Numerische Mathematik
https://doi.org/10.1007/s00211-025-01455-w

Numerische Mathematik

Cost-optimal adaptive FEM with linearization and algebraic solver for semilinear elliptic PDEs

Maximilian Brunner[1] · Dirk Praetorius[1] · Julian Streitberger[1]

Received: 21 January 2024 / Revised: 13 November 2024 / Accepted: 11 February 2025
© The Author(s) 2025

Abstract

We consider scalar semilinear elliptic PDEs, where the nonlinearity is strongly monotone, but only locally Lipschitz continuous. To linearize the arising discrete nonlinear problem, we employ a damped Zarantonello iteration, which leads to a linear Poisson-type equation that is symmetric and positive definite. The resulting system is solved by a contractive algebraic solver such as a multigrid method with local smoothing. We formulate a fully adaptive algorithm that equibalances the various error components coming from mesh refinement, iterative linearization, and algebraic solver. We prove that the proposed adaptive iteratively linearized finite element method (AILFEM) guarantees convergence with optimal complexity, where the rates are understood with respect to the overall computational cost (i.e., the computational time). Numerical experiments investigate the involved adaptivity parameters.

Mathematics Subject Classification 65N30 · 65N50 · 65N15 · 65Y20 · 41A25

1 Introduction

1.1 Problem setting and main results

Undoubtedly, adaptive finite element methods (AFEMs) are in the canon of reliable numerical methods for the solution of partial differential equations (PDEs). Some of

Dirk Praetorius and Julian Streitberger contributed equally to this work.

✉ Maximilian Brunner
maximilian.brunner@asc.tuwien.ac.at

Dirk Praetorius
dirk.praetorius@asc.tuwien.ac.at

Julian Streitberger
julian.streitberger@asc.tuwien.ac.at

[1] Institute of Analysis and Scientific Computing, TU Wien, Wiedner Hauptstraße 8–10, 1040 Vienna, Austria

Published online: 12 March 2025

the seminal contributions in this still very active area are [1–9] for linear problems, [10–14] for nonlinear problems, and [15] for an abstract framework.

By means of conforming finite elements, this paper is concerned with the cost-optimal computation of the solution $u^\star \in H^1_0(\Omega)$ to the *semilinear* elliptic model problem

$$-\operatorname{div}(\boldsymbol{A}\nabla u^\star) + b(u^\star) = F \quad \text{in } \Omega \quad \text{subject to} \quad u^\star = 0 \quad \text{on } \partial\Omega, \tag{1}$$

with a Lipschitz domain $\Omega \subset \mathbb{R}^d$ for $d \in \{1, 2, 3\}$, an elliptic diffusion coefficient $\boldsymbol{A}\colon \Omega \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$, a monotone nonlinearity $b\colon \Omega \to \mathbb{R}$, and sufficiently regular data F. The assumptions are such that the Browder–Minty theorem ensures existence and uniqueness.

Moreover, the model problem (1) can be recast into the framework of strongly monotone and locally Lipschitz continuous operators such that the abstract model problem reads: For $\mathcal{X} = H^1_0(\Omega)$ with topological dual space $\mathcal{X}' = H^{-1}(\Omega)$ and duality bracket $\langle \cdot\,, \cdot \rangle$, a nonlinear operator $\mathcal{A}\colon \mathcal{X} \to \mathcal{X}'$, and given data $F \in \mathcal{X}'$, we aim to approximate the solution $u^\star \in \mathcal{X}$ to

$$\langle \mathcal{A}u^\star\,,\, v \rangle = \langle F\,,\, v \rangle \quad \text{for all } v \in \mathcal{X}. \tag{2}$$

To this end, we employ conforming piecewise polynomial finite element spaces $\mathcal{X}_H \subset \mathcal{X}$ with the corresponding discrete solution $u^\star_H \in \mathcal{X}_H$ to

$$\langle \mathcal{A}u^\star_H\,,\, v_H \rangle = \langle F\,,\, v_H \rangle \quad \text{for all } v_H \in \mathcal{X}_H, \tag{3}$$

which, however, can hardly be computed exactly, since (3) is still a discrete nonlinear system of equations.

The major difficulty of such problems is that the Lipschitz constant of $\mathcal{A}$ depends on the considered functions v and w in the sense that for $\vartheta > 0$, it holds that

$$\|\mathcal{A}v - \mathcal{A}w\|_{\mathcal{X}'} \le L[\vartheta]\, |||v - w||| \quad \text{for all } v, w \in \mathcal{X} \text{ with } \max\big\{|||v|||, |||w|||\big\} \le \vartheta. \tag{LIP$'$}$$

Moreover, this dependence also appears in the stability constant of the residual-based a posteriori error estimator [16, 17].

Hence, for such a problem class, any approximate numerical scheme must ensure uniform boundedness of all computed approximations $u^\star_H \approx u_H \in \mathcal{X}_H$ throughout the algorithm. This constitutes the first main result: The developed *adaptive iteratively linearized FEM* (AILFEM) algorithm (more detailed in Algorithm 1 below) guarantees a uniform upper bound on all iterates (see Theorem 4 below). In particular, the algorithm steers the decision whether it is more preferable to refine the mesh adaptively or to do an additional step of linearization or a further algebraic solver step instead.

Once uniform boundedness is established, we prove full R-linear convergence (Theorem 5 below) as the second main result. Full R-linear convergence establishes contraction in each step of the algorithm regardless of the algorithmic decision. At the

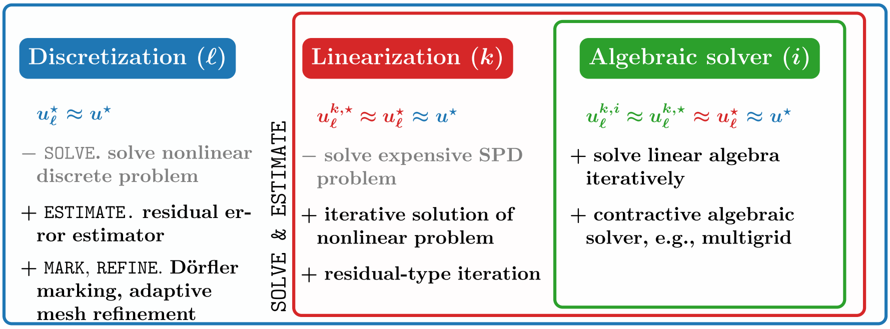


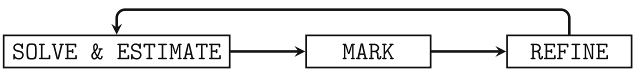
Fig. 1 Depiction of the nested loops of the AILFEM algorithm 1 below

expense of a more challenging analysis that links energy arguments with the energy norm of the algebraic solver, full R-linear convergence is guaranteed for all mesh levels $\ell \geq \ell_0 = 0$ while prior works [17, 18] used compactness arguments which only guaranteed the existence of the index $\ell_0 \in \mathbb{N}_0$ (and not necessarily $\ell_0 = 0$). As a consequence of uniform boundedness and full R-linear convergence, the third main result proves optimal rates both understood with respect to the degrees of freedom and with respect to the overall computational cost (Corollary 1 and Theorem 6) of the proposed algorithm.

Compared to existing results in the literature [14, 19–21], all three main results require a suitable adaptation of the stopping criteria of the linearization loop as well as sufficiently many iterations in the algebra loop, together with subtle technical challenges, in particular, for the proof of full R-linear convergence.

1.2 From AFEM to AILFEM

On each mesh level (with mesh index ℓ), the arising discrete nonlinear problems cannot be solved exactly in practice as supposed in classical AFEM [10–13]. To deal with this issue, we follow [22–24] and consider the so-called *Zarantonello iteration* from [25] as a linearization method (with index k). The Zarantonello iteration is a Richardson-type iteration where only a Laplace-type problem has to be solved in each iteration. Since the arising large SPD systems are still expensive to solve exactly, we employ a contractive algebraic solver as a nested loop to solve the Zarantonello system inexactly (with iteration index i). The loops thus come with a natural nestedness (see Fig. 1), where the overall schematic loop of the algorithm reads



Since the proposed adaptive loop depends on all previous computations, optimal convergence rates should be understood with respect to the overall computational cost. This idea of *optimal complexity* originates from the wavelet community [26, 27] was used in the context of AFEM in [5] for the Poisson model problem and [28] for the Poisson eigenvalue problem, both under realistic assumptions on generic iterative solvers.

AILFEMs with iterative and/or inexact solver with a posteriori error estimators are found in, e.g., [22, 29–33] and references therein. Besides the Zarantonello iteration, for globally Lipschitz continuous nonlinearities, the works [20, 24, 34] analyze also other linearizations such as the Kačanov iteration or damped Newton schemes. Optimal complexity of the Zarantonello loop that is coupled with an algebraic loop is analyzed in [18] for nonsymmetric second-order linear elliptic PDEs and for strongly monotone (and globally Lipschitz continuous) model problems in [14, 19–21, 23].

The literature on AILFEMs for locally Lipschitz continuous problems is scarce and closing this gap is the aim of this work. The semilinear model problem is treated in, e.g., [33] by a damped Newton iteration and in [35] by an energy-based approach with experimentally observed optimal rates. We also refer to the own work [36] for an AILFEM with optimal rates with respect to the overall computational cost, however, under the assumption that the arising linear systems can be solved at linear cost. More precisely, compared to the previous work [36], in this paper we also take an optimal algebraic solver for the linearized problem into account and propose an adaptive algorithm ensuring optimal convergence rates with respect to the computation time. Moreover, compared to [36] that elaborates the proof of full R-linear convergence along the lines of [14], we provide a much simpler proof inspired by [21]. However, the work [21] employs a general quasi-orthogonality from [37] that is not available for nonlinear problems in general since the proof relies on a stable LU-decomposition of the linear problem. Therefore, to avoid compactness arguments like in [18], we employ orthogonality in the underlying energy.

1.3 Outline

This paper is structured as follows: Sect. 2 introduces the abstract framework on locally Lipschitz continuous operators. In Sect. 3, we formulate the (idealized) AILFEM algorithm (Algorithm 1). We prove uniform boundedness for the final iterates of the algebraic solver (Theorem 4). Section 4 presents the second main result: Full R-linear convergence (Theorem 5). In particular, rates with respect to the degrees of freedom coincide with rates with respect to the computational cost (Corollary 1). In Sect. 5, we prove the main result on optimal complexity of the proposed AILFEM algorithm (Theorem 6). In Sect. 6, we present numerical experiments of the proposed AILFEM strategy and investigate its optimal complexity for various choices of the adaptivity parameters.

2 Strongly monotone operators

This section introduces an abstract framework of strongly monotone and locally Lipschitz continuous operators. This class of operators covers the model problem (1) of semilinear elliptic PDEs with monotone semilinearity.

2.1 Abstract model problem

Let $\mathcal{X}$ be a Hilbert space over $\mathbb{R}$ with scalar product $\langle\!\langle \cdot\,,\,\cdot \rangle\!\rangle$ and induced norm $|||\cdot|||$. Let $\mathcal{X}_H \subseteq \mathcal{X}$ be a closed subspace. Let $\mathcal{X}'$ be the dual space with norm $\|\cdot\|_{\mathcal{X}'}$ and denote by $\langle \cdot, \cdot \rangle$ the duality bracket on $\mathcal{X}' \times \mathcal{X}$. Let $\mathcal{A}\colon \mathcal{X} \to \mathcal{X}'$ be a nonlinear operator. We suppose that $\mathcal{A}$ is **strongly monotone**, i.e., there exists a monotonicity constant $\alpha > 0$ such that

$$\alpha\, |||v - w|||^2 \le \langle \mathcal{A}v - \mathcal{A}w\,,\, v - w \rangle \quad \text{for all } v, w \in \mathcal{X}. \tag{SM}$$

Moreover, we suppose that $\mathcal{A}$ is **locally Lipschitz continuous**, i.e., for all $\vartheta > 0$, there exists $L[\vartheta] > 0$ such that

$$\langle \mathcal{A}v - \mathcal{A}w\,,\, \varphi \rangle \le L[\vartheta]\, |||v - w|||\, |||\varphi||| \;\; \text{for all } v, w, \varphi \in \mathcal{X} \text{ with } \max\big\{|||v|||, |||v - w|||\big\} \le \vartheta. \tag{LIP}$$

Remark 1 We remark that local Lipschitz continuity is often defined differently in the existing literature, cf. [38, p. 565]: For all $\Theta > 0$, there exists $L'[\Theta] > 0$ such that

$$\langle \mathcal{A}v - \mathcal{A}w\,,\, \varphi \rangle \le L'[\Theta]\, |||v - w|||\, |||\varphi||| \text{ for all } v, w, \varphi \in \mathcal{X} \text{ with } \max\big\{|||v|||, |||w|||\big\} \le \Theta. \tag{LIP$'$}$$

We note that the conditions (LIP) and (LIP$'$) are indeed equivalent in the sense that (LIP) yields (LIP$'$) with $\Theta = 2\vartheta$, and, conversely, (LIP$'$) yields (LIP) with $\vartheta = 2\,\Theta$. However, condition (LIP) is better suited for the inductive proof of Proposition 2 which is the main ingredient to guarantee uniform boundedness in Theorem 4.

Without loss of generality, we may suppose that $\mathcal{A}0 \neq F \in \mathcal{X}'$. We consider the operator equation: Seek $u^\star \in \mathcal{X}$ that solves (2). For any closed subspace $\mathcal{X}_H \subseteq \mathcal{X}$, we consider the corresponding Galerkin discretization (3). We note existence and uniqueness of the solutions to (2)–(3) and a Céa-type estimate.

Proposition 1 ([36, Proposition 2]) *Suppose that $\mathcal{A}$ satisfies (SM) and (LIP). Then, (2)–(3) admit unique solutions $u^\star \in \mathcal{X}$ and $u_H^\star \in \mathcal{X}_H$, respectively, and*

$$\max\big\{|||u^\star|||, |||u_H^\star|||\big\} \le M := \frac{1}{\alpha}\, \|F - \mathcal{A}0\|_{\mathcal{X}'} > 0 \tag{4}$$

as well as

$$|||u^\star - u_H^\star||| \le C_{\text{Céa}} \min_{v_H \in \mathcal{X}_H} |||u^\star - v_H||| \quad \textit{with} \quad C_{\text{Céa}} = L[2M]/\alpha. \quad \square \tag{5}$$

Finally, we suppose that $\mathcal{A}$ has a potential $\mathcal{P}$: There exists a Gâteaux differentiable function $\mathcal{P}\colon \mathcal{X} \to \mathbb{R}$ such that its derivative $\mathrm{d}\mathcal{P}\colon \mathcal{X} \to \mathcal{X}'$ coincides with $\mathcal{A}$, i.e.,

$$\langle \mathcal{A}w\,,\ v\rangle = \langle \mathrm{d}\mathcal{P}(w)\,,\ v\rangle = \lim_{\substack{t\to 0\\ t\in\mathbb{R}}} \frac{\mathcal{P}(w+tv)-\mathcal{P}(w)}{t} \quad \text{for all } v, w \in \mathcal{X}. \tag{POT}$$

With the energy $\mathcal{E}(v) := (\mathcal{P} - F)v$, there holds the following classical equivalence.

Lemma 1 (see, e.g., [23, Lemma 5.1]) *Let $\mathcal{X}_H \subseteq \mathcal{X}$ be a closed subspace (where also $\mathcal{X}_H$ replaced by $\mathcal{X}$ is admissible). Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and (POT). Let $\vartheta \geq M$. Let $v_H \in \mathcal{X}_H$ with $|||v_H - u_H^\star||| \leq \vartheta$. Then, it holds that*

$$\frac{\alpha}{2}\,|||v_H - u_H^\star|||^2 \leq \mathcal{E}(v_H) - \mathcal{E}(u_H^\star) \leq \frac{L[\vartheta]}{2}\,|||v_H - u_H^\star|||^2. \tag{6}$$

In particular, the solution $u_H^\star$ of the variational formulation (3) is indeed the unique minimizer of $\mathcal{E}$ in $\mathcal{X}_H$, i.e.,

$$\mathcal{E}(u_H^\star) \leq \mathcal{E}(v_H) \quad \textit{for all } v_H \in \mathcal{X}_H. \tag{7}$$

In particular, it holds that

$$\mathcal{E}(v_H) - \mathcal{E}(u^\star) = \big[\mathcal{E}(v_H) - \mathcal{E}(u_H^\star)\big] + \big[\mathcal{E}(u_H^\star) - \mathcal{E}(u^\star)\big] \quad \textit{for all } v_H \in \mathcal{X}_H \tag{8}$$

and all these energy differences are nonnegative. □

2.2 Iterative linearization and algebraic solver

Let $\mathcal{X}_H \subset \mathcal{X}$ be a finite-dimensional (and hence closed) subspace of $\mathcal{X}$. In order to solve the arising nonlinear discrete problems (3), we will incorporate a linearization method as well as an algebraic solver into the proposed algorithm.

Linearization by Zarantonello iteration. For a detailed discussion of the Zarantonello iteration, we refer to [36, Sect. 2.2–2.4]. For a damping parameter $\delta > 0$ and $w_H \in \mathcal{X}_H$, let $\Phi_H(\delta; w_H) \in \mathcal{X}_H$ solve

$$\langle\!\langle \Phi_H(\delta; w_H)\,,\ v_H \rangle\!\rangle = \langle\!\langle w_H\,,\ v_H\rangle\!\rangle + \delta\big[F(v_H) - \langle \mathcal{A}w_H\,,\ v_H\rangle\big] \quad \text{for all } v_H \in \mathcal{X}_H. \tag{9}$$

The Lax–Milgram lemma proves existence and uniqueness of $\Phi_H(\delta; w_H)$, i.e., the Zarantonello operator $\Phi_H(\delta;\cdot)\colon \mathcal{X}_H \to \mathcal{X}_H$ is well-defined. In particular, $u_H^\star = \Phi(\delta; u_H^\star)$ is the unique fixed point of $\Phi_H(\delta;\cdot)$ for any damping parameter $\delta > 0$. Moreover, for sufficiently small $\delta > 0$, the Zarantonello operator is norm-contractive.

Proposition 2 (see, e.g., [36, Proposition 4]) *Suppose that $\mathcal{A}$ satisfies* (SM) *and* (LIP). *Let $\vartheta > 0$ and $v_H, w_H \in \mathcal{X}_H$ with* $\max\big\{|||v_H|||, |||v_H - w_H|||\big\} \leq \vartheta$. *Then, for all $0 < \delta < 2\alpha/L[\vartheta]^2$ and $0 < q_{\mathrm{Zar}}^\star[\delta,\vartheta]^2 := 1 - \delta\,(2\alpha - \delta L[\vartheta]^2) < 1$, it holds that*

$$|||\Phi_H(\delta; v_H) - \Phi_H(\delta; w_H)||| \le q^\star_{\mathrm{Zar}}[\delta, \vartheta]\, |||v_H - w_H|||. \tag{10}$$

We note that $q^\star_{\mathrm{Zar}}[\delta, \vartheta] \to 1$ *as* $\delta \to 0$*. For known* α *and* $L[\vartheta]$*, the contraction constant* $q^\star_{\mathrm{Zar}}[\delta, \vartheta]^2 = 1 - \alpha^2/L[\vartheta]^2 = 1 - \alpha\,\delta$ *is minimal and only attained for* $\delta = \alpha/L[\vartheta]^2$. □

Algebraic solver. The Zarantonello system (9) leads to an SPD system of equations to compute $\Phi_H(\delta; u_H)$. Since large SPD problems are still computationally expensive, we employ an iterative algebraic solver with process function $\Psi_H : \mathcal{X}' \times \mathcal{X}_H \to \mathcal{X}_H$ to solve the arising system (9). More precisely, given a linear functional $\varphi \in \mathcal{X}'$ and an approximation $w_H \in \mathcal{X}_H$ of the exact solutions $w^\star_H \in \mathcal{X}_H$ to

$$\langle\!\langle w^\star_H\,,\, v_H \rangle\!\rangle = \varphi(v_H) \quad \text{for all } v_H \in \mathcal{X}_H,$$

the algebraic solver returns an improved approximation $\Psi_H(\varphi; w_H) \in \mathcal{X}_H$ in the sense that there exists a uniform constant $0 < q_{\mathrm{alg}} < 1$ independent of φ and $\mathcal{X}_H$ such that

$$|||w^\star_H - \Psi_H(\varphi; w_H)||| \le q_{\mathrm{alg}}\, |||w^\star_H - w_H||| \quad \text{for all } w_H \in \mathcal{X}_H. \tag{11}$$

To simplify notation when the right-hand side φ is complicated or lengthy (as for the Zarantonello iteration (9)), we shall write $\Psi_H(w^\star_H; \cdot)$ instead of $\Psi_H(\varphi; \cdot)$, even though $w^\star_H$ is unknown and will never be computed.

2.3 Mesh refinement

Henceforth, let $\mathcal{T}_0$ be an initial triangulation of Ω into compact triangles. For mesh refinement, we use newest vertex bisection (NVB); cf. [39] for $d \ge 2$ with admissible $\mathcal{T}_0$ as well as [40] for $d = 2$ and [41] for $d \ge 2$ with nonadmissible $\mathcal{T}_0$. For $d = 1$, we refer to [42]. For each triangulation $\mathcal{T}_H$ and marked elements $\mathcal{M}_H \subseteq \mathcal{T}_H$, let $\mathcal{T}_h := \mathtt{refine}(\mathcal{T}_H, \mathcal{M}_H)$ be the coarsest refinement of $\mathcal{T}_H$ such that at least all elements $T \in \mathcal{M}_H$ have been refined, i.e., $\mathcal{M}_H \subseteq \mathcal{T}_H \backslash \mathcal{T}_h$. We write $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ if $\mathcal{T}_h$ can be obtained from $\mathcal{T}_H$ by finitely many steps of NVB, and, for $N \in \mathbb{N}_0$, we write $\mathcal{T}_h \in \mathbb{T}_N(\mathcal{T}_H)$ if $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ and $\#\mathcal{T}_h - \#\mathcal{T}_H \le N$. To abbreviate notation, let $\mathbb{T} := \mathbb{T}(\mathcal{T}_0)$. Throughout, any $\mathcal{T}_H \in \mathbb{T}$ is associated with a finite-dimensional space $\mathcal{X}_H \subset \mathcal{X}$ such that nestedness of meshes $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ implies nestedness of the associated spaces $\mathcal{X}_H \subseteq \mathcal{X}_h$.

2.4 Axioms of adaptivity and a posteriori error estimator

For $\mathcal{T}_H \in \mathbb{T}$, $T \in \mathcal{T}_H$, and $v_H \in \mathcal{X}_H$, let $\eta_H(T, v_H) \in \mathbb{R}_{\ge 0}$ be the local contributions of an a posteriori error estimator and abbreviate

$$\eta_H(v_H) := \eta_H(\mathcal{T}_H, v_H), \text{ where } \eta_H(\mathcal{U}_H, v_H) := \Big(\sum_{T \in \mathcal{U}_H} \eta_H(T, v_H)^2\Big)^{1/2} \text{for all } \mathcal{U}_H \subseteq \mathcal{T}_H. \tag{12}$$

We suppose that the error estimator η_H satisfies the following axioms of adaptivity from [15] with a slightly relaxed variant of stability (A1) in the spirit of [17].

(A1) **stability:** For all $\vartheta > 0$ and all $\mathcal{U}_H \subseteq \mathcal{T}_h \cap \mathcal{T}_H$, there exists $C_{\rm stab}[\vartheta] > 0$ such that for all $v_h \in \mathcal{X}_h$ and $v_H \in \mathcal{X}_H$ with $\max\big\{|||v_h|||, |||v_h - v_H|||\big\} \le \vartheta$, it holds that

$$\big|\eta_h(\mathcal{U}_H, v_h) - \eta_H(\mathcal{U}_H, v_H)\big| \le C_{\rm stab}[\vartheta]\, |||v_h - v_H|||.$$

(A2) **reduction:** With $0 < q_{\rm red} < 1$, it holds that

$$\eta_h(\mathcal{T}_h\backslash\mathcal{T}_H, v_H) \le q_{\rm red}\, \eta_H(\mathcal{T}_H\backslash\mathcal{T}_h, v_H) \quad \text{for all } v_H \in \mathcal{X}_H.$$

(A3) **reliability:** There exists $C_{\rm rel} > 0$ such that

$$|||u^\star - u_H^\star||| \le C_{\rm rel}\, \eta_H(u_H^\star).$$

(A4) **discrete reliability:** There exists $C_{\rm drel} > 0$ such that

$$|||u_h^\star - u_H^\star||| \le C_{\rm drel}\, \eta_H(\mathcal{T}_H\backslash\mathcal{T}_h, u_H^\star).$$

2.5 Application of abstract framework (2) to semilinear PDEs (1)

In the following, we comment on how the semilinear PDE (1) fits into the abstract framework in Sect. 2.1–2.4. Let $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, be a bounded Lipschitz domain with polygonal boundary. The weak formulation of the semilinear model problem (1) reads: Given $F \in H^{-1}(\Omega)$, find $u^\star \in \mathcal{X} := H_0^1(\Omega)$ such that

$$\langle \boldsymbol{A}\,\nabla u^\star\,,\ \nabla v\rangle_\Omega + \langle b(u^\star)\,,\ v\rangle_\Omega = \langle F\,,\ v\rangle \quad \text{for all } v \in H_0^1(\Omega), \tag{13}$$

where $\langle\cdot,\,\cdot\rangle_\Omega$ denotes the $L^2(\Omega)$-scalar product. Note that (13) coincides with (2), where $\mathcal{A}u := \langle \boldsymbol{A}\,\nabla u\,,\ \nabla\,\cdot\rangle_\Omega + \langle b(u)\,,\ \cdot\rangle_\Omega$ with $u \in \mathcal{X}$. As a means of discretization, we consider Lagrange finite element spaces of piecewise polynomial functions of a fixed polynomial degree $p \in \mathbb{N}$ on a conforming triangulation $\mathcal{T}_H$ of Ω, namely $\mathcal{X}_H := \mathbb{S}_0^p(\mathcal{T}_H) := \{v_H \in H_0^1(\Omega)\colon v_H|_T \text{ is polynomial of degree } \le p \text{ for all } T \in \mathcal{T}_H\}$. This discretization leads to nested spaces $\mathcal{X}_H \subseteq \mathcal{X}_h$ whenever $\mathcal{T}_H \in \mathbb{T}$ and $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$. The precise assumptions on the model problem are given as follows.

Assumptions on the right-hand side. We suppose the following.

(RHS) Let $\langle F\,,\ v\rangle := \langle f\,,\ v\rangle_\Omega + \langle \boldsymbol{f}\,,\ \nabla v\rangle_\Omega$ with given $f \in L^2(\Omega)$ and $\boldsymbol{f} \in [L^2(\Omega)]^d$.

Assumptions on the diffusion coefficient. The diffusion coefficient $\boldsymbol{A}$ satisfies the following standard assumptions:

(ELL) $\boldsymbol{A} \in L^\infty(\Omega; \mathbb{R}^{d\times d}_{\mathrm{sym}})$, where $\boldsymbol{A}(x)$ is a symmetric and uniformly positive definite matrix, i.e., the minimal and maximal eigenvalues satisfy

$$0 < \mu_0 := \operatorname*{ess\,inf}_{x\in\Omega} \lambda_{\min}(\boldsymbol{A}(x)) \le \operatorname*{ess\,sup}_{x\in\Omega} \lambda_{\max}(\boldsymbol{A}(x)) =: \mu_1 < \infty.$$

In particular, the $\boldsymbol{A}$-induced energy scalar product $\langle\!\langle v\,,\ w\rangle\!\rangle := \langle \boldsymbol{A}\nabla v\,,\ \nabla w\rangle_\Omega$ induces an equivalent norm $|\!|\!| v |\!|\!| := \langle\!\langle v\,,\ v\rangle\!\rangle^{1/2}$ on $H^1_0(\Omega)$.

Assumptions on the nonlinear reaction coefficient. The nonlinearity $b(\cdot)$ satisfies the following assumptions from [43, (A1)–(A3)]:

(CAR) $b\colon \Omega\times\mathbb{R}\to\mathbb{R}$ is a ***Carathéodory*** function, i.e., for all $n\in\mathbb{N}_0$, the n-th derivative $b^{(n)} := \partial^n_\xi b$ of b with respect to the second argument ξ satisfies that

- ▹ for any $\xi\in\mathbb{R}$, the function $x\mapsto b^{(n)}(x,\xi)$ is measurable on Ω,
- ▹ for any $x\in\Omega$, the function $\xi\mapsto b^{(n)}(x,\xi)$ exists and is continuous in ξ.

(MON) We assume **monotonicity** in the second argument, i.e., $b'(x,\xi) := b^{(1)}(x,\xi)\ge 0$ for all $x\in\Omega$ and $\xi\in\mathbb{R}$. By considering $\tilde b(v) := b(v)-b(0)$ and $\tilde f := f - b(0)$, we assume without loss of generality that $b(x,0)=0$.

To establish continuity of $v\mapsto \langle b(v),\ w\rangle_\Omega$, we impose the following **growth condition** on $b(v)$; see, e.g., [44, Chapter III, (12)] or [43, (A4)]:

(GC) There exist $R>0$ and $N\in\mathbb{N}$ with $N\le 5$ for $d=3$ such that

$$|b^{(N)}(x,\xi)| \le R \quad \text{for a.e. } x\in\Omega \text{ and all } \xi\in\mathbb{R}.$$

These assumptions suffice to prove that the operator $\mathcal{A} := \mathcal{X}\to\mathcal{X}' = H^{-1}(\Omega)$ associated with the model problem (13) is strongly monotone (SM) and locally Lipschitz continuous (LIP) in the sense of Sect. 2.1; see [36, Lemma 20].

Energy minimization. Associated with the semilinear model problem (13), we consider the **energy**

$$\mathcal{E}(v) = \frac{1}{2}\int_\Omega |\boldsymbol{A}^{1/2}\nabla v|^2\,\mathrm{d}x + \int_\Omega\int_0^{v(x)} b(s)\,\mathrm{d}s\,\mathrm{d}x - \int_\Omega f v\,\mathrm{d}x - \int_\Omega \boldsymbol{f}\cdot\nabla v\,\mathrm{d}x \quad \text{for } v\in H^1_0(\Omega).$$

To ensure the well-posedness of integrals, we require the following stronger growth condition (guaranteeing **compactness** of the nonlinear reaction term). Indeed, the same assumption is also required for stability (A1) of the residual error estimator (14) below.

(CGC) There holds (GC), if $d\in\{1,2\}$. If $d=3$, there holds (GC) with the stronger assumption $N\in\{2,3\}$.

Residual error estimator. To guarantee well-posedness, we additionally require that $\boldsymbol{A}|_T \in [W^{1,\infty}(T)]^{d\times d}$ and $\boldsymbol{f}|_T\in[W^{1,\infty}(T)]^d$ for all $T\in\mathcal{T}_0$, where $\mathcal{T}_0$ is the initial triangulation of the adaptive algorithm. Then, for $\mathcal{T}_H\in\mathbb{T}$ and $v_H\in\mathcal{X}_H$, the local contributions of the standard residual error estimator (12) for the semilinear model problem (13) read

$$\eta_H(T, v_H)^2 := h_T^2 \, \|f + \mathrm{div}(\boldsymbol{A}\,\nabla v_H - \boldsymbol{f}) - b(v_H)\|^2_{L^2(T)} \\ + h_T \, \|[\![(\boldsymbol{A}\,\nabla v_H - \boldsymbol{f}) \cdot \boldsymbol{n}]\!]\|^2_{L^2(\partial T \cap \Omega)}, \tag{14}$$

where $h_T = |T|^{1/d}$ and where $[\![\cdot]\!]$ denotes the jump across edges (for $d = 2$) resp. faces (for $d = 3$) and $\boldsymbol{n}$ denotes the outer unit normal vector. For $d = 1$, these jumps vanish, i.e., $[\![\cdot]\!] = 0$. The *axioms of adaptivity* are established for the present setting in [17].

Proposition 3 ([17, Proposition 15]) *Suppose* (RHS), (ELL), (CAR), (MON), *and* (CGC). *Suppose that NVB is employed as a refinement strategy. Then, the residual error estimator from* (14) *satisfies* (A1)–(A4) *from Sect.* 2.4. *The constant* C_{rel} *depends only on* d, μ_0, *and uniform shape regularity of the initial mesh* $\mathcal{T}_0$. *The constant* C_{drel} *depends, in addition, on the polynomial degree* p, *and* $C_{\mathrm{stab}}[\vartheta]$ *depends furthermore on* $|\Omega|$, ϑ, N, R, *and* $\boldsymbol{A}$. □

Algebraic solver. As an algebraic solver, we employ a norm-contractive solver to solve the Zarantonello system (9). Possible choices are, e.g., an optimally preconditioned conjugate gradient method [45] or an optimal geometric multigrid [46, 47]. More precisely, the numerical experiments below employ the hp-robust multigrid method from [47], which is well-defined owing to ellipticity (ELL).

3 Fully adaptive algorithm

In this section, we present the adaptive iterative linearized finite element method (AILFEM). As a first main result, we prove that the iterates from the proposed algorithm are uniformly bounded.

3.1 Fully adaptive algorithm

In this section, we introduce a fully adaptive algorithm that steers mesh refinement (ℓ), linearization (k) and the algebraic solver (i). The algorithm utilizes specific stopping indices denoted by an underline, namely $\underline{\ell}$, $\underline{k}[\ell]$, $\underline{i}[\ell, k]$. However, we may omit the dependence when it is apparent from the context, such as in the abbreviation $u_\ell^{k,\underline{i}} := u_\ell^{k,\underline{i}[\ell,k]}$.

For the analysis of Algorithm 1, we define the countably infinite index set

$$\mathcal{Q} := \{(\ell, k, i) \in \mathbb{N}_0^3 \colon u_\ell^{k,i} \text{ is used in Algorithm 1}\},$$

Algorithm 1 adaptive iterative linearized FEM (AILFEM)

Input: Initial mesh $\mathcal{T}_0$, marking parameters $0 < \theta \le 1$, $C_{\mathrm{mark}} \ge 1$, solver parameters $\lambda_{\mathrm{lin}}, \lambda_{\mathrm{alg}} > 0$, minimal number of algebraic solver steps $i_{\min} \in \mathbb{N}$, initial guess $u_0^{0,0} := u_0^{0,\star} := u_0^{0,\underline{i}} \in \mathcal{X}_0$ with $|||u_0^{0,0}||| \le 2M$, and Zarantonello damping parameter $\delta > 0$.

Adaptive loop: For all $\ell = 0, 1, 2, \ldots$, repeat the following steps (I)–(III):

(I) SOLVE & ESTIMATE. For all $k = 1, 2, 3, \ldots$, repeat steps (a)–(c):

(a) Define $u_\ell^{k,0} := u_\ell^{k-1,\underline{i}}$ and, for only theoretical reasons, $u_\ell^{k,\star} := \Phi_\ell(\delta; u_\ell^{k-1,\underline{i}})$.

(b) For all $i = 1, 2, 3, \ldots$ repeat steps (i)–(ii):

(i) Compute $u_\ell^{k,i} := \Psi_\ell(u_\ell^{k,\star}; u_\ell^{k,i-1})$ and error estimator $\eta_\ell(u_\ell^{k,i})$.

(ii) Terminate the i-loop and define $\underline{i}[\ell, k] := i$ if

$$|||u_\ell^{k,i-1} - u_\ell^{k,i}||| \le \lambda_{\mathrm{alg}} \left[\lambda_{\mathrm{lin}}\, \eta_\ell(u_\ell^{k,i}) + |||u_\ell^{k,i} - u_\ell^{k,0}|||\right] \;\;\texttt{AND}\;\; i_{\min} \le i. \tag{15}$$

(c) Terminate the k-loop and define $\underline{k}[\ell] := k$ if

$$\mathcal{E}(u_\ell^{k,0}) - \mathcal{E}(u_\ell^{k,\underline{i}}) \le \lambda_{\mathrm{lin}}^2\, \eta_\ell(u_\ell^{k,\underline{i}})^2 \;\;\texttt{AND}\;\; |||u_\ell^{k,\underline{i}}||| \le 2M. \tag{16}$$

(II) MARK. Find a set $\mathcal{M}_\ell \in \mathbb{M}_\ell[\theta, u_\ell^{\underline{k},\underline{i}}] := \{\mathcal{U}_\ell \subseteq \mathcal{T}_\ell : \theta\, \eta_\ell(u_\ell^{\underline{k},\underline{i}})^2 \le \eta_\ell(\mathcal{U}_\ell, u_\ell^{\underline{k},\underline{i}})^2\}$ such that

$$\#\mathcal{M}_\ell \le C_{\mathrm{mark}} \min_{\mathcal{U}_\ell \in \mathbb{M}_\ell[\theta, u_\ell^{\underline{k},\underline{i}}]} \#\mathcal{U}_\ell. \tag{17}$$

(III) REFINE. Generate the new mesh $\mathcal{T}_{\ell+1} := \mathtt{refine}(\mathcal{M}_\ell, \mathcal{T}_\ell)$ by employing NVB and define $u_{\ell+1}^{0,0} := u_{\ell+1}^{0,\underline{i}} := u_{\ell+1}^{0,\star} := u_\ell^{\underline{k},\underline{i}}$ (nested iteration).

Output: Sequences of successively refined triangulations $\mathcal{T}_\ell$, discrete approximations $u_\ell^{k,i}$ and corresponding error estimators $\eta_\ell(u_\ell^{k,i})$.

where, for any $(\ell, 0, 0) \in \mathcal{Q}$, the final indices are defined as

$$\begin{aligned}
\underline{\ell} &:= \sup\{\ell \in \mathbb{N}_0 : (\ell, 0, 0) \in \mathcal{Q}\} \;\in \mathbb{N}_0 \cup \{\infty\},\\
\underline{k}[\ell] &:= \sup\{k \in \mathbb{N} : (\ell, k, 0) \in \mathcal{Q}\} \;\in \mathbb{N} \cup \{\infty\},\\
\underline{i}[\ell, k] &:= \sup\{i \in \mathbb{N} : (\ell, k, i) \in \mathcal{Q}\} \;\in \mathbb{N} \cup \{\infty\}.
\end{aligned}$$

We note, first, that these definitions are consistent with those of Algorithm 1, second, that Lemma 2 below proves that $\underline{i}[\ell, k] < \infty$, and, third, that hence either $\underline{\ell} = \infty$ or $\underline{\ell} < \infty$ with $\underline{k}[\underline{\ell}] = \infty$. For all $(\ell, k, i) \in \mathcal{Q}$, we introduce the total step counter $|\cdot, \cdot, \cdot|$ defined by

$$\begin{aligned}
|\ell, k, i| &:= \#\{(\ell', k', i') \in \mathcal{Q} : (\ell', k', i') < (\ell, k, i)\}\\
&= \sum_{\ell'=0}^{\ell-1} \sum_{k'=1}^{\underline{k}[\ell']} \sum_{i'=1}^{\underline{i}[\ell',k']} 1 + \sum_{k'=1}^{k-1} \sum_{i'=1}^{\underline{i}[\ell,k']} 1 + \sum_{i'=1}^{i-1} 1.
\end{aligned}$$

We note that this definition provides a lexicographic ordering on $\mathcal{Q}$.

In the later application to AILFEM for semilinear elliptic PDEs, every step of Algorithm 1 can be performed in linear complexity as the following arguments show.

▹ SOLVE. The employed algebraic solver is an hp-robust multigrid [47] and hence each algebraic solver step requires only $\mathcal{O}(\#\mathcal{T}_\ell)$ operations.
▹ ESTIMATE. The simultaneous computation of the standard error indicators $\eta_\ell(T, u_\ell^{k,i})$ for all $T \in \mathcal{T}_\ell$ can be done at the cost of $\mathcal{O}(\#\mathcal{T}_\ell)$.
▹ MARK. The employed Dörfler marking (and the involved determination of $\mathcal{M}_\ell$) is indeed a linear complexity problem; see [5] for $C_{\mathrm{mark}} = 2$ and [48] for $C_{\mathrm{mark}} = 1$.
▹ REFINE. The refinement of $\mathcal{T}_\ell$ is based on NVB and, owing to the mesh-closure estimate [4, 39], requires only linear cost $\mathcal{O}(\#\mathcal{T}_\ell)$.

Thus, the total work until and including the computation of $u_\ell^{k,i}$ is proportional to

$$\mathtt{cost}(\ell, k, i) := \sum_{\substack{(\ell',k',i')\in\mathcal{Q} \\ |\ell',k',i'|\le|\ell,k,i|}} \#\mathcal{T}_{\ell'} = \sum_{\ell'=0}^{\ell-1}\sum_{k'=1}^{\underline{k}[\ell']}\sum_{i'=1}^{\underline{i}[\ell',k']} \#\mathcal{T}_{\ell'} + \sum_{k'=1}^{k-1}\sum_{i'=1}^{\underline{i}[\ell,k']} \#\mathcal{T}_\ell + \sum_{i'=1}^{i} \#\mathcal{T}_\ell. \tag{18}$$

An important observation is that the algebraic solver loop always terminates.

Lemma 2 *Independently of the adaptivity parameters θ, λ_{lin}, and λ_{alg}, the i-loop of Algorithm 1 always terminates, i.e., $\underline{i}[\ell, k] < \infty$ for all $(\ell, k, 0) \in \mathcal{Q}$.*

Proof We argue as in [18, Lemma 3.2]. Let $(\ell, k, 0) \in \mathcal{Q}$. We argue by contradiction and assume that the i-loop stopping criterion (15) in Algorithm 1(I.b.ii) always fails and hence $\underline{i}[\ell, k] = \infty$. By assumption (11), the algebraic solver $\Psi_\ell(u_\ell^{k,\star}; \cdot)$ is contractive and hence convergent with limit $u_\ell^{k,\star} := \Phi_\ell(\delta; u_\ell^{k-1,\underline{i}})$ from Algorithm 1(I.a). Moreover, by failure of the stopping criterion (15) in Algorithm 1(I.b.ii), we thus obtain that

$$\eta_\ell(u_\ell^{k,i}) + |||u_\ell^{k,i} - u_\ell^{k,0}||| \overset{(15)}{\lesssim} |||u_\ell^{k,i} - u_\ell^{k,i-1}||| \xrightarrow{i\to\infty} 0.$$

This yields $|||u_\ell^{k,\star} - u_\ell^{k,0}||| = 0$ and hence $u_\ell^{k,\star} = u_\ell^{k,i}$ for all $i \in \mathbb{N}_0$, since the algebraic solver is contractive. Consequently, the i-loop stopping criterion (15) in Algorithm 1(I.b.ii) will be satisfied for $i = i_{\min}$. This contradicts our assumption, and hence we conclude that $\underline{i}[\ell, k] < \infty$. □

3.2 Energy contraction for the inexact Zarantonello iteration

In this section, we prove uniform boundedness of the iterates $u_\ell^{k,i}$ from Algorithm 1: Note that the algorithm does not compute the Zarantonello iterate $u_\ell^{k,\star} := \Phi_\ell(\delta; u_\ell^{k-1,\underline{i}})$ exactly, but relies on an approximation $u_\ell^{k,\underline{i}} \approx u_\ell^{k,\star}$. We prove that this inexact Zarantonello iteration is contractive with respect to the energy, which is the case if at least $i_{\min} \in \mathbb{N}$ steps of the contractive algebraic solver are performed, i.e., $\underline{i}[\ell, k] \ge i_{\min}$. In particular, a suitable choice of the damping parameter $\delta > 0$ and the index $i_{\min}$ are derived in the following.

Theorem 4 *Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and (POT). With M from (4), define $\tau := M + 3M\big(\frac{L[3M]}{\alpha}\big)^{1/2} \geq 4M$. Let $\lambda_{\mathrm{lin}}, \lambda_{\mathrm{alg}} > 0$ and $0 < \theta \leq 1$ be arbitrary. Suppose that $|||u_\ell^{0,0}||| = |||u_\ell^{0,\underline{i}}||| \leq 2\,M$ with $M > 0$ from (4). Choose $i_{\min} \in \mathbb{N}$ such that*

$$q_{\mathrm{alg}}^{i_{\min}} \leq 1/3. \tag{19}$$

Then, for any choice of $\delta > 0$ satisfying $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$, there exists a uniform energy contraction constant $0 < q_{\mathcal{E}} = q_{\mathcal{E}}[\delta, \tau] < 1$ (see (32b) below) such that the following holds.

$$\triangleright\ \textbf{\textit{nested iteration:}}\quad |||u_\ell^{\underline{k},\underline{i}}||| \leq 2M \qquad \text{if } (\ell, \underline{k}, \underline{i}) \in \mathcal{Q}; \tag{20}$$

$$\triangleright\ \underline{i}\textbf{\textit{-uniform bound:}}\quad |||u_\ell^{k,\underline{i}}||| \leq \tau \qquad \text{if } (\ell, k, \underline{i}) \in \mathcal{Q}; \tag{21}$$

$$\triangleright\ \mathcal{E}\textbf{\textit{-contraction:}}\quad \mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star) \leq q_{\mathcal{E}}^2\,\big(\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big) \quad \text{if } (\ell, k+1, \underline{i}) \in \mathcal{Q}. \tag{22}$$

With (20)–(22), we obtain for all iterates the

$$\triangleright\ \textbf{\textit{uniform bound:}}\quad |||u_\ell^{k,i}||| \leq 5\tau \qquad \text{if } (\ell, k, i) \in \mathcal{Q}. \tag{23}$$

Moreover, there exists an index $k_0 = k_0[\delta, \tau, \alpha, L[3M], M] \in \mathbb{N}$ independently of the mesh refinement index ℓ such that, for all $k' \geq k_0$, the nested iteration condition $|||u_\ell^{k',\underline{i}}||| \leq 2\,M$ in the k-loop stopping criterion (16) is always met.

The main observation of the following lemma is that the uniform boundedness is passed on by the inexact Zarantonello iteration along the k-loop indices.

Lemma 3 *Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and (POT). Let $\lambda_{\mathrm{lin}}, \lambda_{\mathrm{alg}} > 0$ be arbitrary and define $\tau := M + 3\,M\big(\frac{L[3\,M]}{\alpha}\big)^{1/2} \geq 4\,M$. Let $k \in \mathbb{N}_0$ with $0 \leq k < \underline{k}[\ell]$ and*

$$|||u_\ell^{k,\underline{i}}||| \leq \tau. \tag{24}$$

Then, for $i_{\min} \in \mathbb{N}$ satisfying (19) and for any $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$, it holds that

$$\begin{aligned} 0 \leq \Big(\frac{1}{2\delta} - \frac{L[5\tau]}{2}\Big)\, |||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}|||^2 &\leq \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) \\ &\leq \Big(\frac{1}{\delta\,(1 - q_{\mathrm{alg}}^{i_{\min}})} - \frac{\alpha}{2}\Big)\, |||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}|||^2 \text{ for all } (\ell, k+1, \underline{i}) \in \mathcal{Q}. \end{aligned} \tag{25}$$

Proof The proof is subdivided into five steps.

Step 1 (choice of $i_{\min}$). We note that for any $i_{\min} \in \mathbb{N}$, the property (19) is indeed equivalent to

$$\frac{1}{2} \overset{!}{\leq} \frac{1 - 2\,q_{\mathrm{alg}}^{i}}{1 - q_{\mathrm{alg}}^{i}} \quad \text{for all } i \geq i_{\min}. \tag{26}$$

Step 2 (boundedness). Define $e_\ell^{k+1} := u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}$. Recall that for $0 < \delta < 2\alpha/L[2\tau]^2$, the Zarantonello iteration satisfies contraction (10). Hence, the contraction of the algebraic solver (11), the triangle inequality, nested iteration $u_\ell^{k+1,0} = u_\ell^{k,\underline{i}}$, assumption (24), and $4\,M \leq \tau$ show that

$$\begin{aligned}
|||e_\ell^{k+1}||| &\leq |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| + |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \\
&\overset{(11)}{\leq} q_{\mathrm{alg}}^{\underline{i}[\ell,k+1]}\, |||u_\ell^{k+1,\star} - u_\ell^{k+1,0}||| + |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \\
&\leq 2\,|||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \leq 2\left[|||u_\ell^{\star} - u_\ell^{k,\underline{i}}||| + |||u_\ell^{\star} - u_\ell^{k+1,\star}|||\right] \\
&\overset{(10)}{\leq} 2\,(1 + q_{\mathrm{Zar}}^{\star}[\delta, 2\tau])\,|||u_\ell^{\star} - u_\ell^{k,\underline{i}}||| \overset{(24)}{\leq} 4(M + \tau) \leq 5\tau.
\end{aligned}$$

With the convexity of the norm and $|||u_\ell^{k,\underline{i}}||| \leq \tau \leq 5\tau$, we also obtain that

$$|||u_\ell^{k+1,\underline{i}}||| \leq \max_{0 \leq t \leq 1} |||u_\ell^{k,\underline{i}} - t\,e_\ell^{k+1}||| \leq 5\tau. \tag{27}$$

Step 3. Since the energy $\mathcal{E} = \mathcal{P} - F$ from (POT) is Gâteaux differentiable, it follows that $\varphi(t) := \mathcal{E}(u_\ell^{k,\underline{i}} + t\,e_\ell^{k+1})$ is differentiable with

$$\varphi'(t) = \langle \mathrm{d}\mathcal{E}(u_\ell^{k,\underline{i}} + t\,e_\ell^{k+1})\,,\, e_\ell^{k+1}\rangle = \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\,e_\ell^{k+1}) - F\,,\, e_\ell^{k+1}\rangle. \tag{28}$$

The fundamental theorem of calculus and the exact Zarantonello iteration (9) show that

$$\begin{aligned}
&\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) = \varphi(0) - \varphi(1) \\
&= -\int_0^1 \varphi'(t)\,\mathrm{d}t \overset{(28)}{=} -\int_0^1 \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\,e_\ell^{k+1}) - F\,,\, e_\ell^{k+1}\rangle\,\mathrm{d}t \\
&= -\int_0^1 \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\,e_\ell^{k+1}) - \mathcal{A}(u_\ell^{k,\underline{i}})\,,\, e_\ell^{k+1}\rangle\,\mathrm{d}t - \langle \mathcal{A}(u_\ell^{k,\underline{i}}) - F\,,\, e_\ell^{k+1}\rangle \\
&\overset{(9)}{=} -\int_0^1 \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\,e_\ell^{k+1}) - \mathcal{A}(u_\ell^{k,\underline{i}})\,,\, e_\ell^{k+1}\rangle\,\mathrm{d}t + \frac{1}{\delta}\,\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}\,,\, e_\ell^{k+1}\rangle\!\rangle.
\end{aligned} \tag{29}$$

Step 4 (proof of lower bound in (25)). For any $i \in \mathbb{N}$ with $i \leq \underline{i}[\ell, k]$, the contraction (11) of the algebraic solver and nested iteration $u_\ell^{k,\underline{i}} = u_\ell^{k+1,0}$ prove that

$$\begin{aligned} |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| &\overset{(11)}{\leq} q_{\mathrm{alg}}^{\underline{i}[\ell,k+1]} \, |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \\ &\leq q_{\mathrm{alg}}^{i} \, |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| + q_{\mathrm{alg}}^{i} \, |||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}|||. \end{aligned}$$

This gives rise to the a posteriori estimate

$$|||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| \leq \frac{q_{\mathrm{alg}}^{i}}{1 - q_{\mathrm{alg}}^{i}} \, |||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}||| = \frac{q_{\mathrm{alg}}^{i}}{1 - q_{\mathrm{alg}}^{i}} \, |||e_\ell^{k+1}|||. \tag{30}$$

With (30), $i_{\min} \leq i \leq \underline{i}[\ell, k+1]$, and (26), we derive

$$\begin{aligned} &\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}, \, e_\ell^{k+1} \rangle\!\rangle = \langle\!\langle u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}, \, e_\ell^{k+1} \rangle\!\rangle + \langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}, \, e_\ell^{k+1} \rangle\!\rangle \\ &= |||e_\ell^{k+1}|||^2 + \langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}, \, e_\ell^{k+1} \rangle\!\rangle \geq |||e_\ell^{k+1}|||^2 - |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| \, |||e_\ell^{k+1}||| \\ &\overset{(30)}{\geq} |||e_\ell^{k+1}||| \Big[|||e_\ell^{k+1}||| - \frac{q_{\mathrm{alg}}^{i}}{1 - q_{\mathrm{alg}}^{i}} \, |||e_\ell^{k+1}||| \Big] = \Big(\frac{1 - 2\,q_{\mathrm{alg}}^{i}}{1 - q_{\mathrm{alg}}^{i}} \Big) |||e_\ell^{k+1}|||^2 \\ &\overset{(26)}{\geq} \frac{1}{2} \, |||e_\ell^{k+1}|||^2 \geq 0. \end{aligned} \tag{31}$$

With the local Lipschitz continuity (LIP) and (27), it follows from (29) that

$$\begin{aligned} \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) &\overset{\text{(LIP)}}{\geq} -\Big(\int_0^1 t L[5\tau] \, \mathrm{d}t \Big) \, |||e_\ell^{k+1}|||^2 + \frac{1}{\delta} \, \langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}, \, e_\ell^{k+1} \rangle\!\rangle \\ &\overset{(31)}{\geq} \Big[\frac{1}{2\,\delta} - \frac{L[5\tau]}{2} \Big] \, |||e_\ell^{k+1}|||^2. \end{aligned}$$

Since $0 < \delta < 1/L[5\tau]$, the last expression is positive.

Step 5 (proof of upper bound in (25)). To derive the upper equivalence constant, we infer from Step 4 that

$$\begin{aligned} \langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}, \, e_\ell^{k+1} \rangle\!\rangle &\leq |||e_\ell^{k+1}|||^2 + |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| \, |||e_\ell^{k+1}||| \\ &\overset{(30)}{\leq} |||e_\ell^{k+1}||| \Big[|||e_\ell^{k+1}||| + \frac{q_{\mathrm{alg}}^{i}}{1 - q_{\mathrm{alg}}^{i}} \, |||e_\ell^{k+1}||| \Big] \\ &= \Big(\frac{1}{1 - q_{\mathrm{alg}}^{i}} \Big) |||e_\ell^{k+1}|||^2. \end{aligned}$$

Combined with Step 3, we obtain that

$$\begin{aligned}\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) &\overset{(29)}{=} -\int_0^1 \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\, e_\ell^{k+1}) - \mathcal{A}(u_\ell^{k,\underline{i}}),\; e_\ell^{k+1}\rangle \,\mathrm{d}t \\ &\quad + \frac{1}{\delta}\, \langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}},\; e_\ell^{k+1}\rangle\!\rangle \\ &\overset{(\mathrm{SM})}{\le} -\Big(\int_0^1 t\alpha\,\mathrm{d}t\Big)\, |||e_\ell^{k+1}|||^2 + \frac{1}{\delta}\,\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}},\; e_\ell^{k+1}\rangle\!\rangle \\ &\le \Big(\frac{1}{\delta\,(1-q_{\mathrm{alg}}^{i_{\min}})} - \frac{\alpha}{2}\Big)\, |||e_\ell^{k+1}|||^2.\end{aligned}$$

This concludes the proof. □

Lemma 4 (Energy contraction) *Suppose the assumptions of Lemma 3. Recall* $i_{\min} \in \mathbb{N}$ *from* (19)*. Then, for* $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$*, it holds that*

$$0 \le \mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star) \le q_{\mathcal{E}}[\delta,\tau]^2\, [\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)] \tag{32a}$$

with the contraction constant

$$0 \le q_{\mathcal{E}}[\delta,\tau]^2 := 1 - \Big(\frac{1}{\delta} - L[5\tau]\Big)\, \frac{(1-q_{\mathrm{alg}}^{i_{\min}})^2\, \delta^2\alpha^2}{L[2\tau]} < 1. \tag{32b}$$

We note that $q_{\mathcal{E}}[\delta,\tau] \to 1$ *as* $\delta \to 0$*. In particular, it holds that*

$$(1 - q_{\mathcal{E}}[\delta,\tau]^2)\, \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big] \le \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) \le \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star). \tag{33}$$

Proof First, we observe that

$$\begin{aligned}\alpha\, |||u_\ell^\star - u_\ell^{k,\underline{i}}|||^2 &\overset{(\mathrm{SM})}{\le} \langle \mathcal{A}u_\ell^\star - \mathcal{A}u_\ell^{k,\underline{i}},\; u_\ell^\star - u_\ell^{k,\underline{i}}\rangle \overset{(3)}{=} \langle F - \mathcal{A}u_\ell^{k,\underline{i}},\; u_\ell^\star - u_\ell^{k,\underline{i}}\rangle \\ &\overset{(9)}{=} \frac{1}{\delta}\, \langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}},\; u_\ell^\star - u_\ell^{k,\underline{i}}\rangle\!\rangle \\ &\le \frac{1}{\delta}\, |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}|||\; |||u_\ell^\star - u_\ell^{k,\underline{i}}|||.\end{aligned} \tag{34}$$

The inverse triangle inequality and contraction (11) of the algebraic solver prove that

$$\begin{aligned}|||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}||| &\ge |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| - |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| \\ &\overset{(11)}{\ge} (1-q_{\mathrm{alg}}^{i_{\min}})\, |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \overset{(34)}{\ge} (1-q_{\mathrm{alg}}^{i_{\min}})\,\delta\,\alpha\, |||u_\ell^\star - u_\ell^{k,\underline{i}}|||.\end{aligned} \tag{35}$$

Since $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$, it follows that

$$
\begin{aligned}
0 &\overset{(6)}{\le} \mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star) = \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star) - \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}})\big] \\
&\overset{(25)}{\le} \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star) - \Big(\frac{1}{2\delta} - \frac{L[5\tau]}{2}\Big) |||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}|||^2 \\
&\overset{(35)}{\le} \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star) - \Big(\frac{1}{2\delta} - \frac{L[5\tau]}{2}\Big) (1 - q_{\mathrm{alg}}^{i_{\min}})^2\, \delta^2\, \alpha^2\, |||u_\ell^\star - u_\ell^{k,\underline{i}}|||^2 \\
&\overset{(6)}{\le} \Big(1 - \big[1 - \delta\, L[5\tau]\big] \frac{(1 - q_{\mathrm{alg}}^{i_{\min}})^2\, \alpha^2\, \delta}{L[2\tau]}\Big) \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big] \\
&=: q_{\mathcal{E}}[\delta, \tau]^2 \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big].
\end{aligned}
$$

We may rewrite $q_{\mathcal{E}}[\delta, \tau]^2 = 1 - C\delta + C\, L[5\tau]\, \delta^2$ with $C = \frac{(1-q_{\mathrm{alg}}^{i_{\min}})^2\, \alpha^2}{L[2\tau]}$. Since $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\} \le \frac{1}{L[5\tau]}$, we obtain that $0 < q_{\mathcal{E}}[\delta, \tau] < 1$. This proves (32). The lower inequality in (33) follows from the triangle inequality. The upper inequality in (33) holds due to $0 \le \mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star)$ and hence $\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) = \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star) + \mathcal{E}(u_\ell^\star) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) \le \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)$. This concludes the proof. □

Proof of Theorem 4 The proof consists of four steps.

Step 1 (proof of (21)–(22) for $k = 0$ and all $\ell \in \mathbb{N}_0$). Let $\ell \in \mathbb{N}_0$ with $\ell \le \underline{\ell}$ be arbitrary, but fixed. From the initial guess $u_0^{0,0}$ or Algorithm 1(I.c) and $u_\ell^{0,\underline{i}} = u_\ell^{0,0} = u_{\ell-1}^{\underline{k},\underline{i}}$ for any $\ell \in \mathbb{N}$, we have that $|||u_\ell^{0,0}||| \le 2\, M$ and a fortiori $|||u_\ell^{0,0}||| \le \tau$. This proves (21) for $k = 0$ and all $\ell \in \mathbb{N}_0$ with $\ell \le \underline{\ell}$ (even with the stronger bound $2M \le \tau$).

In particular, we may apply Lemma 4 to obtain that $\mathcal{E}(u_\ell^{1,\underline{i}}) - \mathcal{E}(u_\ell^\star) \le q_{\mathcal{E}}[\delta, \tau]^2 \big[\mathcal{E}(u_\ell^{0,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]$, which proves (22) for $k = 0$ and $\ell \in \mathbb{N}_0$.

Step 2 (proof of (21)–(22) for $k \ge 0$ and all $\ell \in \mathbb{N}_0$). Let $\ell \in \mathbb{N}_0$ with $\ell \le \underline{\ell}$. We argue by induction on k, where Step 1 proves the base case $k = 0$. Hence, we may assume that boundedness (21) holds for all $0 \le k' \le k$. Lemma 4 applied separately for all $0 \le k' \le k$ yields energy contraction (32) for the indices $0 \le k' \le k$. Overall, we obtain that

$$
\begin{aligned}
\mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star) &\overset{(32)}{\le} q_{\mathcal{E}}[\delta, \tau]^2 \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big] \\
&\overset{(32)}{\le} q_{\mathcal{E}}[\delta, \tau]^{2(k+1)} \big[\mathcal{E}(u_\ell^{0,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big],
\end{aligned} \tag{36}
$$

where we only used energy contraction (32) for $0 \le k' \le k$, i.e., for indices that are covered by the induction hypothesis. From (36), $|||u_\ell^\star||| \le M$ from (4), and $|||u_\ell^{0,\underline{i}}||| \le 2\, M$ and $u_\ell^{0,\underline{i}} = u_\ell^{0,0}$ from Step 1, we obtain that

$$
\begin{aligned}
|||u_\ell^{k+1,\underline{i}}||| &\le |||u_\ell^\star||| + |||u_\ell^\star - u_\ell^{k+1,\underline{i}}||| \\
&\overset{(6)}{\le} M + \Big(\frac{2}{\alpha}\Big)^{1/2} \big[\mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \\
&\overset{(36)}{\le} M + q_{\mathcal{E}}^{k+1} \Big(\frac{2}{\alpha}\Big)^{1/2} \big[\mathcal{E}(u_\ell^{0,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \\
&\overset{(6)}{\le} M + q_{\mathcal{E}}^{k+1} \Big(\frac{L[3M]}{\alpha}\Big)^{1/2} |||u_\ell^\star - u_\ell^{0,\underline{i}}||| \le M \\
&\quad + q_{\mathcal{E}}^{k+1} \Big(\frac{L[3M]}{\alpha}\Big)^{1/2} 3M \le \tau. \qquad (37)
\end{aligned}
$$

Thus, boundedness (21) is satisfied for $0 \le k' \le k+1$. Again, Lemma 8 yields energy contraction for $0 \le k' \le k+1$. This completes the induction argument and concludes that (21)–(22) hold for all $\ell \in \mathbb{N}_0$ and all $k \in \mathbb{N}_0$.

Step 3 (uniform boundedness). Contraction of the algebraic solver (11), the straightforward estimate from the exact Zarantonello iteration (9), $|||u^\star||| \le M \le \tau$ from (4), $|||u_\ell^{k,0}||| \le \tau$ from (21), and the constraint $\delta < \min\{1/L[5\tau], 2\alpha/L[2\tau]^2\}$ which ensures that $\delta L[2\tau] \le \delta L[5\tau] < 1$, yield that

$$
|||u_\ell^{k,\star} - u_\ell^{k,0}||| = |||\Phi_\ell(\delta; u_\ell^{k,0}) - u_\ell^{k,0}||| \le \delta \, \|F - \mathcal{A}(u_\ell^{k,0})\|_{\mathcal{X}'} \overset{\text{(LIP)}}{\le} \delta \, L[2\tau] \, |||u^\star - u_\ell^{k,0}||| < 2\tau.
$$

With $|||u_\ell^{k,\star}||| \le |||u_\ell^{k,0}||| + |||u_\ell^{k,\star} - u_\ell^{k,0}||| \le 3\tau$ owing to (20), it follows that

$$
|||u_\ell^{k,i}||| \overset{(11)}{\le} |||u_\ell^{k,\star}||| + q_{\mathrm{alg}}^i \, |||u_\ell^{k,\star} - u_\ell^{k,0}||| \le 5\tau \quad \text{for all } (\ell, k, i) \in \mathcal{Q}.
$$

Step 4 (existence of k_0). Let $\ell \in \mathbb{N}_0$ with $\ell \le \underline{\ell}$. As in (37) from Step 2, we obtain

$$
|||u_\ell^{k,\underline{i}}||| \le M + q_{\mathcal{E}}^k \Big(\frac{L[3M]}{\alpha}\Big)^{1/2} 3M.
$$

Clearly, there exists a minimal integer $k_0 = k_0[q_{\mathcal{E}}, \alpha, L[3\,M]] = k_0[\delta, \tau, \alpha, L[3\,M], M] \in \mathbb{N}$ such that, for all $k \ge k_0$, it holds that

$$
M + q_{\mathcal{E}}^k \Big(\frac{L[3M]}{\alpha}\Big)^{1/2} 3M \le 2M.
$$

In particular, k_0 is independent of the mesh level ℓ and $|||u_\ell^{k,\underline{i}}||| \le 2\,M$ for all $k_0 \le k \le \underline{k}[\ell]$. This concludes the proof. □

Remark 2 (i) According to uniform boundedness (23), all involved Lipschitz constants or stability constants are uniformly bounded by $L[10\tau]$ and $C_{\mathrm{stab}}[10\tau]$, respectively.

(ii) Under the assumption that $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$, energy contraction (22) and the lower bound in the norm-energy equivalence (25) are even equivalent, i.e.,

$$
(22) \quad \Longleftrightarrow \quad (25).
$$

To see this, recall that the proof of energy contraction (22) in Lemma 4 exploits (25). The converse implication is obtained as follows: First, energy contraction yields

$$\begin{aligned}\mathcal{E}(u_\ell^{k+1,\star}) - \mathcal{E}(u_\ell^\star) &\le q_{\mathcal{E}}^2\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]\\ &= q_{\mathcal{E}}^2\,\big\{\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\star})\big]+\big[\mathcal{E}(u_\ell^{k+1,\star}) - \mathcal{E}(u_\ell^\star)\big]\big\}\end{aligned} \tag{38}$$

which gives rise to the a posteriori estimate

$$0 \le \mathcal{E}(u_\ell^{k+1,\star}) - \mathcal{E}(u_\ell^\star) \le \frac{q_{\mathcal{E}}^2}{1-q_{\mathcal{E}}^2}\,\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\star})\big]. \tag{39}$$

In particular, we note that the energy difference on the right-hand side is nonnegative. Exploiting uniform boundedness (21), the last inequality yields that

$$\begin{aligned}|||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}|||^2 &\lesssim |||u_\ell^\star - u_\ell^{k+1,\star}|||^2 + |||u_\ell^\star - u_\ell^{k,\underline{i}}|||^2\\ &\overset{(9)}{\le} \big(1 + (q_{\mathrm{Zar}}^\star[\delta, 2\tau])^2\big)\,|||u_\ell^\star - u_\ell^{k,\underline{i}}|||^2\\ &\overset{(6)}{\lesssim} \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\star})\big] + \big[\mathcal{E}(u_\ell^{k+1,\star}) - \mathcal{E}(u_\ell^\star)\big]\\ &\overset{(39)}{\le} \frac{1}{1-q_{\mathcal{E}}^2}\,\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\star})\big].\end{aligned}$$

This concludes the argument.

Remark 3 (i) The stopping criteria (15) and (16) read schematically

`[accuracy criterion] AND [iteration criterion].`

(ii) The accuracy criterion in (16) is heuristically motivated by the fact that the discretization error (estimated by $\eta_\ell(\cdot)$) shall dominate the linearization error

$$\begin{aligned}\frac{\alpha}{2}\,|||u_\ell^\star - u_\ell^{k+1,\underline{i}}|||^2 &\overset{(6)}{\le} \mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\\ &\overset{(22)}{\le} \frac{q_{\mathcal{E}}^2}{1-q_{\mathcal{E}}^2}\,\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}})\big] \overset{(16)}{\lesssim} \lambda_{\mathrm{lin}}^2\,\eta_\ell(u_\ell^{k+1,\underline{i}})^2.\end{aligned} \tag{40}$$

This allows a posteriori error control over the linearization error by means of computable energy differences.

(iii) The accuracy criterion (15) is satisfied given that the discretization and linearization error dominate the algebraic error in the sense of

$$\begin{aligned}|||u_\ell^{k,\star} - u_\ell^{k,i}||| &\overset{(11)}{\le} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}}\,|||u_\ell^{k,i} - u_\ell^{k,i-1}|||\\ &\overset{(15)}{\le} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}}\,\lambda_{\mathrm{alg}}\,\big[\lambda_{\mathrm{lin}}\,\eta_\ell(u_\ell^{k,i}) + |||u_\ell^{k,i} - u_\ell^{k,0}|||\big].\end{aligned} \tag{41}$$

Once the i-loop is stopped, the equivalence (25) and nested iteration $u_\ell^{k,0} = u_\ell^{k-1,\underline{i}}$ yield $|||u_\ell^{k,\underline{i}} - u_\ell^{k,0}|||^2 = |||u_\ell^{k,\underline{i}} - u_\ell^{k-1,\underline{i}}|||^2 \simeq \mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^{k,\underline{i}})$.

4 Full R-linear convergence

We prove full R-linear convergence of Algorithm 1 by adapting the analysis of [19, 21]. The new result extends [36, Theorem 13], where an exact solve for the Zarantonello iteration (9) is supposed. The new proof is built on a summability argument, but the stopping criteria (15)–(16) with iteration count criteria require further analysis to prove full R-linear convergence even (and unlike [19, 21]) for arbitrary adaptivity parameters $0 < \theta \leq 1$, $\lambda_{\rm lin} > 0$ and $\lambda_{\rm alg} > 0$.

Theorem 5 (Full R-linear convergence of Algorithm 1) *Suppose the assumptions of Theorem 4. Suppose the axioms of adaptivity (A1)–(A3). Let* $\lambda_{\rm lin}, \lambda_{\rm alg} > 0$, $0 < \theta \leq 1$, $C_{\rm mark} \geq 1$, *and* $u_0^{0,0} \in \mathcal{X}_0$ *with* $|||u_0^{0,0}||| \leq 2M$. *Then, Algorithm 1 guarantees full R-linear convergence of the quasi-error*

$$\mathrm{H}_\ell^{k,i} := |||u_\ell^\star - u_\ell^{k,i}||| + |||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i}), \tag{42}$$

i.e., there exist constants $0 < q_{\rm lin} < 1$ *and* $C_{\rm lin} > 0$ *such that*

$$\mathrm{H}_\ell^{k,i} \leq C_{\rm lin} q_{\rm lin}^{|\ell,k,i|-|\ell',k',i'|}\, \mathrm{H}_{\ell'}^{k',i'} \text{ for all } (\ell',k',i'), (\ell,k,i) \in \mathcal{Q} \text{ with } |\ell',k',i'| < |\ell,k,i|. \tag{43}$$

The constant $q_{\rm lin}$ *depends only on* θ, $q_{\rm red}$ *from (A2),* $q_{\rm Zar}^\star[\delta, 2\tau]$ *from Proposition 2,* $q_{\mathcal{E}}$ *from Theorem 4, and* $q_{\rm alg}$ *from (11). The constant* $C_{\rm lin}$ *depends only on* M, α, $C_{\text{Céa}}[2M]$, $q_{\rm Zar}^\star[\delta; 2\tau]$, $\lambda_{\rm lin}$, $q_{\rm alg}$, $\lambda_{\rm alg}$, $C_{\rm rel}$, $C_{\rm stab}[10\tau]$, *and* $i_{\rm min}$.

Proof of Theorem 5 The proof is split into seven steps.

Step 1 (equivalences of quasi-error quantities). Throughout the proof, we approach $\mathrm{H}_\ell^{k,i}$ from (42) after introducing auxiliary quantities such as

$$\mathrm{H}_\ell^k := [\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)]^{1/2} + \eta_\ell(u_\ell^{k,\underline{i}}) \quad \text{for all } (\ell,k,\underline{i}) \in \mathcal{Q} \tag{44}$$

and

$$\mathrm{H}_\ell := [\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_\ell^\star)]^{1/2} + \gamma\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}) \overset{(44)}{\simeq} \mathrm{H}_\ell^{\underline{k}} \quad \text{for all } (\ell,\underline{k},\underline{i}) \in \mathcal{Q}, \tag{45}$$

where $0 < \gamma < 1$ is a free parameter to be fixed later in (51) below. In the following, we show that $\mathrm{H}_\ell^{k,\underline{i}} \simeq \mathrm{H}_\ell^k \overset{(45)}{\simeq} \mathrm{H}_\ell$. First, note that the equivalence of energy and norm from (6) (with $L[2\tau]$ from boundedness (21) and (4)) yields that

$$\mathrm{H}_\ell^k \leq \mathrm{H}_\ell^k + |||u_\ell^{k,\star} - u_\ell^{k,\underline{i}}||| \overset{(6)}{\simeq} \mathrm{H}_\ell^{k,\underline{i}} \quad \text{for all } (\ell,k,\underline{i}) \in \mathcal{Q}. \tag{46}$$

The a posteriori estimate (41) for the algebraic solver from Remark 3(iii), norm-energy equivalence (25), and the stopping criterion (16) show that

$$|||u_\ell^{k,\star} - u_\ell^{k,\underline{i}}||| \overset{(41)}{\le} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} \lambda_{\mathrm{alg}} \big[\lambda_{\mathrm{lin}}\, \eta_\ell(u_\ell^{k,\underline{i}}) + |||u_\ell^{k,\underline{i}} - u_\ell^{k,0}|||\big]$$
$$\overset{(25)}{\lesssim} \eta_\ell(u_\ell^{k,\underline{i}}) + \big[\mathcal{E}(u_\ell^{k,0}) - \mathcal{E}(u_\ell^{k,\underline{i}})\big]^{1/2} \overset{(16)}{\lesssim} \eta_\ell(u_\ell^{k,\underline{i}}) \le \mathrm{H}_\ell^{k}.$$

With (46), we conclude that $\mathrm{H}_\ell \simeq \mathrm{H}_\ell^{\underline{k}} \simeq \mathrm{H}_\ell^{\underline{k},\underline{i}}$.

Step 2 (estimator reduction). The axioms (A1)–(A2) and Dörfler marking (17) prove the estimator reduction estimate (cf., e.g., [14, Equation (52)])

$$\eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{i}}) \le q_\theta\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}) + C_{\mathrm{stab}}[4M]\, |||u_{\ell+1}^{\underline{k},\underline{i}} - u_\ell^{\underline{k},\underline{i}}||| \quad \text{for all } \ell \in \mathbb{N}_0, \tag{47}$$

where $4M$ stems from nested iteration (20) from Theorem 4. Moreover, the triangle inequality, the equivalence (6), and energy contraction (22) give that

$$|||u_{\ell+1}^{\underline{k},\underline{i}} - u_\ell^{\underline{k},\underline{i}}||| \le |||u_{\ell+1}^{\star} - u_{\ell+1}^{\underline{k},\underline{i}}||| + |||u_{\ell+1}^{\star} - u_\ell^{\underline{k},\underline{i}}|||$$
$$\overset{(6)}{\le} \Big(\frac{2}{\alpha}\Big)^{1/2} \big[\mathcal{E}(u_{\ell+1}^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + \Big(\frac{2}{\alpha}\Big)^{1/2} \big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2}$$
$$\overset{(22)}{\le} (1 + q_{\mathcal{E}}^{\underline{k}[\ell+1]}) \Big(\frac{2}{\alpha}\Big)^{1/2} \big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2}.$$

Combined with the estimator reduction estimate (47) and with $1 + q_{\mathcal{E}} < 2$, we obtain with $C_1 := 2\,(2/\alpha)^{1/2}\, C_{\mathrm{stab}}[4M]$ that

$$\eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{i}}) \le q_\theta\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}) + C_1 \big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} \quad \text{for all } 0 \le \ell < \underline{\ell}. \tag{48}$$

Step 3 (tail summability with respect to ℓ). Since $1 \le \underline{k}[\ell+1]$, nested iteration $u_{\ell+1}^{0,\underline{i}} = u_\ell^{\underline{k},\underline{i}}$ proves that

$$\mathrm{H}_{\ell+1} \overset{(45)}{=} \big[\mathcal{E}(u_{\ell+1}^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + \gamma\, \eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{i}}) \tag{49}$$
$$\overset{(22)}{\le} q_{\mathcal{E}} \big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + \gamma\, \eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{i}})$$
$$\overset{(48)}{\le} (q_{\mathcal{E}} + C_1\gamma) \big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + q_\theta\, \gamma\, \eta_\ell(u_\ell^{\underline{k},\underline{i}})$$
$$\le \max\{q_{\mathcal{E}} + C_1\gamma,\, q_\theta\} \big(\big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + \gamma\, \eta_\ell(u_\ell^{\underline{k},\underline{i}})\big) \text{ for all } (\ell+1, \underline{k}, \underline{i}) \in \mathcal{Q}. \tag{50}$$

With $0 < q_\theta < 1$, we choose $0 < \gamma < (1 - q_{\mathcal{E}})/C_1 < 1$ to guarantee that

$$0 < \widetilde{q} := \max\{q_{\mathcal{E}} + C_1\gamma,\, q_\theta\} < 1. \tag{51}$$

With the triangle inequality, (49) leads us to

$$
\begin{aligned}
a_{\ell+1} &:= \big[\mathcal{E}(u_{\ell+1}^{k,\underline{i}}) - \mathcal{E}(u_{\ell+1}^\star)\big]^{1/2} + \gamma\,\eta_{\ell+1}(u_{\ell+1}^{k,\underline{i}}) \\
&\overset{(49)}{\le} \widetilde{q}\,\big(\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \gamma\,\eta_\ell(u_\ell^{k,\underline{i}})\big) + \widetilde{q}\,\big[\mathcal{E}(u_\ell^\star) - \mathcal{E}(u_{\ell+1}^\star)\big]^{1/2} \\
&=: \widetilde{q}\,a_\ell + b_\ell \quad \text{for all } (\ell,\underline{k},\underline{i}) \in \mathcal{Q}.
\end{aligned}
\tag{52}
$$

By exploiting the equivalence (6) and stability (A1) (since all $u_\ell^{k,\underline{i}}$ are uniformly bounded by nested iteration (20)), the Céa lemma (5), and reliability (A3) prove that

$$
\begin{aligned}
&\big[\mathcal{E}(u_{\ell'}^\star) - \mathcal{E}(u_{\ell''}^\star)\big]^{1/2} \overset{(6)}{\simeq} |||u_{\ell''}^\star - u_{\ell'}^\star||| \overset{(5)}{\lesssim} |||u^\star - u_\ell^\star||| \overset{(A3)}{\lesssim} \eta_\ell(u_\ell^\star) \\
&\overset{(A1)}{\lesssim} |||u_\ell^\star - u_\ell^{k,\underline{i}}||| + \eta_\ell(u_\ell^{k,\underline{i}}) \\
&\overset{(6)}{\simeq} \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \eta_\ell(u_\ell^{k,\underline{i}}) \simeq a_\ell \text{ for all } \ell \le \ell' \le \ell'' \le \underline{\ell} \text{ with} (\ell,\underline{k},\underline{i}) \in \mathcal{Q}.
\end{aligned}
\tag{53}
$$

Hence, we infer that $b_{\ell+N} \lesssim a_\ell$ for all $0 \le \ell \le \ell + N \le \underline{\ell}$ with $(\ell,\underline{k},\underline{i}) \in \mathcal{Q}$, where the hidden stability constant $C_{\mathrm{stab}}[3\,M]$ depends on $3\,M$ due to (4) and nested iteration (20).

The energy $\mathcal{E}$ from (POT) (and its Pythagorean identity that leads to a telescoping sum) as well as the minimization property (7) for $\mathcal{X}_H = \mathcal{X}$ allow for the estimate

$$
\begin{aligned}
\sum_{\ell'=\ell}^{\ell+N-1} b_{\ell'}^2 &\lesssim \sum_{\ell'=\ell}^{\underline{\ell}-1} \big[\mathcal{E}(u_{\ell'}^\star) - \mathcal{E}(u_{\ell'+1}^\star)\big] \le \mathcal{E}(u_\ell^\star) - \mathcal{E}(u_{\underline{\ell}}^\star) \overset{(7)}{\le} \mathcal{E}(u_\ell^\star) - \mathcal{E}(u^\star) \\
&\overset{(6)}{\le} \frac{L[2M]}{2}\,|||u^\star - u_\ell^\star|||^2 \overset{(A3)}{\le} C_{\mathrm{rel}}^2\,\frac{L[2M]}{2}\,\eta_\ell(u_\ell^\star)^2 \overset{(53)}{\lesssim} a_\ell^2 \text{ for all } 0 \le \ell < \ell + N \le \underline{\ell},
\end{aligned}
\tag{54}
$$

where the hidden stability constant C_{stab} depends on $3\,M$ due to (4) and nested iteration (20).

With (52)–(54), the assumptions for the tail summability criterion from [21, Lemma 6] are met. We thus conclude tail summability of $\mathrm{H}_{\ell+1} \simeq \mathrm{H}_\ell^{\underline{k}} \simeq a_\ell$, i.e.,

$$
\boxed{\sum_{\ell'=\ell+1}^{\underline{\ell}-1} \mathrm{H}_{\ell'}^{\underline{k}} \lesssim \mathrm{H}_\ell^{\underline{k}} \quad \text{for all } 0 \le \ell < \underline{\ell}.}
\tag{55}
$$

Step 4 (quasi-contraction in k). We distinguish three cases.
Case 4.1: Evaluation of (16) yields `TRUE` ∧ `FALSE`. This gives rise to

$$
2M \overset{(16)}{<} |||u_\ell^{k,\underline{i}}||| \le |||u_\ell^\star||| + |||u_\ell^\star - u_\ell^{k,\underline{i}}||| \overset{(4)}{\le} M + |||u_\ell^\star - u_\ell^{k,\underline{i}}|||
$$

and hence, we conclude that $M < |||u_\ell^\star - u_\ell^{k,\underline{i}}|||$. Thus,

$$1 = \frac{M}{M} < \frac{|||u_\ell^\star - u_\ell^{k,\underline{i}}|||}{M} \overset{(6)}{\leq} \frac{1}{M}\Big(\frac{2}{\alpha}\Big)^{1/2}\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(22)}{\leq} \frac{q_{\mathcal{E}}}{M}\Big(\frac{2}{\alpha}\Big)^{1/2}\big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2}. \tag{56}$$

We recall from (4) that $|||u_\ell^\star||| \leq M$ and $|||u_\ell^\star - u_0^\star||| \leq 2\,M$ independently of ℓ. Moreover, there holds quasi-monotonicity of the estimators in the sense that

$$\eta_\ell(u_\ell^\star) \leq C_{\mathrm{mon}}\,\eta_0(u_0^\star) \quad \text{with } C_{\mathrm{mon}} = \big[2 + 8\,C_{\mathrm{stab}}[2M]^2(1 + C_{\mathrm{Céa}}[2M]^2)\,C_{\mathrm{rel}}^2\big]^{1/2}; \tag{57}$$

cf. [15, Lemma 3.6] or [36, Equation (42)] for the locally Lipschitz continuous setting. In particular, estimate (57) holds also for the discrete limit space $\mathcal{X}_{\underline{\ell}} := \mathrm{closure}\big(\bigcup_{\ell=0}^{\underline{\ell}} \mathcal{X}_\ell\big)$. Additionally, we note that the estimate (57) admits

$$\eta_\ell(u_\ell^\star) \overset{(57)}{\leq} C_{\mathrm{mon}}\,\eta_0(u_0^\star) \overset{(A1)}{\leq} C_{\mathrm{mon}}\,\eta_0(0) + C_{\mathrm{mon}}\,C_{\mathrm{stab}}[M]\,|||u_0^\star||| \overset{(56)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2}. \tag{58}$$

The estimate (58), stability ((A1))with stability constant $C_{\mathrm{stab}}[2\tau]$ due to (21) and (4), and energy contraction (22) yield that

$$\eta_\ell(u_\ell^{k,\underline{i}}) \overset{(A1)}{\leq} \eta_\ell(u_\ell^\star) + C_{\mathrm{stab}}[2\tau]\,|||u_\ell^\star - u_\ell^{k,\underline{i}}||| \overset{(6)}{\lesssim} \eta_\ell(u_\ell^\star) + \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(58)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(22)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2}. \tag{59}$$

For $0 \leq k' < k < \underline{k}[\ell]$, the definition (44), energy contraction (22), and (59) prove

$$\mathrm{H}_\ell^k \overset{(22)}{\lesssim} q_{\mathcal{E}}\big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \eta_\ell(u_\ell^{k,\underline{i}}) \overset{(59)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(22)}{\lesssim} q_{\mathcal{E}}^{(k-1)-k'}\big[\mathcal{E}(u_\ell^{k',\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(44)}{\lesssim} q_{\mathcal{E}}^{k-k'}\,\mathrm{H}_\ell^{k'}. \tag{60}$$

This concludes Case 4.1. ◇

Case 4.2: Evaluation of (16) **yields `FALSE` ∧ `FALSE`** or **`FALSE` ∧ `TRUE`**. For $0 \leq k' < k < \underline{k}[\ell]$, the definition (44), the failure of the accuracy condition in the stopping criterion for the inexact Zarantonello linearization (16), energy minimization (7), and energy contraction (22) prove that

$$\mathrm{H}_\ell^k \overset{(16)}{<} \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \lambda_{\mathrm{lin}}^{-1}\big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^{k,\underline{i}})\big]^{1/2} \overset{(7),\,(22)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(22)}{\lesssim} q_{\mathcal{E}}^{(k-1)-k'}\big[\mathcal{E}(u_\ell^{k',\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(44)}{\lesssim} q_{\mathcal{E}}^{k-k'}\,\mathrm{H}_\ell^{k'}. \tag{61}$$

This concludes Case 4.2. ⋄

Case 4.3: Evaluation of (16) yields TRUE ∧ TRUE. The equivalence (25), boundedness (21), and energy minimization (7) prove that

$$\begin{aligned}\mathrm{H}_\ell^{k} &\overset{(A1)}{\lesssim} \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + |||u_\ell^{k,\underline{i}} - u_\ell^{k-1,\underline{i}}||| + \eta_\ell(u_\ell^{k-1,\underline{i}}) \\ &\overset{(25)}{\lesssim} \mathrm{H}_\ell^{k-1} + \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^{k,\underline{i}})\big]^{1/2} \overset{(7)}{\leq} 2\,\mathrm{H}_\ell^{k-1} \quad \text{for all } (\ell, \underline{k}, \underline{i}) \in \mathcal{Q}.\end{aligned} \tag{62}$$

Since $k = \underline{k}[\ell] - 1$ is covered by Case 4.1 or Case 4.2, estimate (62) leads to

$$\mathrm{H}_\ell^{k} \overset{(62)}{\lesssim} \frac{q_\mathcal{E}}{q_\mathcal{E}}\,\mathrm{H}_\ell^{k-1} \lesssim q_\mathcal{E}\,\mathrm{H}_\ell^{k-1} \overset{(60),\,(61)}{\lesssim} q_\mathcal{E}^{\underline{k}[\ell]-k'}\,\mathrm{H}_\ell^{k',\underline{i}}. \tag{63}$$

This concludes Case 4.3. ⋄

Overall, the estimates (60), (61) and (63) result in

$$\boxed{\mathrm{H}_\ell^{k} \lesssim q_\mathcal{E}^{k-k'}\mathrm{H}_\ell^{k'} \quad \text{for all } (\ell, k, j) \in \mathcal{Q} \text{ with } 0 \leq k' \leq k \leq \underline{k}[\ell],} \tag{64}$$

where the hidden constant depends only on M, $C_{\mathrm{stab}}[2\tau]$, α, $L[2M]$, $C_{\mathrm{Céa}}[2M]$, C_{rel}, λ_{lin}, and $q_\mathcal{E}$. Furthermore, we recall from (53) that $\big[\mathcal{E}(u_{\ell-1}^\star) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \lesssim \mathrm{H}_{\ell-1}^{\underline{k}}$. Together with nested iteration $u_{\ell-1}^{\underline{k},\underline{i}} = u_\ell^{0,\underline{i}} = u_\ell^{0,\star}$, this yields that

$$\mathrm{H}_\ell^{0} = \big[\mathcal{E}(u_{\ell-1}^{\underline{k},\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \eta_\ell(u_{\ell-1}^{\underline{k},\underline{i}}) \lesssim \big[\mathcal{E}(u_{\ell-1}^\star) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \mathrm{H}_{\ell-1}^{\underline{k}} \leq \mathrm{H}_{\ell-1}^{\underline{k}}$$

and thus

$$\boxed{\mathrm{H}_\ell^{0} \lesssim \mathrm{H}_{\ell-1}^{\underline{k}} \quad \text{for all } (\ell, 0, 0) \in \mathcal{Q} \text{ with } \ell \geq 1.} \tag{65}$$

Step 5 (tail summability with respect to ℓ and k). The estimates (64)–(65) from Step 4 as well as (55) from Step 3 and the geometric series prove that

$$\begin{aligned}\sum_{\substack{(\ell',k',\underline{i})\in\mathcal{Q}\\ |\ell',k',\underline{i}|>|\ell,k,\underline{i}|}} \mathrm{H}_{\ell'}^{k'} &= \sum_{k'=k+1}^{\underline{k}[\ell]} \mathrm{H}_\ell^{k'} + \sum_{\ell'=\ell+1}^{\underline{\ell}} \sum_{k'=0}^{\underline{k}[\ell']} \mathrm{H}_{\ell'}^{k'} \overset{(64)}{\lesssim} \mathrm{H}_\ell^{k} + \sum_{\ell'=\ell+1}^{\underline{\ell}} \mathrm{H}_{\ell'}^{0} \\ &\overset{(65)}{\lesssim} \mathrm{H}_\ell^{k} + \sum_{\ell'=\ell}^{\underline{\ell}-1} \mathrm{H}_{\ell'}^{\underline{k}} \overset{(55)}{\lesssim} \mathrm{H}_\ell^{k} + \mathrm{H}_\ell^{\underline{k}} \overset{(64)}{\lesssim} \mathrm{H}_\ell^{k} \quad \text{for all } (\ell, k, \underline{i}) \in \mathcal{Q}.\end{aligned} \tag{66}$$

Step 6 (contraction in i). For $i = 0$ and $k = 0$, we recall that $u_\ell^{0,0} = u_\ell^{0,\underline{i}} = u_\ell^{0,\star}$ by definition and hence $\mathrm{H}_\ell^{0,0} \overset{(6)}{\simeq} \mathrm{H}_\ell^{0}$. For $k \geq 1$, nested iteration $u_\ell^{k,0} = u_\ell^{k-1,\underline{i}}$, contraction of the exact Zarantonello iteration (10), and energy equivalence (6) imply that

$$\begin{aligned} |||u_\ell^{k,\star} - u_\ell^{k,0}||| &\le |||u_\ell^\star - u_\ell^{k,\star}||| + |||u_\ell^\star - u_\ell^{k-1,\underline{i}}||| \\ &\overset{(10)}{\le} (q_{\mathrm{Zar}}^\star[\delta; 3M] + 1)\, |||u_\ell^\star - u_\ell^{k-1,\underline{i}}||| \overset{(6)}{\lesssim} 2\,\mathrm{H}_\ell^{k-1}. \end{aligned}$$

Therefore, by using the equivalence (6) once more, we obtain that

$$\boxed{\mathrm{H}_\ell^{k,0} \lesssim \mathrm{H}_\ell^{(k-1)_+} \quad \text{for all } (\ell, k, 0) \in \mathcal{Q}, \quad \text{where } (k-1)_+ := \max\{0, k-1\}.} \tag{67}$$

Let $(\ell, k, i) \in \mathcal{Q}$. It holds that

$$\begin{aligned} \mathrm{H}_\ell^{k,i} &\overset{(42)}{=} |||u_\ell^\star - u_\ell^{k,i}||| + |||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i}) \\ &\overset{(A1)}{\le} \mathrm{H}_\ell^{k,i-1} + (2 + C_{\mathrm{stab}}[10\tau])\, |||u_\ell^{k,i} - u_\ell^{k,i-1}||| \\ &\overset{(11)}{\le} \mathrm{H}_\ell^{k,i-1} + (2 + C_{\mathrm{stab}}[10\tau])(q_{\mathrm{alg}} + 1)\, |||u_\ell^{k,\star} - u_\ell^{k,i-1}||| \overset{(42)}{\lesssim} \mathrm{H}_\ell^{k,i-1}, \end{aligned} \tag{68}$$

where $C_{\mathrm{stab}}[10\tau]$ stems from the uniform bound (23) from Theorem 4. Hence, we obtain

$$\mathrm{H}_\ell^{k,i} \lesssim \mathrm{H}_\ell^{k,i'} \simeq q_{\mathrm{alg}}^{i-i'}\, \mathrm{H}_\ell^{k,i'} \quad \text{for all } (\ell, k, i) \in \mathcal{Q} \text{ with } 0 \le i' \le i \le i_{\min}.$$

For all $0 \le i' < i_{\min} \le i < \underline{i}[\ell, k]$, we obtain with an a posteriori estimate based on the contraction of the Zarantonello iteration (10) (where $q_{\mathrm{Zar}}^\star = q_{\mathrm{Zar}}^\star[\delta, 2\tau]$ depends on τ from (21)), the a posteriori estimate (41) for the algebraic solver, the failure of the accuracy criterion of (15), and the contraction of the algebraic solver (11) that

$$\begin{aligned} \mathrm{H}_\ell^{k,i} &\overset{(42)}{=} |||u_\ell^\star - u_\ell^{k,i}||| + |||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i}) \\ &\le |||u_\ell^\star - u_\ell^{k,\star}||| + 2\, |||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i}) \\ &\le \frac{q_{\mathrm{Zar}}^\star[\delta; 2\tau]}{1 - q_{\mathrm{Zar}}^\star[\delta; 2\tau]}\, |||u_\ell^{k,i} - u_\ell^{k-1,\underline{i}}||| \\ &\quad + \Big(2 + \frac{q_{\mathrm{Zar}}^\star[\delta; 2\tau]}{1 - q_{\mathrm{Zar}}^\star[\delta; 2\tau]}\Big)\, |||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i}) \\ &\overset{(41)}{\lesssim} |||u_\ell^{k,i} - u_\ell^{k-1,\underline{i}}||| + |||u_\ell^{k,i} - u_\ell^{k,i-1}||| + \eta_\ell(u_\ell^{k,i}) \\ &\overset{(15)}{\lesssim} |||u_\ell^{k,i} - u_\ell^{k,i-1}||| \overset{(11)}{\lesssim} |||u_\ell^{k,\star} - u_\ell^{k,i-1}||| \\ &\overset{(11)}{\lesssim} q_{\mathrm{alg}}^{i-i'}\, |||u_\ell^{k,\star} - u_\ell^{k,i'}||| \le q_{\mathrm{alg}}^{i-i'}\, \mathrm{H}_\ell^{k,i'}, \end{aligned} \tag{69}$$

Altogether, the combination of (68) and (69) proves that

$$\boxed{\mathrm{H}_\ell^{k,i} \lesssim q_{\mathrm{alg}}^{i-i'}\, \mathrm{H}_\ell^{k,i'} \quad \text{for all } (\ell,k,i) \in \mathcal{Q} \quad \text{with} \quad 0 \le i' \le i \le \underline{i}[\ell,k],} \tag{70}$$

where the hidden constant depends only on $q_{\mathrm{Zar}}^\star[\delta; 2\tau]$, q_{alg}, λ_{alg}, $C_{\mathrm{stab}}[10\tau]$, and $i_{\min}$.
Step 7 (tail summability with respect to ℓ, k, and i). Finally, we observe that

$$\begin{aligned}
\sum_{\substack{(\ell',k',i')\in\mathcal{Q}\\ |\ell',k',i'|>|\ell,k,i|}} \mathrm{H}_{\ell'}^{k',i'} &= \sum_{i'=i+1}^{\underline{i}[\ell,k]} \mathrm{H}_\ell^{k,i'} + \sum_{k'=k+1}^{\underline{k}[\ell]} \sum_{i'=0}^{\underline{i}[\ell,k']} \mathrm{H}_\ell^{k',i'} + \sum_{\ell'=\ell+1}^{\underline{\ell}} \sum_{k'=0}^{\underline{k}[\ell']} \sum_{i'=0}^{\underline{i}[\ell',k']} \mathrm{H}_{\ell'}^{k',i'} \\
&\overset{(70)}{\lesssim} \mathrm{H}_\ell^{k,i} + \sum_{k'=k+1}^{\underline{k}[\ell]} \mathrm{H}_\ell^{k',0} + \sum_{\ell'=\ell+1}^{\underline{\ell}} \sum_{k'=0}^{\underline{k}[\ell']} \mathrm{H}_{\ell'}^{k',0} \overset{(67)}{\lesssim} \mathrm{H}_\ell^{k,i} + \sum_{\substack{(\ell',k',\underline{i})\in\mathcal{Q}\\ |\ell',k',\underline{i}|>|\ell,k,\underline{i}|}} \mathrm{H}_{\ell'}^{k'} \\
&\overset{(66)}{\lesssim} \mathrm{H}_\ell^{k,i} + \mathrm{H}_\ell^{k} \overset{(46)}{\lesssim} \mathrm{H}_\ell^{k,i} + \mathrm{H}_\ell^{k,\underline{i}} \overset{(70)}{\lesssim} \mathrm{H}_\ell^{k,i} \quad \text{for all } (\ell,k,i) \in \mathcal{Q}.
\end{aligned}$$

Since $\mathcal{Q}$ is countable and linearly ordered, [15, Lemma 4.9] applies and proves R-linear convergence (43) of $\mathrm{H}_\ell^{k,i}$. This concludes the proof. □

Given full R-linear convergence from Theorem 5, then convergence rates with respect to the degrees of freedom coincide with rates with respect to the overall computational cost, where we recall $\mathtt{cost}(\ell,k,i)$ from (18). Since all essential arguments are provided, the proof follows verbatim from [21, Corollary 16].

Corollary 1 *(rates $\widehat{=}$ complexity) Suppose full R-linear convergence* (43)*. Recall* $\mathtt{cost}(\ell,k,i)$ *from* (18)*. Then, for any $s > 0$, it holds that*

$$M(s) := \sup_{(\ell,k,i)\in\mathcal{Q}} (\#\mathcal{T}_\ell)^s\, \mathrm{H}_\ell^{k,i} \le \sup_{(\ell,k,i)\in\mathcal{Q}} \mathtt{cost}(\ell,k,i)^s\, \mathrm{H}_\ell^{k,i} \le C_{\mathrm{cost}}\, M(s), \tag{71}$$

where the constant $C_{\mathrm{cost}} > 0$ depends only on C_{lin}, q_{lin}, and s. Moreover, there exists $s_0 > 0$ such that $M(s) < \infty$ for all $0 < s \le s_0$. □

5 Optimal complexity

A formal approach to optimal complexity relies on the notion of approximation classes [4–6, 15], which reads as follows: For $s > 0$, define

$$\|u^\star\|_{\mathbb{A}_s} := \sup_{N\in\mathbb{N}_0} \Big[(N+1)^s \min_{\mathcal{T}_{\mathrm{opt}}\in\mathbb{T}_N} \eta_{\mathrm{opt}}(u_{\mathrm{opt}}^\star)\Big],$$

where $u_{\mathrm{opt}}^\star$ denotes the exact discrete solution associated with the optimal triangulation $\mathcal{T}_{\mathrm{opt}} \in \mathbb{T}_N(\mathcal{T})$. For $s > 0$, we note that $\|u^\star\|_{\mathbb{A}_s} < \infty$ means that the sequence of estimators along optimally chosen meshes decreases at least as fast as $(N+1)^{-s} \simeq N^{-s}$.

Finally, we are in the position to present the third main result of this paper, namely optimal complexity of Algorithm 1. Its proof relies, in essence, on perturbation arguments. More precisely, sufficiently small θ and λ_{lin} are required to ensure that Algorithm 1 guarantees convergence rate s with respect to the overall computational cost (and time) if the solution $u^\star$ of (2) can be approximated at rate s in the sense of $\|u^\star\|_{\mathbb{A}_s} < \infty$.

Theorem 6 (Optimal complexity) *Define* $\tau := M + 3\,M\big(\frac{L[3\,M]}{\alpha}\big)^{1/2} \geq 4\,M$ *with* M *from* (4). *Let* $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$ *to ensure validity of Theorem 4. Define*

$$\lambda_{\mathrm{lin}}^\star := \min\Big\{1, \Big(\frac{\alpha\,(1-q_{\mathcal{E}}^2)}{2\,q_{\mathcal{E}}^2}\Big)^{1/2}/C_{\mathrm{stab}}[3M]\Big\}. \tag{72}$$

Suppose the axioms (A1)–(A4). Let $0 < \theta < 1$, $0 < \lambda_{\mathrm{alg}}$, *and* $0 < \lambda_{\mathrm{lin}} < \lambda_{\mathrm{lin}}^\star$ *such that*

$$0 < \theta_{\mathrm{mark}} := \frac{(\theta^{1/2} + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)^2}{(1-\lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)^2} < \theta^\star := (1 + C_{\mathrm{stab}}[2M]^2\,C_{\mathrm{rel}}^2)^{-1} < 1. \tag{73}$$

Then, Algorithm 1 guarantees, for all $s > 0$, *that*

$$\sup_{(\ell,k,j)\in\mathcal{Q}} \texttt{cost}(\ell,k,i)^s\,\mathrm{H}_\ell^{k,j} \leq C_{\mathrm{opt}}\,\max\{\|u^\star\|_{\mathbb{A}_s}, \mathrm{H}_0^{0,0}\}. \tag{74}$$

The constant $C_{\mathrm{opt}} > 0$ *depends only on* $q_{\mathcal{E}}$, α, $C_{\mathrm{stab}}[10\tau]$, C_{rel}, C_{drel}, C_{mark}, C_{mesh}, C_{lin}, q_{lin}, $\#\mathcal{T}_0$, *and* s. *In particular, there holds optimal complexity of Algorithm 1.*

To prove the theorem, we require the following results on the estimator, which relies on sufficiently small adaptivity parameter $\lambda_{\mathrm{lin}} > 0$.

Lemma 5 (Estimator equivalence) *Suppose the assumptions of Theorem 4. Recall* $\lambda_{\mathrm{lin}}^\star$ *from* (72). *Then, for all* $(\ell, \underline{k}, \underline{i}) \in \mathcal{Q}$ *with* $\underline{k}[\ell] < \infty$, *it holds that*

$$\eta_\ell(u_\ell^\star) \leq (1 + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)\,\eta_\ell(u_\ell^{\underline{k},\underline{i}}), \tag{75a}$$

and, for $0 < \lambda_{\mathrm{lin}} < \lambda_{\mathrm{lin}}^\star$, *we furthermore have that*

$$(1 - \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)\eta_\ell(u_\ell^{\underline{k},\underline{i}}) \leq \eta_\ell(u_\ell^\star). \tag{75b}$$

For $0 < \lambda_{\mathrm{lin}} < \lambda_{\mathrm{lin}}^\star$, *Dörfler marking for* $u_\ell^\star$ *with parameter* θ_{mark} *from* (73) *implies Dörfler marking for* $u_\ell^{\underline{k},\underline{i}}$ *with parameter* θ, *i.e., for any* $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$, *there holds the implication*

$$\theta_{\mathrm{mark}}\,\eta_\ell(u_\ell^\star)^2 \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^\star)^2 \quad\Longrightarrow\quad \theta\,\eta_\ell(u_\ell^{\underline{k},\underline{i}})^2 \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^{\underline{k},\underline{i}})^2. \tag{76}$$

Proof The proof consists of two steps.

Step 1. First, we obtain from Remark 3(ii) that

$$\frac{\alpha}{2}\,|||u_\ell^\star - u_\ell^{k,\underline{i}}|||^2 \overset{(40)}{\le} \frac{\lambda_{\mathrm{lin}}^2\, q_{\mathcal{E}}^2}{1-q_{\mathcal{E}}^2}\,\eta_\ell(u_\ell^{k,\underline{i}})^2.$$

Exploiting this together with stability (A1), nested iteration (22), and boundedness of the exact discrete solution (4), we obtain for any $\mathcal{U}_\ell \subseteq \mathcal{T}_\ell$ that

$$\begin{aligned}\eta_\ell(\mathcal{U}_\ell; u_\ell^\star) &\overset{(A1)}{\le} \eta_\ell(\mathcal{U}_\ell; u_\ell^{k,\underline{i}}) + C_{\mathrm{stab}}[3M]\,|||u_\ell^\star - u_\ell^{k,\underline{i}}||| \\ &\le \eta_\ell(\mathcal{U}_\ell; u_\ell^{k,\underline{i}}) + \lambda_{\mathrm{lin}}\, C_{\mathrm{stab}}[3M]\Big(\frac{2\,q_{\mathcal{E}}^2}{\alpha\,(1-q_{\mathcal{E}}^2)}\Big)^{1/2}\eta_\ell(u_\ell^{k,\underline{i}}) \\ &= \eta_\ell(\mathcal{U}_\ell; u_\ell^{k,\underline{i}}) + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star\,\eta_\ell(u_\ell^{k,\underline{i}}).\end{aligned} \tag{77}$$

The choice $\mathcal{U}_\ell = \mathcal{T}_\ell$ yields (75a). The same arguments prove that

$$\eta_\ell(\mathcal{U}_\ell; u_\ell^{k,\underline{i}}) \le \eta_\ell(\mathcal{U}_\ell; u_\ell^\star) + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star\,\eta_\ell(u_\ell^{k,\underline{i}}). \tag{78}$$

For $0 < \lambda_{\mathrm{lin}} < \lambda_{\mathrm{lin}}^\star$ and $\mathcal{U}_\ell = \mathcal{T}_\ell$, the rearrangement of (78) proves (75b).
Step 2. Let $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$ satisfy $\theta_{\mathrm{mark}}^{1/2}\,\eta_\ell(u_\ell^\star) \le \eta_\ell(\mathcal{R}_\ell; u_\ell^\star)$. Then, (77)–(78) prove

$$\begin{aligned}\big[1-\lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star\big]\,\theta_{\mathrm{mark}}^{1/2}\,\eta_\ell(u_\ell^{k,\underline{i}}) &\overset{(75b)}{\le} \theta_{\mathrm{mark}}^{1/2}\,\eta_\ell(u_\ell^\star) \le \eta_\ell(\mathcal{R}_\ell; u_\ell^\star) \\ &\overset{(77)}{\le} \eta_\ell(\mathcal{R}_\ell; u_\ell^{k,\underline{i}}) + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star\,\eta_\ell(u_\ell^{k,\underline{i}}) \\ &\overset{(73)}{=} \eta_\ell(\mathcal{R}_\ell; u_\ell^{k,\underline{i}}) + \big[\theta_{\mathrm{mark}}^{1/2}\,(1-\lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star) - \theta^{1/2}\big]\,\eta_\ell(u_\ell^{k,\underline{i}}).\end{aligned}$$

This yields $\theta^{1/2}\,\eta_\ell(u_\ell^{k,\underline{i}}) \le \eta_\ell(\mathcal{R}_\ell; u_\ell^{k,\underline{i}})$ and concludes the proof. □

Proof of Theorem 6 By Corollary 1, it is enough to show

$$\sup_{(\ell,k,i)\in\mathcal{Q}} \big(\#\mathcal{T}_\ell\big)^s\, \mathrm{H}_\ell^{k,i} \lesssim \max\{\|u^\star\|_{\mathbb{A}_s}, \mathrm{H}_0^{0,0}\}. \tag{79}$$

Without loss of generality, we may suppose that $\|u^\star\|_{\mathbb{A}_s} < \infty$. The proof is subdivided into two steps.

Step 1. Let $0 < \theta_{\mathrm{mark}} := (\theta^{1/2} + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)^2\,(1-\lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)^{-2} < \theta^\star := (1 + C_{\mathrm{stab}}[2\,M]^2\,C_{\mathrm{rel}}^2)^{-1}$ and fix any $0 \le \ell' \le \underline{\ell} - 1$. The validity of (A4) and [15, Lemma 4.14] guarantee the existence of a set $\mathcal{R}_{\ell'} \subseteq \mathcal{T}_{\ell'}$ such that

$$\begin{aligned}\#\mathcal{R}_{\ell'} &\lesssim \|u^\star\|_{\mathbb{A}_s}^{1/s}\,[\eta_{\ell'}(u_{\ell'}^\star)]^{-1/s}, \\ \theta_{\mathrm{mark}}\,\eta_{\ell'}(u_{\ell'}^\star) &\le \eta_{\ell'}(\mathcal{R}_{\ell'}, u_{\ell'}^\star),\end{aligned} \tag{80}$$

where the hidden constant depends only on (A1)–(A4). By means of (76) in Lemma 5, we infer that $\mathcal{R}_{\ell'}$ satisfies the Dörfler marking (17) in Algorithm 1 with θ, i.e., $\theta\,\eta_{\ell'}(u_{\ell'}^{k,\underline{i}})^2 \le \eta_{\ell'}(\mathcal{R}_{\ell'}; u_{\ell'}^{k,\underline{i}})^2$. Hence, since $0 < \theta < \theta_{\mathrm{mark}} < \theta^\star$, the optimality of Dörfler marking proves

$$\#\mathcal{M}_{\ell'} \le C_{\mathrm{mark}}\,\#\mathcal{R}_{\ell'} \overset{(80)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\big[\eta_{\ell'}(u_{\ell'}^\star)\big]^{-1/s}. \tag{81}$$

Moreover, full R-linear convergence (43) together with a posteriori error estimates for the final iterates (40) and (41), the norm-energy equivalence (25), and estimator equivalence (75a) prove

$$\begin{aligned} \mathrm{H}_{\ell'+1}^{0,\underline{i}} \overset{(43)}{\lesssim} \mathrm{H}_{\ell'}^{k,\underline{i}} &\overset{(42)}{=} |||u_{\ell'}^\star - u_{\ell'}^{k,\underline{i}}||| + |||u_{\ell'}^{k,\star} - u_{\ell'}^{k,\underline{i}}||| + \eta_{\ell'}(u_{\ell'}^{k,\underline{i}}) \\ &\overset{(41),\,(25)}{\lesssim} |||u_{\ell'}^\star - u_{\ell'}^{k,\underline{i}}||| + [\mathcal{E}(u_{\ell'}^{k,0}) - \mathcal{E}(u_{\ell'}^{k,\underline{i}})]^{1/2} + \eta_{\ell'}(u_{\ell'}^{k,\underline{i}}) \\ &\overset{(40),\,(16)}{\lesssim} \eta_{\ell'}(u_{\ell'}^{k,\underline{i}}) \overset{(75a)}{\lesssim} \eta_{\ell'}(u_{\ell'}^\star). \end{aligned} \tag{82}$$

Consequently, a combination of (81) and (82) concludes that

$$\#\mathcal{M}_{\ell'} \overset{(81)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\big[\eta_{\ell'}(u_{\ell'}^\star)\big]^{-1/s} \overset{(82)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\big[\mathrm{H}_{\ell'+1}^{0,\underline{i}}\big]^{-1/s}. \tag{83}$$

Step 2. For $(\ell,k,i) \in \mathcal{Q}$, full R-linear convergence (43) and the geometric series prove

$$\sum_{\substack{(\ell',k',i')\in\mathcal{Q}\\ |\ell',k',i'|\le|\ell,k,i|}} (\mathrm{H}_{\ell'}^{k',i'})^{-1/s} \overset{(43)}{\lesssim} (\mathrm{H}_{\ell}^{k,i})^{-1/s} \sum_{\substack{(\ell',k',i')\in\mathcal{Q}\\ |\ell',k',i'|\le|\ell,k,i|}} (q_{\mathrm{lin}}^{1/s})^{|\ell,k,i|-|\ell',k',i'|} \lesssim (\mathrm{H}_{\ell}^{k,i})^{-1/s}. \tag{84}$$

We recall the mesh-closure estimate [4, 39–41]

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 \le C_{\mathrm{mesh}} \sum_{\ell'=0}^{\ell-1} \#\mathcal{M}_{\ell'} \quad \text{for all } \ell \ge 0, \tag{85}$$

where $C_{\mathrm{mesh}} > 1$ depends only on $\mathcal{T}_0$ and hence in particular on the dimension d. For $(\ell, k, i) \in \mathcal{Q}$, the preceding estimates show that

$$
\begin{aligned}
\#\mathcal{T}_\ell - \#\mathcal{T}_0 &\overset{(85)}{\lesssim} \sum_{\ell'=0}^{\ell-1} \#\mathcal{M}_{\ell'} \\
&\overset{(83)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s} \sum_{\ell'=0}^{\ell-1} \big(\mathrm{H}_{\ell'+1}^{0,\underline{i}}\big)^{-1/s} \\
&\le \|u^\star\|_{\mathbb{A}_s}^{1/s} \sum_{\substack{(\ell',k',i') \in \mathcal{Q} \\ |\ell',k',i'| \le |\ell,k,i|}} (\mathrm{H}_{\ell'}^{k',i'})^{-1/s} \overset{(84)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s} \, (\mathrm{H}_{\ell}^{k,i})^{-1/s}.
\end{aligned}
$$

Note that $1 \le \#\mathcal{T}_\ell - \#\mathcal{T}_0$ yields $\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 \le 2\,(\#\mathcal{T}_\ell - \#\mathcal{T}_0)$. Hence, we get that

$$
(\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \, \mathrm{H}_\ell^{k,i} \lesssim \|u^\star\|_{\mathbb{A}_s} \quad \text{for all } (\ell, k, i) \in \mathcal{Q} \text{ with } \ell \ge 1.
$$

Theorem 5 proves that

$$
(\#\mathcal{T}_\ell - \#\mathcal{T}_\ell + 1)^s \, \mathrm{H}_\ell^{k,i} = \mathrm{H}_0^{k,i} \overset{(43)}{\lesssim} \mathrm{H}_0^{0,0} \quad \text{for all } (\ell, k, i) \in \mathcal{Q} \text{ with } \ell = 0.
$$

For all $\mathcal{T}_\ell \in \mathbb{T}$, elementary calculation [49, Lemma 22] shows that

$$
\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 \le \#\mathcal{T}_\ell \le \#\mathcal{T}_0\,(\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1). \tag{87}
$$

For all $(\ell, k, i) \in \mathcal{Q}$, we thus arrive at

$$
(\#\mathcal{T}_\ell)^s \, \mathrm{H}_\ell^{k,i} \overset{(87)}{\lesssim} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \, \mathrm{H}_\ell^{k,i} \overset{(86)}{\lesssim} \max\{\|u^\star\|_{\mathbb{A}_s}, \mathrm{H}_0^{0,0}\}.
$$

This concludes the proof of (79). □

6 Numerical experiments

The experiments are performed with the open-source software package MooAFEM [50]. In the following, Algorithm 1 employs the optimal local hp-robust multigrid method [47] as algebraic solver. We remark that in our implementation the condition (19) is slightly relaxed to $|\mathcal{E}(u_\ell^{k,0}) - \mathcal{E}(u_\ell^{k,i})| < 10^{-12} =: \texttt{tol}$.

Experiment 1 (modified sine-Gordon equation [35, Experiment 5.1]) For $\Omega = (0, 1)^2$, we consider

$$
-\Delta u^\star + (u^\star)^3 + \sin(u^\star) = f \quad \text{in } \Omega \quad \text{subject to} \quad u^\star = 0 \text{ on } \partial\Omega \tag{88}
$$

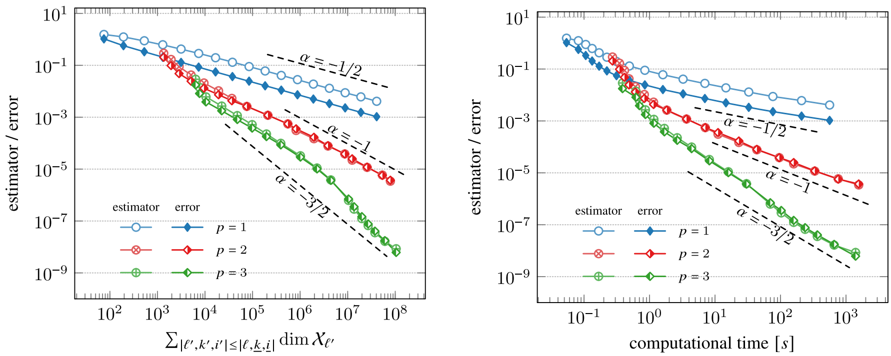


Fig. 2 Experiment 1: Convergence plots of the error $|||u^\star - u_\ell^{\underline{k},\underline{i}}|||$ (diamond) and the error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{i}})$ (circle) over $\texttt{cost}(\ell, \underline{k}, \underline{i})$ (left) and over computational time in seconds (right)

with the monotone semilinearity $b(v) = v^3 + \sin(v)$, which fits into the locally Lipschitz continuous framework (cf. [36, Experiment 26]). We choose f such that

$$u^\star(x) = \sin(\pi x_1)\sin(\pi x_2).$$

For $T \in \mathcal{T}_H$, the refinement indicators $\eta_H(T;\cdot)$ read

$$\eta_H(T, v_H)^2 := h_T^2 \, \|f + \Delta v_H - b(v_H)\|_{L^2(T)}^2 + h_T \, \|[\![\nabla v_H \cdot \boldsymbol{n}]\!]\|_{L^2(\partial T \cap \Omega)}^2. \quad (89)$$

For $p = 2$, damping parameter $\delta = 0.3$, and $i_{\min} = 1$, we stop the computation as soon as $\eta_\ell(u_\ell^{\underline{k},\underline{i}}) < 10^{-4}$. Table 1 depicts the values of the weighted cost

$$\eta_\ell(u_\ell^{\underline{k},\underline{i}}) \, \texttt{cost}(\ell, \underline{k}, \underline{i})^{p/2} \quad (90)$$

to determine the best parameter choice. For a fair comparison, the weighted cost from (90) balances the overachievement of the prescribed tolerance with the associated cumulative computational cost. We observe that the parameters $\theta \in \{0.3, 0.4\}$ and $\lambda_{\mathrm{lin}} \ge 0.5$ perform comparably well. The parameter λ_{alg} may be used for fine-tuning, but for moderate $\theta \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$ and as soon as λ_{lin} is set, the influence is comparably low.

For the following experiments, we set $\delta = 0.3$, $\theta = 0.3$, $\lambda_{\mathrm{lin}} = 0.7$, and $\lambda_{\mathrm{alg}} = 0.3$. Figure 2 depicts the error $|||u^\star - u_\ell^{\underline{k},\underline{i}}|||$ and the estimator $\eta_\ell(u_\ell^{\underline{k},\underline{i}})$ over $\texttt{cost}(\ell, \underline{k}, \underline{i})$ (left) and over the cumulative time in seconds (right) for the displayed polynomial degrees $p \in \{1, 2, 3\}$.

In both plots, the decay rate is of (expected) optimal order $p/2$ for $p \in \{1, 2, 3\}$.

Experiment 2 We consider a globally Lipschitz continuous example from [20, Section 5.3] with Lipschitz constant $L = 2$ and monotonicity constant $\alpha = 1 - 2\exp(-\frac{3}{2})$ and hence $\delta = \alpha/L^2 \approx 0.138434919925785$ is a viable choice. For $d = 2$ and the

Table 1 The weighted cost (90) with $p = 2$ of the sine-Gordon problem (88) for different adaptivity parameters $\lambda_{\text{lin}}, \lambda_{\text{alg}}, \theta \in \{0.1, 0.2, \dots, 0.9\}$ and fixed damping parameter $\delta = 0.3$, where the mesh refinement is stopped if $\eta_\ell(u_\ell^{k,i}) < 10^{-4}$, where the θ-blockwise minimal values are highlighted in green and the overall minimal value in red (with white font)

$\delta = 0.3$	$\theta = 0.1$					$\theta = 0.2$					$\theta = 0.3$				
λ_{alg} \ λ_{lin}	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9	0.1	0.3	0.5	0.7	0.9
0.1	1306	650	660	660	660	735	639	347	347	347	724	659	373	373	373
0.3	928	660	660	660	660	545	269	269	269	269	505	333	241	241	241
0.5	654	654	654	654	654	534	274	274	273	273	462	278	262	262	262
0.7	649	617	617	617	617	293	262	262	262	262	420	298	259	259	259
0.9	676	646	646	646	646	268	269	269	269	269	422	321	247	247	247

	$\theta = 0.4$					$\theta = 0.5$					$\theta = 0.6$				
0.1	807	643	357	357	357	816	658	337	350	350	882	600	332	361	361
0.3	533	375	252	252	252	532	448	266	266	266	663	466	293	293	293
0.5	464	346	253	253	253	572	399	278	278	278	643	389	292	292	292
0.7	487	377	247	247	247	573	427	293	293	293	606	402	296	296	296
0.9	502	390	264	264	264	520	417	288	288	288	563	512	288	288	288

	$\theta = 0.7$					$\theta = 0.8$					$\theta = 0.9$				
0.1	856	634	361	337	337	985	741	413	375	375	1028	710	466	344	344
0.3	663	457	321	321	321	673	471	328	328	328	735	551	349	349	349
0.5	705	446	299	299	299	638	452	340	340	340	700	542	374	374	374
0.7	630	541	338	338	338	752	518	343	343	343	680	586	352	352	352
0.9	639	518	347	347	347	770	579	373	373	373	722	667	367	367	367

L-shaped domain $(-1, 1)^2 \backslash \big([0, 1] \times [-1, 0]\big) \subset \mathbb{R}^2$, we seek $u^\star \in H_0^1(\Omega)$ such that

$$-\operatorname{div}(\mu(|\nabla u^\star|^2)\nabla u^\star) = f \quad \text{in } \Omega,$$

where f is chosen such that $u^\star$ reads in polar coordinates $(r, \varphi) \in \mathbb{R}_{>0} \times [0, 2\pi)$

$$u^\star(r, \varphi) = r^{2/3} \sin\left(\frac{2\varphi}{3}\right) (1 - r\cos\varphi)\,(1 + r\cos\varphi)\,(1 - r\sin\varphi)\,(1 + r\sin\varphi)\,\cos\varphi.$$

This example has a singularity at the origin. We consider $p = 1$, since stability (A1) in the quasilinear case remains open for $p > 1$. Moreover, the parameters are $\theta = 0.3$, $\lambda_{\text{lin}} = 0.7$, $\lambda_{\text{alg}} = 0.3$, and $i_{\min} = 1$.

In Fig. 3, we plot a sample solution (right) as well as convergence results of various error components (left) over the degrees of freedom. We observe that after a preasymptotic phase, optimal convergence rate $-1/2$ is restored for the exact error (diamond),

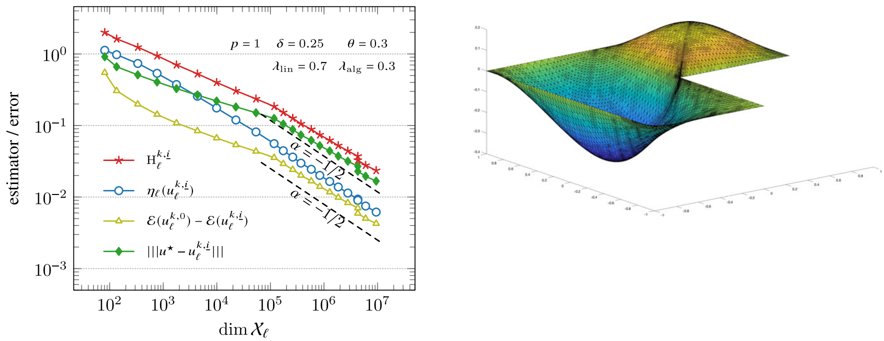


Fig. 3 Experiment 2: Convergence plots of various error components over the degrees of freedom (left). Right: Plot of the approximate solution $u_{13}^{1,1}$ on $\mathcal{X}_{13}$ with $\#\mathcal{X}_{13} = 10209$

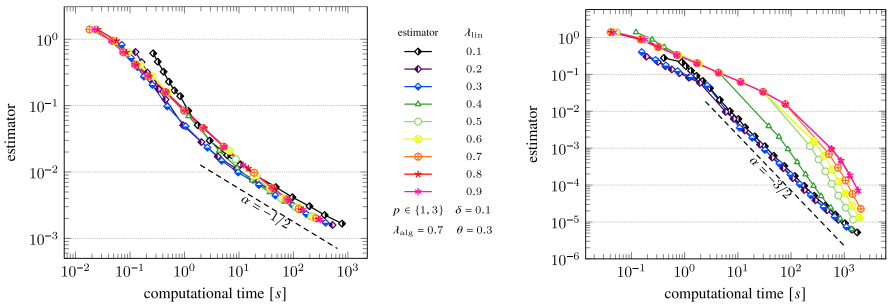


Fig. 4 Convergence plots of the error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{i}})$ over computational time of Experiment 3. Left: Convergence plot for $p = 1$ Right: Convergence plot for $p = 3$

the quasi-error $\mathrm{H}_\ell^{k,\underline{i}}$, the linearization error $\mathcal{E}(u_\ell^{k,0}) - \mathcal{E}(u_\ell^{k,\underline{i}})$ (triangle), and the error estimator $\eta_\ell(u_\ell^{k,\underline{i}})$ (circle).

Experiment 3 (singularly perturbed sine-Gordon equation) This example is a variant of [35, Experiment 5.2]. For $d = 2$ and the L-shaped domain $(-1, 1)^2 \backslash \big([0, 1] \times [-1, 0]\big) \subset \mathbb{R}^2$, let $\varepsilon = 10^{-5}$ and consider

$$-\varepsilon \Delta u^\star + u^\star + (u^\star)^3 + \sin(u^\star) = 1 \quad \text{in } \Omega \quad \text{subject to} \quad u^\star = 0 \text{ on } \partial\Omega,$$

with the monotone semilinearity $b(v) = v^3 + \sin(v)$. In this case, the exact solution $u^\star$ is unknown. We use the energy norm $||| \cdot |||^2 = \varepsilon \, \langle \nabla \cdot \, , \, \nabla \cdot \rangle + \langle \cdot \, , \, \cdot \rangle$. The experiment is conducted with damping parameter $\delta = 0.1$, $\lambda_{\rm alg} = 0.7$, $\theta = 0.3$, and $i_{\min} = 1$. The refinement indicator (89) is modified along the lines of [16, Remark 4.14] to

$$\eta_H(T, v_H)^2 := \hbar_T^2 \, \| f + \varepsilon \Delta v_H - v_H - b(v_H) \|_{L^2(T)}^2 + \hbar_T \, \| [\![\varepsilon \, \nabla v_H \, \cdot \, \boldsymbol{n}]\!] \|_{L^2(\partial T \cap \Omega)}^2,$$

where the scaling factors $\hbar_T = \min\{\varepsilon^{-1/2} \, h_T, 1\}$ ensure ε-robustness of the estimator.

In Fig. 4, we plot the error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{i}})$ for all $(\ell, \underline{k}, \underline{i}) \in \mathcal{Q}$ against the computational time for $\lambda_{\rm lin} \in \{0.1, 0.2, ..., 0.9\}$ and polynomial degrees $p \in \{1, 3\}$.

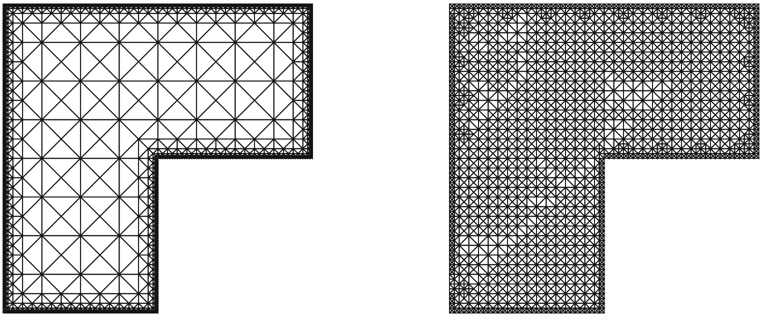


Fig. 5 Mesh plot of Experiment 3 for $p = 3$. Left: Adaptivity parameter $\lambda_{\text{lin}} = 0.2$. Right: Adaptivity parameter $\lambda_{\text{lin}} = 0.7$

The decay rate is of (expected) optimal order $p/2$. The choice of λ_{lin} does not play a major role in Fig. 4 (left) for $p = 1$, but significantly prolongs the preasymptotic phase for $p = 3$; see Fig. 4 (right). Figure 5 shows meshes with #`nDof` $= 12475$ for $\lambda_{\text{lin}} = 0.2$ and #`nDof` $= 12152$ for $\lambda_{\text{lin}} = 0.7$. We see that the selection of a large $\lambda_{\text{lin}} = 0.7$ results in fewer linearization steps as well as fewer and algebraic solver steps but is subsequently taken care of by the mesh adaptivity. Hence, we observe stronger refinement in the interior.

This experiment shows that Algorithm 1 is suitable for a setting with dominating reaction given that a suitable norm on $\mathcal{X}$ is chosen. A large choice of λ_{lin} seems possible, but pays off only after a long preasymptotic phase.

Acknowledgements The authors acknowledge TU Wien Bibliothek for financial support through its Open Access Funding Programme.

Funding Open access funding provided by TU Wien (TUW). This research was funded in whole or in part by the Austrian Science Fund (FWF) [10.55776/F65, 10.55776/I6802, and 10.55776/P33216]. For open access purposes, the author has applied a CC BY public copyright license to any author accepted manuscript version arising from this submission.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by/4.0/.

References

1. Babuška, I., Vogelius, M.: Feedback and adaptive finite element solution of one-dimensional boundary value problems. Numer. Math. **44**(1), 75–102 (1984). https://doi.org/10.1007/BF01389757
2. Dörfler, W.: A convergent adaptive algorithm for Poisson's equation. SIAM J. Numer. Anal. **33**(3), 1106–1124 (1996). https://doi.org/10.1137/0733054
3. Morin, P., Nochetto, R.H., Siebert, K.G.: Data oscillation and convergence of adaptive FEM. SIAM J. Numer. Anal. **38**(2), 466–488 (2000). https://doi.org/10.1137/S0036142999360044

4. Binev, P., Dahmen, W., DeVore, R.: Adaptive finite element methods with convergence rates. Numer. Math. **97**(2), 219–268 (2004). https://doi.org/10.1007/s00211-003-0492-7
5. Stevenson, R.: Optimality of a standard adaptive finite element method. Found. Comput. Math. **7**(2), 245–269 (2007). https://doi.org/10.1007/s10208-005-0183-0
6. Cascon, J.M., Kreuzer, C., Nochetto, R.H., Siebert, K.G.: Quasi-optimal convergence rate for an adaptive finite element method. SIAM J. Numer. Anal. **46**(5), 2524–2550 (2008). https://doi.org/10.1137/07069047X
7. Kreuzer, C., Siebert, K.G.: Decay rates of adaptive finite elements with Dörfler marking. Numer. Math. **117**(4), 679–716 (2011). https://doi.org/10.1007/s00211-010-0324-5
8. Cascón, J.M., Nochetto, R.H.: Quasioptimal cardinality of AFEM driven by nonresidual estimators. IMA J. Numer. Anal. **32**(1), 1–29 (2012). https://doi.org/10.1093/imanum/drr014
9. Feischl, M., Führer, T., Praetorius, D.: Adaptive FEM with optimal convergence rates for a certain class of nonsymmetric and possibly nonlinear problems. SIAM J. Numer. Anal. **52**(2), 601–625 (2014). https://doi.org/10.1137/120897225
10. Veeser, A.: Convergent adaptive finite elements for the nonlinear Laplacian. Numer. Math. **92**(4), 743–770 (2002). https://doi.org/10.1007/s002110100377
11. Diening, L., Kreuzer, C.: Linear convergence of an adaptive finite element method for the p-Laplacian equation. SIAM J. Numer. Anal. **46**(2), 614–638 (2008). https://doi.org/10.1137/070681508
12. Belenki, L., Diening, L., Kreuzer, C.: Optimality of an adaptive finite element method for the p-Laplacian equation. IMA J. Numer. Anal. **32**(2), 484–510 (2012). https://doi.org/10.1093/imanum/drr016
13. Garau, E.M., Morin, P., Zuppa, C.: Quasi-optimal convergence rate of an AFEM for quasi-linear problems of monotone type. Numer. Math. Theory Methods Appl. **5**(2), 131–156 (2012). https://doi.org/10.4208/nmtma.2012.m1023
14. Gantner, G., Haberl, A., Praetorius, D., Schimanko, S.: Rate optimality of adaptive finite element methods with respect to overall computational costs. Math. Comput. **90**(331), 2011–2040 (2021). https://doi.org/10.1090/mcom/3654
15. Carstensen, C., Feischl, M., Page, M., Praetorius, D.: Axioms of adaptivity. Comput. Math. Appl. **67**(6), 1195–1253 (2014). https://doi.org/10.1016/j.camwa.2013.12.003
16. Verfürth, R.: A Posteriori Error Estimation Techniques for Finite Element Methods. Oxford University Press, Oxford (2013). https://doi.org/10.1093/acprof:oso/9780199679423.001.0001
17. Becker, R., Brunner, M., Innerberger, M., Melenk, J.M., Praetorius, D.: Rate-optimal goal-oriented adaptive FEM for semilinear elliptic PDEs. Comput. Math. Appl. **118**, 18–35 (2022). https://doi.org/10.1016/j.camwa.2022.05.008
18. Brunner, M., Innerberger, M., Miraçi, A., Praetorius, D., Streitberger, J., Heid, P.: Adaptive FEM with quasi-optimal overall cost for nonsymmetric linear elliptic pdes. IMA J. Numer. Anal. **44**(3), 1560–1596 (2024). https://doi.org/10.1093/imanum/drad039
19. Haberl, A., Praetorius, D., Schimanko, S., Vohralík, M.: Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver. Numer. Math. **147**(3), 679–725 (2021). https://doi.org/10.1007/s00211-021-01176-w
20. Heid, P., Praetorius, D., Wihler, T.P.: Energy contraction and optimal convergence of adaptive iterative linearized finite element methods. Comput. Methods Appl. Math. **21**(2), 407–422 (2021). https://doi.org/10.1515/cmam-2021-0025
21. Bringmann, P., Feischl, M., Miraçi, A., Praetorius, D., Streitberger, J.: On full linear convergence and optimal complexity of adaptive FEM with inexact solver (2023) arXiv:2311.15738
22. Congreve, S., Wihler, T.P.: Iterative Galerkin discretizations for strongly monotone problems. J. Comput. Appl. Math. **311**, 457–472 (2017). https://doi.org/10.1016/j.cam.2016.08.014
23. Gantner, G., Haberl, A., Praetorius, D., Stiftner, B.: Rate optimal adaptive FEM with inexact solver for nonlinear operators. IMA J. Numer. Anal. **38**(4), 1797–1831 (2018). https://doi.org/10.1093/imanum/drx050
24. Heid, P., Wihler, T.P.: On the convergence of adaptive iterative linearized Galerkin methods. Calcolo **57**(3), 24 (2020). https://doi.org/10.1007/s10092-020-00368-4
25. Zarantonello, E.: Solving functional equations by contractive averaging. Math. Res. Center Report **160** (1960)
26. Cohen, A., Dahmen, W., DeVore, R.: Adaptive wavelet methods for elliptic operator equations: convergence rates. Math. Comput. **70**(233), 27–75 (2001). https://doi.org/10.1090/S0025-5718-00-01252-7

27. Cohen, A., Dahmen, W., Devore, R.: Adaptive wavelet schemes for nonlinear variational problems. SIAM J. Numer. Anal. **41**(5), 1785–1823 (2003). https://doi.org/10.1137/S0036142902412269
28. Carstensen, C., Gedicke, J.: An adaptive finite element eigenvalue solver of asymptotic quasi-optimal computational complexity. SIAM J. Numer. Anal. **50**(3), 1029–1057 (2012). https://doi.org/10.1137/090769430
29. Becker, R., Mao, S., Shi, Z.: A convergent nonconforming adaptive finite element method with quasi-optimal complexity. SIAM J. Numer. Anal. **47**(6), 4639–4659 (2010). https://doi.org/10.1137/070701479
30. El Alaoui, L., Ern, A., Vohralík, M.: Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems. Comput. Methods Appl. Mech. Eng. **200**(37–40), 2782–2795 (2011). https://doi.org/10.1016/j.cma.2010.03.024
31. Arioli, M., Georgoulis, E.H., Loghin, D.: Stopping criteria for adaptive finite element solvers. SIAM J. Sci. Comput. **35**(3), 1537–1559 (2013). https://doi.org/10.1137/120867421
32. Ern, A., Vohralík, M.: Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. SIAM J. Sci. Comput. **35**(4), 1761–1791 (2013). https://doi.org/10.1137/120896918
33. Amrein, M., Wihler, T.P.: Fully adaptive Newton–Galerkin methods for semilinear elliptic partial differential equations. SIAM J. Sci. Comput. **37**(4), 1637–1657 (2015). https://doi.org/10.1137/140983537
34. Heid, P., Wihler, T.P.: Adaptive iterative linearization Galerkin methods for nonlinear problems. Math. Comput. **89**(326), 2707–2734 (2020). https://doi.org/10.1090/mcom/3545
35. Amrein, M., Heid, P., Wihler, T.P.: A numerical energy reduction approach for semilinear diffusion–reaction boundary value problems based on steady-state iterations. SIAM J. Numer. Anal. **61**(2), 755–783 (2023). https://doi.org/10.1137/22M1478586
36. Becker, R., Brunner, M., Innerberger, M., Melenk, J.M., Praetorius, D.: Cost-optimal adaptive iterative linearized FEM for semilinear elliptic PDEs. ESAIM Math. Model. Numer. Anal. **57**(4), 2193–2225 (2023). https://doi.org/10.1051/m2an/2023036
37. Feischl, M.: Inf-sup stability implies quasi-orthogonality. Math. Comput. **91**(337), 2059–2094 (2022). https://doi.org/10.1090/mcom/3748
38. Zeidler, E.: Nonlinear Functional Analysis and Its Applications. Part II/B. Springer, New York (1990). https://doi.org/10.1007/978-1-4612-0985-0
39. Stevenson, R.: The completion of locally refined simplicial partitions created by bisection. Math. Comput. **77**(261), 227–241 (2008). https://doi.org/10.1090/S0025-5718-07-01959-X
40. Karkulik, M., Pavlicek, D., Praetorius, D.: On 2D newest vertex bisection: optimality of mesh-closure and H^1-stability of L_2-projection. Constr. Approx. **38**(2), 213–234 (2013). https://doi.org/10.1007/s00365-013-9192-4
41. Diening, L., Gehring, L., Storn, J.: Adaptive mesh refinement for arbitrary initial triangulations (2023) arXiv:2306.02674
42. Aurada, M., Feischl, M., Führer, T., Karkulik, M., Praetorius, D.: Efficiency and optimality of some weighted-residual error estimator for adaptive 2D boundary element methods. Comput. Methods Appl. Math. **13**(3), 305–332 (2013). https://doi.org/10.1515/cmam-2013-0010
43. Bank, R.E., Holst, M., Szypowski, R., Zhu, Y.: Finite element error estimates for critical growth semilinear problems without angle conditions (2011) arXiv:1108.3661
44. Fučík, S., Kufner, A.: Nonlinear Differential Equations. Studies in Applied Mechanics, vol. 2, p. 359. Elsevier, Amsterdam (1980)
45. Chen, L., Nochetto, R.H., Xu, J.: Optimal multilevel methods for graded bisection grids. Numer. Math. **120**(1), 1–34 (2012). https://doi.org/10.1007/s00211-011-0401-4
46. Wu, J., Zheng, H.: Uniform convergence of multigrid methods for adaptive meshes. Appl. Numer. Math. **113**, 109–123 (2017). https://doi.org/10.1016/j.apnum.2016.11.005
47. Innerberger, M., Miraçi, A., Praetorius, D., Streitberger, J.: hp-robust multigrid solver on locally refined meshes for FEM discretizations of symmetric elliptic PDEs. ESAIM Math. Model. Numer. Anal. **58**(1), 247–272 (2024). https://doi.org/10.1051/m2an/2023104
48. Pfeiler, C.-M., Praetorius, D.: Dörfler marking with minimal cardinality is a linear complexity problem. Math. Comput. **89**(326), 2735–2752 (2020). https://doi.org/10.1090/mcom/3553
49. Bespalov, A., Haberl, A., Praetorius, D.: Adaptive FEM with coarse initial mesh guarantees optimal convergence rates for compactly perturbed elliptic problems. Comput. Methods Appl. Mech. Eng. **317**, 318–340 (2017). https://doi.org/10.1016/j.cma.2016.12.014

50. Innerberger, M., Praetorius, D.: MooAFEM: an object oriented Matlab code for higher-order adaptive FEM for (nonlinear) elliptic PDEs. Appl. Math. Comput. **442**, 127731 (2023). https://doi.org/10.1016/j.amc.2022.127731

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.