SIAM J. NUMER. ANAL.
Vol. 53, No. 1, pp. 43–62


# CONVERGENCE AND OPTIMALITY OF ADAPTIVE LEAST SQUARES FINITE ELEMENT METHODS*

CARSTEN CARSTENSEN† AND EUN-JAE PARK‡

**Abstract.** The first-order div least squares finite element methods (LSFEMs) allow for an immediate a posteriori error control by the computable residual of the least squares functional. This paper establishes an adaptive refinement strategy based on some equivalent refinement indicators. Since the first-order div LSFEM measures the flux errors in $H(\mathrm{div})$, the data resolution error measures the $L^2$ norm of the right-hand side $f$ minus the piecewise polynomial approximation $\Pi f$ without a mesh-size factor. Hence the data resolution term is neither an oscillation nor of higher order and consequently requires a particular treatment, e.g., by the thresholding second algorithm due to Binev and DeVore. The resulting novel adaptive LSFEM with separate marking converges with optimal rates relative to the notion of a nonlinear approximation class.

**Key words.** least squares finite element method, adaptive algorithm, optimal convergence rates, approximation class, a posteriori error estimates, discrete reliability

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

**DOI.** 10.1137/130949634

**1. Introduction.** The adaptive finite element method (AFEM) has been understood to converge with optimal convergence rates with respect to the concept of a nonlinear approximation class [4, 22, 15, 16]. Although optimal convergence rates are often observed in many numerical experiments [1, 3, 21], even the plain convergence is not known for the natural adaptive least squares finite element method (ALSFEM) and small bulk parameter.

One striking advantage of the least squares finite element method (LSFEM) [6] is the immediate reliable and efficient error control via the least squares functional $LS(f; \mathbf{p}_\ell, u_\ell)$ evaluated at the discrete approximations $(\mathbf{p}_\ell, u_\ell)$ in standard Raviart–Thomas and Courant finite element subspaces of the Sobolev spaces $H(\mathrm{div};\Omega)\times H^1_0(\Omega)$ with respect to a triangulation $\mathcal{T}_\ell$. It is expected that the elementwise evaluation of the least squares functional leads to an effective ALSFEM [1, 3, 21]. One difficulty in the convergence analysis of those schemes is the question of whether the least squares residual is indeed strictly reduced provided the mesh is refined [1].

This paper introduces novel a posteriori error estimators and so motivates a new adaptive mesh-refining strategy ALSFEM with a modification of the standard adaptive loop `SOLVE`, `ESTIMATE`, `MARK`, `REFINE` by separate marking; details follow in section 4. For the purpose of this introduction, it suffices to understand *separate marking* in that there are two cases on each level $\ell$ with a triangulation $\mathcal{T}_\ell$ and the discrete solution $(\mathbf{p}_\ell, u_\ell)$. In Case A for $\|f - f_\ell\|^2 \le \kappa\eta_\ell^2$ with the piecewise constant integral means $f_\ell := \Pi_\ell f$ of the right-hand side $f \in L^2(\Omega)$, the step `MARK` consists of the Dörfler marking with the elementwise contribution

*Received by the editors December 18, 2013; accepted for publication October 10, 2014; published electronically January 8, 2015. The research of the authors was supported in part by the WCU program through NRF (R31-2008-000-10049-0) and NRF-2012R1A2A2A01046471.
http://www.siam.org/journals/sinum/53-1/94963.html
†Institut für Mathematik, Humboldt-Universität zu Berlin, 10099 Berlin, Germany, and Department of Computational Science and Engineering, Yonsei University, Seoul 120-749, Korea (cc@math.hu-berlin.de).
‡Department of Mathematics, and Department of Computational Science and Engineering, Yonsei University, Seoul 120-749, Korea (ejpark@yousei.ac.kr).

$$\eta_\ell^2(T) := \|(1-\Pi_\ell)\mathbf{p}_\ell\|_{L^2(T)}^2 + |T|^{1/2} \sum_{E\in\mathcal{E}(T)\cap\mathcal{E}_\ell(\Omega)} \|[\partial u_\ell/\partial\nu_E]_E\|_{L^2(E)}^2 + |T|^{1/2}\left(\sum_{E\in\mathcal{E}(T)\cap\mathcal{E}_\ell(\Omega)} \|[\mathbf{p}_\ell]_E\cdot\tau_E\|_{L^2(E)}^2 + \sum_{E\in\mathcal{E}(T)\cap\mathcal{E}_\ell(\partial\Omega)} \|\mathbf{p}_\ell\cdot\tau_E\|_{L^2(E)}^2\right). \tag{1.1}$$

This refinement indicator involves the normal components of the jumps of the piecewise gradients $\nabla u_\ell$ of the displacement variable as in standard finite element methods [7, 23] and the tangential components of the jumps of the flux approximation $\mathbf{p}_\ell$ over the interior edges $\mathcal{E}_\ell(\Omega)$ with appropriate modifications along the boundary edges $\mathcal{E}_\ell(\partial\Omega)$.

In the remaining Case B, the adaptive algorithm employs some optimal algorithm to reduce the data resolution term such that the triangulation of the new level $\ell+1$ leads to a data resolution $\|f-\Pi_{\ell+1}f\|$ on the finer triangulation smaller than or equal to a reduction factor $\varrho$ times the data resolution $\|f-\Pi_\ell f\|$ of the current one. One possible realization is the thresholding second algorithm due to Binev and DeVore [5].

The *first* main result of this paper is the *a posteriori error control* through those error estimators in the sense of the equivalence

$$\eta_\ell^2 + ||f-\Pi_\ell f||^2 \approx LS(f;\mathbf{p}_\ell,u_\ell). \tag{1.2}$$

Here and throughout this paper, $a\lesssim b$ denotes $a\le C\ b$ with some generic constant $C$ which may depend on the domain and the initial coarse mesh $\mathcal{T}_0$ but which is independent of the level $\ell$. Moreover, $a\approx b$ abbreviates $a\lesssim b\lesssim a$.

The *second* main result is the *R-linear convergence* of three variants of the aforementioned adaptive LSFEM, where some of the contributions in (1.1) are neglected. In particular, there exists some $0<\varrho_{\rm red}<1$ for the error estimator reduction property

$$\eta_\ell \lesssim \varrho_{\rm red}^\ell(\eta_0 + ||f-\Pi_0 f||) \quad \text{for all } \ell=0,1,2,\ldots. \tag{1.3}$$

The *third* main result is the *optimality* of the novel ALSFEM with the refinement indicator (1.1) with respect to the concept of nonlinear approximations classes following ideas from the finite element community [19] based on nonlinear approximation classes [4]. The proof is based on arguments from [14] to control the overall overhead in the closure algorithm of [4] even for separate marking with Cases A and B.

The remaining parts of this paper are organized as follows. Section 2 introduces the Poisson model problem (PMP) and its least squares discretization with Raviart–Thomas-type flux approximations. Section 3 introduces three alternative a posteriori error estimators $\eta_\ell$ and their reliability and efficiency a posteriori error analysis as indicated in (1.2). Section 4 presents the associated three new ALSFEMs with three parameter choices $(\Xi_1,\Xi_2)=(1,0)$, $(0,1)$, and $(1,1)$ and establishes the error estimator reduction property (1.3) even for small bulk parameters and arbitrarily coarse initial meshes.

Some modified discrete reliability in section 5 holds for $\Xi_1=1=\Xi_2$, which corresponds to (1.1). This allows the proof of the quasi-optimal convergence in section 6.

Standard notation on Lebesgue and Sobolev spaces and norms is employed throughout this paper: $||\cdot||$ denotes the $L^2$ norm and $|||\cdot|||$ denotes the $H^1$ seminorm over the entire domain $\Omega$, while $|||\cdot||||_{NC} := ||\nabla_{NC}\cdot||$ is some piecewise version thereof. The piecewise mesh-size $h_\ell\in L^\infty(\Omega)$ is defined by $h_\ell|_T := h_T = |T|^{1/2}$ for the area $|T|$ of a triangle $T\in\mathcal{T}_\ell$ in the triangulation $\mathcal{T}_\ell$.

The measure $|\cdot|$ is context sensitive and refers to the number of elements of some finite set (e.g., the number $|\mathcal{T}|$ of triangles in a triangulation $\mathcal{T}$) or the length $|E|$ of an edge $E$ or the area $|T|$ of some domain $T$ and not just the modulus of a real number or the Euclidean length of a vector.

The analysis is exploited for polygonal domains in two space dimensions with newest-vertex bisections (NVBs), while the generalization to three dimensions appears incremental.

**2. PMP and its least squares discretisation.** Given $f \in L^2(\Omega)$ on a simply connected bounded polygonal Lipschitz domain $\Omega \subset \mathbb{R}^2$, the PMP seeks some function $u \in C_0(\Omega) \cap H^2_{\mathrm{loc}}(\Omega)$ with

$$-\Delta u = f \text{ in } \Omega \quad \text{and} \quad u = 0 \text{ on } \partial\Omega. \tag{2.1}$$

The least squares methods consider the equivalent first-order system

$$-\operatorname{div} \mathbf{p} = f \quad \text{and} \quad \mathbf{p} - \nabla u = \mathbf{0} \quad \text{in } \Omega. \tag{2.2}$$

The weak form involves the $L^2$ inner product $(\cdot,\cdot)_{L^2(\Omega)}$ and its $L^2$ norm $\|\cdot\|$ over $\Omega$. Standard notation is employed for the Sobolev space $H^1(\Omega)$ with seminorm $|||\cdot|||$ and $V := H^1_0(\Omega)$. The Hilbert space

$$H(\operatorname{div};\Omega) = \{\mathbf{q} \in L^2(\Omega;\mathbb{R}^2) : \operatorname{div} \mathbf{q} \in L^2(\Omega)\}$$

consists of all $L^2$ vector functions $\mathbf{q} = (q_1, q_2)$ with weak divergence $\operatorname{div} \mathbf{q} := \partial_x q_1 + \partial_y q_2$ in $L^2(\Omega)$ and associated norm $\|\cdot\|_{H(\operatorname{div})}$ [7, 9, 10, 17]. The least squares method solves system (2.2) by minimizing the residual functional defined, for any $(\mathbf{q}, v) \in H(\operatorname{div};\Omega) \times V$, by

$$LS(f;\mathbf{q},v) := \|f + \operatorname{div} \mathbf{q}\|^2 + \|\mathbf{q} - \nabla v\|^2. \tag{2.3}$$

The associated Euler–Lagrange equations lead to the weak problem: Seek $(\mathbf{p}, u) \in H(\operatorname{div};\Omega) \times V$ such that, for all $(\mathbf{q}, v) \in H(\operatorname{div};\Omega) \times V$,

$$\int_\Omega (f + \operatorname{div} \mathbf{p}) \operatorname{div} \mathbf{q}\, dx + \int_\Omega (\mathbf{p} - \nabla u)\cdot(\mathbf{q} - \nabla v) dx = 0. \tag{2.4}$$

The well-established equivalence of the norm in $H(\operatorname{div};\Omega) \times V$ with the least squares functional

$$LS(0;\mathbf{q},v) \approx \|\mathbf{q}\|^2_{H(\operatorname{div})} + |||v|||^2 \quad \text{for all } (\mathbf{q},v) \in H(\operatorname{div};\Omega) \times V \tag{2.5}$$

leads to the unique existence of a minimizer of $LS(f;\cdot)$ and weak solution $(\mathbf{p}, u) \in H(\operatorname{div};\Omega) \times V$ [6]. Moreover, the conforming discretization leads to a quasi-optimal convergence.

The prototype example for a discretization is the lowest-order Raviart–Thomas function space $RT_0(\mathcal{T})$ based on a regular triangulation $\mathcal{T}$ of $\Omega$ in closed triangles in the sense of Ciarlet [9, 7], i.e., $\cup\mathcal{T} = \bar{\Omega}$ and any two distinct triangles in $\mathcal{T}$ are either disjoint or share exactly one vertex or one common edge. Given any regular triangulation $\mathcal{T}$ of $\Omega$ into triangles, let

$$\begin{aligned} V(\mathcal{T}) &:= P_1(\mathcal{T}) \cap V, \\ RT_0(\mathcal{T}) &:= \{\mathbf{q} \in P_1(\mathcal{T};\mathbb{R}^2) \cap H(\operatorname{div};\Omega) : \forall\, T \in \mathcal{T}, \exists a_T, b_T, c_T \in \mathbb{R} \\ &\qquad \forall\, x \in T,\, \mathbf{q}(x) = (a_T, b_T) + c_T\, x\}. \end{aligned}$$

There exists a unique minimizer $(\mathbf{p}_{LS}, u_{LS})$ of $LS(f;\cdot)$ in $RT_0(\mathcal{T}) \times V(\mathcal{T})$ and this is characterized as the weak solution of the discrete analogue (2.6) of (2.4). In other words, the LSFEM solution $(\mathbf{p}_{LS}, u_{LS}) \in RT_0(\mathcal{T}) \times V(\mathcal{T}) \subset H(\mathrm{div};\Omega) \times V$ satisfies, for all $(\mathbf{q}_{RT}, v_C) \in RT_0(\mathcal{T}) \times V(\mathcal{T}) \subset H(\mathrm{div};\Omega) \times V$, that

$$\int_\Omega (f + \operatorname{div} \mathbf{p}_{LS}) \operatorname{div} \mathbf{q}_{RT} dx + \int_\Omega (\mathbf{p}_{LS} - \nabla u_{LS}) \cdot (\mathbf{q}_{RT} - \nabla v_C) dx = 0. \tag{2.6}$$

The Céa lemma leads to the best approximation property

$$\|\mathbf{p} - \mathbf{p}_{LS}\|_{H(\mathrm{div})} + |||u - u_{LS}||| \lesssim \min_{\mathbf{q}_{RT} \in RT_0(\mathcal{T})} \|\mathbf{p} - \mathbf{q}_{RT}\|_{H(\mathrm{div})} + \min_{v_C \in V(\mathcal{T})} |||u - v_C|||.$$

Provided the exact solution $u$ belongs to $H^2(\Omega)$ (e.g., for a convex domain $\Omega$), standard approximation results lead to linear convergence in the maximal mesh-size. However, in case of reduced elliptic regularity (e.g., for a nonconvex domain), appropriate mesh-refining strategies are required to avoid suboptimal convergence rates for less regular problems.

This section concludes with some representation result frequently employed throughout the paper. Let $\Pi$ denote the $L^2$ orthogonal projection onto the piecewise constants $P_0(\mathcal{T};\mathbb{R}^m)$ for $m = 1, 2$ with respect to the present triangulation $\mathcal{T}$. Let $CR_0^1(\mathcal{T})$ denote the functions in $P_1(\mathcal{T})$ which are continuous at the midpoints of all interior edges $\mathcal{E}(\Omega)$ and vanish at the midpoints of all boundary edges $\mathcal{E}(\partial\Omega)$. Let $\nabla_{NC}$ denote the piecewise action of the gradient.

Proposition 2.1. *Any Raviart–Thomas function $\mathbf{q}_{RT} \in RT_0(\mathcal{T})$ reads*

$$\mathbf{q}_{RT} = \Pi \mathbf{q}_{RT} + \operatorname{div} \mathbf{q}_{RT}/2(\bullet - \mathrm{mid}(\mathcal{T})) \quad \text{a.e. in } \Omega \tag{2.7}$$

*(where $\bullet - \mathrm{mid}(\mathcal{T})$ abbreviates $x - \mathrm{mid}(T)$ at any $x \in T \in \mathcal{T}$ with center of inertia $\mathrm{mid}(T)$) and satisfies, for unique $v_{CR} \in CR_0^1(\mathcal{T})$ and $w_C \in (P_1(\mathcal{T}) \cap C(\overline{\Omega}))/\mathbb{R}$, that*

$$\Pi \mathbf{q}_{RT} = \nabla_{NC} v_{CR} + \operatorname{Curl} w_C. \tag{2.8}$$

*Therein, $v_{CR} \in CR_0^1(\mathcal{T})$ is the Crouzeix–Raviart solution of the PMP with right-hand side $-\operatorname{div} \mathbf{q}_{RT} \in L^2(\Omega)$, i.e., $v_{CR}$ solves the nonconforming finite element problem*

$$\int_\Omega \nabla_{NC} v_{CR} \cdot \nabla_{NC} w_{CR} dx = -\int_\Omega w_{CR} \operatorname{div} \mathbf{q}_{RT} dx \quad \text{for all } w_{CR} \in CR_0^1(\mathcal{T}). \tag{2.9}$$

*Moreover, for any discrete solution $\mathbf{q}_{RT}$ of a mixed finite element problem or any LSFEM solution $\mathbf{q}_{RT} := \mathbf{p}_{LS}$ of (2.6), $w_C \equiv 0$ holds in (2.8). In other words, those particular Raviart–Thomas fluxes are $L^2$ orthogonal onto $\operatorname{Curl}(P_1(\mathcal{T}) \cap C(\overline{\Omega}))$.*

*Proof.* The identities (2.7)–(2.9) are proved in [18] but have been essentially known since [2]. The formula (2.7) follows from the very definition of the Raviart–Thomas functions. The formula (2.8) is a discrete Helmholtz decomposition for simply connected domains of any piecewise constant vector field.

The proof of the $L^2$ orthogonality follows from the observation that any function in $\operatorname{Curl}(P_1(\mathcal{T}) \cap C(\overline{\Omega}))$ is a divergence-free Raviart–Thomas function; the converse holds as well for the simply connected domain. This plus the discrete equation with such a test function leads to the asserted $L^2$ orthogonality. ∎

**3. Alternative a posteriori error control.** The equivalence (2.5) plus (2.2) imply the error-residual equivalence

$$LS(f;\mathbf{p}_{LS},u_{LS}) \approx \|\mathbf{p}-\mathbf{p}_{LS}\|^2_{H(\mathrm{div})} + |||u-u_{LS}|||^2. \tag{3.1}$$

Hence the least squares functional $LS(f;\mathbf{p}_{LS},u_{LS})$ is an explicit residual-based a posteriori error estimator (even for inexact solve). To design an optimal adaptive LSFEM, this section introduces alternative a posteriori error estimators and presents their reliability and efficiency analysis.

Throughout this section, $\mathcal{T}$ denotes a fixed shape-regular triangulation of the domain $\Omega$ with LSFEM solutions $(\mathbf{p}_{LS},u_{LS})$ with mesh-size $h_\mathcal{T}\in P_0(\mathcal{T})$ defined by $h_\mathcal{T}|_T := |T|^{1/2}$ for all $T\in\mathcal{T}$ $\nu$, and let $\Pi$ denote the associated $L^2$ orthogonal projection onto $P_0(\mathcal{T};\mathbb{R}^m)$.

For each interior edge $E=\partial T_+\cap\partial T_-$ shared by the two triangles $T_+$ and $T_-$ with unit normal vector $\nu_E$ (and unit tangent vector $\tau_E$), let $[\cdot]_E$ denote the jump across $E$; for instance,

$$[\nabla u_{LS}]_E(x) := \lim_{T_+\ni x_+\to x}\nabla u_{LS}(x_+) - \lim_{T_-\ni x_-\to x}\nabla u_{LS}(x_-)\quad\text{for all } x\in E.$$

Given such interior edge $E\in\mathcal{E}(\Omega)$ of length $|E|=\mathrm{diam}(E)$ with edge-patch $\omega_E := \mathrm{int}(T_+\cup T_-)$, set

$$\eta^2(E) := |E|\,\|[\nabla u_{LS}]_E\cdot\nu_E\|^2_{L^2(E)}. \tag{3.2}$$

The normal component of the Raviart–Thomas functions is continuous across the edge $E\in\mathcal{E}(\Omega)$, while its tangential component is some generally discontinuous piecewise affine function which allows for the jump

$$[\mathbf{p}_{LS}]_E(x) := \lim_{T_+\ni x_+\to x}\mathbf{p}_{LS}(x_+) - \lim_{T_-\ni x_-\to x}\mathbf{p}_{LS}(x_-)\quad\text{for all } x\in E\in\mathcal{E}(\Omega).$$

The tangential unit vector $\tau_E$ along the edge $E$ determines the error estimator contribution

$$\mu^2(E) := |E|\,\|[\mathbf{p}_{LS}]_E\cdot\tau_E\|^2_{L^2(E)}\quad\text{for all } E\in\mathcal{E}(\Omega).$$

There is no jump residual of $\nabla u_{LS}$ along the boundary edges and those contributions do not arise owing to test functions which vanish along the boundary $\partial\Omega$; hence,

$$\eta^2(E) := 0\quad\text{for all } E\in\mathcal{E}(\partial\Omega) := \{E\in\mathcal{E} : E\subset\partial\Omega\}.$$

The boundary condition $u=0$ along $\partial\Omega$ is reflected in $\mathbf{p}\cdot\tau_E=0$ for all boundary edges $E\in\mathcal{E}(\partial\Omega)$. This gives rise to the residual

$$\mu^2(E) := |E|\,\|\mathbf{p}_{LS}\cdot\tau_E\|^2_{L^2(E)}\quad\text{for all } E\in\mathcal{E}(\partial\Omega).$$

The novel estimators are reliabile and efficient in the form

$$\eta(\mathcal{E}) := \left(\sum_{E\in\mathcal{E}}\eta^2(E)\right)^{1/2}\quad\text{and}\quad \mu(\mathcal{E}) := \left(\sum_{E\in\mathcal{E}}\mu^2(E)\right)^{1/2}.$$

THEOREM 3.1. *Let $(\mathbf{p},u)\in H\times V$ solve* (2.2)*, let $(\mathbf{p}_{LS},u_{LS})\in RT_0(\mathcal{T})\times V(\mathcal{T})$ solve* (2.4)*, and adopt the aforementioned notation on $\mathcal{T}$, $h_\mathcal{T}$, $\eta(\mathcal{E})$, and $\mu(\mathcal{E})$. Then it holds that*

$$\begin{aligned}\|\mathbf{p}-\mathbf{p}_{LS}\|_{H(\mathrm{div})} + |||u-u_{LS}||| &\approx LS(f;\mathbf{p}_{LS},u_{LS})^{1/2}\\ &\approx \|h_\mathcal{T}\,\mathrm{div}\,\mathbf{p}_{LS}\| + \|f-\Pi f\| + \eta(\mathcal{E}) \qquad (3.3)\\ &\approx \|h_\mathcal{T}\,\mathrm{div}\,\mathbf{p}_{LS}\| + \|f-\Pi f\| + \mu(\mathcal{E}). \qquad (3.4)\end{aligned}$$

Three comments are in order before the proof of Theorem 3.1 concludes this section.

*Remark* 3.2 (interpretation of $\mu(\mathcal{E})$ and $\eta(\mathcal{E})$). One striking observation is that either contribution $\mu(\mathcal{E})$ or $\eta(\mathcal{E})$ may be left out in (3.3) or (3.4). The term $\|\mathbf{p}_{LS} - \nabla u_{LS}\|$ has the interpretation of either a consistency error of $\mathbf{p}_{LS}$ (that is, its distance to gradients of Sobolev functions in $V$) and then leads to $\mu(\mathcal{E})$ or an interpretation as an equilibrium error of $\nabla u_{LS}$ (that is, its distance to the functions in $H(\mathrm{div},\Omega)$ with divergence $\mathrm{div}\,\mathbf{p}_{LS}$) and then leads to $\eta(\mathcal{E})$.

*Remark* 3.3 (methodology). The reliability proof in Theorem 3.1, that is, the assertion $\lesssim$ in (3.3) and in (3.4), follows with standard arguments like integration by parts and stability and approximation properties of quasi interpolation. The proof works without any Helmholtz decomposition but employs Proposition 2.1. The efficiency, that is, the assertion $\gtrsim$ in (3.3) and in (3.4), follows from Verfürth's inverse estimate technology [23].

*Remark* 3.4 (exact solve required). One essential difference between the reliable and efficient a posteriori error control of (3.1) and that of Theorem 3.1 is that the latter equivalence holds for the *exact* least squares solutions only. The proof below relies on the fact that (2.6) holds exactly (i.e., the linear system is solved exactly or at least in sufficiently high accuracy). In contrast to that, (3.1) holds for *any* discrete approximation $(\mathbf{p}_{LS}, u_{LS})$ and so allows for *inexact solve*.

*Proof of Theorem* 3.1. *The first step of the proof* concerns the bilinear form $b : L^2(\Omega) \times H(\mathrm{div},\Omega) \to \mathbb{R}$, defined by

$$b(v,\mathbf{q}) := \int_\Omega v \ \mathrm{div}\,\mathbf{q}\ dx \quad \text{for all } (v,\mathbf{q}) \in L^2(\Omega) \times H(\mathrm{div},\Omega).$$

The inf-sup condition of $b$ is well understood in the context of mixed FEMs [7, 9, 10]. In particular, the LBB condition for $P_0(\mathcal{T}) \times RT_0(\mathcal{T})$ is well established and shows for any piecewise constant function, like $\Pi f + \mathrm{div}\,\mathbf{p}_{LS}$, that there exists $\mathbf{q}_{RT} \in RT_0(\mathcal{T})$ with

$$b(\Pi f + \mathrm{div}\,\mathbf{p}_{LS}, \mathbf{q}_{RT}) = \|\Pi f + \mathrm{div}\,\mathbf{p}_{LS}\|^2 \quad \text{and} \quad \|\mathbf{q}_{RT}\|_{H(\mathrm{div})} \lesssim \|\Pi f + \mathrm{div}\,\mathbf{p}_{LS}\|.$$

For the chosen $\mathbf{q}_{RT}$ and $v_C := 0$, the solution of LSFEM satisfies

$$\begin{aligned} &b(\Pi f + \mathrm{div}\,\mathbf{p}_{LS}, \mathbf{q}_{RT}) \\ &\quad = \int_\Omega (f + \mathrm{div}\,\mathbf{p}_{LS})\,\mathrm{div}\,\mathbf{q}_{RT}\,dx \\ &\quad = -\int_\Omega (\mathbf{p}_{LS} - \nabla u_{LS}) \cdot \mathbf{q}_{RT}\,dx. \end{aligned}$$

The combination of the previous arguments plus a Cauchy inequality show

$$\|\Pi f + \mathrm{div}\,\mathbf{p}_{LS}\|^2 \lesssim \|\mathbf{p}_{LS} - \nabla u_{LS}\| \|\Pi f + \mathrm{div}\,\mathbf{p}_{LS}\|.$$

The resulting estimate plus a triangle inequality prove that

$$\|f + \mathrm{div}\,\mathbf{p}_{LS}\| \lesssim \|f - \Pi f\| + \|\mathbf{p}_{LS} - \nabla u_{LS}\|. \tag{3.5}$$

One interpretation of the estimate (3.5) in the equivalence (3.1) is some split of the residuals in the data resolution error $\|f - \Pi f\|$ and the residual $\|\mathbf{p}_{LS} - \nabla u_{LS}\|$. $\square$

*The second step of the proof* employs Proposition 2.1 with

$$\mathbf{p}_{LS} = \nabla_{NC}\widehat{u}_{CR} + \frac{\operatorname{div}\mathbf{p}_{LS}}{2}(\bullet - \operatorname{mid}(\mathcal{T}))$$

and the Crouzeix–Raviart solution $\widehat{u}_{CR}$ of the PMP with right-hand side $-\operatorname{div}\mathbf{p}_{LS}$. This implies the decomposition

$$\|\mathbf{p}_{LS} - \nabla u_{LS}\|^2 = |||\widehat{u}_{CR} - u_{LS}|||_{NC}^2 + \|(1-\Pi)\mathbf{p}_{LS}\|^2. \quad \square \tag{3.6}$$

*Step three of the proof* studies $v_{CR} := \widehat{u}_{CR} - u_{LS}$ and its conforming quasi–interpolation $v_1$ from [8]. Given $v_{CR} \in CR_0^1(\mathcal{T})$ and any interior node $z \in \mathcal{N}(\Omega)$, define the averages $v_1(z)$ of the (possibly) discontinuous values of $v_{CR}$. Linear interpolation of those values defines $v_1 \in P_1(\mathcal{T}) \cap C_0(\Omega)$ with the approximation and stability estimate [8]

$$\|(v_{CR} - v_1)/h_{\mathcal{T}}\| + |||v_1|||_{NC} \lesssim |||v_{CR}|||_{NC} \le \|\mathbf{p}_{LS} - \nabla u_{LS}\|. \tag{3.7}$$

The aforementioned split and the Galerkin orthogonality of $\mathbf{p}_{LS} - \nabla u_{LS}$ onto $\nabla v_1$ show

$$\begin{aligned}|||\widehat{u}_{CR} - u_{LS}|||_{NC}^2 &= \int_\Omega (\mathbf{p}_{LS} - \nabla u_{LS}) \cdot \nabla_{NC} v_{CR}\, dx \\ &= \int_\Omega (\mathbf{p}_{LS} - \nabla u_{LS}) \cdot \nabla_{NC}(v_{CR} - v_1)\, dx.\end{aligned}$$

A piecewise integration by parts with a careful rearrangement of the edge contributions shows that this last term equals

$$\int_\Omega (v_1 - v_{CR}) \operatorname{div}\mathbf{p}_{LS}\, dx + \sum_{E\in\mathcal{E}(\Omega)} \int_E (v_1 - v_{CR})[\partial u_{LS}/\partial \nu_E]_E\, ds.$$

The trace theorem along each edge plus (3.7) show that the edge contributions are

$$\lesssim \left( \sum_{E\in\mathcal{E}(\Omega)} |E|\, \|[\partial u_{LS}/\partial\nu_E]_E\|_{L^2(E)}^2 \right)^{1/2} |||\widehat{u}_{CR} - u_{LS}|||_{NC}.$$

This and some weighted Cauchy inequality with (3.7) lead to

$$|||\widehat{u}_{CR} - u_{LS}|||_{NC} \lesssim \eta(\mathcal{E}) + \|h_{\mathcal{T}} \operatorname{div}\mathbf{p}_{LS}\|. \quad \square$$

*Step four of the proof* is the equivalence $\|(1-\Pi)\mathbf{p}_{LS}\| \approx \|h_{\mathcal{T}} \operatorname{div}\mathbf{p}_{LS}\|$. This follows from the representation of Raviart–Thomas functions in Proposition 2.1. The combination of the aforementioned estimates concludes the proof of $\lesssim$ in (3.3). $\square$

*Step five of the proof* starts with the decomposition (3.6) from step three and focuses on the interpretation of the remaining term

$$|||\widehat{u}_{CR} - u_{LS}|||_{NC} = \min_{v_C\in V(\mathcal{T})} |||\widehat{u}_{CR} - v_C|||_{NC}$$

as the best-approximation error when $\widehat{u}_{CR}$ is approximated by a piecewise affine conforming function. This consistency error is well understood in the a posteriori error analysis of Crouzeix–Raviart nonconforming finite elements. An explicit proof and related references of

$$\min_{v_C\in V(\mathcal{T})}|||\widehat{u}_{CR}-v_C|||_{NC}^2 \approx \min_{v\in V}|||\widehat{u}_{CR}-v|||_{NC}^2 \approx \sum_{E\in\mathcal{E}}|E|\,||[\partial\widehat{u}_{CR}/\partial s]_E||_{L^2(S)}^2$$

can be found in [12, Theorem 5.1]. The relation of $\nabla_{NC}\widehat{u}_{CR}$ to $\mathbf{p}_{LS}$ from step two implies

$$\sum_{E\in\mathcal{E}}|E|\,||[\partial\widehat{u}_{CR}/\partial s]_E||_{L^2(E)}^2 \lesssim \mu^2(\mathcal{E}) + \|h_{\mathcal{T}}\operatorname{div}\mathbf{p}_{LS}\|^2.$$

This follows for each edge $E\in\mathcal{E}$ with edge-patch $\omega_E$ from the estimate

$$\left|\int_E [\operatorname{div}\mathbf{p}_{LS}(x-\operatorname{mid}(\mathcal{T}))]_E\cdot\tau_E\,ds_x\right| \lesssim |E|\,\|\operatorname{div}\mathbf{p}_{LS}\|_{L^2(\omega_E)}.$$

The combination of the aforementioned estimates concludes the proof of $\lesssim$ in (3.4).

Efficiency will be proved for each of the local contributions to the estimator. ◻

*Step six* concerns the affine function $(\mathbf{p}_{LS}-\nabla u_{LS})|_T$ on some fixed $T\in\mathcal{T}$. The inverse estimate

$$\|\operatorname{div}\mathbf{p}_{LS}\|_{L^2(T)} = \|\operatorname{div}(\mathbf{p}_{LS}-\nabla u_{LS})\|_{L^2(T)} \lesssim h_T^{-1}\|\mathbf{p}_{LS}-\nabla u_{LS}\|_{L^2(T)}$$

proves a local form of the efficiency (in terms of the residual (2.4))

$$\|h_{\mathcal{T}}\operatorname{div}\mathbf{p}_{LS}\| \lesssim \|\mathbf{p}_{LS}-\nabla u_{LS}\|. \qquad \square$$

*Step seven* is the simple observation $\|f-\Pi f\|\le\|f+\operatorname{div}\mathbf{p}_{LS}\|$. ◻

*Step eight* establishes the efficiency of $\eta(E)$ for an interior edge $E=\partial T_+\cap\partial T_-$ shared by the two triangles $T_+$ and $T_-$. Define the edge-bubble function $b_E\in W_0^{1,\infty}(\omega_E)$ as four times the product of the two barycentric coordinates of $T_\pm$ associated to the two vertices at the tips of $E$. Then, the maximum of $b_E$ equals one and is attained at the midpoint of $E$. In fact, since $[\nabla u_{LS}]_E$ is constant, up to a proper sign $\sigma=\pm1$,

$$\sigma\,\eta(E) = \int_E [\nabla u_{LS}]_E\cdot\nu_E\,ds \approx \int_E b_E[\nabla u_{LS}]_E\cdot\nu_E\,ds.$$

An integration by parts on $T_+$ and $T_-$ shows

$$\begin{aligned}\int_E b_E[\nabla u_{LS}]_E\cdot\nu_E\,ds &= \int_E b_E[\nabla u_{LS}-\mathbf{p}_{LS}]_E\cdot\nu_E\,ds\\ &= \int_{\omega_E}\big((\nabla u_{LS}-\mathbf{p}_{LS})\cdot\nabla b_E - b_E\operatorname{div}\mathbf{p}_{LS}\big)dx.\end{aligned}$$

The combination of the previous estimates followed by Cauchy inequalities with $\|b_E\|\approx|E|$ and $\|\nabla b_E\|=1$ leads eventually to

$$\eta(E)\lesssim\|\mathbf{p}_{LS}-\nabla u_{LS}\|_{L^2(\omega_E)}+\|h_{\mathcal{T}}\operatorname{div}\mathbf{p}_{LS}\|_{L^2(\omega_E)}.$$

The efficiency of the edge contribution therefore follows from that of $\|h_{\mathcal{T}}\operatorname{div}\mathbf{p}_{LS}\|$. ◻

*Step nine of the proof* deduces the efficiency of $\mu(E)$ for an edge $E \in \mathcal{E}$ from the literature. The discrete local efficiency proof of the Raviart–Thomas mixed finite element method of [11, section 4] establishes an estimate for the PMP. The arguments apply here as well when the right-hand side is replaced by $-\operatorname{div} \mathbf{p}_{LS}$ and $p_h - p_H$ in [11, Theorem 4.1] is replaced by $\nabla u_{LS} - \mathbf{p}_{LS}$. In fact, the proof in [11, Lemma 4.2] solely requires $p_h \equiv \nabla u_{LS}$ to be orthogonal to some local rotation. In conclusion, the arguments of [11, section 4] prove in the present situation and notation that

$$\mu(E) \lesssim \|\mathbf{p}_{LS} - \nabla u_{LS}\|_{L^2(\omega_E)} + \|h_{\mathcal{T}} \operatorname{div} \mathbf{p}_{LS}\|_{L^2(\omega_E)}.$$

The analysis is provided in [11, section 4] for an interior edge only, but the arguments apply to some boundary edge as well. Further details are omitted. $\square$

*Step ten of the proof* finishes it. In fact, the local efficiency results combine to the claimed remaining inequalities. Further details are omitted for brevity. $\square$

**4. A LSFEM and the contraction property.** This section is devoted to three ALSFEMs with separate marking for the error estimator and the data resolution. The versions add or neglect the contribution $\mu_\ell(\mathcal{E})$ or $\eta_\ell(\mathcal{E})$ as seen from Theorem 3.1 via some global parameter $\Xi_1, \Xi_2 \in \{0,1\}$ with $\Xi_1 + \Xi_2 \geq 1$.

**Adaptive algorithm (ALSFEM).** Input: Initial regular triangulation $\mathcal{T}_0$ of the polygonal domain $\Omega$ into triangles and parameters $0 < \Theta \leq 1$, $0 < \varrho < 1$, $0 < \kappa < \infty$, and $\Xi_1, \Xi_2 \in \{0,1\}$ with $1 \leq \Xi_1 + \Xi_2$.

For any level $\ell = 0, 1, 2, 3, \ldots$
`SOLVE` LSFEM based on regular triangulation $\mathcal{T}_\ell$ with solution $(\mathbf{p}_\ell, u_\ell)$.
`ESTIMATE` $\eta_\ell^2 := \sum_{T \in \mathcal{T}_\ell} \eta_\ell^2(T)$ where, for any $T \in \mathcal{T}_\ell$,

$$\eta_\ell^2(T) := \|(1 - \Pi_\ell)\mathbf{p}_\ell\|^2_{L^2(T)} + \Xi_1 |T|^{1/2} \sum_{E \in \mathcal{E}(T) \cap \mathcal{E}_\ell(\Omega)} \|[\partial u_\ell / \partial \nu_E]_E\|^2_{L^2(E)}$$
$$+ \Xi_2 |T|^{1/2} \left( \sum_{E \in \mathcal{E}(T) \cap \mathcal{E}_\ell(\Omega)} \|[\mathbf{p}_\ell]_E \cdot \tau_E\|^2_{L^2(E)} + \sum_{E \in \mathcal{E}(T) \cap \mathcal{E}_\ell(\partial\Omega)} \|\mathbf{p}_\ell \cdot \tau_E\|^2_{L^2(E)} \right).$$

($\Pi_\ell$ denotes $L^2$ orthogonal projection onto $P_0(\mathcal{T}_\ell; \mathbb{R}^m)$; $f_\ell := \Pi_\ell f$.)
`MARK` in Case A for $\|f - f_\ell\|^2 \leq \kappa \eta_\ell^2$: Compute subset $\mathcal{M}_\ell$ of $\mathcal{T}_\ell$ of (almost) minimal cardinality $|\mathcal{M}_\ell|$ with

$$\Theta \eta_\ell^2 \leq \eta_\ell^2(\mathcal{M}_\ell) := \sum_{T \in \mathcal{M}_\ell} \eta_\ell^2(T). \tag{4.1}$$

`REFINE` Compute the smallest regular refinement $\mathcal{T}_{\ell+1}$ of $\mathcal{T}_\ell$ with $\mathcal{M}_\ell \subset \mathcal{T}_\ell \backslash \mathcal{T}_{\ell+1}$ by NVB.
Case B for $\kappa \eta_\ell^2 < \|f - f_\ell\|^2$: Compute an admissible refinement of $\mathcal{T}_{\ell+1}$ with (almost) minimal cardinality and

$$\|f - f_{\ell+1}\|^2 \leq \varrho \|f - f_\ell\|^2. \tag{4.2}$$

Output: Sequence of discrete solutions $(\mathbf{p}_\ell, u_\ell)_{\ell \in \mathbb{N}_0}$ and data approximations $(f_\ell)_{\ell \in \mathbb{N}_0}$.

In the further analysis of this paper, all three ALSFEMs are convergent while solely the most comprehensive one with $\Xi_1 = 1 = \Xi_2$ is proved to converge with optimal rates.

THEOREM 4.1. *For any of the three choices* $(\Xi_1, \Xi_2) \in \{(1,0), (0,1), (1,1)\}$, *there exist some constants* $0 < \varrho_{red} < 1$ *and* $0 < \Lambda_{red} < \infty$ *which depend only on* $\mathcal{T}_0, \Theta, \kappa, \varrho$ *such that, for all* $\ell \in \mathbb{N}_0$,

$$\xi_\ell^2 := LS(f;\mathbf{p}_\ell,u_\ell)+\|f-f_\ell\|^2+\Lambda_{red}\eta_\ell^2 \quad \textit{satisfies} \quad \xi_{\ell+1}^2 \le \varrho_{red}\xi_\ell^2.$$

*Remark* 4.2 (parameters). The contraction property does *not* require any restriction on $0<\Theta\le 1$, on $0<\varrho<1$, or on $0<\kappa<\infty$. This is important because those parameters have to be chosen as input of the adaptive algorithm. The constants $\varrho_{\rm red}$ and $\Lambda_{\rm red}$ will depend on the choice but not the fact $\varrho_{\rm red}<1$. The proof below reveals, furthermore, that $\Lambda_{\rm red}$ could be chosen smaller at the expense that $\varrho_{\rm red}$ increases. Indeed, $\Lambda_{\rm red}\to 0^+$ implies $\varrho_{\rm red}\to 1^-$.

*Remark* 4.3 (data reduction). The thresholding second algorithm of [5] is one possible realization of an optimal refinement in Case B of ALSFEM. Any other (quasi-) optimal algorithm for the data reduction may be employed in the algorithm and the analysis. Note that some extra overlay computation may be required which is included in the optimality analysis.

*Proof of Theorem* 4.1. The Dörfler marking in Case A is well-understood in the context of adaptive FEM [15] and leads to $0<\varrho_{\rm red}(\ell):=\varrho'_{\rm red}<1$ and $\Lambda'_{\rm red}\approx 1$ with

$$\eta_{\ell+1}\le \varrho_{\rm red}(\ell)\eta_\ell+\Lambda'_{\rm red}(\|\mathbf{p}_{\ell+1}-\mathbf{p}_\ell\|+|||u_{\ell+1}-u_\ell|||). \tag{4.3}$$

The proof of this is essentially a triangle inequality plus the fact that the weights in the error estimates are given in terms of the area $|T|$ of a triangle $T\in\mathcal{T}_{\ell+1}\backslash\mathcal{T}_\ell$ (and hence reduced by a factor $\le 1/2$). The bulk criterion controls the majority of triangles and leads to (4.3); for further details on the reduction of the flux terms see [14, 15].

Case B does not allow any control over the error estimator contributions in $\eta_\ell$ and hence leads to (4.3) with $\varrho_{\rm red}(\ell):=1$.

On the other hand, the data resolution is not controlled in Case A and so

$$\|f-f_{\ell+1}\|^2\le\varrho(\ell)\|f-f_\ell\|^2 \tag{4.4}$$

holds for $\varrho(\ell):=1$. In Case B, (4.2) leads to (4.4) with $\varrho(\ell):=\varrho$.

The above estimates (4.3)–(4.4) hold in each of Cases A and B for the respective values of $0<\varrho_{\rm red}(\ell)\le 1$ and $0<\varrho(\ell)\le 1$ with $\min\{\varrho_{\rm red}(\ell),\varrho(\ell)\}=\max\{\varrho'_{\rm red},\varrho\}<1$ and so for any $\ell\in\mathbb{N}_0$.

Straightforward algebra with the Galerkin orthogonality of the LSFEM shows that

$$LS(0;\mathbf{p}_{\ell+1}-\mathbf{p}_\ell,u_{\ell+1}-u_\ell)=LS(f;\mathbf{p}_\ell,u_\ell)-LS(f;\mathbf{p}_{\ell+1},u_{\ell+1}). \tag{4.5}$$

This and (2.5) prove

$$\|\mathbf{p}_{\ell+1}-\mathbf{p}_\ell\|^2+|||u_{\ell+1}-u_\ell|||^2\lesssim LS(f;\mathbf{p}_\ell,u_\ell)-LS(f;\mathbf{p}_{\ell+1},u_{\ell+1}).$$

For any $0<\lambda<\infty$, there exits $\Lambda''_{\rm red}\approx 1$ such that the last estimate and (4.3) lead to

$$\eta_{\ell+1}^2\le\varrho_{\rm red}^2(\ell)(1+\lambda)\eta_\ell^2+(1+1/\lambda)\Lambda''_{\rm red}(LS(f;\mathbf{p}_\ell,u_\ell)-LS(f;\mathbf{p}_{\ell+1},u_{\ell+1})).$$

The multiplication with $\Lambda_{\rm red}:=1/(\Lambda''_{\rm red}(1+1/\lambda))$ results in

$$\begin{aligned}\xi_{\ell+1}^2&=LS(f;\mathbf{p}_{\ell+1},u_{\ell+1})+\|f-f_{\ell+1}\|^2+\Lambda_{\rm red}\eta_{\ell+1}^2\\&\le LS(f;\mathbf{p}_\ell,u_\ell)+\|f-f_{\ell+1}\|^2+\varrho_{\rm red}^2(\ell)(1+\lambda)\Lambda_{\rm red}\eta_\ell^2.\end{aligned}$$

For $0 < \epsilon < \min\{1, \Lambda_{\text{red}}\}$, this is equivalent to

$$\begin{aligned}\xi_{\ell+1}^2 \leq{}& (1-\epsilon) LS(f;\mathbf{p}_\ell, u_\ell) + (1-\epsilon)\|f - f_\ell\|^2 + (\Lambda_{\text{red}} - \epsilon)\eta_\ell^2 \\ &+ \epsilon LS(f;\mathbf{p}_\ell, u_\ell) + \|f - f_{\ell+1}\|^2 - (1-\epsilon)\|f - f_\ell\|^2 \\ &+ (\epsilon - \Lambda_{\text{red}} + \varrho_{\text{red}}^2(\ell)\,(1+\lambda)\Lambda_{\text{red}})\eta_\ell^2.\end{aligned} \tag{4.6}$$

The first three terms on the right-hand side of (4.6) are less than or equal to $\varrho_{\text{red}}\,\xi_\ell^2$ for $\varrho_{\text{red}} := \max\{1-\epsilon, 1-\epsilon/\Lambda_{\text{red}}\} < 1$, while the last four terms are analyzed in Cases A and B separately.

Case A implies $0 < \varrho_{\text{red}}(\ell) = \varrho'_{\text{red}} < 1$ and

$$\|f - f_{\ell+1}\|^2 \leq \|f - f_\ell\|^2 \leq \kappa \eta_\ell^2.$$

For all $(\Xi_1, \Xi_2) \in \{(1,0),(0,1),(1,1)\}$, Theorem 3.1 implies reliability of $\eta_\ell$ in the sense that there exists some $C_{rel} \approx 1$ with

$$LS(f;\mathbf{p}_\ell, u_\ell) \leq C_{rel}(\eta_\ell^2 + \|f - f_\ell\|^2).$$

The previous two estimates prove that the four remaining terms are smaller than or equal to $\eta_\ell^2$ times the factor $\epsilon(1+\kappa)(1+C_{rel}) - \Lambda_{\text{red}}(1 - \varrho'_{\text{red}}(1+\lambda))$. This factor is nonpositive for any sufficiently small but positive $\epsilon$ and $\lambda$. This concludes the proof in Case A.

Case B is characterized by $\varrho_{\text{red}}(\ell) = 1$ and

$$\max\{\|f - f_{\ell+1}\|^2/\varrho, \kappa\eta_\ell^2\} \leq \|f - f_\ell\|^2.$$

The reliability of Theorem 3.1 plus the bound on $\eta_\ell$ show that the last four terms in (4.6) are smaller than or equal to $\|f - f_\ell\|^2$ times $\epsilon(1+C_{rel})(1+1/\kappa) + \lambda\Lambda_{\text{red}}/\kappa + \varrho - 1$. Since $0 < \varrho < 1$, this factor is nonpositive for sufficiently small $\epsilon$ and $\lambda$. This concludes the proof in Case B.

Altogether, (4.6) proves $\xi_{\ell+1}^2 \leq \varrho_{\text{red}}\,\xi_\ell^2$ in all different cases. ∎

**5. Discrete reliability.** The essential ingredient in the optimality proof for the adaptive algorithm with $\Xi_1 = 1 = \Xi_2$ is some new version of the discrete reliability with some extra term on the right-hand side. For any two regular triangulations $\mathcal{T}_\ell$ and $\mathcal{T}_{\ell+m}$ with respect to some admissible refinement $\mathcal{T}_{\ell+m}$ of $\mathcal{T}_\ell$ over $m$ levels, let $(\mathbf{p}_\ell, u_\ell)$ and $(\mathbf{p}_{\ell+m}, u_{\ell+m})$ denote the respective LSFEM solutions.

THEOREM 5.1 (discrete reliability). *The error estimator $\eta_\ell$ of section* 6 *for* $(\Xi_1, \Xi_2) = (1,1)$ *satisfies*

$$\begin{aligned}&\|\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \nabla(u_{\ell+m} - u_\ell)\|^2 + \|\operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell)\|^2 \\ &\quad\lesssim \eta_\ell^2(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+m}) + \|(1 - \Pi_\ell)\operatorname{div}\mathbf{p}_{\ell+m}\|^2.\end{aligned}$$

The remainder of this section is devoted to the proof of Theorem 5.1 with three lowest-order Raviart–Thomas mixed finite element solutions of the PMP with the respective right-hand sides

$$-\operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell),\ -\Pi_\ell \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell),\ -\Pi_\ell \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell)$$

and the respective triangulations from the levels $\ell+m, \ell+m, \ell$ (indicated through the lower index) which lead to the respective fluxes $\mathbf{q}_{\ell+m} \in RT_0(\mathcal{T}_{\ell+m})$, $\mathbf{q}^*_{\ell+m} \in RT_0(\mathcal{T}_{\ell+m})$, $\mathbf{q}_\ell \in RT_0(\mathcal{T}_\ell)$.

LEMMA 5.2. *There exists some $\beta_{\ell+m} \in C(\Omega) \cap P_1(\mathcal{T}_{\ell+m})$ such that, for any $v_\ell \in V(\mathcal{T}_\ell)$,*

$$\begin{aligned}
&\|\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \nabla(u_{\ell+m} - u_\ell)\|^2 + \|\Pi_\ell \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell)\|^2 \\
&\qquad = \int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot (\nabla(u_{\ell+m} - u_\ell - v_\ell) - \operatorname{Curl} \beta_{\ell+m}) dx \\
&\qquad\quad + \int_\Omega (\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \nabla(u_{\ell+m} - u_\ell)) \cdot (\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) dx.
\end{aligned}$$

*Proof*. The LSFEM on level $\ell + m$ followed by elementary algebra with the $L^2$ projection $\Pi_\ell$ onto $P_0(\mathcal{T}_\ell)$ (or vectors thereof) results in

$$\begin{aligned}
&- \int_\Omega (\mathbf{p}_{\ell+m} - \nabla u_{\ell+m}) \cdot (\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \nabla(u_{\ell+m} - u_\ell)) dx \\
&\qquad = \int_\Omega (f_{\ell+m} + \operatorname{div} \mathbf{p}_{\ell+m}) \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell) dx \\
&\qquad = \int_\Omega (f_{\ell+m} - f_\ell) \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell) dx + \|\operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell)\|^2 \\
&\qquad\quad + \int_\Omega (f_\ell + \operatorname{div} \mathbf{p}_\ell) \Pi_\ell \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell) dx.
\end{aligned}$$

Since $\operatorname{div} \mathbf{q}_\ell = \Pi_\ell \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell)$, the aforementioned identity plus the LSFEM on level $\ell$ with test function $\mathbf{q}_\ell$ shows for any $v_\ell \in V(\mathcal{T}_\ell)$ that

$$\begin{aligned}
&\|\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \nabla(u_{\ell+m} - u_\ell)\|^2 + \|\operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell)\|^2 \\
&\qquad = - \int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot (\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \mathbf{q}_\ell - \nabla(u_{\ell+m} - u_\ell - v_\ell)) dx \\
&\qquad\quad - \int_\Omega (f_{\ell+m} - f_\ell) \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell) dx.
\end{aligned}$$

The term $\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \mathbf{q}_\ell$ is split into $\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \mathbf{q}_{\ell+m} + \mathbf{q}^*_{\ell+m} - \mathbf{q}_\ell$ plus $\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}$. The Raviart–Thomas function $\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \mathbf{q}_{\ell+m} + \mathbf{q}^*_{\ell+m} - \mathbf{q}_\ell$ is divergence free and hence (cf. Proposition 2.1) equals $\operatorname{Curl} \beta_{\ell+m}$ for some $\beta_{\ell+m} \in C(\Omega) \cap P_1(\mathcal{T}_{\ell+m})$. Hence, the preceding identity reads

$$\begin{aligned}
&\|\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \nabla(u_{\ell+m} - u_\ell)\|^2 + \|\operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell)\|^2 \\
&\qquad = \int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot (\nabla(u_{\ell+m} - u_\ell - v_\ell) - \operatorname{Curl} \beta_{\ell+m}) dx \\
&\qquad\quad - \int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot (\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) dx - \int_\Omega (f_{\ell+m} - f_\ell) \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell) dx.
\end{aligned}$$

Simple algebra proves the identity

$$\begin{aligned}
&- \int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot (\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) dx \\
&\qquad = \int_\Omega (\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \nabla(u_{\ell+m} - u_\ell)) \cdot (\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) dx \\
&\qquad\quad - \int_\Omega (\mathbf{p}_{\ell+m} - \nabla u_{\ell+m}) \cdot (\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) dx.
\end{aligned}$$

The LSFEM on level $\ell+m$ shows that

$$\begin{aligned}
&-\int_\Omega (\mathbf{p}_{\ell+m} - \nabla u_{\ell+m}) \cdot (\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) dx \\
&\qquad = \int_\Omega (f_{\ell+m} + \operatorname{div} \mathbf{p}_{\ell+m}) \operatorname{div}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) dx \\
&\qquad = \int_\Omega (f_{\ell+m} + \operatorname{div} \mathbf{p}_{\ell+m})(1-\Pi_\ell) \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell) dx \\
&\qquad = \int_\Omega (f_{\ell+m} - f_\ell) \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell) dx + \|(1-\Pi_\ell) \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell)\|^2.
\end{aligned}$$

The combination of the last three identities completes the proof. ∎

The difference $\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}$ is super-close.

Lemma 5.3. *It holds that*

$$\|\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}\| \lesssim \|h_\ell (1-\Pi_\ell) \operatorname{div} \mathbf{p}_{\ell+m}\|.$$

*Proof.* The discrete Helmholtz decomposition (2.8) asserts that

$$\Pi_{\ell+m}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) = \nabla_{NC} \tilde{\alpha}_{\ell+m} + \operatorname{Curl} \tilde{\beta}_{\ell+m}$$

for some $\tilde{\alpha}_{\ell+m} \in V(\mathcal{T}_{\ell+m})$ and some $\tilde{\beta}_{\ell+m} \in C(\Omega) \cap P_1(\mathcal{T}_{\ell+m})$. Since $\operatorname{Curl} \tilde{\beta}_{\ell+m} \in RT_0(\mathcal{T}_{\ell+m})$ is divergence free, it is $L^2$ orthogonal onto $\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}$ and hence on $\Pi_{\ell+m}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m})$. It follows $\tilde{\beta}_{\ell+m} = 0$ in agreement with Proposition 2.1. Hence,

$$\|\Pi_{\ell+m}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m})\|^2 = \int_\Omega (\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) \cdot \nabla_{NC} \tilde{\alpha}_{\ell+m} \, dx.$$

Since $\operatorname{div}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}) = (1-\Pi_\ell) \operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell) = (1-\Pi_\ell) \operatorname{div} \mathbf{p}_{\ell+m}$, an integration by parts proves

$$\|\Pi_{\ell+m}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m})\|^2 = -\int_\Omega ((1-\Pi_\ell)\tilde{\alpha}_{\ell+m})(1-\Pi_\ell) \operatorname{div} \mathbf{p}_{\ell+m} \, dx.$$

Some piecewise Poincaré inequality leads to

$$\|\Pi_{\ell+m}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m})\|^2 \leq |||\tilde{\alpha}_{\ell+m}||| \, ||h_\ell (1-\Pi_\ell) \operatorname{div} \mathbf{p}_{\ell+m}||.$$

Since $|||\tilde{\alpha}_{\ell+m}||| = \|\Pi_{\ell+m}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m})\|$, this proves

$$\|\Pi_{\ell+m}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m})\| \leq ||h_\ell (1-\Pi_\ell) \operatorname{div} \mathbf{p}_{\ell+m}||.$$

The identity (2.7) with the constant divergence of $\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m}$ on a fine triangle $T \in \mathcal{T}_{\ell+m}$ implies

$$\|(1-\Pi_{\ell+m})(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m})\| \leq \|h_{\ell+m} \operatorname{div}(\mathbf{q}_{\ell+m} - \mathbf{q}^*_{\ell+m})\| \leq \|h_\ell (1-\Pi_\ell) \operatorname{div} \mathbf{p}_{\ell+m}\|.$$

This concludes the proof. ∎

The subsequent lemma utilizes $\Xi_1 = 1$ in the definition of $\eta_\ell$.

Lemma 5.4. *There exists some $v_\ell \in V(\mathcal{T}_\ell)$ which satisfies*

$$\int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot \nabla(u_{\ell+m} - u_\ell - v_\ell) dx \lesssim \eta_\ell(\mathcal{T}_\ell \backslash \mathcal{T}_{\ell+m}) |||u_{\ell+m} - u_\ell|||.$$

*Proof.* The proof utilizes the Scott–Zhang interpolation [20] $v_\ell$ of $u_{\ell+m} - u_\ell$ and follows the discrete reliability proof from the AFEM literature [19]. In this quasi interpolation, for each node $z \in \mathcal{M}_\ell(\Omega)$, one may select some edge $E$ in $\mathcal{E}_\ell(z) \cap \mathcal{E}_{\ell+m}(z)$ whenever possible. The standard approximation and stability properties of quasi interpolations lead after an integration by parts for $w := u_{\ell+m} - u_\ell - v_\ell$ to

$$\begin{aligned}
&\int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot \nabla w dx \\
&\quad = -\int_\Omega w \operatorname{div} \mathbf{p}_\ell dx - \sum_{E \in \mathcal{E}_\ell(\Omega)} \int_E w \left[\partial u_\ell / \partial \nu_E\right]_E ds \\
&\quad \lesssim (\|h_\ell \operatorname{div} \mathbf{p}_\ell\| + \eta_\ell)\, |||u_{\ell+m} - u_\ell|||.
\end{aligned}$$

The point is that $w$ vanishes on any $T \in \mathcal{T}_\ell \cap \mathcal{T}_{\ell+m}$. This allows the reduction of the contributions in the a posteriori error estimators to the compliment $\mathcal{T}_\ell \backslash \mathcal{T}_{\ell+m}$ of $\mathcal{T}_\ell \cap \mathcal{T}_{\ell+m}$ and concludes the proof. ∎

The subsequent lemma utilizes $\Xi_2 = 1$ in the definition of $\eta_\ell$.

Lemma 5.5. *The function $\beta_{\ell+m}$ from Lemma 5.2 satisfies*

$$-\int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot \operatorname{Curl} \beta_{\ell+m} dx \lesssim \eta_\ell(\mathcal{T}_\ell \backslash \mathcal{T}_{\ell+m}) \|\mathbf{p}_{\ell+m} - \mathbf{p}_\ell\|_{H(\mathrm{div})}.$$

*Proof.* Since $\mathbf{p}_\ell - \nabla u_\ell$ is $L^2$ orthogonal to $\operatorname{Curl} \beta_\ell$ for the Scott–Zhang quasi interpolation $\beta_\ell \in C(\Omega) \cap P_1(\mathcal{T}_\ell)$ of $\beta_{\ell+m}$, standard arguments (piecewise integration by parts and trace inequalities for some edge contributions) show

$$\begin{aligned}
&-\int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot \operatorname{Curl} \beta_{\ell+m}\, dx \\
&\quad = \int_\Omega (\mathbf{p}_\ell - \nabla u_\ell) \cdot \operatorname{Curl}(\beta_{\ell+m} - \beta_\ell) dx \\
&\quad \lesssim \Big( \sum_{E \in \mathcal{E}_\ell \backslash \mathcal{E}_{\ell+m}} |E|\, \|[\mathbf{p}_\ell] \cdot \tau_E\|_{L^2(E)}^2 \Big)^{1/2} |||\beta_{\ell+m}|||.
\end{aligned}$$

The design of $\beta_{\ell+m}$ in the proof of Lemma 5.2 shows

$$\begin{aligned}
|||\beta_{\ell+m}||| &= \|\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \mathbf{q}_{\ell+m} + \mathbf{q}_{\ell+m}^* - \mathbf{q}_\ell\| \\
&\le \|\mathbf{p}_{\ell+m} - \mathbf{p}_\ell\| + \|\mathbf{q}_\ell\| + \|\mathbf{q}_{\ell+m} - \mathbf{q}_{\ell+m}^*\|.
\end{aligned}$$

All three terms are bounded by $\lesssim \|\mathbf{p}_{\ell+m} - \mathbf{p}_\ell\|_{H(\mathrm{div})}$. This is obvious for the first term, follows from the definition of $\mathbf{q}_\ell$ and the stability of the mixed finite element schemes for the second, and for the third term follows from Lemma 5.3. This and $\Xi_2 = 1$ conclude the proof. ∎

*Proof of Theorem* 5.1. The combination of Lemmas 5.2–5.5 shows an estimate for

$$\begin{aligned}
&\|(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell, u_{\ell+m} - u_\ell)\|_{H(\mathrm{div}) \times H_0^1(\Omega)}^2 \approx LS(0; \mathbf{p}_{\ell+m} - \mathbf{p}_\ell, u_{\ell+m} - u_\ell) \\
&= \|\mathbf{p}_{\ell+m} - \mathbf{p}_\ell - \nabla(u_{\ell+m} - u_\ell)\|^2 + \|\operatorname{div}(\mathbf{p}_{\ell+m} - \mathbf{p}_\ell)\|^2.
\end{aligned}$$

This and some standard rearrangements conclude the proof. ∎

**6. Quasi-optimal convergence.** The initial triangulation $\mathcal{T}_0$ specifies the geometry in the PMP with right-hand side $f \in L^2(\Omega)$ and exact solution $u \in H^1_0(\Omega)$. All possible markings by the NVB define the set of all admissible triangulations $\mathbb{T}$ which includes $\mathcal{T}_0$ and all its possible shape-regular refinements. Given a natural number $N$, the subset $\mathbb{T}(N)$ consists of all admissible triangulations $\mathcal{T}$ with $|\mathcal{T}| \leq |\mathcal{T}_0| + N$. The least squares functional is minimal for an optimal (but unknown) triangulation with minimal value

$$E(u, f, N) := \min_{\mathcal{T}\in\mathbb{T}(N)} \min_{(\mathbf{p}_\mathcal{T}, u_\mathcal{T})\in RT_0(\mathcal{T})\times V(\mathcal{T})} LS(f; \mathbf{p}_\mathcal{T}, u_\mathcal{T}).$$

It appears too costly to compute such an optimal mesh with corresponding optimal LSFEM solution, but it serves as a reference value for the performance of the computed ones.

This paper proves that ALSFEM computes a sequence $(\mathbf{p}_\ell, u_\ell)_\ell$ of approximations which are not much less accurate than the fictitious optimal values in $(E(u, f, N))_N$. The convergence rates are compared in terms of nonlinear approximation classes as in [4, 22, 15]. Given any $0 < \sigma < \infty$, the set $\mathcal{A}_\sigma$ consists of all pairs $(u, f) \in H^1_0(\Omega) \times L^2(\Omega)$ such that

$$|(u, f)|^2_{\mathcal{A}_\sigma} := \sup_{N\in\mathbb{N}} N^{2\sigma}\, E(u, f, N) < \infty. \tag{6.1}$$

The following theorem states that the convergence rate of the ALSFEM is optimal for sufficiently small parameters $\Theta_0$ and $\kappa_0$ but for an arbitrarily coarse initial triangulation $\mathcal{T}_0$.

THEOREM 6.1. *There exist some maximal bulk parameter* $0 < \Theta_0 < 1$ *and maximal separation parameter* $0 < \kappa_0 < \infty$ *which depend exclusively on* $\mathcal{T}_0$ *such that for all* $0 < \Theta \leq \Theta_0$*, for all* $0 < \varrho < 1$*, for all* $0 < \sigma < \infty$*, and for all* $0 < \kappa \leq \kappa_0$*, there exists some constant* $C_{qo}$ *which satisfies the following. The output* $(\mathbf{p}_\ell, u_\ell)_\ell$ *of ALSFEM with* $(\Xi_1, \Xi_2) = (1, 1)$ *in the PMP with* $(u, f) \in \mathcal{A}_\sigma$ *satisfies*

$$\sup_{\ell\in\mathbb{N}} (|\mathcal{T}_\ell| - |\mathcal{T}_0|)^{2\sigma}\, LS(f; \mathbf{p}_\ell, u_\ell) \leq C_{qo} |(u, f)|^2_{\mathcal{A}_\sigma}. \tag{6.2}$$

Some comments are in order before the proof of Theorem 6.1 concludes this section.

*Remark* 6.2 (quasi optimality). Notice that the definition of the seminorm $|(u, f)|_{\mathcal{A}_\sigma}$ implies $1 \lesssim C_{qo}$ in (6.2) for *any* choice of $(\mathbf{p}_\ell, u_\ell)$ with respect to *any* sequence of nested triangulations $(\mathcal{T}_\ell)$. This motivates the name *quasi-optimal convergence* for Theorem 6.1 which guarantees *optimal convergence rates* and is analogous to optimality results in the FEM literature [4, 22, 15, 14]; it is a key result for *optimal computational complexity*.

*Remark* 6.3 (separate marking necessary). One important difference between the adaptive algorithm in this paper and other existing adaptive finite element schemes is that the data resolution term $\|f - \Pi f\|$ is not an oscillation nor can it be reduced in one-level refinements. Recall that $\|f - \Pi f\|$ is enforced by the $H(\text{div})$ norm which leads to the contribution $\|f - \Pi f\|$ in the least squares functional. This is why ALSFEM utilizes a separate marking strategy with some optimal scheme [5] as in [14].

*Remark* 6.4 (nonlinear approximation class). The notion of a nonlinear approximation class has been introduced in [4] to the optimality analysis of adaptive FEMs. The version here involves the least squares functional and hence seemingly includes some procedural error related to LSFEM. The medius analysis of [13] shows that, in fact, $(u,f) \in \mathcal{A}_\sigma$ is equivalent (that means the sets coincide and the norms are equivalent) to $(u,f) \in \tilde{\mathcal{A}}_\sigma$, where $\tilde{\mathcal{A}}_\sigma$ is defined as above through the seminorm (6.1) with $E(u,f,N)$ replaced by

$$\tilde{E}(u,f,N) := \min_{\mathcal{T}\in\mathbb{T}(N)} \big( \min_{f_\mathcal{T}\in P_0(\mathcal{T})} \|f - f_\mathcal{T}\|^2 + \min_{u_\mathcal{T}\in V(\mathcal{T})} |||u - u_\mathcal{T}|||^2 \big).$$

The resulting nonlinear approximation class $\tilde{\mathcal{A}}_\sigma = \mathcal{A}_\sigma$ clearly deserves this name and is related to some Besov space [4]. The aforementioned equivalence from [13] allows also the corresponding modification in the assertion of Theorem 6.1.

*Proof of Theorem* 6.1. The proof follows the outline of the FEM literature while the separate marking in ALSFEM causes some extra difficulties which are overcome with arguments from [14] for sufficiently small $\Theta$ and $\kappa$. It is a novel argument to allow the term $\|(1-\Pi)\operatorname{div}\mathbf{p}_{\ell+m}\|$ on the right-hand side in the discrete reliability.

Recall $\xi_\ell^2 := LS(f;\mathbf{p}_\ell,u_\ell) + \|f - f_\ell\|^2 + \Lambda_{\mathrm{red}}\eta_\ell^2$ from Theorem 4.1 for any $\ell \in \mathbb{N}_0$ and observe that the contraction therein does not leave anything to prove in the pathological situation $\xi_0 = 0$. Hence we may and will assume $\xi_0 > 0$ and may consider any parameter $0 < \tau < |(u,f)|_{\mathcal{A}\sigma}/\xi_0$. Given any $\ell \in \mathbb{N}$, let $N(\ell)$ denote the minimal natural number with

$$|(u,f)|_{\mathcal{A}_\sigma} \le \varepsilon(\ell) N(\ell)^\sigma \quad \text{for } \varepsilon(\ell) := \tau\xi_\ell.$$

*Step one of the proof* consists of the demonstration of

$$N(\ell) \le 2|(u,f)|_{\mathcal{A}_\sigma}^{1/\sigma}\varepsilon(\ell)^{-1/\sigma}. \tag{6.3}$$

For $N(\ell) = 1$, Theorem 4.1 leads to the contradiction

$$|(u,f)|_{\mathcal{A}_\sigma} \le \varepsilon(\ell) = \tau\xi_\ell \le \tau\xi_0 < |(u,f)|_{\mathcal{A}_\sigma}.$$

Consequently, $2 \le N(\ell)$. This and the minimality of $N(\ell)$ show

$$\varepsilon(\ell)(N(\ell)/2)^\sigma \le \varepsilon(\ell)(N(\ell)-1)^\sigma < |(u,f)|_{\mathcal{A}_\sigma}. \qquad \square$$

*Step two of the proof* is the design of a reference triangulation $\widehat{\mathcal{T}}_\ell$ which is some admissible refinement of $\mathcal{T}_\ell$ (over possibly many levels) with associated LSFEM solutions $(\widehat{\mathbf{p}}_\ell, \widehat{u}_\ell)$ and

$$LS(f;\widehat{\mathbf{p}}_\ell,\widehat{u}_\ell) \le \varepsilon(\ell)^2 = \tau^2\xi_\ell^2; \tag{6.4}$$

$$|\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell| \lesssim |(u,f)|_{\mathcal{A}_\sigma}^{1/\sigma}\varepsilon(\ell)^{-1/\sigma}. \tag{6.5}$$

Indeed, the definition of $E(u,f,N(\ell))$ implies the existence of an optimal triangulation $\mathcal{T}_\ell^* \in \mathbb{T}(N(\ell))$ such that the corresponding LSFEM solution $(\mathbf{p}_\ell^*, u_\ell^*)$ satisfies

$$E(u,f,N(\ell)) = LS(f;\mathbf{p}_\ell^*,u_\ell^*).$$

This and the definition of $|(u,f)|_{\mathcal{A}_\sigma}$ as a supremum plus the design of $N(\ell)$ lead to

$$LS(f;\mathbf{p}_\ell^*,u_\ell^*) \le N(\ell)^{-2\sigma}|(u,f)|_{\mathcal{A}_\sigma}^2 \le \varepsilon(\ell)^2 = \tau^2\xi_\ell^2.$$

This proves (6.4) for all refinements $\widehat{\mathcal{T}}_\ell$ of $\mathcal{T}_\ell$ such as the overlay $\widehat{\mathcal{T}}_\ell := \mathcal{T}_\ell \otimes \mathcal{T}_\ell^*$ of $\mathcal{T}_\ell$ and $\mathcal{T}_\ell^*$ defined as their smallest common refinement. It is known [22, 15] that the overlay $\widehat{\mathcal{T}}_\ell$ is a regular triangulation with

$$|\widehat{\mathcal{T}}_\ell| + |\mathcal{T}_0| \le |\mathcal{T}_\ell| + |\mathcal{T}_\ell^*|.$$

It is also known (and much more elementary to prove) that

$$|\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell| \le |\widehat{\mathcal{T}}_\ell| - |\mathcal{T}_\ell|.$$

The combination of the previous inequalities implies

$$|\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell| \le |\mathcal{T}_\ell^*| - |\mathcal{T}_0| \le N(\ell).$$

This and (6.3) conclude step two of the proof. ∎

*Step three of the proof* verifies for sufficiently small $\Theta_0$ and $\kappa_0$ with an appropriate choice of $\tau \approx 1$ that Case A on level $\ell$ of ALSFEM leads to

$$|\mathcal{M}_\ell|^{2\sigma} \lesssim |(u,f)|^2_{\mathcal{A}_\sigma} \xi_\ell^{-2}. \tag{6.6}$$

The core of the proof is the discrete reliability of Theorem 5.1 for the triangulation $\mathcal{T}_\ell$ and its admissible refinement $\widehat{\mathcal{T}}_\ell$. The notation of section 7 replaces $\widehat{\mathbf{p}}_\ell, \widehat{u}_\ell, \widehat{f}_\ell, \widehat{\Pi}_\ell$ by $\mathbf{p}_{\ell+m}, u_{\ell+m}$, $f_{\ell+m}$, $\Pi_{\ell+m}$, respectively, and shows (in the present notation of this proof) that

$$\begin{aligned} &\|(\widehat{\mathbf{p}}_\ell - \mathbf{p}_\ell, \widehat{u}_\ell - u_\ell)\|^2_{H(\operatorname{div})\times H^1_0(\Omega)} \\ &\quad\lesssim \eta_\ell^2(\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell) + \|\widehat{f}_\ell - f_\ell\|^2 + \|(1-\Pi_\ell)\operatorname{div}\widehat{\mathbf{p}}_\ell\|^2. \end{aligned} \tag{6.7}$$

The equivalence (2.5) shows that the left-hand side in (6.7) is equivalent to $LS(0; \widehat{\mathbf{p}}_\ell - \mathbf{p}_\ell, \widehat{u}_\ell - u_\ell)$. Straightforward algebra with the Galerkin orthogonality of the LSFEMs as in (4.5) applies to the latter term and shows that it is equal to $LS(f; \mathbf{p}_\ell, u_\ell) - LS(f; \widehat{\mathbf{p}}_\ell, \widehat{u}_\ell)$.

The triangle inequality and the definition of the least squares functional show that

$$\begin{aligned} 1/2\,\|(1-\Pi_\ell)\operatorname{div}\widehat{\mathbf{p}}_\ell\|^2 &\le \|f - f_\ell\|^2 + \|(1-\Pi_\ell)(f + \operatorname{div}\widehat{\mathbf{p}}_\ell)\|^2 \\ &\le \|f - f_\ell\|^2 + LS(f; \widehat{\mathbf{p}}_\ell, \widehat{u}_\ell). \end{aligned}$$

The combination of those arguments with (6.7) proves with $C_{dRel} \approx 1$ that

$$LS(f; \mathbf{p}_\ell, u_\ell) \le C_{dRel}\left(\eta_\ell^2(\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell) + \|f - f_\ell\|^2 + LS(f; \widehat{\mathbf{p}}_\ell, \widehat{u}_\ell)\right).$$

Recall that Case A and the definition of $\xi_\ell$ imply

$$\|f - f_\ell\|^2 \le \kappa_0 \eta_\ell^2 \le \kappa_0/\Lambda_{\mathrm{red}}\,\xi_\ell^2.$$

The efficiency of Theorem 3.1 and the equivalence with $\xi_\ell$ leads to some $C_{eff} \approx 1$ with

$$\xi_\ell^2 \le C_{eff}\, LS(f; \mathbf{p}_\ell, u_\ell).$$

The combination of the preceding three displayed estimates with (6.4) results in

$$\xi_\ell^2/(C_{eff} C_{dRel}) \le \eta_\ell^2(\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell) + \left(\tau^2 + \kappa_0/\Lambda_{\mathrm{red}}\right)\xi_\ell^2.$$

Some proper choice of $\tau^2 \le 1/(4C_{eff}C_{dRel})$ and $\kappa_0 \le \Lambda_{\mathrm{red}}/(4C_{eff}C_{dRel})$ leads to

$$\Lambda_{\mathrm{red}}\,\eta_\ell^2 \le \xi_\ell^2 \le 2C_{eff}C_{dRel}\,\eta_\ell^2(\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell).$$

The point is that this implies $\Theta\eta_\ell^2 \le \eta_\ell^2(\mathcal{M}_\ell)$ for $\Theta \le \Theta_0 := \Lambda_{\rm red}/(2C_{eff}C_{dRel})$. In other words, the set $\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell$ is in the competition of the marking strategy in Case A of ALSFEM. Among all those subsets, the chosen set $\mathcal{M}_\ell$ has (almost) minimal cardinality $|\mathcal{M}_\ell|$. In comparison with $\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell$,

$$|\mathcal{M}_\ell| \lesssim |\mathcal{T}_\ell \setminus \widehat{\mathcal{T}}_\ell|.$$

This and (6.5) imply (6.6). $\square$

*Step four of the proof* establishes (6.6) in Case B on level $\ell$ of ALSFEM with some set $\mathcal{M}_\ell := \mathcal{M}_\ell^{(0)} \cup \cdots \cup \mathcal{M}_\ell^{(K(\ell))}$ of triangles with the outcome $\mathcal{T}_{\ell+1} := \mathcal{T}_\ell^{(K(\ell)+1)}$ of the successive refinements with NVB. The initialization $\mathcal{T}_\ell^{(0)} := \mathcal{T}_\ell$ and the loop

$$\mathcal{T}_\ell^{(k+1)} := \text{ NVB( in } \mathcal{T}_\ell^{(k)} \text{ with marked set } \mathcal{M}_\ell^{(k)}) \text{ for } k = 0, \ldots, K(\ell)$$

define an algorithm to identify the pairwise disjoint sets of triangles $\mathcal{M}_\ell^{(0)}, \ldots, \mathcal{M}_\ell^{(K(\ell))}$ as in [14, section 3]. The remaining parts in step four of the proof report on arguments from [14] which link the optimality of the thresholding second algorithm from [5] with (6.6). Since $f$ is part of $(u,f) \in \mathcal{A}_\sigma$, the triangulation $\mathcal{T}_{\ell+1} = \mathcal{T}_\ell \otimes \mathcal{T}$ for the output $\mathcal{T}$ of the thresholding second algorithm with given tolerance $\mathrm{Tol} := \varrho^{1/2}||f - f_\ell||$ is almost optimal in the sense that it satisfies

$$||f - f_{\ell+1}|| \le \mathrm{Tol} \quad \text{and} \quad |\mathcal{T}| - |\mathcal{T}_0| \lesssim |(u,f)|_{\mathcal{A}\sigma}^{1/\sigma}\, \mathrm{Tol}^{-1/\sigma}.$$

Theorem 3.1 in [14] furthermore ensures that

$$\sum_{k=0}^{K(\ell)} |\mathcal{M}_\ell^{(k)}| = |\mathcal{M}_\ell| \le |\mathcal{T}| - |\mathcal{T}_0|.$$

Since Case B is characterized by $\kappa\eta_\ell^2 \le ||f - f_\ell||^2$,

$$\xi_\ell^2 \approx \eta_\ell^2 + ||f - f_\ell||^2 \le (1 + 1/\kappa)||f - f_\ell||^2 = (1 + 1/\kappa)/\varrho\, \mathrm{Tol}^2.$$

The combination of the preceding three displayed formulas proves

$$|\mathcal{M}_\ell|^{2\sigma} \lesssim (1 + 1/\kappa)/\varrho\, |(u,f)|_{\mathcal{A}_\sigma}^2 \xi_\ell^{-2}.$$

This shows (6.6) for $\kappa \approx 1 \approx \varrho$. More details can be found in [14] (where oscillations may be substituted by the present data resolution errors). $\square$

*Step five is the end of the proof.* The overhead control of [4] shows that NVB satisfies (with $C_{BDV} \approx 1$ which depends exclusively on $\mathcal{T}_0$)

$$|\mathcal{T}_\ell| - |\mathcal{T}_0| \le C_{BDV} \sum_{k=0}^{\ell-1} |\mathcal{M}_k| \quad \text{for all } \ell = 0, 1, 2, \ldots.$$

This result was intended exclusively for level-oriented refinements where one step divides one triangle at most three times. The design of the intermediate refinements in step four after [14], however, verbatim provides this overhead control also for the separate marking of this paper.

The contraction property of Theorem 4.1 shows that

$$\xi_k^{-2} \le \varrho_{\rm red}^{\ell-k} \xi_\ell^{-2} \quad \text{for all } 0 \le k \le \ell.$$

This and (6.6) prove

$$|\mathcal{M}_k| \lesssim |(u,f)|_{\mathcal{A}_\sigma}^{1/\sigma} \xi_k^{-1/\sigma} \leq |(u,f)|_{\mathcal{A}_\sigma}^{1/\sigma} \varrho_{\mathrm{red}}^{(\ell-k)/(2\sigma)} \xi_\ell^{-1/\sigma}.$$

The combination with the overhead control leads to

$$|\mathcal{T}_\ell| - |\mathcal{T}_0| \lesssim |(u,f)|_{\mathcal{A}_\sigma}^{1/\sigma} \xi_\ell^{-1/\sigma} \sum_{k=0}^{\ell-1} \varrho_{\mathrm{red}}^{(\ell-k)/(2\sigma)}.$$

This, the series $\sum_{m=0}^{\infty} \varrho_{\mathrm{red}}^{m/(2\sigma)} \lesssim 1$ and $LS(f; \mathbf{p}_\ell, u_\ell) \leq \xi_\ell^2$ imply

$$|\mathcal{T}_\ell| - |\mathcal{T}_0| \lesssim |(u,f)|_{\mathcal{A}_\sigma}^{1/\sigma} \xi_\ell^{-1/} \leq |(u,f)|_{\mathcal{A}_\sigma}^{1/\sigma} LS(f; \mathbf{p}_\ell, u_\ell)^{-1/(2\sigma)}. \qquad \square$$

**Acknowledgment.** The authors thank Eunjung Lee of the Department of Computational Science and Engineering of Yonsei University in Seoul, Korea, for valuable discussions.

## REFERENCES

[1] J. H. Adler, T. A. Manteuffel, S. F. McCormick, J. W. Nolting, J. W. Ruge, and L. Tang, *Efficiency based adaptive local refinement for first-order system least-squares formulations*, SIAM J. Sci. Comput., 33 (2011), pp. 1–24.

[2] D. Arnold and F. Brezzi, *Mixed and nonconforming finite element methods: Implementation, postprocessing and error estimates*, RAIRO Model. Math. Anal. Numer., 19 (1985), pp. 7–32.

[3] M. Berndt, T. A. Manteuffel, and S. F. McCormick, *Local error estimates and adaptive refinement for first-order system least squares (FOSLS)*, Electron. Trans. Numer. Anal., 6 (1997), pp. 35–43.

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

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

[6] P. B. Bochev and M. D. Gunzburger, *Least-Squares Finite Element Methods,* Appl. Math. Sci. 166, Springer, New York, 2009.

[7] D. Braess, *Finite Elements*, 3rd ed., Cambridge University Press, Cambridge, UK, 2007.

[8] S. C. Brenner, *Two-level additive Schwarz preconditioners for nonconforming finite element methods*, Math. Comp., 65 (1996), pp. 897–921.

[9] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods,* 3rd ed., Texts in Appl. Math. 15, Springer, New York, 2008.

[10] F. Brezzi and M. Fortin, *Mixed and Hybrid Finite Element Methods,* Springer Ser. Comput. Math. 15, Springer, New York, 1991.

[11] C. Carstensen and R. W. H. Hoppe, *Error reduction and convergence for an adaptive mixed finite element method*, Math. Comp. 75 (2006), pp. 1033–1042.

[12] C. Carstensen, R. W. H. Hoppe, M. Eigel, and C. Löbhard, *A review of unified a posteriori finite element error control*, Numer. Math. Theor. Meth. Appl, 5 (2012), pp. 509–558.

[13] C. Carstensen, D. Peterseim, and M. Schedensack, *Comparison results of finite element methods for the Poisson model problem*, SIAM J. Numer. Anal. 50 (2012), pp. 2803–2823.

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

[15] J. M. Cascon, Ch. 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.

[16] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., 33 (1996), pp. 1106–1124.

[17] V. Girault and P.-A. Raviart, *Finite Element Methods for Navier-Stokes Equations,* Springer Ser. Comput. Math. 5, Springer, New York, 1986.

[18] L. D. Marini, *An inexpensive method for the evaluation of the solution of the lowest order Raviart-Thomas mixed method*, SIAM J. Numer. Anal., 22 (1985), pp. 493–496.

[19] R. H. Nochetto, K. G. Siebert, and A. Veeser, *Theory of adaptive finite element methods: An introduction*, in Multiscale, Nonlinear and Adaptive Approximation, Springer, Berlin, 2009, pp. 409–542.

[20] L. R. Scott and S. Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Math. Comp., 54 (1990), pp. 483–493.

[21] G. Starke, *An adaptive least-squares mixed finite element method for elasto-plasticity*, SIAM J. Numer. Anal., 45(1):371–388 (electronic), 2007.

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

[23] R. Verfürth, *A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Wiley, New York, 1996.