

COST-OPTIMAL ADAPTIVE ITERATIVE LINEARIZED FEM FOR SEMILINEAR ELLIPTIC PDES

ROLAND BECKER¹, MAXIMILIAN BRUNNER^{2,*}, MICHAEL INNERBERGER²,
JENS MARKUS MELENK² AND DIRK PRAETORIUS²

Abstract. We consider scalar semilinear elliptic PDEs where the nonlinearity is strongly monotone, but only locally Lipschitz continuous. We formulate an adaptive iterative linearized finite element method (AILFEM) which steers the local mesh refinement as well as the iterative linearization of the arising nonlinear discrete equations. To this end, we employ a damped Zarantonello iteration so that, in each step of the algorithm, only a linear Poisson-type equation has to be solved. We prove that the proposed AILFEM strategy guarantees convergence with optimal rates, where rates are understood with respect to the overall computational complexity (*i.e.*, the computational time). Moreover, we formulate and test an adaptive algorithm where also the damping parameter of the Zarantonello iteration is adaptively adjusted. Numerical experiments underline the theoretical findings.

Mathematics Subject Classification. 65N30, 65N50, 65N15, 65Y20, 41A25.

Received November 8, 2022. Accepted April 23, 2023.

1. INTRODUCTION

1.1. State of the art

Cost-optimal computation of a discrete solution with an error below a given tolerance is the prime aim of any numerical method. Since convergence of numerical schemes is usually (but not necessarily) spoiled by singularities of the (given) data or the (unknown) solution, *a posteriori* error estimation and adaptive mesh refinement schemes are pivotal to reliable and efficient numerical approximation. This is the foundation of adaptive finite element methods (AFEM), for which the mathematical understanding of convergence and optimality is fairly mature; we refer to [7, 12, 14, 15, 18, 21, 33, 35, 36, 38] for linear elliptic equations, to [10, 17, 23, 26, 40] for certain quasi-linear PDEs, and to [13] for an overview of available results on rate-optimal AFEM.

In particular, for nonlinear PDEs, the arising discrete equations must be solved iteratively. The interplay of adaptive mesh refinement and iterative solvers has been treated extensively in the literature; we refer, *e.g.*, to [3, 4, 8, 38] for algebraic solvers for linear PDEs, to [1, 19, 23, 25, 28, 29, 31] for the iterative linearization of nonlinear PDEs, and to [20, 27] for fully adaptive schemes including linearization and algebraic solver. For the

Keywords and phrases. Adaptive iterative linearized finite element method, Semilinear PDEs, Iterative solver, A posteriori error estimation, Convergence, Optimal convergence rates, Cost-optimality.

¹ Université de Pau et des Pays de l'Adour, IPRA-LMAP, Avenue de l'Université, BP 1155, 64013 PAU Cedex, France.

² TU Wien, Institute of Analysis and Scientific Computing, Wiedner Hauptstr. 8–10/E101/4, 1040 Vienna, Austria.

*Corresponding author: maximilian.brunner@asc.tuwien.ac.at

latter works, the consideration is usually restricted to the class of strongly monotone and globally Lipschitz continuous nonlinearities; see [25] for the first plain convergence result, [28] for an abstract framework for plain convergence of adaptive iteratively linearized finite element methods (AILFEM), [23, 24] for rate-optimality of AILFEM based on the Zarantonello iteration (as proposed in [16]), and [30] for rate-optimality for other linearization strategies including the Kačanov iteration as well as the damped Newton method. In particular, we note that [24, 27, 30] prove optimal convergence rates with respect to the overall computational cost. For more general nonlinear operators, optimal convergence rates are empirically observed (*e.g.*, [20]), but the quest for a sound mathematical analysis is still ongoing.

1.2. Contributions of the present work

We prove optimal convergence of AILFEM for strongly monotone, but only locally Lipschitz continuous operators, where our interest stems from the treatment of semilinear elliptic PDEs. For $d \in \{1, 2, 3\}$ and a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$, our model problem reads: Find the (unique) solution $u^* \in H_0^1(\Omega)$ to the (scalar) semilinear elliptic PDE

$$-\operatorname{div}(\mathbf{A}\nabla u^*) + b(u^*) = f - \operatorname{div} \mathbf{f} \text{ in } \Omega \quad \text{subject to} \quad u^* = 0 \text{ on } \partial\Omega, \quad (1)$$

where we refer to Section 3 for a discussion of the precise assumptions on the diffusion matrix \mathbf{A} , the semilinearity b , and the given data f and \mathbf{f} . The presented AILFEM algorithm employs the Zarantonello linearization with a damping parameter $\delta > 0$, requiring only to solve a *linear* Poisson-type problem in each linearization step. The AILFEM algorithm takes the form



where the first step represents an inner loop of the Zarantonello iteration and error estimation by a residual *a posteriori* error estimator. This inner loop is stopped when the linearization error (measured in terms of the energy difference of discrete Zarantonello iterates) is small with respect to the discretization error (measured in terms of the error estimator). However, since the PDE operator is only locally Lipschitz continuous, the stopping criterion must be slightly extended when compared to that of [24, 28, 30] for globally Lipschitz continuous operators. As usual in this context, we employ the Dörfler marking to single out elements for refinement, and mesh refinement relies on newest vertex bisection.

We prove that the solver iterates are uniformly bounded, provided that the Zarantonello parameter δ is chosen appropriately (Cor. 10). For arbitrary adaptivity parameters (θ for marking and λ for stopping the Zarantonello iteration), we then prove *full* linear convergence (Thm. 13), *i.e.*, linear convergence regardless of the algorithmic decision for yet another solver step or mesh refinement. For sufficiently small marking parameters, this even guarantees *rate-optimality with respect to the number of degrees of freedom* (Thm. 16) and *cost-optimality, i.e.*, rate-optimality with respect to the overall computational cost (Cor. 18).

1.3. Outline

This work is organized as follows: In Section 2, we present our adaptive iterative linearized finite element method (Algo. A) and the details of its individual steps. This includes the discussion of the abstract Hilbert space setting, the precise assumptions for the iterative solver, and a discussion of the extended stopping criterion. Finally, we prove full linear convergence of the proposed AILFEM algorithm (Thm. 13) and optimal rates both with respect to the degrees of freedom (Thm. 16) as well as the overall computational cost (Cor. 18). In Section 3, we introduce and discuss semilinear elliptic PDEs, which fit into the abstract framework of Section 2. Section 4 presents a practical extension of our AILFEM strategy (Algo. B), which includes the adaptive choice of the Zarantonello damping parameter δ . In Section 5, we support our theoretical findings with numerical

experiments. Finally, Appendix A concludes the work by providing additional material, which allows us to apply the abstract setting to a wider range of problems like non-scalar semilinear PDEs.

1.4. General notation

Without ambiguity, we use $|\cdot|$ to denote the absolute value $|\lambda|$ of a scalar $\lambda \in \mathbb{R}$, the Euclidean norm $|x|$ of a vector $x \in \mathbb{R}^d$, and the Lebesgue measure $|\omega|$ of a set $\omega \subseteq \mathbb{R}^d$, depending on the respective context. Furthermore, $\#\mathcal{U}$ denotes the cardinality of a finite set \mathcal{U} .

2. STRONGLY MONOTONE OPERATORS

In this section, we present the mathematical heart of our analysis, which will later be applied to strongly monotone semilinear PDEs.

2.1. Abstract model problem

Let \mathcal{X} be a Hilbert space over \mathbb{R} with scalar product $\langle \cdot, \cdot \rangle$ 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 $\mathcal{A}: \mathcal{X} \rightarrow \mathcal{X}'$ be a nonlinear operator. We suppose that \mathcal{A} is **strongly monotone**, *i.e.*, there exists $\alpha > 0$ such that

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

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

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

Remark 1. ([43], p. 565) defines local Lipschitz continuity as follows: For all $\Theta > 0$, there exists $L'[\Theta] > 0$ such that

$$\langle \mathcal{A}v - \mathcal{A}w, \varphi \rangle \leq L'[\Theta] \|v - w\| \|\varphi\| \quad \text{for all } v, w, \varphi \in \mathcal{X} \text{ with } \max\{\|v\|, \|w\|\} \leq \Theta. \tag{2}$$

Conditions (LIP) and (2) are indeed equivalent in the sense that

$$\begin{aligned} \max\{\|v\|, \|w\|\} \leq \max\{\|v\|, \|v - w\| + \|w\|\} &\leq 2\vartheta, \\ \max\{\|v\|, \|v - w\|\} \leq \max\{\|v\|, \|v\| + \|w\|\} &\leq 2\Theta. \end{aligned}$$

However, (LIP) is better suited for the inductive structure in the proof of Corollary 5.

Without loss of generality, we may suppose that $\mathcal{A}0 \neq F \in \mathcal{X}'$. We consider the operator equation

$$\mathcal{A}u^* = F. \tag{3}$$

For any closed subspace $\mathcal{X}_H \subseteq \mathcal{X}$, we consider the corresponding Galerkin discretization

$$\langle \mathcal{A}u_H^*, v_H \rangle = \langle F, v_H \rangle \quad \text{for all } v_H \in \mathcal{X}_H. \tag{4}$$

We observe that the setting of strongly monotone and locally Lipschitz operators yields existence and uniqueness of the solutions to (3) and (4) as well as a Céa-type estimate.

Proposition 2. *Suppose that \mathcal{A} satisfies (SM) and (LIP). Then, (3) and (4) admit unique solutions $u^* \in \mathcal{X}$ and $u_H^* \in \mathcal{X}_H$, respectively, and it holds that*

$$\max\{\|u^*\|, \|u_H^*\|\} \leq M := \frac{1}{\alpha} \|F - \mathcal{A}0\|_{\mathcal{X}'} \neq 0 \tag{5}$$

as well as

$$\|u^* - u_H^*\| \leq C_{\text{Céa}} \min_{v_H \in \mathcal{X}_H} \|u^* - v_H\| \quad \text{with} \quad C_{\text{Céa}} = L[2M]/\alpha. \tag{6}$$

Proof. Since \mathcal{A} is (even locally Lipschitz) continuous, existence of u_H^* follows from the Browder–Minty theorem on monotone operators ([43], Thm. 26.A). Uniqueness of u_H^* follows from strong monotonicity, since any two solutions $u_H^*, u_H \in \mathcal{X}_H$ to (4) satisfy

$$\alpha \|u_H^* - u_H\|^2 \stackrel{\text{(SM)}}{\leq} \langle \mathcal{A}u_H^* - \mathcal{A}u_H, u_H^* - u_H \rangle \stackrel{\text{(4)}}{=} 0$$

and hence $u_H^* = u_H$. Boundedness (5) follows from

$$\alpha \|u_H^*\|^2 \stackrel{\text{(SM)}}{\leq} \langle \mathcal{A}u_H^* - \mathcal{A}0, u_H^* \rangle = \langle F - \mathcal{A}0, u_H^* \rangle \leq \|F - \mathcal{A}0\|_{\mathcal{X}'} \|u_H^*\|.$$

Since (3) is equivalent to (4) with $\mathcal{X} = \mathcal{X}_H$, the foregoing results also cover $u^* \in \mathcal{X}$. This concludes the proof of (5). To see the Céa-type estimate (6), recall the Galerkin orthogonality

$$\langle \mathcal{A}u^* - \mathcal{A}u_H^*, v_H \rangle = 0 \quad \text{for all } v_H \in \mathcal{X}_H. \tag{7}$$

For $v_H \in \mathcal{X}_H$, standard reasoning leads us to

$$\begin{aligned} \alpha \|u^* - u_H^*\|^2 &\stackrel{\text{(SM)}}{\leq} \langle \mathcal{A}u^* - \mathcal{A}u_H^*, u^* - u_H^* \rangle \stackrel{\text{(7)}}{=} \langle \mathcal{A}u^* - \mathcal{A}u_H^*, u^* - v_H \rangle \\ &\stackrel{\text{(LIP)}}{\leq} L[2M] \|u^* - u_H^*\| \|u^* - v_H\|. \end{aligned}$$

Rearranging the last estimate, we prove (6), where the minimum is attained since \mathcal{X}_H is closed. This concludes the proof. \square

Finally, we suppose that the operator \mathcal{A} possesses a potential \mathcal{P} : there exists a Gâteaux differentiable function $\mathcal{P}: \mathcal{X} \rightarrow \mathbb{R}$ such that its derivative $d\mathcal{P}: \mathcal{X} \rightarrow \mathcal{X}'$ coincides with \mathcal{A} , i.e., it holds that

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

We define the energy $\mathcal{E}(v) := (\mathcal{P} - F)v$, where F is the right-hand side from (3).

Note that the energy \mathcal{E} trivially satisfies that

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

and all these energy differences are non-negative; see (10).

Moreover, assumption (POT) admits the following classical equivalence:

Lemma 3 (see e.g., ([23], Lemma 5.1)). *Suppose that \mathcal{A} satisfies (SM), (LIP), and (POT). Let $\vartheta \geq M$. Let $v_H \in \mathcal{X}_H$ with $\|v_H - u_H^*\| \leq \vartheta$. Then, it holds that*

$$\frac{\alpha}{2} \|v_H - u_H^*\|^2 \leq \mathcal{E}(v_H) - \mathcal{E}(u_H^*) \leq \frac{L[\vartheta]}{2} \|v_H - u_H^*\|^2. \tag{9}$$

In particular, the solution u_H^* of (4) is indeed the unique minimizer of \mathcal{E} in \mathcal{X}_H , i.e.,

$$\mathcal{E}(u_H^*) \leq \mathcal{E}(v_H) \quad \text{for all } v_H \in \mathcal{X}_H, \tag{10}$$

and, therefore, (4) can equivalently be reformulated as an energy minimization problem:

$$\text{Find } u_H^* \in \mathcal{X}_H \text{ such that } \mathcal{E}(u_H^*) = \min_{v_H \in \mathcal{X}_H} \mathcal{E}(v_H). \quad \square$$

2.2. Zarantonello iteration

Let $\mathcal{X}_H \subseteq \mathcal{X}$ be a closed subspace. For given damping parameter $\delta > 0$, we define the Zarantonello mapping $\Phi_H(\delta; \cdot): \mathcal{X}_H \rightarrow \mathcal{X}_H$ by

$$\langle\langle \Phi_H(\delta; w_H), v_H \rangle\rangle = \langle\langle w_H, v_H \rangle\rangle + \delta \langle F - \mathcal{A}w_H, v_H \rangle \quad \text{for all } v_H \in \mathcal{X}_H. \tag{11}$$

Clearly, existence and uniqueness of $\Phi_H(\delta; w_H) \in \mathcal{X}_H$ and hence well-posedness of $\Phi_H(\delta; \cdot)$ follows from the Riesz theorem. The following two estimates are obvious: first,

$$\|\|\Phi_H(\delta; w_H) - w_H\|\| \leq \delta \|F - \mathcal{A}w_H\|_{\mathcal{X}'}^2 = \delta \sup_{v \in \mathcal{X} \setminus \{0\}} \frac{\langle F - \mathcal{A}w_H, v \rangle}{\|v\|} \quad \text{for all } w_H \in \mathcal{X}_H, \tag{12}$$

second,

$$\|\|\Phi_H(\delta; v_H) - \Phi_H(\delta; w_H)\|\| \leq \|v_H - w_H\| + \delta \|\mathcal{A}v_H - \mathcal{A}w_H\|_{\mathcal{X}'}^2 \quad \text{for all } v_H, w_H \in \mathcal{X}_H. \tag{13}$$

Due to the local Lipschitz continuity (LIP) of \mathcal{A} , this proves that also $\Phi_H(\delta; \cdot)$ is locally Lipschitz continuous. By definition, $u_H^* \in \mathcal{X}_H$ solves (4) if and only if it is a fixed point of $\Phi_H(\delta; \cdot)$, i.e., $u_H^* = \Phi_H(\delta; u_H^*)$.

2.3. Zarantonello iteration and norm contraction

Let $\mathcal{X}_H \subseteq \mathcal{X}$ be a closed subspace. The next proposition ([43], Sect. 25.4) proves local contraction of $\Phi_H(\delta; \cdot)$ with respect to the energy norm. For the convenience of the reader, we include the proof to highlight that local Lipschitz continuity suffices.

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

$$\|\|\Phi_H(\delta; v_H) - \Phi_H(\delta; w_H)\|\| \leq q_N[\delta] \|v_H - w_H\|. \tag{14}$$

We note that $q_N[\delta] \rightarrow 1$ as $\delta \rightarrow 0$. Moreover, for known α and $L[\vartheta]$, the contraction constant $q_N[\delta]^2 = 1 - \alpha^2/L[\vartheta]^2 = 1 - \alpha\delta$ is minimal and only attained for $\delta = \alpha/L[\vartheta]^2$.

Proof. Recall that the Riesz mapping

$$I_H: \mathcal{X}_H \rightarrow \mathcal{X}'_H, \quad v_H \mapsto I_H(v_H) := \langle\langle \cdot, v_H \rangle\rangle \quad \text{for all } v_H \in \mathcal{X}_H \tag{15}$$

is an isometric isomorphism; cf., e.g., ([42], Chap. III.6). Therefore, a reformulation of the Zarantonello iteration reads

$$\langle\langle \Phi_H(\delta; w_H), \varphi_H \rangle\rangle = \langle\langle w_H, \varphi_H \rangle\rangle + \delta \langle\langle \varphi_H, I_H^{-1}(F - \mathcal{A}w_H) \rangle\rangle \quad \text{for all } \varphi_H, w_H \in \mathcal{X}_H.$$

Given $v_H, w_H \in \mathcal{X}_H$ with $\max\{\|v_H\|, \|v_H - w_H\|\} \leq \vartheta$, we exploit the last equality for $\Phi_H(\delta; v_H)$ by subtraction of $\Phi_H(\delta; w_H)$ and use $\varphi_H = \Phi_H(\delta; v_H) - \Phi_H(\delta; w_H)$ to arrive at

$$\begin{aligned} \|\|\Phi_H(\delta; v_H) - \Phi_H(\delta; w_H)\|\|^2 &= \|v_H - w_H\|^2 - 2\delta \langle\langle v_H - w_H, I_H^{-1}(\mathcal{A}v_H - \mathcal{A}w_H) \rangle\rangle \\ &\quad + \delta^2 \|I_H^{-1}(\mathcal{A}v_H - \mathcal{A}w_H)\|^2. \end{aligned}$$

The isometry property of I_H implies that

$$\|I_H^{-1}(\mathcal{A}v_H - \mathcal{A}w_H)\|^2 \stackrel{(15)}{=} \|\mathcal{A}v_H - \mathcal{A}w_H\|_{\mathcal{X}'}^2 \stackrel{(LIP)}{\leq} L[\vartheta]^2 \|v_H - w_H\|^2.$$

Moreover, it holds that

$$\langle\langle v_H - w_H, I_H^{-1}(\mathcal{A}v_H - \mathcal{A}w_H) \rangle\rangle \stackrel{(15)}{=} \langle\mathcal{A}v_H - \mathcal{A}w_H, v_H - w_H\rangle \stackrel{(SM)}{\geq} \alpha \|v_H - w_H\|^2.$$

Combining these observations, we see that

$$0 \leq \|\Phi_H(\delta; v_H) - \Phi_H(\delta; w_H)\|^2 \leq [1 - 2\delta\alpha + \delta^2 L[\vartheta]^2] \|v_H - w_H\|^2.$$

Rearranging $q_N[\delta]^2 := 1 - 2\delta\alpha + \delta^2 L[\vartheta]^2 = 1 - \delta(2\alpha - \delta L[\vartheta]^2)$, we conclude the first claim. Finally, it follows from elementary calculus that $\delta = \alpha/L[\vartheta]^2$ is the unique minimizer of the quadratic polynomial $q_N[\delta]$ if α and $L[\vartheta]^2$ are fixed. This concludes the proof. \square

Corollary 5. *Suppose that \mathcal{A} satisfies (SM) and (LIP). Let $u_H^0 \in \mathcal{X}_H$ with $\|u_H^0\| \leq 2M$. Let $0 < \delta < 2\alpha/L[3M]^2$ and let $0 < q_N[\delta] < 1$ be chosen according to Proposition 4, where $\vartheta = 3M$. Define*

$$u_H^{k+1} := \Phi_H(\delta; u_H^k) \quad \text{for all } k \in \mathbb{N}_0. \quad (16)$$

Then, it holds that

$$(1 - q_N[\delta]) \|u_H^* - u_H^k\| \leq \|u_H^{k+1} - u_H^k\| \leq (1 + q_N[\delta]) \|u_H^* - u_H^k\| \quad (17)$$

and

$$\|u_H^* - u_H^{k+1}\| \leq q_N[\delta] \|u_H^* - u_H^k\| \leq q_N[\delta]^{k+1} \|u_H^* - u_H^0\| \leq 3M \quad \text{for all } k \in \mathbb{N}_0. \quad (18)$$

In particular, it follows that

$$\|u_H^k\| \leq 4M \quad \text{for all } k \in \mathbb{N}_0. \quad (19)$$

Proof. The claim (18) is proved by induction on k . By recalling (5), it holds that $\|u_H^*\| \leq M$ as well as $\|u_H^* - u_H^0\| \leq \|u_H^*\| + \|u_H^0\| \leq 3M$. Therefore, Proposition 4 proves that

$$\|u_H^* - u_H^1\| = \|\Phi_H(\delta; u_H^*) - \Phi_H(\delta; u_H^0)\| \stackrel{(14)}{\leq} q_N[\delta] \|u_H^* - u_H^0\| \leq 3M.$$

This proves (18) for $k = 0$. In the induction step, we know that $\|u_H^* - u_H^k\| \leq 3M$. As before, (14) from Proposition 4 and the induction hypothesis prove that

$$\begin{aligned} \|u_H^* - u_H^{k+1}\| &= \|\Phi_H(\delta; u_H^*) - \Phi_H(\delta; u_H^k)\| \stackrel{(14)}{\leq} q_N[\delta] \|u_H^* - u_H^k\| \\ &\leq q_N[\delta]^{k+1} \|u_H^* - u_H^0\| \leq 3M. \end{aligned}$$

This proves (18) for general $k \in \mathbb{N}_0$, and the inequalities (17) follow from (14) and the triangle inequality. Moreover, the triangle inequality yields that

$$\|u_H^k\| \leq \|u_H^*\| + \|u_H^* - u_H^k\| \leq 4M.$$

This concludes the proof. \square

Corollary 6. *Suppose that \mathcal{A} satisfies (SM) and (LIP). Let $u_H^0 \in \mathcal{X}_H$ with $\|u_H^0\| \leq 2M$. Let $0 < \delta < 2\alpha/L[6M]^2$ and let $0 < q_N[\delta] < 1$ be chosen according to Proposition 4, where $\vartheta = 6M$. Then, the Zarantonello iterates from (16) satisfy (17)–(19) as well as*

$$\|u_H^{k+1} - u_H^k\| \leq q_N[\delta] \|u_H^k - u_H^{k-1}\| \leq q_N[\delta]^k \|u_H^1 - u_H^0\| \leq 6M \quad \text{for all } k \in \mathbb{N}. \quad (20)$$

Proof. Since $L[3M] \leq L[6M]$, it only remains to prove (20). We argue by induction and note that

$$\|u_H^1 - u_H^0\| \leq \|u_H^1\| + \|u_H^0\| \stackrel{(19)}{\leq} 6M.$$

Therefore, Proposition 4 proves that

$$\| \|u_H^2 - u_H^1\| \| = \| \Phi_H(\delta; u_H^1) - \Phi_H(\delta; u_H^0) \| \stackrel{(14)}{\leq} q_N[\delta] \| \|u_H^1 - u_H^0\| \| \leq 6M.$$

This proves (20) for $k = 1$. In the induction step, we know that $\| \|u_H^{k+1} - u_H^k\| \| \leq 6M$. Therefore, Proposition 4 and the induction hypothesis prove that

$$\| \|u_H^{k+2} - u_H^{k+1}\| \| = \| \Phi_H(\delta; u_H^{k+1}) - \Phi_H(\delta; u_H^k) \| \stackrel{(14)}{\leq} q_N[\delta] \| \|u_H^{k+1} - u_H^k\| \| \leq 6M.$$

This proves (20) for general $k \in \mathbb{N}$ and concludes the proof. □

2.4. Zarantonello iteration and energy contraction

Let $\mathcal{X}_H \subseteq \mathcal{X}$ be a closed subspace. The next result extends the abstract lower bound from ([28] Prop. 1) to the Zarantonello iteration in the locally Lipschitz continuous setting.

Lemma 7. *Suppose that \mathcal{A} satisfies (SM), (LIP), and (POT). Let $u_H^0 \in \mathcal{X}_H$ with $\| \|u_H^0\| \| \leq 2M$. Then, for $0 < \delta < 2\alpha/L[6M]^2$, the Zarantonello iteration (11) yields that*

$$0 \leq \kappa[\delta] \| \|u_H^{k+1} - u_H^k\| \|^2 \leq \mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1}) \leq K[\delta] \| \|u_H^{k+1} - u_H^k\| \|^2, \tag{21}$$

where $\kappa[\delta] = (\delta^{-1} - L[6M]/2) > 0$ and $K[\delta] = (\delta^{-1} - \alpha/2)$.

Proof. Define $e_H^{k+1} := u_H^{k+1} - u_H^k$ for all $k \in \mathbb{N}_0$. Then, (POT) guarantees that $\mathcal{E} = \mathcal{P} - F$ is Gâteaux differentiable. Define $\varphi(t) := \mathcal{E}(u_H^k + t e_H^{k+1})$ for $t \in [0, 1]$ and observe that

$$\varphi'(t) = \langle d\mathcal{E}(u_H^k + t e_H^{k+1}), e_H^{k+1} \rangle = \langle \mathcal{A}(u_H^k + t e_H^{k+1}) - F, e_H^{k+1} \rangle.$$

For $0 < \delta < 2\alpha/L[6M]^2$, Corollary 6 together with the boundedness $\| \|u_H^k\| \| \leq 4M$ from (19) and the convexity of the norm show that

$$\max\{ \| \|e_H^{k+1}\| \|, \| \|u_H^k - t e_H^{k+1}\| \| \} \leq 6M \quad \text{for all } k \in \mathbb{N}_0. \tag{22}$$

With the fundamental theorem of calculus and the Zarantonello iteration (11), we see that

$$\begin{aligned} \mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1}) &= - \int_0^1 \langle \mathcal{A}(u_H^k + t e_H^{k+1}) - F, e_H^{k+1} \rangle dt \\ &= - \int_0^1 \langle \mathcal{A}(u_H^k + t e_H^{k+1}) - \mathcal{A}u_H^k, e_H^{k+1} \rangle dt - \langle \mathcal{A}u_H^k - F, e_H^{k+1} \rangle \\ &\stackrel{(11)}{=} - \int_0^1 \langle \mathcal{A}(u_H^k + t e_H^{k+1}) - \mathcal{A}u_H^k, e_H^{k+1} \rangle dt + \frac{1}{\delta} \langle \langle e_H^{k+1}, e_H^{k+1} \rangle \rangle \\ &\stackrel{(LIP)}{\geq} \left(\frac{1}{\delta} - \int_0^1 tL[6M] dt \right) \| \|u_H^{k+1} - u_H^k\| \|^2 = \left(\frac{1}{\delta} - \frac{L[6M]}{2} \right) \| \|u_H^{k+1} - u_H^k\| \|^2. \end{aligned}$$

Since $\delta < 2\alpha/L[6M]^2 \leq 2/L[6M]$, it follows that $\kappa[\delta] = (1/\delta - L[6M]/2) > 0$. This proves the lower bound in (21). Moreover, the same argument also yields that

$$\begin{aligned} \mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1}) &\stackrel{(11)}{=} - \int_0^1 \langle \mathcal{A}(u_H^k + t e_H^{k+1}) - \mathcal{A}u_H^k \rangle e_H^{k+1} dt + \frac{1}{\delta} \langle \langle e_H^{k+1}, e_H^{k+1} \rangle \rangle \\ &\stackrel{(SM)}{\leq} \left(\frac{1}{\delta} - \int_0^1 \alpha t dt \right) \| \|u_H^{k+1} - u_H^k\| \|^2 = \left(\frac{1}{\delta} - \frac{\alpha}{2} \right) \| \|u_H^{k+1} - u_H^k\| \|^2. \end{aligned}$$

This concludes the proof. □

The Zarantonello iterates are also contractive with respect to the energy difference.

Proposition 8 (Energy contraction). *Suppose that \mathcal{A} satisfies (SM), (LIP), and (POT). Then, for $0 < \delta < 2\alpha/L[6M]^2$, it holds that*

$$0 \leq \mathcal{E}(u_H^{k+1}) - \mathcal{E}(u_H^*) \leq q_{\mathcal{E}}[\delta]^2 [\mathcal{E}(u_H^k) - \mathcal{E}(u_H^*)] \quad \text{for all } k \in \mathbb{N}_0 \quad (23a)$$

with contraction constant

$$0 \leq q_{\mathcal{E}}[\delta]^2 := 1 - \left(1 - \frac{\delta L[6M]}{2}\right) \frac{2\delta\alpha^2}{L[3M]} < 1. \quad (23b)$$

We note that $q_{\mathcal{E}}[\delta] \rightarrow 1$ as $\delta \rightarrow 0$. Furthermore, for all $k \in \mathbb{N}_0$, it holds that

$$(1 - q_{\mathcal{E}}[\delta]^2) [\mathcal{E}(u_H^k) - \mathcal{E}(u_H^*)] \leq \mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1}) \leq (1 + q_{\mathcal{E}}[\delta]^2) [\mathcal{E}(u_H^k) - \mathcal{E}(u_H^*)]. \quad (24)$$

Proof. First, we observe that

$$\begin{aligned} \alpha \|u_H^* - u_H^k\|^2 &\leq \langle \mathcal{A}u_H^* - \mathcal{A}u_H^k, u_H^* - u_H^k \rangle \stackrel{(4)}{=} \langle F - \mathcal{A}u_H^k, u_H^* - u_H^k \rangle \\ &\stackrel{(11)}{=} \frac{1}{\delta} \langle u_H^{k+1} - u_H^k, u_H^* - u_H^k \rangle \leq \frac{1}{\delta} \|u_H^{k+1} - u_H^k\| \|u_H^* - u_H^k\|. \end{aligned} \quad (25)$$

Since $0 < \delta < 2\alpha/L[6M]^2$, it follows that

$$\begin{aligned} 0 &\stackrel{(9)}{\leq} \mathcal{E}(u_H^{k+1}) - \mathcal{E}(u_H^*) = \mathcal{E}(u_H^k) - \mathcal{E}(u_H^*) - [\mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1})] \\ &\stackrel{(21)}{\leq} \mathcal{E}(u_H^k) - \mathcal{E}(u_H^*) - \left(\frac{1}{\delta} - \frac{L[6M]}{2}\right) \|u_H^{k+1} - u_H^k\|^2 \\ &\stackrel{(25)}{\leq} \mathcal{E}(u_H^k) - \mathcal{E}(u_H^*) - \left(\frac{1}{\delta} - \frac{L[6M]}{2}\right) \delta^2 \alpha^2 \|u_H^* - u_H^k\|^2 \\ &\stackrel{(9)}{\leq} \left[1 - \left(1 - \frac{\delta L[6M]}{2}\right) \frac{2\delta\alpha^2}{L[3M]}\right] [\mathcal{E}(u_H^k) - \mathcal{E}(u_H^*)], \end{aligned}$$

where (9) holds due to (18) from Corollary 5. This proves (23). The inequalities (24) follow from the triangle inequality. This concludes the proof. \square

Remark 9. For a globally Lipschitz continuous \mathcal{A} with Lipschitz constant L , we observe that the energy contraction factor is minimal for $\delta = 1/L$, where $q_{\mathcal{E}}[\delta]^2 = 1 - \frac{\alpha^2}{L^2}$. In contrast, the optimal norm contraction factor $q_{\mathcal{N}}[\delta]^2 = 1 - \frac{\alpha^2}{L^2}$ is obtained for $\delta = \frac{\alpha}{L^2}$; cf. Proposition 4. To allow a larger damping parameter $\delta > 0$, energy contraction is preferred.

2.5. Mesh refinement

From now on, let \mathcal{T}_0 be a given conforming triangulation of the polyhedral Lipschitz domain $\Omega \subset \mathbb{R}^d$ with $d \geq 1$. For mesh refinement, we employ newest vertex bisection (NVB) for $d \geq 2$ (see e.g., [39]), or the 1D bisection from [5] for $d = 1$. For each triangulation \mathcal{T}_H and a set of marked elements $\mathcal{M}_H \subseteq \mathcal{T}_H$, let $\mathcal{T}_h := \text{refine}(\mathcal{T}_H, \mathcal{M}_H)$ be the coarsest triangulation such that all $T \in \mathcal{M}_H$ have been refined, i.e., $\mathcal{M}_H \subseteq \mathcal{T}_H \setminus \mathcal{T}_h$. We write $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$, if \mathcal{T}_h results from \mathcal{T}_H by finitely many steps of refinement. To abbreviate notation, let $\mathbb{T} := \mathbb{T}(\mathcal{T}_0)$.

Throughout, each triangulation $\mathcal{T}_H \in \mathbb{T}$ is associated with a conforming finite-dimensional space $\mathcal{X}_H \subset \mathcal{X}$, and we suppose that mesh refinement $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ implies nestedness $\mathcal{X}_H \subseteq \mathcal{X}_h \subset \mathcal{X}$.

2.6. Axioms of adaptivity and a posteriori error estimator

For $\mathcal{T}_H \in \mathbb{T}$ and $v_H \in \mathcal{X}_H$, let

$$\eta_H(T, \cdot): \mathcal{X}_H \rightarrow \mathbb{R}_{\geq 0} \quad \text{for all } T \in \mathcal{T}_H \tag{26}$$

be the local contributions of an a posteriori error estimator

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

We suppose that the error estimator η_H satisfies the following axioms of adaptivity from [13] with a slightly relaxed variant of stability (A1) from [9].

(A1) Stability: For all $\vartheta > 0$ and all¹ $\mathcal{U}_H \subseteq \mathcal{T}_h \cap \mathcal{T}_H$, there exists $C_{\text{stab}}[\vartheta] > 0$ such that for all $v_h \in \mathcal{X}_h$ and $v_H \in \mathcal{X}_H$ with $\max \{ \| \|v_h\| \|, \| \|v_h - v_H\| \| \} \leq \vartheta$, it holds that

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

(A2) Reduction: With $0 < q_{\text{red}} < 1$, it holds that

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

(A3) Reliability: There exists $C_{\text{rel}} > 0$ such that

$$\| \|u^* - u_H^* \| \| \leq C_{\text{rel}} \eta_H(u_H^*).$$

(A4) Discrete reliability: There exists $C_{\text{drel}} > 0$ such that

$$\| \|u_h^* - u_H^* \| \| \leq C_{\text{drel}} \eta_H(\mathcal{T}_H \setminus \mathcal{T}_h, u_H^*).$$

2.7. Idealized adaptive algorithm

In the following, we formulate and analyze an AILFEM algorithm in the spirit of [24], but with an extended stopping criterion in Algorithm A(i.b), i.e.,

$$|\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \leq \lambda^2 \eta_\ell(u_\ell^k)^2 \wedge \| \|u_\ell^k \| \| \leq 2M. \tag{i.b}$$

Clearly, if the stopping criterion from Algorithm A(i.b) holds, then also the simpler stopping criterion $|\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \leq \lambda^2 \eta_\ell(u_\ell^k)^2$ from ([24], Algo. 2) holds.

The proposed algorithm is idealized in the sense that an appropriate parameter $\delta > 0$ is chosen a priori; see Theorem 16 below.

Following [24], the analysis of Algorithm A requires the ordered index set

$$\mathcal{Q} := \{(\ell, k) \in \mathbb{N}_0^2 \mid \text{index pair } (\ell, k) \text{ occurs in Algorithm A and } k < \underline{k}(\ell)\}, \tag{27}$$

where $\underline{k}(\ell) \geq 1$ counts the number of solver steps for each ℓ . The pair $(\ell, \underline{k}(\ell))$ is excluded from \mathcal{Q} , since either $(\ell + 1, 0) \in \mathcal{Q}$ and $u_{\ell+1}^0 = u_\ell^{\underline{k}(\ell)}$ or even $\underline{k}(\ell) := \infty$ if the k -loop does not terminate after finitely many steps. Since Algorithm A is sequential, the index set \mathcal{Q} is lexicographically ordered: For (ℓ, k) and $(\ell', k') \in \mathcal{Q}$, we

¹While ([9], Prop. 15) states stability only for $\mathcal{T}_h \cap \mathcal{T}_H$, the inspection of the proof reveals that indeed arbitrary subsets $\mathcal{U}_H \subseteq \mathcal{T}_h \cap \mathcal{T}_H$ are admissible.

Algorithm A. Idealized AILFEM with energy contraction.

Input: initial triangulation \mathcal{T}_0 , initial guess $u_0^0 := 0$ with $M = \frac{1}{\alpha} \|F - \mathcal{A}0\|_{\mathcal{X}'} < \infty$ according to (5), marking parameters $0 < \theta \leq 1$ and $1 \leq C_{\text{mark}} < \infty$, damping parameter $\delta > 0$, solver parameter $\lambda > 0$.

Loop: For $\ell = 0, 1, 2, \dots$, repeat the following steps (i)–(iv):

- (i) For all $k = 1, 2, 3, \dots$, repeat the following steps (a)–(b):
 - (a) Compute $u_\ell^k := \Phi_\ell(\delta; u_\ell^{k-1})$ and $\eta_\ell(T, u_\ell^k)$ for all $T \in \mathcal{T}_\ell$.
 - (b) Terminate k -loop if $(|\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \leq \lambda^2 \eta_\ell(u_\ell^k)^2 \wedge \|u_\ell^k\| \leq 2M)$.
- (ii) Upon termination of the k -loop, define $\underline{k}(\ell) := k$.
- (iii) Determine a set $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ with up to the multiplicative factor C_{mark} minimal cardinality such that $\theta \eta_\ell(u_\ell^{\underline{k}(\ell)})^2 \leq \sum_{T \in \mathcal{M}_\ell} \eta_\ell(T, u_\ell^{\underline{k}(\ell)})^2$.
- (iv) Generate $\mathcal{T}_{\ell+1} := \text{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ and define $u_{\ell+1}^0 := u_\ell^{\underline{k}(\ell)}$.

write $(\ell', k') < (\ell, k)$ if and only if (ℓ', k') appears earlier in Algorithm A than (ℓ, k) . Given this ordering, we define the *total step counter*

$$|(\ell, k)| := \#\{(\ell', k') \in \mathcal{Q} \mid (\ell', k') < (\ell, k)\} = k + \sum_{\ell'=0}^{\ell-1} \underline{k}(\ell'),$$

which provides the total number of solver steps up to the computation of u_ℓ^k .

Moreover, we define $\overline{\mathcal{Q}} := \mathcal{Q} \cup \{(\ell, \underline{k}(\ell)) \mid \ell \in \mathbb{N}_0 \text{ with } (\ell+1, 0) \in \mathcal{Q}\}$. Note that $\overline{\mathcal{Q}} \subset \mathbb{N}_0 \times \mathbb{N}_0$ is a countably infinite index set such that, for all $(\ell, k) \in \mathbb{N}_0 \times \mathbb{N}_0$,

$$\begin{aligned} (\ell+1, 0) \in \overline{\mathcal{Q}} &\implies (\ell, \underline{k}(\ell)) \in \overline{\mathcal{Q}} \text{ and } \underline{k}(\ell) = \max\{k \in \mathbb{N}_0 \mid (\ell, k) \in \overline{\mathcal{Q}}\}, \\ (\ell, k+1) \in \overline{\mathcal{Q}} &\implies (\ell, k) \in \mathcal{Q}. \end{aligned}$$

With $\underline{\ell} := \sup\{\ell \in \mathbb{N}_0 \mid (\ell, 0) \in \mathcal{Q}\}$, it then follows that either $\underline{\ell} = \infty$ or $\underline{k}(\underline{\ell}) = \infty$. From now on and throughout the paper, we employ the abbreviations $(\ell, \underline{k}) := (\ell, \underline{k}(\ell))$ and $u_\ell^{\underline{k}} := u_\ell^{\underline{k}(\ell)}$.

Corollary 10. *Suppose that \mathcal{A} satisfies (SM), (LIP), and (POT). Suppose the axioms of adaptivity (A1)–(A3). Let $\lambda > 0$ and $0 < \theta \leq 1$ be arbitrary. Then, there exists a choice of the parameter $\delta > 0$ in Algorithm A such that there exist $0 < q_{\mathbb{N}} < 1$ and $0 < q_{\mathcal{E}} < 1$ such that the following properties hold:*

– **nested iteration:** $\|u_\ell^0\| \leq 2M$ for all $(\ell, 0) \in \mathcal{Q}$; (28)

– **boundedness:** $\|u_\ell^k\| \leq 4M$ for all $(\ell, k) \in \mathcal{Q}$; (29)

– **norm contraction:** $\|u_\ell^* - u_\ell^{k+1}\| \leq q_{\mathbb{N}} \|u_\ell^* - u_\ell^k\|$ for all $(\ell, k) \in \mathcal{Q}$; (30)

– **energy contraction:** $\mathcal{E}(u_\ell^{k+1}) - \mathcal{E}(u_\ell^*) \leq q_{\mathcal{E}}^2 [\mathcal{E}(u_\ell^k) - \mathcal{E}(u_\ell^*)]$ for all $(\ell, k) \in \mathcal{Q}$. (31)

Moreover, this guarantees (17) and (18) for all $(\ell, k) \in \mathcal{Q}$ with $q_{\mathbb{N}}[\delta]$ replaced by $q_{\mathbb{N}}$. Furthermore, there exists $k_0 \in \mathbb{N}_0$ such that $\|u_\ell^k\| \leq 2M$ for all $(\ell, k) \in \mathcal{Q}$ with $k \geq k_0$.

Proof. Let $0 < \delta < 2\alpha/L[6M]^2$ be arbitrary but fixed. From Algorithm A and $u_0^0 := 0$, we have that $\|u_\ell^0\| \leq 2M$. Then, $\|u_\ell^* - u_\ell^0\| \leq 3M$. Choose $0 < q_{\mathbb{N}} := q_{\mathbb{N}}[\delta] < 1$ according to Proposition 4, where $\vartheta = 3M$ as well as $0 < q_{\mathcal{E}} := q_{\mathcal{E}}[\delta] < 1$ according to Proposition 8. This proves norm contraction (30) as well as energy contraction (31) for all $(\ell, k) \in \mathcal{Q}$. Furthermore, for all $(\ell, k) \in \mathcal{Q}$, it follows that

$$\|u_\ell^k\| \leq \|u_\ell^*\| + \|u_\ell^* - u_\ell^k\| \stackrel{(18)}{\leq} M + q_{\mathbb{N}}^k \|u_\ell^* - u_\ell^0\| \leq M + q_{\mathbb{N}}^k 3M \leq 4M, \quad (32)$$

which proves boundedness (29). Moreover, (32) together with $0 < q_N < 1$ from (30) proves that there exists $k_0 \in \mathbb{N}_0$, which is independent of ℓ , such that, for all $k \geq k_0$, it holds that

$$\|u_\ell^k\| \stackrel{(32)}{\leq} M + q_N^k 3M \stackrel{!}{\leq} 2M.$$

This shows for $(\ell, 0) \in \mathcal{Q}$ that the stopping criterion $\|u_\ell^k\| \leq 2M$ is met for all $(\ell, k) \in \mathcal{Q}$ with $k \geq k_0$. This concludes the proof. \square

2.8. AILFEM under the assumption of energy contraction (31)

Norm contraction (30) is the critical ingredient in the proof of Corollary 10 — leading to boundedness (Coro. 5), which is key to the proof of energy contraction (31) (cf. (22)). Thus, norm contraction (30) is sufficient for obtaining nested iteration (28), boundedness (29), and energy contraction (31). However, supposing (31) already suffices to obtain uniform constants in the energy norm as the next result shows. Thus, throughout the rest of this paper, we suppose that energy contraction (31) holds for all $(\ell, k) \in \mathcal{Q}$.

Lemma 11. *Suppose that \mathcal{A} satisfies (SM), (LIP), and (POT). Suppose that the choice of $\delta > 0$ guarantees that Algorithm A satisfies energy contraction (31). Then, it holds that*

$$\|u_\ell^k\| \leq M + 3M \frac{L[3M]}{\alpha} =: \frac{\tau}{2} \quad \text{for all } (\ell, k) \in \mathcal{Q}. \quad (33a)$$

Moreover, it holds that

$$\|u_\ell^k - u_\ell^{k'}\| \leq \tau \quad \text{for all } (\ell, k), (\ell, k') \in \mathcal{Q}. \quad (33b)$$

Furthermore, there exists $k_0 \in \mathbb{N}_0$, which is independent of ℓ , such that

$$\|u_\ell^k\| \leq 2M \quad \text{for all } (\ell, k) \in \mathcal{Q} \text{ with } k \geq k_0. \quad (34)$$

Proof. From Algorithm A and $u_0^0 := 0$, we have that $\|u_\ell^0\| \leq 2M$. With $\|u_\ell^*\| \leq M$ from (5), it holds that $\|u_\ell^* - u_\ell^0\| \leq 3M$. For all $(\ell, k) \in \mathcal{Q}$, it follows that

$$\begin{aligned} \|u_\ell^k\| &\leq \|u_\ell^*\| + \|u_\ell^* - u_\ell^k\| \stackrel{(9)}{\leq} M + \left(\frac{2}{\alpha}\right)^{1/2} (\mathcal{E}(u_\ell^k) - \mathcal{E}(u_\ell^*))^{1/2} \\ &\stackrel{(31)}{\leq} M + q_\mathcal{E}^k \left(\frac{2}{\alpha}\right)^{1/2} (\mathcal{E}(u_\ell^0) - \mathcal{E}(u_\ell^*))^{1/2} \stackrel{(9)}{\leq} M + q_\mathcal{E}^k 3M \left(\frac{L[3M]}{\alpha}\right)^{1/2} \end{aligned} \quad (35)$$

$$\stackrel{(31)}{\leq} M + 3M \left(\frac{L[3M]}{\alpha}\right)^{1/2} =: \frac{\tau}{2}. \quad (36)$$

This and the triangle inequality prove (33b). Moreover, inequality (35) together with $0 < q_\mathcal{E} < 1$ from energy contraction (31) proves that there exists $k_0 \in \mathbb{N}_0$, which is independent of ℓ , such that

$$\|u_\ell^k\| \stackrel{(35)}{\leq} M + q_\mathcal{E}^k 3M \left(\frac{L[3M]}{\alpha}\right)^{1/2} \stackrel{!}{\leq} 2M \quad \text{for all } (\ell, k) \in \mathcal{Q} \text{ with } k \geq k_0. \quad (37)$$

This concludes the proof. \square

Remark 12. (i) From Lemma 11, we infer that the stopping criterion can fail only finitely many times due to the energy norm criterion $\|u_\ell^k\| \leq 2M$.

(ii) Under the assumption of energy contraction (31), we note that (33b) shows that τ provides a uniform upper bound for the involved stability and Lipschitz constants $C_{\text{stab}}[\tau]$ and $L[\tau]$, respectively. Indeed, it will become apparent later that stability and local Lipschitz continuity will only be exploited for the differences $\|u_H^k - u_H^{k-1}\|$, $\|u_H^* - u_H^{k+1}\|$, or $\|u^* - u_H^*\|$ in (A1), (9), and (21).

2.9. Main results

Given the Pythagoras identity (8) and energy contraction (31), the first main theorem states full linear convergence of the quasi-error

$$\Delta_\ell^k := \|\|u^* - u_\ell^k\|\| + \eta_\ell(u_\ell^k). \quad (38)$$

Theorem 13 (Full linear convergence). *Suppose that \mathcal{A} satisfies (SM), (LIP), and (POT). Suppose the axioms of adaptivity (A1)–(A3) and orthogonality (8), where \mathcal{X}_H is understood as \mathcal{X}_ℓ for $(\ell, k) \in \mathcal{Q}$. Let $0 < \theta \leq 1$, $1 \leq C_{\text{mark}} \leq \infty$, and $\lambda > 0$. Suppose that the choice of $\delta > 0$ guarantees that Algorithm A satisfies energy contraction (31). Then, there exist $C_{\text{lin}} > 0$ and $0 < q_{\text{lin}} < 1$ such that Algorithm A leads to*

$$\Delta_\ell^k \leq C_{\text{lin}} q_{\text{lin}}^{|\ell, k| - |(\ell', k')|} \Delta_{\ell'}^{k'} \quad \text{for all } (\ell, k), (\ell', k') \in \mathcal{Q} \text{ with } (\ell', k') < (\ell, k). \quad (39)$$

The constants C_{lin} and q_{lin} depend only on M , $L[\tau/2]$, α , $C_{\text{stab}}[\tau]$, q_{red} , C_{rel} , and $q_\mathcal{E}$ as well as on the adaptivity parameters $0 < \theta \leq 1$ and $\lambda > 0$.

The proof of Theorem 13 extends that of ([24], Thm. 4), since the stopping criterion from Algorithm A(i.b) requires further analysis to cover all cases. To ease notation, we introduce the shorthand

$$\mathfrak{d}(v, w)^2 = |\mathcal{E}(v) - \mathcal{E}(w)| \quad \text{for all } v, w \in \mathcal{X}.$$

The following lemma provides the essential step in the proof of Theorem 13.

Lemma 14. *Under the assumptions of Theorem 13, there exist constants $\mu > 0$ and $0 < q_{\text{lin}} < 1$ such that*

$$\Lambda_\ell^k := \mathfrak{d}(u^*, u_\ell^k)^2 + \mu \eta_\ell(u_\ell^k)^2 \quad \text{for all } (\ell, k) \in \mathcal{Q} \quad (40)$$

satisfies the following statements (i) and (ii):

- (i) $\Lambda_\ell^{k+1} \leq q_{\text{lin}}^2 \Lambda_\ell^k$ for all $(\ell, k+1) \in \mathcal{Q}$.
- (ii) $\Lambda_{\ell+1}^0 \leq q_{\text{lin}}^2 \Lambda_\ell^{k-1}$ for all $(\ell+1, 0) \in \mathcal{Q}$.

The constants μ and q_{lin} depend only on M , $L[2M]$, α , $C_{\text{stab}}[\tau]$, q_{red} , C_{rel} , and $q_\mathcal{E}$ as well as on the adaptivity parameters $0 < \theta \leq 1$ and $\lambda > 0$.

Proof. For $k \in \mathbb{N}$ such that $1 \leq k \leq \underline{k}(\ell)$, the stopping criterion of Algorithm A(i.b), i.e.,

$$\mathfrak{d}(u_\ell^{k-1}, u_\ell^k)^2 = |\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \leq \lambda^2 \eta_\ell(u_\ell^k)^2 \wedge \|\|u_\ell^k\|\| \leq 2M, \quad (i.b)$$

comprises four cases. Statement (i) contains the cases **true** \wedge **false**, **false** \wedge **false**, and **false** \wedge **true**. Statement (ii) consists of the remaining case **true** \wedge **true**.

Case 1: Evaluation of (i.b) returns **true** \wedge **false**.

This case investigates (i.b) for $k+1 < \underline{k}(\ell)$. First, we note that

$$\|\|u^* - u_\ell^*\|\|^2 \stackrel{(A3)}{\leq} C_{\text{rel}}^2 \eta_\ell(u_\ell^*)^2 \stackrel{(A1), (33a)}{\leq} 2 C_{\text{rel}}^2 \eta_\ell(u_\ell^{k+1})^2 + 2 C_{\text{rel}}^2 C_{\text{stab}}^2[\tau] \|\|u_\ell^* - u_\ell^{k+1}\|\|^2.$$

Together with (9), this leads us to

$$\mathfrak{d}(u^*, u_\ell^*)^2 \stackrel{(9)}{\leq} \frac{L[2M]}{2} \|\|u^* - u_\ell^*\|\|^2 \stackrel{(9)}{\leq} C_1 \eta_\ell(u_\ell^{k+1})^2 + C_2 \mathfrak{d}(u_\ell^*, u_\ell^{k+1})^2,$$

where we define $C_1 := L[2M]C_{\text{rel}}^2$ and $C_2 := 2\alpha^{-1}L[2M]C_{\text{rel}}^2C_{\text{stab}}^2[\tau]$. For $0 < \varepsilon < 1$, we obtain that

$$\begin{aligned} \mathfrak{d}(u^*, u_\ell^{k+1})^2 &\stackrel{(8)}{=} (1 - \varepsilon) \mathfrak{d}(u^*, u_\ell^*)^2 + \varepsilon \mathfrak{d}(u^*, u_\ell^*)^2 + \mathfrak{d}(u_\ell^*, u_\ell^{k+1})^2 \\ &\leq (1 - \varepsilon) \mathfrak{d}(u^*, u_\ell^*)^2 + \varepsilon C_1 \eta_\ell(u_\ell^{k+1})^2 + (1 + \varepsilon C_2) \mathfrak{d}(u_\ell^*, u_\ell^{k+1})^2 \\ &\stackrel{(31)}{\leq} (1 - \varepsilon) \mathfrak{d}(u^*, u_\ell^*)^2 + \varepsilon C_1 \eta_\ell(u_\ell^{k+1})^2 + (1 + \varepsilon C_2) q_{\mathcal{E}}^2 \mathfrak{d}(u_\ell^*, u_\ell^k)^2. \end{aligned}$$

We use the last inequality for the quasi-error Λ_ℓ^{k+1} to obtain that

$$\begin{aligned} \Lambda_\ell^{k+1} &= \mathfrak{d}(u^*, u_\ell^{k+1})^2 + \mu \eta_\ell(u_\ell^{k+1})^2 \\ &\leq (1 - \varepsilon) \mathfrak{d}(u^*, u_\ell^*)^2 + (\mu + \varepsilon C_1) \eta_\ell(u_\ell^{k+1})^2 + (1 + \varepsilon C_2) q_{\mathcal{E}}^2 \mathfrak{d}(u_\ell^*, u_\ell^k)^2. \end{aligned} \quad (41)$$

We need four auxiliary estimates:

First, since $\|u_\ell^*\| \leq M$ and $\|u_\ell^* - u_0^*\| \leq 2M$ hold independently of ℓ , the axioms (A1)–(A3) and Proposition 2 imply quasi-monotonicity of the estimators, *i.e.*,

$$\eta_\ell(u_\ell^*) \leq C_{\text{mon}} \eta_0(u_0^*) \quad \text{with} \quad C_{\text{mon}} = (2 + 8C_{\text{stab}}[2M]^2(1 + C_{\text{Céa}}^2)C_{\text{rel}}^2)^{1/2}; \quad (42)$$

cf. ([13], Lemma 3.6). With $C_0 := C_{\text{mon}} \max\{1, C_{\text{stab}}[M]M\}$, we infer that

$$\eta_\ell(u_\ell^*) \stackrel{(42)}{\leq} C_{\text{mon}} \eta_0(u_0^*) \stackrel{(A1)}{\leq} C_{\text{mon}} \eta_0(0) + C_{\text{mon}} C_{\text{stab}}[M] \|u_0^*\| \leq C_0(\eta_0(0) + 1). \quad (43)$$

Second, with $C_3 := 2C_0(\eta_0(0) + 1)$ and $C_4 := 4\alpha^{-1}C_{\text{stab}}[\tau]^2 q_{\mathcal{E}}^2$, it holds that

$$\begin{aligned} \eta_\ell(u_\ell^{k+1})^2 &\stackrel{(A1)}{\leq} 2\eta_\ell(u_\ell^*)^2 + 2C_{\text{stab}}[\tau]^2 \|u_\ell^* - u_\ell^{k+1}\|^2 \stackrel{(9)}{\leq} 2\eta_\ell(u_\ell^*)^2 + \frac{4}{\alpha} C_{\text{stab}}[\tau]^2 \mathfrak{d}(u_\ell^*, u_\ell^{k+1})^2 \\ &\stackrel{(31)}{\leq} 2\eta_\ell(u_\ell^*)^2 + \frac{4}{\alpha} C_{\text{stab}}[\tau]^2 q_{\mathcal{E}}^2 \mathfrak{d}(u_\ell^*, u_\ell^k)^2 \stackrel{(43)}{\leq} C_3 + C_4 \mathfrak{d}(u_\ell^*, u_\ell^k)^2. \end{aligned} \quad (44)$$

Third, the error estimator allows for the following estimate with an arbitrary but fixed Young parameter $0 < \gamma < 1$:

$$\begin{aligned} \eta_\ell(u_\ell^{k+1})^2 &\stackrel{(A1)}{\leq} (1 + \gamma) \eta_\ell(u_\ell^k)^2 + (1 + \gamma^{-1}) C_{\text{stab}}[\tau]^2 \|u_\ell^{k+1} - u_\ell^k\|^2 \\ &\leq (1 + \gamma) \eta_\ell(u_\ell^k)^2 + 2(1 + \gamma^{-1}) C_{\text{stab}}[\tau]^2 [\|u_\ell^* - u_\ell^{k+1}\|^2 + \|u_\ell^* - u_\ell^k\|^2] \\ &\stackrel{(9)}{\leq} (1 + \gamma) \eta_\ell(u_\ell^k)^2 + \frac{4}{\alpha} (1 + \gamma^{-1}) C_{\text{stab}}[\tau]^2 [\mathfrak{d}(u_\ell^*, u_\ell^{k+1})^2 + \mathfrak{d}(u_\ell^*, u_\ell^k)^2] \\ &\stackrel{(31)}{\leq} (1 + \gamma) \eta_\ell(u_\ell^k)^2 + C_5 \mathfrak{d}(u_\ell^*, u_\ell^k)^2, \end{aligned} \quad (45)$$

where $C_5 := 4\alpha^{-1}(1 + \gamma^{-1})C_{\text{stab}}[\tau]^2(1 + q_{\mathcal{E}}^2)$.

Fourth, we observe that the case **true** \wedge **false** yields that

$$2M < \|u_\ell^{k+1}\| \leq \|u_\ell^*\| + \|u_\ell^* - u_\ell^{k+1}\| \leq M + \|u_\ell^* - u_\ell^{k+1}\|$$

and hence $M < \|u_\ell^* - u_\ell^{k+1}\|$. With $C_6 := 2\alpha^{-1}M^{-2}q_{\mathcal{E}}^2$, this observation leads us to

$$1 < \frac{\|u_\ell^* - u_\ell^{k+1}\|^2}{M^2} \stackrel{(9)}{\leq} 2\alpha^{-1}M^{-2} \mathfrak{d}(u_\ell^*, u_\ell^{k+1})^2 \stackrel{(31)}{\leq} C_6 \mathfrak{d}(u_\ell^*, u_\ell^k)^2. \quad (46)$$

Recall that $0 < \varepsilon < 1$ and define $0 < \sigma := \frac{\varepsilon + \gamma}{1 + \gamma} < 1$. This choice of σ ensures that

$$(1 - \sigma)(1 + \gamma) = 1 - \varepsilon. \quad (47)$$

We apply these observations to the term $(\mu + \varepsilon C_1) \eta_\ell(u_\ell^{k+1})^2$ of (41) to arrive at

$$\begin{aligned} (\mu + \varepsilon C_1) \eta_\ell(u_\ell^{k+1})^2 &= (1 - \sigma) \mu \eta_\ell(u_\ell^{k+1})^2 + (\sigma \mu + \varepsilon C_1) \eta_\ell(u_\ell^{k+1})^2 \\ &\stackrel{(45)}{\leq} (1 - \sigma)(1 + \gamma) \mu \eta_\ell(u_\ell^k)^2 + (1 - \sigma) \mu C_5 \mathfrak{d}(u_\ell^*, u_\ell^k)^2 + (\sigma \mu + \varepsilon C_1) \eta_\ell(u_\ell^{k+1})^2 \\ &\stackrel{(44)}{\leq} (1 - \sigma)(1 + \gamma) \mu \eta_\ell(u_\ell^k)^2 + (1 - \sigma) \mu C_5 \mathfrak{d}(u_\ell^*, u_\ell^k)^2 + (\sigma \mu + \varepsilon C_1) \left[C_3 + C_4 \mathfrak{d}(u_\ell^*, u_\ell^k)^2 \right] \\ &\stackrel{(46)}{<} (1 - \sigma)(1 + \gamma) \mu \eta_\ell(u_\ell^k)^2 + (1 - \sigma) \mu C_5 \mathfrak{d}(u_\ell^*, u_\ell^k)^2 + (\sigma \mu + \varepsilon C_1) (C_3 C_6 + C_4) \mathfrak{d}(u_\ell^*, u_\ell^k)^2 \\ &\stackrel{(47)}{=} (1 - \varepsilon) \mu \eta_\ell(u_\ell^k)^2 + [(1 - \sigma) C_5 + \sigma C_7] \mu \mathfrak{d}(u_\ell^*, u_\ell^k)^2 + \varepsilon C_1 C_7 \mathfrak{d}(u_\ell^*, u_\ell^k)^2, \end{aligned} \quad (48)$$

where $C_7 := C_3 C_6 + C_4$. Together with (41), we obtain that

$$\begin{aligned} \Lambda_\ell^{k+1} &\stackrel{(41)}{\leq} (1 - \varepsilon) \mathfrak{d}(u^*, u_\ell^*)^2 + (\mu + \varepsilon C_1) \eta_\ell(u_\ell^{k+1})^2 + (1 + \varepsilon C_2) q_\mathcal{E}^2 \mathfrak{d}(u_\ell^*, u_\ell^k)^2 \\ &\stackrel{(48)}{\leq} (1 - \varepsilon) \mathfrak{d}(u^*, u_\ell^*)^2 + (1 - \varepsilon) \mu \eta_\ell(u_\ell^k)^2 \\ &\quad + \{ [(1 - \sigma) C_5 + \sigma C_7] \mu + \varepsilon C_1 C_7 + (1 + \varepsilon C_2) q_\mathcal{E}^2 \} \mathfrak{d}(u_\ell^*, u_\ell^k)^2 \\ &\leq (1 - \varepsilon) \mathfrak{d}(u^*, u_\ell^*)^2 + (1 - \varepsilon) \mu \eta_\ell(u_\ell^k)^2 \\ &\quad + \{ \mu \max \{ C_5, C_7 \} + \varepsilon C_1 C_7 + (1 + \varepsilon C_2) q_\mathcal{E}^2 \} \mathfrak{d}(u_\ell^*, u_\ell^k)^2. \end{aligned}$$

Note that C_1, \dots, C_7 depend only on the problem setting. Provided that

$$\mu \max \{ C_5, C_7 \} + \varepsilon C_1 C_7 + (1 + \varepsilon C_2) q_\mathcal{E}^2 \leq 1 - \varepsilon, \quad (49)$$

we conclude that

$$\begin{aligned} \Lambda_\ell^{k+1} &\leq (1 - \varepsilon) \left[\mathfrak{d}(u^*, u_\ell^*)^2 + \mathfrak{d}(u_\ell^*, u_\ell^k)^2 + \mu \eta_\ell(u_\ell^k)^2 \right] \\ &\stackrel{(8)}{=} (1 - \varepsilon) \left[\mathfrak{d}(u^*, u_\ell^k)^2 + \mu \eta_\ell(u_\ell^k)^2 \right] = (1 - \varepsilon) \Lambda_\ell^k. \end{aligned}$$

Case 2: Evaluation of (i.b) returns false \wedge false or false \wedge true.

These cases follow from the arguments found in ([24], Lemma 10(i)). There, the proof is based in essence on the estimate

$$\eta_\ell(u_\ell^{k+1})^2 < \lambda^{-2} \mathfrak{d}(u_\ell^{k+1}, u_\ell^k)^2 \stackrel{(31)}{\leq} \lambda^{-2} (1 + q_\mathcal{E}^2) \mathfrak{d}(u_\ell^*, u_\ell^k)^2,$$

to obtain an upper bound of the quasi-error Λ_ℓ^{k+1} in terms of the linearization error $\mathfrak{d}(u_\ell^*, u_\ell^k)^2$. With $C_8 := \lambda^{-2} (1 + q_\mathcal{E}^2)$ and provided that

$$(\mu + \varepsilon C_1) C_8 + (1 + \varepsilon C_2) q_\mathcal{E}^2 \leq 1 - \varepsilon, \quad (50)$$

([24], Lemma 10(i)) then proves that

$$\Lambda_\ell^{k+1} \leq (1 - \varepsilon) \Lambda_\ell^k.$$

Up to the final choice of $\mu, \varepsilon > 0$, this concludes the proof of these cases and statement (i).

Case 3: Evaluation of (i.b) returns true \wedge true.

The case **true \wedge true** is analyzed in ([24], Lemma 10(ii)) and is based on the contractivity of the error estimator given that the Dörfler marking is employed.

Define $q_\theta := (1 - (1 - q_{\text{red}}^2) \theta)$ and $C_9 := 4\alpha^{-1}(1 + q_\mathcal{E}^2) C_{\text{stab}}[\tau]^2$. Let $0 < \omega < 1$ be arbitrary. Note that $C_1, C_2, C_9 > 0$ and $0 < q_\theta < 1$ depend only on the problem setting. Provided that

$$\varepsilon C_1 \mu^{-1} + q_\theta (1 + \delta) \leq 1 - \varepsilon \quad \text{and} \quad \varepsilon C_2 + q_\mathcal{E}^2 + \mu q_\theta (1 + \omega^{-1}) C_9 \leq 1 - \varepsilon, \quad (51)$$

we obtain from ([24], Lemma 10(ii)) that

$$\Lambda_{\ell+1}^0 \leq (1 - \varepsilon) \Lambda_\ell^{k-1}.$$

Up to the final choice of $\omega, \mu, \varepsilon > 0$, this concludes the proof of Lemma 14(ii).

Choice of parameters

We proceed as follows:

- (1) Choose $\omega > 0$ such that $(1 + \omega) q_\theta < 1$.
- (2) Choose $\mu > 0$ such that $q_\mathcal{E}^2 + \mu \max\{C_5, C_7\} < 1$, $q_\mathcal{E}^2 + \mu C_8 < 1$, and $q_\mathcal{E}^2 + \mu q_\theta (1 + \omega)^{-1} C_9 < 1$.
- (3) Finally, choose $\varepsilon > 0$ sufficiently small such that (49)–(51) are satisfied.

This concludes the proof of Lemma 14 with $q_{\text{fin}}^2 := (1 - \varepsilon)$. \square

Proof of Theorem 13. According to (9), it holds that $\Delta_\ell^k \simeq (\Lambda_\ell^k)^{1/2}$, where the hidden constants depend only on μ, α , and $L[\tau/2]$. We use (9) for the term $\|u_\ell^* - u_\ell^k\|$, and hence the dependency $L[\tau/2]$ is justified by (36). Then, linear convergence (39) follows from Lemma 14 and induction, since the set \mathcal{Q} is linearly ordered with respect to the total step counter $|(\cdot, \cdot)|$. \square

Remark 15. (i) Provided that energy contraction (31) holds and that the adaptivity parameter $\lambda > 0$ is sufficiently small, the stopping criterion

$$|\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \leq \lambda^2 \eta_\ell (u_\ell^k)^2 \quad (\text{i.b}')$$

from [24] is a viable alternative to the stopping criterion of Algorithm A(i.b). The main difficulty is to ensure nested iteration (28). This relies, in essence, on the estimate

$$\begin{aligned} \frac{\alpha}{2} \|u_\ell^* - u_\ell^k\|^2 &\stackrel{(9)}{\leq} \mathcal{E}(u_\ell^k) - \mathcal{E}(u_\ell^*) \stackrel{(31), (24)}{\leq} \frac{q_\mathcal{E}[\delta]^2}{1 - q_\mathcal{E}[\delta]^2} \left[\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k) \right] \\ &\stackrel{(\text{i.b}')}{\leq} \frac{q_\mathcal{E}[\delta]^2}{1 - q_\mathcal{E}[\delta]^2} \lambda^2 \eta_\ell (u_\ell^k)^2 \stackrel{(\text{A1})}{\leq} 2 \frac{q_\mathcal{E}[\delta]^2}{1 - q_\mathcal{E}[\delta]^2} \lambda^2 \left[\eta_\ell (u_\ell^*)^2 + C_{\text{stab}}[\tau/2]^2 \|u_\ell^* - u_\ell^k\|^2 \right], \end{aligned}$$

where $\|u_\ell^* - u_\ell^k\| \leq \tau/2$ stems from (36). Using a uniform estimate for the error estimator as in (43), the last estimate, and the observation that $\|u_\ell^k\| \leq M + \|u_\ell^* - u_\ell^k\|$ lead us to

$$\|u_{\ell+1}^0\| = \|u_\ell^k\| \leq M + \lambda \frac{r[\delta] C_0 (\eta_0(0) + 1)}{[1 - \lambda^2 r[\delta]^2 C_{\text{stab}}[\tau/2]^2]^{1/2}} \stackrel{!}{\leq} 2M \quad \text{with } r[\delta]^2 := \frac{4}{\alpha} \frac{q_\mathcal{E}[\delta]^2}{1 - q_\mathcal{E}[\delta]^2},$$

where a sufficiently small λ such that $\lambda^2 r[\delta]^2 C_{\text{stab}}[\tau/2]^2 < 1$ is required and where $C_0 := C_{\text{mon}} \max\{1, C_{\text{stab}}[M] M\}$. We see that a sufficiently small $\lambda > 0$ ensures nested iteration (28). In contrast, (i.b) leads to full linear convergence for arbitrary $\lambda > 0$.

- (ii) Theorem 13 proves linear convergence, and hence in particular plain convergence $\Delta_\ell^k \rightarrow 0$ as $|(\ell, k)| \rightarrow \infty$. In Appendix A, it is shown that plain convergence also holds for Algorithm A with the modified stopping criterion

$$\| \|u_\ell^k - u_\ell^{k+1}\| \| \leq \lambda \eta_\ell(u_\ell^k) \quad \wedge \quad \| \|u_\ell^k\| \| < 2M \quad (\text{i.b}'')$$

(instead of Algorithm A(i.b)) in the strongly monotone and locally Lipschitz continuous setting without (POT). Due to the lack of an energy \mathcal{E} , the result relies on norm contraction (30) instead of energy contraction (31).

To formulate our main result on optimal convergence rates, we need some additional notation. For $N \in \mathbb{N}_0$, let $\mathbb{T}_N := \{\mathcal{T} \in \mathbb{T} \mid \#\mathcal{T} - \#\mathcal{T}_0 \leq N\}$ denote the (finite) set of all refinements of \mathcal{T}_0 which have at most N elements more than \mathcal{T}_0 . For $s > 0$, we define

$$\| \|u^*\|_{\mathbb{A}_s} := \sup_{N \in \mathbb{N}_0} \left((N+1)^s \min_{\mathcal{T}_{\text{opt}} \in \mathbb{T}_N} [\| \|u^* - u_{\text{opt}}^*\| \| + \eta_{\text{opt}}(u_{\text{opt}}^*)] \right) \in \mathbb{R}_{\geq 0} \cup \{\infty\}. \quad (52)$$

Here, $u_{\text{opt}}^* \in \mathcal{X}_{\text{opt}}$ denotes the exact Galerkin solution (4) with respect to the optimal mesh \mathcal{T}_{opt} , where optimality is understood with respect to the quasi-error Δ_{opt}^* from (38) (consisting of the energy norm error plus error estimator). In explicit terms, $\| \|u^*\|_{\mathbb{A}_s} < \infty$ means that an algebraic convergence rate $\mathcal{O}(N^{-s})$ for the quasi-error Δ_{opt}^* is possible, if the optimal triangulations are chosen.

The second main theorem states optimal convergence rates of the quasi-error (38) with respect to the number of degrees of freedom. As usual in this context (see *e.g.*, [13]), the result requires that the adaptivity parameters $0 < \theta \leq 1$ and $\lambda > 0$ are sufficiently small. The proof is found in, *e.g.*, ([24], Thm. 8). A careful inspection of the proof reveals that it requires only estimates of the form

$$\text{d}(u_\ell^k, u_\ell^{k-1}) \leq \lambda \eta_\ell(u_\ell^k),$$

as well as linear convergence (39), which are satisfied for Algorithm A. The results from [24] are proven for a uniform Lipschitz and stability constant; in the present setting, this follows from Remark 12(ii).

Theorem 16 (Rate-optimality w.r.t. degrees of freedom). *Suppose that \mathcal{A} satisfies (SM), (LIP), and (POT) as well as the axioms of adaptivity (A1)–(A4). Suppose that the choice of $\delta > 0$ guarantees that Algorithm A satisfies energy contraction (31). Define*

$$\lambda_{\text{opt}} := \frac{1 - q_\mathcal{E}}{q_\mathcal{E} C_{\text{stab}}[\tau]} \left(\frac{\alpha}{2} \right)^{1/2}, \quad (53)$$

with τ from (33). Let $0 < \theta \leq 1$ and $0 < \lambda < \lambda_{\text{opt}} \theta$ such that

$$0 < \theta' := \frac{\theta + \lambda/\lambda_{\text{opt}}}{1 - \lambda/\lambda_{\text{opt}}} < (1 + C_{\text{stab}}[\tau]^2 C_{\text{rel}}^2)^{-1/2}. \quad (54)$$

Let $s > 0$. Then, there exist $c_{\text{opt}}, C_{\text{opt}} > 0$ such that

$$c_{\text{opt}}^{-1} \| \|u^*\|_{\mathbb{A}_s} \leq \sup_{(\ell, k) \in \mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \Delta_\ell^k \leq C_{\text{opt}} \max\{\| \|u^*\|_{\mathbb{A}_s}, \Delta_0^0\}, \quad (55)$$

where $\| \|u^*\|_{\mathbb{A}_s}$ is defined in (52). The constant $c_{\text{opt}} > 0$ depends only on $C_{\text{Céa}} = L[2M]/\alpha$, fine properties of NVB refinement, $C_{\text{stab}}[\tau]$, C_{rel} , $\#\mathcal{T}_0$, and s , and additionally on $\underline{\ell}$ or ℓ_0 , if $\underline{\ell} < \infty$ or $\eta_{\ell_0}(u_{\ell_0}^k) = 0$ for some $(\ell_0, 0) \in \mathcal{Q}$, respectively. The constant $C_{\text{opt}} > 0$ depends only on fine properties of NVB refinement, α , $C_{\text{stab}}[\tau]$, q_{red} , C_{rel} , C_{drel} , $1 - \lambda/\lambda_{\text{opt}}$ (and hence on energy contraction $q_\mathcal{E}$), C_{mark} , C_{lin} , q_{lin} , and s .

To estimate the work necessary to compute $u_\ell^k \in \mathcal{X}_\ell$, we make the following assumptions which are usually satisfied in practice:

- The computation of all indicators $\eta_\ell(T, u_\ell^k)$ for $T \in \mathcal{T}_\ell$ requires $\mathcal{O}(\#\mathcal{T}_\ell)$ operations;
- The marking in Algorithm A(iii) can be performed at linear cost $\mathcal{O}(\#\mathcal{T}_\ell)$ (cf. [38], or the algorithm from [37] providing \mathcal{M}_ℓ with minimal cardinality);
- We have linear cost $\mathcal{O}(\#\mathcal{T}_\ell)$ to generate the new mesh $\mathcal{T}_{\ell+1}$ in Algorithm A(iv).

In addition, we make the following “idealized” assumption, but refer to Remark 17(ii):

- The solutions $u_\ell^k \in \mathcal{X}_\ell$ of the linearized problems in Algorithm A(i.a) can be computed in linear complexity $\mathcal{O}(\#\mathcal{T}_\ell)$.

Since a step $(\ell, k) \in \mathcal{Q}$ of Algorithm A depends on the full history of preceding steps, the total work spent to compute $u_\ell^k \in \mathcal{X}_\ell$ is then of order

$$\text{work}(\ell, k) := \sum_{\substack{(\ell', k') \in \mathcal{Q} \\ (\ell', k') \leq (\ell, k)}} \#\mathcal{T}_{\ell'} \quad \text{for all } (\ell, k) \in \mathcal{Q}. \tag{56}$$

Remark 17. (i) In order to avoid the computation of $\eta_{\ell+1}(u_{\ell+1}^k)$ in each step of the inner loop, *i.e.*, for all k such that $(\ell + 1, k) \in \mathcal{Q}$, one may use $\eta_\ell(u_\ell^k)$ instead. While the proof of linear convergence with the adapted stopping criterion is possible, the proof of optimality remains an open question that goes beyond this work.

(ii) The idealized assumption that the cost of solving the linearized discrete system in Algorithm A(i.a) is linear, can be avoided with an extended algorithm (and refined analysis) in the spirit of [27]. There, an algebraic solve procedure is built into the presented adaptive algorithm as an additional inner loop, taking into account not only discretization and linearization errors but also algebraic errors. In this setting, the “idealized” assumption on the solver would be reduced to the assumption that one solver step has linear cost, which is feasible in the context of FEM. To keep the length of the present manuscript reasonable, we have decided to focus only on the linearization. The details follow along the lines of [27] and are omitted.

The next corollary states the equivalence of rate-optimality with respect to the number of degrees of freedom and rate-optimality with respect to the total work, *i.e.*, the overall computational cost.

Corollary 18 (Rate-optimality w.r.t. computational cost). *Let $(\mathcal{T}_\ell)_{\ell \in \mathbb{N}_0}$ be the sequence generated by Algorithm A. Suppose full linear convergence (39) with respect to the quasi-error Δ_ℓ^k from (38). Then, for all $s > 0$, it holds that*

$$C_{\text{rate}} := \sup_{(\ell, k) \in \mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \Delta_\ell^k \leq \sup_{(\ell, k) \in \mathcal{Q}} \text{work}(\ell, k)^s \Delta_\ell^k \leq \frac{(\#\mathcal{T}_0)^s C_{\text{lin}}}{(1 - q_{\text{lin}}^{1/s})^s} C_{\text{rate}}. \tag{57}$$

Consequently, rate-optimality with respect to the number of elements (55) yields that

$$c_{\text{opt}}^{-1} \|u^*\|_{\mathbb{A}_s} \leq \sup_{(\ell, k) \in \mathcal{Q}} \text{work}(\ell, k)^s \Delta_\ell^k \leq C_{\text{opt}} \frac{(\#\mathcal{T}_0)^s C_{\text{lin}}}{(1 - q_{\text{lin}}^{1/s})^s} \max\{\|u^*\|_{\mathbb{A}_s}, \Delta_0^0\}. \tag{58}$$

Proof. The first inequality in (57) is obvious. To obtain the upper bound, let $(\ell, k) \in \mathcal{Q}$. Elementary calculus (see ([11], Lemma 22)) proves that

$$\#\mathcal{T}_H \leq \#\mathcal{T}_0 (\#\mathcal{T}_H - \#\mathcal{T}_0 + 1) \quad \text{for all } \mathcal{T}_H \in \mathbb{T}.$$

Moreover, linear convergence (39) and the geometric series lead us to

$$\sum_{\substack{(\ell',k') \in \mathcal{Q} \\ (\ell',k') \leq (\ell,k)}} (\Delta_{\ell'}^{k'})^{-1/s} \stackrel{(39)}{\leq} C_{\text{lin}}^{1/s} (\Delta_{\ell}^k)^{-1/s} \sum_{\substack{(\ell',k') \in \mathcal{Q} \\ (\ell',k') \leq (\ell,k)}} (q_{\text{lin}}^{1/s})^{|\ell,k| - |(\ell',k')|} \leq \frac{C_{\text{lin}}^{1/s} (\Delta_{\ell}^k)^{-1/s}}{1 - q_{\text{lin}}^{1/s}}.$$

Combining the last two inequalities, we obtain that

$$\begin{aligned} \sum_{\substack{(\ell',k') \in \mathcal{Q} \\ (\ell',k') \leq (\ell,k)}} \#\mathcal{T}_{\ell'} &\leq (\#\mathcal{T}_0) \sum_{\substack{(\ell',k') \in \mathcal{Q} \\ (\ell',k') \leq (\ell,k)}} (\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1) \leq (\#\mathcal{T}_0) C_{\text{rate}}^{1/s} \sum_{\substack{(\ell',k') \in \mathcal{Q} \\ (\ell',k') \leq (\ell,k)}} (\Delta_{\ell'}^{k'})^{-1/s} \\ &\leq (\#\mathcal{T}_0) \frac{C_{\text{lin}}^{1/s}}{1 - q_{\text{lin}}^{1/s}} (\Delta_{\ell}^k)^{-1/s} C_{\text{rate}}^{1/s}. \end{aligned}$$

Rearranging this estimate, we obtain the upper bound in (57). □

3. SEMILINEAR MODEL PROBLEM

3.1. Model problem

For $d \in \{1, 2, 3\}$, let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Given $f \in L^2(\Omega)$ and $\mathbf{f} \in [L^2(\Omega)]^d$, we aim to approximate the weak solution $u^* \in \mathcal{X} := H_0^1(\Omega)$ of the semilinear elliptic PDE

$$-\operatorname{div}(\mathbf{A}\nabla u^*) + b(u^*) = f - \operatorname{div} \mathbf{f} \quad \text{in } \Omega \quad \text{subject to } u^* = 0 \text{ on } \partial\Omega. \tag{59}$$

While the precise assumptions on the coefficients $\mathbf{A}: \Omega \rightarrow \mathbb{R}_{\text{sym}}^{d \times d}$ and $b: \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ are given in Sections 3.3 and 3.4, we note that, here and below, we abbreviate $\mathbf{A}\nabla u^* \equiv \mathbf{A}(\cdot)\nabla u^*(\cdot): \Omega \rightarrow \mathbb{R}^d$ and $b(u^*) \equiv b(\cdot, u^*(\cdot)): \Omega \rightarrow \mathbb{R}$.

Let $\langle \cdot, \cdot \rangle_{\Omega}$ denote the $L^2(\Omega)$ -scalar product $\langle v, w \rangle_{\Omega} := \int_{\Omega} vw \, dx$ and let $\langle\langle v, w \rangle\rangle := \langle \mathbf{A}\nabla v, \nabla w \rangle_{\Omega}$ be the \mathbf{A} -induced energy scalar product on $H_0^1(\Omega)$. Then, the weak formulation of (59) reads as follows: Find $u^* \in H_0^1(\Omega)$ such that

$$\langle \mathbf{A}, u^* \rangle v := \langle\langle u^*, v \rangle\rangle + \langle b(u^*), v \rangle_{\Omega} = \langle f, v \rangle_{\Omega} + \langle \mathbf{f}, \nabla v \rangle_{\Omega} =: \langle F, v \rangle \quad \text{for all } v \in H_0^1(\Omega). \tag{60}$$

Existence and uniqueness of the solution $u^* \in H_0^1(\Omega)$ of (60) follow from the Browder–Minty theorem on monotone operators (see Sect. 3.6 for details).

Based on conforming triangulations \mathcal{T}_H of Ω and fixed polynomial degree $m \in \mathbb{N}$, let $\mathcal{X}_H := \{v_H \in H_0^1(\Omega) \mid \forall T \in \mathcal{T}_H: v_H|_T \text{ is a polynomial of degree } \leq m\}$. Then, the FEM discretization of (60) reads: find $u_H^* \in \mathcal{X}_H$ such that

$$\langle\langle u_H^*, v_H \rangle\rangle + \langle b(u_H^*), v_H \rangle_{\Omega} = \langle F, v_H \rangle \quad \text{for all } v_H \in \mathcal{X}_H. \tag{61}$$

The FEM solution u_H^* approximates the sought exact solution u^* .

3.2. General notation

For $1 \leq p \leq \infty$, let $1 \leq p' \leq \infty$ be the conjugate Hölder index which ensures that $\|\phi\psi\|_{L^1(\Omega)} \leq \|\phi\|_{L^p(\Omega)}\|\psi\|_{L^{p'}(\Omega)}$ for $\phi \in L^p(\Omega)$ and $\psi \in L^{p'}(\Omega)$, *i.e.*, $1/p + 1/p' = 1$ with the convention that $p' = 1$ for $p = \infty$ and *vice versa*. Moreover, for $1 \leq p < d$, let $1 \leq p^* := dp/(d - p) < \infty$ denote the critical Sobolev exponent of p in dimension $d \in \mathbb{N}$. We recall the Gagliardo–Nirenberg–Sobolev inequality (see *e.g.*, ([22], Thm. 16.6))

$$\|v\|_{L^r(\Omega)} \leq C_{\text{GNS}} \|\nabla v\|_{L^p(\Omega)} \quad \text{for all } v \in W_0^{1,p}(\Omega) \tag{62}$$

with a constant $C_{\text{GNS}} = C_{\text{GNS}}(|\Omega|, d, p, r)$. With $\mathcal{X} = H_0^1(\Omega)$, we restrict to $p = 2$. If $d \in \{1, 2\}$, (62) holds for any $1 \leq r < \infty$. If $d = 3$, (62) holds for all $1 \leq r \leq p^* = 6$, where $r = p^*$ is the largest possible exponent such that the embedding $W^{1,p}(\Omega) \hookrightarrow L^r(\Omega)$ is continuous.

3.3. Assumptions on diffusion coefficient

The diffusion coefficient $\mathbf{A}: \Omega \rightarrow \mathbb{R}_{\text{sym}}^{d \times d}$ satisfies the following standard assumptions:

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

$$0 < \mu_0 := \inf_{x \in \Omega} \lambda_{\min}(\mathbf{A}(x)) \leq \sup_{x \in \Omega} \lambda_{\max}(\mathbf{A}(x)) =: \mu_1 < \infty.$$

In particular, the \mathbf{A} -induced energy scalar product $\langle\langle v, w \rangle\rangle := \langle \mathbf{A} \nabla v, \nabla w \rangle_\Omega$ induces an equivalent norm $\|v\| := \langle\langle v, v \rangle\rangle^{1/2}$ on $H_0^1(\Omega)$.

To guarantee later that the residual *a posteriori* error estimators are well-defined, we additionally require that $A|_T \in [W^{1,\infty}(T)]^{d \times d}$ for all $T \in \mathcal{T}_0$, where \mathcal{T}_0 is the initial triangulation of the adaptive algorithm.

3.4. Assumptions on the nonlinear reaction coefficient

The nonlinearity $b: \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ satisfies the following assumptions, which follow ([6], (A1)–(A3)):

(CAR) $b: \Omega \times \mathbb{R} \rightarrow \mathbb{R}$ is a *Carathéodory* function, *i.e.*, for all $n \in \mathbb{N}_0$, the n th derivative $b^{(n)} := \partial_\xi^n b$ 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.*, $b'(x, \xi) := b^{(1)}(x, \xi) \geq 0$ for all $x \in \Omega$ and $\xi \in \mathbb{R}$. Without loss of generality², we assume 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.*, ([22], Chap. III, (12)) or ([6], (A4)):

(GC) If $d \in \{1, 2\}$, there exists $N \in \mathbb{N}$ such that $1 \leq N < \infty$. For $d = 3$, there exists $N \in \mathbb{N}$ such that $1 \leq N \leq 5$. Suppose that, for $d \in \{1, 2, 3\}$, there exists $R > 0$ such that

$$|b^{(N)}(x, \xi)| \leq R \quad \text{for a.e. } x \in \Omega \text{ and all } \xi \in \mathbb{R}. \tag{63}$$

While (GC) turns out to be sufficient for plain convergence of the later AILFEM algorithm, we require the following stronger assumption for linear convergence and optimal convergence rates.

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

Remark 19. (i) Let $v, w \in H_0^1(\Omega)$. To establish continuity of $(v, w) \mapsto \langle b(v), w \rangle_\Omega$, we apply the Hölder inequality with Hölder conjugates $1 \leq s, s' \leq \infty$ to obtain that

$$|\langle b(v), w \rangle_\Omega| \leq \|b(v)\|_{L^{s'}(\Omega)} \|w\|_{L^s(\Omega)}. \tag{64}$$

The smoothness assumption (CAR) admits a Taylor expansion for b . Together with $b(0) = 0$ from (MON), this yields that

$$b(v) \stackrel{\text{(MON)}}{=} \sum_{n=1}^{N-1} \frac{b^{(n)}(0)}{n!} v^n + \left(\int_0^1 \frac{(1-\xi)^{N-1}}{(N-1)!} b^{(N)}(\xi v) \, d\xi \right) v^N. \tag{65}$$

With $\|v^n\|_{L^{s'}(\Omega)} = \|v\|_{L^{ns'}(\Omega)}^n$, it follows that

$$\|b(v)\|_{L^{s'}(\Omega)} \stackrel{\text{(GC)}}{\lesssim} \sum_{n=1}^{N-1} \|v^n\|_{L^{s'}(\Omega)} + \|v^N\|_{L^{s'}(\Omega)} = \sum_{n=1}^{N-1} \|v\|_{L^{ns'}(\Omega)}^n + \|v\|_{L^{Ns'}(\Omega)}^N$$

²Otherwise, consider $\tilde{b}(v) := b(v) - b(0)$ and $\tilde{f} := f - b(0)$ instead.

$$\lesssim \sum_{n=1}^N \|v\|_{L^{Ns'}}^n \leq N \max\left\{1, \|v\|_{L^{Ns'}}^{N-1}\right\} \|v\|_{L^{Ns'}(\Omega)},$$

where the second to last estimate exploits the L^p -space inclusions for bounded Ω . To guarantee that $|\langle b(v), w \rangle_\Omega| < \infty$, condition (GC) should ensure that the embedding

$$H_0^1(\Omega) \hookrightarrow L^r(\Omega) \quad \text{is continuous} \quad \text{for} \quad r = s \quad \text{and} \quad r = Ns'. \tag{66}$$

If $d \in \{1, 2\}$, (66) follows if $1 \leq r < \infty$ and hence arbitrary $1 < s < \infty$ and $N \in \mathbb{N}$. If $d = 3$, $r = s = 2^* = 6$ is the maximal index in (66). Hence, it follows that $N \leq 2^*/s' = 2^*/2^{*'} = 2^* - 1 = 5$. Altogether, we conclude continuity of $(v, w) \mapsto \langle b(v), w \rangle_\Omega$ for all $N \in \mathbb{N}$ if $d \in \{1, 2\}$, and $N \leq 5$ if $d = 3$.

(ii) The definition of ([9], (GC)) uses

$$|b^{(n)}(x, \xi)| \leq R(1 + |\xi|^{N-n}) \quad \text{for all } x \in \Omega, \text{ all } \xi \in \mathbb{R}, \text{ and all } 0 \leq n \leq N$$

instead of (63). However, the following observation replaces the estimates for all $b^{(n)}$ with $0 \leq n < N$. Due to the smoothness assumption (CAR), we may apply a Taylor expansion for an admissible σ such that $(N - n)\sigma < \infty$ if $d = 1, 2$ and $(N - n)\sigma \leq 6$ if $d = 3$. Together with $\|v^n\|_{L^\sigma(\Omega)} = \|v\|_{L^{n\sigma}(\Omega)}^n$, this leads us to

$$\begin{aligned} \|b^{(n)}(v)\|_{L^\sigma(\Omega)} &\leq \sum_{j=n}^{N-1} \frac{b^{(j)}(0)}{(j-n)!} \|v^{j-n}\|_{L^\sigma(\Omega)} + \left(\int_0^1 \frac{(1-\xi)^{N-1-n}}{(N-1-n)!} b^{(N)}(\xi v) \, d\xi \right) \|v^{N-n}\|_{L^\sigma(\Omega)} \\ &\stackrel{\text{(GC)}}{\lesssim} \sum_{j=n}^{N-1} \|v\|_{L^{(j-n)\sigma}(\Omega)}^{j-n} + \|v\|_{L^{(N-n)\sigma}(\Omega)}^{N-n} \lesssim \sum_{j=n}^N \|v\|_{L^{(N-n)\sigma}(\Omega)}^{j-n} \\ &\leq (N-n) \left(1 + \|v\|_{L^{(N-n)\sigma}(\Omega)}^{N-n} \right) \lesssim (N-n) (1 + \|v\|^{N-n}), \end{aligned} \tag{67}$$

where the additive constant stems from the fact that $b^{(n)}(0) \neq 0$ in general (in contrast to the reasoning in (i)). This estimate plays a central role in proving the local Lipschitz continuity of b and thus of the overall semilinear model problem; see Lemma 20 below and the discussion thereafter.

3.5. Assumptions on the right-hand sides

For $d = 1$, the exact solution u^* from (60) below satisfies an L^∞ -bound, since H^1 -functions are absolutely continuous. For $d \in \{2, 3\}$, we need the following assumption:

(RHS) We suppose that the right-hand side fulfils that

$$\mathbf{f} \in [L^p(\Omega)]^d \text{ for some } p > d \geq 2 \quad \text{and} \quad \mathbf{f} \in L^q(\Omega) \text{ where } 1/q := 1/p + 1/d.$$

To guarantee later that the residual a *posteriori* error estimator from (74) is well-defined, we additionally require that $\mathbf{f}|_T \in H(\text{div}, T)$ and $\mathbf{f}|_T \cdot \mathbf{n} \in L^2(\partial T)$ for all $T \in \mathcal{T}_0$, where \mathcal{T}_0 is the initial triangulation of the adaptive algorithm.

3.6. Well-posedness and applicability of abstract framework

Let $v, w \in H_0^1(\Omega)$. We consider the operator \mathcal{A} , where $H^{-1}(\Omega) := H_0^1(\Omega)'$ is used to denote the dual space of $H_0^1(\Omega)$,

$$\mathcal{A}: H_0^1(\Omega) \rightarrow H^{-1}(\Omega), \quad \mathcal{A}w := \langle w, \cdot \rangle + \langle b(w), \cdot \rangle_\Omega. \tag{68}$$

Since $b'(x, \zeta) \geq 0$ according to (MON), this implies that

$$(b(x, \xi_2) - b(x, \xi_1))(\xi_2 - \xi_1) \geq 0 \quad \text{for all } x \in \Omega \quad \text{and } \xi_1, \xi_2 \in \mathbb{R}.$$

Together with (ELL) and for $v, w \in H_0^1(\Omega)$, we thus see that

$$\langle \mathcal{A}w - \mathcal{A}v, w - v \rangle = \langle\langle w - v, w - v \rangle\rangle + \langle b(w) - b(v), w - v \rangle_\Omega \geq \|w - v\|^2. \tag{69}$$

This proves that \mathcal{A} is strongly monotone with $\alpha = 1$ with respect to the energy norm $\|\cdot\|$. The following lemma is crucial to prove local Lipschitz continuity.

Lemma 20. *Suppose (RHS), (ELL), (CAR), (MON), and (GC). Let $\vartheta > 0$ and let $v, w \in H_0^1(\Omega)$ with $\max\{\|w\|, \|w - v\|\} \leq \vartheta < \infty$. Then, it holds that*

$$\langle b(w) - b(v), z \rangle_\Omega \leq \tilde{L}[\vartheta] \|w - v\| \|z\| \quad \text{for all } z \in H_0^1(\Omega) \tag{70}$$

with $\tilde{L}[\vartheta] = \tilde{L}(|\Omega|, d, \vartheta, N, R, \mu_0)$.

Proof. Due to the smoothness assumption (CAR), we may consider the Taylor expansion

$$b(v) = \sum_{n=0}^{N-1} b^{(n)}(w) \frac{(v-w)^n}{n!} + \frac{(v-w)^N}{(N-1)!} \int_0^1 (1-\xi)^{N-1} b^{(N)}(w + (v-w)\xi) d\xi. \tag{71}$$

In order to apply the generalized Hölder inequality for three terms $\phi, \varphi, \psi \in H_0^1(\Omega)$

$$\langle \phi \varphi, \psi \rangle_\Omega \leq \|\phi\|_{L^{t''}(\Omega)} \|\varphi\|_{L^t(\Omega)} \|\psi\|_{L^t(\Omega)},$$

where $1 = 1/t + 1/t + 1/t''$, we choose $t > 2$ arbitrarily for $d \in \{1, 2\}$ and $t = 6$ and hence $t'' = 3/2$ for $d = 3$. In both cases, we see that

$$\begin{aligned} \langle b(w) - b(v), z \rangle_\Omega &\leq \sum_{n=1}^{N-1} \frac{1}{n!} \|b^{(n)}(w)(w-v)^{n-1}\|_{L^{t''}(\Omega)} \|w-v\|_{L^t(\Omega)} \|z\|_{L^t(\Omega)} \\ &\quad + \left\| \frac{(w-v)^{N-1}}{(N-1)!} \int_0^1 (1-\xi)^{N-1} b^{(N)}(w + (v-w)\xi) d\xi \right\|_{L^{t''}(\Omega)} \|w-v\|_{L^t(\Omega)} \|z\|_{L^t(\Omega)} \\ &\stackrel{\text{(GC)}}{\lesssim} \left(\sum_{n=1}^{N-1} \|b^{(n)}(w)(w-v)^{n-1}\|_{L^{t''}(\Omega)} + \|w-v\|_{L^{(N-1)t''}(\Omega)}^{N-1} \right) \|w-v\| \|z\|, \end{aligned}$$

where the hidden constant depends on R from (GC). Since $H_0^1(\Omega) \hookrightarrow L^{(N-1)t''}(\Omega)$ for $d \in \{1, 2, 3\}$, it remains to prove that

$$\|b^{(n)}(w)(w-v)^{n-1}\|_{L^{t''}(\Omega)} \leq C[\vartheta] \quad \text{for all } n = 1, \dots, N-1. \tag{72}$$

To this end, choose $t_1 = (N-1)t''/(N-n)$ and $t_2 = (N-1)t''/(n-1)$ and note that

$$\frac{1}{t''} = \frac{1}{t''} \left(\frac{N-n}{N-1} + \frac{n-1}{N-1} \right) = \frac{1}{t_1} + \frac{1}{t_2}.$$

Using the Hölder inequality, we arrive at

$$\|b^{(n)}(w)(w-v)^{n-1}\|_{L^{t''}(\Omega)} \leq \|b^{(n)}(w)\|_{L^{t_1}(\Omega)} \|(w-v)^{n-1}\|_{L^{t_2}(\Omega)}.$$

Since $\|\varphi^j\|_{L^\sigma(\Omega)} = \|\varphi\|_{L^{j\sigma}(\Omega)}^j$ and $(N-1)t'' < \infty$ if $d \in \{1, 2\}$ and $(N-1)t'' \leq 6$ if $d = 3$ guarantee admissibility as in Remark 19(ii), we apply the Sobolev embedding to obtain that

$$\|b^{(n)}(w)\|_{L^{t_1}(\Omega)} \stackrel{(67)}{\lesssim} 1 + \|w\|_{L^{(N-n)t_1}(\Omega)}^{N-n} = 1 + \|w\|_{L^{(N-1)t''}(\Omega)}^{N-n} \lesssim 1 + \|w\|^{N-n}$$

and

$$\|(w-v)^{n-1}\|_{L^{t_2}(\Omega)} = \|w-v\|_{L^{(n-1)t_2}(\Omega)}^{n-1} = \|w-v\|_{L^{(N-1)t''}(\Omega)}^{n-1} \lesssim \|w-v\|^{n-1}.$$

The last estimates together with the assumptions $\|w-v\| \leq \vartheta$ and $\|w\| \leq \vartheta$ conclude the proof with hidden constant $\tilde{L}[\vartheta] = \tilde{L}(|\Omega|, d, \vartheta, N, R, \mu_0) > 0$. □

To see the local Lipschitz continuity of \mathcal{A} , let $v, w, \psi \in H_0^1(\Omega)$ and observe that

$$\langle \mathcal{A}w - \mathcal{A}v, \psi \rangle = \langle w - v, \psi \rangle + \langle b(w) - b(v), \psi \rangle_\Omega \stackrel{(70)}{\leq} (1 + \tilde{L}[\vartheta]) \|w - v\| \|\psi\|,$$

provided that $\|w\| \leq \vartheta$ and $\|w - v\| \leq \vartheta$. This shows that \mathcal{A} is locally Lipschitz continuous with Lipschitz constant $L[\vartheta] := 1 + \tilde{L}[\vartheta]$. Hence, \mathcal{A} fits into the abstract setting of Section 2.

Furthermore, following [2], we note that the energy for the semilinear model problem (59) of Section 3 for $v \in H_0^1(\Omega)$ is given by

$$\mathcal{E}(v) = \frac{1}{2} \int_\Omega |\mathbf{A}^{1/2} \nabla v|^2 \, dx + \int_\Omega \int_0^{v(x)} b(s) \, ds \, dx - \int_\Omega f v \, dx - \int_\Omega \mathbf{f} \cdot \nabla v \, dx. \tag{73}$$

To see that the second integral is well-defined, note that the integration of the Taylor expansion (65) gives rise to a term s^{N+1} evaluated at $s = v(x)$ and $s = 0$. Its integrability $\|v^{N+1}\|_{L^1(\Omega)} = \|v\|_{L^{(N+1)}(\Omega)}^{N+1} < \infty$ is ensured by (CGC).

3.7. Residual error estimators

For $\mathcal{T}_H \in \mathbb{T}$ and $v_H \in \mathcal{X}_H$, the local contributions of the standard residual error estimator for the semilinear model problem (60) read

$$\begin{aligned} \eta_H(T, v_H)^2 &:= h_T^2 \|f + \operatorname{div}(\mathbf{A} \nabla v_H - \mathbf{f}) - b(v_H)\|_{L^2(T)}^2 \\ &\quad + h_T \|[(\mathbf{A} \nabla v_H - \mathbf{f}) \cdot \mathbf{n}]\|_{L^2(\partial T \cap \Omega)}^2, \end{aligned} \tag{74}$$

where $[\![\cdot]\!]$ denotes the jump across edges (for $d = 2$) resp. faces (for $d = 3$) and \mathbf{n} denotes the outer unit normal vector. For $d = 1$, these jumps vanish, i.e., $[\![\cdot]\!] = 0$. ([9], Prop. 15) proves the axioms of adaptivity (A1)–(A4) for the present setting.

Proposition 21 ([9], Prop. 15). *Suppose (RHS), (ELL), (CAR), (MON), and (CGC). Then, the residual error estimator from (74) satisfies (A1)–(A4) from Section 2.6. The constant C_{rel} depends only on d, μ_0 , and uniform shape regularity of the meshes $\mathcal{T}_H \in \mathbb{T}$. The constant C_{drel} depends, in addition, on the polynomial degree m , and $C_{\text{stab}}[\vartheta]$ depends furthermore on $|\Omega|, \vartheta, n, R$, and \mathbf{A} .*

4. PRACTICAL ALGORITHM

For the semilinear problem (59) of Section 3, it holds that $\alpha = 1$ according to (69). The optimal damping parameter $\delta > 0$ as well as $L[6M]$ are unknown in practice. In this section, we present a practical algorithm which is formulated with computable quantities only.

Algorithm B. Practical AILFEM.

Input: initial triangulation \mathcal{T}_0 , initial guess $u_0^0 := 0$ and $M = \|F - \mathcal{A}0\|_{\mathcal{X}'} < \infty$ according to (5), marking parameters $0 < \theta \leq 1$ and $C_{\text{mark}} \geq 1$, solver termination parameter $\lambda > 0$, and solver parameters $L_0 := 1$ and $\beta := \sqrt{2}$.

Loop: For $\ell = 0, 1, 2, \dots$, repeat the following steps (i)–(v):

- (i) Calculate $\delta_\ell \leftarrow 1/L_\ell$ and $q_\ell^2 \leftarrow 1 - \delta_\ell^2$.
 - (ii) For all $k = 1, 2, \dots$, repeat the following steps (a)–(c):
 - (a) Compute $u_\ell^k := \Phi_\ell(\delta_\ell; u_\ell^{k-1})$ and $\eta_\ell(T, u_\ell^k)$ for all $T \in \mathcal{T}_\ell$.
 - (b) Terminate k -loop if $\left(|\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \leq \lambda^2 \eta_\ell(u_\ell^k)^2 \wedge \|u_\ell^k\| \leq 2M \right)$
 - (c) If $(\mathcal{E}(u_\ell^k) > q_\ell^2 \mathcal{E}(u_\ell^{k-1}))$, then
 - (c1) Discard the computed u_ℓ^k and set $k \leftarrow k - 1$.
 - (c2) Increase $L_\ell \leftarrow \beta L_\ell$.
 - (c3) Update $\delta_\ell \leftarrow 1/L_\ell$ and $q_\ell^2 \leftarrow 1 - \delta_\ell^2$.
 - (iii) Upon termination of the k -loop, define $\underline{k}(\ell) := k$.
 - (iv) Determine $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ with $\theta \eta_\ell(u_\ell^{\underline{k}(\ell)})^2 \leq \sum_{T \in \mathcal{M}_\ell} \eta_\ell(T, u_\ell^{\underline{k}(\ell)})^2$.
 - (v) Generate $\mathcal{T}_{\ell+1} := \text{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ and define $u_{\ell+1}^0 := u_\ell^{\underline{k}(\ell)}$.
-

4.1. AILFEM and contraction of damped Zarantonello iteration

Instead of adaptively choosing $\delta > 0$, we adapt the local Lipschitz constant L . Since $\alpha = 1$, this already determines the optimal choice $\delta = 1/L$ and $q[\delta]^2 = 1 - \delta^2$; see Remark 9.

Remark 22. The motivation of the criterion in Algorithm B(ii.c) is based on the equivalence

$$\mathcal{E}(u_\ell^k) - \mathcal{E}(u_\ell^*) \leq q_\ell^2 [\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^*)] \iff \mathcal{E}(u_\ell^k) - q_\ell^2 \mathcal{E}(u_\ell^{k-1}) \leq (1 - q_\ell^2) \mathcal{E}(u_\ell^*). \tag{75}$$

The energy minimization property from Lemma 3 and $b(0) = 0$ from (MON) show that $\mathcal{E}(u_\ell^*) \leq \mathcal{E}(0) = 0$; cf. (73). As a necessary criterion for energy contraction (31), we thus obtain $\mathcal{E}(u_\ell^k) \leq q_\ell^2 \mathcal{E}(u_\ell^{k-1})$, which is enforced by Algorithm B(ii.c).

Remark 23. Note that $\lambda > 0$ is arbitrary but fixed and remains unchanged throughout the algorithm. In the numerical experiments below, the particular choice $\lambda = 0.1$ is motivated by the following heuristic argument: the estimator $\eta_\ell(u_\ell^*)$ and hence approximately $\eta_\ell(u_\ell^{\underline{k}})$ controls the discretization error, while $\|u_\ell^* - u_\ell^{\underline{k}}\|^2 \stackrel{(9)}{\simeq} \mathcal{E}(u_\ell^{\underline{k}}) - \mathcal{E}(u_\ell^*) \stackrel{(24)}{\lesssim} \mathcal{E}(u_\ell^{\underline{k}-1}) - \mathcal{E}(u_\ell^{\underline{k}}) \stackrel{(21)}{\simeq} \|u_\ell^{\underline{k}} - u_\ell^{\underline{k}-1}\|^2$ controls the linearization error – at least if δ_H is sufficiently small. Hence, $\mathcal{E}(u_\ell^{\underline{k}-1}) - \mathcal{E}(u_\ell^{\underline{k}}) \leq 0.1^2 \eta_\ell(u_\ell^{\underline{k}})^2$ heuristically aims at limiting the linearization error to be at most 10% of the current discretization error.

The next result states that Algorithm B(ii.c) will not lead to an infinite loop.

Proposition 24. *Suppose that \mathcal{A} satisfies (SM), (LIP), and (POT). Let $u_H^0 \in \mathcal{X}_H$ with $\|u_H^0\| \leq 2M$. Set $L_0, L_H \leftarrow 1$ and define $\beta := \sqrt{2}$. Compute $\delta_H = 1/L_H$ and $q_H^2 = 1 - \delta_H^2$. Starting with $k \leftarrow 1$ and $u_H^1 := \Phi_H(\delta_H; u_H^0) \in \mathcal{X}_H$, we proceed as follows:*

- Given $u_H^k \in \mathcal{X}_H$ for $k \geq 1$, compute $u_H^{k+1} := \Phi_H(\delta_H; u_H^k) \in \mathcal{X}_H$ and check if

$$\mathcal{E}(u_H^{k+1}) \leq q_H^2 \mathcal{E}(u_H^k). \tag{76}$$

- If (76) holds, then increase $k \leftarrow k + 1$.
- If (76) fails, then increase $L_H \leftarrow \beta L_H$ and update $\delta_H \leftarrow 1/L_H$ and $q_H^2 \leftarrow 1 - \delta_H^2$. Discard the computed u_H^{k+1} .

Then, the condition (76) fails only finitely often so that this simple algorithm defines the sequence of iterates $(u_H^k)_{k \in \mathbb{N}_0}$.

Proof. Step 1. Given the initial $L_0 = 1$, there exists a minimal number $j \in \mathbb{N}_0$ such that

$$\frac{L[6M]^2}{2\alpha} < \beta^j L_0 = L_H(j) \quad \text{and thus} \quad \delta_H := \delta_H(j) = \frac{1}{\beta^j L_0} < \frac{2\alpha}{L[6M]^2}.$$

Define $q_H[\delta_H(k)]^2 := 1 - \delta_H(k)^2$. Recall $q_\varepsilon[\delta_H]$ from (23b) and observe that

$$q_\varepsilon[\delta_H(k)]^2 = 1 - \left(1 - \frac{\delta_H(k)L[6M]}{2}\right) \frac{2\delta_H(k)\alpha^2}{L[3M]} \simeq 1 - \delta_H(k) + \delta_H(k)^2 \quad \text{for } \delta_H(k) \rightarrow 0.$$

Since $\delta_H(k) \rightarrow 0$ for $k \rightarrow \infty$, there exists a minimal number $k_0 \in \mathbb{N}$ with $k_0 \geq j$ such that

$$q_\varepsilon[\delta_H(k_0)]^2 < q_H^2[\delta_H(k_0)] = 1 - \frac{1}{\beta^{2k_0} L_0^2} < 1 \quad \text{as well as} \quad \delta_H(k_0) = \frac{1}{\beta^{k_0} L_0} < \frac{2\alpha}{L[6M]^2}.$$

This implies that Proposition 8 holds for the theoretical sequence $\tilde{u}_H^0 := u_H^{k_0}$ and $\tilde{u}_H^{k+1} := \Phi_H(\delta_H; \tilde{u}_H^k)$. In particular, we conclude that energy contraction (31) holds with $q_H^2 = 1 - \delta_H^2$. Moreover, Remark 22 shows that the necessary criterion (76) is guaranteed to hold for the iterates $(\tilde{u}_H^k)_{k \in \mathbb{N}_0}$ as soon as (31) holds.

Step 2. Since the failure of (76) increases the current value of L to βL , it follows from Step 1 that (76) can fail only finitely often, until the recomputed sequence $(u_H^k)_{k \in \mathbb{N}_0}$ satisfies (76) for all $k \in \mathbb{N}_0$ with $k \geq k_0$. □

Remark 25. The optimality results for Algorithm A are expected to carry over – at least asymptotically – to Algorithm B; see Proposition 24. The major difficulty lies in algorithmically determining whether the correct estimate of the Lipschitz constant (and thus δ_H) is preasymptotic or not, *i.e.*, determining k in Step 2 from the last proposition by means of computable quantities only. However, it is ensured that δ_H remains uniformly bounded from below.

5. NUMERICAL EXPERIMENTS

In this section, we test and illustrate Algorithm B with numerical experiments. All experiments were implemented using the Matlab code *MooAFEM* [32]. Throughout, $\Omega \subset \mathbb{R}^2$ and we use $x = (x_1, x_2) \in \Omega$ to denote the Cartesian coordinates. In all experiments, we consider equation (59) with isotropic diffusion $\mathbf{A} = \begin{pmatrix} \varepsilon & 0 \\ 0 & \varepsilon \end{pmatrix}$ with $0 < \varepsilon \leq 1$. The adaptivity parameter is set to $\theta = 0.5$ and $C_{\text{mark}} = 1$. Moreover, recall the definition of the overall computational cost from (56), which reads

$$\text{work}(\ell, k) = \sum_{\substack{(\ell', k') \in \mathcal{Q} \\ (\ell', k') \leq (\ell, k)}} \#\mathcal{T}_{\ell'} = k \#\mathcal{T}_\ell + \sum_{\ell'=0}^{\ell-1} k(\ell') \#\mathcal{T}_{\ell'}.$$

Experiment 26 (Nonlinear variant of the sine-Gordon equation ([2], Expt. 5.1)).

For $\Omega = (0, 1)^2$, let $\mathcal{X} = H_0^1(\Omega)$ with $\|\cdot\|^2 = \langle \nabla \cdot, \nabla \cdot \rangle$ (*i.e.*, $\varepsilon = 1$) and consider

$$-\Delta u^* + (u^*)^3 + \sin(u^*) = f \quad \text{in } \Omega \quad \text{subject to} \quad u^* = 0 \quad \text{on } \partial\Omega, \tag{77}$$

with the monotone semilinearity $b(v) = v^3 + \sin(v)$, which satisfies (ELL), (CAR), (MON), and (GC). We set $\mathbf{f} = 0$ and choose f in such a way that

$$u^*(x) = \sin(\pi x_1) \sin(\pi x_2),$$

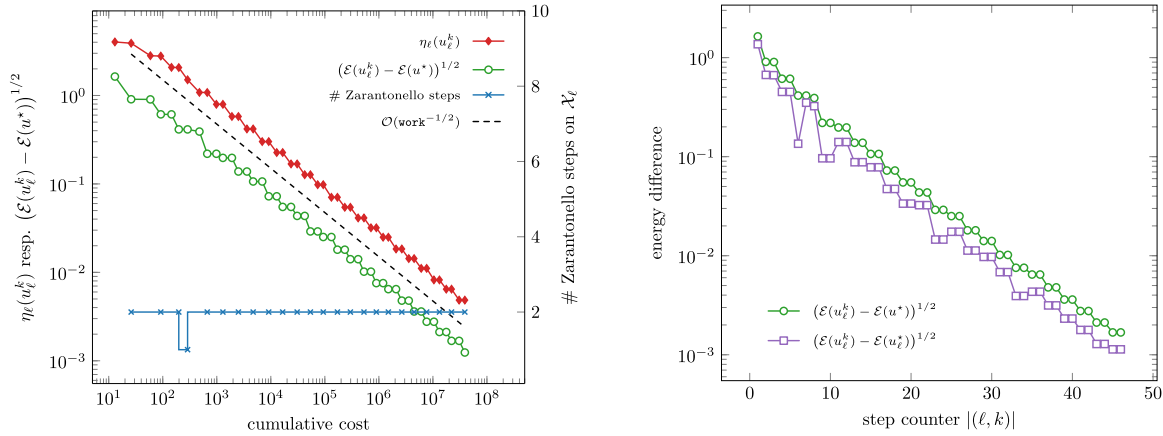


FIGURE 1. Results of Experiment 26 with polynomial degree $m = 1$. *Left*: error estimator $\eta_\ell(u_\ell^k)$ (diamond, left ordinate) and energy difference of iterative solutions $(\mathcal{E}(u_\ell^k) - \mathcal{E}(u^*))^{1/2}$ (circle, left ordinate) against $\text{work}(\ell, k)$ and the number of Zarantonello steps on \mathcal{X}_ℓ (cross, right ordinate). *Right*: energy difference of $\mathcal{E}(u_\ell^k)$ to $\mathcal{E}(u^*)$ (circle) and to $\mathcal{E}(u_\ell^*)$ (square) over the total step counter $|(\ell, k)|$. Throughout, $\mathcal{E}(u^*)$ is obtained by Aitken extrapolation and $\mathcal{E}(u_\ell^*)$ by sufficient Zarantonello steps on each level ℓ .

which satisfies (RHS). In Figure 1, we plot the *a posteriori* estimator $\eta_\ell(u_\ell^k)$ and the energy difference of the iterative solutions $(\mathcal{E}(u_\ell^k) - \mathcal{E}(u^*))^{1/2}$ against the $\text{work}(\ell, k)$ for lowest order FEM $m = 1$, where we approximate $\mathcal{E}(u^*)$ by means of Aitken convergence acceleration on uniform meshes with up to $\#\mathcal{T}_{\text{final}} = 67\,108\,864$ degrees of freedom on the finest mesh. The decay rate is of (expected) optimal order $\mathcal{O}(\text{work}(\ell, k)^{-1/2})$ as $|(\ell, k)| \rightarrow \infty$. Moreover, the experimentally observed number of sufficient linearization steps $k(\ell)$ is two. Furthermore, in Figure 1, we plot the difference of $\mathcal{E}(u_\ell^k)$ to the approximated reference energy $\mathcal{E}(u^*)$ using Aitken’s acceleration and to the energy $\mathcal{E}(u_\ell^*)$ on \mathcal{X}_ℓ over the step counter $|(\ell, k)|$. The reference energy $\mathcal{E}(u_\ell^*)$ is calculated by a sufficient number of Zarantonello iterations on each level ℓ until the energy difference of successive iterates is below the tolerance $\text{tol} < 10^{-15}$.

Experiment 27 (Singularly perturbed sine-Gordon equation).

This example is a variant of ([2], Expt. 5.2). For $d = 2$ and $\Omega = (0, 1)^2$, let $\varepsilon = 10^{-5}$ and consider

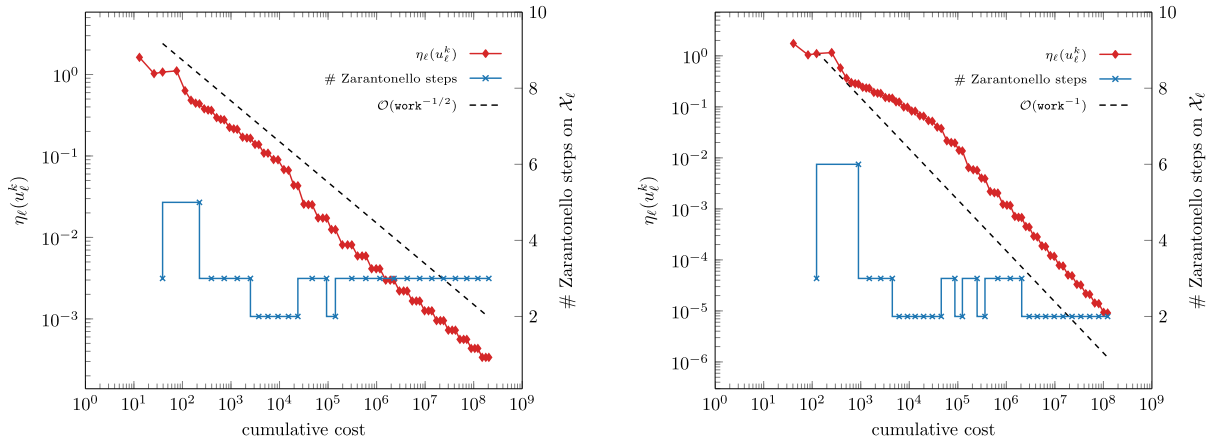
$$-\varepsilon\Delta u^* + 2u^* + \sin(u^*) = 1 \quad \text{in } \Omega \quad \text{subject to } u^* = 0 \text{ on } \partial\Omega,$$

with the monotone semilinearity $b(v) = v + \sin(v)$. In this case, the exact solution u^* is unknown. The used \mathcal{X} -norm is given by $\|\cdot\|_{\mathcal{X}}^2 = \varepsilon \langle \nabla \cdot, \nabla \cdot \rangle + \langle \cdot, \cdot \rangle$. The particular choice of the \mathcal{X} -norm allows for $\alpha = 1$ due to the monotonicity of $b(v)$. The problem clearly satisfies (ELL), (CAR), (MON), and (GC). Moreover, $f = 1$ and $\mathbf{f} = 0$ satisfy (RHS). In this experiment, we employ a slight modification of the error estimator (74) following ([41], Rmk. 4.14)

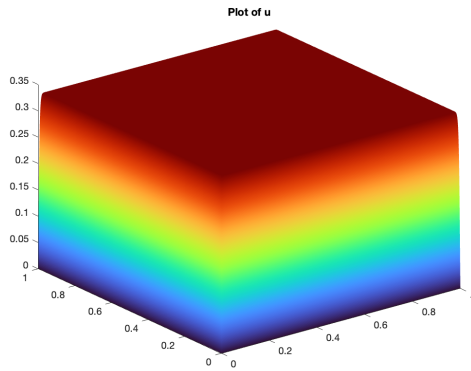
$$\eta_H(T, v_H)^2 := \tilde{h}_T^2 \|f + \varepsilon\Delta v_H - b(v_H)\|_{L^2(T)}^2 + \tilde{h}_T \|\llbracket \varepsilon \nabla v_H \cdot \mathbf{n} \rrbracket\|_{L^2(\partial T \cap \Omega)},$$

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

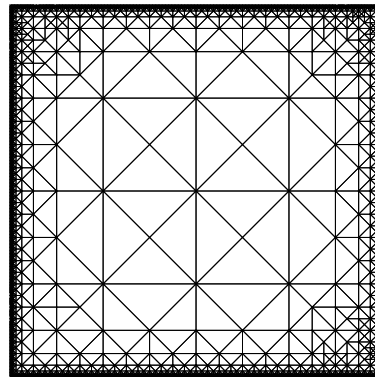
In Figure 2A, we plot the error estimator $\eta_\ell(u_\ell^k)$ for all $(\ell, k) \in \mathcal{Q}$ against the $\text{work}(\ell, k)$ for polynomial degrees $m \in \{1, 2\}$. The decay rate is of (expected) optimal order $\mathcal{O}(\text{work}(\ell, k)^{-m/2})$ as $|(\ell, k)| \rightarrow \infty$. The number of Zarantonello steps on each mesh refinement level ℓ stabilizes for $m \in \{1, 2\}$ at three ($m = 1$) and two ($m = 2$) after an initial phase. For $m = 2$, Figure 2B shows the approximate solution u_ℓ^k , where $\ell = 28$ and



(A)



(B)



(C)

FIGURE 2. Using the norm $\|\cdot\|^2 = \varepsilon \langle \nabla \cdot, \nabla \cdot \rangle + \langle \cdot, \cdot \rangle$ in Experiment 27. *Top*: convergence plot of the error estimator $\eta_\ell(u_\ell^k)$ over $\text{work}(\ell, k)$ and number of Zarantonello iterations on \mathcal{X}_ℓ over work for $m = 1$ (*top, left*) and $m = 2$ (*top, right*). *Bottom*: plot of the approximate solution u_ℓ^k (*bottom, left*) and plot of a sample mesh (*bottom, right*). (A) Error estimator $\eta_\ell(u_\ell^k)$ over work (diamond, left ordinate) and number of Zarantonello iteration steps on \mathcal{X}_ℓ over work (cross, right ordinate) for $m = 1$ (*left*) and $m = 2$ (*right*), (B) approximated solutions u_ℓ^k , where $\ell = 28$, $k(28) = 2$, and $m = 2$ and (C) mesh with $\#\mathcal{T}_\ell = 4295$, where $\ell = 11$ and $m = 1$.

$k(28) = 2$. Figure 2C depicts a mesh plot for $\#\mathcal{T}_\ell = 4295$ for $\ell = 11$ and $m = 1$. In particular, this experiment shows that Algorithm B is suitable for a setting with dominating nonlinear reaction given that a suitable norm on \mathcal{X} is chosen. Furthermore, we remark that the nonlinearity $b(v) = v + \sin(v)$ is globally Lipschitz continuous with Lipschitz constant $L = 2$. In our experiments, δ_ℓ is decreased twice, *i.e.*, δ_ℓ decreases from 1 to $0.5 = 1/L$, which is optimal according to Remark 9 and remains uniformly bounded from below; *cf.* Remark 25.

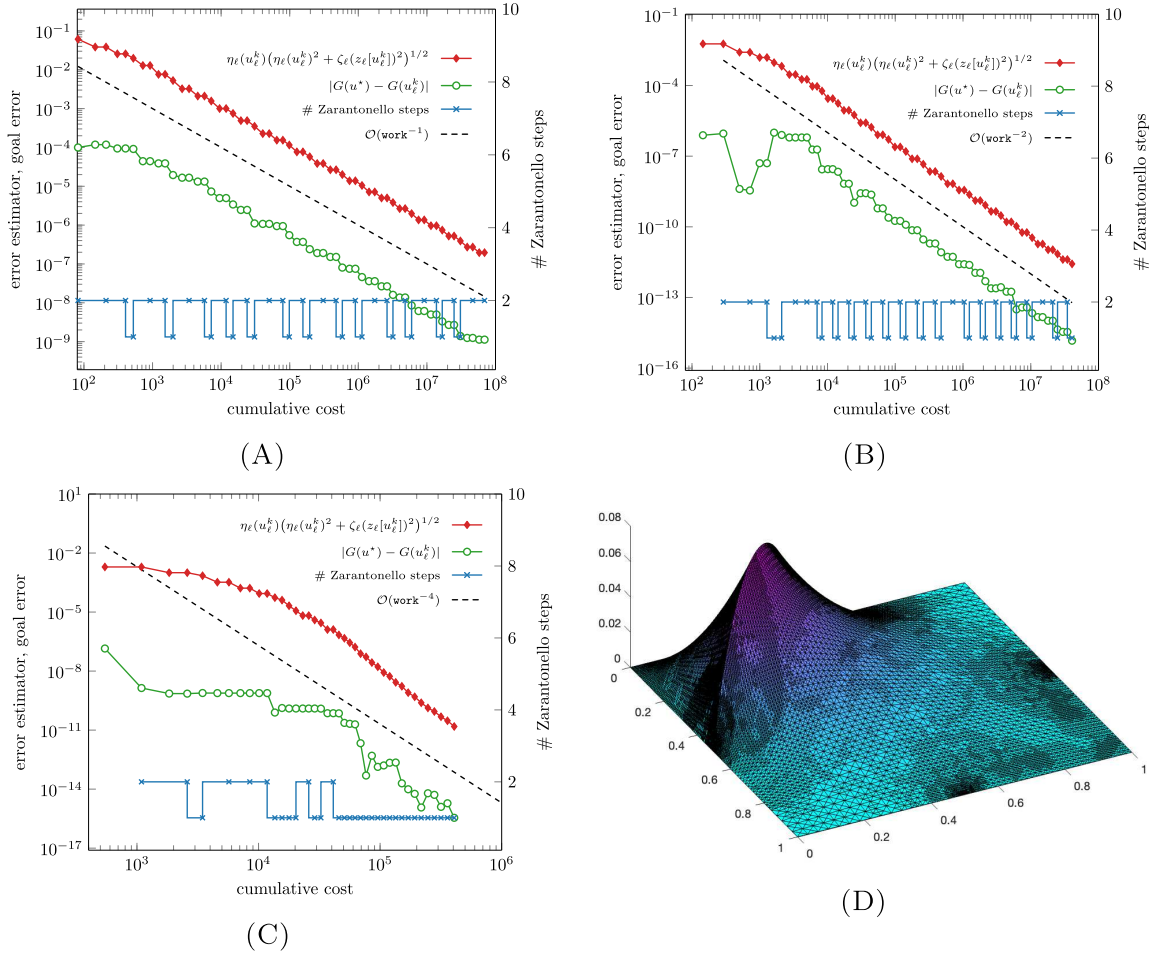


FIGURE 3. (A)–(C) Product error estimator $\eta_\ell(u_\ell^k) [\eta_\ell(u_\ell^k)^2 + \zeta_\ell(z_\ell[u_\ell^k])^2]^{1/2}$ (diamond, left ordinate), absolute goal error $|G(u^*) - G(u_\ell^k)|$ (circle, left ordinate), and number of Zarantonello steps on \mathcal{X}_ℓ over work (cross, right ordinate) for $m = 1$ (top, left), $m = 2$ (top, right), and $m = 4$ (bottom, left). (D) Plot of an iterative solution u_ℓ^k (bottom, right). (A) Results for $m = 1$, (B) results for $m = 2$, (C) results for $m = 4$, and (D) plot of iterative solution u_ℓ^k , $\ell = 16$, $\underline{k}(16) = 2$, $\dim(\mathcal{X}_\ell) = 14599$, and $m = 1$.

Experiment 28 (Goal-oriented AILFEM (GAILFEM)).

We also test a canonical extension of Algorithm B in a goal-oriented setting similar to that of ([34], Example 7.3). A thorough treatment of this problem (and the assumptions thereof) is found in ([9], Example 35). We use the proposed practical Algorithm B as the solve module for the semilinear primal problem in the GOAFEM algorithm ([9], Algo. 17). Let $\Omega = (0, 1)^2$ and $\varepsilon = 1$. The weak formulation of the primal problem reads: Find $u^* \in H_0^1(\Omega)$ such that

$$\langle \nabla u^*, \nabla v \rangle + \langle b(u^*), v \rangle = \int_\Omega \mathbf{f} \cdot \nabla v \, dx, \quad \text{for all } v \in H_0^1(\Omega), \tag{78}$$

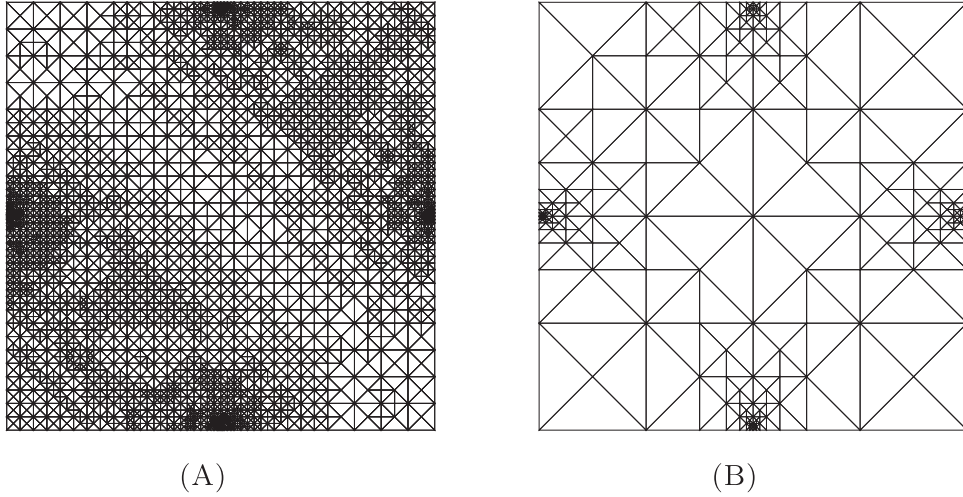


FIGURE 4. Generated GAILFEM meshes for $m = 1$ and $m = 4$. (A) Mesh generated for $m = 1$, where $\dim \mathcal{X}_\ell = 3092$ and $\ell = 12$ and (B) mesh generated for $m = 4$, where $\dim \mathcal{X}_\ell = 3081$ and $\ell = 12$.

where $b(v) = v^3$ and $\mathbf{f} = \chi_{\Omega_f}(-1, 0)$ with the characteristic function χ_{Ω_f} of $\Omega_f = \{x \in \Omega \mid x_1 + x_2 \leq \frac{1}{2}\}$. The weak formulation of the practical dual problem for the linearization point $w \in H_0^1(\Omega)$ reads: find $z^*[w] \in H_0^1(\Omega)$ such that

$$\langle \nabla z^*[w], \nabla v \rangle + \langle b'(w)z^*[w], v \rangle = \int_{\Omega} \mathbf{g} \cdot \nabla v \, dx, \quad \text{for all } v \in H_0^1(\Omega),$$

where $b'(v) = 3v^2$ and $\mathbf{g} = \chi_{\Omega_g}(-1, 0)$ with $\Omega_g = \{x \in \Omega \mid x_1 + x_2 \geq \frac{3}{2}\}$. The goal functional thus reads

$$G(v) := - \int_{\Omega_g} \frac{\partial v}{\partial x_1} \, dx \quad \text{for all } v \in H_0^1(\Omega).$$

Since $\text{div}(\mathbf{g}) = 0$ on every element $T \in \mathcal{T}_0$, the associated error estimator for the dual problem reads

$$\zeta_H(w; T, v_H)^2 := h_T^2 \|\Delta v_H - b'(w)(v_H)\|_{L^2(T)}^2 + h_T \|\llbracket (\nabla v_H - \mathbf{g}) \cdot \mathbf{n} \rrbracket\|_{L^2(\partial T \cap \Omega)}^2. \tag{79}$$

We used $\llbracket \cdot \rrbracket^2 = \langle \cdot, \cdot \rangle$ as the \mathcal{X} -norm. For various polynomial degrees $m \in \{1, 2, 4\}$, Figure 3A–C shows the results of the proposed GAILFEM algorithm driven by the product estimator $\eta_\ell(u_\ell^k) \left[\eta_\ell(u_\ell^k)^2 + \zeta_\ell(z_\ell[u_\ell^k])^2 \right]^{1/2}$, which is an upper bound to the goal error difference $G(u^*) - G(u_\ell^k)$ and a viable way to recover optimal convergence rates; cf. [9]. We plot the estimator product $\eta_\ell(u_\ell^k) \left[\eta_\ell(u_\ell^k)^2 + \zeta_\ell(z_\ell[u_\ell^k])^2 \right]^{1/2}$, the number of Zarantonello steps, and the absolute goal error difference $|G(u^*) - G(u_\ell^k)|$ over the $\text{work}(\ell, k)$, where $G(u^*) = -0.0015849518088245$ serves as a reference value; see ([9], Example 35). In Figure 3D, we plot the sample solution u_ℓ^k , where $\ell = 13$, $k(13) = 2$, and $m = 1$.

The decay rate is of (expected) optimal order $\mathcal{O}(\text{work}(\ell, k)^{-m})$ for $|(\ell, k)| \rightarrow \infty$, where $m \in \{1, 2, 4\}$ is the polynomial degree of the FEM space \mathcal{X}_ℓ . The number of Zarantonello steps does not exceed two for $m = \{1, 2, 4\}$ and stabilizes after an initial phase at one for $m = 4$, respectively. Figure 4 depicts two meshes for $m = 1$ and $m = 4$.

APPENDIX A. CONVERGENCE FOR VECTOR-VALUED SEMILINEAR PDES

This appendix aims to extend the analysis from Section 2 to problems where the monotone operator does not have a potential, *e.g.*, vector-valued semilinear PDEs. We prove plain convergence of Algorithm A without the assumption (POT) and with the modified stopping criterion

$$\| \| u_\ell^k - u_\ell^{k-1} \| \| \leq \lambda \eta_\ell(u_\ell^k) \quad \wedge \quad \| \| u_\ell^k \| \| \leq 2M \tag{i.b''}$$

replacing Algorithm A(i.b). The proof requires some preliminary observations: First, the convergence of the exact discrete solutions u_ℓ^* towards the exact solution u_∞^* in the so-called discrete limit space, which dates back to the seminal work [7]. Second, we need to show that the approximate discrete solutions u_ℓ^k converge to the same limit.

Lemma 29. *Suppose that \mathcal{A} satisfies (SM) and (LIP). With the discrete subspaces $\mathcal{X}_\ell \subset \mathcal{X}$ from Algorithm A (with or without the modified stopping criterion (i.b'')), define the discrete limit space $\mathcal{X}_\infty := \overline{\bigcup_{\ell=0}^{\underline{\ell}} \mathcal{X}_\ell}$, where we recall that $\underline{\ell} = \sup\{\ell \in \mathbb{N}_0 \mid (\ell, 0) \in \mathcal{Q}\}$. Then, there exists a unique $u_\infty^* \in \mathcal{X}_\infty$ which solves*

$$\langle \mathcal{A}u_\infty^*, v_\infty \rangle = \langle F, v_\infty \rangle \quad \text{for all } v_\infty \in \mathcal{X}_\infty. \tag{A.1}$$

Moreover, given the exact discrete solutions $u_\ell^* \in \mathcal{X}_\ell$, it holds that

$$\| \| u_\infty^* - u_\ell^* \| \| \rightarrow 0 \quad \text{as } \ell \rightarrow \underline{\ell}. \tag{A.2}$$

Additionally, suppose (A1)–(A3) and suppose that the choice of $\delta > 0$ in Algorithm A ensures norm contraction (30). Then, the approximations u_ℓ^k computed in Algorithm A fulfil that

$$\| \| u_\infty^* - u_\ell^k \| \| \rightarrow 0 \quad \text{as } (\ell, k) \in \mathcal{Q} \quad \text{with } |(\ell, k)| \rightarrow \infty. \tag{A.3}$$

Proof. The proof consists of three steps.

Step 1 (Exact solutions). Since $\mathcal{X}_\ell \subseteq \mathcal{X}_{\ell+1} \subset \mathcal{X}$, the discrete limit space $\mathcal{X}_\infty := \overline{\bigcup_{\ell=0}^{\underline{\ell}} \mathcal{X}_\ell}$ is a closed subspace of \mathcal{X} . Proposition 2 proves the existence of a unique $u_\infty^* \in \mathcal{X}_\infty$ satisfying (A.1). The Galerkin solutions u_ℓ^* from (4) are also Galerkin approximations of u_∞^* . Hence, there holds the C ea-type estimate

$$\| \| u_\infty^* - u_\ell^* \| \| \stackrel{(6)}{\leq} C_{\text{C ea}} \min_{v_\ell \in \mathcal{X}_\ell} \| \| u_\infty^* - v_\ell \| \| \xrightarrow{\ell \rightarrow \underline{\ell}} 0, \tag{A.4}$$

where convergence follows by definition of \mathcal{X}_∞ .

Step 2 (Approximate solutions for $\underline{\ell} = \infty$). The norm contraction (30) and $u_{\ell+1}^0 = u_\ell^k$ reveal that

$$0 \leq \| \| u_{\ell+1}^* - u_{\ell+1}^{k(\ell+1)} \| \| \stackrel{(30)}{\leq} q_N^{k(\ell+1)} \| \| u_{\ell+1}^* - u_{\ell+1}^0 \| \| \leq q_N \left[\| \| u_\ell^* - u_\ell^{k(\ell)} \| \| + \| \| u_{\ell+1}^* - u_\ell^* \| \| \right].$$

From Step 1, we infer that $(u_\ell^*)_{\ell \in \mathbb{N}_0}$ is a Cauchy sequence. Defining $a_\ell := \| \| u_\ell^* - u_\ell^{k(\ell)} \| \|$ and $b_\ell := q_N \| \| u_{\ell+1}^* - u_\ell^* \| \|$, the last estimate can be rewritten as

$$0 \leq a_{\ell+1} \leq q_N a_\ell + b_\ell, \quad \text{where } \lim_{\ell \rightarrow \infty} b_\ell = 0.$$

It follows from elementary calculus (*cf.* ([13], Coro. 4.8) that

$$0 = \lim_{\ell \rightarrow \infty} a_\ell = \lim_{\ell \rightarrow \infty} \| \| u_\ell^* - u_\ell^{k(\ell)} \| \|.$$

Altogether, we obtain that

$$\begin{aligned} \|\|u_\infty^* - u_\ell^k\|\| &\leq \|\|u_\infty^* - u_\ell^*\|\| + \|\|u_\ell^* - u_\ell^k\|\| \stackrel{(30)}{\leq} \|\|u_\infty^* - u_\ell^*\|\| + \|\|u_\ell^* - u_\ell^0\|\| \\ &\leq \|\|u_\infty^* - u_\ell^*\|\| + \|\|u_\ell^* - u_{\ell-1}^*\|\| + \|\|u_{\ell-1}^* - u_{\ell-1}^k\|\| \rightarrow 0 \quad \text{as } \ell \rightarrow \infty. \end{aligned}$$

Step 3 (Approximate solutions for $\underline{\ell} < \infty$ and $\underline{k}(\ell) = \infty$). It holds that $u_\infty^* = u_{\underline{\ell}}^*$ and hence, due to (30),

$$\|\|u_\infty^* - u_\ell^k\|\| = \|\|u_{\underline{\ell}}^* - u_{\underline{\ell}}^k\|\| \rightarrow 0 \quad \text{as } |(\ell, k)| \rightarrow \infty.$$

This concludes the proof. □

The following theorem states plain convergence in the abstract setting of the proposed AILFEM algorithm.

Theorem 30 (Plain convergence). *Suppose that \mathcal{A} satisfies (SM) and (LIP). Suppose the axioms of adaptivity (A1)–(A3). Suppose that the choice of $\delta > 0$ in Algorithm A ensures (30). Then, for any choice of the marking parameters $0 < \theta \leq 1$, $\lambda > 0$, and $1 \leq C_{\text{mark}} \leq \infty$, Algorithm A with modified stopping criterion (i.b'') guarantees convergence of the quasi-error from (38), i.e.,*

$$\Delta_\ell^k = \|\|u^* - u_\ell^k\|\| + \eta_\ell(u_\ell^k) \rightarrow 0 \quad \text{as } (\ell, k) \in \mathcal{Q} \text{ with } |(\ell, k)| \rightarrow \infty. \tag{A.5}$$

Proof. The assertion $|(\ell, k)| \rightarrow \infty$ consists of two cases:

Case 1 ($\underline{\ell} = \infty$). Recall the generalized estimator reduction ([13], Lemma 4.7): Let $\omega > 0$. Given the Dörfler marking in Algorithm A(iii), it follows that

$$\eta_{\ell+1}(u_{\ell+1}^k)^2 \leq q_{\text{est}} \eta_\ell(u_\ell^k)^2 + C_{\text{est}} \|\|u_{\ell+1}^k - u_\ell^k\|\|^2, \tag{A.6}$$

where $0 < q_{\text{est}} := (1 + \omega) [1 - (1 - q_{\text{red}}^2)\theta] < 1$ and $C_{\text{est}} := (1 + \omega^{-1}) C_{\text{stab}}[4M]^2$ with $\omega > 0$ being sufficiently small and where $4M$ stems from nested iteration (28). From Lemma 29, we infer that $\|\|u_{\ell+1}^k - u_\ell^k\|\| \rightarrow 0$ as $\ell \rightarrow \infty$. Hence, it follows from elementary calculus (cf. ([13], Coro. 4.8)) that $\eta_\ell(u_\ell^k) \rightarrow 0$ as $\ell \rightarrow \infty$. Moreover, this and Lemma 29 prove that

$$\begin{aligned} \|\|u^* - u_\ell^k\|\| &\stackrel{(A3)}{\leq} C_{\text{rel}} \eta_\ell(u_\ell^*) + \|\|u_\ell^* - u_\ell^k\|\| \stackrel{(A1)}{\leq} C_{\text{rel}} \eta_\ell(u_\ell^k) + (1 + C_{\text{rel}} C_{\text{stab}}[3M]) \|\|u_\ell^* - u_\ell^k\|\| \\ &\leq C_{\text{rel}} \eta_\ell(u_\ell^k) + (1 + C_{\text{rel}} C_{\text{stab}}[3M]) \left[\|\|u_\ell^* - u_\infty^*\|\| + \|\|u_\infty^* - u_\ell^k\|\| \right] \xrightarrow{\ell \rightarrow \infty} 0. \end{aligned}$$

We conclude that $\|\|u^* - u_\ell^k\|\| + \eta_\ell(u_\ell^k) + \eta_\ell(u_\ell^*) \rightarrow 0$ as $\ell \rightarrow \infty$. Due to (18) together with Lemma 29 and for $C'_{\text{rel}} := 1 + C_{\text{rel}}$, this yields for all $(\ell, k) \in \mathcal{Q}$ that

$$\begin{aligned} \Delta_\ell^k &\leq C'_{\text{rel}} \eta_\ell(u_\ell^*) + [1 + C_{\text{stab}}[3M]] \|\|u_\ell^* - u_\ell^k\|\| \stackrel{(30)}{\leq} C'_{\text{rel}} \eta_\ell(u_\ell^*) + [1 + C_{\text{stab}}[3M]] \|\|u_\ell^* - u_\ell^0\|\| \\ &\leq C'_{\text{rel}} \eta_\ell(u_\ell^*) + [1 + C_{\text{stab}}[3M]] \left[\|\|u_\ell^* - u_{\ell-1}^*\|\| + \|\|u_{\ell-1}^* - u_{\ell-1}^k\|\| \right] \xrightarrow{\ell \rightarrow \infty} 0. \end{aligned}$$

This concludes the proof of the first case.

Case 2 ($\underline{\ell} < \infty$ and $\underline{k}(\ell) = \infty$). Since $\underline{k}(\ell) = \infty$, at least one of the cases is met:

$$\#\{k \in \mathbb{N}_0 \mid \|\|u_\ell^k\|\| > 2M\} = \infty \quad \text{or} \quad \#\{k \in \mathbb{N}_0 \mid \lambda \eta_\ell(u_\ell^k) < \|\|u_\ell^k - u_\ell^{k-1}\|\|\} = \infty.$$

Since norm contraction (30) holds, the arguments to obtain (32) prove the existence of $k_0 \in \mathbb{N}$ such that, for all $k \geq k_0$, it holds that

$$\|\|u_\ell^k\|\| \leq 2M.$$

We deduce from the (not met) stopping criterion in Algorithm A(i.b'') and (30) that

$$\lambda \eta_{\underline{\ell}}(u_{\underline{\ell}}^k) \stackrel{\text{(i.b'')}}{<} \|\|u_{\underline{\ell}}^k - u_{\underline{\ell}}^{k-1}\|\| \xrightarrow{k \rightarrow \infty} 0.$$

With contraction (30), we see that

$$\|\|u^* - u_{\underline{\ell}}^k\|\| \stackrel{\text{(A3)}}{\leq} C_{\text{rel}} \eta_{\underline{\ell}}(u_{\underline{\ell}}^*) + \|\|u_{\underline{\ell}}^* - u_{\underline{\ell}}^k\|\| \stackrel{\text{(A1)}}{\leq} C_{\text{rel}} \eta_{\underline{\ell}}(u_{\underline{\ell}}^k) + (1 + C_{\text{stab}}[3M]) \|\|u_{\underline{\ell}}^* - u_{\underline{\ell}}^k\|\| \xrightarrow{k \rightarrow \infty} 0.$$

This concludes the proof of the second case and the proof is complete. □

The next corollary states that the exact solution $u^* = u_{\underline{\ell}}^*$ is discrete if $\underline{\ell} < \infty$. Moreover, if there exists ℓ with $\eta_{\ell}(u_{\ell}^k) = 0$, then the exact solution u^* coincides with u_{ℓ}^k .

Corollary 31. *Under the assumptions of Theorem 30, there hold the following implications:*

- (i) *If $\underline{\ell} = \sup\{\ell \in \mathbb{N}_0 \mid (\ell, 0) \in \mathcal{Q}\} < \infty$, then $u^* = u_{\underline{\ell}}^*$ and $\eta_{\underline{\ell}}(u_{\underline{\ell}}^*) = 0$.*
- (ii) *If $\ell \in \mathbb{N}_0$ with $\underline{k} < \infty$ and $\eta_{\ell}(u_{\ell}^k) = 0$, then $u_{\ell}^k = u^* = u_{\ell}^*$.*

Proof. (i) According to Theorem 30, it holds that

$$\Delta_{\underline{\ell}}^k = \|\|u^* - u_{\underline{\ell}}^k\|\| + \eta_{\underline{\ell}}(u_{\underline{\ell}}^k) \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

Norm contraction (30) proves that

$$\|\|u_{\underline{\ell}}^* - u_{\underline{\ell}}^k\|\| \leq q_N^k \|\|u_{\underline{\ell}}^* - u_{\underline{\ell}}^0\|\| \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

Uniqueness of the limit yields that $u^* = u_{\underline{\ell}}^*$. With stability (A1), we obtain that

$$0 \leq \eta_{\underline{\ell}}(u_{\underline{\ell}}^*) \leq \eta_{\underline{\ell}}(u_{\underline{\ell}}^k) + C_{\text{stab}}[3M] \|\|u_{\underline{\ell}}^* - u_{\underline{\ell}}^k\|\| \rightarrow 0 \quad \text{as } k \rightarrow \infty.$$

This concludes the proof of (i).

(ii) Note that the stopping criterion in Algorithm A(i.b'') implies that $\|\|u_{\underline{\ell}}^k - u_{\underline{\ell}}^{k-1}\|\| \leq \lambda \eta_{\underline{\ell}}(u_{\underline{\ell}}^k) = 0$ by assumption. Thus, $u_{\underline{\ell}}^k = u_{\underline{\ell}}^{k-1}$. This implies that $u_{\underline{\ell}}^{k-1}$ is a fixed point of $\Phi_{\underline{\ell}}(\delta; \cdot)$. Since the fixed point is unique, we infer that $u_{\underline{\ell}}^k = u_{\underline{\ell}}^{k-1} = u_{\underline{\ell}}^*$. With reliability (A3), we thus obtain that

$$\|\|u^* - u_{\underline{\ell}}^*\|\| \stackrel{\text{(A3)}}{\leq} C_{\text{rel}} \eta_{\underline{\ell}}(u_{\underline{\ell}}^*) = C_{\text{rel}} \eta_{\underline{\ell}}(u_{\underline{\ell}}^k) = 0.$$

This concludes the proof. □

Plain convergence is required to obtain results proving weak convergence in the spirit of ([9], Lemma 28). This is pivotal for achieving quasi-orthogonality along the lines of ([9], Lemma 29), which can substitute (8) in the proof of full linear convergence. Details are omitted.

Acknowledgements. The authors thankfully acknowledge support by the Austrian Science Fund (FWF) through the doctoral school *Dissipation and dispersion in nonlinear PDEs* (grant W1245) and the stand-alone projects *Computational nonlinear PDEs* (grant P33216) and *Analysis of \mathcal{H} -matrices* (grant P28367). Michael Innerberger, Jens Markus Melenk, and Dirk Praetorius are supported by the SFB *Taming complexity in partial differential systems* (grant SFB F65). Additionally, Maximilian Brunner and Michael Innerberger are supported by the *Vienna School of Mathematics*. The authors thank the anonymous reviewers for their constructive feedback which helped to improve results and presentation.

REFERENCES

- [1] M. Amrein and T.P. Wihler, Fully adaptive Newton-Galerkin methods for semilinear elliptic partial differential equations. *SIAM J. Sci. Comput.* **37** (2015) A1637–A1657.
- [2] M. Amrein, P. Heid and T.P. Wihler, A numerical energy minimisation approach for semilinear diffusion-reaction boundary value problems based on steady state iterations. Preprint [arXiv:2202.07398](https://arxiv.org/abs/2202.07398) (2022).
- [3] M. Arioli, E.H. Georgoulis and D. Loghin, Stopping criteria for adaptive finite element solvers. *SIAM J. Sci. Comput.* **35** (2013) A1537–A1559.
- [4] M. Arioli, J. Liesen, A. Mičldar and Z. Strakoš, Interplay between discretization and algebraic computation in adaptive numerical solution of elliptic PDE problems. *GAMM-Mitt.* **36** (2013) 102–129.
- [5] M. Aurada, M. Feischl, T. Führer, M. Karkulik and D. Praetorius, Efficiency and optimality of some weighted-residual error estimator for adaptive 2D boundary element methods. *Comput. Methods Appl. Math.* **13** (2013) 305–332.
- [6] R.E. Bank, M. Holst, R. Szypowski and Y. Zhu, Finite element error estimates for critical growth semilinear problems without angle conditions. Preprint [arXiv:1108.3661](https://arxiv.org/abs/1108.3661) (2011).
- [7] I. Babuška and M. Vogelius, Feedback and adaptive finite element solution of one-dimensional boundary value problems. *Numer. Math.* **44** (1984) 75–102.
- [8] R. Becker, S. Mao and Z. Shi, A convergent nonconforming adaptive finite element method with quasi-optimal complexity. *SIAM J. Numer. Anal.* **47** (2010) 4639–4659.
- [9] R. Becker, M. Brunner, M. Innerberger, J.M. Melenk and D. Praetorius, Rate-optimal goal-oriented adaptive FEM for semilinear elliptic PDEs. *Comput. Math. Appl.* **118** (2022) 18–35.
- [10] L. Belenki, L. Diening and C. Kreuzer, Optimality of an adaptive finite element method for the p -Laplacian equation. *IMA J. Numer. Anal.* **32** (2012) 484–510.
- [11] A. Bespalov, A. Haberl and D. Praetorius, Adaptive FEM with coarse initial mesh guarantees optimal convergence rates for compactly perturbed elliptic problems. *Comput. Methods Appl. Mech. Eng.* **317** (2017) 318–340.
- [12] P. Binev, W. Dahmen and R. DeVore, Adaptive finite element methods with convergence rates. *Numer. Math.* **97** (2004) 219–268.
- [13] C. Carstensen, M. Feischl, M. Page and D. Praetorius, Axioms of adaptivity. *Comput. Math. Appl.* **67** (2014) 1195–1253.
- [14] J.M. Cascón and R.H. Nochetto, Quasioptimal cardinality of AFEM driven by nonresidual estimators. *IMA J. Numer. Anal.* **32** (2012) 1–29.
- [15] J.M. Cascon, C. Kreuzer, R.H. Nochetto and K.G. Siebert, Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.* **46** (2008) 2524–2550.
- [16] S. Congreve and T.P. Wihler, Iterative Galerkin discretizations for strongly monotone problems. *J. Comput. Appl. Math.* **311** (2017) 457–472.
- [17] L. Diening and C. Kreuzer, Linear convergence of an adaptive finite element method for the p -Laplacian equation. *SIAM J. Numer. Anal.* **46** (2008) 614–638.
- [18] W. Dörfler, A convergent adaptive algorithm for Poisson’s equation. *SIAM J. Numer. Anal.* **33** (1996) 1106–1124.
- [19] L. El Alaoui, A. Ern and M. Vohralík, Guaranteed and robust a posteriori error estimates and balancing discretization and linearization errors for monotone nonlinear problems. *Comput. Methods Appl. Mech. Eng.* **200** (2011) 2782–2795.
- [20] A. Ern and M. Vohralík, Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. *SIAM J. Sci. Comput.* **35** (2013) A1761–A1791.
- [21] M. Feischl, T. Führer and D. Praetorius, Adaptive FEM with optimal convergence rates for a certain class of nonsymmetric and possibly nonlinear problems. *SIAM J. Numer. Anal.* **52** (2014) 601–625.
- [22] S. Fučík and A. Kufner, Nonlinear differential equations, in *Studies in Applied Mechanics*. Vol. 2. Elsevier, Amsterdam (1980).
- [23] G. Gantner, A. Haberl, D. Praetorius and B. Stiftner, Rate optimal adaptive FEM with inexact solver for nonlinear operators. *IMA J. Numer. Anal.* **38** (2018) 1797–1831.
- [24] G. Gantner, A. Haberl, D. Praetorius and S. Schimanko, Rate optimality of adaptive finite element methods with respect to overall computational costs. *Math. Comput.* **90** (2021) 2011–2040.
- [25] E.M. Garau, P. Morin and C. Zuppa, Convergence of an adaptive Kačanov FEM for quasi-linear problems. *Appl. Numer. Math.* **61** (2011) 512–529.
- [26] E.M. Garau, P. Morin and C. Zuppa, Quasi-optimal convergence rate of an AFEM for quasi-linear problems of monotone type. *Numer. Math. Theory Methods Appl.* **5** (2012) 131–156.
- [27] A. Haberl, D. Praetorius, S. Schimanko and M. Vohralík, Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver. *Numer. Math.* **147** (2021) 679–725.
- [28] P. Heid and T.P. Wihler, Adaptive iterative linearization Galerkin methods for nonlinear problems. *Math. Comput.* **89** (2020) 2707–2734.
- [29] P. Heid and T.P. Wihler, On the convergence of adaptive iterative linearized Galerkin methods. *Calcolo* **57** (2020) 24.
- [30] P. Heid, D. Praetorius and T.P. Wihler, Energy contraction and optimal convergence of adaptive iterative linearized finite element methods. *Comput. Methods Appl. Math.* **21** (2021) 407–422.
- [31] P. Houston and T.P. Wihler, An hp -adaptive Newton-discontinuous-Galerkin finite element approach for semilinear elliptic boundary value problems. *Math. Comput.* **87** (2018) 2641–2674.
- [32] M. Innerberger and D. Praetorius, MooAFEM: an object oriented Matlab code for higher-order adaptive FEM for (nonlinear) elliptic PDEs. *Appl. Math. Comput.* **442** (2023) 127731.

- [33] C. Kreuzer and K.G. Siebert, Decay rates of adaptive finite elements with Dörfler marking. *Numer. Math.* **117** (2011) 679–716.
- [34] M.S. Mommer and R. Stevenson, A goal-oriented adaptive finite element method with convergence rates. *SIAM J. Numer. Anal.* **47** (2009) 861–886.
- [35] P. Morin, R.H. Nochetto and K.G. Siebert, Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.* **38** (2000) 466–488.
- [36] P. Morin, K.G. Siebert and A. Veiser, A basic convergence result for conforming adaptive finite elements. *Math. Models Methods Appl. Sci.* **18** (2008) 707–737.
- [37] C.-M. Pfeiler and D. Praetorius, Dörfler marking with minimal cardinality is a linear complexity problem. *Math. Comput.* **89** (2020) 2735–2752.
- [38] R. Stevenson, Optimality of a standard adaptive finite element method. *Found. Comput. Math.* **7** (2007) 245–269.
- [39] R. Stevenson, The completion of locally refined simplicial partitions created by bisection. *Math. Comput.* **77** (2008) 227–241.
- [40] A. Veiser, Convergent adaptive finite elements for the nonlinear Laplacian. *Numer. Math.* **92** (2002) 743–770.
- [41] R. Verfürth, A Posteriori Error Estimation Techniques for Finite Element Methods. Oxford University Press, Oxford (2013).
- [42] K. Yosida, Functional analysis, in Classics in Mathematics. Springer, Berlin (1995). Reprint of the sixth (1980) edition.
- [43] E. Zeidler, Nonlinear Functional Analysis and Its Applications. Part II/B. Springer-Verlag, New York (1990).



Please help to maintain this journal in open access!

This journal is currently published in open access under the Subscribe to Open model (S2O). We are thankful to our subscribers and supporters for making it possible to publish this journal in open access in the current year, free of charge for authors and readers.

Check with your library that it subscribes to the journal, or consider making a personal donation to the S2O programme by contacting subscribers@edpsciences.org.

More information, including a list of supporters and financial transparency reports, is available at <https://edpsciences.org/en/subscribe-to-open-s2o>.