MATHEMATICS OF COMPUTATION
https://doi.org/10.1090/mcom/3474
Article electronically published on September 4, 2019

COLLECTIVE MARKING FOR ADAPTIVE LEAST-SQUARES FINITE ELEMENT METHODS WITH OPTIMAL RATES

CARSTEN CARSTENSEN

Abstract. All previously known optimal adaptive least-squares finite element methods (LSFEMs) combine two marking strategies with a separate L^2 data approximation as a consequence of the natural norms equivalent to the least-squares functional. The algorithm and its analysis in this paper circumvent the natural norms in a div-LSFEM model problem with lowest-order conforming and mixed finite element functions and allow for a simple collective Dörfler marking for the first time. A refined analysis provides discrete reliability and quasi-orthogonality in the weaker norms $L^2 \times H^1$ rather than $H(\mathrm{div}) \times H^1$ and replaces data approximation terms by data oscillations. The optimal convergence rates then follow for the lowest-order version from the axioms of adaptivity for the newest-vertex bisection without restrictions on the initial mesh-size in any space dimension.

1. Introduction

In its prominent divergence form, the least-squares finite element method (LSFEM) for the Poisson model problem minimises the least-squares functional

$$LS(f;\tau_{LS},v_{LS}) := \|\tau_{LS} - \nabla v_{LS}\|^2_{L^2(\Omega)} + \|f + \operatorname{div}\tau_{LS}\|^2_{L^2(\Omega)} \tag{1}$$

over all $(\tau_{LS},v_{LS}) \in X(\mathcal{T}) := RT_0(\mathcal{T}) \times S^1_0(\mathcal{T})$ for a given right-hand side $f \in L^2(\Omega)$ in a polyhedral bounded Lipschitz domain $\Omega \subset \mathbb{R}^n$ partitioned into shape-regular simplices in the triangulation $\mathcal{T}$. The lowest-order Raviart–Thomas function space $RT_0(\mathcal{T}) \subset H(\mathrm{div},\Omega)$ and the conforming first-order polynomials $S^1_0(\mathcal{T}) \subset H^1_0(\Omega)$ [3–5] allow for a unique minimizer (σ_{LS},u_{LS}) of the least-squares functional $LS(f;\bullet)$ with quasi-optimal convergence towards the solution $(\sigma,u) := (\nabla u,u) \in X := H(\mathrm{div},\Omega)\times H^1_0(\Omega)$ to $f + \Delta u = 0$ in Ω. The quasi-optimal convergence follows in the norms in X from the well-established equivalence

$$LS(f;\tau_{LS},v_{LS}) \approx \|\sigma - \tau_{LS}\|^2_{H(\mathrm{div})} + \|\!|u - v_{LS}|\!\|^2 \tag{2}$$

for all $(\tau_{LS},v_{LS}) \in X(\mathcal{T})$ (and even for all test functions in X) [2]; cf. [14] for the equivalence constants in (2) and asymptotic exactness, where $H^1_0(\Omega)$ is endowed with the energy norm $\|\!|\bullet|\!\| := |\bullet|_{H^1(\Omega)}$.

Received by the editor February 24, 2018, and, in revised form, November 23, 2018.
2010 *Mathematics Subject Classification.* Primary 65N30, 65N12, 65N15.
Key words and phrases. Adaptivity, finite element method, least-squares, nonconforming, a priori, a posteriori, adaptive mesh-refinement, optimal convergence rates, axioms of adaptivity.
This work was supported in the project *Foundation and application of generalized mixed FEM towards nonlinear problems in solid mechanics* (CA 151/22-2) by the DFG via the priority program 1748 *Reliable Simulation Techniques in Solid Mechanics, Development of Non-standard Discretization Methods, Mechanical and Mathematical Analysis*.

The natural adaptive LSFEM evaluates the contribution $\|\sigma_{LS} - \nabla u_{LS}\|^2_{L^2(T)} + \|f + \operatorname{div}\sigma_{LS}\|^2_{L^2(T)}$ of a simplex $T \in \mathcal{T}$ to the least-squares functional as a refinement indicator in a (collective) Dörfler marking strategy. The corresponding analysis in [12] guarantees convergence only for a *large* bulk parameter Θ, while all positive theoretical results of optimal convergence rates require some *very small* Θ [1, 7, 13, 15, 17]. This conflict in the choice of Θ led to the design of alternative error estimators in a series of papers [6, 11–13], in which the norm of X enforces a separate marking strategy [13] for the data error $\|f - \Pi f\|_{L^2(\Omega)}$ of the right-hand side f and its piecewise constant approximation Πf.

The refined analysis on the collective marking of this paper in weaker norms can circumvent this data approximation term $\|f - \Pi f\|_{L^2(\Omega)}$ in this form but generates the (much smaller) data oscillation $\operatorname{osc}(f, \mathcal{T}) := \|h_\mathcal{T}(f - \Pi f)\|_{L^2(\Omega)}$ with an extra piecewise constant factor $h_\mathcal{T}$, the mesh-size $h_\mathcal{T}|_K := h_K := |K|^{1/n}$ for any simplex $K \in \mathcal{T}$ of volume $|K|$. The point of departure is a novel equivalence

$$LS(\Pi f; \sigma_{LS}, u_{LS}) + \operatorname{osc}^2(f, \mathcal{T}) \approx \|\sigma - \sigma_{LS}\|^2_{L^2(\Omega)} + \|\!|u - u_{LS}|\!\|^2 + \operatorname{osc}^2(f, \mathcal{T}) \tag{3}$$

for exact solve in the sense that (σ_{LS}, u_{LS}) is the unique minimizer of (1) in $X(\mathcal{T})$. (Notice that $LS(f; \sigma_{LS}, u_{LS}) - LS(\Pi f; \sigma_{LS}, u_{LS}) = \|f - \Pi f\|^2_{L^2(\Omega)}$, while the right-hand side of (3) involves the smaller $\operatorname{osc}^2(f, \mathcal{T})$. The point is that $\sigma - \sigma_{LS}$ appears on the, resp., right-hand sides in the $H(\operatorname{div})$ norm in (2) and in the L^2 norm in (3).)

This paper introduces an alternative error analysis for the $L^2 \times H^1$ norms in (3) that satisfies the axioms of adaptivity [7, 13] and therefore leads to optimal convergence rates in adaptive mesh-refinements. In the weaker norms at hand, the quasi-orthogonality becomes less trivial and relies on a detailed analysis as in [10] with a surprise: The collective marking adaptive LSFEM converges with optimal rates *without any further restrictions* on the initial mesh $\mathcal{T}_0$.

The collective marking solely utilizes the residual-based explicit error estimator

$$\begin{aligned} \eta^2(T) &:= |T|^{2/n}\|\operatorname{div}\sigma_{LS}\|^2_{L^2(T)} + \operatorname{osc}^2(f, T) \\ &\quad + |T|^{1/n} \sum_{E \in \mathcal{E}(T)} \Big(\|[\sigma_{LS} - \nabla u_{LS}]_E\|^2_{L^2(E\setminus\partial\Omega)} + \|(\sigma_{LS} - \nabla u_{LS}) \times \nu_E\|^2_{L^2(E\cap\partial\Omega)} \Big) \end{aligned} \tag{4}$$

for each $T \in \mathcal{T}$ with jumps $[\bullet]_E$ across the side E of T with unit normal ν_E. The sum $\eta^2(\mathcal{T})$ of those contributions $\eta^2(T)$ for all $T \in \mathcal{T}$ defines the alternative error estimator, which is reliable and efficient in the sense that

$$\begin{aligned} LS(\Pi f; \sigma_{LS}, u_{LS}) &\lesssim \|\sigma - \sigma_{LS}\|^2_{L^2(\Omega)} + \|\!|u - u_{LS}|\!\|^2 \\ &\lesssim \eta^2(\mathcal{T}) \lesssim LS(\Pi f; \sigma_{LS}, u_{LS}) + \operatorname{osc}^2(f, \mathcal{T}). \end{aligned} \tag{5}$$

This paper establishes optimal convergence rates of the subsequent novel adaptive algorithm ACLSFEM based on the axioms of adaptivity [7, 13] with collective (Dörfler) marking.

Algorithm ACLSFEM

Input: regular triangulation $\mathcal{T}_0$ and parameter $0 < \Theta \ll 1$
for $\ell = 0, 1, 2, \ldots$ **do**
Compute the minimizer (σ_ℓ, u_ℓ) of the least-squares functional (1) in $X(\mathcal{T}_\ell)$
Compute $\eta_\ell(T)$ for any $T \in \mathcal{T}_\ell$ from (4)
Mark (almost) minimal subset $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ with $\Theta \sum_{T\in\mathcal{T}_\ell} \eta_\ell^2(T) \leq \sum_{T\in\mathcal{M}_\ell} \eta_\ell^2(T)$
Refine $\mathcal{T}_\ell$ with newest-vertex bisection to compute $\mathcal{T}_{\ell+1}$ with $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell \setminus \mathcal{T}_{\ell+1}$
end for
Output: sequence of triangulations $\mathcal{T}_\ell$ with (σ_ℓ, u_ℓ) and $\eta_\ell(\mathcal{T}_\ell)$

Undisplayed numerical experiments confirm the optimal convergence rates and also show the different performance of the alternatives adaptive schemes from [11] with separate marking.

The remaining parts of the paper are organized as follows. Section 2 introduces the notation and the technical tools required throughout the error analysis in this paper and summarizes the axioms and the precise optimality statement. The proofs follow in Section 3. The inner mechanism of the least-squares minimization in Subsection 3.2 allows a proof of a discrete reliability, but seemingly restricts the analysis to the lowest-order case but works for any space dimension $n \geq 2$.

Standard notation on Lebesgue and Sobolev spaces $L^2(\Omega)$, $H^k(\Omega)$, $H(\mathrm{div}, \Omega)$ and the corresponding spaces of vector- or matrix-valued functions $L^2(\Omega; \mathbb{R}^n)$, $L^2(\Omega; \mathbb{R}^{n\times n})$, etc., with the L^2 scalar product $(\bullet, \bullet)_{L^2(\Omega)}$ applies throughout this paper. For a regular triangulation $\mathcal{T}$ of Ω, let $H^k(\mathcal{T}) := \prod_{T\in\mathcal{T}} H^k(T) \equiv \{v \in L^2(\Omega) \mid \forall T \in \mathcal{T},\ v|_T \in H^k(T) := H^k(\mathrm{int}(T))\}$ denote the piecewise Sobolev spaces, and let $(\nabla_{\mathrm{NC}} v)|_T = \nabla(v|_T)$ on $T \in \mathcal{T}$ denote the piecewise gradient for $v \in H^1(\mathcal{T})$. Let $|\!|\!| \bullet |\!|\!|_{NC} := |\bullet|_{H^1(\Omega)} = \|\nabla \bullet\|_{L^2(\Omega)}$ and $|\!|\!| \bullet |\!|\!|_{NC} := \|\nabla_{\mathrm{NC}} \bullet\|_{L^2(\Omega)}$ abbreviate the (nonconforming) energy (semi) norm. The L^2 projection onto the piecewise constants $P_0(\mathcal{T})$ reads $\Pi v|_T := ⨍_T v \,\mathrm{d}x := \int_T v \,\mathrm{d}x/|T|$ for any $v \in L^2(\Omega)$ and $T \in \mathcal{T}$ and applies componentwise to vector-valued functions.

The orientation of the unit normal vector ν_E of a side E is fixed to define the jump $[q]_E$ of a piecewise smooth vector field q with the normal component $[q]_E \cdot \nu_E$ and the tangential components $[q]_E \times \nu_E$ (with natural interpretations for $n \neq 3$).

The measure $|\bullet|$ is context sensitive and refers to the number of elements of some finite set or the measure $|E|$ of a side E, etc., and not just the modulus of a real number or the Euclidean length of a vector.

The inequality $A \lesssim B$ abbreviates the relation $A \leq CB$ with a generic positive constant C independent of the underlying triangulation, but solely dependent on the initial triangulation $\mathcal{T}_0$; $A \approx B$ abbreviates $A \lesssim B \lesssim A$.

2. Preliminaries

2.1. **Triangulations and finite element spaces.** The newest vertex bisection (NVB) applies throughout this paper with the set $\mathbb{T}$ of admissible triangulations $\mathcal{T}$ of a bounded polyhedral Lipschitz domain $\Omega \subset \mathbb{R}^n$ into simplices computed by successive admissible refinements of an initial triangulation $\mathcal{T}_0$ with maximal mesh-size h_0 (plus some initialization of tagged simplices as in [18]). All triangulations in this paper are admissible (i.e., successively generated by NVB from $\mathcal{T}_0$) and so regular in the sense of Ciarlet [3–5, 16] (without hanging nodes or edges, etc.), and they are also shape-regular (i.e. locally quasi-uniform).

For a triangulation $\mathcal{T} \in \mathbb{T}$, let $\mathcal{E}$ (resp., $\mathcal{E}(T)$) denote the set of all sides in the triangulation (resp., of a simplex $T \in \mathcal{T}$); side is the notion for a proper subsimplex of n vertices of T and means an edge in $n = 2$ and triangle in $n = 3$ dimension. Let $P_k(\mathcal{T})$ denote the (generally discontinuous) piecewise polynomials of degree at most $k \in \mathbb{N}_0$; $\operatorname{mid}(T)$ (resp., $\operatorname{mid}(E)$) denotes the center of mass of $T \in \mathcal{T}$ (resp., $E \in \mathcal{E}$); $\mathcal{E}(\Omega)$ denotes the set of interior sides, while $\mathcal{E}(\partial\Omega)$ is the remaining set of sides along the boundary in $\mathcal{T}$.

The lowest-order finite element spaces named after Courant, Crouzeix–Raviart, and Raviart-Thomas read

$$
\begin{aligned}
S_0^1(\mathcal{T}) &:= P_1(\mathcal{T}) \cap H_0^1(\Omega),\\
CR_0^1(\mathcal{T}) &:= \{v_{\mathrm{CR}} \in P_1(\mathcal{T}) \,|\, v_{\mathrm{CR}} \text{ continous at} \operatorname{mid}(E) \text{ for all } E \in \mathcal{E}(\Omega)\\
&\qquad\qquad \text{and } v_{\mathrm{CR}}(\operatorname{mid}(E)) = 0 \text{ for all } E \in \mathcal{E}(\partial\Omega)\},\\
RT_0(\mathcal{T}) &:= \{A + b(\bullet - \operatorname{mid}(\mathcal{T})) \in H(\operatorname{div},\Omega) \,|\, A \in P_0(\mathcal{T};\mathbb{R}^n),\, b \in P_0(\mathcal{T})\}.
\end{aligned}
$$

It will be used throughout this paper that the piecewise integration by parts formula $(v_{\mathrm{CR}}, \operatorname{div}\tau_{RT})_{L^2(\Omega)} = -(\tau_{RT}, \nabla_{\mathrm{NC}} v_{\mathrm{CR}})_{L^2(\Omega)}$ holds without jump or boundary contributions for all Crouzeix–Raviart and Raviart–Thomas functions $v_{\mathrm{CR}} \in CR_0^1(\mathcal{T})$ and $\tau_{RT} \in RT_0(\mathcal{T})$ in the same regular triangulation $\mathcal{T}$. (The proof studies the boundary terms along a side F along which the lowest-order Raviart–Thomas functions have a continuous constant normal component (no jump), while the jump of the Crouzeix–Raviart function arises with an integral mean zero on F.)

Each simplex T has an outer unit normal ν_T and each side $E \in \mathcal{E}$ is attached to a unique orientation with $\nu_E = \nu_{T^+} = -\nu_{T^-}$ for an interior side $E = \partial T^+ \cap \partial T^-$ shared by two simplices so that $[w_h]_E := w_h|_{T^+} - w_h|_{T^-}$ defines the jump for any piecewise H^1 function w_h and $\omega_E := \operatorname{int}(T^+ \cup T^-)$. Along the boundary $E \subset \partial\Omega$ (with homogeneous boundary conditions), $[w_h]_E := w_h|_{T^+}$ is the trace w_h and $\omega_E := \operatorname{int}(T^+)$.

2.2. **Discrete inequalities.** Besides the Poincaré–Friedrichs inequality $\|v\|_{L^2(\Omega)} \le C_F \|\!|v|\!\|$ for all $v \in H_0^1(\Omega)$, there is a discrete Friedrichs inequality for $\|v_{\mathrm{CR}}\|_{L^2(\Omega)} \le C_{dF} \|\!|v_{\mathrm{CR}}|\!\|_{NC}$ for all $v_{\mathrm{CR}} \in CR_0^1(\mathcal{T})$ [5, p. 301] with a constant $C_{dF} \approx 1$ estimated in [9]. The constant C_{dP} in the discrete Poincaré inequality,

$$
\|v_{\mathrm{CR}} - v_K\|_{L^2(K)} \le C_{dP} h_K \, \|\nabla_{\mathrm{NC}} v_{\mathrm{CR}}\|_{L^2(K)}, \tag{6}
$$

for a Crouzeix–Raviart finite element function v_{CR} on a fine triangulation of the simplex K and its integral mean $v_K := \fint_K v_{\mathrm{CR}} dx$, is estimated in [9] as $C_{dP} = \sqrt{3/2\cot(\omega_0)}$ with the minimal angle ω_0 in the triangulation for $n = 2$ and $C_{dP} = C_{\mathrm{sr}}\sqrt{5}/3$ with $\operatorname{diam}(K) \le C_{\mathrm{sr}} h_K = C_{\mathrm{sr}} |K|^{1/n}$ and the shape regularity constant $C_{\mathrm{sr}} \approx 1$ for $n = 3$; the continuous version leads to a constant C_{P}. The reader is referred to textbooks [3–5, 16] for other standard estimates from Cauchy to (discrete) trace inequalities.

2.3. **Two level notation.** This subsection provides the necessary notation for a discretization of two levels for further reference throughout this paper. Given an admissible refinement $\widehat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ of a triangulation $\mathcal{T} \in \mathbb{T}$, let $(\sigma_{LS}, u_{LS}) \in X(\mathcal{T})$

(resp., $(\widehat{\sigma_{LS}}, \widehat{u_{LS}}) \in X(\widehat{\mathcal{T}})$) be the minimizer of (1) based on $\mathcal{T}$ (resp., $\widehat{\mathcal{T}}$). Define

$$\delta^2(\mathcal{T},\widehat{\mathcal{T}}) := LS(\widehat{\Pi} f - \Pi f; \widehat{\sigma_{LS}} - \sigma_{LS}, \widehat{u_{LS}} - u_{LS}) \tag{7}$$

and $\delta(\mathcal{T},\widehat{\mathcal{T}}) := \delta^2(\mathcal{T},\widehat{\mathcal{T}})^{1/2}$ with the L^2 projection Π (resp., $\widehat{\Pi}$) onto $P_0(\mathcal{T})$ (resp., $P_0(\widehat{\mathcal{T}})$). The error estimator $\eta(T) := \eta^2(T)^{1/2}$ of (4) is based on $\mathcal{T}$ and the analog definition applies to $\widehat{\mathcal{T}}$ and defines $\widehat{\eta}(\widehat{T})$ for each $\widehat{T} \in \widehat{\mathcal{T}}$.

The subsequent summation rule frequently abbreviates particular sums

$$\eta^2(\mathcal{M}) := \sum_{T\in\mathcal{M}} \eta^2(T) \quad \text{for any subset } \mathcal{M} \subset \mathcal{T} \tag{8}$$

and $\eta(\mathcal{M}) := \eta^2(\mathcal{M})^{1/2}$ (with an analog abbreviation for $\widehat{\eta}^2(\widehat{\mathcal{M}})$ and $\widehat{\mathcal{M}} \subset \widehat{\mathcal{T}}$).

Based on $\mathcal{T}$ and $\widehat{\mathcal{T}}$, one defines the set $\mathcal{R}_0 := \mathcal{T} \setminus \widehat{\mathcal{T}}$ of refined simplices and its supersets $\mathcal{R}_1$ (resp., $\mathcal{R}_2$) as $\mathcal{T} \setminus \widehat{\mathcal{T}}$ plus one (resp., two) layers of neighbouring simplices: $\mathcal{R}_{j+1} := \{T \in \mathcal{T} : \operatorname{dist}(T, \mathcal{R}_j) = 0\}$ for $j = 0, 1$. Those sets of simplices $\mathcal{T} \setminus \widehat{\mathcal{T}} \equiv \mathcal{R}_0 \subset \mathcal{R}_1 \subset \mathcal{R}_2 \subset \mathcal{T}$ cover certain subdomains $\Omega' := \operatorname{int}(\cup(\mathcal{T} \setminus \widehat{\mathcal{T}})) \subset \Omega'' := \operatorname{int}(\cup\mathcal{R}_1)$.

Based on $\mathcal{T}$ and its refinement $\widehat{\mathcal{T}}$, the subtriangulation $\widehat{\mathcal{T}}(K) := \{T \in \widehat{\mathcal{T}} : T \subset K\}$ is a shape-regular triangulation of $K \in \mathcal{T}$ into one or more simplices.

2.4. **Axioms of adaptivity.** The first three axioms concern an admissible refinement $\widehat{\mathcal{T}} \in \mathbb{T}(\mathcal{T})$ of a triangulation $\mathcal{T} \in \mathbb{T}$ and the respective estimators $\widehat{\eta}$ and η as well as the distance $\delta(\mathcal{T},\widehat{\mathcal{T}})$ and the set $\mathcal{R}_2$ from the previous subsection and adopt the sum convention (8). With universal positive constants $\Lambda_1, \dots, \Lambda_5, \widehat{\Lambda}_3$ and $\varrho_2 < 1$, suppose stability (A1), reduction (A2), discrete reliability (A3), and quasi-monotonicity (QM) as follows:

$$|\widehat{\eta}(\mathcal{T}\cap\widehat{\mathcal{T}}) - \eta(\mathcal{T}\cap\widehat{\mathcal{T}})| \le \Lambda_1 \delta(\mathcal{T},\widehat{\mathcal{T}}); \tag{A1}$$

$$\widehat{\eta}(\widehat{\mathcal{T}} \setminus \mathcal{T}) \le \varrho_2 \eta(\mathcal{T}\setminus\widehat{\mathcal{T}}) + \Lambda_2 \delta(\mathcal{T},\widehat{\mathcal{T}}); \tag{A2}$$

$$\delta^2(\mathcal{T},\widehat{\mathcal{T}}) \le \Lambda_3 \eta^2(\mathcal{R}_2) + \widehat{\Lambda}_3 \widehat{\eta}^2(\widehat{\mathcal{T}}); \tag{A3}$$

$$\widehat{\eta}(\widehat{\mathcal{T}}) \le \Lambda_5 \eta(\mathcal{T}). \tag{QM}$$

The quasiorthogonality (A4) concerns the discrete solutions (σ_ℓ, u_ℓ) and $\eta_\ell(\mathcal{T}_\ell)$ from the output of ACLSFEM from the introduction and assumes

$$\sum_{k=\ell}^{\infty} \delta^2(\mathcal{T}_k, \mathcal{T}_{k+1}) \le \Lambda_4 \eta_\ell^2 \quad \text{for all } \ell \in \mathbb{N}_0. \tag{A4}$$

2.5. **Optimal convergence.** Under the assumptions (A1)–(A4) and (QM) from the previous subsection, the techniques from [1, 7, 13, 15, 17] imply the following statement of optimal rates: For any bulk parameter $\Theta < \Theta_0 := 1/(1 + \Lambda_1^2\Lambda_3)$, the output of ACLSFEM satisfies for any $s > 0$ that

$$\sup_{\ell\in\mathbb{N}_0} (1 + |\mathcal{T}_\ell| - |\mathcal{T}_0|)^s \eta_\ell(\mathcal{T}_\ell) \lesssim \sup_{N\in\mathbb{N}_0} (1+N)^s \min\{\eta(\mathcal{T}) \,|\, \mathcal{T} \in \mathbb{T} \text{ with } |\mathcal{T}| - |\mathcal{T}_0| \le N\}. \tag{9}$$

(Here $\eta(\mathcal{T})$ is the error estimator with respect to all simplices in a competing triangulation $\mathcal{T}$ defined by the minimizer (σ_{LS}, u_{LS}) of (1) in $X(\mathcal{T})$ through (4) and the sum convention (8).) The equivalence constants in (9) also depend on Θ and s and prove that the optimal rate (on the right-hand) is equal [7] to the rate for ACLSFEM (on the left-hand side).

The proof of (9) is indeed contained in [7, 13] (for $\widehat{\Lambda}_3 = 0$ in [7] and for the general case in [13] provided the data approximation $\mu \equiv 0$ is neglected for collective marking). The notion of optimal rates in approximation classes involves the total error with the best approximation of the solution (σ, u) in $X(\mathcal{T})$ and the data error $\operatorname{osc}(f, \mathcal{T})$. Because of the equivalence (3), optimality in terms of nonlinear approximation classes (as in [1, 7, 15, 17]) follows from (9).

3. Proofs

Adopt the notation from Subsection 2.3, so $(\sigma_{LS}, u_{LS}) \in X(\mathcal{T})$ (resp., $(\widehat{\sigma_{LS}}, \widehat{u_{LS}}) \in X(\widehat{\mathcal{T}})$) is the minimizer of (1) based on $\mathcal{T}$ (resp., its refinement $\widehat{\mathcal{T}}$).

3.1. **New equivalence.** This subsection establishes (3) of the introduction, where $\operatorname{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}})$ abbreviates the square root of $\sum_{T \in \mathcal{T} \setminus \widehat{\mathcal{T}}} \operatorname{osc}^2(\widehat{\Pi} f, T)$ with the piecewise mean $\widehat{\Pi} f \in P_0(\widehat{\mathcal{T}})$ of $f \in L^2(\Omega)$ with respect to the finer triangulation $\widehat{\mathcal{T}}$.

Lemma 1. *It holds that*

$$\|\widehat{\sigma_{LS}} - \sigma_{LS}\|_{L^2(\Omega)} + \|\!|\widehat{u_{LS}} - u_{LS}|\!\| \lesssim \delta(\mathcal{T}, \widehat{\mathcal{T}}) + \operatorname{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}).$$

Proof. Recall (1), (7), expand binomial formulas, and integrate by parts to verify

$$\begin{aligned}
&\|\widehat{\sigma_{LS}} - \sigma_{LS}\|^2_{L^2(\Omega)} + \|\!|\widehat{u_{LS}} - u_{LS}|\!\|^2 - \delta^2(\mathcal{T}, \widehat{\mathcal{T}}) + \|\widehat{\Pi} f - \Pi f + \operatorname{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|^2_{L^2(\Omega)} \\
&\quad = 2(\widehat{\sigma_{LS}} - \sigma_{LS}, \nabla(\widehat{u_{LS}} - u_{LS}))_{L^2(\Omega)} = -2(\operatorname{div}(\widehat{\sigma_{LS}} - \sigma_{LS}), \widehat{u_{LS}} - u_{LS})_{L^2(\Omega)} \\
&\quad = 2(\widehat{\Pi} f - \Pi f, \widehat{u_{LS}} - u_{LS})_{L^2(\Omega)} - 2(\widehat{\Pi} f - \Pi f + \operatorname{div}(\widehat{\sigma_{LS}} - \sigma_{LS}), \widehat{u_{LS}} - u_{LS})_{L^2(\Omega)}.
\end{aligned}$$

A piecewise Poincaré inequality (6) in the second-to-last term eventually results in

$$(\widehat{\Pi} f - \Pi f, \widehat{u_{LS}} - u_{LS})_{L^2(\Omega)} \le C_{dP} \operatorname{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}) \|\!|\widehat{u_{LS}} - u_{LS}|\!\|. \tag{10}$$

The proof of (10) focuses on one simplex $K \in \mathcal{T} \setminus \widehat{\mathcal{T}}$ and the integral mean f_K of $\widehat{\Pi} f$ on K. The Crouzeix–Raviart function $v_{\mathrm{CR}} := (\widehat{u_{LS}} - u_{LS})|_K \in \mathrm{CR}^1(\widehat{\mathcal{T}}(K))$ with integral mean v_K arises in

$$\begin{aligned}
(\widehat{\Pi} f - \Pi f, \widehat{u_{LS}} - u_{LS})_{L^2(K)} &= (\widehat{\Pi} f - f_K, v_{\mathrm{CR}} - v_K)_{L^2(K)} \\
&\le \widehat{\Pi} f - f_K\|_{L^2(K)} \|v_{\mathrm{CR}} - v_K\|_{L^2(K)}.
\end{aligned}$$

The term $\|v_{\mathrm{CR}} - v_K\|_{L^2(K)}$ is bounded in (6) and leads to the extra mesh-size factor required in the oscillations of (10).

This, Friedrichs's, and Cauchy's inequality prove that the first displayed terms are

$$\le 2 \left(C_{dP} \operatorname{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}) + C_F \|\widehat{\Pi} f - \Pi f + \operatorname{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|_{L^2(\Omega)} \right) \|\!|\widehat{u_{LS}} - u_{LS}|\!\|.$$

This and some rearrangements conclude the proof. $\square$

Theorem 2. *The exact (resp., discrete) solution (σ, u) (resp. (σ_{LS}, u_{LS})) and the right-hand side $f \in L^2(\Omega)$ satisfy* (3).

Proof. The assertion "$\gtrsim$" follows from Lemma 1 for the fixed $\mathcal{T}$ and some sequence of successive uniform refinements to generate a sequence of finer triangulations $\widehat{\mathcal{T}}$ with maximal mesh-size $\max h_{\widehat{\mathcal{T}}} \to 0$. Indeed, the convergence of the least-squares finite element scheme in $H(\operatorname{div}) \times H^1(\Omega)$ as $\max h_{\widehat{\mathcal{T}}} \to 0$ leads to $\delta^2(\mathcal{T}, \widehat{\mathcal{T}}) \to LS(\Pi f; \sigma_{LS}, u_{LS})$. This and Lemma 1 conclude the proof.

The proof of the converse estimate "$\lesssim$" starts with a triangle inequality $\|\sigma_{LS} - \nabla u_{LS}\|_{L^2(\Omega)} \le \|\sigma - \sigma_{LS}\|_{L^2(\Omega)} + \|u - u_{LS}\|$. The remaining estimate of $\Pi f + \operatorname{div} \sigma_{LS} \in P_0(\mathcal{T})$ in L^2 utilizes the inf-sup stability of the divergence operator [3, 4, 16]

$$1 \lesssim \inf_{v_0 \in P_0(\mathcal{T})\setminus\{0\}} \sup_{\tau_{RT} \in RT_0(\mathcal{T})\setminus\{0\}} (v_0, \operatorname{div} \tau_{RT})_{L^2(\Omega)} / (\|v_0\|_{L^2(\Omega)} \|\tau_{RT}\|_{H(\operatorname{div})})$$

with a continuous right inverse of $\operatorname{div} : RT_0(\mathcal{T}) \to P_0(\mathcal{T})$ to select $\tau_{RT} \in RT_0(\mathcal{T})$ with

$$\Pi f + \operatorname{div} \sigma_{LS} = \operatorname{div} \tau_{RT} \quad \text{and} \quad \|\tau_{RT}\|_{L^2(\Omega)} \lesssim \|\Pi f + \operatorname{div} \sigma_{LS}\|_{L^2(\Omega)}.$$

The discrete Euler–Lagrange equations associated to (1) at the discrete minimizer (σ_{LS}, u_{LS}) follow from differentiation and read

$$(\sigma_{LS} - \nabla u_{LS}, \tau_{RT} - \nabla v_C)_{L^2(\Omega)} + (\Pi f + \operatorname{div} \sigma_{LS}, \operatorname{div} \tau_{RT})_{L^2(\Omega)} = 0 \tag{11}$$

for all $(\tau_{RT}, v_C) \in RT_0(\mathcal{T}) \times S^1_0(\mathcal{T})$. The aforementioned choice of τ_{RT} and its stability prove in (11) that

$$\begin{aligned}\|\Pi f + \operatorname{div} \sigma_{LS}\|^2_{L^2(\Omega)} &= -(\sigma_{LS} - \nabla u_{LS}, \tau_{RT})_{L^2(\Omega)} \\ &\lesssim \|\sigma_{LS} - \nabla u_{LS}\|_{L^2(\Omega)} \|\Pi f + \operatorname{div} \sigma_{LS}\|_{L^2(\Omega)}.\end{aligned}$$

This and the aforementioned triangle inequality conclude the proof. □

3.2. **An internal variable.** The internal variable u_{CR} of Lemma 3 describes an algebraic link between the two residuals in the least squares functional. Its analog $\widehat{u_{\mathrm{CR}}}$ on the fine level of $\widehat{\mathcal{T}}$ enables a miraculous local control of some key terms in Lemma 4 below. The definition of the piecewise constant $s(\mathcal{T}) \in P_0(\mathcal{T})$ involves the affine function $\bullet - \operatorname{mid}(T)$, that equals $x - \operatorname{mid}(T)$ at $x \in T \in \mathcal{T}$, and the integral mean $\Pi|\bullet - \operatorname{mid}(T)|^2$ on $T \in \mathcal{T}$ of its squared norm $|\bullet - \operatorname{mid}(T)|^2 \in P_2(T)$,

$$h^2_{\mathcal{T}} \approx s(\mathcal{T}) := n^{-2}\, \Pi|\bullet - \operatorname{mid}(\mathcal{T})|^2 \in P_0(\mathcal{T}). \tag{12}$$

Lemma 3. *The piecewise constant L^2 best approximation $\Pi\sigma_{LS}$ of the Raviart–Thomas solution σ_{LS} is equal to the piecewise gradient $\nabla_{NC} u_{CR} = \Pi\sigma_{LS}$ of a Crouzeix–Raviart function $u_{CR} \in CR^1_0(\mathcal{T})$ with*

$$(1 + s(\mathcal{T})) \operatorname{div} \sigma_{LS} = \Pi(u_{CR} - u_{LS} - f) \quad \text{a.e. in } \Omega. \tag{13}$$

Proof. The L^2 scalar product $(\sigma_{LS}, \tau_{RT})_{L^2(\Omega)}$ of the Raviart–Thomas functions in (11) concerns the piecewise constant contribution $\Pi\tau_{RT}$ and the L^2 orthogonal remainder $(1 - \Pi)\tau_{RT} = n^{-1} \operatorname{div} \tau_{RT} (\bullet - \operatorname{mid}(\mathcal{T}))$; the weight $s(\mathcal{T})$ enters in the L^2 scalar product

$$((1 - \Pi)\sigma_{LS}, (1 - \Pi)\tau_{RT})_{L^2(\Omega)} = (s(\mathcal{T}) \operatorname{div} \sigma_{LS}, \operatorname{div} \tau_{RT})_{L^2(\Omega)}. \tag{14}$$

Consequently, (11) implies

$$(\Pi f + (1 + s(\mathcal{T})) \operatorname{div} \sigma_{LS}, \operatorname{div} \tau_{RT})_{L^2(\Omega)} = (\nabla u_{LS} - \Pi\sigma_{LS}, \tau_{RT})_{L^2(\Omega)}$$

for any $\tau_{RT} \in RT_0(\mathcal{T})$. Equation (11) also shows that $\nabla u_{LS} - \Pi\sigma_{LS}$ is L^2 orthogonal to the divergence-free functions in $RT_0(\mathcal{T})$ as well as to $\nabla S^1_0(\mathcal{T})$. Consequently, $\Pi\sigma_{LS} - \nabla u_{LS}$ is equal to the piecewise gradient of some Crouzeix–Raviart function v_{CR}, and then $u_{\mathrm{CR}} := v_{\mathrm{CR}} + u_{LS}$. This is standard in $n = 2$ space dimensions from a discrete Helmholtz decomposition [11] and follows in general, also for multiply connected domains, as follows: Define v_{CR} as the Riesz representation of the functional $(\Pi\sigma_{LS} - \nabla u_{LS}, \nabla_{\mathrm{NC}}\bullet)_{L^2(\Omega)}$ in the Hilbert space $CR^1_0(\mathcal{T})$ with the scalar product $(\nabla_{\mathrm{NC}}\bullet, \nabla_{\mathrm{NC}}\bullet)_{L^2(\Omega)}$. Then $\varrho_{RT} := \Pi\sigma_{LS} - \nabla_{\mathrm{NC}} u_{\mathrm{CR}} \in P_0(\mathcal{T}; \mathbb{R}^n)$ is perpendicular to the piecewise gradient $\nabla_{\mathrm{NC}}\psi_E$ in L^2 for the side-oriented basis function ψ_E in $CR_0^1(\mathcal{T})$ of an interior side $E \in \mathcal{E}(\Omega)$. A piecewise integration by parts in this orthogonality with $\int_F [\psi_E]_F ds = 0$ for all sides F leads to $\int_E [\varrho_{RT}]_E \cdot \nu_E \,\mathrm{d}s = 0$. Since $[\varrho_{RT}]_E \cdot \nu_E$ is constant along E, it has to vanish. In other words, the normal components of ϱ_{RT} are continuous in Ω, and so $\varrho_{RT} \in RT_0(\mathcal{T})$ is a divergence-free test function. The orthogonality of $\Pi\sigma_{LS} - \nabla u_{LS}$ and $\varrho_{RT} \in RT_0(\mathcal{T})$ on the one hand and $(\nabla_{\mathrm{NC}} v_{\mathrm{CR}}, \varrho_{RT})_{L^2(\Omega)} = 0$ for the divergence-free $\varrho_{RT} \in RT_0(\mathcal{T})$ on the other imply $(\varrho_{RT}, \varrho_{RT})_{L^2(\Omega)} = 0$. Thus $\varrho_{RT} = 0$ implies the claimed representation.

It follows that the previous right-hand side $(\nabla u_{LS} - \Pi\sigma_{LS}, \tau_{RT})_{L^2(\Omega)}$ is equal to $-(\nabla_{\mathrm{NC}} v_{\mathrm{CR}}, \tau_{RT})_{L^2(\Omega)}$. An integration by parts shows this is $(\Pi v_{\mathrm{CR}}, \operatorname{div}\tau_{RT})_{L^2(\Omega)}$ and so proves

$$(\Pi f + (1 + s(\mathcal{T}))\operatorname{div}\sigma_{LS} - \Pi v_{\mathrm{CR}}, \operatorname{div}\tau_{RT})_{L^2(\Omega)} = 0 \quad \text{for all } \tau_{RT} \in RT_0(\mathcal{T}).$$

Since the divergence operator maps $RT_0(\mathcal{T})$ onto $P_0(\mathcal{T})$, this concludes the proof. $\square$

This subsection concludes with an application of Lemma 3 to the last term in

$$\delta^2(\mathcal{T}, \widehat{\mathcal{T}}) = \|\widehat{\sigma_{LS}} - \sigma_{LS} - \nabla(\widehat{u_{LS}} - u_{LS})\|^2_{L^2(\Omega)} + \|\Pi \operatorname{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|^2_{L^2(\Omega)}$$
$$+ \|(\widehat{\Pi} - \Pi)(f + \operatorname{div}\widehat{\sigma_{LS}})\|^2_{L^2(\Omega)}. \tag{15}$$

The first two terms on the right-hand side in (15) will be rewritten in Lemma 5 of the subsequent subsection, while the last term is controlled with (13) even locally. This serves as a novel and elementary estimate of some key oscillation $(\widehat{\Pi} - \Pi)f$ (with f shifted to $f + \operatorname{div}\widehat{\sigma_{LS}}$) *without* the mesh-size factor $h_{\mathcal{T}}$ by some first-order residual contribution $\|\widehat{\sigma_{LS}} - \nabla\widehat{u_{LS}}\|$ *with* an additional mesh-size factor $h_{\mathcal{T}}$.

Lemma 4. *For any $T \in \mathcal{T}$,*

$$\|(\widehat{\Pi} - \Pi)(f + \operatorname{div}\widehat{\sigma_{LS}})\|_{L^2(T)} \lesssim h_T \|\widehat{\sigma_{LS}} - \nabla\widehat{u_{LS}}\|_{L^2(T)}. \tag{16}$$

Proof. The substitution of $\mathcal{T}$ by $\widehat{\mathcal{T}}$ in Lemma 3 and $\widehat{v_{\mathrm{CR}}} := \widehat{u_{\mathrm{CR}}} - \widehat{u_{LS}}$ prove that

$$\widehat{\Pi} f + \operatorname{div}\widehat{\sigma_{LS}} = \widehat{\Pi}\widehat{v_{\mathrm{CR}}} - s(\widehat{\mathcal{T}})\operatorname{div}\widehat{\sigma_{LS}}.$$

The two arguments for the estimate of

$$\|(\widehat{\Pi} - \Pi)(f + \operatorname{div}\widehat{\sigma_{LS}})\|^2_{L^2(T)} = ((\widehat{\Pi} - \Pi)(f + \operatorname{div}\widehat{\sigma_{LS}}), \widehat{v_{\mathrm{CR}}} - s(\widehat{\mathcal{T}})\operatorname{div}\widehat{\sigma_{LS}})_{L^2(T)}$$

are very different. The discrete Poincaré inequality allows for

$$\begin{aligned}&((\widehat{\Pi} - \Pi)(f + \operatorname{div}\widehat{\sigma_{LS}}), \widehat{v_{\mathrm{CR}}})_{L^2(T)}\\ &\quad \le C_{dP} h_T \|\nabla_{\mathrm{NC}}\widehat{v_{\mathrm{CR}}}\|_{L^2(T)} \|(\widehat{\Pi} - \Pi)(f + \operatorname{div}\widehat{\sigma_{LS}})\|_{L^2(T)}.\end{aligned}$$

This, a Cauchy inequality, the definition of (12) in the second term, namely

$$\|s(\widehat{\mathcal{T}})\operatorname{div}\widehat{\sigma_{LS}}\|_{L^2(T)} \le C_{sr} h_T/(n+1) \|\sqrt{s(\widehat{\mathcal{T}})}\operatorname{div}\widehat{\sigma_{LS}}\|_{L^2(T)},$$

and a division by $\|(\widehat{\Pi} - \Pi)(f + \operatorname{div}\widehat{\sigma_{LS}})\|_{L^2(T)}$ show that

$$\|(\widehat{\Pi} - \Pi)(f + \operatorname{div}\widehat{\sigma_{LS}})\|_{L^2(T)} \lesssim h_T \left(\|\nabla_{\mathrm{NC}}\widehat{v_{\mathrm{CR}}}\|_{L^2(T)} + \|\sqrt{s(\widehat{\mathcal{T}})}\operatorname{div}\widehat{\sigma_{LS}}\|_{L^2(T)} \right).$$

The L^2 orthogonal split $\widehat{\sigma_{LS}} - \nabla\widehat{u_{LS}} = \widehat{\Pi}\widehat{\sigma_{LS}} - \nabla\widehat{u_{LS}} + (1 - \widehat{\Pi})\widehat{\sigma_{LS}}$ into piecewise averages and $\widehat{\Pi}\widehat{\sigma_{LS}} - \nabla\widehat{u_{LS}} = \nabla_{\mathrm{NC}}\widehat{v_{\mathrm{CR}}}$ from Lemma 3 prove that

$$\widehat{\sigma_{LS}} - \nabla\widehat{u_{LS}} = \nabla_{\mathrm{NC}}\widehat{v_{\mathrm{CR}}} + (1 - \widehat{\Pi})\widehat{\sigma_{LS}} \quad \text{a.e. in } T.$$

This and the computation in (14) (with respect to the fine level of $\widehat{\mathcal{T}}$ and with σ_{LS} and τ_{RT} replaced by $\widehat{\sigma_{LS}}$) imply for each $\widehat{T} \in \widehat{\mathcal{T}}(T)$ (i.e., $\widehat{T} \in \widehat{\mathcal{T}}$ with $\widehat{T} \subset T$) that

$$\|\nabla_{\mathrm{NC}} \widehat{v_{\mathrm{CR}}}\|^2_{L^2(\widehat{T})} + \|\sqrt{s(\widehat{\mathcal{T}})}\,\mathrm{div}\,\widehat{\sigma_{LS}}\|^2_{L^2(\widehat{T})} = \|\widehat{\sigma_{LS}} - \nabla \widehat{u_{LS}}\|^2_{L^2(\widehat{T})}.$$

The substitution of this in the second-to-last displayed estimate concludes the proof. $\square$

3.3. **Three auxiliary stresses.** The a posteriori error analysis of mixed and LS-FEM frequently utilizes three mixed finite element solutions $\widehat{\tau_{LS}}, \tau^*_{LS} \in RT_0(\widehat{\mathcal{T}})$, and $\tau_{LS} \in RT_0(\mathcal{T})$ defined as mixed finite element solutions to the Poisson model problem with respective right-hand sides

$$\mathrm{div}\,\widehat{\tau_{LS}} = \mathrm{div}(\widehat{\sigma_{LS}} - \sigma_{LS}) \quad \text{and} \quad \mathrm{div}\,\tau^*_{LS} = \Pi\,\mathrm{div}(\widehat{\sigma_{LS}} - \sigma_{LS}) = \mathrm{div}\,\tau_{LS}.$$

This allows for a divergence-free $\widehat{\varrho_{RT}} := \widehat{\sigma_{LS}} - \sigma_{LS} - \widehat{\tau_{LS}} + \tau^*_{LS} - \tau_{LS} \in RT_0(\widehat{\mathcal{T}})$. The role of the three auxiliary stress functions is visible in the subsequent identity.

Lemma 5. *Any $v_C \in S^1_0(\mathcal{T})$ satisfies*

$$\begin{aligned}
&\|\widehat{\sigma_{LS}} - \sigma_{LS} - \nabla(\widehat{u_{LS}} - u_{LS})\|^2_{L^2(\Omega)} + \|\Pi\ \mathrm{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|^2_{L^2(\Omega)}\\
&\quad = (\sigma_{LS} - \nabla u_{LS}, \nabla(\widehat{u_{LS}} - u_{LS} - v_C) - \widehat{\varrho_{RT}})_{L^2(\Omega)}\\
&\qquad + (\widehat{\sigma_{LS}} - \sigma_{LS} - \nabla(\widehat{u_{LS}} - u_{LS}), \widehat{\tau_{LS}} - \tau^*_{LS})_{L^2(\Omega)}.
\end{aligned}$$

Proof. Appropriate test functions lead in (11) (and in its nondisplayed analog with respect to $\widehat{\mathcal{T}}$) to the lemma with elementary algebra; more details can be found (in different notation) in [11, Lemma 5.2]. $\square$

The auxiliary stresses are controlled as follows.

Lemma 6. *It holds that*

$$\begin{aligned}
&\|\widehat{\tau_{LS}} - \tau^*_{LS}\|_{L^2(\Omega)} \lesssim \mathrm{osc}(\mathrm{div}\,\widehat{\sigma_{LS}}, \mathcal{T} \setminus \widehat{\mathcal{T}}),\\
&\qquad \|\widehat{\varrho_{RT}}\|_{L^2(\Omega)} \lesssim \delta(\mathcal{T}, \widehat{\mathcal{T}}) + \mathrm{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}) + \|\Pi\,\mathrm{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|_{L^2(\Omega)}.
\end{aligned}$$

Proof. The first inequality is from [11, Lemma 5.3] in 2D and we give a more general proof for completeness. Let $\widehat{w_{\mathrm{CR}}} \in \mathrm{CR}^1_0(\widehat{\mathcal{T}})$ solve

$$(\nabla_{\mathrm{NC}} \widehat{w_{\mathrm{CR}}}, \nabla_{\mathrm{NC}} \widehat{\psi_{\mathrm{CR}}})_{L^2(\Omega)} = (g, \widehat{\psi_{\mathrm{CR}}})_{L^2(\Omega)} \quad \text{for all } \widehat{\psi_{\mathrm{CR}}} \in \mathrm{CR}^1_0(\widehat{\mathcal{T}})$$

for $g := -\,\mathrm{div}(\widehat{\tau_{LS}} - \tau^*_{LS}) \in P_0(\widehat{\mathcal{T}})$, which is L^2 orthogonal to $P_0(\mathcal{T})$. An integration by parts and the mixed finite element equations in the definition of $\widehat{\tau_{LS}}$ and τ^*_{LS} show that the piecewise constant vector field

$$\widehat{\Pi}(\widehat{\tau_{LS}} - \tau^*_{LS}) - \nabla_{\mathrm{NC}} \widehat{w_{\mathrm{CR}}} \perp \nabla_{\mathrm{NC}} \mathrm{CR}^1_0(\widehat{\mathcal{T}}) + \{\widehat{\tau_{RT}} \in RT_0(\widehat{\mathcal{T}}) : \mathrm{div}\,\widehat{\tau_{RT}} = 0\}$$

(with $\perp$ denoting L^2 orthogonality). Since this vector field belongs to $RT_0(\widehat{\mathcal{T}})$ (from an integration by parts) and is divergence-free, the orthogonality shows that it has to vanish, i.e., $\widehat{\Pi}(\widehat{\tau_{LS}} - \tau^*_{LS}) = \nabla_{\mathrm{NC}} \widehat{w_{\mathrm{CR}}}$. This and a standard argument for the aforementioned nonconforming solution to the Poisson equation with the right-hand side $g \perp P_0(\mathcal{T})$ show (with a piecewise discrete Poincaré inequality in the final step) that

$$\begin{aligned}
\|\nabla_{\mathrm{NC}} \widehat{w_{\mathrm{CR}}}\|^2_{L^2(\Omega)} &= (g, \widehat{w_{\mathrm{CR}}})_{L^2(\Omega)}\\
&= (g, \widehat{w_{\mathrm{CR}}} - \Pi \widehat{w_{\mathrm{CR}}})_{L^2(\Omega)} \lesssim \mathrm{osc}(g, \mathcal{T}) \|\nabla_{\mathrm{NC}} \widehat{w_{\mathrm{CR}}}\|_{L^2(\Omega)}.
\end{aligned}$$

This concludes the proof of the first asserted inequality. The proof of the second starts with a triangle inequality followed by the L^2 orthogonality of $\widehat{\tau_{LS}}$ (resp., τ^*_{LS}) and the divergence-free function $\widehat{\sigma_{LS}} - \sigma_{LS} - \widehat{\tau_{LS}}$ (resp., $\tau^*_{LS} - \tau_{LS}$):

$$\begin{aligned}1/2\,\|\widehat{\varrho_{RT}}\|^2_{L^2(\Omega)} &\le \|\widehat{\sigma_{LS}} - \sigma_{LS} - \widehat{\tau_{LS}}\|^2_{L^2(\Omega)} + \|\tau^*_{LS} - \tau_{LS}\|^2_{L^2(\Omega)}\\ &= \|\widehat{\sigma_{LS}} - \sigma_{LS}\|^2_{L^2(\Omega)} - \|\widehat{\tau_{LS}}\|^2_{L^2(\Omega)} + \|\tau_{LS}\|^2_{L^2(\Omega)} - \|\tau^*_{LS}\|^2_{L^2(\Omega)}.\end{aligned}$$

The stability of the auxiliary mixed finite element solutions results in

$$\|\widehat{\varrho_{RT}}\|_{L^2(\Omega)} \lesssim \|\widehat{\sigma_{LS}} - \sigma_{LS}\|_{L^2(\Omega)} + \|\Pi\,\mathrm{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|_{L^2(\Omega)}.$$

This and Lemma 1 conclude the proof. □

3.4. **A posteriori error analysis.** This subsection collects some core arguments for the proof of the discrete reliability. Recall (4) from the introduction and the definitions of $\mathcal{R}_0$, $\mathcal{R}_1$, $\mathcal{R}_2$ and $\Omega' := \mathrm{int}(\cup(\mathcal{T}\setminus\widehat{\mathcal{T}}))$, etc. from Subsection 2.3.

Proposition 7. *There exists a universal constant $\Lambda > 0$ that depends exclusively on $\mathcal{T}_0$ such that*

$$\begin{aligned}&\|\widehat{\sigma_{LS}} - \sigma_{LS} - \nabla(\widehat{u_{LS}} - u_{LS})\|^2_{L^2(\Omega)} + \|\Pi\,\mathrm{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|^2_{L^2(\Omega)}\\ &\quad\le \frac{1}{2}\delta^2(\mathcal{T},\widehat{\mathcal{T}}) + \Lambda\left(\eta^2(\mathcal{R}_2) + \mathrm{osc}^2(\mathrm{div}\,\widehat{\sigma_{LS}}, \mathcal{T}\setminus\widehat{\mathcal{T}})\right).\end{aligned}$$

Proof. A Cauchy inequality in the last term of the identity in Lemma 5 allows us to absorb one factor, and, with Lemma 6 in the final step, this leads to

$$\begin{aligned}&\|\widehat{\sigma_{LS}} - \sigma_{LS} - \nabla(\widehat{u_{LS}} - u_{LS})\|^2_{L^2(\Omega)} + \|\Pi\,\mathrm{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|^2_{L^2(\Omega)}\\ &\lesssim (\sigma_{LS} - \nabla u_{LS}, \nabla(\widehat{u_{LS}} - u_{LS} - v_C) - \widehat{\varrho_{RT}})_{L^2(\Omega)} + \|\widehat{\tau_{LS}} - \tau^*_{LS}\|^2_{L^2(\Omega)}\\ &\lesssim (\sigma_{LS} - \nabla u_{LS}, \nabla(\widehat{u_{LS}} - u_{LS} - v_C) - \widehat{\varrho_{RT}})_{L^2(\Omega)} + \mathrm{osc}^2(\mathrm{div}\,\widehat{\sigma_{LS}}, \mathcal{T}\setminus\widehat{\mathcal{T}}).\end{aligned}$$

The two contributions in the second-to-last term require different arguments, which can be collected from the literature and are merely outlined here. First, the appropriate choice of $v_C \in S^1_0(\mathcal{T})$ as a Scott–Zhang quasi-interpolation of $\widehat{u_{LS}} - u_{LS}$ in the second-to-last term leads to $\widehat{u_{LS}} - u_{LS} = v_C$ on all simplices in $\mathcal{T}\cap\widehat{\mathcal{T}}$ outside Ω'. The approximation and stability properties of $\widehat{u_{LS}} - u_{LS} - v_C$ allow standard arguments after a piecewise integration by parts (as in [11] and as for many a posteriori error estimation with conforming finite element functions) to deduce that

$$\begin{aligned}&(\sigma_{LS} - \nabla u_{LS}, \nabla(\widehat{u_{LS}} - u_{LS} - v_C))_{L^2(\Omega)} \lesssim \|\!|\widehat{u_{LS}} - u_{LS}|\!\|\\ &\times\left(\|h_{\mathcal{T}}\,\mathrm{div}\,\sigma_{LS}\|^2_{L^2(\Omega')} + \sum_{E\in\mathcal{E}(\Omega')} h_E\|[\sigma_{LS} - \nabla u_{LS}]_E\cdot\nu_E\|^2_{L^2(E)}\right)^{1/2}\end{aligned}$$

with a reduced domain Ω' and the set $\mathcal{E}(\Omega')$ of sides interior in Ω'.

Second, the divergence-free Raviart–Thomas function $\widehat{\varrho_{RT}}$ is piecewise constant with respect to the fine triangulation, and so the other contribution reads

$$\begin{aligned}&(\nabla u_{LS} - \widehat{\Pi}\sigma_{LS}, \widehat{\varrho_{RT}})_{L^2(\Omega)}\\ &\quad= (\Pi\sigma_{LS} - \widehat{\Pi}\sigma_{LS}, \widehat{\varrho_{RT}})_{L^2(\Omega)} + (\nabla u_{LS} - \Pi\sigma_{LS}, \widehat{\varrho_{RT}})_{L^2(\Omega)}\\ &\quad\le \|\sigma_{LS} - \Pi\sigma_{LS}\|_{L^2(\Omega')}\|\widehat{\varrho_{RT}}\|_{L^2(\Omega)} - (\nabla_{\mathrm{NC}} v_{\mathrm{CR}}, \widehat{\varrho_{RT}})_{L^2(\Omega)}\end{aligned}$$

with $v_{\mathrm{CR}} := u_{\mathrm{CR}} - u_{LS}$ and u_{CR} from Lemma 3. The ansatz of the lowest-order Raviart–Thomas functions shows $(\sigma_{LS} - \Pi\sigma_{LS})(x) = (n^{-1}\,\mathrm{div}\,\sigma_{LS})(x - \mathrm{mid}(T))$ for

$x \in T \in \mathcal{T}$ and so implies $\|\sigma_{LS} - \Pi\sigma_{LS}\|_{L^2(\Omega')} \lesssim \|h_\mathcal{T} \operatorname{div} \sigma_{LS}\|_{L^2(\Omega')}$. The remaining contribution $(\nabla_{\mathrm{NC}} v_{\mathrm{CR}}, \widehat{\varrho_{RT}})_{L^2(\Omega)}$ is standard in the context of an a posteriori nonconforming finite element error analysis. The point is that any Crouzeix–Raviart function $\widehat{v_{\mathrm{CR}}}$ on the fine triangulation leads to $(\nabla_{\mathrm{NC}} \widehat{v_{\mathrm{CR}}}, \widehat{\varrho_{RT}})_{L^2(\Omega)} = 0$. The choice of $\widehat{v_{\mathrm{CR}}}$ from [8, Theorem 3.2] leads to

$$\|\nabla_{\mathrm{NC}}(\widehat{v_{\mathrm{CR}}} - v_{\mathrm{CR}})\|^2_{L^2(\Omega)} \lesssim \sum_{E \in \overline{\mathcal{E}}(\Omega'')} h_E \|[\nabla_{\mathrm{NC}} v_{\mathrm{CR}}]_E \times \nu_E\|^2_{L^2(E)}$$

with the domain $\Omega'' := \operatorname{int}(\cup \mathcal{R}_1)$ and the set $\mathcal{E}(\Omega'')$ of interior sides in Ω'' and its superset $\overline{\mathcal{E}}(\Omega'') := \mathcal{E}(\Omega'') \cup (\mathcal{E}(\partial\Omega) \cap \mathcal{E}(\overline{\Omega''}))$ with extra sides on $\partial\Omega \cap \partial\Omega''$. This proves

$$\begin{aligned}
&(\nabla u_{LS} - \sigma_{LS}, \widehat{\varrho_{RT}})_{L^2(\Omega)} \lesssim \|\widehat{\varrho_{RT}}\|_{L^2(\Omega)} \\
&\quad \times \left(\|h_\mathcal{T} \operatorname{div} \sigma_{LS}\|^2_{L^2(\Omega')} + \sum_{E \in \overline{\mathcal{E}}(\Omega'')} h_E \|[\nabla_{\mathrm{NC}} v_{\mathrm{CR}}]_E \times \nu_E\|^2_{L^2(E)} \right)^{1/2}.
\end{aligned}$$

Since Lemma 6 controls $\|\widehat{\varrho_{RT}}\|_{L^2(\Omega)}$ and Lemma 1 controls $\|\!|\widehat{u_{LS}} - u_{LS}|\!\|$, the left-hand side of the assertion is $\lesssim \operatorname{osc}^2(\operatorname{div} \widehat{\sigma_{LS}}, \mathcal{T} \setminus \widehat{\mathcal{T}})$ plus

$$\Big(\delta(\mathcal{T}, \widehat{\mathcal{T}}) + \operatorname{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}) + \|\Pi \operatorname{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|_{L^2(\Omega)}\Big) \times \text{ estimators.}$$

Since the square of $\|\Pi \operatorname{div}(\widehat{\sigma_{LS}} - \sigma_{LS})\|_{L^2(\Omega)}$ appears in the controlled left-hand side, it can be absorbed; notice that $\operatorname{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}) \le \operatorname{osc}(f, \mathcal{T} \setminus \widehat{\mathcal{T}})$. Hence the remaining proof focusses on the estimators, i.e., on

$$\begin{aligned}
\text{estimators}^2 := {}& \|h_\mathcal{T} \operatorname{div} \sigma_{LS}\|^2_{L^2(\Omega')} + \sum_{E \in \mathcal{E}(\Omega')} h_E \|[\sigma_{LS} - \nabla u_{LS}]_E \cdot \nu_E\|^2_{L^2(E)} \\
&+ \sum_{E \in \overline{\mathcal{E}}(\Omega'')} h_E \|[\Pi\sigma_{LS} - \nabla u_{LS}]_E \times \nu_E\|^2_{L^2(E)}.
\end{aligned}$$

Since $(\sigma_{LS} - \Pi\sigma_{LS})(x) = (n^{-1} \operatorname{div} \sigma_{LS})(x - \operatorname{mid}(T))$ for $x \in T \in \mathcal{T}$, triangle and trace inequalities show that

$$\begin{aligned}
&\|[\Pi\sigma_{LS} - \nabla u_{LS}]_E \times \nu_E\|_{L^2(E)} - \|[\sigma_{LS} - \nabla u_{LS}]_E \times \nu_E\|_{L^2(E)} \\
&\le \|[\sigma_{LS} - \Pi\sigma_{LS}]_E \times \nu_E\|_{L^2(E)} \lesssim h_E^{1/2} \|\operatorname{div} \sigma_{LS}\|_{L^2(\omega_E)}
\end{aligned}$$

for the side patch ω_E of the at most two simplices T with the side $E \in \mathcal{E}(T)$. The last contribution leads to $\|h_\mathcal{T} \operatorname{div} \sigma_{LS}\|^2_{L^2(\omega_E)}$ for all $E \in \overline{\mathcal{E}}(\Omega'')$ and eventually to an enlargement of the set from $\mathcal{R}_1$ to $\mathcal{R}_2$. In conclusion, it follows that estimators $\lesssim \eta(\mathcal{R}_2)$ in terms of the error estimator in (4). ☐

3.5. **Reliability.** The first corollary from the previous analysis is reliability.

Theorem 8. *The exact* (σ, u) *and the discrete solution* (σ_{LS}, u_{LS}) *and the error estimator* (4) *satisfy* (5).

Proof. Given the fixed $\mathcal{T} \in \mathbb{T}$, consider the sequence of its successive uniform refinements and let $\widehat{\mathcal{T}}$ be one member of this sequence with maximal mesh-size

$\max h_{\widehat{\mathcal{T}}}$. Passing to the limit with $\max h_{\widehat{\mathcal{T}}} \to 0$, the convergence of the least-squares finite element scheme in $H(\operatorname{div}) \times H^1(\Omega)$ allows for the substitution of $(\widehat{\sigma_{LS}}, \widehat{u_{LS}})$ by (σ, u) in Proposition 7. Consequently,

$$\begin{aligned}&\|\sigma_{LS} - \nabla u_{LS}\|^2_{L^2(\Omega)} + \|\Pi f + \operatorname{div}\sigma_{LS}\|^2_{L^2(\Omega)}\\ &\quad\le \frac{1}{2} LS(\Pi f; \sigma_{LS}, u_{LS}) + \Lambda\left(\eta^2(\mathcal{T}) + \operatorname{osc}^2(f,\mathcal{T})\right).\end{aligned}$$

Since the left-hand side is $LS(\Pi f; \sigma_{LS}, u_{LS})$ and since $\operatorname{osc}^2(f,\mathcal{T}) \le \eta^2(\mathcal{T})$,

$$LS(\Pi f; \sigma_{LS}, u_{LS}) \le 4\Lambda\eta^2(\mathcal{T}).$$

This and Theorem 2 conclude the proof of the reliability. The efficiency holds even in a local form for (a) $|T|^{1/n}\|\operatorname{div}\sigma_{LS}\|_{L^2(T)} \approx \|\sigma_{LS} - \Pi\sigma_{LS}\|_{L^2(T)} \le \|\sigma_{LS} - \nabla u_{LS}\|_{L^2(T)}$ and for (b) the jump terms can be controlled by triangle and trace inequalities. □

3.6. **Discrete reliability.** The previous subsections allow the proof of (A3).

Theorem 9. *The discrete solution* (σ_{LS}, u_{LS}) *(resp.,* $(\widehat{\sigma_{LS}}, \widehat{u_{LS}})$*) on the triangulation* $\mathcal{T} \in \mathbb{T}$ *(resp., its admissible refinement* $\widehat{\mathcal{T}}$*) and the error estimator* (4) *satisfy* (A3) *with a universal constant* Λ_3 *and* $\widehat{\Lambda}_3 := h^2\,\Lambda_3$ *for* $h := \max_{T\in\mathcal{T}} h_T$.

Proof. The combination of Lemma 4 and Proposition 7 in the split (15) shows

$$\delta^2(\mathcal{T},\widehat{\mathcal{T}}) \lesssim \eta^2(\mathcal{R}_2) + \operatorname{osc}^2(\operatorname{div}\widehat{\sigma_{LS}}, \mathcal{T}\setminus\widehat{\mathcal{T}}) + \|h_{\mathcal{T}}(\widehat{\sigma_{LS}} - \nabla\widehat{u_{LS}})\|^2_{L^2(\Omega)}.$$

Since $\operatorname{osc}(\widehat{\Pi} f, \mathcal{T}\setminus\widehat{\mathcal{T}}) \le \eta(\mathcal{R}_2)$, the triangle inequality shows

$$\operatorname{osc}(\operatorname{div}\widehat{\sigma_{LS}}, \mathcal{T}\setminus\widehat{\mathcal{T}}) \le \operatorname{osc}(\widehat{\Pi} f + \operatorname{div}\widehat{\sigma_{LS}}, \mathcal{T}\setminus\widehat{\mathcal{T}}) + \eta(\mathcal{R}_2).$$

Notice the extra factor $h_{\mathcal{T}}$ in the oscillation term so that the previous estimate reads

$$\delta^2(\mathcal{T},\widehat{\mathcal{T}}) \lesssim \eta^2(\mathcal{R}_2) + h^2 LS(\widehat{\Pi} f; \widehat{\sigma_{LS}}, \widehat{u_{LS}}).$$

Theorem 8 applies to the fine triangulation as well, and so $h^2\widehat{\eta}^2(\widehat{\mathcal{T}})$ controls the last term to conclude the proof. □

3.7. **Quasiorthogonality.** The main novel argument in the proof is an inequality on the least-squares functional. Straightforward algebra with the least-squares functionals and (11) lead to the (well-) known identity [11, Eq (4.5)]

$$LS(0; \widehat{\sigma_{LS}} - \sigma_{LS}, \widehat{u_{LS}} - u_{LS}) + LS(\widehat{\Pi} f; \widehat{\sigma_{LS}}, \widehat{u_{LS}}) = LS(\widehat{\Pi} f; \sigma_{LS}, u_{LS}). \tag{17}$$

The point is that the data approximation $\widehat{\Pi} f$ is the same on both sides of the equality and so leads to the 0 in the leading term of (17). This paper suggests a modification.

Theorem 10. *It holds that*

$$\begin{aligned}&\delta^2(\mathcal{T},\widehat{\mathcal{T}}) + LS(\widehat{\Pi} f; \widehat{\sigma_{LS}}, \widehat{u_{LS}}) - LS(\Pi f; \sigma_{LS}, u_{LS})\\ &\qquad\lesssim LS^{1/2}(\widehat{\Pi} f; \widehat{\sigma_{LS}}, \widehat{u_{LS}})\operatorname{osc}(\widehat{\Pi} f, \mathcal{T}\setminus\widehat{\mathcal{T}}).\end{aligned}$$

Proof. An elementary modification of (17) shows that

$$\begin{aligned}&LS(\widehat{\Pi} f - \Pi f; \widehat{\sigma_{LS}} - \sigma_{LS}, \widehat{u_{LS}} - u_{LS}) + LS(\widehat{\Pi} f; \widehat{\sigma_{LS}}, \widehat{u_{LS}})\\ &\quad= LS(\Pi f; \sigma_{LS}, u_{LS}) + 2(\widehat{\Pi} f - \Pi f, \widehat{\Pi} f + \operatorname{div}\widehat{\sigma_{LS}})_{L^2(\Omega)}.\end{aligned}\tag{18}$$

The identity (13) applies to the finer level as well and allows the substitution of $\widehat{\Pi} f + \operatorname{div} \widehat{\sigma_{LS}} = \widehat{\Pi}\widehat{v_{\mathrm{CR}}} - s(\widehat{\mathcal{T}}) \operatorname{div} \widehat{\sigma_{LS}}$ in the last term of (18):

$$(\widehat{\Pi} f - \Pi f, \widehat{\Pi} f + \operatorname{div} \widehat{\sigma_{LS}})_{L^2(\Omega)} = (\widehat{\Pi} f - \Pi f, \widehat{v_{\mathrm{CR}}} - s(\widehat{\mathcal{T}}) \operatorname{div} \widehat{\sigma_{LS}})_{L^2(\Omega)}.$$

The first contribution in the last term allows a discrete Poincaré inequality as in (10) and leads to

$$(\widehat{\Pi} f - \Pi f, \widehat{v_{\mathrm{CR}}} - \Pi \widehat{v_{\mathrm{CR}}})_{L^2(\Omega)} \leq C_{dP} \operatorname{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}) \| \widehat{v_{\mathrm{CR}}} \|_{NC}.$$

The second contribution relies on $\sqrt{s(\widehat{\mathcal{T}})} \leq \sqrt{s(\mathcal{T})} \leq C_{\mathrm{sr}} h_{\mathcal{T}}$ a.e. and reads

$$-(\widehat{\Pi} f - \Pi f, s(\widehat{\mathcal{T}}) \operatorname{div} \widehat{\sigma_{LS}})_{L^2(\Omega)} \leq C_{\mathrm{sr}} \operatorname{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}) \, \| \sqrt{s(\widehat{\mathcal{T}})} \operatorname{div} \widehat{\sigma_{LS}} \|_{L^2(\Omega')}.$$

The sum of the two contributions utilizes $\nabla_{\mathrm{NC}} \widehat{v_{\mathrm{CR}}} = \widehat{\Pi} \widehat{\sigma_{LS}} - \nabla \widehat{u_{LS}}$ and

$$\| \sqrt{s(\widehat{\mathcal{T}})} \operatorname{div} \widehat{\sigma_{LS}} \|^2_{L^2(\Omega')} + \| \widehat{v_{\mathrm{CR}}} \|^2_{NC} = \| \widehat{\sigma_{LS}} - \nabla \widehat{u_{LS}} \|^2_{L^2(\Omega)}.$$

Consequently,

$$(\widehat{\Pi} f - \Pi f, \widehat{\Pi} f + \operatorname{div} \widehat{\sigma_{LS}})_{L^2(\Omega)} \lesssim \operatorname{osc}(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}) \| \widehat{\sigma_{LS}} - \nabla \widehat{u_{LS}} \|_{L^2(\Omega)}.$$

The substitution of this in (18) concludes the proof. □

3.8. **Quasi-monotonicity.** For small $\widehat{\Lambda}_3$ (e.g., for a sufficiently fine initial triangulation) (QM) holds [13]. In the general situation, Theorem 10 implies that

$$LS(\widehat{\Pi} f; \widehat{\sigma_{LS}}, \widehat{u_{LS}}) \lesssim LS(\Pi f; \sigma_{LS}, u_{LS}) + \operatorname{osc}^2(\widehat{\Pi} f, \mathcal{T} \setminus \widehat{\mathcal{T}}).$$

This and the equivalence (5) with respect to the coarse triangulation $\mathcal{T}$ and to the fine triangulation $\widehat{\mathcal{T}}$ imply (QM). □

3.9. **Finish of the proof.** The axioms (A3) and (QM) are proven explicitly in this paper. The axioms (A1) and (A2) follow with standard arguments like triangle, (discrete) trace, and Cauchy inequalities as in [13, 15]. The axiom (A4) requires a generalisation (A4$_\varepsilon$): Theorem 10 leads to a universal constant $C_1 > 0$ such that, for any $\varepsilon > 0$,

$$\begin{aligned} \delta^2(\mathcal{T}_k, \mathcal{T}_{k+1}) + (1-\varepsilon) LS(\Pi_{k+1} f; \sigma_{k+1}, u_{k+1}) \\ \leq LS(\Pi_k f; \sigma_k, u_k) + C_1/\varepsilon \operatorname{osc}^2(\Pi_{k+1} f, \mathcal{T}_k). \end{aligned}$$

Given any $\ell, m \in \mathbb{N}_0$, the sum over all $k = \ell, \ell+1, \ldots, \ell+m$ leads to

$$\begin{aligned} \sum_{k=\ell}^{\ell+m} \delta^2(\mathcal{T}_k, \mathcal{T}_{k+1}) \leq LS(\Pi_\ell f; \sigma_\ell, u_\ell) + \varepsilon \sum_{k=\ell}^{\ell+m} LS(\Pi_{k+1} f; \sigma_{k+1}, u_{k+1}) \\ + C_1/\varepsilon \sum_{k=\ell}^{\ell+m} \operatorname{osc}^2(\Pi_{k+1} f, \mathcal{T}_k). \end{aligned}$$

Up to the mesh-size weight factors, the oscillations $\operatorname{osc}(\Pi_{k+1} f, \mathcal{T}_k)$ contain pairwise orthogonal terms $(\Pi_{k+1} - \Pi_k) f$ which add up to $(\Pi_{\ell+m+1} - \Pi_\ell) f$. This leads to a proof that the sum over all (squared) oscillations in the last displayed inequality is $\leq \operatorname{osc}^2(\Pi_{\ell+m+1} f, \mathcal{T}_\ell)$. The reliability of (5) bounds $LS(\Pi_k f; \sigma_k, u_k) \leq C_2 \eta_k^2(\mathcal{T}_k)$. The combination of the aforementioned arguments proves

$$\sum_{k=\ell}^{\ell+m} \delta^2(\mathcal{T}_k, \mathcal{T}_{k+1}) \leq (C_1/\varepsilon + C_2)\eta_\ell^2 + \varepsilon C_2 \sum_{k=\ell+1}^{\ell+m+1} \eta_k^2(\mathcal{T}_k).$$

This and the application of (QM) to $\eta_{\ell+m+1}^2 \leq \Lambda_5 \eta_\ell^2$ provides the generalisation (A4$_\varepsilon$) in [7, 13], where it is shown that the other axioms plus (A4$_\varepsilon$) for some sufficiently small $\varepsilon > 0$ imply (A4). □

References

[1] P. Binev, W. Dahmen, and R. DeVore, *Adaptive finite element methods with convergence rates*, Numerische Mathematik **97** (2004), 219–268, DOI 10.1007/s00211-003-0492-7.

[2] P. B. Bochev and M. D. Gunzburger, *Least-squares Finite Element Methods*, Applied Mathematical Sciences, vol. 166, Springer, New York, 2009. MR2490235

[3] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*, Springer Series in Computational Mathematics, vol. 44, Springer, Heidelberg, 2013. MR3097958

[4] D. Braess, *Finite Elements: Theory, fast solvers, and applications in elasticity theory; Translated from the German by Larry L. Schumaker*, 3rd ed., Cambridge University Press, Cambridge, 2007. MR2322235

[5] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Texts in Applied Mathematics, vol. 15, Springer, New York, 2008. MR2373954

[6] P. Bringmann, C. Carstensen, and G. Starke, *An adaptive least-squares FEM for linear elasticity with optimal convergence rates*, SIAM J. Numer. Anal. **56** (2018), no. 1, 428–447, DOI 10.1137/16M1083797. MR3757107

[7] C. Carstensen, M. Feischl, M. Page, and D. Praetorius, *Axioms of adaptivity*, Comput. Math. Appl. **67** (2014), no. 6, 1195–1253, DOI 10.1016/j.camwa.2013.12.003. MR3170325

[8] C. Carstensen, D. Gallistl, and M. Schedensack, *Discrete reliability for Crouzeix-Raviart FEMs*, SIAM J. Numer. Anal. **51** (2013), 2935–2955, DOI 10.1137/130915856.

[9] C. Carstensen and F. Hellwig, *Constants in discrete Poincaré and Friedrichs inequalities and discrete quasi-interpolation*, Comput. Methods Appl. Math. **18** (2018), no. 3, 433–450, DOI 10.1515/cmam-2017-0044. MR3824773

[10] C. Carstensen and F. Hellwig, *Optimal convergence rates for adaptive lowest-order discontinuous Petrov-Galerkin schemes*, SIAM J. Numer. Anal. **56** (2018), 1091–1111, DOI 10.1137/17M1146671, `https://doi.org/10.1137/17M1146671`.

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

[12] C. Carstensen, E.-J. Park, and P. Bringmann, *Convergence of natural adaptive least squares finite element methods*, Numer. Math. **136** (2017), 1097–1115, DOI 10.1007/s00211-017-0866-x.

[13] C. Carstensen and H. Rabus, *Axioms of adaptivity with separate marking for data resolution*, SIAM J. Numer. Anal. **55** (2017), no. 6, 2644–2665, DOI 10.1137/16M1068050. MR3719030

[14] C. Carstensen and J. Storn, *Asymptotic Exactness of the least-squares finite element residual*, SIAM J. Numer. Anal. **56** (2018), 2008–2028, DOI 10.1137/17M1125972, `https://doi.org/10.1137/17M1125972`.

[15] 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), 2524–2550, DOI 10.1137/07069047X.

[16] D. A. Di Pietro and A. Ern, *Mathematical Aspects of Discontinuous Galerkin Methods*, Mathématiques & Applications (Berlin) [Mathematics & Applications], vol. 69, Springer, Heidelberg, 2012. MR2882148

[17] R. Stevenson, *Optimality of a standard adaptive finite element method*, Found. Comput. Math. **7** (2007), no. 2, 245–269, DOI 10.1007/s10208-005-0183-0. MR2324418

[18] R. Stevenson, *The completion of locally refined simplicial partitions created by bisection*, Math. Comp. **77** (2008), no. 261, 227–241, DOI 10.1090/S0025-5718-07-01959-X. MR2353951

Institut für Mathematik, Humboldt-Universität zu Berlin, Unter den Linden 6, D-10099 Berlin, Germany
Email address: cc@math.hu-berlin.de