ESAIM: M2AN 53 (2019) 1841–1870
https://doi.org/10.1051/m2an/2019039

**ESAIM: Mathematical Modelling and Numerical Analysis**
www.esaim-m2an.org

# ADAPTIVE UZAWA ALGORITHM FOR THE STOKES EQUATION

Giovanni Di Fratta[1], Thomas Führer[2], Gregor Gantner[1,*] and Dirk Praetorius[1]

**Abstract.** Based on the Uzawa algorithm, we consider an adaptive finite element method for the Stokes system. We prove linear convergence with optimal algebraic rates for the residual estimator (which is equivalent to the total error), if the arising linear systems are solved iteratively, *e.g.*, by PCG. Our analysis avoids the use of discrete efficiency of the estimator. Unlike prior work, our adaptive Uzawa algorithm can thus avoid to discretize the given data and does not rely on an interior node property for the refinement.

**Mathematics Subject Classification.** 65N30, 65N50, 65N15, 41A25.

Received December 20, 2018. Accepted May 24, 2019.

## 1. Introduction

The mathematical analysis of adaptive finite element methods (AFEMs) has significantly increased over the last years. Nowadays, AFEMs are recognized as a powerful and rigorous tool to efficiently solve partial differential equations arising in physics and engineering.

### 1.1. Model problem

In this paper, we focus on an adaptive algorithm for the solution of the steady-state Stokes equations, which after a suitable normalization read

$$\begin{aligned} -\Delta \boldsymbol{u} + \nabla p &= \boldsymbol{f} \quad \text{in } \Omega, \\ \nabla \cdot \boldsymbol{u} &= 0 \quad \text{in } \Omega, \\ \boldsymbol{u} &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{1.1}$$

In the literature, the first equation is referred to as *momentum equation*, the second as *mass equation*, and the third as *no-slip boundary condition*. Here, $\Omega \subset \mathbb{R}^d$ with $d \in \{2,3\}$ is a bounded polygonal resp. polyhedral Lipschitz domain. Given the body force $\boldsymbol{f}$, one seeks the velocity field $\boldsymbol{u}$ of an incompressible fluid and the

*Keywords and phrases.* Adaptive finite element method, optimal convergence, Uzawa algorithm, Stokes equation.

[1] TU Wien, Institute for Analysis and Scientific Computing, Wiedner Hauptstr. 8–10/E101/4, 1040 Wien, Austria.
[2] Pontificia Universidad Católica de Chile, Facultad de Matemáticas, Vickuña Mackenna, 4860 Santiago, Chile.
*Corresponding author: gregor.gantner@asc.tuwien.ac.at

 

associated pressure $p$. With

$$\mathbb{V} := H_0^1(\Omega)^d, \quad \mathbb{P} := \left\{ q \in L^2(\Omega) : \int_\Omega q \, \mathrm{d}x = 0 \right\}, \tag{1.2}$$

it is well-known that the Stokes problem admits a unique solution $(\boldsymbol{u}, p) \in \mathbb{V} \times \mathbb{P}$, where $p$ can be characterized as the unique null average solution of the elliptic Schur complement equation; see, *e.g.*, [8]. More precisely, the pressure solves the elliptic equation

$$Sp = \nabla \cdot \Delta^{-1} \boldsymbol{f} \quad \text{with the Schur complement operator} \quad S := \nabla \cdot \Delta^{-1} \nabla : \mathbb{P} \to \mathbb{P}. \tag{1.3}$$

The latter equation can be reformulated as a fixpoint problem for the operator

$$N_\alpha : \mathbb{P} \to \mathbb{P}, \quad q \mapsto (I - \alpha S) q + \alpha \nabla \cdot \Delta^{-1} \boldsymbol{f}. \tag{1.4}$$

Note that $S$ is self-adjoint. Since the norm of self-adjoint operators coincides with their spectral radius and $S$ has positive spectrum, one has that $\|I - \alpha S\| < 1$ whenever $|1 - \alpha \|S\|| < 1$. It follows that $N_\alpha$ is a contraction for $0 < \alpha < 2 \|S\|^{-1}$; see Appendix A. Moreover, elementary calculation proves that $\|S\| \leq 1$. Hence, for all $0 < \alpha < 2$ and any initial guess $p^0 \in \mathbb{P}$, the generalized Richardson iteration

$$p^{j+1} := N_\alpha p^j = (I - \alpha S) p^j + \alpha \nabla \cdot \Delta^{-1} \boldsymbol{f} \tag{1.5}$$

converges to the exact pressure of the Stokes problem. It follows that $\boldsymbol{u} = \lim_{j \to \infty} \boldsymbol{u}[p^j]$ in $\mathbb{V}$ with $\boldsymbol{u}[p^j] := -\Delta^{-1}(\boldsymbol{f} - \nabla p^j)$, so that, at the continuous level, the full iterative process can be expressed in the form

$$\begin{aligned} \boldsymbol{u}[p^j] &= -\Delta^{-1}(\boldsymbol{f} - \nabla p^j), \\ p^{j+1} &= p^j - \alpha \nabla \cdot \boldsymbol{u}[p^j]. \end{aligned} \tag{1.6}$$

In the spirit of Kondratyuk and Stevenson [24], the iterative scheme (1.6), usually referred to as *Uzawa algorithm* for the Stokes problem, is the starting point of our AFEM analysis.

### 1.2. State of the art

Although AFEMs for the analysis of mixed variational problems issuing from fluid dynamics have a long history in the engineering and physics literature, only in the last decade, Dahlke *et al.* [12] introduced an adaptive wavelet method based on the Uzawa algorithm for solving the Stokes problem. In [1], the adaptive wavelet method is replaced by an AFEM. Their numerical experiments suggested that the latter algorithm leads to optimal algebraic convergence rates. Indeed, by addition of a mesh-coarsening step to this method, Kondratyuk [23] proved optimal convergence rates. Later, in [24], the original algorithm of Bänsch *et al.* [1] was modified by adding an additional loop, which separately controls the triangulations on which the pressure is discretized.

We also note that for a standard conforming AFEM with Taylor–Hood elements, the first proofs of plain convergence were presented in [25, 26]. The work [17] gives an optimality proof under the assumption that some *general quasi-orthogonality* is satisfied. This assumption has only recently been verified in [14]. For adaptive nonconforming finite element methods, convergence and optimal rates have been investigated and proved in [2, 9, 21].

### 1.3. Adaptive Uzawa FEM

In this work, we further investigate the algorithm of Kondratyuk and Stevenson [24], which is described in the following: Given a possibly non-conforming partition $\mathcal{P}_i$ of $\Omega$, we denote by $p_i \in \mathbb{P}_i$ the best approximation to $p$, with respect to the $S$-induced energy norm $\|\cdot\|_{\mathbb{P}}$, from the corresponding discrete space $\mathbb{P}_i \subset \mathbb{P}$ of piecewise polynomials of degree $m-1$ with vanishing integral mean. With the corresponding velocity $\boldsymbol{u}_i := \boldsymbol{u}[p_i]$ defined analogously to (1.6) and the $L^2$-orthogonal projection $\Pi_i : L^2(\Omega) \to \mathbb{P}_i$, one can show that $(\boldsymbol{u}_i, p_i)$ is the unique solution of the reduced problem

$$\begin{aligned} -\Delta \boldsymbol{u}_i + \nabla p_i &= \boldsymbol{f} \quad \text{in } \Omega, \\ \Pi_i \nabla \cdot \boldsymbol{u}_i &= 0 \quad \text{in } \Omega, \\ \boldsymbol{u}_i &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{1.7}$$

In general, the velocity $\boldsymbol{u}_i$ is not discrete, and hence this problem can still not be solved in practice. It is thus approximated by some FEM approximation $\boldsymbol{U}_{ijk} \in \mathbb{V}_{ijk}$ (the use of three indices being motivated by the structure of the algorithm based on three different iterations) *via* a standard adaptive algorithm of the form

$$\boxed{\text{SOLVE}} \longrightarrow \boxed{\text{ESTIMATE}} \longrightarrow \boxed{\text{MARK}} \longrightarrow \boxed{\text{REFINE}}$$

for the vector-valued Poisson problem steered by a weighted-residual error estimator $\eta_{ijk}$. Here, $\mathbb{V}_{ijk} \subset \mathbb{V}$ denotes the space of all continuous piecewise polynomials on some conforming triangulation $\mathcal{T}_{ijk}$, which is a refinement of the possibly non-conforming $\mathcal{P}_i$. In the next loop, we apply a discretized version of the Uzawa algorithm (1.6) to obtain an approximation $P_{ij} \in \mathbb{P}_i$ of $p_i$. Here, the update reads $P_{i(j+1)} = P_{ij} - \Pi_i \nabla \cdot \boldsymbol{U}_{ijk}$. The last loop employs an adaptive tree approximation algorithm from [5] to obtain a better approximation $p_{i+1} \in \mathbb{P}_{i+1}$ of $p$ on a refinement $\mathcal{P}_{i+1}$ of the partition $\mathcal{P}_i$ such that $\vartheta \, \|\nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega \le \|\Pi_{i+1} \nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega$ for some bulk parameter $0 < \vartheta < 1$. We will see in Section 3.1 that $\|\nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega$ is related to $\|p - p_i\|_{\mathbb{P}}$ and $\|\Pi_{i+1} \nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega$ to $\|p_{i+1} - p_i\|_{\mathbb{P}}$. In contrast to [24], in [1] the latter loop was not present, since the same triangulation for the discretization of the pressure and the velocity, *i.e.*, $\mathcal{P}_i = \mathcal{T}_{ijk}$ was used.

Under the assumption that the right-hand side $\boldsymbol{f}$ is a piecewise polynomial of degree $m-1$, Kondratyuk and Stevenson [24] proved that the approximations $\boldsymbol{U}_{ijk}$ and $P_{ij}$ converge with optimal algebraic rate to the exact solutions $\boldsymbol{u}$ and $p$. To generalize this result for arbitrary $\boldsymbol{f}$, as in the seminal work [27], which proves optimal convergence of a standard AFEM for the Poisson problem, Kondratyuk and Stevenson [24] apply an additional loop to resolve the data oscillations appropriately. However, Kondratyuk and Stevenson [24] only apply the proof of this generalization. Moreover, as in the seminal work [27], the analysis of Kondratyuk and Stevenson [24] hinges on the following interior node property: Given marked elements $\mathcal{M}_{ijk}$ of the current velocity triangulation $\mathcal{T}_{ijk}$, the next velocity triangulation $\mathcal{T}_{ij(k+1)}$ is the coarsest refinement *via* newest vertex bisection (NVB) such that all $T \in \mathcal{M}_{ijk}$ and all $T' \in \mathcal{T}_{ijk}$, which share a common $(n-1)$-dimensional hyperface, contain a vertex of $\mathcal{T}_{ij(k+1)}$ in their interior. In particular for $n = 3$, this property is highly demanding; see, *e.g.*, the 3D refinement pattern in [13].

## 1.4. Contributions of present work

In the spirit of Cascon *et al.* [11], who generalize [27], we prove that the algorithm of Kondratyuk and Stevenson [24] without the data approximation loop leads to convergence of the combined error estimator $\eta_{ijk} + \|\nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega$ (which is equivalent to the error plus data oscillations) at optimal algebraic rate with respect to the number of elements $\#\mathcal{T}_{ijk}$ if one uses standard newest vertex bisection (without interior node property) for the velocity triangulations. We also prove that the combined estimator sequence converges linearly in each step, *i.e.*, it essentially contracts uniformly in each step. Moreover, our algorithm allows for the inexact solution of the arising linear systems for the discrete velocities by iterative solvers like PCG.

On a conceptual level, our proofs show that even for general saddle point problems and adaptive strategies based on Richardson-type iterations, the analysis of rate optimal adaptivity can be conducted without exploiting discrete efficiency estimates of the corresponding *a posteriori* error estimators.

## 1.5. Outline

The paper is organized as follows: Section 2 rewrites the Stokes problem in its variational form, introduces newest vertex bisection, and fixes some notation for the discrete ansatz spaces. In Section 3, we consider the reduced Stokes problem and the corresponding Galerkin approximations, recall some well-known results on *a posteriori* error estimation, and introduce the tree approximation Algorithm 3.6 from [5] as well as our variant of the adaptive Uzawa Algorithm 3.6 from [24]. In Section 4, we state and prove linear convergence of the resulting combined error estimator in each step of the algorithm (Thm. 4.1). To this end, we show that each increase of either $i, j$, or $k$ essentially leads to a uniform contraction of the combined error estimator. Finally, Section 5 is dedicated to the main Theorem 5.3 on optimal convergence rates for the combined error estimator and its proof. As an auxiliary result of general interest, Lemma 5.1 proves that the two different definitions of approximation classes from the literature, which are either based on the accuracy $\varepsilon > 0$ (see, *e.g.*, [24, 28]) or the number of elements $N$ (see, *e.g.*, [10, 11]), are exactly the same.

While all constants in statements of theorems, lemmas, etc. are explicitly given, we abbreviate the notation in proofs: For scalar terms $A$ and $B$, we write $A \lesssim B$ to abbreviate $A \le C\, B$, where the generic constant $C > 0$ is clear from the context. Moreover, $A \simeq B$ abbreviates $A \lesssim B \lesssim A$.

## 2. Preliminaries

### 2.1. Continuous Stokes problem

The vector-valued velocity fields $\boldsymbol{v} \in \mathbb{V}$ are denoted in boldface, the scalar pressures $q \in \mathbb{P}$ in normal font. Let $\langle \cdot\,, \cdot \rangle_\Omega$ be the $L^2(\Omega)$ scalar product with the corresponding $L^2(\Omega)$ norm $\|\cdot\|_\Omega$. With the bilinear forms $a : \mathbb{V} \times \mathbb{V} \to \mathbb{R}$ and $b : \mathbb{V} \times \mathbb{P} \to \mathbb{R}$ defined by

$$a(\boldsymbol{w}, \boldsymbol{v}) := \langle \nabla \boldsymbol{w}\,,\, \nabla \boldsymbol{v} \rangle_\Omega \quad \text{and} \quad b(\boldsymbol{v}, q) := -\langle \nabla \cdot \boldsymbol{v}\,,\, q \rangle_\Omega,$$

the mixed variational formulation of the Stokes problem (1.1) reads as follows: Given $\boldsymbol{f} \in L^2(\Omega)^d$, let $(\boldsymbol{u}, p) \in \mathbb{V} \times \mathbb{P}$ be the unique solution to

$$\begin{aligned} a(\boldsymbol{u}, \boldsymbol{v}) + b(\boldsymbol{v}, p) &= \langle \boldsymbol{f}\,,\, \boldsymbol{v} \rangle_\Omega && \text{for all } \boldsymbol{v} \in \mathbb{V}, \\ b(\boldsymbol{u}, q) &= 0 && \text{for all } q \in \mathbb{P}. \end{aligned} \tag{2.1}$$

On the velocity space $\mathbb{V}$, we consider the $a(\cdot,\cdot)$-induced energy norm $\|\boldsymbol{v}\|_{\mathbb{V}} := a(\boldsymbol{v}, \boldsymbol{v})^{1/2} = \|\nabla \boldsymbol{v}\|_\Omega \simeq \|\boldsymbol{v}\|_{H^1(\Omega)}$. We note that $\nabla \cdot \boldsymbol{v} \in \mathbb{P}$ for all $\boldsymbol{v} \in \mathbb{V}$ and

$$\|\nabla \cdot \boldsymbol{v}\|_\Omega \le \|\nabla \boldsymbol{v}\|_\Omega = \|\boldsymbol{v}\|_{\mathbb{V}} \quad \text{for all } \boldsymbol{v} \in \mathbb{V}, \tag{2.2}$$

which follows from integration by parts; see Appendix B.

Define the operators $A : \mathbb{V} \to \mathbb{V}^*$, $B : \mathbb{V} \to \mathbb{P}^*$, and $B' : \mathbb{P} \to \mathbb{V}^*$ by

$$A\boldsymbol{v} := a(\boldsymbol{v}, \cdot), \quad B\boldsymbol{v} := b(\boldsymbol{v}, \cdot), \quad B'q := b(\cdot, q).$$

Then, the Schur complement operator $S := BA^{-1}B' : \mathbb{P} \to \mathbb{P}^* \sim \mathbb{P}$ is bounded, symmetric, and elliptic; see Lemma 2.2 of [24]. Thus, it holds that $\|q\|_{\mathbb{P}} := \langle Sq\,,\, q \rangle_\Omega^{1/2} \simeq \|q\|_\Omega$ on $\mathbb{P}$. More precisely, there exists a constant $C_{\rm div} \ge 1$, which depends only on $\Omega$, such that

$$C_{\rm div}^{-1} \|q\|_\Omega \le \|q\|_{\mathbb{P}} \le \|q\|_\Omega \quad \text{for all } q \in \mathbb{P}. \tag{2.3}$$

Here, the upper bound with constant 1 follows from $\|S\| \le 1$, which itself follows from (2.2).

### 2.2. Partitions, triangulations, and newest vertex bisection (NVB)

Throughout, $\mathcal{P}$ is a finite (possibly non-conforming) partition of $\Omega$ into compact (non-degenerate) simplices, which is used to discretize $\mathbb{P}$, while $\mathcal{T}$ is a finite (conforming) triangulation of $\Omega$ into compact (non-degenerate) simplices, which is used to discretize $\mathbb{V}$. Throughout, we use NVB refinement; see, *e.g.*, [22, 28] for the precise mesh-refinement rules.

We write $\mathcal{P}' := \texttt{bisect}(\mathcal{P}, \mathcal{M})$ for the partition obtained by *one* bisection of all marked elements $\mathcal{M} \subseteq \mathcal{P}$, *i.e.*, $\mathcal{M} = \mathcal{P} \backslash \mathcal{P}'$ and $\#\mathcal{M} = \#\mathcal{P}' - \#\mathcal{P}$. We write $\mathcal{P}' \in \mathbb{T}^{\mathrm{nc}}(\mathcal{P})$, if there exists $J \in \mathbb{N}_0$ and partitions $\mathcal{P}_j$ and $\mathcal{M}_j \subseteq \mathcal{P}_j$ for all $j = 0, \ldots, J$, such that

$$\mathcal{P} = \mathcal{P}_0, \quad \mathcal{P}_j = \texttt{bisect}(\mathcal{P}_{j-1}, \mathcal{M}_{j-1}) \text{ for all } j = 1, \ldots, J, \quad \text{and} \quad \mathcal{P}' = \mathcal{P}_J.$$

We write $\mathcal{T}' := \texttt{refine}(\mathcal{T}, \mathcal{M})$ for the coarsest conforming triangulation such that (at least) all marked elements $\mathcal{M} \subseteq \mathcal{T}$ have been bisected, *i.e.*, $\mathcal{M} \subseteq \mathcal{T} \backslash \mathcal{T}'$. We write $\mathcal{T}' \in \mathbb{T}^{\mathrm{c}}(\mathcal{T})$, if there exists $J \in \mathbb{N}_0$ and triangulations $\mathcal{T}_j$ and $\mathcal{M}_j \subseteq \mathcal{T}_j$ for all $j = 0, \ldots, J$, such that

$$\mathcal{T} = \mathcal{T}_0, \quad \mathcal{T}_j = \texttt{refine}(\mathcal{T}_{j-1}, \mathcal{M}_{j-1}) \text{ for all } j = 1, \ldots, J, \quad \text{and} \quad \mathcal{T}' = \mathcal{T}_J.$$

Let $\mathcal{T}_{\mathrm{init}}$ be a given initial (conforming) triangulation of $\Omega$. We define the sets

$$\mathbb{T}^{\mathrm{nc}} := \mathbb{T}^{\mathrm{nc}}(\mathcal{T}_{\mathrm{init}}) \quad \text{and} \quad \mathbb{T}^{\mathrm{c}} := \mathbb{T}^{\mathrm{c}}(\mathcal{T}_{\mathrm{init}}) \tag{2.4}$$

of all non-conforming and conforming NVB refinements of $\mathcal{T}_{\mathrm{init}}$. Clearly, $\mathbb{T}^{\mathrm{c}} \subset \mathbb{T}^{\mathrm{nc}}$. We write $\mathcal{T} := \texttt{close}(\mathcal{P})$ if $\mathcal{P} \in \mathbb{T}^{\mathrm{nc}}$ is a partition and $\mathcal{T} \in \mathbb{T}^{\mathrm{c}}$ is the coarsest (conforming) refinement of $\mathcal{P}$. Existence and uniqueness of $\mathcal{T}$ follow from the fact that NVB is a binary refinement rule, and the order of the bisections does not matter. In particular, this also implies that $\texttt{refine}(\mathcal{T}, \mathcal{M}) = \texttt{close}(\texttt{bisect}(\mathcal{T}, \mathcal{M}))$ for all $\mathcal{T} \in \mathbb{T}^{\mathrm{c}}$ and $\mathcal{M} \subseteq \mathcal{T}$.

It follows from elementary geometric observations that NVB refinement leads only to finitely many shapes of simplices $T$; see, *e.g.*, [28]. Hence, all NVB refinements are uniformly $\gamma$-shape regular, *i.e.*,

$$\gamma := \sup_{\mathcal{P} \in \mathbb{T}^{\mathrm{nc}}} \max_{T \in \mathcal{P}} \frac{\mathrm{diam}(T)}{|T|^{1/d}} < \infty. \tag{2.5}$$

Finally, we recall the following properties of NVB, where $C_{\mathrm{son}}, C_{\mathrm{cls}} > 0$ are constants, which depend only on $\mathcal{T}_{\mathrm{init}}$ and the space dimension $d \geq 2$:

- **(M1)** **overlay estimate:** For all $\mathcal{P}, \mathcal{P}' \in \mathbb{T}^{\mathrm{nc}}$, there exists a (unique) coarsest common refinement $\mathcal{P} \oplus \mathcal{P}' \in \mathbb{T}^{\mathrm{nc}}(\mathcal{P}) \cap \mathbb{T}^{\mathrm{nc}}(\mathcal{P}')$. It holds that $\#(\mathcal{P} \oplus \mathcal{P}') \leq \#\mathcal{P} + \#\mathcal{P}' - \#\mathcal{T}_{\mathrm{init}}$. If $\mathcal{P}, \mathcal{P}' \in \mathbb{T}^{\mathrm{c}}$ are conforming, it also holds that $\mathcal{P} \oplus \mathcal{P}' \in \mathbb{T}^{\mathrm{c}}$.
- **(M2)** **finite number of sons:** For all $\mathcal{T} \in \mathbb{T}^{\mathrm{c}}$, $\mathcal{M} \subseteq \mathcal{T}$, and $\mathcal{T}' := \texttt{refine}(\mathcal{T}, \mathcal{M})$, it holds that $\bigcup\{T' \in \mathcal{T}' : T' \subseteq T\} = T$ and $\#\{T' \in \mathcal{T}' : T' \subseteq T\} \leq C_{\mathrm{son}}$ for all $T \in \mathcal{T}$.
- **(M3)** **mesh-closure estimate:** For all sequences $\mathcal{T}_j \in \mathbb{T}^{\mathrm{c}}$ such that $\mathcal{T}_0 = \mathcal{T}_{\mathrm{init}}$ and $\mathcal{T}_j = \texttt{refine}(\mathcal{T}_{j-1}, \mathcal{M}_{j-1})$ with $\mathcal{M}_{j-1} \subseteq \mathcal{T}_{j-1}$ for all $j \in \mathbb{N}$, it holds that

$$\#\mathcal{T}_J - \#\mathcal{T}_{\mathrm{init}} \leq C_{\mathrm{cls}} \sum_{j=0}^{J-1} \#\mathcal{M}_j \quad \text{for all } J \in \mathbb{N}_0. \tag{2.6}$$

- **(M4)** **conformity estimate:** For all partitions $\mathcal{P} \in \mathbb{T}^{\mathrm{nc}}$, it holds that

$$\#\texttt{close}(\mathcal{P}) - \#\mathcal{T}_{\mathrm{init}} \leq C_{\mathrm{cls}}(\#\mathcal{P} - \#\mathcal{T}_{\mathrm{init}}). \tag{2.7}$$

The overlay estimate (M1) is first proved in [27] for $d = 2$, but the proof transfers to arbitrary dimension $d \geq 2$; see [11]. For $d = 2$, (M2) obviously holds with $C_{\mathrm{son}} = 4$, while it is proved in [16] for general dimension $d \geq 2$. The closure estimate (M3) is first proved in [6] for $d = 2$. For general $d \geq 2$, it is proved in [28]. While the proofs of [6, 28] require an admissibility condition on $\mathcal{T}_{\mathrm{init}}$, the work [22] proves (M3) for $d = 2$, but arbitrary conforming triangulation $\mathcal{T}_{\mathrm{init}}$. It is easy to check that (M3) implies (M4); see Lemma 2.5 of [6] for a proof in 2D, which, however, directly transfers to arbitrary dimension $d \geq 2$.

## 2.3. Discrete function spaces

Given a fixed polynomial degree $m \in \mathbb{N}$ as well as $\mathcal{P} \in \mathbb{T}^{\mathrm{nc}}$ and $\mathcal{T} \in \mathbb{T}^{\mathrm{c}}$, we consider the discrete spaces

$$\begin{aligned} \mathbb{P}(\mathcal{P}) &:= \{Q_{\mathcal{P}} \in \mathbb{P} \ : \forall T \in \mathcal{P} \quad Q_{\mathcal{P}}|_T \text{ is polynomial of degree } \leq m-1\}, \\ \mathbb{V}(\mathcal{T}) &:= \{\boldsymbol{V}_{\mathcal{T}} \in \mathbb{V} : \forall T \in \mathcal{T} \quad \boldsymbol{V}_{\mathcal{T}}|_T \text{ is polynomial of degree } \leq m\}, \end{aligned} \tag{2.8}$$

which consist of piecewise polynomials.

**Remark 2.1.** We note that our analysis in principle permits to choose the polynomial degree for the pressure space $\mathbb{P}(\mathcal{P})$ larger than $m-1$. Indeed, the analysis of Kondratyuk and Stevenson [24] only exploits the assumption that the degree is not larger than $m-1$ to prove the local efficiency ([24], Prop. 5.2), which we do not require; see also Remark 3.1 of [24]. However, since we investigate (optimal) convergence of error quantities consisting of pressure as well as velocity terms, enlarging only the degree of the pressure space will in general not affect the best possible convergence rate; see also Remark 5.4. Moreover, both the present paper and [24] do not allow for degrees smaller than $m-1$, since otherwise the property $\mathcal{T} \in \mathbb{T}^{\mathrm{nc}}(\mathcal{P}') \cap \mathbb{T}^{\mathrm{c}}$ could no longer be guaranteed by Algorithm 3.6, and this condition is essential to ensure that the pressure meshes of the adaptive Algorithm 3.7 are coarser than the velocity meshes.

## 2.4. Auxiliary problems

Let $\mathcal{P} \in \mathbb{T}^{\mathrm{nc}}$. Then, $p_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$ denotes the best approximation of the exact pressure $p$ with respect to $\|\cdot\|_{\mathbb{P}}$, *i.e.*,

$$\|p - p_{\mathcal{P}}\|_{\mathbb{P}} = \min_{Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})} \|p - Q_{\mathcal{P}}\|_{\mathbb{P}}. \tag{2.9}$$

It is well-known that $p_{\mathcal{P}}$ can be obtained *via* the unique solution $(\boldsymbol{u}_{\mathcal{P}}, p_{\mathcal{P}}) \in \mathbb{V} \times \mathbb{P}(\mathcal{P})$ of the *reduced Stokes problem*

$$\begin{aligned} a(\boldsymbol{u}_{\mathcal{P}}, \boldsymbol{v}) + b(\boldsymbol{v}, p_{\mathcal{P}}) &= \langle \boldsymbol{f} \,,\, \boldsymbol{v}\rangle_\Omega && \text{for all } \boldsymbol{v} \in \mathbb{V}, \\ b(\boldsymbol{u}_{\mathcal{P}}, Q_{\mathcal{P}}) &= 0 && \text{for all } Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P}); \end{aligned} \tag{2.10}$$

see Section 4 of [24]. Note that the second condition can equivalently be stated as $\Pi_{\mathcal{P}} \nabla \cdot \boldsymbol{u}_{\mathcal{P}} = 0$ in $\Omega$, where $\Pi_{\mathcal{P}} : L^2(\Omega) \to \mathbb{P}(\mathcal{P})$ is the orthogonal projection with respect to $\|\cdot\|_\Omega$. Thus, (2.10) is just the variational formulation of (1.7).

Even though $p_{\mathcal{P}}$ is a discrete function, it cannot be computed since $\mathbb{V}$ is infinite dimensional. Given $q \in \mathbb{P}$, let $\boldsymbol{u}[q] \in \mathbb{V}$ be the unique solution to the (vector-valued) Poisson equation

$$a(\boldsymbol{u}[q], \boldsymbol{v}) = \langle \boldsymbol{f} \,,\, \boldsymbol{v}\rangle_\Omega - b(\boldsymbol{v}, q) \quad \text{for all } \boldsymbol{v} \in \mathbb{V}. \tag{2.11}$$

Note that $\boldsymbol{u}_{\mathcal{P}} = \boldsymbol{u}[p_{\mathcal{P}}]$.

Finally, let $\mathcal{T} \in \mathbb{T}^{\mathrm{nc}}(\mathcal{P}) \cap \mathbb{T}^{\mathrm{c}}$ be a conforming refinement of $\mathcal{P}$. Then, $\boldsymbol{U}_{\mathcal{T}}[q] \in \mathbb{V}(\mathcal{T})$ is the unique solution to the Galerkin discretization of (2.11)

$$a(\boldsymbol{U}_{\mathcal{T}}[q], \boldsymbol{V}_{\mathcal{T}}) = \langle \boldsymbol{f} \,,\, \boldsymbol{V}_{\mathcal{T}}\rangle_\Omega - b(\boldsymbol{V}_{\mathcal{T}}, q) \quad \text{for all } \boldsymbol{V}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T}). \tag{2.12}$$

Note that $\boldsymbol{U}_{\mathcal{T}}[q]$ is the Galerkin approximation to $\boldsymbol{u}[q]$ in $\mathbb{V}(\mathcal{T})$. Since $\|\cdot\|_{\mathbb{V}}$ denotes the energy norm corresponding to $a(\cdot,\cdot)$, there holds the Céa lemma

$$\|\boldsymbol{u}[q] - \boldsymbol{U}_{\mathcal{T}}[q]\|_{\mathbb{V}} = \min_{\boldsymbol{V}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T})} \|\boldsymbol{u}[q] - \boldsymbol{V}_{\mathcal{T}}\|_{\mathbb{V}}, \tag{2.13}$$

Recall the operators $A, B, B'$ from Section 2.1. Note that $\boldsymbol{u}[q] - \boldsymbol{u}[r] = A^{-1}B'(r-q)$ for arbitrary $q, r \in \mathbb{P}$, which yields that $\|\boldsymbol{u}[q] - \boldsymbol{u}[r]\|_{\mathbb{V}}^2 = \langle B'(r-q) \,,\, A^{-1}B'(r-q)\rangle_{\mathbb{V}^* \times \mathbb{V}}$. By definition of the operator $S = BA^{-1}B'$ and the norm $\|\cdot\|_{\mathbb{P}}$, we thus see that

$$\|\boldsymbol{U}_{\mathcal{T}}[q] - \boldsymbol{U}_{\mathcal{T}}[r]\|_{\mathbb{V}} \leq \|\boldsymbol{u}[q] - \boldsymbol{u}[r]\|_{\mathbb{V}} = \|q - r\|_{\mathbb{P}}. \tag{2.14}$$

### 2.5. Notational conventions

Throughout this work, $(\boldsymbol{u}, p) \in \mathbb{V} \times \mathbb{P}$ denotes the exact solution of the continuous Stokes problem (2.1). All occurring functions $\boldsymbol{u}_{\mathcal{P}}$, $\boldsymbol{u}[q]$, and $\boldsymbol{U}_{\mathcal{T}}[q]$ are approximations of $\boldsymbol{u}$. All occurring functions $p_{\mathcal{P}}$ and $P_{\mathcal{P}}$ are approximations of $p$. We employ bold face symbols for velocity functions, *e.g.*, $\boldsymbol{v} \in \mathbb{V}$ or $\boldsymbol{V}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T})$, and normal font for pressure functions, *e.g.*, $q \in \mathbb{P}$, $Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$. Finally, small letters indicate functions, which are continuous or not computable, *e.g.*, $\boldsymbol{u}$, $p$, and $p_{\mathcal{P}}$, while *computable discrete* functions are written with capital letters, *e.g.*, $\boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}]$. The corresponding partitions $\mathcal{P} \in \mathbb{T}^{\rm nc}$ resp. triangulations $\mathcal{T} \in \mathbb{T}^{\rm c}$ are always indicated by indices. The most important symbols are listed in Appendix D.

### 2.6. Abbreviate notation for adaptive algorithm

The adaptive algorithm below generates nested partitions $\mathcal{P}_i \in \mathbb{T}^{\rm nc}$ and triangulations $\mathcal{T}_{ijk} \in \mathbb{T}^{\rm c}$ for certain indices $(i,j,k) \in \mathcal{Q} \subset \mathbb{N}_0^3$ such that $\mathcal{T}_{ijk} \in \mathbb{T}^{\rm nc}(\mathcal{P}_i) \cap \mathbb{T}^{\rm c}$. Furthermore, it provides approximations

$$p \approx P_{ij} \in \mathbb{P}_i := \mathbb{P}(\mathcal{P}_i) \quad \text{as well as} \quad \boldsymbol{u} \approx \boldsymbol{U}_{ijk} \in \mathbb{V}_{ijk} := \mathbb{V}(\mathcal{T}_{ijk}). \tag{2.15}$$

More precisely and with the notation from Section 2.4, it holds that[3]

$$P_{ij} \approx p_i := p_{\mathcal{P}_i} \quad \text{as well as} \quad \boldsymbol{U}_{ijk} \approx \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}] \approx \boldsymbol{u}[P_{ij}] =: \boldsymbol{u}_{ij}. \tag{2.16}$$

Besides this notation, let

$$\Pi_i := \Pi_{\mathcal{P}_i} : L^2(\Omega) \to \mathbb{P}(\mathcal{P}_i) \tag{2.17}$$

be the $L^2(\Omega)$-orthogonal projection (with respect to $\|\cdot\|_\Omega$) and let

$$\eta_{ijk} := \eta(\mathcal{T}_{ijk}; \boldsymbol{U}_{ijk}, P_{ij}) \approx \eta(\mathcal{T}_{ijk}; \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}], P_{ij}) \tag{2.18}$$

be the computable *a posteriori* error estimator from Section 3.1 below.

## 3. Adaptive Uzawa algorithm

### 3.1. *A posteriori* error estimation

Throughout this section, let $\mathcal{P} \in \mathbb{T}^{\rm nc}$ be a partition of $\Omega \subset \mathbb{R}^d$ and $\mathcal{T} \in \mathbb{T}^{\rm nc}(\mathcal{P}) \cap \mathbb{T}^{\rm c}$ be a conforming refinement. We recall the residual *a posteriori* error estimator: For $T \in \mathcal{T}$, $Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$, and $\boldsymbol{V}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T})$, define

$$\eta_T(\boldsymbol{V}_{\mathcal{T}}, Q_{\mathcal{P}})^2 := |T|^{2/n} \, \|\boldsymbol{f} - \nabla Q_{\mathcal{P}} + \Delta \boldsymbol{V}_{\mathcal{T}}\|_T^2 + |T|^{1/n} \, \|[\![ Q_{\mathcal{P}} \, \boldsymbol{n} - \nabla \boldsymbol{V}_{\mathcal{T}} \cdot \boldsymbol{n} ]\!]\|_{\partial T \cap \Omega}^2, \tag{3.1}$$

where $[\![\cdot]\!]$ denotes the jump of its argument over $\partial T$. Then, the error estimator reads

$$\eta(\mathcal{M}; \boldsymbol{V}_{\mathcal{T}}, Q_{\mathcal{P}})^2 := \sum_{T \in \mathcal{M}} \eta_T(\boldsymbol{V}_{\mathcal{T}}, Q_{\mathcal{P}})^2 \quad \text{for all } \mathcal{M} \subset \mathcal{T}. \tag{3.2}$$

In the following, we recall some important properties of $\eta$ from [11,24]. We start with the available reliability results.

**Lemma 3.1** (*reliability* [24], Props. 5.1, 5.5)**.** *There exists a constant $C_{\rm rel} > 0$ such that, for all $Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$, it holds that*

$$\|\boldsymbol{u}[Q_{\mathcal{P}}] - \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}]\|_{\mathbb{V}} \le C_{\rm rel} \, \eta(\mathcal{T}; \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}], Q_{\mathcal{P}}). \tag{3.3}$$

[3] Do not confuse the pressure $p_i$ with the iterates $p^j$ of the exact Uzawa algorithm (1.6).

*Moreover, it holds that*

$$\|\boldsymbol{u}_{\mathcal{P}} - \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}]\|_{\mathbb{V}} + \|p_{\mathcal{P}} - Q_{\mathcal{P}}\|_{\mathbb{P}} \le C_{\mathrm{rel}}\left(\eta(\mathcal{T}; \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}], Q_{\mathcal{P}}) + \|\Pi_{\mathcal{P}} \nabla \cdot \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}]\|_{\Omega}\right) \tag{3.4}$$

*as well as*

$$\|\boldsymbol{u} - \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}]\|_{\mathbb{V}} + \|p - Q_{\mathcal{P}}\|_{\mathbb{P}} \le C_{\mathrm{rel}}\left(\eta(\mathcal{T}; \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}], Q_{\mathcal{P}}) + \|\nabla \cdot \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}]\|_{\Omega}\right). \tag{3.5}$$

*The constant $C_{\mathrm{rel}}$ depends only on $\gamma$-shape regularity.*

For some fixed discrete pressure $Q_{\mathcal{P}}$, we recall the localized upper bound in the current form of [11], which improves Proposition 5.1 of [24].

**Lemma 3.2** (*discrete reliability* [11], Lem. 3.6)**.** *Let $\widehat{\mathcal{T}} \in \mathbb{T}^{\mathrm{c}}(\mathcal{T})$. There exists a constant $C_{\mathrm{drel}} > 0$ such that, for all $Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$, it holds that*

$$\|\boldsymbol{U}_{\widehat{\mathcal{T}}}[Q_{\mathcal{P}}] - \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}]\|_{\mathbb{V}} \le C_{\mathrm{drel}}\, \eta(\mathcal{T} \setminus \widehat{\mathcal{T}}; \boldsymbol{U}[Q_{\mathcal{P}}], Q_{\mathcal{P}}). \tag{3.6}$$

*The constant $C_{\mathrm{drel}}$ depends only on $\gamma$-shape regularity.*

Next, we note that the estimator depends Lipschitz continuously on the arguments. The result is slightly stronger than Proposition 5.4 of [24], but the proof is standard [11].

**Lemma 3.3** (*stability* [11], Prop. 3.3)**.** *Let $\widehat{\mathcal{T}} \in \mathbb{T}^{\mathrm{c}}(\mathcal{T})$. There exists a constant $C_{\mathrm{stab}} > 0$ such that, for all $\boldsymbol{V}_{\widehat{\mathcal{T}}} \in \mathbb{V}(\widehat{\mathcal{T}})$, $\boldsymbol{W}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T})$, and $Q_{\mathcal{P}}, R_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$, it holds that*

$$|\eta(\mathcal{T} \cap \widehat{\mathcal{T}}; \boldsymbol{V}_{\widehat{\mathcal{T}}}, Q_{\mathcal{P}}) - \eta(\mathcal{T} \cap \widehat{\mathcal{T}}; \boldsymbol{W}_{\mathcal{T}}, R_{\mathcal{P}})| \le C_{\mathrm{stab}}\left(\|\boldsymbol{V}_{\widehat{\mathcal{T}}} - \boldsymbol{W}_{\mathcal{T}}\|_{\mathbb{V}} + \|Q_{\mathcal{P}} - R_{\mathcal{P}}\|_{\mathbb{P}}\right). \tag{3.7}$$

*The constant $C_{\mathrm{stab}}$ depends only on the polynomial degree $m$ and $\gamma$-shape regularity.*

The following reduction property follows from the reduction of the mesh-size on refined elements. The proof is standard [11].

**Lemma 3.4** (*reduction* [11], Proof of Cor. 3.4)**.** *Let $\widehat{\mathcal{T}} \in \mathbb{T}^{\mathrm{c}}(\mathcal{T})$. Let $Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$. Then, with $q_{\mathrm{red}} = 2^{-1/(n+1)}$, there holds the reduction property*

$$\eta(\widehat{\mathcal{T}} \setminus \mathcal{T}; \boldsymbol{U}_{\widehat{\mathcal{T}}}[Q_{\mathcal{P}}], Q_{\mathcal{P}}) \le q_{\mathrm{red}}\, \eta(\mathcal{T} \setminus \widehat{\mathcal{T}}; \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}], Q_{\mathcal{P}}) + C_{\mathrm{red}}\, \|\boldsymbol{U}_{\widehat{\mathcal{T}}}[Q_{\mathcal{P}}] - \boldsymbol{U}_{\mathcal{T}}[Q_{\mathcal{P}}]\|_{\mathbb{V}}. \tag{3.8}$$

*The constant $C_{\mathrm{red}} > 0$ depends only on the polynomial degree $m$ and $\gamma$-shape regularity.*

Finally, for the divergence contribution to the Stokes error estimator, we recall the following equivalence. The result is slightly stronger than Proposition 5.7 of [24].

**Lemma 3.5.** *Let $C_{\mathrm{div}} \ge 1$ be the norm equivalence constant from (2.3). Let $\Pi_{\mathcal{P}} : L^2(\Omega) \to \mathbb{P}(\mathcal{P})$ be the $L^2(\Omega)$-orthogonal projection. If $Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$, then it holds that*

$$\|\Pi_{\mathcal{P}} \nabla \cdot \boldsymbol{u}[Q_{\mathcal{P}}]\|_{\Omega} \le \|\nabla \cdot (\boldsymbol{u}_{\mathcal{P}} - \boldsymbol{u}[Q_{\mathcal{P}}])\|_{\Omega} \le \|p_{\mathcal{P}} - Q_{\mathcal{P}}\|_{\mathbb{P}} \le C_{\mathrm{div}}\, \|\Pi_{\mathcal{P}} \nabla \cdot \boldsymbol{u}[Q_{\mathcal{P}}]\|_{\Omega}. \tag{3.9}$$

*If $q \in \mathbb{P}$, it holds that*

$$\|\nabla \cdot \boldsymbol{u}[q]\|_{\Omega} \le \|p - q\|_{\mathbb{P}} \le C_{\mathrm{div}}\, \|\nabla \cdot \boldsymbol{u}[q]\|_{\Omega}. \tag{3.10}$$

*Proof.* From the definition of the Schur complement operator, we have that

$$\nabla\cdot(\boldsymbol{u}_{\mathcal{P}}-\boldsymbol{u}[Q_{\mathcal{P}}]) = S(p_{\mathcal{P}}-Q_{\mathcal{P}}). \tag{3.11}$$

Taking into account (2.3), we obtain that

$$\begin{aligned}\|\nabla\cdot(\boldsymbol{u}_{\mathcal{P}}-\boldsymbol{u}[Q_{\mathcal{P}}])\|_\Omega^2 &\overset{(3.11)}{=} \langle S(p_{\mathcal{P}}-Q_{\mathcal{P}})\,,\,\nabla\cdot(\boldsymbol{u}_{\mathcal{P}}-\boldsymbol{u}[Q_{\mathcal{P}}])\rangle_\Omega\\ &= \langle p_{\mathcal{P}}-Q_{\mathcal{P}}\,,\,\nabla\cdot(\boldsymbol{u}_{\mathcal{P}}-\boldsymbol{u}[Q_{\mathcal{P}}])\rangle_{\mathbb{P}} \le \|p_{\mathcal{P}}-Q_{\mathcal{P}}\|_{\mathbb{P}}\,\|\nabla\cdot(\boldsymbol{u}_{\mathcal{P}}-\boldsymbol{u}[Q_{\mathcal{P}}])\|_{\mathbb{P}}\\ &\overset{(2.3)}{\le} \|p_{\mathcal{P}}-Q_{\mathcal{P}}\|_{\mathbb{P}}\,\|\nabla\cdot(\boldsymbol{u}_{\mathcal{P}}-\boldsymbol{u}[Q_{\mathcal{P}}])\|_\Omega.\end{aligned}$$

Together with $\Pi_{\mathcal{P}}\nabla\cdot\boldsymbol{u}_{\mathcal{P}}=0$, this proves that

$$\|\Pi_{\mathcal{P}}\nabla\cdot\boldsymbol{u}[Q_{\mathcal{P}}]\|_\Omega \le \|\nabla\cdot(\boldsymbol{u}_{\mathcal{P}}-\boldsymbol{u}[Q_{\mathcal{P}}])\|_\Omega \le \|p_{\mathcal{P}}-Q_{\mathcal{P}}\|_{\mathbb{P}}.$$

On the other hand, note that $\Pi_{\mathcal{P}}(p_{\mathcal{P}}-Q_{\mathcal{P}}) = p_{\mathcal{P}}-Q_{\mathcal{P}}$. The norm equivalence (2.3) and the Cauchy–Schwarz inequality thus imply that

$$\begin{aligned}C_{\mathrm{div}}\|p_{\mathcal{P}}-Q_{\mathcal{P}}\|_{\mathbb{P}}\,\|\Pi_{\mathcal{P}}\nabla\cdot\boldsymbol{u}[Q_{\mathcal{P}}]\|_\Omega &\overset{(2.3)}{\ge} \|p_{\mathcal{P}}-Q_{\mathcal{P}}\|_\Omega\,\|\Pi_{\mathcal{P}}\nabla\cdot\boldsymbol{u}[Q_{\mathcal{P}}]\|_\Omega\\ &\ge -\langle p_{\mathcal{P}}-Q_{\mathcal{P}}\,,\,\Pi_{\mathcal{P}}\nabla\cdot\boldsymbol{u}[Q_{\mathcal{P}}]\rangle_\Omega \;=\; \langle p_{\mathcal{P}}-Q_{\mathcal{P}}\,,\,\Pi_{\mathcal{P}}\nabla\cdot(\boldsymbol{u}_{\mathcal{P}}-\boldsymbol{u}[Q_{\mathcal{P}}])\rangle_\Omega\\ &= \langle p_{\mathcal{P}}-Q_{\mathcal{P}}\,,\,\nabla\cdot(\boldsymbol{u}_{\mathcal{P}}-\boldsymbol{u}[Q_{\mathcal{P}}])\rangle_\Omega \overset{(3.11)}{=} \langle S(p_{\mathcal{P}}-Q_{\mathcal{P}})\,,\,p_{\mathcal{P}}-Q_{\mathcal{P}}\rangle_\Omega = \|p_{\mathcal{P}}-Q_{\mathcal{P}}\|_{\mathbb{P}}^2\end{aligned}$$

and therefore $\|p_{\mathcal{P}}-Q_{\mathcal{P}}\|_{\mathbb{P}} \le C_{\mathrm{div}}\|\Pi_{\mathcal{P}}\nabla\cdot\boldsymbol{u}[Q_{\mathcal{P}}]\|_\Omega$. This concludes the proof of (3.9). The proof of (3.10) follows along the same lines (with $p=p_{\mathcal{P}}$ and hence $0=\nabla\cdot\boldsymbol{u}=\nabla\cdot\boldsymbol{u}_{\mathcal{P}}$, and $q=Q_{\mathcal{P}}$). ☐

## 3.2. Adaptive refinement of pressure triangulation

To refine the partitions $\mathcal{P}_i$, we apply the following algorithm from Section 2 of [4] (which slightly differs from the well-known *thresholding second algorithm* of [5]). We will use it in Algorithm 3.7 with parameters $\mathcal{P}_i, \mathcal{T}_{ijk}, \boldsymbol{U}_{ijk} \approx \boldsymbol{u}[P_{ij}]$, and $\vartheta$. In this context, the idea of Algorithm 3.6 is to achieve that $\|p_{i+1}-P_{ij}\|_{\mathbb{P}} \simeq \|\Pi_{i+1}\nabla\cdot u[P_{ij}]\|_\Omega$ dominates $\|p-P_{ij}\|_{\mathbb{P}} \simeq \|\nabla\cdot u[P_{ij}]\|_\Omega$ (see Lem. 3.5), and to subsequently proceed to the iteration in $j$ and improve the Uzawa approximation.

**Algorithm 3.6.** Input: Partition $\mathcal{P}' := \mathcal{P} \in \mathbb{T}^{\mathrm{nc}}$, triangulation $\mathcal{T} \in \mathbb{T}^{\mathrm{nc}}(\mathcal{P}) \cap \mathbb{T}^{\mathrm{c}}$, function $\boldsymbol{V}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T})$, adaptivity parameter $0<\vartheta\le 1$.
Loop: Iterate the following steps (i)–(iii) until $\vartheta\,\|\nabla\cdot\boldsymbol{V}_{\mathcal{T}}\|_\Omega \le \|\Pi_{\mathcal{P}'}\nabla\cdot\boldsymbol{V}_{\mathcal{T}}\|_\Omega$:

- (i) Compute $e(T) := \inf\{\|\nabla\cdot\boldsymbol{V}_{\mathcal{T}}-Q\|_T^2 \,:\, Q$ polynomial of degree $m-1\}$ for all $T\in\mathcal{P}'$, for which $e(T)$ has not been already computed.
- (ii) For all $T\in\mathcal{P}'$ for which $\widetilde{e}(T)$ has not been already defined, define $\widetilde{e}(T) := e(T)$ if $T\in\mathcal{P}$ and $\widetilde{e}(T) := e(T)\widetilde{e}(\widetilde{T})/(e(T)+\widetilde{e}(\widetilde{T}))$ otherwise, where $\widetilde{T}$ denotes the unique father element of $T$.
- (iii) Choose one element $T\in\mathcal{P}'$ with $\widetilde{e}(T) = \max_{T'\in\mathcal{P}'}\widetilde{e}(T')$ and employ newest vertex bisection to generate $\mathcal{P}' := \mathtt{bisect}(\mathcal{P}',\{T\})$.

Output: Triangulation $\mathcal{P}' = \mathtt{binev}(\mathcal{P},\mathcal{T},\boldsymbol{V}_{\mathcal{T}};\vartheta) \in \mathbb{T}^{\mathrm{nc}}(\mathcal{P})$ with $\mathcal{T}\in\mathbb{T}^{\mathrm{nc}}(\mathcal{P}')\cap\mathbb{T}^{\mathrm{c}}$.

According to Theorem 2.1 of [4], the output $\mathcal{P}'$ is a quasi-optimal mesh in $\mathbb{T}^{\mathrm{nc}}(\mathcal{P})$ with $\vartheta\,\|\nabla\cdot\boldsymbol{V}_{\mathcal{T}}\|_\Omega \le \|\Pi_{\mathcal{P}'}\nabla\cdot\boldsymbol{V}_{\mathcal{T}}\|_\Omega$: This means that for all $\vartheta<\vartheta'<1$ and all $\widetilde{\mathcal{P}}\in\mathbb{T}^{\mathrm{nc}}(\mathcal{P})$ with $\vartheta'\,\|\nabla\cdot\boldsymbol{V}_{\mathcal{T}}\|_\Omega \le \|\Pi_{\widetilde{\mathcal{P}}}\nabla\cdot\boldsymbol{V}_{\mathcal{T}}\|_\Omega$, it holds that $\#\mathcal{P}'-\#\mathcal{P} \le C_{\mathrm{bin}}\,(\#\widetilde{\mathcal{P}}-\#\mathcal{P})$ for some $C_{\mathrm{bin}}>1$, which depends only on the ratio $(1-\vartheta'^2)/(1-\vartheta^2)$. The same reference states that the effort of Algorithm 3.6 is $\mathcal{O}(\#\mathcal{T}\log(\#\mathcal{T}))$ if $0<\vartheta<1$.

To obtain optimal algebraic convergence rates of the error estimator, one has to choose $\vartheta$ sufficiently small and $\vartheta'$ sufficiently close to $\vartheta$; see Theorem 5.3 below.

### 3.3. Adaptive Uzawa algorithm

We investigate the following adaptive Uzawa algorithm, which goes back to Section 7 of [24].

**Algorithm 3.7.** INPUT: Conforming initial triangulation $\mathcal{P}_0 := \mathcal{T}_{000} := \mathcal{T}_{\rm init}$ of $\Omega$, initial approximation $P_{00} = 0$, counters $i = j = k = 0$, adaptivity parameters $0 \le \kappa_1 < 1$, $0 < \kappa_2 < 1$, $0 < \kappa_3 < 1$, $0 < \vartheta \le 1$, $0 < \theta \le 1$, and $C_{\rm mark} \ge 1$.
LOOP: Iterate the following steps (i)–(iv):

(i) Compute $\boldsymbol{U}_{ijk} \in \mathbb{V}_{ijk}$ as well as (all local contributions of) the corresponding error estimator $\eta_{ijk} = \eta(\mathcal{T}_{ijk}; \boldsymbol{U}_{ijk}, P_{ij})$ such that the exact Galerkin approximation $\boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}] \in \mathbb{V}_{ijk}$ of $\boldsymbol{u}_{ij}$ satisfies that $\|\boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}] - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} \le \kappa_1\, \eta_{ijk}$.

(ii) `while` $\eta_{ijk} + \|\Pi_i \nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega \le \kappa_2 \left(\eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega\right)$ `do`
- Determine $\mathcal{P}_{i+1} := \texttt{binev}(\mathcal{P}_i, \mathcal{T}_{ijk}, \boldsymbol{U}_{ijk}; \vartheta)$ by Algorithm 3.6.
- Define $P_{(i+1)0} := P_{ij}$, and $\mathcal{T}_{(i+1)00} := \mathcal{T}_{ijk}$.
- Update counters $(i, j, k) \mapsto (i+1, 0, 0)$.

`end while`

(iii) `if` $\eta_{ijk} \le \kappa_3 \|\Pi_i \nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega$ `then`
- Define $P_{i(j+1)} := P_{ij} - \Pi_i \nabla\cdot \boldsymbol{U}_{ijk} \in \mathbb{P}_i$, and $\mathcal{T}_{i(j+1)0} := \mathcal{T}_{ijk}$.
- Update counters $(i, j, k) \mapsto (i, j+1, 0)$.

(iv) `else`
- Determine a set $\mathcal{M}_{ijk} \subseteq \mathcal{T}_{ijk}$ of (up to the fixed factor $C_{\rm mark}$) minimal cardinality, which satisfies the Dörfler marking criterion

$$\theta\, \eta_{ijk}^2 \le \eta(\mathcal{M}_{ijk}; P_{ij}, \boldsymbol{U}_{ijk})^2. \tag{3.12}$$

- Generate $\mathcal{T}_{ij(k+1)} := \texttt{refine}(\mathcal{T}_{ijk}, \mathcal{M}_{ijk})$.
- Update counters $(i, j, k) \mapsto (i, j, k+1)$.

`end if`

**Remark 3.8.** The actual implementation of Algorithm 3.7 will replace the triple indices $(i, j, k)$ by one single index $n \in \mathbb{N}_0$, which is increased in each step (ii)–(iv). However, the present statement of the algorithm makes the numerical analysis more accessible.

**Lemma 3.9.** *Define the index set $\mathcal{Q} := \{(i, j, k) \in \mathbb{N}_0^3 : \boldsymbol{U}_{ijk}$ is defined by Algorithm 3.7$\}$. Then, for $(i, j, k) \in \mathbb{N}_0^3$, there hold the following assertions* (a)–(c)*:*

(a) *If $(i, j, k+1) \in \mathcal{Q}$, then $(i, j, k) \in \mathcal{Q}$.*
(b) *If $(i, j+1, 0) \in \mathcal{Q}$, then $(i, j, 0) \in \mathcal{Q}$ and $\underline{k}(i, j) := \max\{k \in \mathbb{N}_0 : (i, j, k) \in \mathcal{Q}\} < \infty$.*
(c) *If $(i+1, 0, 0) \in \mathcal{Q}$, then $(i, 0, 0) \in \mathcal{Q}$ and $\underline{j}(i) := \max\{j \in \mathbb{N}_0 : (i, j, 0) \in \mathcal{Q}\} < \infty$.*

*Throughout, we shall make the following conventions for the triple index: If we write $\eta_{ij\underline{k}}$ etc. (see, e.g., Lem. 4.5), then (implicitly) $\underline{k} = \underline{k}(i, j)$. If we write $\eta_{i\underline{j}\underline{k}}$ etc. (see, e.g., Lem. 4.6), then (implicitly) $\underline{j} = \underline{j}(i)$ and $\underline{k} = \underline{k}(i, \underline{j})$.*

*Proof.* Each step (ii)–(iv) of the algorithm increases either $i$ or $j$ or $k$ by one. □

**Remark 3.10.** Unlike the algorithm from [24], our formulation of the adaptive Uzawa algorithm avoids any special treatment of the data oscillations (*i.e.*, to resolve $\boldsymbol{f}$ by a piecewise polynomial in an additional loop).

**Remark 3.11.** We note that the choice $\boldsymbol{U}_{ijk} := \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}]$ (*i.e.*, $\kappa_1 = 0$) is admissible in step (i) of Algorithm 3.7. In the spirit of [15], one can also employ the PCG algorithm ([20], Algorithm 11.5.1) with optimal preconditioner. With $\kappa_1'$ and an additional index $\ell \in \mathbb{N}_0$ for the PCG iteration and initially $\ell := 0$, repeat the following three steps, until $U_{ijk} := U_{ijk(\ell+1)}$ satisfies that $\|\boldsymbol{U}_{ijk(\ell+1)} - \boldsymbol{U}_{ijk\ell}\|_{\mathbb{V}} \le \kappa_1'\, \eta_{ijk(\ell+1)}$:

– Do one PCG step to obtain $\boldsymbol{U}_{ijk(\ell+1)} \in \mathbb{V}_{ijk}$ from $\boldsymbol{U}_{ijk\ell} \in \mathbb{V}_{ijk}$.
– Compute (all local contributions of) the estimator $\eta_{ijk(\ell+1)} := \eta(\mathcal{T}_{ijk}; \boldsymbol{U}_{ijk(\ell+1)}, P_{ij})$.
– Update counters $(i,j,k,\ell) \mapsto (i,j,k,\ell+1)$.

If the preconditioner is optimal, *i.e.*, the preconditioned linear system has uniformly bounded condition number, then it follows that PCG is a uniform contraction ([15], Sect. 2.6): There exists $0 < q_{\rm pcg} < 1$ such that

$$\|U_{\mathcal{T}_{ijk}}[P_{ij}] - \boldsymbol{U}_{ijk(\ell+1)}\|_{\mathbb{V}} \le q_{\rm pcg}\, \|U_{\mathcal{T}_{ijk}}[P_{ij}] - \boldsymbol{U}_{ijk\ell}\|_{\mathbb{V}} \quad \text{for all } \ell \in \mathbb{N}_0.$$

Hence, the PCG loop terminates, and the triangle inequality proves that

$$\|U_{\mathcal{T}_{ijk}}[P_{ij}] - \boldsymbol{U}_{ijk(\ell+1)}\|_{\mathbb{V}} \le \frac{q_{\rm pcg}}{1-q_{\rm pcg}}\, \|\boldsymbol{U}_{ijk(\ell+1)} - \boldsymbol{U}_{ijk\ell}\|_{\mathbb{V}} \le \frac{q_{\rm pcg}}{1-q_{\rm pcg}}\, \kappa_1'\, \eta_{ijk(\ell+1)},$$

*i.e.*, the criterion of step (i) of Algorithm 3.7 is satisfied for $\kappa_1 := \kappa_1' q_{\rm pcg}/(1-q_{\rm pcg})$.

## 4. Convergence

### 4.1. Main theorem on linear convergence

To state linear convergence, we need an ordering of the set $\mathcal{Q}$ from Lemma 3.9: For $(i,j,k),(i',j',k') \in \mathcal{Q}$, write $(i',j',k') < (i,j,k)$ if the index $(i',j',k')$ appears earlier in Algorithm 3.7 than $(i,j,k)$. Define

$$|(i,j,k)| := \#\{(i',j',k') \in \mathcal{Q} \,:\, (i',j',k') < (i,j,k)\} \in \mathbb{N}_0. \tag{4.1}$$

Note that $|(i,j,k)|$ coincides with the single index $n$ from Remark 3.8. Then, we have the following theorem. The proof is given in Section 4.3.

**Theorem 4.1.** *Let $0 < \kappa_1 < \theta^{1/2}/C_{\rm stab}$. Suppose that $0 < \kappa_2, \kappa_3 < 1$ are sufficiently small as in Lemmas 4.5 and 4.6 below. Let $0 < \vartheta \le 1$ and $0 < \theta \le 1$. Then, there exist constants $C_{\rm lin} > 0$ and $0 < q_{\rm lin} < 1$ such that*

$$\eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega \le C_{\rm lin} q_{\rm lin}^{|(i,j,k)|-|(i',j',k')|} \left(\eta_{i'j'k'} + \|\nabla\cdot \boldsymbol{U}_{i'j'k'}\|_\Omega\right) \tag{4.2}$$

*for all $(i',j',k'),(i,j,k) \in \mathcal{Q}$ with $(i',j',k') < (i,j,k)$. The constants $C_{\rm lin}$ and $q_{\rm lin}$ depend only on the domain $\Omega$, $\gamma$-shape regularity, the polynomial degree $m$, and the parameters $\kappa_1, \kappa_2$, $\kappa_3$, $\vartheta$, and $\theta$.*

**Remark 4.2.** The adaptive Uzawa algorithm from Bänsch *et al.* [1] employs only one triangulation for both, the pressure and the velocity. Similarly, we can additionally update $\mathcal{P}_i := \mathcal{T}_{ij(k+1)}$ in step (iv) of Algorithm 3.7. Since $0 < \kappa_2 < 1$ and $\Pi_i \nabla\cdot U_{ijk} = \nabla U_{ijk}$, then the condition in (ii) will always fail. We note that the convergence analysis of Section 4.2 and in particular, linear convergence (Thm. 4.1) clearly remain valid for this modified algorithm, while our proof of optimal convergence rates (Thm. 5.3) fails.

### 4.2. Auxiliary results

The first lemma provides links between the exact Galerkin solutions $\boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}]$ and its approximations $\boldsymbol{U}_{ijk}$.

**Lemma 4.3.** *Let $(i,j,k) \in \mathcal{Q}$. For all $\mathcal{S} \subseteq \mathcal{T}_{ijk}$, it holds that*

$$|\eta(\mathcal{S}; \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}], P_{ij}) - \eta(\mathcal{S}; \boldsymbol{U}_{ijk}, P_{ij})| \le \kappa_1\, C_{\rm stab}\, \eta_{ijk}, \tag{4.3}$$

*where $C_{\rm stab} > 0$ is the constant from Lemma 3.3. This particularly yields the equivalence*

$$(1-\kappa_1\, C_{\rm stab})\, \eta_{ijk} \le \eta(\mathcal{T}_{ijk}; \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}], P_{ij}) \le (1+\kappa_1\, C_{\rm stab})\, \eta_{ijk}. \tag{4.4}$$

*as well as the reliability estimates*

$$\|\boldsymbol{u}_{ij} - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} \le C'_{\rm rel}(\kappa_1)\,\eta_{ijk}, \tag{4.5}$$
$$\|\boldsymbol{u}_{i} - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} + \|p_i - P_{ij}\|_{\mathbb{P}} \le C'_{\rm rel}(\kappa_1)\,\big(\eta_{ijk} + \|\Pi_i \nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega\big), \tag{4.6}$$
$$\|\boldsymbol{u} - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} + \|p - P_{ij}\|_{\mathbb{P}} \le C'_{\rm rel}(\kappa_1)\,\big(\eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega\big), \tag{4.7}$$

*where* $C'_{\rm rel}(\kappa_1) := ((1+\kappa_1 C_{\rm stab})C_{\rm rel} + \kappa_1(C_{\rm rel}+1)) \ge C_{\rm rel}$ *with* $C_{\rm rel} > 0$ *from Lemma 3.1.*

*Proof.* To shorten notation, we set $\eta^\star_{ijk} := \eta(\mathcal{T}_{ijk}; \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}], P_{ij})$. The stability (4.3) follows from Lemma 3.3 and $\|\boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}] - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} \le \kappa_1\,\eta_{ijk}$, which is guaranteed by step (i) of Algorithm 3.7. Taking $\mathcal{S} = \mathcal{T}_{ijk}$, (4.4) is an immediate consequence. To see (4.5), we use reliability (3.3), step (i) of Algorithm 3.7, and (4.4) to see that

$$\|\boldsymbol{u}_{ij} - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} \overset{(3.3)}{\le} C_{\rm rel}\,\eta^\star_{ijk} + \|\boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}] - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} \overset{(4.4)}{\le} ((1+\kappa_1 C_{\rm stab})C_{\rm rel} + \kappa_1)\,\eta_{ijk}.$$

To prove (4.6), we apply (3.4)

$$\begin{aligned}\|\boldsymbol{u}_{i} - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} + \|p_i - P_{ij}\|_{\mathbb{P}} &\overset{(3.4)}{\le} C_{\rm rel}\big(\eta^\star_{ijk} + \|\Pi_i\nabla\cdot\boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}]\|_\Omega\big) + \|\boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}] - \boldsymbol{U}_{ijk}\|_{\mathbb{V}}\\ &\overset{(4.4)}{\le} ((1+\kappa_1 C_{\rm stab})C_{\rm rel} + \kappa_1)\,\eta_{ijk} + C_{\rm rel}\,\|\Pi_i\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega.\end{aligned}$$

Similarly, (4.7) follows from (3.5). ☐

The following three lemmas prove that Algorithm 3.7 leads to contraction if either $i$, $j$, or $k$ is increased. Throughout, let $0 < \vartheta \le 1$, $0 < \theta \le 1$, and, if not stated otherwise, $0 \le \kappa_1 < 1$, $0 < \kappa_2, \kappa_3 < 1$.

**Lemma 4.4.** *Let* $(i,j,0) \in \mathcal{Q}$ *and define* $\underline{k} := \max\{k \in \mathbb{N}_0 \colon (i,j,k) \in \mathcal{Q}\} \in \mathbb{N}_0 \cup \{\infty\}$. *If* $0 \le \kappa_1 < \theta^{1/2}/C_{\rm stab}$, *then, there exist constants* $0 < q_1 < 1$ *and* $C_1 > 0$, *which depend only on* $\gamma$*-shape regularity, the polynomial degree* $m$, $\kappa_1$, *and* $\theta$, *such that*

$$\eta_{ij(k+n)} \le C_1\,q_1^n\,\eta_{ijk} \quad \textit{for all } k,n \in \mathbb{N}_0 \textit{ with } k \le k+n \le \underline{k}. \tag{4.8}$$

*Moreover, it holds that*

$$\eta_{ijk} \le \eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega \le \frac{1}{\kappa_2}\left(1 + \frac{1}{\kappa_3}\right)\eta_{ijk} \quad \textit{for all } 0 \le k < \underline{k}. \tag{4.9}$$

*If* $\underline{k} = \infty$, *this yields that* $\|\boldsymbol{u} - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} + \|p - P_{ij}\|_{\mathbb{P}} \to 0$ *as* $k \to \infty$ *with* $p = p_i = P_{ij}$.

*Proof.* We split the proof into three steps.

**Step 1.** If $\boldsymbol{U}_{ijk} = \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}]$ for all $(i,j,k) \in \mathcal{Q}$, step (iv) of Algorithm 3.7 is the usual adaptive step in an adaptive algorithm for, *e.g.*, the (vector-valued) Poisson model problem. Hence, (4.8) follows from reliability (3.3), stability (3.7) and reduction (3.8); see, *e.g.*, Theorem 4.1 (i) of [10]. For general $\boldsymbol{U}_{ijk}$, (4.8) follows from Theorem 7.2 of [10] under the constraint $0 \le \kappa_1 < \theta^{1/2}/C_{\rm stab}$.

**Step 2.** If $k < \underline{k}$, the structure of Algorithm 3.7 implies that the conditions in step (ii) and (iii) are false, *i.e.*,

$$\eta_{ijk} + \|\Pi_i\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega > \kappa_2\,\big(\eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega\big) \quad \text{and} \quad \eta_{ijk} > \kappa_3\,\|\Pi_i\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega.$$

Hence,

$$\eta_{ijk} \le \eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega < \frac{1}{\kappa_2}\big(\eta_{ijk} + \|\Pi_i\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega\big) < \frac{1}{\kappa_2}\left(1 + \frac{1}{\kappa_3}\right)\eta_{ijk}$$

which proves (4.9).

**Step 3.** For $\underline{k} = \infty$, the estimates (4.8)–(4.9) imply that

$$\|\boldsymbol{u} - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} + \|p - P_{ij}\|_{\mathbb{P}} \overset{(4.7)}{\lesssim} \eta_{ijk} + \|\nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega \overset{(4.9)}{\simeq} \eta_{ijk} \xrightarrow{k\to\infty} 0.$$

Note that $\underline{k} = \infty$ also implies that neither $i$ nor $j$ are increased, *i.e.*, $P_{ij}$ remains constant as $k \to \infty$. Hence, $p = P_{ij} \in \mathbb{P}_i$ and therefore also $p = p_i$. ☐

**Lemma 4.5.** *Let $(i, 0, 0) \in \mathcal{Q}$ and define $\underline{j} := \max\{j \in \mathbb{N}_0 : (i, j, 0) \in \mathcal{Q}\} \in \mathbb{N}_0 \cup \{\infty\}$. If $0 < \kappa_3 \ll 1$ is sufficiently small (see (4.18) in the proof below), then there exist constants $0 < q_2 < 1$ and $C_2 > 0$ such that*

$$\|p_i - P_{i(j+n)}\|_{\mathbb{P}} \le q_2^n \, \|p_i - P_{ij}\|_{\mathbb{P}} \quad \textit{for all } j, n \in \mathbb{N}_0 \textit{ with } j \le j + n \le \underline{j}. \tag{4.10}$$

*Moreover, it holds that*

$$C_2^{-1} \, \|p_i - P_{ij}\|_{\mathbb{P}} \le \eta_{ij\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega \le C_2 \, \|p_i - P_{ij}\|_{\mathbb{P}} \quad \textit{for all } 0 \le j < \underline{j}. \tag{4.11}$$

*If $\underline{j} = \infty$, this yields convergence $\|\boldsymbol{u} - \boldsymbol{U}_{ij\underline{k}}\|_{\mathbb{V}} + \|p - P_{ij}\|_{\mathbb{P}} \to 0$ as $j \to \infty$. While $q_2$ depends only on the domain $\Omega$, $\gamma$-shape regularity, $\kappa_1$, and $\kappa_3$, the constant $C_2$ depends additionally on $\kappa_2$.*

*Proof.* We split the proof into three steps.
**Step 1.** If $j < \underline{j}(i)$ and $k = \underline{k}(i, j)$, then necessarily $\underline{k}(i, j) < \infty$. The structure of Algorithm 3.7 implies that the condition in step (ii) is false, while the condition in step (iii) is true, *i.e.*,

$$\eta_{ij\underline{k}} + \|\Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega > \kappa_2 \left(\eta_{ij\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega\right) \quad \text{and} \quad \eta_{ij\underline{k}} \le \kappa_3 \, \|\Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega. \tag{4.12}$$

First, this proves that

$$\begin{aligned} \kappa_2 \left(\eta_{ij\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega\right) &< \eta_{ij\underline{k}} + \|\Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega \le (1 + \kappa_3) \, \|\Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega \\ &\le (1 + \kappa_3) \, \|\nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega \le (1 + \kappa_3) \left(\eta_{ij\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega\right). \end{aligned} \tag{4.13}$$

Second, reliability (4.5) gives that

$$\|\Pi_i \nabla \cdot \left(\boldsymbol{u}_{ij} - \boldsymbol{U}_{ij\underline{k}}\right)\|_\Omega \le \|\boldsymbol{u}_{ij} - \boldsymbol{U}_{ij\underline{k}}\|_{\mathbb{V}} \overset{(4.5)}{\le} C'_{\rm rel}(\kappa_1) \, \eta_{ijk} \overset{(4.12)}{\le} \kappa_3 C'_{\rm rel}(\kappa_1) \, \|\Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega. \tag{4.14}$$

The triangle inequality yields that

$$(1 - \kappa_3 C'_{\rm rel}(\kappa_1)) \, \|\Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega \overset{(4.14)}{\le} \|\Pi_i \nabla \cdot \boldsymbol{u}_{ij}\|_\Omega \overset{(4.14)}{\le} (1 + \kappa_3 C'_{\rm rel}(\kappa_1)) \|\Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega. \tag{4.15}$$

This leads us to

$$\begin{aligned} C_{\rm div}^{-1} \frac{1 - \kappa_3 C'_{\rm rel}(\kappa_1)}{1 + \kappa_3 C'_{\rm rel}(\kappa_1)} \|p_i - P_{ij}\|_{\mathbb{P}} &\overset{(3.9)}{\le} \frac{1 - \kappa_3 C'_{\rm rel}(\kappa_1)}{1 + \kappa_3 C'_{\rm rel}(\kappa_1)} \, \|\Pi_i \nabla \cdot \boldsymbol{u}_{ij}\|_\Omega \\ &\overset{(4.15)}{\le} (1 - \kappa_3 C'_{\rm rel}(\kappa_1)) \, \|\Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega \overset{(4.15)}{\le} \|\Pi_i \nabla \cdot \boldsymbol{u}_{ij}\|_\Omega \overset{(3.9)}{\le} \|p_i - P_{ij}\|_{\mathbb{P}}. \end{aligned} \tag{4.16}$$

If $\kappa_3 C'_{\rm rel}(\kappa_1) < 1$, the combination of (4.16) and (4.13) proves (4.11).

**Step 2.** Starting from $P_{ij}$, one step of the *exact* Uzawa iteration for the *reduced* Stokes problem (leading to the auxiliary quantity $p_{i(j+1)}$) guarantees the existence of some $0 < q_{\rm Uzawa} < 1$ such that the following contraction holds (see [24], Eq. (4.3)):

$$\|p_i - p_{i(j+1)}\|_{\mathbb{P}} \le q_{\rm Uzawa} \, \|p_i - P_{ij}\|_{\mathbb{P}} \quad \text{with} \quad p_{i(j+1)} = P_{ij} - \Pi_i \nabla \cdot \boldsymbol{u}_{ij}. \tag{4.17}$$

The contraction constant $q_{\mathrm{Uzawa}}$ is the norm of the operator from (1.4) with $\alpha = 1$. Indeed, the proof of (4.17) works exactly as in Appendix A if $S : \mathbb{P} \to \mathbb{P}$ is replaced by the operator $\Pi_i S : \mathbb{P}_i \to \mathbb{P}_i$. In particular, $q_{\mathrm{Uzawa}}$ does neither depend on $i$ nor on $j$. Since $P_{i(j+1)} = P_{ij} - \Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}$, we are thus led to

$$\begin{aligned}
\|p_i - P_{i(j+1)}\|_{\mathbb{P}} &\le \|p_i - p_{i(j+1)}\|_{\mathbb{P}} + \|p_{i(j+1)} - P_{i(j+1)}\|_{\mathbb{P}} \\
&\le q_{\mathrm{Uzawa}} \|p_i - P_{ij}\|_{\mathbb{P}} + \|\Pi_i \nabla \cdot (\boldsymbol{u}_{ij} - \boldsymbol{U}_{ij\underline{k}})\|_{\mathbb{P}} \\
&\overset{(4.14)}{\le} q_{\mathrm{Uzawa}} \|p_i - P_{ij}\|_{\mathbb{P}} + \kappa_3 C'_{\mathrm{rel}}(\kappa_1) \|\Pi_i \nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_{\Omega} \\
&\overset{(4.16)}{\le} \left( q_{\mathrm{Uzawa}} + \frac{\kappa_3 C'_{\mathrm{rel}}(\kappa_1)}{1 - \kappa_3 C'_{\mathrm{rel}}(\kappa_1)} \right) \|p_i - P_{ij}\|_{\mathbb{P}} =: q_2 \|p_i - P_{ij}\|_{\mathbb{P}}.
\end{aligned}$$

Let $0 < \kappa_3 \ll 1$ be sufficiently small, *i.e.*,

$$0 < \kappa_3 C'_{\mathrm{rel}}(\kappa_1) < 1 \quad \text{and} \quad 0 < q_2 := q_{\mathrm{Uzawa}} + \frac{\kappa_3 C'_{\mathrm{rel}}(\kappa_1)}{1 - \kappa_3 C'_{\mathrm{rel}}(\kappa_1)} < 1. \tag{4.18}$$

Then, induction proves that $\|p_i - P_{i(j+n)}\|_{\mathbb{P}} \le q_2^n \|p_i - P_{ij}\|_{\mathbb{P}}$ for every $j, n \in \mathbb{N}_0$ with $j \le j + n \le \underline{j}$. This proves (4.10).

**Step 3.** For $\underline{j} = \infty$, the estimates (4.10)–(4.11) imply that

$$\|\boldsymbol{u} - \boldsymbol{U}_{ij\underline{k}}\|_{\mathbb{V}} + \|p - P_{ij}\|_{\mathbb{P}} \overset{(4.7)}{\lesssim} \eta_{ij\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_{\Omega} \overset{(4.11)}{\simeq} \|p_i - P_{ij}\|_{\mathbb{P}} \xrightarrow{j \to \infty} 0.$$

This concludes the proof. ☐

Note that $\underline{i} := \max\{i \in \mathbb{N}_0 : (i, 0, 0) \in \mathcal{Q}\} < \infty$ in Algorithm 3.7 implies that either $\underline{j} := \underline{j}(\underline{i}) = \infty$ or $\underline{k}(\underline{i}, \underline{j}) = \infty$. According to Lemma 4.4 (for $\underline{k} = \infty$) and Lemma 4.5 (for $\underline{j} = \infty$), it only remains to analyze the case $\underline{i} = \infty$.

**Lemma 4.6.** *Let $\underline{i} := \max\{i \in \mathbb{N}_0 : (i, 0, 0) \in \mathcal{Q}\} \in \mathbb{N}_0 \cup \{\infty\}$. If $0 < \kappa_2 \ll 1$ is sufficiently small (see (4.24) in the proof below), then there exist constants $0 < q_3 < 1$ and $C_3 > 0$ such that*

$$\|p - P_{(i+n)\underline{j}}\|_{\mathbb{P}} \le q_3^n \|p - P_{i\underline{j}}\|_{\mathbb{P}} \quad \text{for all } i, n \in \mathbb{N}_0 \text{ with } i \le i + n \le \underline{i}. \tag{4.19}$$

*Moreover, it holds that*

$$C_3^{-1} \|p - P_{i\underline{j}}\|_{\mathbb{P}} \le \eta_{i\underline{j}\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{i\underline{j}\underline{k}}\|_{\Omega} \le C_3 \|p - P_{i\underline{j}}\|_{\mathbb{P}} \quad \text{for all } 0 \le i < \underline{i}. \tag{4.20}$$

*While $C_3$ depends only on the domain $\Omega$, $\gamma$-shape regularity, $\kappa_1$ and $\kappa_2$, the contraction constant $q_3$ depends additionally on $0 < \vartheta \le 1$. If $\underline{i} = \infty$, this yields convergence $\|\boldsymbol{u} - \boldsymbol{U}_{i\underline{j}\underline{k}}\|_{\mathbb{V}} + \|p - P_{i\underline{j}}\|_{\mathbb{P}} \to 0$ as $i \to \infty$.*

*Proof.* We split the proof into five steps.

**Step 1.** According to Algorithm 3.7, it holds that

$$\eta_{i\underline{j}\underline{k}} + \|\Pi_i \nabla \cdot \boldsymbol{U}_{i\underline{j}\underline{k}}\|_{\Omega} \le \kappa_2 \left( \eta_{i\underline{j}\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{i\underline{j}\underline{k}}\|_{\Omega} \right). \tag{4.21}$$

For $0 < \kappa_2 < 1$, this implies that

$$\eta_{i\underline{j}\underline{k}} + \|\Pi_i \nabla \cdot \boldsymbol{U}_{i\underline{j}\underline{k}}\|_{\Omega} \le \frac{\kappa_2}{1 - \kappa_2} \|\nabla \cdot \boldsymbol{U}_{i\underline{j}\underline{k}}\|_{\Omega}.$$

Recall that

$$\|\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega \le \|\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega + \|\nabla\cdot(\boldsymbol{u}_{i\underline{j}}-\boldsymbol{U}_{i\underline{j}\underline{k}})\|_\Omega \overset{(4.5)}{\le} \|\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega + C'_{\rm rel}(\kappa_1)\,\eta_{i\underline{j}\underline{k}}.$$

We abbreviate $C(\kappa_1,\kappa_2) := C'_{\rm rel}(\kappa_1)\,\kappa_2/(1-\kappa_2)$. For sufficiently small $0<\kappa_2\ll 1$ with $0<C(\kappa_1,\kappa_2)<1$, the combination of the last two estimates implies that $\|\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega \le (1-C(\kappa_1,\kappa_2))^{-1}\,\|\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega$. With

$$C'(\kappa_1,\kappa_2) := \frac{C(\kappa_1,\kappa_2)}{1-C(\kappa_1,\kappa_2)},$$

we are hence led to

$$\begin{aligned}\|\boldsymbol{u}_{i\underline{j}}-\boldsymbol{U}_{i\underline{j}\underline{k}}\|_{\mathbb{V}} &\overset{(4.5)}{\le} C'_{\rm rel}(\kappa_1)\left(\eta_{i\underline{j}\underline{k}} + \|\Pi_i\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega\right) \le C(\kappa_1,\kappa_2)\,\|\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega \\ &\le C'(\kappa_1,\kappa_2)\,\|\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega \overset{(3.10)}{\le} C'(\kappa_1,\kappa_2)\,\|p-P_{i\underline{j}}\|_{\mathbb{P}}.\end{aligned}\tag{4.22}$$

Conversely,

$$\begin{aligned}\|p-P_{i\underline{j}}\|_{\mathbb{P}} &\overset{(3.10)}{\le} C_{\rm div}\|\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega \le C_{\rm div}\big(\|\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega + \|\nabla\cdot(\boldsymbol{u}_{i\underline{j}}-\boldsymbol{U}_{i\underline{j}\underline{k}})\|_\Omega\big) \\ &\overset{(4.5)}{\le} \max\{1,C'_{\rm rel}(\kappa_1)\}\,C_{\rm div}\left(\|\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega + \eta_{i\underline{j}\underline{k}}\right).\end{aligned}$$

In particular, this proves (4.20).

**Step 2.** Recall from Step 1 that

$$\begin{aligned}\|\nabla\cdot(\boldsymbol{u}_{i\underline{j}}-\boldsymbol{U}_{i\underline{j}\underline{k}})\|_\Omega + \|\Pi_i\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega &\overset{(4.5)}{\le} \max\{1,C'_{\rm rel}(\kappa_1)\}\left(\eta_{i\underline{j}\underline{k}} + \|\Pi_i\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega\right) \\ &\overset{(4.22)}{\le} \max\{1,C'_{\rm rel}(\kappa_1)\}\,C'(\kappa_1,\kappa_2)\,\|p-P_{i\underline{j}}\|_{\mathbb{P}}.\end{aligned}\tag{4.23}$$

We hence observe that

$$\begin{aligned}\|p_i-P_{i\underline{j}}\|_{\mathbb{P}} &\overset{(3.9)}{\le} C_{\rm div}\|\Pi_i\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega \le C_{\rm div}\big(\|\Pi_i\nabla\cdot(\boldsymbol{u}_{i\underline{j}}-\boldsymbol{U}_{i\underline{j}\underline{k}})\|_\Omega + \|\Pi_i\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega\big) \\ &\overset{(4.23)}{\le} C_{\rm div}\max\{1,C'_{\rm rel}(\kappa_1)\}\,C'(\kappa_1,\kappa_2)\,\|p-P_{i\underline{j}}\|_{\mathbb{P}}.\end{aligned}$$

**Step 3.** From Algorithm 3.6, we obtain that

$$\vartheta\,\|\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega \le \|\Pi_{i+1}\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega.$$

According to (4.22), it holds that

$$\|\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega \le \|\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega + \|\nabla\cdot(\boldsymbol{u}_{i\underline{j}}-\boldsymbol{U}_{i\underline{j}\underline{k}})\|_\Omega \overset{(4.22)}{\le} (1+C(\kappa_1,\kappa_2))\,\|\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega,$$

as well as

$$\|\Pi_{i+1}\nabla\cdot(\boldsymbol{u}_{i\underline{j}}-\boldsymbol{U}_{i\underline{j}\underline{k}})\|_\Omega \le \|\boldsymbol{u}_{i\underline{j}}-\boldsymbol{U}_{i\underline{j}\underline{k}}\|_{\mathbb{V}} \overset{(4.22)}{\le} C'(\kappa_1,\kappa_2)\,\|\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega.$$

Combining the last three estimates, we see that

$$\begin{aligned}\|\Pi_{i+1}\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega &\ge \|\Pi_{i+1}\nabla\cdot\boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega - \|\Pi_{i+1}\nabla\cdot(\boldsymbol{u}_{i\underline{j}}-\boldsymbol{U}_{i\underline{j}\underline{k}})\|_\Omega \\ &\ge \Big(\frac{\vartheta}{1+C(\kappa_1,\kappa_2)} - C'(\kappa_1,\kappa_2)\Big)\,\|\nabla\cdot\boldsymbol{u}_{i\underline{j}}\|_\Omega.\end{aligned}$$

Recall the constant $C_{\mathrm{div}} \geq 1$ from (2.3). If $0 < \kappa_2 \ll 1$ is sufficiently small, it holds that $C''(\kappa_1, \kappa_2, \vartheta) := \big(\frac{\vartheta}{1+C(\kappa_1,\kappa_2)} - C'(\kappa_1,\kappa_2)\big)/C_{\mathrm{div}} > 0$. This implies that

$$\begin{aligned}\|p_{i+1} - P_{i\underline{j}}\|_{\mathbb{P}} &\overset{(3.9)}{\geq} \|\Pi_{i+1} \nabla \cdot \boldsymbol{u}_{i\underline{j}}\|_\Omega \geq \Big(\frac{\vartheta}{1 + C(\kappa_1,\kappa_2)} - C'(\kappa_1,\kappa_2)\Big) \|\nabla \cdot \boldsymbol{u}_{i\underline{j}}\|_\Omega \\ &\overset{(3.10)}{\geq} C''(\kappa_1,\kappa_2,\vartheta)\, \|p - P_{i\underline{j}}\|_{\mathbb{P}}.\end{aligned}$$

Together with the Pythagoras theorem, we are hence led to

$$\|p - p_{i+1}\|_{\mathbb{P}}^2 = \|p - P_{i\underline{j}}\|_{\mathbb{P}}^2 - \|p_{i+1} - P_{i\underline{j}}\|_{\mathbb{P}}^2 \leq (1 - C''(\kappa_1,\kappa_2,\vartheta)^2)\, \|p - P_{i\underline{j}}\|_{\mathbb{P}}^2.$$

**Step 4.** Combining Step 2 and Step 3, we obtain that

$$\begin{aligned}\|p - P_{(i+1)\underline{j}}\|_{\mathbb{P}}^2 &= \|p - p_{i+1}\|_{\mathbb{P}}^2 + \|p_{i+1} - P_{(i+1)\underline{j}}\|_{\mathbb{P}}^2 \\ &\leq \big(1 - C''(\kappa_1,\kappa_2,\vartheta)^2\big)\, \|p - P_{i\underline{j}}\|_{\mathbb{P}}^2 + C_{\mathrm{div}}^2 \max\{1, C'_{\mathrm{rel}}(\kappa_1)^2\}\, C'(\kappa_1,\kappa_2)^2\, \|p - P_{(i+1)\underline{j}}\|_{\mathbb{P}}^2.\end{aligned}$$

For sufficiently small $0 < \kappa_2 \ll 1$, *i.e.*,

$$C(\kappa_1,\kappa_2) = \frac{C'_{\mathrm{rel}}(\kappa_1)\kappa_2}{1-\kappa_2} < 1, \tag{4.24}$$

$$0 < C''(\kappa_1,\kappa_2,\vartheta) = \Big(\frac{\vartheta}{1 + C(\kappa_1,\kappa_2)} - \frac{C(\kappa_1,\kappa_2)}{1 - C(\kappa_1,\kappa_2)}\Big)\, C_{\mathrm{div}}^{-1}, \tag{4.25}$$

$$0 < q_3^2 := \frac{1 - C''(\kappa_1,\kappa_2,\vartheta)^2}{1 - C_{\mathrm{div}}^2 \max\{1, C'_{\mathrm{rel}}(\kappa_1)^2\}\, C'(\kappa_1,\kappa_2)^2} < 1, \tag{4.26}$$

we hence see that

$$\|p - P_{(i+1)\underline{j}}\|_{\mathbb{P}}^2 \leq q_3^2\, \|p - P_{i\underline{j}}\|_{\mathbb{P}}^2.$$

By induction, we conclude (4.19).

**Step 5.** For $\underline{i} = \infty$, the estimates (4.19)–(4.20) imply that

$$\|\boldsymbol{u} - \boldsymbol{U}_{i\underline{j}\underline{k}}\|_{\mathbb{V}} + \|p - P_{i\underline{j}}\|_{\mathbb{P}} \overset{(4.7)}{\lesssim} \eta_{i\underline{j}\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{i\underline{j}\underline{k}}\|_\Omega \overset{(4.20)}{\lesssim} \|p - P_{i\underline{j}}\|_{\mathbb{P}} \xrightarrow{i\to\infty} 0.$$

This concludes the proof. □

### 4.3. Proof of Theorem 4.1

To prove Theorem 4.1, we need the following two lemmas. A slightly weaker version of the first lemma is already proved in Lemma 4.9 from [10]. The elementary proof, however, immediately extends to the following generalization and is therefore omitted. The second lemma states certain quasi-monotonicities for the output of the adaptive algorithm.

**Lemma 4.7.** *Let $(a_\ell)_{\ell\in\mathbb{N}_0}$ be a sequence with $a_\ell \geq 0$ for all $\ell \in \mathbb{N}_0$. With the convention $0^{-1/s} := \infty$, the following three statements are pairwise equivalent:*

(a) *There exist a constant $C > 0$ such that $\sum_{n=\ell}^\infty a_n \leq C a_\ell$ for all $\ell \in \mathbb{N}_0$.*
(b) *For all $s > 0$, there exists $C > 0$ such that $\sum_{n=0}^\ell a_n^{-1/s} \leq C a_\ell^{-1/s}$ for all $\ell \in \mathbb{N}_0$.*
(c) *There exist $0 < q < 1$ and $C > 0$ such that $a_{\ell+n} \leq C q^n a_\ell$ for all $n, \ell \in \mathbb{N}_0$.*

*Here, in each statement, the constants $C > 0$ may differ.*

**Lemma 4.8.** *Let* $0 < \kappa_1 < \theta^{1/2}/C_{\rm stab}$. *Suppose that* $\kappa_2, \kappa_3$ *are sufficiently small as in Lemmas 4.5 and 4.6. Let* $(i,j,0) \in \mathcal{Q}$. *Then, there hold the assertions* (a)–(d)*:*

(a) *If* $i \geq 1$, *then* $\eta_{i00} + \|\nabla \cdot \boldsymbol{U}_{i00}\|_\Omega \leq C_{\rm mon}\big(\eta_{(i-1)\underline{j}\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{(i-1)\underline{j}\underline{k}}\|_\Omega\big)$.
(b) *If* $j \geq 1$, *then* $\eta_{ij0} + \|\nabla \cdot \boldsymbol{U}_{ij0}\|_\Omega \leq C_{\rm mon}\big(\eta_{i(j-1)\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}\|_\Omega\big)$.
(c) $\eta_{ijk} + \|\nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega \leq C_{\rm mon}\big(\eta_{ijk'} + \|\nabla \cdot \boldsymbol{U}_{ijk'}\|_\Omega\big)$ *for all* $0 \leq k' \leq k \leq \underline{k}(i,j)$.
(d) $\eta_{ij\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{ij\underline{k}}\|_\Omega \leq C_{\rm mon}\big(\eta_{ij'\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{ij'\underline{k}}\|_\Omega\big)$ *for all* $0 \leq j' \leq j < \underline{j}(i)$.

*The constant* $C_{\rm mon} > 0$ *depends only on* $\Omega$, $C_{\rm stab}$, $C_{\rm rel}$, $C_1$, *and* $C_2$.

*Proof.* To shorten notation, we set $\eta^\star_{ijk} := \eta(\mathcal{T}_{ijk}; \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}], P_{ij})$ and $\boldsymbol{U}^\star_{ijk} := \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}]$. To prove (a), recall from step (ii) of Algorithm 3.7 that $\mathcal{T}_{i00} = \mathcal{T}_{(i-1)\underline{j}\underline{k}}$ as well as $P_{i0} = P_{(i-1)\underline{j}}$. Hence, $\boldsymbol{U}^\star_{i00} = \boldsymbol{U}^\star_{(i-1)\underline{j}\underline{k}}$ and consequently $\eta^\star_{i00} = \eta^\star_{(i-1)\underline{j}\underline{k}}$ as well as $\|\nabla \cdot \boldsymbol{U}^\star_{i00}\|_\Omega = \|\nabla \cdot \boldsymbol{U}^\star_{(i-1)\underline{j}\underline{k}}\|_\Omega$. Since $\kappa_1 < \theta^{1/2} C_{\rm stab}^{-1} \leq C_{\rm stab}^{-1}$, we can apply the equivalence (4.4) in both directions. With step (i) of Algorithm 3.7, we see that

$$\begin{aligned}
\eta_{i00} + \|\nabla \cdot \boldsymbol{U}_{i00}\|_\Omega &\overset{(4.4)}{\lesssim} \eta^\star_{i00} + \|\nabla \cdot \boldsymbol{U}^\star_{i00}\|_\Omega + \|\boldsymbol{U}^\star_{i00} - \boldsymbol{U}_{i00}\|_{\mathbb{V}} \lesssim \eta^\star_{i00} + \|\nabla \cdot \boldsymbol{U}^\star_{i00}\|_\Omega + \eta_{i00} \\
&\overset{(4.4)}{\lesssim} \eta^\star_{i00} + \|\nabla \cdot \boldsymbol{U}^\star_{i00}\|_\Omega = \eta^\star_{(i-1)\underline{j}\underline{k}} + \|\nabla \cdot \boldsymbol{U}^\star_{(i-1)\underline{j}\underline{k}}\|_\Omega \overset{(4.4)}{\lesssim} \eta_{(i-1)\underline{j}\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{(i-1)\underline{j}\underline{k}}\|_\Omega \\
&\quad + \|\boldsymbol{U}^\star_{(i-1)\underline{j}\underline{k}} - \boldsymbol{U}_{(i-1)\underline{j}\underline{k}}\|_{\mathbb{V}} \lesssim \eta_{(i-1)\underline{j}\underline{k}} + \|\nabla \cdot \boldsymbol{U}_{(i-1)\underline{j}\underline{k}}\|_\Omega.
\end{aligned}$$

To prove (b), recall from step (iii) of Algorithm 3.7 that $\mathcal{T}_{ij0} = \mathcal{T}_{i(j-1)\underline{k}}$ and $P_{ij} = P_{i(j-1)} - \Pi_i \nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}$. According to the discrete variational form (2.12), it holds that

$$a(\boldsymbol{U}^\star_{ij0} - \boldsymbol{U}^\star_{i(j-1)\underline{k}}, \boldsymbol{V}_{ij0}) = b(\boldsymbol{V}_{ij0}, \Pi_i \nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}) \quad \text{for all } \boldsymbol{V}_{ij0} \in \mathbb{V}(\mathcal{T}_{ij0}) = \mathbb{V}(\mathcal{T}_{(i-1)\underline{j}\underline{k}}).$$

This proves that $\|\boldsymbol{U}^\star_{ij0} - \boldsymbol{U}^\star_{i(j-1)\underline{k}}\|_{\mathbb{V}} \lesssim \|\Pi_i \nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}\|_\Omega \leq \|\nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}\|_\Omega$. First, it follows that

$$\begin{aligned}
\|\nabla \cdot \boldsymbol{U}_{ij0}\|_\Omega &\leq \|\nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}\|_\Omega + \|\boldsymbol{U}_{ij0} - \boldsymbol{U}_{i(j-1)\underline{k}}\|_{\mathbb{V}} \leq \|\nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}\|_\Omega + \|\boldsymbol{U}^\star_{ij0} - \boldsymbol{U}^\star_{i(j-1)\underline{k}}\|_{\mathbb{V}} \\
&\quad + \|\boldsymbol{U}^\star_{ij0} - \boldsymbol{U}_{ij0}\|_{\mathbb{V}} + \|\boldsymbol{U}^\star_{i(j-1)\underline{k}} - \boldsymbol{U}_{i(j-1)\underline{k}}\|_{\mathbb{V}} \leq \|\nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}\|_\Omega + \kappa_1\, \eta_{ij0} + \kappa_1\, \eta_{i(j-1)\underline{k}}.
\end{aligned}$$

Second, stability of the error estimator (Lem. 3.3), $\mathcal{T}_{ij0} = \mathcal{T}_{i(j-1)\underline{k}}$ and the previous estimate prove that

$$\begin{aligned}
\eta_{ij0} &\overset{(3.7)}{\leq} \eta_{i(j-1)\underline{k}} + C_{\rm stab}\big(\|\boldsymbol{U}_{ij0} - \boldsymbol{U}_{i(j-1)\underline{k}}\|_{\mathbb{V}} + \|\Pi_i \nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}\|_\Omega\big) \\
&\leq (1 + \kappa_1 C_{\rm stab})\, \eta_{i(j-1)\underline{k}} + C_{\rm stab}\, \|\nabla \cdot \boldsymbol{U}_{i(j-1)\underline{k}}\|_\Omega + \kappa_1 C_{\rm stab}\, \eta_{ij0}.
\end{aligned}$$

Recall that $\kappa_1 C_{\rm stab} < \theta^{1/2} \leq 1$. Thus, combining the last two estimates, we conclude the proof of (b).

To prove (c), note that Lemma 4.4 implies that

$$\eta_{ijk} \overset{(4.8)}{\leq} C_1\, \eta_{ijk'} \quad \text{for all } 0 \leq k' < k \leq \underline{k} := \underline{k}(i,j). \tag{4.27}$$

Moreover, the Pythagoras theorem, reliability (3.3), and the equivalence (4.4) prove that

$$\begin{aligned}
\|\nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega &\leq \|\nabla \cdot \boldsymbol{U}_{ijk'}\|_\Omega + \|\boldsymbol{U}^\star_{ijk} - \boldsymbol{U}^\star_{ijk'}\|_{\mathbb{V}} + \|\boldsymbol{U}^\star_{ijk} - \boldsymbol{U}_{ijk}\|_{\mathbb{V}} + \|\boldsymbol{U}^\star_{ijk'} - \boldsymbol{U}_{ijk'}\|_{\mathbb{V}} \\
&\leq \|\nabla \cdot \boldsymbol{U}_{ijk'}\|_\Omega + \|\boldsymbol{u}_{ij} - \boldsymbol{U}^\star_{ijk'}\|_{\mathbb{V}} + \kappa_1\, \eta_{ijk} + \kappa_1\, \eta_{ijk'} \\
&\overset{(3.3)+(4.27)}{\lesssim} \|\nabla \cdot \boldsymbol{U}_{ijk'}\|_\Omega + \eta^\star_{ijk'} + \eta_{ijk'} \\
&\overset{(4.4)}{\lesssim} \|\nabla \cdot \boldsymbol{U}_{ijk'}\|_\Omega + \eta_{ijk'}.
\end{aligned}$$

To prove (d), note that Lemma 4.5 implies that

$$\eta_{ij\underline{\underline{k}}} + \|\nabla \cdot \boldsymbol{U}_{ij\underline{\underline{k}}}\|_\Omega \overset{(4.11)}{\simeq} \|p_i - P_{ij}\|_{\mathbb{P}} \overset{(4.10)}{\leq} \|p_i - P_{ij'}\|_{\mathbb{P}} \overset{(4.11)}{\simeq} \eta_{ij'\underline{\underline{k}}} + \|\nabla \cdot \boldsymbol{U}_{ij'\underline{\underline{k}}}\|_\Omega.$$

This concludes the proof. □

*Proof of Theorem 4.1.* For all $0 \leq i' \leq i \leq \underline{i}$, define $\underline{\underline{j}}(i) \in \mathbb{N}_0$ by

$$\underline{\underline{j}}(i) := \begin{cases} 0 & \text{if } i' < i, \\ j' & \text{if } i' = i. \end{cases}$$

For all $0 \leq i' \leq i \leq \underline{i}$ and all $\underline{\underline{j}}(i) \leq j \leq \underline{j}(i)$, define $\underline{\underline{k}}(i,j) \in \mathbb{N}_0$ by

$$\underline{\underline{k}}(i,j) := \begin{cases} 0 & \text{if } i' < i \text{ or } j' < j, \\ k' & \text{if } i' = i \text{ and } j' = j. \end{cases}$$

As for $\underline{j}$ and $\underline{k}$, we write $\underline{\underline{j}} = \underline{\underline{j}}(i)$ and $\underline{\underline{k}} = \underline{\underline{k}}(i,j)$ if $i$ and $j$ are clear from the context. Further, we abbreviate

$$\mu_{ijk} := \eta_{ijk} + \|\nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega.$$

With this notation and according to Lemma 4.7, (4.2) is equivalent to

$$\sum_{\substack{(i,j,k)\in\mathcal{Q} \\ (i',j',k')\leq(i,j,k)}} \mu_{ijk} = \sum_{i=i'}^{\underline{i}} \sum_{j=\underline{\underline{j}}(i)}^{\underline{j}(i)} \sum_{k=\underline{\underline{k}}(i,j)}^{\underline{k}(i,j)} \mu_{ijk} \lesssim \mu_{i'j'k'} \quad \text{for all } (i',j',k') \in \mathcal{Q}. \tag{4.28}$$

We prove (4.28) in the following three steps.

**Step 1.** For $\underline{\underline{k}}(i,j) < \underline{k}(i,j) < \infty$, Lemma 4.8 (c) proves that $\mu_{ij\underline{k}} \lesssim \mu_{ij\underline{\underline{k}}}$ Hence, Lemma 4.4 in combination with the geometric series allows to estimate the sum over $k$

$$\begin{aligned} \sum_{i=i'}^{\underline{i}} \sum_{j=\underline{\underline{j}}(i)}^{\underline{j}(i)} \sum_{k=\underline{\underline{k}}(i,j)}^{\underline{k}(i,j)} \mu_{ijk} &\overset{(c)}{\lesssim} \sum_{i=i'}^{\underline{i}} \sum_{j=\underline{\underline{j}}(i)}^{\underline{j}(i)} \sum_{k=\underline{\underline{k}}(i,j)}^{\underline{k}(i,j)-1} \mu_{ijk} \overset{(4.9)}{\simeq} \sum_{i=i'}^{\underline{i}} \sum_{j=\underline{\underline{j}}(i)}^{\underline{j}(i)} \sum_{k=\underline{\underline{k}}(i,j)}^{\underline{k}(i,j)-1} \eta_{ijk} \overset{(4.8)}{\lesssim} \sum_{i=i'}^{\underline{i}} \sum_{j=\underline{\underline{j}}(i)}^{\underline{j}(i)} \eta_{ij\underline{\underline{k}}} \\ &\leq \sum_{i=i'}^{\underline{i}} \sum_{j=\underline{\underline{j}}(i)}^{\underline{j}(i)} \mu_{ij\underline{\underline{k}}} = \sum_{j=\underline{\underline{j}}(i')}^{\underline{j}(i')} \mu_{i'j\underline{\underline{k}}} + \sum_{i=i'+1}^{\underline{i}} \sum_{j=\underline{\underline{j}}(i)}^{\underline{j}(i)} \mu_{ij\underline{\underline{k}}} = \sum_{j=j'}^{\underline{j}(i')} \mu_{i'j\underline{\underline{k}}} + \sum_{i=i'+1}^{\underline{i}} \sum_{j=0}^{\underline{j}(i)} \mu_{ij0}. \end{aligned} \tag{4.29}$$

**Step 2.** In this step, we bound the first summand of (4.29) by $\mu_{i'j'k'}$. It holds that

$$\sum_{j=j'}^{\underline{j}(i')} \mu_{i'j\underline{\underline{k}}} = \mu_{i'j'\underline{\underline{k}}} + \sum_{j=j'+1}^{\underline{j}(i')} \mu_{i'j\underline{\underline{k}}} = \mu_{i'j'k'} + \sum_{j=j'+1}^{\underline{j}(i')} \mu_{i'j0}.$$

Lemmas 4.8(b) and 4.5 in combination with the geometric series show that

$$\sum_{j=j'+1}^{\underline{j}(i')} \mu_{i'j0} \overset{(b)}{\lesssim} \sum_{j=j'+1}^{\underline{j}(i')} \mu_{i'(j-1)\underline{k}} = \sum_{j=j'}^{\underline{j}(i')-1} \mu_{i'j\underline{k}} \overset{(4.11)}{\simeq} \sum_{j=j'}^{\underline{j}(i')-1} \|p_{i'} - P_{i'j}\|_{\mathbb{P}} \overset{(4.10)}{\lesssim} \|p_{i'} - P_{i'j'}\|_{\mathbb{P}} \overset{(3.4)}{\lesssim} \mu_{i'j'k'}.$$

**Step 3.** In this step, we bound the second summand of (4.29) by $\mu_{i'j'k'}$. First, we consider only the terms where $j > 0$. As in Step 2, Lemmas 4.8(b) and 4.5 in combination with the geometric series show that

$$\sum_{i=i'+1}^{\underline{i}} \sum_{j=1}^{\underline{j}(i)} \mu_{ij0} \overset{\text{(b)}}{\lesssim} \sum_{i=i'+1}^{\underline{i}} \sum_{j=1}^{\underline{j}(i)} \mu_{i(j-1)\underline{k}} = \sum_{i=i'+1}^{\underline{i}} \sum_{j=0}^{\underline{j}(i)-1} \mu_{ij\underline{k}} \overset{\text{Lem. 4.5}}{\lesssim} \sum_{i=i'+1}^{\underline{i}} \mu_{i0\underline{k}} \overset{\text{(c)}}{\lesssim} \sum_{i=i'+1}^{\underline{i}} \mu_{i00}.$$

Hence, it holds that

$$\sum_{i=i'+1}^{\underline{i}} \sum_{j=0}^{\underline{j}(i)} \mu_{ij0} = \sum_{i=i'+1}^{\underline{i}} \mu_{i00} + \sum_{i=i'+1}^{\underline{i}} \sum_{j=1}^{\underline{j}(i)} \mu_{ij0} \lesssim \sum_{i=i'+1}^{\underline{i}} \mu_{i00}.$$

Lemmas 4.8(a) and 4.6 in combination with the geometric series show that

$$\sum_{i=i'+1}^{\underline{i}} \mu_{i00} \overset{\text{(a)}}{\lesssim} \sum_{i=i'+1}^{\underline{i}} \mu_{(i-1)\underline{j}\underline{k}} = \sum_{i=i'}^{\underline{i}-1} \mu_{i\underline{j}\underline{k}} \overset{(4.20)}{\simeq} \sum_{i=i'}^{\underline{i}-1} \|p - P_{i\underline{j}}\|_{\mathbb{P}} \overset{(4.19)}{\lesssim} \|p - P_{i'\underline{j}}\|_{\mathbb{P}} \overset{(3.5)}{\lesssim} \mu_{i'\underline{j}\underline{k}}.$$

If $j' = \underline{j}(i')$, then Lemma 4.8(c) yields that $\mu_{i'\underline{j}\underline{k}} = \mu_{i'j'\underline{k}} \lesssim \mu_{i'j'k'}$. Otherwise, if $j' < \underline{j}(i')$, then Lemma 4.8(b)–(d) yield that

$$\mu_{i'\underline{j}\underline{k}} \overset{\text{(c)}}{\lesssim} \mu_{i'\underline{j}0} \overset{\text{(b)}}{\lesssim} \mu_{i'(\underline{j}-1)\underline{k}} \overset{\text{(d)}}{\lesssim} \mu_{i'j'\underline{k}} \overset{\text{(c)}}{\lesssim} \mu_{i'j'k'}.$$

Altogether, we have derived (4.28), which concludes the proof. $\square$

## 5. Convergence rates

### 5.1. Main theorem on optimal convergence rates

The first lemma relates two different characterizations of approximation classes from the literature, which are either based on the accuracy $\varepsilon > 0$ (see, *e.g.*, [24,28]) or the number of elements $N$ (see, *e.g.*, [10,11]).

**Lemma 5.1.** *Recall that* $\mathbb{T}^{\mathrm{c}} = \mathbb{T}^{\mathrm{c}}(\mathcal{T}_{\mathrm{init}})$. *Let* $\varrho : \mathbb{T}^{\mathrm{c}} \to \mathbb{R}_{\geq 0}$ *satisfy that* $\inf_{\mathcal{T}\in\mathbb{T}^{\mathrm{c}}} \varrho(\mathcal{T}) = 0$. *Let* $s > 0$ *and define*

$$\mathbb{A}_s^{\mathrm{c}}(\varrho) := \sup_{N\in\mathbb{N}_0} \Big( (N+1)^s \min_{\mathcal{T}\in\mathbb{T}_N^{\mathrm{c}}} \varrho(\mathcal{T}) \Big), \ \textit{where } \mathbb{T}_N^{\mathrm{c}} := \{\mathcal{T} \in \mathbb{T}^{\mathrm{c}} \,:\, \#\mathcal{T} - \#\mathcal{T}_{\mathrm{init}} \leq N\}. \tag{5.1}$$

*With* $\mathbb{T}_\varepsilon^{\mathrm{c}}(\varrho) := \{\mathcal{T} \in \mathbb{T}^{\mathrm{c}} \,:\, \varrho(\mathcal{T}) \leq \varepsilon\} \neq \emptyset$ *for* $\varepsilon > 0$, *there holds the equality*

$$\mathbb{A}_s^{\mathrm{c}}(\varrho) = \sup_{\varepsilon>0} \Big( \varepsilon \min_{\mathcal{T}\in\mathbb{T}_\varepsilon^{\mathrm{c}}(\varrho)} (\#\mathcal{T} - \#\mathcal{T}_{\mathrm{init}})^s \Big). \tag{5.2}$$

*The minimum in* (5.1) *exists, since all* $\mathbb{T}_N^{\mathrm{c}}$ *are finite sets. The minimum in* (5.2) *exists, since the cardinality is a mapping* $\# : \mathbb{T}^{\mathrm{nc}} \to \mathbb{N}$. *In either case, the minimizers might not be unique. If* $\mathbb{T}^{\mathrm{c}} = \mathbb{T}^{\mathrm{c}}(\mathcal{T}_{\mathrm{init}})$ *is replaced by* $\mathbb{T}^{\mathrm{nc}} = \mathbb{T}^{\mathrm{nc}}(\mathcal{T}_{\mathrm{init}})$, *one can define* $\mathbb{A}_s^{\mathrm{nc}}$, $\mathbb{T}_N^{\mathrm{nc}}$, *and* $\mathbb{T}_\varepsilon^{\mathrm{nc}}(\varrho)$ *similarly, and the assertion* (5.2) *holds accordingly.*

*Proof.* We only consider the set $\mathbb{T}^{\mathrm{c}}$ of conforming triangulations, the proof for the set $\mathbb{T}^{\mathrm{nc}}$ of non-conforming triangulations follows along the same lines. For $N \in \mathbb{N}_0$, define $\varepsilon_N := \min_{\mathcal{T}\in\mathbb{T}_N^{\mathrm{c}}} \varrho(\mathcal{T}) \geq 0$.

**Step 1.** To prove "$\geq$" in (5.2), let $\varepsilon > 0$. If $0 < \varepsilon < \varepsilon_0$, there exists a minimal $N \in \mathbb{N}_0$ such that $\min_{\mathcal{T}\in\mathbb{T}_N^{\mathrm{c}}} \varrho(\mathcal{T}) \leq \varepsilon$. In particular, it follows that $N > 0$, $\mathbb{T}_N^{\mathrm{c}} \cap \mathbb{T}_\varepsilon^{\mathrm{c}}(\varrho) \neq \emptyset$, and $\varepsilon < \min_{\mathcal{T}\in\mathbb{T}_{N-1}^{\mathrm{c}}} \varrho(\mathcal{T})$. This yields that

$$\varepsilon \min_{\mathcal{T}\in\mathbb{T}_\varepsilon^{\mathrm{c}}(\varrho)} (\#\mathcal{T} - \#\mathcal{T}_{\mathrm{init}})^s \leq \min_{\mathcal{T}\in\mathbb{T}_{N-1}^{\mathrm{c}}} \varrho(\mathcal{T})\, N^s \leq \sup_{N\in\mathbb{N}_0} \Big( (N+1)^s \min_{\mathcal{T}\in\mathbb{T}_N^{\mathrm{c}}} \varrho(\mathcal{T}) \Big) = \mathbb{A}_s^{\mathrm{c}}(\varrho). \tag{5.3}$$

If $\varepsilon_0 \le \varepsilon$, then $\mathcal{T}_{\rm init} \in \mathbb{T}^{\rm c}_{\varepsilon_0}(\varrho) \subseteq \mathbb{T}^{\rm c}_{\varepsilon}(\varrho)$ and hence the left-hand side of (5.3) is zero, and (5.3) thus remains true. Taking the supremum over all $\varepsilon > 0$, we prove "$\ge$" in (5.2).

**Step 2.** To prove "$\le$" in (5.2), let $N \in \mathbb{N}_0$. If $\varepsilon_N > 0$, the definition of $\varepsilon_N$ yields that $\#\mathcal{T} - \#\mathcal{T}_{\rm init} \ge N+1$ for all $\mathcal{T} \in \mathbb{T}^{\rm c}_{\lambda\varepsilon_N}(\varrho)$ and all $0 < \lambda < 1$. This proves that

$$(N+1)^s \min_{\mathcal{T}\in\mathbb{T}^{\rm c}_N} \varrho(\mathcal{T}) \le \min_{\mathcal{T}\in\mathbb{T}^{\rm c}_{\lambda\varepsilon_N}(\varrho)} (\#\mathcal{T} - \#\mathcal{T}_{\rm init})^s\, \varepsilon_N \le \frac{1}{\lambda} \sup_{\varepsilon>0} \big(\varepsilon \min_{\mathcal{T}\in\mathbb{T}^{\rm c}_{\varepsilon}(\varrho)} (\#\mathcal{T} - \#\mathcal{T}_{\rm init})^s\big). \tag{5.4}$$

If $\varepsilon_N = 0$, then the left-hand side of (5.4) is zero, and the overall estimate thus remains true. Taking the supremum over all $N \in \mathbb{N}_0$, we prove "$\le$" in (5.2) for the limit $\lambda \to 1$. $\square$

The following lemma specifies $\varrho(\mathcal{T})$ and hence introduces the precise approximation class of the present work.

**Lemma 5.2.** *For $s > 0$, let*

$$\mathbb{A}^{\rm c}_s := \mathbb{A}^{\rm c}_s(\varrho), \quad \textit{where} \quad \varrho(\mathcal{T}) := \eta(\mathcal{T}; \boldsymbol{U}_\mathcal{T}[p_\mathcal{T}], p_\mathcal{T}) + \|\nabla\cdot \boldsymbol{U}_\mathcal{T}[p_\mathcal{T}]\|_\Omega \quad \textit{for } \mathcal{T} \in \mathbb{T}^{\rm c}. \tag{5.5}$$

*Then, $\varrho$ satisfies the assumptions of Lemma 5.1. Moreover, there exists a constant $C > 0$, which depends only on $C_{\rm stab}$ and $C_{\rm rel}$, such that*

$$\varrho(\mathcal{T}) \le C \min_{Q_\mathcal{T}\in\mathbb{P}(\mathcal{T})} \big(\eta(\mathcal{T}; \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}], Q_\mathcal{T}) + \|\nabla\cdot \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}]\|_\Omega\big). \tag{5.6}$$

*Proof.* Let $Q_\mathcal{T} \in \mathbb{P}(\mathcal{T})$. According to (2.14), we have that $\|\boldsymbol{U}_\mathcal{T}[p_\mathcal{T}] - \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}]\|_\mathbb{V} \le \|p_\mathcal{T} - Q_\mathcal{T}\|_\mathbb{P}$. Since $p_\mathcal{T}$ is the best approximation of $p$ in $\mathbb{P}(\mathcal{T})$, it holds that $\|p_\mathcal{T} - Q_\mathcal{T}\|_\mathbb{P} \le \|p - Q_\mathcal{T}\|_\mathbb{P}$. Hence, stability (3.7) and reliability (3.5) of the error estimator prove that

$$\begin{aligned}
\varrho(\mathcal{T}) &= \eta(\mathcal{T}; \boldsymbol{U}_\mathcal{T}[p_\mathcal{T}], p_\mathcal{T}) + \|\nabla\cdot \boldsymbol{U}_\mathcal{T}[p_\mathcal{T}]\|_\Omega \\
&\overset{(3.7)}{\lesssim} \eta(\mathcal{T}; \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}], Q_\mathcal{T}) + \|\boldsymbol{U}_\mathcal{T}[p_\mathcal{T}] - \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}]\|_\mathbb{V} + \|p_\mathcal{T} - Q_\mathcal{T}\|_\mathbb{P} + \|\nabla\cdot \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}]\|_\Omega \\
&\lesssim \eta(\mathcal{T}; \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}], Q_\mathcal{T}) + \|\nabla\cdot \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}]\|_\Omega + \|p - Q_\mathcal{T}\|_\mathbb{P}. \\
&\overset{(3.5)}{\lesssim} \eta(\mathcal{T}; \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}], Q_\mathcal{T}) + \|\nabla\cdot \boldsymbol{U}_\mathcal{T}[Q_\mathcal{T}]\|_\Omega.
\end{aligned}$$

This proves (5.6). With linear convergence (Thm. 4.1), this yields that

$$\inf_{\mathcal{T}\in\mathbb{T}^{\rm c}} \varrho(\mathcal{T}) \le \inf_{(i,j,k)\in\mathcal{Q}} \varrho(\mathcal{T}_{ijk}) \lesssim \inf_{(i,j,k)\in\mathcal{Q}} \big(\eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega\big) = 0.$$

This concludes the proof. $\square$

Together with Theorem 4.1, the following theorem is the main result of this work. It states optimal convergence of Algorithm 3.7. The proof is given in Section 5.2.

**Theorem 5.3.** *Let $0 < \vartheta < C_{\rm div}^{-1}$ and $0 < \theta < \theta_{\rm opt} := (1 + C_{\rm stab}^2 C_{\rm drel}^2)^{-1}$. Suppose that*

$$\kappa_1 < \theta^{1/2} C_{\rm stab} \quad \textit{and} \quad \theta < \sup_{\delta>0} \frac{(1-\kappa_1 C_{\rm stab})^2\theta_{\rm opt} - (1+\delta^{-1})\kappa_1^2 C_{\rm stab}^2}{1+\delta}, \tag{5.7}$$

*i.e., $0 \le \kappa_1 < 1$ is sufficiently small. Moreover, let $0 < \kappa_2, \kappa_3 < 1$ be sufficiently small in the sense of Lemmas 4.5, 4.6, and 5.6 below. Then, for all $s > 0$, it holds that*

$$\mathbb{A}^{\rm c}_s < \infty \quad \Longleftrightarrow \quad \sup_{(i,j,k)\in\mathcal{Q}} \big(\eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega\big)\big(\#\mathcal{T}_{ijk} - \#\mathcal{T}_{\rm init} + 1\big)^s < \infty. \tag{5.8}$$

The following remark relates our definition of the approximation class from Lemma 5.2 to that of the so-called total error. We refer to Appendix C for the proof.

**Remark 5.4.** (i) The seminal work [24] employs two approximation classes:

- $\mathbb{A}_s^{\mathrm{c}}(\boldsymbol{u}) := \mathbb{A}_s^{\mathrm{c}}(\varrho_{\boldsymbol{u}})$ for $\varrho_{\boldsymbol{u}}(\mathcal{T}) := \min_{\boldsymbol{V}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T})} \|\boldsymbol{u} - \boldsymbol{V}_{\mathcal{T}}\|_{\mathbb{V}}$.
- $\mathbb{A}_s^{\mathrm{nc}}(p) := \mathbb{A}_s^{\mathrm{nc}}(\varrho_p)$ for $\varrho_p(\mathcal{P}) := \min_{Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})} \|p - Q_{\mathcal{P}}\|_{\mathbb{P}} = \|p - p_{\mathcal{P}}\|_{\mathbb{P}}$.

With the data oscillations for any $\mathcal{P} \in \mathbb{T}^{\mathrm{nc}}$, $\mathrm{osc}^2 := \sum_{T \in \mathcal{P}} \mathrm{osc}_T^2$ where $\mathrm{osc}_T^2 := |T|^{2/n} \, \|(1 - \Pi_{\mathcal{P}})\boldsymbol{f}\|_T^2$ for all $T \in \mathcal{P}$, we additionally define the approximation class:

- $\mathbb{A}_s^{\mathrm{nc}}(\boldsymbol{f}) := \mathbb{A}_s^{\mathrm{nc}}(\varrho_{\boldsymbol{f}})$ for $\varrho_{\boldsymbol{f}}(\mathcal{P}) := \mathrm{osc}(\mathcal{P})$.

Clearly, the definitions of $\varrho_p$, $\varrho_{\boldsymbol{u}}$, and $\varrho_{\boldsymbol{f}}$ satisfy the assumptions of Lemma 5.1. Moreover,

$$\mathbb{A}_s^{\mathrm{nc}}(p) \simeq \mathbb{A}_s^{\mathrm{c}}(p) := \mathbb{A}_s^{\mathrm{c}}(\varrho_p) \quad \text{and} \quad \mathbb{A}_s^{\mathrm{nc}}(\boldsymbol{f}) \simeq \mathbb{A}_s^{\mathrm{c}}(\boldsymbol{f}) := \mathbb{A}_s^{\mathrm{c}}(\varrho_{\boldsymbol{f}}). \tag{5.9}$$

(ii) If we additionally define

- $\mathbb{A}_s^{\mathrm{c}}(\boldsymbol{u}, p, \boldsymbol{f}) := \mathbb{A}_s^{\mathrm{c}}(\varrho_{\boldsymbol{u},p,\boldsymbol{f}})$ for $\varrho_{\boldsymbol{u},p,\boldsymbol{f}}(\mathcal{T}) := \varrho_{\boldsymbol{u}}(\mathcal{T}) + \varrho_p(\mathcal{T}) + \varrho_{\boldsymbol{f}}(\mathcal{T})$,

then it holds for all $s > 0$ that

$$\frac{1}{3}\big(\mathbb{A}_s^{\mathrm{c}}(\boldsymbol{u}) + \mathbb{A}_s^{\mathrm{c}}(p) + \mathbb{A}_s^{\mathrm{c}}(\boldsymbol{f})\big) \le \mathbb{A}_s^{\mathrm{c}}(\boldsymbol{u}, p, \boldsymbol{f}) \le 3^s \big(\mathbb{A}_s^{\mathrm{c}}(\boldsymbol{u}) + \mathbb{A}_s^{\mathrm{c}}(p) + \mathbb{A}_s^{\mathrm{c}}(\boldsymbol{f})\big). \tag{5.10}$$

In the literature, *cf.* [10,11], the term $\varrho_{\boldsymbol{u},p,\boldsymbol{f}}(\mathcal{T})$ is usually referred to as *total error.*
(iii) There hold efficiency and reliability in the sense that

$$\mathbb{A}_s^{\mathrm{c}} \lesssim \mathbb{A}_s^{\mathrm{c}}(\boldsymbol{u}, p, \boldsymbol{f}) \le C_{\mathrm{rel}}\, \mathbb{A}_s^{\mathrm{c}}, \tag{5.11}$$

*i.e.*, our approximation class coincides with the one of the total error. In particular, if the volume force $\boldsymbol{f}$ is a $\mathcal{T}_{\mathrm{init}}$-piecewise polynomial of degree less or equal than $m-1$, the oscillations vanish and our approximation class also coincides with that of Section 7 from [24].

(iv) Note that for smooth $\boldsymbol{u}, p$, and $\boldsymbol{f}$ and uniform mesh-refinement, one expects an optimal algebraic convergence rate of $s = m/d$. For non-smooth data and adaptive mesh-refinement, the involved approximation classes can be characterized in terms of Besov regularity; see, *e.g.*, [7,18,19].

## 5.2. Proof of Theorem 5.3

We start with an auxiliary lemma, which was originally proved in Lemma 6.3 from [24].

**Lemma 5.5.** *Let $0 < \vartheta < \vartheta' < C_{\mathrm{div}}^{-1}$. Let $0 < \omega < 1$ be sufficiently small such that*

$$0 < q := C_{\mathrm{div}} \frac{\omega + \vartheta'}{1 - \omega} < 1, \tag{5.12}$$

*Let $\mathcal{P} \in \mathbb{T}^{\mathrm{nc}}$ and $\mathcal{T} \in \mathbb{T}^{\mathrm{c}}(\mathcal{P})$. Let $Q_{\mathcal{P}} \in \mathbb{P}(\mathcal{P})$. Let $\boldsymbol{V}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T})$ satisfy that*

$$\|\nabla \cdot (\boldsymbol{u}[Q_{\mathcal{P}}] - \boldsymbol{V}_{\mathcal{T}})\|_\Omega \le \omega \, \|\nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega. \tag{5.13}$$

*Then,* $\texttt{binev}(\mathcal{P}, \mathcal{T}, \boldsymbol{V}_{\mathcal{T}}; \vartheta)$ *from Algorithm 3.6 returns $\mathcal{P}' \in \mathbb{T}^{\mathrm{nc}}(\mathcal{P})$ such that the following implication is satisfied for all $\overline{\mathcal{P}} \in \mathbb{T}^{\mathrm{nc}}(\mathcal{P})$*

$$\|p - p_{\overline{\mathcal{P}}}\|_{\mathbb{P}}^2 \le (1 - q^2)\, \|p - Q_{\mathcal{P}}\|_{\mathbb{P}}^2 \quad \Longrightarrow \quad \#\mathcal{P}' - \#\mathcal{P} \le C_{\mathrm{bin}}\, (\#\overline{\mathcal{P}} - \#\mathcal{T}_{\mathrm{init}}). \tag{5.14}$$

*Proof.* To see (5.14), let $\overline{\mathcal{P}} \in \mathbb{T}^{\rm nc}(\mathcal{P})$ with $\|p - p_{\overline{\mathcal{P}}}\|_{\mathbb{P}}^2 \le (1-q^2)\,\|p - Q_{\mathcal{P}}\|_{\mathbb{P}}^2$. Note that

$$\|p - p_{\widetilde{\mathcal{P}}}\|_{\mathbb{P}}^2 \le \|p - p_{\overline{\mathcal{P}}}\|_{\mathbb{P}}^2 \le (1-q^2)\,\|p - Q_{\mathcal{P}}\|_{\mathbb{P}}^2, \quad \text{where} \quad \widetilde{\mathcal{P}} := \mathcal{P} \oplus \overline{\mathcal{P}} \in \mathbb{T}^{\rm nc}(\mathcal{P}). \tag{5.15}$$

The triangle inequality and assumption (5.13) show that

$$\|\nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega \le \|\nabla \cdot u[Q_{\mathcal{P}}]\|_\Omega + \|\nabla \cdot (u[Q_{\mathcal{P}}] - V_{\mathcal{T}})\|_\Omega \overset{(5.13)}{\le} \|\nabla \cdot \boldsymbol{u}[Q_{\mathcal{P}}]\|_\Omega + \omega\,\|\nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega.$$

Hence, Lemma 3.5 yields that

$$\begin{aligned} q^2(1-\omega)^2\,\|\nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega^2 &\le q^2\,\|\nabla \cdot \boldsymbol{u}[Q_{\mathcal{P}}]\|_\Omega^2 \\ &\overset{(3.10)}{\le} q^2\,\|p - Q_{\mathcal{P}}\|_{\mathbb{P}}^2 \overset{(5.15)}{\le} \|p - Q_{\mathcal{P}}\|_{\mathbb{P}}^2 - \|p - p_{\widetilde{\mathcal{P}}}\|_{\mathbb{P}}^2 = \|p_{\widetilde{\mathcal{P}}} - Q_{\mathcal{P}}\|_{\mathbb{P}}^2 \overset{(3.9)}{\le} C_{\rm div}^2 \|\Pi_{\widetilde{\mathcal{P}}} \nabla \cdot \boldsymbol{u}[Q_{\mathcal{P}}]\|_\Omega^2. \end{aligned}$$

The triangle inequality together with (5.13) shows that

$$\|\Pi_{\widetilde{\mathcal{P}}} \nabla \cdot \boldsymbol{u}[Q_{\mathcal{P}}]\|_\Omega \le \|\Pi_{\widetilde{\mathcal{P}}} \nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega + \|\Pi_{\widetilde{\mathcal{P}}} \nabla \cdot (\boldsymbol{u}[Q_{\mathcal{P}}] - \boldsymbol{V}_{\mathcal{T}})\|_\Omega \overset{(5.13)}{\le} \|\Pi_{\widetilde{\mathcal{P}}} \nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega + \omega\,\|\nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega.$$

Altogether, we derive that

$$q(1-\omega)\,\|\nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega \le C_{\rm div} \|\Pi_{\widetilde{\mathcal{P}}} \nabla \cdot \boldsymbol{u}[Q_{\mathcal{P}}]\|_\Omega \le C_{\rm div}\big(\|\Pi_{\widetilde{\mathcal{P}}} \nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega + \omega\,\|\nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega\big).$$

By choice of $q$ in (5.12), this is equivalent to

$$\vartheta'\,\|\nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega = \frac{q(1-\omega) - C_{\rm div}\,\omega}{C_{\rm div}}\,\|\nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega \le \|\Pi_{\widetilde{\mathcal{P}}} \nabla \cdot \boldsymbol{V}_{\mathcal{T}}\|_\Omega.$$

By definition, Algorithm 3.6 returns $\mathcal{P}' \in \mathbb{T}^{\rm nc}(\mathcal{P})$ such that

$$\#\mathcal{P}' - \#\mathcal{P} \le C_{\rm bin}\,(\#\widetilde{\mathcal{P}} - \#\mathcal{P}) \overset{(M1)}{\le} C_{\rm bin}\,(\#\overline{\mathcal{P}} - \#\mathcal{T}_{\rm init}).$$

This concludes the proof. □

The heart of the proof of Theorem 4.1 is the following auxiliary lemma.

**Lemma 5.6.** *Let $(i,j,k) \in \mathcal{Q}$ with $k < \underline{k}(i,j)$ and $s > 0$. Let $0 < \vartheta < C_{\rm div}^{-1}$ and $0 < \theta < \theta_{\rm opt} = (1 + C_{\rm stab}^2 C_{\rm drel}^2)^{-1}$. Let $0 \le \kappa_1 < 1$ be sufficiently small such that (5.7) is satisfied. For sufficiently small $0 < \kappa_2 \ll 1$ (see (5.24) in the proof below), there exists $C_{\rm comp}$ such that*

$$\#\mathcal{M}_{ijk} \le C_{\rm comp}(1 + (\mathbb{A}_s^{\rm c})^{1/s})\big(\eta_{ijk} + \|\nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega\big)^{-1/s}. \tag{5.16}$$

*The constant $C_{\rm comp} > 0$ depends only on the domain $\Omega$, $\gamma$-shape regularity, the polynomial degree $m$, the parameters $\kappa_1, \kappa_2, \kappa_3, \vartheta, \theta$ $C_{\rm mark}$, and $s$.*

*Proof.* The proof is split into five steps.
**Step 1.** Choose

$$\varepsilon := \eta_{ijk} + \|\nabla \cdot \boldsymbol{U}_{ijk}\|_\Omega. \tag{5.17}$$

Without loss of generality, we may assume that $\varepsilon > 0$ and $\mathbb{A}_s^{\rm c} < \infty$. Then, Lemmas 5.1 and 5.2 guarantee the existence of $\overline{\mathcal{T}} \in \mathbb{T}^{\rm c}$ such that

$$\#\overline{\mathcal{T}} - \#\mathcal{T}_{\rm init} \le (\mathbb{A}_s^{\rm c}/\varepsilon)^{1/s} \quad \text{and} \quad \eta(\overline{\mathcal{T}}; \boldsymbol{U}_{\overline{\mathcal{T}}}[p_{\overline{\mathcal{T}}}], p_{\overline{\mathcal{T}}}) + \|\nabla \cdot \boldsymbol{U}_{\overline{\mathcal{T}}}[p_{\overline{\mathcal{T}}}]\|_\Omega \le \varepsilon. \tag{5.18}$$

**Step 2.** Define the uniformly refined triangulations

$$\widehat{\mathcal{T}}_0 := \texttt{close}(\mathcal{P}_i) \oplus \overline{\mathcal{T}} \quad \text{and} \quad \widehat{\mathcal{T}}_{n+1} := \texttt{refine}(\widehat{\mathcal{T}}_n, \widehat{\mathcal{T}}_n) \quad \text{for all } n \in \mathbb{N}_0.$$

Note that $P_{ij} \in \mathbb{P}(\mathcal{P}_i) \subseteq \mathbb{P}(\widehat{\mathcal{T}}_n)$. We recall some standard arguments for adaptive mesh-refinement for the (vector-valued) Poisson model problem. Reliability (3.3), stability (3.7), and reduction (3.8) guarantee the existence of $C_{\rm ctr} > 0$ and $0 < q_{\rm ctr} < 1$ such that

$$\eta(\widehat{\mathcal{T}}_n; \boldsymbol{U}_{\widehat{\mathcal{T}}_n}[P_{ij}], P_{ij}) \leq C_{\rm ctr}\, q_{\rm ctr}^n\, \eta(\widehat{\mathcal{T}}_0; \boldsymbol{U}_{\widehat{\mathcal{T}}_0}[P_{ij}], P_{ij});$$

see, *e.g.*, Theorem 4.1 (i) of [10]. According to, *e.g.*, Section 3.4 of [10], there exists $C'_{\rm mon} > 0$ such that for all $\widehat{\mathcal{T}} \in \mathbb{T}^{\rm c}$, $\widehat{\mathcal{T}}' \in \mathbb{T}^{\rm c}(\widehat{\mathcal{T}})$, $P_{\widehat{\mathcal{T}}} \in \mathbb{P}(\widehat{\mathcal{T}})$

$$\eta(\widehat{\mathcal{T}}'; \boldsymbol{U}_{\widehat{\mathcal{T}}'}[P_{\widehat{\mathcal{T}}}], P_{\widehat{\mathcal{T}}}) \leq C'_{\rm mon}\eta(\widehat{\mathcal{T}}; \boldsymbol{U}_{\widehat{\mathcal{T}}}[P_{\widehat{\mathcal{T}}}], P_{\widehat{\mathcal{T}}}) \tag{5.19}$$

Note that $C_{\rm ctr}$, $q_{\rm ctr}$, and $C'_{\rm mon}$ depend only on $\gamma$-shape regularity and the polynomial degree $m$. With stability (3.7) and quasi-monotonicity (5.19), it follows that

$$\begin{aligned}
\eta(\widehat{\mathcal{T}}_n; \boldsymbol{U}_{\widehat{\mathcal{T}}_n}[P_{ij}], P_{ij}) &\leq C_{\rm ctr}\, q_{\rm ctr}^n\, \eta(\widehat{\mathcal{T}}_0; \boldsymbol{U}_{\widehat{\mathcal{T}}_0}[P_{ij}], P_{ij}) \\
&\overset{(3.7)}{\leq} C_{\rm ctr}\, q_{\rm ctr}^n \left[\eta(\widehat{\mathcal{T}}_0; \boldsymbol{U}_{\widehat{\mathcal{T}}_0}[p_{\overline{\mathcal{T}}}], p_{\overline{\mathcal{T}}}) + C_{\rm stab}\big(\|\boldsymbol{U}_{\widehat{\mathcal{T}}_0}[P_{ij}] - \boldsymbol{U}_{\widehat{\mathcal{T}}_0}[p_{\overline{\mathcal{T}}}]\|_{\mathbb{V}} + \|P_{ij} - p_{\overline{\mathcal{T}}}\|_{\mathbb{P}}\big)\right] \\
&\overset{(5.19)}{\leq} C_{\rm ctr}\, q_{\rm ctr}^n \left[C'_{\rm mon}\, \eta(\overline{\mathcal{T}}; \boldsymbol{U}_{\overline{\mathcal{T}}}[p_{\overline{\mathcal{T}}}], p_{\overline{\mathcal{T}}}) + C_{\rm stab}\big(\|\boldsymbol{U}_{\widehat{\mathcal{T}}_0}[P_{ij}] - \boldsymbol{U}_{\widehat{\mathcal{T}}_0}[p_{\overline{\mathcal{T}}}]\|_{\mathbb{V}} + \|P_{ij} - p_{\overline{\mathcal{T}}}\|_{\mathbb{P}}\big)\right].
\end{aligned}$$

With (2.14), we hence obtain that

$$\eta(\widehat{\mathcal{T}}_n; \boldsymbol{U}_{\widehat{\mathcal{T}}_n}[P_{ij}], P_{ij}) \leq C_{\rm ctr}\, q_{\rm ctr}^n \left[C'_{\rm mon}\, \eta(\overline{\mathcal{T}}; \boldsymbol{U}_{\overline{\mathcal{T}}}[p_{\overline{\mathcal{T}}}], p_{\overline{\mathcal{T}}}) + 2C_{\rm stab}\, \|P_{ij} - p_{\overline{\mathcal{T}}}\|_{\mathbb{P}}\right].$$

According to the reliability estimates (3.5) and (4.7), it holds that

$$\begin{aligned}
\|P_{ij} - p_{\overline{\mathcal{T}}}\|_{\mathbb{P}} &\leq \|p - p_{\overline{\mathcal{T}}}\|_{\mathbb{P}} + \|p - P_{ij}\|_{\mathbb{P}} \\
&\leq C'_{\rm rel}(\kappa_1)\left\{\big(\eta(\overline{\mathcal{T}}; \boldsymbol{U}_{\overline{\mathcal{T}}}[p_{\overline{\mathcal{T}}}], p_{\overline{\mathcal{T}}}) + \|\nabla\cdot \boldsymbol{U}_{\overline{\mathcal{T}}}[p_{\overline{\mathcal{T}}}]\|_\Omega\big) + \big(\eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega\big)\right\}.
\end{aligned}$$

By choice of $\overline{\mathcal{T}}$ in Step 1 and for $k < \underline{k}(i,j)$, we overall obtain that

$$\begin{aligned}
\eta(\widehat{\mathcal{T}}_n; \boldsymbol{U}_{\widehat{\mathcal{T}}_n}[P_{ij}], P_{ij}) &\leq q_{\rm ctr}^n\, C_{\rm ctr}\left[C'_{\rm mon} + 4\,C_{\rm stab}\,C'_{\rm rel}(\kappa_1)\right]\big(\eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega\big) \\
&\overset{(4.9)}{\leq} q_{\rm ctr}^n\, C_{\rm ctr}\left[C'_{\rm mon} + 4\,C_{\rm stab}\,C'_{\rm rel}(\kappa_1)\right]\frac{1}{\kappa_2}\left(1 + \frac{1}{\kappa_3}\right)\eta_{ijk}.
\end{aligned} \tag{5.20}$$

**Step 3.** To shorten notation, we set $\eta^\star_{ijk} := \eta(\mathcal{T}_{ijk}; \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}], P_{ij})$ and $\boldsymbol{U}^\star_{ijk} := \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}]$. Note that discrete reliability (3.6) and stability (3.7) imply optimality of Dörfler marking (see, *e.g.*, [10], Sect. 4.5): For any $0 < \theta_\star < \theta_{\rm opt}$, there exists some $0 < \lambda = \lambda(\theta_\star) \ll 1$ such that, for all $\check{\mathcal{T}} \in \mathbb{T}^{\rm c}(\mathcal{T}_{ijk})$, it holds that

$$\eta(\check{\mathcal{T}}; \boldsymbol{U}_{\check{\mathcal{T}}}[P_{ij}], P_{ij}) \leq \lambda\, \eta^\star_{ijk} \quad \Longrightarrow \quad \theta_\star\, (\eta^\star_{ijk})^2 \leq \eta(\mathcal{T}_{ijk} \setminus \check{\mathcal{T}}; \boldsymbol{U}_{ijk}, P_{ij})^2. \tag{5.21}$$

The second inequality in (5.21), Lemma 4.3, and the Young inequality imply for $\delta > 0$ that

$$\begin{aligned}
(1 - \kappa_1 C_{\rm stab})^2 \theta_\star\, \eta_{ijk}^2 &\overset{(4.4)}{\leq} \theta_\star (\eta^\star_{ijk})^2 \overset{(5.21)}{\leq} \eta(\mathcal{T}_{ijk} \setminus \check{\mathcal{T}}; \boldsymbol{U}^\star_{ijk}, P_{ij})^2 \\
&\overset{(4.3)}{\leq} (1 + \delta)\eta(\mathcal{T}_{ijk} \setminus \check{\mathcal{T}}; \boldsymbol{U}_{ijk}, P_{ij})^2 + (1 + \delta^{-1})\kappa_1^2 C_{\rm stab}^2 \eta_{ijk}^2.
\end{aligned}$$

Due to (5.7), we can choose $0 < \theta_\star < \theta_{\rm opt}$ sufficiently close to $\theta_{\rm opt}$ such that

$$\theta\,\eta_{ijk}^2 \overset{(5.7)}{\leq} \sup_{\delta>0} \frac{(1-\kappa_1 C_{\rm stab})^2\theta_\star - (1+\delta^{-1})\kappa_1^2 C_{\rm stab}^2}{1+\delta}\,\eta_{ijk}^2 \leq \eta(\mathcal{T}_{ijk}\setminus\check{\mathcal{T}};\boldsymbol{U}_{ijk},P_{ij})^2. \tag{5.22}$$

Let $\ell \in \mathbb{N}_0$ be the minimal integer such that

$$q_{\rm ctr}^\ell \frac{C'_{\rm mon}}{1-\kappa_1 C_{\rm stab}} C_{\rm ctr}\left[C'_{\rm mon} + 4C_{\rm stab}C'_{\rm rel}(\kappa_1)\right]\frac{1}{\kappa_2}\Big(1+\frac{1}{\kappa_3}\Big) \leq \lambda.$$

Recall $\widehat{\mathcal{T}}_\ell$ from Step 2. For $\check{\mathcal{T}} := \widehat{\mathcal{T}}_\ell \oplus \mathcal{T}_{ijk}$, it then holds that

$$\eta(\check{\mathcal{T}};\boldsymbol{U}_{\check{\mathcal{T}}}[P_{ij}],P_{ij}) \overset{(5.19)}{\leq} C'_{\rm mon}\,\eta(\widehat{\mathcal{T}}_\ell;\boldsymbol{U}_{\widehat{\mathcal{T}}_\ell}[P_{ij}],P_{ij}) \overset{(5.20)}{\leq} \lambda\,(1-\kappa_1 C_{\rm stab})\,\eta_{ijk} \overset{(4.4)}{\leq} \lambda\eta_{ijk}^\star.$$

Hence, (5.21) and (5.22) imply that $\theta\,\eta_{ijk}^2 \leq \eta(\mathcal{T}_{ijk}\setminus\check{\mathcal{T}};\boldsymbol{U}_{ijk},P_{ij})^2$.

**Step 4.** Since $\mathcal{M}_{ijk} \subseteq \mathcal{T}_{ijk}$ in Algorithm 3.7 (iv) has (up to some fixed factor $C_{\rm mark}$) minimal cardinality, the overlay estimate (M1) implies that

$$\begin{aligned} C_{\rm mark}^{-1}\#\mathcal{M}_{ijk} &\overset{(5.21)}{\leq} \#(\mathcal{T}_{ijk}\setminus\check{\mathcal{T}}) \leq \#\check{\mathcal{T}} - \#\mathcal{T}_{ijk} \overset{(\rm M1)}{\leq} \#\widehat{\mathcal{T}}_\ell - \#\mathcal{T}_{\rm init} \overset{(\rm M2)}{\leq} C_{\rm son}^\ell \#\widehat{\mathcal{T}}_0 \\ &\overset{(\rm M1)}{\leq} C_{\rm son}^\ell\big(\#\texttt{close}(\mathcal{P}_i) + \#\overline{\mathcal{T}} - \#\mathcal{T}_{\rm init}\big) \\ &\overset{(5.18)}{\lesssim} (\mathbb{A}_s^{\rm c})^{1/s}\big(\eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega\big)^{-1/s} + \#\texttt{close}(\mathcal{P}_i). \end{aligned}$$

Elementary calculation (see, *e.g.*, [3], Lem. 22) shows that

$$\#\mathcal{P} - \#\mathcal{T}_{\rm init} + 1 \leq \#\mathcal{P} \leq \#\mathcal{T}_{\rm init}\big(\#\mathcal{P} - \#\mathcal{T}_{\rm init} + 1\big) \quad \text{for all } \mathcal{P} \in \mathbb{T}^{\rm nc}.$$

With $\#\mathcal{T}_{\rm init} \simeq 1 \lesssim \big(\eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega\big)^{-1/s}$, the conformity estimate (M4) yields that

$$\#\texttt{close}(\mathcal{P}_i) \lesssim \big(\eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega\big)^{-1/s} + (\#\mathcal{P}_i - \#\mathcal{T}_{\rm init}).$$

Altogether, this step thus concludes that

$$\#\mathcal{M}_{ijk} \lesssim (1 + (\mathbb{A}_s^{\rm c})^{1/s})\big(\eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega\big)^{-1/s} + (\#\mathcal{P}_i - \#\mathcal{T}_{\rm init}). \tag{5.23}$$

**Step 5.** Reliability (4.5) as well as Algorithm 3.7 (ii) show for all $0 \leq i' < i$ that

$$\|\nabla\cdot(\boldsymbol{u}_{i'\underline{j}} - \boldsymbol{U}_{i'\underline{j}\underline{k}})\|_\Omega \leq \|\boldsymbol{u}_{i'\underline{j}} - \boldsymbol{U}_{i'\underline{j}\underline{k}}\|_{\mathbb{V}} \leq C'_{\rm rel}(\kappa_1)\,\eta_{i'\underline{j}\underline{k}} \leq C'_{\rm rel}(\kappa_1)\,\frac{\kappa_2}{1-\kappa_2}\,\|\nabla\cdot\boldsymbol{U}_{i'\underline{j}\underline{k}}\|_\Omega.$$

Let $0 < \vartheta < \vartheta' < C_{\rm div}^{-1}$ and $\omega := C'_{\rm rel}(\kappa_1)\kappa_2/(1-\kappa_2)$. For $0 < \kappa_2 \ll 1$ with

$$0 < q := C_{\rm div}\,\frac{\omega + \vartheta'}{1-\omega} < 1, \tag{5.24}$$

Lemma 5.5 applies and proves for all $\overline{\mathcal{P}}_{i'} \in \mathbb{T}^{\rm nc}(\mathcal{P}_{i'})$ that

$$\|p - p_{\overline{\mathcal{P}}_{i'}}\|_{\mathbb{P}} \leq (1-q^2)^{1/2}\,\|p - P_{i'\underline{j}}\|_{\mathbb{P}} \quad\Longrightarrow\quad \#\mathcal{P}_{i'+1} - \#\mathcal{P}_{i'} \lesssim \#\overline{\mathcal{P}}_{i'} - \#\mathcal{T}_{\rm init}.$$

We choose $\overline{\mathcal{P}}_{i'}$ from the definition (5.2) of the approximation norm $\mathbb{A}_s^{\rm c}$ such that

$$\#\overline{\mathcal{P}}_{i'} - \#\mathcal{T}_{\rm init} \le (\mathbb{A}_s^{\rm c}/\varepsilon_{i'})^{1/s} \quad \text{with} \quad \eta(\overline{\mathcal{P}}_{i'}; \boldsymbol{U}_{\overline{\mathcal{P}}_{i'}}[p_{\overline{\mathcal{P}}_{i'}}], p_{\overline{\mathcal{P}}_{i'}}) + \|\nabla\cdot \boldsymbol{U}_{\overline{\mathcal{P}}_{i'}}[p_{\overline{\mathcal{P}}_{i'}}]\|_\Omega \le \varepsilon_{i'} := \frac{(1-q^2)^{1/2}}{C'_{\rm rel}(\kappa_1)} \|p - P_{i'\underline{j}}\|_{\mathbb{P}}.$$

Reliability (3.5) shows that $\|p-p_{\overline{\mathcal{P}}_{i'}}\|_{\mathbb{P}} \le C_{\rm rel}\big(\eta(\overline{\mathcal{P}}_{i'}; \boldsymbol{U}_{\overline{\mathcal{P}}_{i'}}[p_{\overline{\mathcal{P}}_{i'}}], p_{\overline{\mathcal{P}}_{i'}}) + \|\nabla\cdot \boldsymbol{U}_{\overline{\mathcal{P}}_{i'}}[p_{\overline{\mathcal{P}}_{i'}}]\|_\Omega\big)$. With $C_{\rm rel} \le C'_{\rm rel}(\kappa_1)$, Lemmas 4.6 and 4.7(b) yield that

$$\#\mathcal{P}_i - \#\mathcal{T}_{\rm init} = \sum_{i'=0}^{i-1} (\#\mathcal{P}_{i'+1} - \#\mathcal{P}_{i'}) \lesssim (\mathbb{A}_s^{\rm c})^{1/s} \sum_{i'=0}^{i-1} \|p - P_{i'\underline{j}}\|_{\mathbb{P}}^{-1/s} \overset{\rm (b)}{\lesssim} (\mathbb{A}_s^{\rm c})^{1/s} \|p - P_{(i-1)\underline{j}}\|_{\mathbb{P}}^{-1/s}.$$

Next, we prove that $\|p - P_{(i-1)\underline{j}}\|_{\mathbb{P}}^{-1/s} \lesssim \big(\eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega\big)^{-1/s}$. To this end, we apply Lemmas 4.8(a)–(d) and 4.6. For $i, j > 0$, it holds that

$$\begin{aligned} \eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega &\overset{\rm (c)}{\lesssim} \eta_{ij0} + \|\nabla\cdot \boldsymbol{U}_{ij0}\|_\Omega \overset{\rm (b)}{\lesssim} \eta_{i(j-1)\underline{k}} + \|\nabla\cdot \boldsymbol{U}_{i(j-1)\underline{k}}\|_\Omega \overset{\rm (d)}{\lesssim} \eta_{i0\underline{k}} + \|\nabla\cdot \boldsymbol{U}_{i0\underline{k}}\|_\Omega \\ &\overset{\rm (c)}{\lesssim} \eta_{i00} + \|\nabla\cdot \boldsymbol{U}_{i00}\|_\Omega \overset{\rm (a)}{\lesssim} \eta_{(i-1)\underline{j}\underline{k}} + \|\nabla\cdot \boldsymbol{U}_{(i-1)\underline{j}\underline{k}}\|_\Omega \overset{(4.20)}{\simeq} \|p - P_{(i-1)\underline{j}}\|_{\mathbb{P}}. \end{aligned}$$

Note that the overall estimate is also true if $j = 0$. This proves that $\#\mathcal{P}_i - \#\mathcal{T}_{\rm init} \lesssim (\mathbb{A}_s^{\rm c})^{1/s}\big(\eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega\big)^{-1/s}$. With (5.23), we obtain that

$$\#\mathcal{M}_{ijk} \lesssim (1 + (\mathbb{A}_s^{\rm c})^{1/s})\big(\eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|_\Omega\big)^{-1/s}.$$

This concludes the proof. $\square$

*Proof of Theorem 5.3.* The proof is split into two steps.

**Step 1.** We show the lower bound in (5.8). Recall that $P_{ij} \in \mathbb{P}(\mathcal{P}_i) \subseteq \mathbb{P}(\mathcal{T}_{ijk})$ for all $(i,j,k) \in \mathcal{Q}$. Therefore, Lemma 5.2 gives that

$$\varrho(\mathcal{T}_{ijk}) \overset{(5.6)}{\lesssim} \eta(\mathcal{T}_{ijk}; \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}], P_{ij}) + \|\nabla\cdot \boldsymbol{U}_{\mathcal{T}_{ijk}}[P_{ij}]\|_\Omega \overset{(4.4)}{\simeq} \eta_{ijk} + \|\nabla\cdot \boldsymbol{U}_{ijk}\|. \tag{5.25}$$

If there exists some $(i,j,k) \in \mathcal{Q}$ such that $\mathcal{T}_{ijk} = \mathcal{T}_{i'j'k'}$ for all $(i',j',k') \in \mathcal{Q}$ with $(i,j,k) \le (i',j',k')$, then, $\varrho(\mathcal{T}_{i'j'k'}) = \varrho(\mathcal{T}_{ijk})$, (5.6), and convergence (4.2) yield that $\varrho(\mathcal{T}_{i'j'k'}) = 0$ and hence $\mathbb{A}_s^{\rm c} < \infty$. Otherwise, let $N \in \mathbb{N}_0$ and let $(i,j,k) \in \mathcal{Q}$ be the largest possible index (with respect to "$\le$") such that $\#\mathcal{T}_{ijk} - \#\mathcal{T}_{\rm init} \le N$, *i.e.*, $\mathcal{T}_{ijk} \in \mathbb{T}_N^{\rm c}$. Clearly, it holds that $k < \underline{k}(i,j)$. Therefore, the son estimate (M2) yields that

$$N + 1 < \#\mathcal{T}_{ij(k+1)} - \#\mathcal{T}_{\rm init} + 1 \simeq \#\mathcal{T}_{ij(k+1)} \overset{\rm (M2)}{\simeq} \#\mathcal{T}_{ijk} \simeq \#\mathcal{T}_{ijk} - \#\mathcal{T}_{\rm init} + 1.$$

Together with (5.25), this leads to

$$\min_{\mathcal{T} \in \mathbb{T}_N^{\rm c}} (N+1)^s \varrho(\mathcal{T}) \lesssim (\#\mathcal{T}_{ijk} - \#\mathcal{T}_{\rm init} + 1)^s \varrho(\mathcal{T}_{ijk}).$$

Taking the supremum over all $(i,j,k) \in \mathcal{Q}$, and then over all $N \in \mathbb{N}_0$, we conclude the first step.

**Step 2.** We show the upper bound in (5.8). According to the closure estimate (M3) and Lemma 5.6, it holds for all $(i',j',k') \in \mathcal{Q}$ with $\mathcal{T}_{i'j'k'} \neq \mathcal{T}_{\mathrm{init}}$ that

$$\begin{aligned}\#\mathcal{T}_{i'j'k'} - \#\mathcal{T}_{\mathrm{init}} + 1 \simeq \#\mathcal{T}_{i'j'k'} - \#\mathcal{T}_{\mathrm{init}} &\overset{\text{(M3)}}{\lesssim} \sum_{\substack{(i,j,k)\leq(i',j',k')\\ k\neq \underline{k}(i,j)}} \#\mathcal{M}_{ijk} \\ &\overset{(5.16)}{\lesssim} (1+(\mathbb{A}_s^{\mathrm{c}})^{1/s}) \sum_{(i,j,k)\leq(i',j',k')} \big(\eta_{ijk} + \|\nabla\cdot\boldsymbol{U}_{ijk}\|_\Omega\big)^{-1/s}.\end{aligned}$$

Hence, linear convergence (4.2) in combination with Lemma 4.7(a) gives that

$$\#\mathcal{T}_{i'j'k'} - \#\mathcal{T}_{\mathrm{init}} + 1 \lesssim (1+(\mathbb{A}_s^{\mathrm{c}})^{1/s})^s \big(\eta_{i'j'k'} + \|\nabla\cdot\boldsymbol{U}_{i'j'k'}\|_\Omega\big)^{-1/s}$$

for all for all $(i',j',k') \in \mathcal{Q}$ with $\mathcal{T}_{i'j'k'} \neq \mathcal{T}_{\mathrm{init}}$. For all other $(i',j',k') \in \mathcal{Q}$ with $\mathcal{T}_{i'j'k'} = \mathcal{T}_{\mathrm{init}}$, the latter estimate is clear. With $(1+(\mathbb{A}_s^{\mathrm{c}})^{1/s})^s \lesssim 1 + \mathbb{A}_s^{\mathrm{c}}$, we conclude the proof. □

## Appendix A. Contraction property of $N_\alpha$

The norm of a self-adjoint operator $T : H \to H$ on a Hilbert space $H$ satisfies that

$$\|T\| = \max\{|\mu|, |M|\}, \quad \text{where} \quad \mu := \inf_{x\in H\setminus\{0\}} \frac{\langle Tx\,,\,x\rangle_H}{\|x\|_H^2} \text{ and } M := \sup_{x\in H\setminus\{0\}} \frac{\langle Tx\,,\,x\rangle_H}{\|x\|_H^2}.$$

If $T$ is positive semi-definite (*i.e.*, $\langle Tx\,,\,x\rangle_H \geq 0$ for all $x \in H$), then

$$\|T\| = \sup_{x\in H\setminus\{0\}} \frac{\langle Tx\,,\,x\rangle_H}{\|x\|_H^2}.$$

Consider $H = \mathbb{P}$. Let $0 < \alpha < 2\,\|S\|^{-1}$. Since the Schur complement operator $S = \nabla\cdot\Delta^{-1}\nabla : \mathbb{P} \to \mathbb{P}$ is self-adjoint, also the operator $T := I - \alpha S$ is self-adjoint. Moreover, $S$ is positive definite. Hence,

$$\mu = \inf_{q\in\mathbb{P}\setminus\{0\}} \frac{\langle (I-\alpha S)q\,,\,q\rangle_\Omega}{\|q\|_\Omega^2} = 1 - \alpha \sup_{q\in\mathbb{P}\setminus\{0\}} \frac{\langle Sq\,,\,q\rangle_\Omega}{\|q\|_\Omega^2} = 1 - \alpha\,\|S\| > -1$$

as well as

$$M = \sup_{q\in\mathbb{P}\setminus\{0\}} \frac{\langle (I-\alpha S)q\,,\,q\rangle_\Omega}{\|q\|_\Omega^2} = 1 - \alpha \inf_{q\in\mathbb{P}\setminus\{0\}} \frac{\langle Sq\,,\,q\rangle_\Omega}{\|q\|_\Omega^2} < 1.$$

Altogether, $\|I - \alpha S\| = \max\{|\mu|, |M|\} < 1$ and thus $N_\alpha : \mathbb{P} \to \mathbb{P}$ from (1.4) is a contraction.

## Appendix B. Proof of (2.2)

It suffices to prove the inequality for $\boldsymbol{v}$ in the dense subspace $C_c^\infty(\Omega)^n \subseteq H_0^1(\Omega) = \mathbb{V}$. Integration by parts and the fact that $\partial_k\partial_j\boldsymbol{v}_j = \partial_j\partial_k\boldsymbol{v}_j$ show that

$$\begin{aligned}\|\nabla\cdot\boldsymbol{v}\|_\Omega^2 &= \sum_{j,k=1}^n \langle \partial_j\boldsymbol{v}_j\,,\,\partial_k\boldsymbol{v}_k\rangle_\Omega = -\sum_{j,k=1}^n \langle \partial_k\partial_j\boldsymbol{v}_j\,,\,\boldsymbol{v}_k\rangle_\Omega = -\sum_{j,k=1}^n \langle \partial_j\partial_k\boldsymbol{v}_j\,,\,\boldsymbol{v}_k\rangle_\Omega \\ &= \sum_{j,k=1}^n \langle \partial_k\boldsymbol{v}_j\,,\,\partial_j\boldsymbol{v}_k\rangle_\Omega \leq \sum_{j,k=1}^n \|\partial_k\boldsymbol{v}_j\|_\Omega\|\partial_j\boldsymbol{v}_k\|_\Omega \leq \frac{1}{2}\sum_{j,k=1}^n \big(\|\partial_k\boldsymbol{v}_j\|_\Omega^2 + \|\partial_j\boldsymbol{v}_k\|_\Omega^2\big) = \|\nabla\boldsymbol{v}\|_\Omega^2.\end{aligned}$$

## Appendix C. Proof of Remark 5.4

**Proof of (5.9).** Let $q \in \{p, \boldsymbol{f}\}$. First, $\mathbb{A}_s^{\rm nc}(q) \le \mathbb{A}_s^{\rm c}(q)$ is trivially satisfied due to $\mathbb{T}^{\rm c} \subseteq \mathbb{T}^{\rm nc}$. To see the converse inequality, let $N \in \mathbb{N}_0$ be arbitrary and $\mathcal{P}' \in \mathbb{T}_N^{\rm nc}$ with $\varrho_q(\mathcal{P}') = \min_{\mathcal{P} \in \mathbb{T}_N^{\rm nc}} \varrho_q(\mathcal{P})$. According to (M4), we have that $\mathtt{close}(\mathcal{P}) \in \mathbb{T}^{\rm c}_{C_{\rm cls}N}$. Thus, monotonicity of $\varrho_q$ gives that

$$\begin{aligned}\min_{\mathcal{T} \in \mathbb{T}^{\rm c}_{\lfloor C_{\rm cls}N \rfloor}} (C_{\rm cls}N+1)^s \varrho_q(\mathcal{T}) &\le (C_{\rm cls}N+1)^s \varrho_q(\mathtt{close}(\mathcal{P}')) \le (C_{\rm cls}+1)^s (N+1)^s \varrho_q(\mathcal{P}') \\ &= (C_{\rm cls}+1)^s (N+1)^s \min_{\mathcal{P} \in \mathbb{T}_N^{\rm nc}} \varrho_q(\mathcal{P}) \le (C_{\rm cls}+1)^s \mathbb{A}_s^{\rm nc}(q).\end{aligned}$$

Finally, elementary estimation yields for arbitrary $M \in \mathbb{N}_0$ and $N := \lfloor M/C_{\rm cls} \rfloor$ that

$$\min_{\mathcal{T} \in \mathbb{T}^{\rm c}_M} (M+1)^s \varrho_q(\mathcal{T}) \lesssim \min_{\mathcal{T} \in \mathbb{T}^{\rm c}_{\lfloor C_{\rm cls}N \rfloor}} (C_{\rm cls}N+1)^s \varrho_q(\mathcal{T}) \le 2^s \mathbb{A}_s^{\rm nc}(q).$$

Taking the supremum over all $M \in \mathbb{N}_0$, we conclude the proof. ☐

**Proof of (5.10).** By definition, we have that $\varrho_{\boldsymbol{u}}(\mathcal{T}) + \varrho_p(\mathcal{T}) + \varrho_{\boldsymbol{f}}(\mathcal{T}) = \varrho_{\boldsymbol{u},p,\boldsymbol{f}}(\mathcal{T})$. Hence,

$$\mathbb{A}_s^{\rm c}(\boldsymbol{u}) + \mathbb{A}_s^{\rm c}(p) + \mathbb{A}_s^{\rm c}(\boldsymbol{f}) \le 3\, \mathbb{A}_s^{\rm c}(\boldsymbol{u}, p, \boldsymbol{f}).$$

Moreover, the overlay estimate (M1) also proves the converse estimate.

To see this, let $N \in \mathbb{N}_0$. If $N \bmod 3 = 0$, choose $n' = n'' = n''' = N/3 \in \mathbb{N}_0$. If $N \bmod 3 = 2$, choose $n' = (N-1)/3$, $n'' = (N-1)/3 \in \mathbb{N}_0$, $n''' = (N+2)/3 \in \mathbb{N}_0$. If $N \bmod 3 = 1$, choose $n' = (N-2)/3$, $n'' = (N+1)/3 \in \mathbb{N}_0$, $n''' = (N+1)/3 \in \mathbb{N}_0$. Choose $\mathcal{T}' \in \mathbb{T}^{\rm c}_{n'}$ such that $\varrho_{\boldsymbol{u}}(\mathcal{T}') = \min_{\mathcal{T} \in \mathbb{T}^{\rm c}_{n'}} \varrho_{\boldsymbol{u}}(\mathcal{T})$. Choose $\mathcal{T}'' \in \mathbb{T}^{\rm c}_{n''}$ such that $\varrho_p(\mathcal{T}'') = \min_{\mathcal{T} \in \mathbb{T}^{\rm c}_{n''}} \varrho_p(\mathcal{T})$. Choose $\mathcal{T}''' \in \mathbb{T}^{\rm c}_{n'''}$ such that $\varrho_p(\mathcal{T}''') = \min_{\mathcal{T} \in \mathbb{T}^{\rm c}_{n'''}} \varrho_p(\mathcal{T})$. Then, $n' + n'' + n''' = N$ and hence $\mathcal{T} := \mathcal{T}' \oplus \mathcal{T}'' \oplus \mathcal{T}''' \in \mathbb{T}^{\rm c}_N$. Moreover, the monotonicity of $\varrho_{\boldsymbol{u}}$, $\varrho_p$, and $\varrho_{\boldsymbol{f}}$ yields that

$$\begin{aligned}(N+1)^s\, \varrho_{\boldsymbol{u},p,\boldsymbol{f}}(\mathcal{T}) &\le \Big(\frac{N+1}{n'+1}\Big)^s (n'+1)^s\, \varrho_{\boldsymbol{u}}(\mathcal{T}') + \Big(\frac{N+1}{n''+1}\Big)^s (n''+1)^s\, \varrho_p(\mathcal{T}'') \\ &\quad + \Big(\frac{N+1}{n'''+1}\Big)^s (n'''+1)^s\, \varrho_{\boldsymbol{f}}(\mathcal{T}''') \le \Big(\frac{N+1}{n'+1}\Big)^s \big(\mathbb{A}_s^{\rm c}(\boldsymbol{u}) + \mathbb{A}_s^{\rm c}(p) + \mathbb{A}_s^{\rm c}(\boldsymbol{f})\big).\end{aligned}$$

Since $(N+1)/(n'+1) \le 3$, this concludes the proof. ☐

**Proof of (5.11).** For all $\mathcal{T} \in \mathbb{T}^{\rm c}$, it holds that $(1 - \Pi_{\mathcal{T}})(-\nabla p_{\mathcal{T}} + \Delta \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}]) = 0$ and thus $\mathrm{osc}(\mathcal{T}) \le \eta(\mathcal{T}; \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}], p_{\mathcal{T}})$. Together with reliability (3.5), this implies that $\mathbb{A}_s^{\rm c}(\boldsymbol{u}, p, \boldsymbol{f}) \le C_{\rm rel}\, \mathbb{A}_s^{\rm c}$. A standard efficiency estimate (see, *e.g.*, [1], Lem. 4.2) together with the triangle inequality and (2.2) show that

$$\begin{aligned}\eta(\mathcal{T}; \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}], p_{\mathcal{T}}) + \|\nabla \cdot \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}]\|_\Omega &\overset{[1]}{\lesssim} \|\boldsymbol{u}[p_{\mathcal{T}}] - \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}]\|_{\mathbb{V}} + \mathrm{osc}(\mathcal{T}) + \|\nabla \cdot \boldsymbol{u}[p_{\mathcal{T}}]\|_\Omega \\ &\overset{(3.10)}{\le} \|\boldsymbol{u} - \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}]\|_{\mathbb{V}} + \|\boldsymbol{u} - \boldsymbol{u}[p_{\mathcal{T}}]\|_{\mathbb{V}} + \|p - p_{\mathcal{T}}\|_{\mathbb{P}} + \mathrm{osc}(\mathcal{T}) \\ &\overset{(2.14)}{=} \|\boldsymbol{u} - \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}]\|_{\mathbb{V}} + 2\|p - p_{\mathcal{T}}\|_{\mathbb{P}} + \mathrm{osc}(\mathcal{T}).\end{aligned}$$

The hidden constant depends only on $\mathcal{T}_{\rm init}$ and the polynomial degree of $m$. Moreover, it holds that $\boldsymbol{U}_{\mathcal{T}} := \mathrm{argmin}_{\boldsymbol{V}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T})} \|\boldsymbol{u} - \boldsymbol{V}_{\mathcal{T}}\|_{\mathbb{V}} = \boldsymbol{U}_{\mathcal{T}}[p]$. Hence, (2.14) shows that

$$\|\boldsymbol{u} - \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}]\|_{\mathbb{V}} \le \|\boldsymbol{u} - \boldsymbol{U}_{\mathcal{T}}\|_{\mathbb{V}} + \|\boldsymbol{U}_{\mathcal{T}}[p] - \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}]\|_{\mathbb{V}} \overset{(2.14)}{\le} \|\boldsymbol{u} - \boldsymbol{U}_{\mathcal{T}}\|_{\mathbb{V}} + \|p - p_{\mathcal{T}}\|_{\mathbb{P}}.$$

Combining the latter two estimates, we prove for $\mathcal{T}_{\rm init}$-piecewise polynomial $\boldsymbol{f}$ that

$$\eta(\mathcal{T}; \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}], p_{\mathcal{T}}) + \|\nabla \cdot \boldsymbol{U}_{\mathcal{T}}[p_{\mathcal{T}}]\|_\Omega \lesssim \min_{\boldsymbol{V}_{\mathcal{T}} \in \mathbb{V}(\mathcal{T})} \|\boldsymbol{u} - \boldsymbol{V}_{\mathcal{T}}\|_{\mathbb{V}} + \min_{Q_{\mathcal{T}} \in \mathbb{P}(\mathcal{T})} \|p - Q_{\mathcal{T}}\|_{\mathbb{P}} + \mathrm{osc}(\mathcal{T}).$$

Overall, we thus get the converse estimate $\mathbb{A}_s^{\rm c} \lesssim \mathbb{A}_s^{\rm c}(\boldsymbol{u}, p, \boldsymbol{f})$ and hence obtain (5.11). ☐

## Appendix D. List of symbols

The most important symbols are listed in the following table.

| Name | Description | First appearance |
|---|---|---|
| $a(\cdot,\cdot)$ | Bilinear form corresponding to $-\Delta$ | Section 2.1 |
| $A$ | Operator corresponding to $-\Delta$ | Section 2.1 |
| $\mathbb{A}_s^{\mathrm{c}}$ | Approximation constant on conforming triangulations | Lemma 5.2 |
| $\mathbb{A}_s^{\mathrm{c}}(\cdot)$ | Approximation constant for given quantity on conforming triangulations | Lemma 5.1 |
| $\mathbb{A}_s^{\mathrm{nc}}$ | Approximation constant on non-conforming triangulations | Lemma 5.2 |
| $\mathbb{A}_s^{\mathrm{nc}}(\cdot)$ | Approximation constant for given quantity on non-conforming triangulations | Lemma 5.1 |
| $b(\cdot,\cdot)$ | Bilinear from corresponding to $-\nabla\cdot$ | Section 2.1 |
| $B$ | Operator corresponding to $-\nabla\cdot$ | Section 2.1 |
| $B'$ | Operator corresponding to $\nabla$ | Section 2.1 |
| $\texttt{binev}(\cdot,\cdot,\cdot;\cdot)$ | Output of Binev algorithm | Algorithm 3.6 |
| $\texttt{bisect}(\cdot,\cdot)$ | Non-conforming refinement function | Section 2.2 |
| $C_1$ | Linear convergence constant in $k$-direction | Lemma 4.4 |
| $C_2$ | Linear convergence constant in $j$-direction | Lemma 4.5 |
| $C_3$ | Linear convergence constant in $i$-direction | Lemma 4.6 |
| $C_{\mathrm{bin}}$ | Binev constant | Section 3.2 |
| $C_{\mathrm{cls}}$ | Constant in closure estimate | Section 2.2 |
| $C_{\mathrm{comp}}$ | Comparison constant | Lemma 5.6 |
| $C_{\mathrm{div}}$ | Equivalence constant for norms on pressure space | Section 2.1 |
| $C_{\mathrm{drel}}$ | Discrete reliability constant | Lemma 3.2 |
| $C_{\mathrm{lin}}$ | Linear convergence constant | Theorem 4.1 |
| $C_{\mathrm{mark}}$ | Marking constant of adaptive algorithm | Algorithm 3.7 |
| $C_{\mathrm{mon}}$ | Monotonicity constant for estimator | Lemma 4.8 |
| $C_{\mathrm{red}}$ | Reduction constant | Lemma 3.4 |
| $C_{\mathrm{rel}}$ | Reliability constant | Lemma 3.1 |
| $C'_{\mathrm{rel}}(\cdot)$ | Reliability constant for adaptive algorithm | Lemma 4.3 |
| $C_{\mathrm{son}}$ | Maximal number of sons | Section 2.2 |
| $C_{\mathrm{stab}}$ | Stability constant for estimator | Lemma 3.3 |
| $\texttt{close}(\cdot)$ | Conforming closure of triangulation | Section 2.2 |
| $d$ | Dimension | Section 1.1 |
| $\eta$ | Error estimator | Section 3.1 |
| $\eta_{ijk}$ | Error estimator of adaptive algorithm | Section 2.6 |
| $\eta_T$ | Error indicator on an element | Section 3.1 |
| $\boldsymbol{f}$ | Given body force | Section 1.1 |
| $\gamma$ | Shape regularity constant | Section 2.2 |
| $\underline{j}$ | Maximal index $j$ for given index $i$ | Lemma 3.9 |
| $\underline{k}$ | Maximal index $k$ for given indices $(i,j)$ | Lemma 3.9 |
| $\kappa_1$ | Parameter of adaptive algorithm to approximate Galerkin approximation | Algorithm 3.7 |
| $\kappa_2$ | Parameter for $i$ direction of adaptive algorithm | Algorithm 3.7 |
| $\kappa_3$ | Parameter for $j$ direction of adaptive algorithm | Algorithm 3.7 |
| $m$ | Polynomial degree | Section 2.3 |
| $\Omega$ | Bounded Lipschitz domain | Section 1.1 |
| $p$ | Exact pressure | Section 1.1 |

| | | |
|---|---|---|
| $p_i$ | Best approximation in discrete pressure space of adaptive algorithm | Section 2.6 |
| $p_{\mathcal{P}}$ | Best approximation in discrete pressure space | Section 2.4 |
| $P_{ij}$ | Approximative pressure of adaptive algorithm | Section 2.6 |
| $\mathbb{P}$ | Pressure space | Section 1.1 |
| $\mathbb{P}(\cdot)$ | Discrete pressure space on non-conforming triangulation | Section 2.3 |
| $\mathbb{P}_i$ | Discrete pressure space of adaptive algorithm | Section 2.6 |
| $\mathcal{P}_i$ | Non-conforming triangulation for pressure of adaptive algorithm | Section 2.6 |
| $\Pi_i$ | $L^2$-orthogonal projection on non-conforming triangulation of adaptive algorithm | Section 2.6 |
| $\Pi_{\mathcal{P}}$ | $L^2$-orthogonal projection on non-conforming triangulation | Section 2.4 |
| $q_1$ | Linear convergence constant in $k$-direction between 0 and 1 | Lemma 4.4 |
| $q_2$ | Linear convergence constant in $j$-direction between 0 and 1 | Lemma 4.5 |
| $q_3$ | Linear convergence constant in $i$-direction between 0 and 1 | Lemma 4.6 |
| $q_{\text{lin}}$ | Linear convergence constant between 0 and 1 | Theorem 4.1 |
| $q_{\text{red}}$ | Reduction constant between 0 and 1 | Lemma 3.4 |
| $\mathcal{Q}$ | Set of possible indices | Lemma 3.9 |
| $\texttt{refine}(\cdot,\cdot)$ | Conforming refinement function | Section 2.2 |
| $S$ | Schur complement operator | Section 2.1 |
| $\mathbb{T}^{\text{c}}$ | Set of conforming triangulations | Section 2.2 |
| $\mathbb{T}^{\text{c}}(\cdot)$ | Set of conforming refinements | Section 2.2 |
| $\mathbb{T}^{\text{c}}_{\varepsilon}(\cdot)$ | Set of conforming triangulations with given quantity below $\varepsilon$ | Lemma 5.1 |
| $\mathbb{T}^{\text{c}}_{N}$ | Set of conforming triangulations with bounded element number | Lemma 5.1 |
| $\mathbb{T}^{\text{nc}}$ | Set of non-conforming triangulations | Section 2.2 |
| $\mathbb{T}^{\text{nc}}(\cdot)$ | Set of non-conforming refinements | Section 2.2 |
| $\mathbb{T}^{\text{nc}}_{\varepsilon}(\cdot)$ | Set of non-conforming triangulations with given quantity below $\varepsilon$ | Lemma 5.1 |
| $\mathbb{T}^{\text{nc}}_{N}$ | Set of non-conforming triangulations with bounded element number | Lemma 5.1 |
| $\mathcal{T}_{ijk}$ | Conforming triangulation for velocity of adaptive algorithm | Section 2.6 |
| $\mathcal{T}_{\text{init}}$ | Initial conforming triangulation | Section 2.2 |
| $\vartheta$ | Parameter of Binev algorithm | Algorithm 3.6 |
| $\theta$ | Dörfler marking parameter of adaptive algorithm | Algorithm 3.7 |
| $\theta_{\text{opt}}$ | Threshold for Dörfler marking parameter | Algorithm 3.7 |
| $\boldsymbol{u}$ | Exact velocity | Section 1.1 |
| $\boldsymbol{u}[\cdot]$ | Exact velocity for given pressure | Section 2.4 |
| $\boldsymbol{u}_{ij}$ | Exact velocity to approximate pressure of adaptive algorithm | Section 2.6 |
| $\boldsymbol{u}_{\mathcal{P}}$ | Exact velocity for best approximation in discrete pressure space | Section 2.4 |
| $\boldsymbol{U}_{ijk}$ | Approximative velocity of adaptive algorithm | Section 2.6 |
| $\boldsymbol{U}_{\mathcal{T}}[\cdot]$ | Galerkin approximation of velocity for given pressure | Section 2.4 |
| $\mathbb{V}$ | Velocity space | Section 1.1 |
| $\mathbb{V}(\cdot)$ | Discrete velocity space on conforming triangulation | Section 2.3 |
| $\mathbb{V}_{ijk}$ | Discrete velocity space of adaptive algorithm | Section 2.6 |
| $\langle\cdot\,,\,\cdot\rangle_{\Omega}$ | $L^2$-scalar product | Section 2.1 |
| $\\|\cdot\\|_{\Omega}$ | $L^2$-norm | Section 2.1 |
| $\\|\cdot\\|_{\mathbb{P}}$ | Norm on pressure space | Section 2.1 |
| $\\|\cdot\\|_{\mathbb{V}}$ | Norm on velocity space | Section 2.1 |
| $\lvert(\cdot,\cdot,\cdot)\rvert$ | Number of iterations to reach given indices | Section 4.1 |
| $\oplus$ | Overlay of two triangulations | Section 2.2 |
| $<$ | Order relation on set of possible indices | Section 4.1 |

*Acknowledgements.* The authors thankfully acknowledge the support by the Austrian Science Fund (FWF) through grant P27005 (DP), P29096 (GG, DP), as well as grant F65 (GDF, DP) and by CONICYT through FONDECYT project P11170050 (TF). Moreover, GG thanks Peter Binev for his explanations on [4, 5].

# References

[1] E. Bänsch, P. Morin and R.H. Nochetto, An adaptive Uzawa FEM for the Stokes problem: convergence without the inf-sup condition. *SIAM J. Numer. Anal.* **40** (2002) 1207–1229.

[2] R. Becker and S. Mao, Quasi-optimality of adaptive nonconforming finite element methods for the Stokes equations. *SIAM J. Numer. Anal.* **49** (2011) 970–991.

[3] A. Bespalov, A. Haberl and D. Praetorius, Adaptive FEM with coarse initial mesh guarantees optimal convergence rates for compactly perturbed elliptic problems. *Comput. Methods Appl. Mech. Eng.* **317** (2017) 318–340.

[4] P. Binev, Tree approximation for hp-adaptivity. SIAM J. Numer. Anal. **56** (2018), 3346-3357.

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

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

[7] P. Binev, W. Dahmen, R. DeVore and P. Petrushev, Approximation classes for adaptive methods. *Serdica Math. J.* **28** (2002) 391–416.

[8] J.H. Bramble, A proof of the inf–sup condition for the Stokes equations on Lipschitz domains. *Math. Models Methods Appl. Sci.* **13** (2003) 361–371.

[9] C. Carstensen, D. Peterseim and H. Rabus, Optimal adaptive nonconforming FEM for the Stokes problem. *Numer. Math.* **123** (2013) 291–308.

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

[11] 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.

[12] S. Dahlke, W. Dahmen and K. Urban, Adaptive wavelet methods for saddle point problems—optimal convergence rates. *SIAM J. Numer. Anal.* **40** (2002) 1230–1262.

[13] C. Erath, G. Gantner and D. Praetorius, Optimal convergence behavior of adaptive FEM driven by simple $(h-h/2)$-type error estimators. Preprint arXiv:1805.00715 (2018).

[14] M. Feischl, Optimal adaptivity for a standard finite element method for the Stokes problem. Preprint arXiv:1710.08289 (2017).

[15] T. Führer, A. Haberl, D. Praetorius and S. Schimanko, Adaptive BEM with inexact PCG solver yields almost optimal computational costs. *Numer. Math.* (2018) published online first.

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

[17] T. Gantumur, On the convergence theory of adaptive mixed finite element methods for the Stokes problem. Preprint arXiv:1403.0895 (2014).

[18] T. Gantumur, Convergence rates of adaptive methods, Besov spaces, and multilevel approximation. *Found. Comput. Math.* **17** (2017) 917–956.

[19] F.D. Gaspoz and P. Morin, Convergence rates for adaptive finite elements. *IMA J. Numer. Anal.* **29** (2008) 917–936.

[20] G.H. Golub and C.F. Van Loan, Matrix Computations, 4th edition. Johns Hopkins University Press, Baltimore, MD (2013).

[21] J. Hu and J. Xu, Convergence and optimality of the adaptive nonconforming linear element method for the Stokes problem. *J. Sci. Comput.* **55** (2013) 125–148.

[22] M. Karkulik, D. Pavlicek and D. Praetorius, On 2D newest vertex bisection: optimality of mesh-closure and $H^1$-stability of $L_2$-projection. *Constr. Approx.* **38** (2013) 213–234.

[23] Y. Kondratyuk, Adaptive Finite Element Algorithms for the Stokes Problem: COnvergence Rates and Optimal Computational Complexity. Preprint, Department of Mathematics, Utrecht University, The Netherlands (2006) 1346.

[24] Y. Kondratyuk and R. Stevenson, An optimal adaptive finite element method for the Stokes problem. *SIAM J. Numer. Anal.* **46** (2008) 747–775.

[25] P. Morin, K.G Siebert and A. Veeser, A basic convergence result for conforming adaptive finite elements. *Math. Models Methods Appl. Sci.* **18** (2008) 707–737.

[26] K.G. Siebert, A convergence proof for adaptive finite elements without lower bound. *IMA J. Numer. Anal.* **31** (2010) 947–970.

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

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