n} \|\,\mathrm{curl}\,\widehat{p_h}\|_{L^2(T)}^2 \\ &+ |T|^{1/n} \sum_{F\in\mathcal{E}(T)} \|[\widehat{p_h}]_F \times \nu_F\|_{L^2(F)}^2\Big).\end{aligned}$$

With the abbreviations $g := \widehat{p_h} - p_h$ and $e := \widehat{u_h} - u_h$ from the previous proof, the triangle inequality in $\mathbb{R}^{(3+n)m}$ and reverse triangle inequalities in the Lebesgue norms over simplices and sides show

$$\eta(\hat{\mathcal{T}}, \hat{\mathcal{T}} \setminus \mathcal{T}) \leq \Bigg(\sum_{\substack{K \in \mathcal{T} \setminus \hat{\mathcal{T}},\\ T \in \hat{\mathcal{T}}(K)}} \Big(|T|^{2/n} \, \|p_h - \nabla u_h\|^2_{L^2(T)} + |T|^{2/n} \, \|\operatorname{curl} p_h\|^2_{L^2(T)}$$
$$+ |T|^{1/n} \sum_{F \in \mathcal{E}(T)} \|[p_h]_F \times \nu_F\|^2_{L^2(F)} \Big) \Bigg)^{1/2}$$
$$+ \Bigg(\sum_{T \in \hat{\mathcal{T}} \setminus \mathcal{T}} \Big(|T|^{2/n} \|g - \nabla e\|^2_{L^2(T)} + |T|^{2/n} \| \operatorname{curl} g\|^2_{L^2(T)}$$
$$+ |T|^{1/n} \sum_{F \in \mathcal{E}(T)} \|[g]_F \times \nu_F\|^2_{L^2(F)} \Big) \Bigg)^{1/2}.$$

Since $[p_h \times \nu_F]_F = 0$ for a side $F \in \hat{\mathcal{E}}(\operatorname{int}(K))$ in the interior of a coarse simplex, and $|T| \leq |K|/2$ for $T \in \mathcal{T}(K)$, the first term on the right-hand side of the above displayed formula $\leq \rho_2 \eta(\mathcal{T}, \mathcal{T} \setminus \hat{\mathcal{T}})$ for $\rho_2 = 2^{-1/(2n)}$. The remaining second term is estimated with Lemma 5.2 as above with $\Lambda_1 = \Lambda_2$. ∎

5.4. Proof of discrete reliability (A3). The proof of (A3) in Theorem 5.1 requires a discrete intermediate solution $(\widetilde{p_h}, \widetilde{u_h}) \in \Sigma(\hat{\mathcal{T}}) \times P_k(\hat{\mathcal{T}})$ with respect to the fine triangulation $\hat{\mathcal{T}}$ to the above Poisson model problem with a right-hand side $\Pi_k f \in P_k(\mathcal{T})$ with respect to the coarse triangulation $\mathcal{T}$. Recall that $-\operatorname{div} \widehat{p_h} = \widehat{\Pi}_k f \in P_k(\hat{\mathcal{T}})$ (the orthogonal projection of f onto $P_k(\hat{\mathcal{T}})$).

LEMMA 5.3. *It holds that* $\|\widehat{p_h} - \widetilde{p_h}\|_{\mathrm{H}(\operatorname{div},\Omega)} + \|\widehat{u_h} - \widetilde{u_h}\|_{L^2(\Omega)} \lesssim \|\widehat{\Pi}_k f - \Pi_k f\|_{L^2(\Omega)}$.

Proof. The inf sup stability of the mixed FEM on the fine level leads to the existence of a test function $(\widehat{q_h}, \widehat{v_h}) \in \Sigma(\hat{\mathcal{T}}) \times U(\hat{\mathcal{T}})$ of bounded norm in $\mathrm{H}(\operatorname{div}, \Omega) \times L^2(\Omega)$ with

$$\begin{aligned}&\|\widehat{p_h} - \widetilde{p_h}\|_{\mathrm{H}(\operatorname{div},\Omega)} + \|\widehat{u_h} - \widetilde{u_h}\|_{L^2(\Omega)}\\ &\qquad = a(\widehat{p_h} - \widetilde{p_h}, \widehat{q_h}) + b(\widehat{q_h}, \widehat{u_h} - \widetilde{u_h}) + b(\widehat{p_h} - \widetilde{p_h}, \widehat{v_h}).\end{aligned}$$

Since $(\widehat{p_h}, \widehat{u_h})$ and $(\widetilde{p_h}, \widetilde{u_h})$ solve discrete problems with respect to the fine level with the test function $(\widehat{q_h}, \widehat{v_h})$, the previous terms are equal to

$$\int_\Omega \widehat{v_h} \operatorname{div}(\widehat{p_h} - \widetilde{p_h}) \, \mathrm{dx} = \int_\Omega \widehat{v_h} (\Pi_k f - \widehat{\Pi}_k f) \, \mathrm{dx} \lesssim \|\widehat{\Pi}_k f - \Pi_k f\|_{L^2(\Omega)}. \qquad ∎$$

The following lemma is key in the L^2 flux error control [CHX09, CR11, HX12].

LEMMA 5.4 (see [HX12]). *It holds that* $\|\widetilde{p_h} - p_h\|_{L^2(\Omega)} \lesssim \eta(\mathcal{T}, \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}}))$.

Proof. This is shown in (3.2)–(3.3) of Lemma 3.1 in [HX12], where the right-hand side f is replaced by $\widehat{\Pi}_k f$. We refer the reader to [CHX09, CR11, HX12] for further details and give merely an overview over the arguments in this paper: Since the L^2 vector function $\widetilde{p_h} - p_h$ is divergence-free, the discrete Helmholtz decomposition (proven in Lemma 2.6 in [HX12]) asserts that it is equal to some rotation $\operatorname{curl} \widehat{\beta_h}$ for some function $\widehat{\beta_h}$ of piecewise polynomials of degree $k+1$ or $k+2$ (with $\operatorname{curl} \widehat{\beta_h} := (-\partial \widehat{\beta_h}/\partial x_2, \partial \widehat{\beta_h}/\partial x_1))$ in $n = 2$ and for some first-kind Nédélec finite element function in $n = 3$ dimensions. An appropriate quasi-interpolation leads to $\operatorname{curl} \beta_h$, and, according to the discrete equations, shows $a(\widetilde{p_h} - p_h, \operatorname{curl}\beta_h) = 0 = a(\widetilde{p_h}, \operatorname{curl}\widehat{\beta_h})$. This and piecewise integration by parts followed by trace inequalities and approximation and stability properties of the quasi-interpolation error $\widehat{\beta_h} - \beta_h$ conclude the proof. $\square$

The new ingredient of the discrete reliability proof for the natural norms is the following lemma based on the surjective operator $\operatorname{div} : H_0^1(\Omega;\mathbb{R}^n) \to L_0^2(\Omega)$.

LEMMA 5.5. *It holds that* $\|\widetilde{u_h}-u_h\|_{L^2(\Omega)} \lesssim \|\widetilde{p_h}-p_h\|_{L^2(\Omega)}+\|h_{\mathcal{T}}(p_h-\nabla_{NC}u_h)\|_{L^2(\Omega')}$ *for the subdomain* Ω' *covered by the refined region* $\mathcal{T}\setminus\hat{\mathcal{T}}$.

Proof. Extend $\widetilde{u_h}-u_h$ by zero outside of Ω, and consider this expansion as a right-hand side of a Poisson problem on a large ball that includes Ω with zero boundary values on the boundary of this ball. Let $z \in H^2(\Omega)$ be the restriction of the smooth solution to Ω to infer

$$-\Delta z = \widetilde{u_h} - u_h \text{ a.e. in } \Omega \quad \text{and} \quad \|z\|_{H^2(\Omega)} \lesssim \|\widetilde{u_h} - u_h\|_{L^2(\Omega)}. \tag{5.3}$$

The mixed FEMs under consideration allow for some interpolation operator $I_F : H^1(\Omega;\mathbb{R}^n) \to \Sigma(\mathcal{T})$ (resp., $\widehat{I_F} : H^1(\Omega;\mathbb{R}^n) \to \Sigma(\hat{\mathcal{T}})$) with the commuting diagram property $\Pi_k \operatorname{div}\Psi = \operatorname{div} I_F\Psi$ and the first-order approximation properties $\|h_{\mathcal{T}}^{-1}(\Psi - I_F\Psi)\| \lesssim \|\Psi\|_{H^1(\Omega)}$ for any $\Psi \in H^1(\Omega;\mathbb{R}^n)$ (and corresponding results for $\widehat{I_F}$ and $\widehat{\Pi_k}$ with respect to $\hat{\mathcal{T}}$). The commuting property leads to

$$\|\widetilde{u_h}-u_h\|^2_{L^2(\Omega)} = -\int_\Omega (\widetilde{u_h}-u_h)\operatorname{div}\nabla z \,\mathrm{dx} = \int_\Omega u_h \operatorname{div} I_F\nabla z \,\mathrm{dx} - \int_\Omega \widetilde{u_h} \operatorname{div}\widehat{I_F}\nabla z \,\mathrm{dx}\,.$$

The discrete equations on the fine and coarse levels lead to

$$\|\widetilde{u_h}-u_h\|^2_{L^2(\Omega)} = a(\widetilde{p_h},\widehat{I_F}\nabla z)-a(p_h,I_F\nabla z) = a(\widetilde{p_h}-p_h,\widehat{I_F}\nabla z)+a(p_h,\widehat{I_F}\nabla z-I_F\nabla z).$$

The Cauchy inequality in L^2, the stability of $\widehat{I_F}$, and the bound (5.3) show

$$a(\widetilde{p_h}-p_h,\widehat{I_F}\nabla z) \lesssim \|\widetilde{u_h}-u_h\|_{L^2(\Omega)}\,\|\widetilde{p_h}-p_h\|_{L^2(\Omega)}.$$

Notice that $\widehat{I_F}\nabla z - I_F\nabla z$ vanishes on $\mathcal{T}\cap\hat{\mathcal{T}}$, and let a' denote the L^2 scalar product over Ω'. Then, the split

$$a(p_h,\widehat{I_F}\nabla z-I_F\nabla z) = a'(p_h,\widehat{I_F}\nabla z-I_F\nabla z) = a'(p_h,\nabla z-I_F\nabla z)-a'(p_h,\nabla z-\widehat{I_F}\nabla z)$$

allows arguments of [CPS16] on the coarse and fine levels. The Raviart–Thomas and the Brezzi–Douglas–Marini finite element spaces lead to the L^2 orthogonality of $\nabla z - \widehat{I_F}$ onto $\nabla_{NC}u_h \in P_{k-1}(\mathcal{T};\mathbb{R}^n)$ (with the convention $P_{-1} := \{0\}$). This, the elementwise first-order approximation property (on the fine level even with some smaller $h_{\hat{\mathcal{T}}}$), and (5.3) in the end show

$$\begin{aligned}
|a'(p_h,\nabla z-\widehat{I_F}\nabla z)| &= |a'(p_h-\nabla_{NC}u_h,\nabla z-\widehat{I_F}\nabla z)| \\
&\le \|h_{\mathcal{T}}(p_h-\nabla_{NC}u_h)\|_{L^2(\Omega')}\,\|h_{\mathcal{T}}^{-1}(\nabla z-\widehat{I_F}\nabla z)\|_{L^2(\Omega')} \\
&\lesssim \|h_{\mathcal{T}}(p_h-\nabla_{NC}u_h)\|_{L^2(\Omega')}\,\|\widetilde{u_h}-u_h\|_{L^2(\Omega)}.
\end{aligned}$$

The same arguments apply verbatim to the term $a'(p_h,\nabla z - I_F\nabla z)$ with the mesh-size $h_{\mathcal{T}}$ on the coarse level as displayed above. The combination of the preceding four displayed estimates concludes the proof:

$$\|\widetilde{u_h}-u_h\|^2_{L^2(\Omega)} \lesssim \|\widetilde{u_h}-u_h\|_{L^2(\Omega)}(\|\widetilde{p_h}-p_h\|_{L^2(\Omega)}+\|h_{\mathcal{T}}(p_h-\nabla_{NC}u_h)\|_{L^2(\Omega')}). \quad \square$$

Proof of (A3) *in Theorem* 5.1. The triangle inequality and the three aforementioned lemmas lead to the asserted discrete reliability. Lemma 5.3, $\widehat{\Pi}_k f = \Pi_k f$ in $\Omega \setminus \Omega'$, and a triangle inequality show

$$\delta(\mathcal{T}, \hat{\mathcal{T}}) \lesssim \|\widetilde{p_h} - p_h\|_{\mathrm{H}(\mathrm{div},\Omega)} + \|\widetilde{u_h} - u_h\|_{L^2(\Omega)} + \|\widehat{\Pi}_k f - \Pi_k f\|_{L^2(\Omega')}.$$

The combination with Lemmas 5.4–5.5 concludes the proof. ∎

5.5. Proof of quasiorthogonality (A4). The flux or stress approximations in the mixed finite element schemes allow for an orthogonality up to some perturbation with data oscillations. This leads to a proof of the axiom (A4$_\varepsilon$) for any $\varepsilon > 0$ and, together with (A1)–(A2), Theorem 3.1 provides (A4). The analysis for the error of the Lagrange multiplier is based on the smaller L^2 error of the L^2 projection onto the discrete space of the error in the displacement variable for sufficiently fine meshes.

LEMMA 5.6. *Given the polyhedral Lipschitz domain* Ω*, there exists some index* $\alpha > 1/2$ *such that any discrete solution* $(p_h, u_h) \in \Sigma(\mathcal{T}) \times P_k(\mathcal{T})$ *and the* L^2 *projection* $\Pi_k u$ *of the displacement* u *of the exact solution* $p = \nabla u$ *onto* $P_k(\mathcal{T})$ *with respect to the triangulation* $\mathcal{T} \in \mathbb{T}$ *satisfies* $\|\Pi_k u - u_h\|_{L^2(\Omega)} \lesssim h_{\max}^\alpha \|p - p_h\|_{L^2(\Omega)} + \mathrm{osc}(f, \mathcal{T})$.

If $k \geq 1$*, then* $\|\Pi_k u - u_h\|_{L^2(\Omega)} \lesssim h_{\max}^\alpha \big(\|p - p_h\|_{L^2(\Omega)} + \mathrm{osc}(f, \mathcal{T})\big)$.

Proof. Adopt the notation and arguments from Lemma 5.5 for the right-hand side $\Pi_k u - u_h \in P_k(\mathcal{T})$ of the Poisson equation applied to Ω with weak solution $z \in H_0^1(\Omega)$. The reduced elliptic regularity of the polyhedral domain Ω [Dau88] leads to the index $0 < \alpha \leq 1$ with

$$-\Delta z = \Pi_k u - u_h \text{ in } \Omega \quad \text{and} \quad \|z\|_{H^{1+\alpha}(\Omega)} \lesssim \|\Pi_k u - u_h\|_{L^2(\Omega)}.$$

The interpolation operator I_F can be extended to $I_F : H^\alpha(\Omega; \mathbb{R}^n) \cap H(\mathrm{div}, \Omega) \to \Sigma(\mathcal{T})$ and allows for the error estimate

$$\|\nabla z - I_F \nabla z\|_{L^2(\Omega)} \lesssim h_{\max}^\alpha \|z\|_{H^{1+\alpha}(\Omega)} \lesssim h_{\max}^\alpha \|\Pi_k u - u_h\|_{L^2(\Omega)}. \tag{5.4}$$

The commuting diagram property shows that the piecewise polynomial right-hand side is equal to $u_h - \Pi_k u = \mathrm{div}\, \nabla z = \mathrm{div}\, I_F \nabla z$. This, the discrete equation for (p_h, u_h), and an integration by parts with $u \in H_0^1(\Omega)$ result in

$$\begin{aligned}\|\Pi_k u - u_h\|_{L^2(\Omega)}^2 &= \int_\Omega (u_h - \Pi_k u) \Delta z \,\mathrm{dx} = \int_\Omega (u_h - u) \,\mathrm{div}\, I_F \nabla z \,\mathrm{dx} \\ &= \int_\Omega (p - p_h) \cdot I_F \nabla z \,\mathrm{dx}.\end{aligned}$$

The approximation properties (5.4) of the Fortin interpolation show

$$\int_\Omega (p_h - p) \cdot (\nabla z - I_F \nabla z) \,\mathrm{dx} \lesssim h_{\max}^\alpha \|p - p_h\|_{L^2(\Omega)} \|\Pi_k u - u_h\|_{L^2(\Omega)}.$$

An integration by parts with $z \in H_0^1(\Omega)$ and the discrete and exact equations lead to

$$\int_\Omega (p - p_h) \cdot \nabla z \,\mathrm{dx} = \int_\Omega z\, (f - \Pi_k f) \,\mathrm{dx} = \int_\Omega (z - \Pi_k z)(f - \Pi_k f) \,\mathrm{dx}.$$

A piecewise Poincaré inequality and the above regularity of z lead to the bound

$$\int_\Omega (p_h - p) \cdot \nabla z \,\mathrm{dx} \lesssim h_{\max}^\beta \,\mathrm{osc}(f, \mathcal{T}) \|\Pi_k u - u_h\|_{L^2(\Omega)}$$

with $\beta = 0$ for $k = 0$ and $\beta = \alpha$ for $k \geq 1$. The combination of the preceding four displayed formulas results in

$$\|\Pi_k u - u_h\|^2_{L^2(\Omega)} \lesssim \|\Pi_k u - u_h\|_{L^2(\Omega)} \big(h^\alpha_{\max}\|p - p_h\|_{L^2(\Omega)} + h^\beta_{\max} \operatorname{osc}(f, \mathcal{T})\big). \qquad \square$$

A corresponding lemma for the flux error is known from [CHX09, CR11, HX12].

LEMMA 5.7 (see [HX12]). *It holds that* $\|\widetilde{p_h} - p_h\|_{L^2(\Omega)} \lesssim \eta(\mathcal{T}, \mathcal{R}(\mathcal{T}, \hat{\mathcal{T}}))$.

Proof. This is shown in (3.1) of Lemma 3.1 in [HX12]. $\square$

The proof of (A4) in Theorem 5.1 recalls the L^2 quasiorthogonality of the flux errors of [CHX09, Theorem 3.2] or [CR11, Lemma 4.3 and (4.4)] in the form

$$\|p_{\ell+1} - p_\ell\|^2_{L^2(\Omega)} + \|p - p_{\ell+1}\|^2_{L^2(\Omega)} - \|p - p_\ell\|^2_{L^2(\Omega)} \lesssim \|p - p_{\ell+1}\|_{L^2(\Omega)} \operatorname{osc}(f_{\ell+1}, \mathcal{T}_\ell).$$

The mixed FEM fixes the divergence $-\operatorname{div} p_\ell = \Pi_\ell f =: f_\ell$ and their orthogonality

$$\|f_{\ell+1} - f_\ell\|^2_{L^2(\Omega)} + \|f - f_{\ell+1}\|^2_{L^2(\Omega)} - \|f - f_\ell\|^2_{L^2(\Omega)} = 0$$

leads (for all $\ell \in \mathbb{N}$) in the aforementioned L^2 quasiorthogonality to

$$\|p_{\ell+1}-p_\ell\|^2_{H(\operatorname{div},\Omega)}+\|p-p_{\ell+1}\|^2_{H(\operatorname{div},\Omega)}-\|p-p_\ell\|^2_{H(\operatorname{div},\Omega)} \lesssim \|p-p_{\ell+1}\|_{L^2(\Omega)} \operatorname{osc}(f_{\ell+1}, \mathcal{T}_\ell).$$

For any $0 < \varepsilon$ with $\varepsilon\Lambda_3 < 1$ and the multiplicative constant $C \approx 1$ hidden in the notation $\lesssim$, the sum of those estimates results for any $\ell, m \in \mathbb{N}_0$ in

$$\begin{aligned}\sum_{k=\ell}^{\ell+m} \|p_{k+1} - p_k\|^2_{H(\operatorname{div},\Omega)} \leq \|p - p_\ell\|^2_{H(\operatorname{div},\Omega)} + \varepsilon/\Lambda_3 \sum_{k=\ell}^{\ell+m-1} \|p - p_{k+1}\|^2_{L^2(\Omega)} \\ + C^2\Lambda_3/\varepsilon \sum_{k=\ell}^{\ell+m} \operatorname{osc}^2(f_{k+1}, \mathcal{T}_k).\end{aligned} \tag{5.5}$$

For a sequence of uniformly refined meshes $\hat{\mathcal{T}}$, the discrete reliability (A3) leads (in the limit $h_{\widetilde{\mathcal{T}}} \to 0$) to the reliability of [CPS16], with the abbreviation $\sigma^2_\ell := \sigma^2(\mathcal{T}_\ell)$,

$$\|p - p_\ell\|^2_{H(\operatorname{div},\Omega)} + \|u - u_\ell\|^2_{L^2(\Omega)} \leq \Lambda_3\, \sigma^2_\ell \quad \text{for all } \ell \in \mathbb{N}_0. \tag{5.6}$$

The oscillation $\operatorname{osc}(f_{k+1}, \mathcal{T}_k) = \|h_\ell(f_{k+1} - f_k)\|_{L^2(\Omega)}$ is bounded by $\|h_\ell\|_{L^\infty(\Omega)}\, \|f_{k+1} - f_k\|_{L^2(\Omega)}$. The L^2 orthogonality of the integrands shows

$$\sum_{k=\ell}^{\ell+m} \operatorname{osc}^2(f_{k+1}, \mathcal{T}_k) \leq h^2_{\max}\, \|f_{\ell+m+1} - f_\ell\|^2_{L^2(\Omega)} \leq h^2_{\max}\, \|f - f_\ell\|^2_{L^2(\Omega)}.$$

The combination of the previous estimates with (5.5) leads to the quasiorthogonality (A4$_\varepsilon$) for the flux contribution in the form

$$\sum_{k=\ell}^{\ell+m} \|p_{k+1} - p_k\|^2_{H(\operatorname{div},\Omega)} \leq \Lambda_3\sigma^2_\ell + \varepsilon \sum_{k=\ell+1}^{\ell+m} \sigma^2_k + C^2\Lambda_3 h^2_{\max}/\varepsilon\, \mu^2(\mathcal{T}_\ell). \tag{5.7}$$

Since the Lagrange multipliers are not orthogonal in general, the critical term is controlled with Lemma 5.6 with the L^2 projection Π_ℓ onto $P_k(\mathcal{T}_\ell)$ with respect to $\mathcal{T}_\ell$ with maximal mesh-size $\leq h_{\max}$ by

$$\|\Pi_\ell u - u_\ell\|_{L^2(\Omega)} \lesssim h^\alpha_{\max}\|p - p_\ell\|_{L^2(\Omega)} + \operatorname{osc}(f, \mathcal{T}_\ell).$$

This estimate on the level $\ell+1$ and elementary algebra result in

$$
\begin{aligned}
&\|u_{\ell+1}-u_\ell\|^2_{L^2(\Omega)} - \|u-u_\ell\|^2_{L^2(\Omega)} + \|u-u_{\ell+1}\|^2_{L^2(\Omega)}\\
&= 2(u-u_{\ell+1}, u_\ell - u_{\ell+1})_{L^2(\Omega)} = 2(\Pi_{\ell+1}u - u_{\ell+1}, u_\ell - u_{\ell+1})_{L^2(\Omega)}\\
&\lesssim \|u_{\ell+1}-u_\ell\|_{L^2(\Omega)}(h^\alpha_{\max}\|p-p_{\ell+1}\|_{L^2(\Omega)} + \operatorname{osc}(f,\mathcal{T}_{\ell+1}))\\
&\lesssim h^\alpha_{\max}\|u_{\ell+1}-u_\ell\|_{L^2(\Omega)}\,\|p-p_{\ell+1}\|_{H(\operatorname{div},\Omega)}.
\end{aligned}
$$

(Utilize $0<\alpha\le 1$ and $\operatorname{div}(p-p_{\ell+1}) = \Pi_{\ell+1}f - f$ in the final step.) The best approximation of the mixed FEM implies $||p-p_{\ell+1}\|_{H(\operatorname{div},\Omega)} \lesssim \|p-p_\ell\|_{H(\operatorname{div},\Omega)} + \|u-u_\ell\|_{L^2(\Omega)}$. This and (5.6) show

$$
\begin{aligned}
&1/2\,\|u_{\ell+1}-u_\ell\|^2_{L^2(\Omega)} - \|u-u_\ell\|^2_{L^2(\Omega)} + \|u-u_{\ell+1}\|^2_{L^2(\Omega)}\\
&\lesssim h^{2\alpha}_{\max}\|p-p_{\ell+1}\|^2_{H(\operatorname{div},\Omega)} \lesssim h^{2\alpha}_{\max}\sigma^2_\ell.
\end{aligned}
$$

The sum over all those inequalities reads

$$
\sum_{k=\ell}^{\ell+m}\|u_{k+1}-u_k\|^2_{L^2(\Omega)} \lesssim \|u-u_\ell\|^2_{L^2(\Omega)} + h^{2\alpha}_{\max}\sum_{k=\ell+1}^{\ell+m}\sigma^2_k.
$$

The combination of (5.6) with the previous flux error control leads to

$$
\sum_{j=\ell}^{\ell+m}\delta^2(\mathcal{T}_j,\mathcal{T}_{j+1}) \lesssim \sigma^2_\ell + h^{2\alpha}_{\max}\sum_{j=\ell+1}^{\ell+m}\sigma^2_j. \tag{5.8}
$$

In other words, some generic constant $\Lambda_4 \approx 1$ and $\varepsilon = \Lambda_4 h^{2\alpha}_{\max}$ satisfy (A4$_\varepsilon$) and $\varepsilon \to 0$ as $h_{\max}\to 0$; further details are omitted for $k\ge 1$.

It remains to prove (A4) for $k=0$ *without* any assumption on the smallness of $h_{\max}$. In case $k=0$, the piecewise derivatives of u_h in $\eta(\mathcal{T})$ vanish, and the error estimator $\eta(\mathcal{T})$ does *not* contain any u_h at all. Set $\delta'(\mathcal{T},\hat{\mathcal{T}}) := \|\widehat{p_h} - p_h\|_{H(\operatorname{div},\Omega)}$, and observe from the arguments of this section that (A1)–(A3) hold with δ replaced by δ' (even with possibly smaller constants). The estimate (5.7) is (A4$_\varepsilon$) for any $0<\varepsilon<1/\Lambda_3$ when δ is replaced by δ'. Utilizing the general theory with δ replaced by δ', Theorem 4.2 implies plain convergence of the error estimator. Hence the right-hand side of (5.8) is $\lesssim \sigma^2_\ell$, and this proves (A4) for the original distance function δ and no reference to $h_{\max}$ small. ∎

Acknowledgment. This work was rewritten while the first author enjoyed the hospitality of the Hausdorff Research Institute of Mathematics in Bonn, Germany, during the Hausdorff Trimester Program *Multiscale Problems: Algorithms, Numerical Analysis and Computation.*

REFERENCES

[Alo96] A. Alonso, *Error estimators for a mixed method*, Numer. Math., 74 (1996), pp. 385–395.

[BBF13] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Springer Ser. Comput. Math. 44, Springer, Heidelberg, 2013.

[BC17] P. Bringmann and C. Carstensen, *An adaptive least-squares FEM for the Stokes equations with optimal convergence rates*, Numer. Math., 135 (2017), pp. 459–492.

[BCS17] P. Bringmann, C. Carstensen, and G. Starke, *An adaptive least-squares FEM for linear elasticity with optimal convergence rates*, SIAM J. Numer. Anal., to appear.

[BDdV04] P. Binev, W. Dahmen, and R. de Vore, *Adaptive finite element methods with convergence rates*, Numer. Math., 97 (2004), pp. 219–268.

[BdV04] P. Binev and R. de Vore, *Fast computation in adaptive tree approximation*, Numer. Math., 97 (2004), pp. 193–217.

[BM08] R. Becker and S. Mao, *An optimally convergent adaptive mixed finite element method*, Numer. Math., 111 (2008), pp. 35–54.

[Car97] C. Carstensen, *A posteriori error estimate for the mixed finite element method*, Math. Comp., 66 (1997), pp. 465–476.

[CDN12] A. Cohen, R. DeVore, and R.H. Nochetto, *Convergence rates of AFEM with H^{-1} data*, Found. Comput. Math., 5 (2012), pp. 671–718.

[CDNP16] C. Carstensen, A. Dond, N. Nataraj, and A.K. Pani, *Error analysis of nonconforming and mixed FEMs for second-order linear non-selfadjoint and indefinite elliptic problems*, Numer. Math., 133 (2016), pp. 557–597.

[CDR16] C. Carstensen, A. Dond, and H. Rabus, *Quasi-optimality of adaptive mixed FEMs for general second-order linear elliptic problems*, in preparation, 2017.

[CFPP14] C. Carstensen, M. Feischl, M. Page, and D. Praetorius, *Axioms of adaptivity*, Comput. Math. Appl., 67 (2014), pp. 1195–1253.

[CHX09] L. Chen, M. Holst, and J. Xu, *Convergence and optimality of adaptive mixed finite element methods*, Math. Comp., 78 (2009), pp. 35–53.

[CKNS08] J.M. Cascon, C. Kreuzer, R.H. Nochetto, and K.G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal., 46 (2008), pp. 2524–2550, https://doi.org/10.1137/07069047X.

[CP15] C. Carstensen and E.-J. Park, *Convergence and optimality of adaptive least squares finite element methods*, SIAM J. Numer. Anal., 53 (2015), pp. 43–62, https://doi.org/10.1137/130949634.

[CPS16] C. Carstensen, D. Peterseim, and A. Schröder, *The norm of a discretized gradient in $H(\mathrm{div})^*$ for a posteriori finite element error analysis*, Numer. Math., 132 (2016), pp. 519–539.

[CR11] C. Carstensen and H. Rabus, *An optimal adaptive mixed finite element method*, Math. Comp., 80 (2011), pp. 649–667.

[Dau88] M. Dauge, *Elliptic Boundary Value Problems on Corner Domains: Smoothness and Asymptotics of Solutions*, Lecture Notes in Math. 1341, Springer, Berlin, 1988.

[DPE12] D.A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Mathématiques & Applications (Berlin) [Mathematics & Applications] 69, Springer, Heidelberg, 2012.

[GSS14] D. Gallistl, M. Schedensack, and R. Stevenson, *A remark on newest vertex bisection in any space dimension*, Comput. Methods Appl. Math., 14 (2014), pp. 317–320.

[HX12] J.G. Huang and Y.F. Xu, *Convergence and complexity of arbitrary order adaptive mixed element methods for the Poisson equation*, Sci. China Math., 55 (2012), pp. 1083–1098.

[Rab15] H. Rabus, *Quasi-optimal convergence of AFEM based on separate marking—Parts* I *and* II, J. Numer. Math., 23 (2015), pp. 137–156, 157–174.

[Ste07] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math., 7 (2007), pp. 245–269.

[Ste08] R. Stevenson, *The completion of locally refined simplicial partitions created by bisection*, Math. Comp., 77 (2008), pp. 227–241.