
DISSERTATION

# On optimal adaptivity for semilinear PDEs

ausgeführt zum Zwecke der Erlangung des akademischen Grades
eines Doktors der technischen Wissenschaften unter der Leitung von

**Univ.-Prof. Dipl.-Math. Dr.techn. Dirk Praetorius**
E101 – Institut für Analysis und Scientific Computing, TU Wien

eingereicht an der Technischen Universität Wien
Fakultät für Mathematik und Geoinformation

von

**Dipl.-Ing. Maximilian Brunner, BSc**
Matrikelnummer: 01528246

Diese Dissertation wurde begutachtet von

**Prof. Dr. Roland Becker**
Laboratoire de mathématiques et de leurs applications, Université de Pau et des Pays de l'Adour

**Prof. Dr. Dirk Praetorius**
Institut für Analysis und Scientific Computing, TU Wien

**Prof. Dr. Thomas Wihler**
Mathematisches Institut, Universität Bern

Wien, am 8. April 2024

# Kurzfassung

Diese Arbeit widmet sich ratenoptimalen adaptiven Finite Elemente Methoden (AFEMs) zur Lösung von semilinearen, elliptischen partiellen Differentialgleichungen (PDEs). Das zugrundeliegende Modellproblem besitzt einen nichtlinearen Reaktionsterm, wobei der mit der PDE assoziierte Operator *lokal* Lipschitz-stetig ist. Diese Arbeit präsentiert und analysiert Algorithmen, welche mehrere Fehlerquellen passend austarieren und dadurch ratenoptimal sind, d.h. eine Fehlergröße fällt mit bestmöglicher Rate über der Anzahl der Freiheitsgrade der Diskretisierung. Die drei Hauptkapitel haben folgenden Inhalt:

Im ersten Hauptkapitel untersuchen wir eine zielgerichtete AFEM (engl. *goal-oriented AFEM*, GOAFEM) für semilineare Probleme mit linearer Zielgröße (engl. *quantity of interest*). Bei GOAFEM steht die ratenoptimale Approximation eines Funktionalwertes der exakten, aber unbekannten Lösung im Vordergrund. Mittels gängiger Dualisierungstechnik ist der Approximationsfehler durch ein Produkt zweier Fehlerkomponenten abschätzbar, wodurch sich Konvergenzraten potentiell addieren. Dadurch ist GOAFEM in der Praxis sehr geschätzt. Der Approximationsfehler im Zielfunktional führt bei nichtlinearen Problemen zu einem nicht-berechenbaren, theoretischen dualen Problem, welches von der exakten Lösung abhängt. Deshalb wird dieses durch ein berechenbares, praktisches duales Problem ersetzt. Passendes Markieren der zu verfeinernden Elemente ermöglicht den Beweis von linearer Konvergenz: Kontraktion des Fehlerprodukts unabhängig davon, welche Fehlerkomponente die markierten Elemente bestimmt. Weiters zeigen wir optimale Konvergenzraten bezüglich der Anzahl der Freiheitsgrade der Diskretisierung. Dies erweitert die Literatur über ratenoptimale GOAFEM erstmalig auf ein Modellproblem mit nichtlinearer PDE.

Um das nichtlineare Modellproblem effizient zu lösen, betrachten wir im zweiten Hauptkapitel eine AFEM, welche auch die Anzahl der Linearisierungsschritte adaptiv steuert. Wir nehmen zunächst an, dass die linearisierten Systeme mit linearem Aufwand exakt gelöst werden können. Dann ist der präsentierte Algorithmus (engl. *adaptive iteratively linearized FEM*, AILFEM) kostenoptimal. Das heißt, eine Fehlergröße fällt mit optimalen Raten über dem kumulativen Rechenaufwand zur Berechnung der numerischen Approximation. Die Hauptschwierigkeit in der numerischen Analysis lokal Lipschitz-stetiger Probleme ist es, die uniforme Beschränktheit aller berechneten Iterierten zu zeigen. Damit können wir volle R-lineare Konvergenz zeigen, d.h. Kontraktion einer Fehlergröße unabhängig von der Wahl der Adaptivitätsparameter und unabhängig davon, ob das Gitter verfeinert wird oder ein weiterer Linearisierungsschritt vollzogen wird. Für hinreichend kleine Adaptivitätsparameter zeigen wir schließlich optimale Konvergenzraten bezüglich des theoretischen Rechenaufwands der präsentierten AILFEM.

Im dritten Hauptkapitel analysieren wir die obige AILFEM, wobei das linearisierte Problem mittels eines algebraischen Lösers approximativ gelöst wird. Die numerische Störung dieser inexakten Linearisierung erhöht signifikant die mathematische Schwierigkeit, die uniforme Beschränktheit aller Iterierten zu zeigen. Wie zuvor beweisen wir volle R-lineare Konvergenz einer Fehlergröße, welche nun Diskretisierungs-, Linearisierungs- und algebraischen Fehler beinhaltet. Wir folgern optimale Raten bezüglich des Rechenaufwands. Da nun alle Schritte der AILFEM rigoros mit linearer Komplexität realisierbar sind, sind Konvergenzraten bezüglich des kumulativen Rechenaufwands auch als Raten bezüglich der Gesamtrechenzeit zu verstehen. Dies zeigt sich auch in numerischen Experimenten.

# Abstract

This thesis is devoted to rate-optimal adaptive finite element methods (AFEMs) for semilinear elliptic partial differential equations (PDEs). It considers a model problem with a nonlinear reaction term, where the operator associated to the PDE is *locally* Lipschitz continuous. By equilibrating various error sources, this thesis proves that all presented algorithms are rate-optimal, i.e., a suitable error quantity converges with optimal decay rate with respect to the number of degrees of freedom of the discretization. The main contributions are the following:

First, we investigate a goal-oriented AFEM (GOAFEM) for the semilinear model problem where the principal aim is to approximate a linear functional (*quantity of interest*) evaluated at the exact, but unknown solution with optimal convergence rates. By means of established duality techniques, the approximation error can be estimated by a product of two approximation errors. This product structure allows that convergence rates add up, contributing substantially to the attractivity of GOAFEMs in practice. For nonlinear problems, the approximation error in the goal first leads to a noncomputable theoretical dual problem that depends on the unavailable exact solution. To make the goal error accessible, we replace this by a computable practical dual problem. A suitable marking strategy for the refinement allows for the proof of R-linear convergence: contraction of the error product regardless of which error component determines the marked elements. Moreover, we show optimal convergence rates with respect to the number of degrees of freedom of the discretization. This, for the first time, extends the literature on rate-optimal GOAFEM to a model problem with underlying nonlinear PDE.

Second, to efficiently solve the nonlinear model problem, we consider an AFEM, where the number of linearization steps is also steered adaptively. Under the assumption that all arising linear systems can be solved at linear cost, the proposed algorithm, coined as adaptive iteratively linearized finite element method (AILFEM), is cost-optimal. This means that the suitable error quantity decays with optimal convergence rates with respect to the (theoretical) overall computational cost that is needed to obtain the numerical approximation. The main challenge in the numerical analysis of locally Lipschitz continuous problems is to ensure that all iterates are uniformly bounded. Having achieved this, we prove full R-linear convergence, i.e., contraction of an error quantity independently of the adaptivity parameters and regardless of whether we refine the mesh or perform a linearization step. For sufficiently small adaptivity parameters, we eventually establish optimal convergence rates with respect to the theoretical computational cost of the proposed AILFEM.

Third, we analyze the preceding AILFEM, where the linearized problem is additionally solved with an iterative algebraic solver. This perturbation of the exact linearization procedure significantly increases the technicalities to verify uniform boundedness of all iterates. As before, we prove full R-linear convergence for an error quantity that now consists of error components stemming from discretization, linearization, and the algebraic solver. We conclude optimal rates with respect to computational complexity. Importantly, all steps in the AILFEM strategy can now rigorously be realized in linear complexity. Hence, optimal convergence rates with respect to overall computational cost can indeed be understood as optimal convergence rates with respect to computation time. This is also observed in numerical experiments.

# Danksagung

Ich möchte mich herzlichst bei Prof. Dirk Praetorius für die letzten Jahre bedanken. Der Start des Doktorats inmitten einer Hochschulschließung verlangte viel ab, wurde aber mit vielen Zoommeetings bestmöglich abgefangen. Zurück im Büro war es immer möglich bei Fragen vorbeizuschauen, Probleme anzusprechen und von Dirks reichhaltiger Erfahrung zu profitieren. Fachliche und persönliche Weiterentwicklung erfordert ehrliches, konstruktives Feedback und eine Atmosphäre, die Fehler erlaubt, welche aber nie zu komfortabel ist. Die adaptive Betreuung, welche auch viel Freiheiten bot, sowie das Umfeld der Arbeitsgruppe ermöglichten genau das. Das häufige, stets konstruktive Feedback hat sichergestellt, *dass die Techniken nun alle da sind*. Danke dafür.

Ich danke ebenso Prof. Markus Melenk und Prof. Roland Becker für die konstruktive Zusammenarbeit. Ganz besonders danke ich auch Prof. Roland Becker und Prof. Thomas Wihler für die Begutachtung der vorliegenden Dissertation.

Ich möchte mich bei Philipp Bringmann für den wertvollen Input zur Disseration bedanken. Die Arbeit an der Universität wäre nur halb so produktiv und halb so lustig ohne Abeitsgruppe. Deshalb möchte ich mich bei der ehemaligen und aktuellen (erweiterten) Arbeitsgruppe bedanken: Michele Aldé, Marie Auzinger, Björn Bahr, Maximilian Bernkopf, Simon Brandstetter, Philipp Bringmann, Ondine Chanon, Andreas Czink, Brigitte Ecker, Markus Faustmann, Michael Feischl, Rebecca Fischer, Giovanni Di Fratta, Alexander Freiszlinger, Hubert Hackl, Valentin Helml, Amanda Huber, Michael Innerberger, Ani Miraçi, Yadana Yu Niang, Carl–Martin Pfeiler, Alexander Rieder, Michele Ruggeri, Andrea Scaglioni, Stefan Schimanko, Ursula Schweigler, Julian Streitberger, David Wörgötter, und Fabian Zehetgruber. Die lustigen, ausgedehnten Mittagspausen, aber auch die Zaubershows von Björn Bahr haben die Zeit des Doktorats versüßt und waren ein gewichtiger Grund, dass ich immer gerne ins Büro gekommen ist. Auch gilt mein Dank der Kaffeepolicy des ASC, welche mir nicht unwesentlich durch das Doktorat getragen hat.

Das Abschließen eines Doktoratsstudiums ist immer eine große Leistung des persönlichen Umfeldes. Deshalb bedanke ich mich bei allen Studienkolleg*Innen für die unzähligen Stunden beim Übung lösen, die WG-Feiern und den starken Zusammenhalt, wenn es schwierig war.

Meiner Familie und ganz besonders meinen Eltern danke ich von ganzem Herzen. Sie haben mir stets freie Wahl gelassen, wie ich mich entwickeln möchte und sie haben mich in allen, aber besonders auch in schwierigen Zeiten bedingungslos unterstützt und an mich geglaubt. Ich danke für dieses Privileg.

Stets unterstützend und verlässlich, mit einem offenem Ohr und viel Geduld, mit viel Einsatz, Liebe und Witz — der größte Dank gilt meiner Freundin Martina Schwarz.

Ich danke dem *österreichischen Wissenschaftsfonds* (FWF), der meine Arbeit über die Projekte *Analysis of $\mathcal{H}$-matrices* (grant P28367) und *Computational nonlinear PDEs* (grant P33216) finanziert hat, sowie der *Vienna School of Mathematics*. Ebenso möchte ich für die Förderung zur Teilnahme an der *Summer school on numerical analysis of nonlinear PDEs* bei Prof. Gallistl bedanken und beim *International Office der TU Wien* für die Konferenzteilnahme an der ENUMATH 2023.

# Eidesstattliche Erklärung

Ich erkläre an Eides statt, dass ich die vorliegende Dissertation selbstständig und ohne fremde Hilfe verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt bzw. die wörtlich oder sinngemäß entnommenen Stellen als solche kenntlich gemacht habe.

Wien, am 8. April 2024

Maximilian Brunner

# Contents

# 1 Introduction

Nowadays in science, partial differential equations (PDEs) are ubiquitous. Their applications range from classical mechanics to electrodynamics, hysteresis phenomena, the Schrödinger equation in quantum physics, but PDEs are also used as the underlying mathematical foundation for the Black–Scholes equation in option pricing. Due to the complex nature of PDEs, their exact solution is in general not available. This makes the development of numerical methods to reliably approximate the unknown solution of highest relevance. Adaptive finite element methods (AFEMs) are computationally particularly effective as underlined by the following quote.

> *In the past three decades* ***self-adaptive discretisation*** *methods have gained enormous importance for the numerical solution of partial differential equations that arise from physical and technical applications. The aim is to obtain a numerical solution within a prescribed tolerance using a* ***minimal amount of work****. [emphasis added in boldface]*
>
> — Rüdiger Verfürth in [Ver13, Preface], 2013

This thesis investigates adaptive finite element methods (AFEMs) for a certain class of nonlinear problems, namely semilinear elliptic PDEs. The principal objective is to prove optimal convergences rates with respect to the number of degrees of freedom of computable approximate solutions towards the exact (unknown) solution or towards a *quantity of interest* that depends on the exact solution. The presented AFEMs have proven optimal convergence rates with respect to the number of degrees of freedom of the finite element space and, by including linearization and algebraic solver, also quasi-optimal computational cost.

The thesis is structured as follows: This introduction discusses the involved concepts, presents the main results, and puts the results in context with the existing literature. In Section 1.1, we present the semilinear model problem considered throughout the thesis and its inherent properties. Section 1.2 introduces the concept of mesh adaptivity and discusses the standard routines that are present in any AFEM routine. We present a schematic AFEM algorithm that also takes linearization and an algebraic solver into account (adaptive iteratively linearized FEM, AILFEM; Algorithm 1.8 below). Section 1.3 explains in which sense optimal convergence is understood and presents the main contributions of this thesis for the presented AILFEMs. Section 1.4 motivates the extension of AFEM to the goal-oriented setting (GOAFEM) and states the main theorem on optimal convergence rates. We conclude the introduction with an outline of this thesis in Section 1.5 and other scientific contributions beyond this thesis that are not presented in detail (Section 1.6).

The main part of the thesis is subdivided into three main chapters.

> [①GOA]: 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.*, 118:18–35, 2022. DOI: 10.1016/j.camwa.2022.05.008

Chapter 2 is based on [①GOA] and investigates a GOAFEM for the semilinear model problem. The main results are optimal convergence rates with respect to the number of degrees of freedom of the finite element space.

[②AIL1]: R. Becker, M. Brunner, M. Innerberger, J. M. Melenk, and D. Praetorius. Cost-optimal adaptive iterative linearized FEM for semilinear elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 57(4):2193–2225, 2023. DOI: 10.1051/m2an/2023036

Chapter 3 recasts the semilinear model problem in an abstract framework of locally Lipschitz continuous operators and proves optimal convergence rates with respect to computational cost under the assumption that the arising linearized systems can be solved in linear complexity. The underlying publication of this chapter is [②AIL1].

[③AIL2]: M. Brunner, D. Praetorius, and J. Streitberger. Cost-optimal adaptive FEM with linearization and algebraic solver for semilinear elliptic PDEs, 2024. arXiv: 2401.06486

Building on [②AIL1], we analyze the proposed AILFEM with an additional nested loop to solve the linearized systems iteratively in linear complexity. Chapter 4 investigates the perturbation from the inexact linearization and is based on [③AIL2]. Since all steps in the proposed algorithm can indeed be realized in linear complexity, we prove optimal convergence rates with respect to computation time.

## 1.1 Semilinear elliptic problem and first results

By means of conforming finite element methods, we seek for a rate-optimal discrete approximation of the solution $u^\star \in H^1_0(\Omega)$ to the *second-order semilinear elliptic* model problem

$$\begin{aligned} -\operatorname{div}(\boldsymbol{A}\nabla u^\star) + b(u^\star) &= f - \operatorname{div}(\boldsymbol{f}) && \text{in } \Omega, \\ u^\star &= 0 && \text{on } \partial\Omega, \end{aligned} \tag{1.1}$$

where the computational domain $\Omega \subset \mathbb{R}^d$ is a Lipschitz domain with $d \in \{1,2,3\}$, the diffusion coefficient $\boldsymbol{A}\colon \Omega \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ is elliptic, the nonlinear reaction coefficient $b\colon \Omega \to \mathbb{R}$ is monotonically increasing, and given data $f \in L^2(\Omega)$ and $\boldsymbol{f} \in [L^2(\Omega)]^d$. A precise discussion of the assumptions is found in [②AIL1, Section 3.3 below]. For the moment, we only highlight that the nonlinear reaction $s \mapsto b(s)$ for $s \in \mathbb{R}$ satisfies a certain growth condition (for further details, cf. (GC) and its more restrictive variant (CGC) below). This growth condition ensures that the semilinear term is a *compact perturbation* of a linear model problem.

In its *weak form*, the model problem (1.1) reads: Find a solution $u^\star \in H^1_0(\Omega)$ that satisfies

$$\langle \mathcal{A}u^\star, v\rangle := \langle \boldsymbol{A}\nabla u^\star, \nabla v\rangle_\Omega + \langle b(u^\star), v\rangle_\Omega = \langle f, v\rangle_\Omega + \langle \boldsymbol{f}, \nabla v\rangle_\Omega =: \langle F, v\rangle \text{ for all } v \in H^1_0(\Omega), \tag{1.2}$$

where $\langle\cdot\,,\cdot\rangle_\Omega$ denotes the $L^2(\Omega)$-scalar product, which is naturally extended to the duality brackets $\langle\cdot\,,\cdot\rangle$ between $H^1_0(\Omega)$ and its topological dual space $H^{-1}(\Omega) = H^1_0(\Omega)'$. The weak formulation is obtained by multiplying the strong form (1.1) with a so-called test function $v \in H^1_0(\Omega)$ and integration by parts. Since $H^1_0(\Omega)$-functions can be characterized by its vanishing trace on the boundary, the boundary condition is already incorporated into the ansatz space $H^1_0(\Omega)$. We note that $F \in H^{-1}(\Omega)$, and oftentimes abbreviate $F(v) := \langle F\,,\, v\rangle$.

Throughout the thesis, we suppose that the linear diffusion coefficient $\boldsymbol{A}(x)$ is bounded, symmetric, and uniformly elliptic. For $v \in H^1_0(\Omega)$, the associated energy norm $|||\cdot|||$ is induced by the principal part of the PDE and defined as

$$|||v|||^2 := \langle \boldsymbol{A}\nabla v\,,\,\nabla v\rangle_\Omega.$$

This is an equivalent norm on $H^1_0(\Omega)$. The induced energy scalar product is denoted by $\langle\!\langle \cdot\,,\,\cdot \rangle\!\rangle = |||\cdot|||^2$.

To guarantee *well-posedness* of (1.2) (and hence the well-posedness of the model problem), we require that the operator $\mathcal{A}\colon H^{-1}(\Omega) \to H^1_0(\Omega)$ is strongly monotone, i.e., there exists $\alpha > 0$ such that

$$\alpha\,|||v-w|||^2 \le \langle \mathcal{A}v - \mathcal{A}w\,,\,v-w\rangle \quad \text{for all } v,w \in H^1_0(\Omega), \tag{SM}$$

and locally Lipschitz continuous, i.e., there exists $L[\max\{|||v|||,|||w|||\}] > 0$ such that

$$\sup_{\varphi\in H^1_0(\Omega)\setminus\{0\}} \frac{\langle \mathcal{A}v - \mathcal{A}w\,,\,\varphi\rangle}{|||\varphi|||} \le L[\max\{|||v|||,|||w|||\}]\,|||v-w||| \quad \text{for all } v,w \in H^1_0(\Omega). \tag{LIP}$$

Then, the Browder–Minty theorem on monotone operators (see [Zei90, Theorem 26.A]) guarantees the existence and uniqueness of the solution $u^\star \in H^1_0(\Omega)$ to (1.2).

The weak form (1.2) can be made approximable through *discretization* techniques. One fundamental advantage of finite element methods is that they allow for fairly general geometries. In order to neglect any error form the partition of the domain, let the underlying domain $\Omega \subset \mathbb{R}^d$ be a polygonal, bounded Lipschitz domain.

Once the domain $\Omega$ is decomposed by a simplicial triangulation $\mathcal{T}_H$ (cf. [EG04, Section 1.3]), we replace $\mathcal{X} := H^1_0(\Omega)$ by piecewise polynomial ansatz spaces with fixed $p \in \mathbb{N}$

$$\mathcal{S}^p(\mathcal{T}_H) := \{v \in H^1(\Omega) \mid \text{for all } T \in \mathcal{T}_H, v_H|_T \text{ is a polynomial of total degree} \le p\}.$$

Furthermore, we define $\mathcal{X}_H := \mathcal{S}^p_0(\mathcal{T}_H) := H^1_0(\Omega) \cap \mathcal{S}^p(\mathcal{T}_H)$. The *discrete formulation* reads: Seek $u^\star_H \in \mathcal{X}_H$ such that

$$\langle \mathcal{A}u^\star_H\,,\,v_H\rangle = \langle F\,,\,v_H\rangle \quad \text{for all } v_H \in \mathcal{X}_H. \tag{1.3}$$

Since $\mathcal{X}_H \subset \mathcal{X}$ is a closed subspace, the Browder–Minty theorem ensures the existence and uniqueness of the discrete solution $u^\star_H \in \mathcal{X}_H$ as well.

The *compact* growth of the semilinearity $b$ ensures that there exists a well-defined *energy functional* $\mathcal{E}$ for the semilinear model problem (1.2) (cf. [②AIL1, Assumption (CGC) and Section 3.3.6 below]). Hence, the semilinear model problem (1.2) can be seen as the Euler–Lagrange equation of an energy minimization problem with a Gâteaux differentiable functional $\mathcal{E}\colon \mathcal{X} \to \mathbb{R}$. More precisely, the operator $\mathcal{A}$ possesses a Gâteaux differentiable potential $\mathcal{P}\colon \mathcal{X} \to \mathbb{R}$ such that its derivative $\mathrm{d}\mathcal{P}\colon \mathcal{X}' \to \mathcal{X}$ equals $\mathcal{A}$, i.e.,

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

The energy $\mathcal{E}$ can be defined as $\mathcal{E}(v) \coloneqq (\mathcal{P} - F)v$ and, for (1.2), reads

$$\mathcal{E}(v) \coloneqq \frac{1}{2}\int_\Omega |\boldsymbol{A}^{1/2}\nabla v|^2 \,\mathrm{d}x + \int_\Omega \int_0^{v(x)} b(s)\,\mathrm{d}s\,\mathrm{d}x - \int_\Omega fv\,\mathrm{d}x - \int_\Omega \boldsymbol{f}\cdot\nabla v\,\mathrm{d}x. \tag{1.4}$$

There holds a classical equivalence that relates energy norm $|||\cdot|||$ and energy differences.

**Lemma 1.1** (see, e.g., [GHPS18, Lemma 5.1]). *Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and (POT). Let $v_H \in \mathcal{X}_H$ with $\max\{|||v_H|||, |||u_H^\star|||\} \leq \vartheta$. Then, it holds that*

$$\frac{\alpha}{2}|||u_H^\star - v_H|||^2 \leq \mathcal{E}(v_H) - \mathcal{E}(u_H^\star) \leq \frac{L[\vartheta]}{2}|||u_H^\star - v_H|||^2. \tag{1.5}$$

*In particular, the discrete formulation to (1.3) is equivalent to the energy minimization problem:*

$$\text{Find} \quad u_H^\star \in \mathcal{X}_H \quad \text{such that} \quad \mathcal{E}(u_H^\star) = \min_{v_H\in\mathcal{X}_H} \mathcal{E}(v_H). \quad \square \tag{1.6}$$

An important feature of strongly monotone and locally Lipschitz continuous problems is that there holds a *quasi-best approximation* property (see [①GOA, Proposition 2.11 below] for the semilinear model problem and [②AIL1, Proposition 3.2 below] for its proof in the abstract setting).

**Proposition 1.2** (Céa). *Suppose (SM) and (LIP). Then, the solution $u^\star \in H^1_0(\Omega)$ from (1.2) and its discrete approximation $u_H^\star \in \mathcal{X}_H$ from (1.3) satisfy*

$$|||u^\star - u_H^\star||| \leq \frac{L[2M]}{\alpha}\min_{v_H\in\mathcal{X}_H}|||u^\star - v_H|||, \text{ where} \quad M \coloneqq \frac{1}{\alpha}\|F - \mathcal{A}0\|_{H^{-1}(\Omega)}. \tag{1.7}$$

*We use $C_{\mathrm{Céa}} = C_{\mathrm{Céa}}[M] \coloneqq L[2M]/\alpha$ to abbreviate the constant. The factor 2 stems from a slightly different but equivalent definition of local Lipschitz continuity (cf. (LIP) and Remark 3.1 in the third chapter below).* $\square$

The investigation of nonlinearities typically asks which (polynomial) growth $N \in \mathbb{N}_0$ of the reaction contribution

$$|b|\colon \mathbb{R}\to\mathbb{R}, \quad \xi \mapsto |b(\xi)| \leq C\,|\xi|^N \quad \text{for } C > 0 \tag{1.8}$$

is possible such that the problem (1.2) remains computationally stable. By choosing suitable norms that are connected to the variational setting, this question can be answered with arguments based on Sobolev embeddings and is related to the *local Lipschitz continuity* in its core. To see this, suppose that $b(0) = 0$. This is without loss of generality, since otherwise the right-hand side of (1.2) may be replaced by $\widetilde{f} \coloneqq f - b(0)$. The suitable *growth condition* on $N$ from [③AIL2, Assumption (GC) below] reads:

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

$$|b^{(N)}(\xi)| \leq R \quad \text{for all } \xi \in \mathbb{R}. \tag{GC}$$

The growth condition (GC) admits the estimate

$$\|b(u_H) - b(0)\|_{H^{-1}(\Omega)} \lesssim |||u_H - 0|||^p =: C_{\mathrm{bnd}}[|||u_H|||]\, |||u_H||| \quad \text{for all appoximations } u_H \approx u^\star.$$

In the existing literature, discrete $L^\infty(\Omega)$-bounds are either assumed for the discrete exact solutions (e.g., [HPZ15; XHYM21]) or derived under the assumption that $u^\star \in H^s(\Omega)$ for $s > 1$ [BHSZ11]. Without additional regularity, a discrete maximum principle is restrictive in an adaptive setting, since it imposes angle constraints on the triangulations. Oftentimes, global Lipschitz continuity of $b$, i.e., a global Lipschitz constant $L > 0$ is also supposed in the literature (e.g., [AW15; HPZ15; XHYM21]).

The preprint [BHSZ11] shows that the global Lipschitz continuity can be replaced by a growth condition. By further improving this approach (without supposing additional regularity), all presented works [①GOA, Chapter 2 below], [②AIL1, Chapter 3 below], and [③AIL2, Chapter 4 below] rely on growth conditions and only on the local Lipschitz continuity of the semilinearity $b$ without discrete $L^\infty(\Omega)$-bounds. This is a novel result and generalizes the existing literature.

The difficulty of requiring the Lipschitz continuity assumption (LIP) only locally is that the Lipschitz constant may vary with the functions $v \in H^1_0(\Omega)$ and $w \in H^1_0(\Omega)$. This is also the case in the energy equivalence from Lemma 1.1 and the Céa lemma 1.2 that exploit (LIP). This local dependence is also passed on to the stability constant of the residual *a posteriori* error estimator that is used to steer the algorithm (see [①GOA, Stability (A1) below]). In conclusion, the local Lipschitz continuity necessitates *uniform boundedness* of all computed quantities in the algorithm.

**Proposition 1.3.** *Suppose (SM), (LIP), and possibly (POT) (depending on the case chosen in Proposition 1.4 below). If the algorithm takes linearization and/or an algebraic solver into account, we additionally suppose the estimator axioms (A1)–(A3) introduced below. Then, there exists $C_{\mathrm{bnd}} = C_{\mathrm{bnd}}[M]$ with $M := \frac{1}{\alpha}\|F - \mathcal{A}0\|_{H^{-1}(\Omega)}$ such that*

$$|||u_H||| \leq C_{\mathrm{bnd}} = C_{\mathrm{bnd}}[M] \quad \text{for all } u_H \text{ that are computed in Algorithm 1.8 below.} \tag{UB}$$

*The bound simplifies to $C_{\mathrm{bnd}} = M$ for the exact solutions; cf [②AIL1, Proposition 3.2] below.*

*Sketch of proof.* This is rigorously proven in [①GOA, Lemma 2.8 below], where the argument exploits a continuous maximum principle for $u^\star$ and does not rely on the estimator axioms (A1)–(A3). In [②AIL1, Corollary 3.11 below] as well as [③AIL2, Theorem 4.8 below], the argument mainly relies on the contraction of the proposed linearization strategy and the contraction of the algebraic solver (for more details, see Section 1.2.6 below). □

**(Quasi-)Pythagorean estimate and compactness.** For the proof of linear convergence (and thus optimal rates), we exploit that the semilinear model problem (1.2) admits a *(quasi-)Pythagorean estimate*. There are two considered metrics: First, the difference of

energy $\mathcal{E}$ from (1.4) and second, the energy norm $|||\cdot|||$ induced by the principal part of the PDE operator. The latter relies on *compactness* arguments, thus imposing further constraints on (GC) for the case of $d = 3$, namely:

> There exist $R > 0$ and $N \in \mathbb{N}$ with $N \in \{2, 3\}$ for $d = 3$ such that
>
> $$|b^{(N)}(\xi)| \leq R \quad \text{for all } \xi \in \mathbb{R}. \tag{CGC}$$

> **Proposition 1.4.** *Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and that $\mathcal{X}_H \subset \mathcal{X}$.*
> (i) *Additionally, we suppose (POT). Then, the energy (1.4) satisfies that all differences appearing in*
>
> $$\mathcal{E}(v_H) - \mathcal{E}(u^\star) = [\mathcal{E}(v_H) - \mathcal{E}(u_H^\star)] + [\mathcal{E}(u_H^\star) - \mathcal{E}(u^\star)] \quad \text{for all } v_H \in \mathcal{X}_H, \tag{O}$$
>
> *are nonnegative.*
> (ii) *Under a suitable notion of compactness of the nonlinearity (CGC), we get a weaker result for the energy norm $|||\cdot|||$: For every $0 < \varepsilon < 1$, there exists a sufficiently fine discrete space $\mathcal{X}_H$ such that, for all spaces $\mathcal{X}_h$ with $\mathcal{X}_h \supseteq \mathcal{X}_H$, there holds quasi-orthogonality, i.e.,*
>
> $$\frac{1}{1+\varepsilon} |||u^\star - v_h|||^2 \leq |||u^\star - u_h^\star|||^2 + |||u_h^\star - v_h|||^2 \leq \frac{1}{1-\varepsilon} |||u^\star - v_h|||^2 \text{ for all } v_h \in \mathcal{X}_h. \tag{QO}$$

*Sketch of proof.* (i) This is immediate from the energy minimization (1.6) in Lemma 1.1 and its counterpart on the continuous level for (1.6).
(ii) The result is proved in [①GOA, Equation 2.78 below] in slightly modified form. The argument applies Galerkin orthogonality and nestedness of spaces for $u_H^\star \in \mathcal{X}_H \subseteq \mathcal{X}_h$, but this can be applied to any $v_h \in \mathcal{X}_h$ (instead of $u_H^\star$). □

## 1.2 Adaptive FEM with linearization and algebraic solver

This section starts with the concept of (mesh) adaptivity in Section 1.2.1. Sections 1.2.2–1.2.4 are devoted to the discussion of the involved modules of standard AFEM. Section 1.2.5 gives an overview of the literature on standard AFEM. Section 1.2.6 extends the standard AFEM to AILFEM, where linearization and an algebraic solver are included into the SOLVE-module of the standard AFEM routine. We conclude with a literature overview on AILFEMs in Section 1.2.7.

### 1.2.1 The concept of adaptivity

The discrete finite element space $\mathcal{X}_H$ is supposed to admit an *a priori* bound depending on the mesh-refinement level $H := \max_{T \in \mathcal{T}_H} |T|^{1/d}$ of the form

$$\|u^\star - u_H^\star\|_{H^1_0(\Omega)} = O(H^r) \quad \text{as} \quad H \to 0, \tag{1.10}$$

where $r > 0$ denotes the *rate* of the (global) approximation. The generic approach is to uniformly reduce the mesh size $H$ of the domain $\mathcal{X}_H$, where the number of elements grow


**Figure 1.1:** AFEM loop where linearization and (possibly) an algebraic solver is steered.

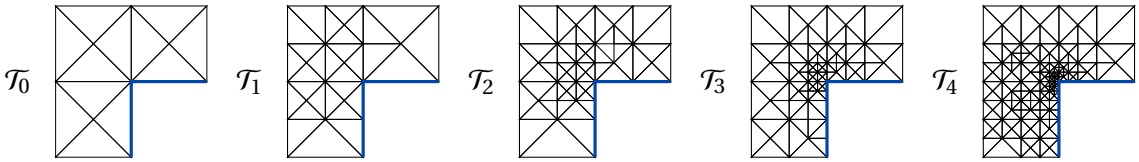



**Figure 1.2:** First meshes in the sequence of adaptively refined meshes for the problem from Experiment 1.5. The refinement focuses on the reentrant corner at $(0,0)$, where the unknown exact solution possesses a singularity.

exponentially. For sufficiently smooth solutions, the convergence rate is only bounded by the polynomial degree of $\mathcal{X}_H$. In nonconvex domains, the geometry leads to singularities at the reentrant corners and the convergence behavior may be spoiled by such (local point-) singularities. Uniform refinement has no means of localization of singularities by only steering the global mesh-refinement level $H$. Consequently, the global refinement requires additional computational effort to efficiently resolve the singularity and overall leads to a deteriorated convergence rate; see [Gri11] for nonconvex geometries and Experiment 1.5 below. By introducing a computable local error measure $\eta_H(T)$ of the approximation error on all triangles $T \in \mathcal{T}_H$ that does not rely on the unknown solution $u^\star$, such *a posteriori* information can indeed detect singularities and steer a localized mesh-refinement feedback loop as displayed in Figure 1.1. This is the concept of mesh *adaptivity*.

The feedback loop from Figure 1.1 can be described as follows: For a given triangulation $\mathcal{T}_H$, the module `SOLVE & ESTIMATE` steers the linearization and the algebraic solver and computes $\eta_H(T)$ for all $T \in \mathcal{T}_H$. The module `MARK` singles out elements, where the local contributions are large, e.g., by employing the Dörfler marking criterion (1.19) below that is first found in the seminal contribution [Dör96]. The `REFINE` module bisects the marked elements and performs a mesh-closure step to avoid hanging nodes. Overall, a new triangulation $\mathcal{T}_h$ is obtained by *a posteriori* information on potential singularities.

A sequence of adaptively refined meshes is depicted in Figure 1.2, where the schematic loop from Figure 1.1 is carried out by a simplified version of Algorithm 1.8 below; cf. Experiment 1.5 for further details.

With the mesh-level index $\ell$ and an approximation $u_\ell \approx u_\ell^\star$ computed by Algorithm 1.8 below, the decay rate of the energy error $|||u^\star - u_\ell|||$ over the number of degrees of freedom of $\mathcal{X}_\ell$ is a suitable measure for a fair comparison of uniform mesh refinement and adaptively refined meshes.

**Experiment 1.5** (Convergence of uniform vs. adaptive refinement)**.** *We consider the L-shaped domain* $\Omega = (-1,1)^2 \setminus [0,1) \times [-1,0) \subset \mathbb{R}^2$*, where the boundary consists of a homogeneous Dirichlet part* $\partial\Omega_D$ *(highlighted in blue and bold in Figure 1.2) and an inho-*

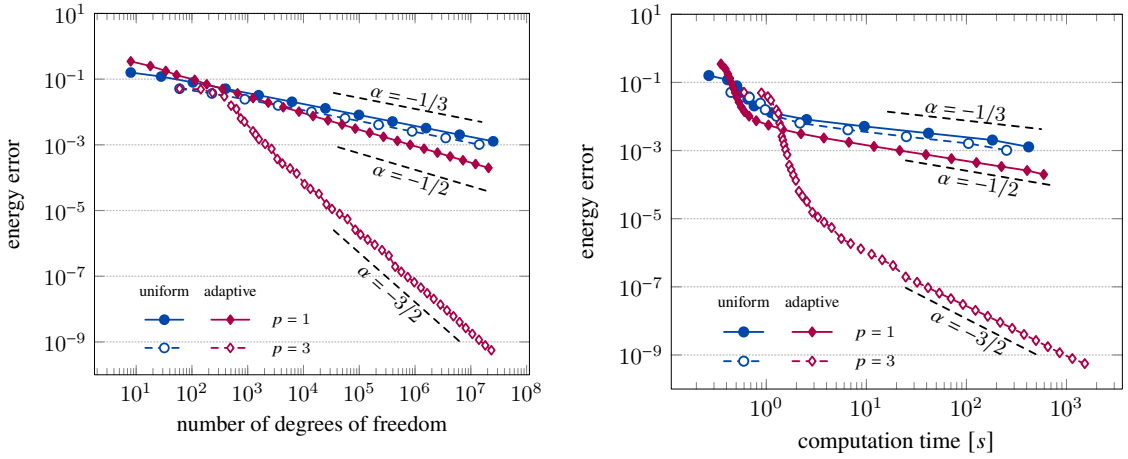



**Figure 1.3:** Adaptive (diamond, red) vs. uniform mesh refinement (circle, blue): Plot of the approximation error $|||u^\star - u_\ell|||$ for an approximation $u_\ell \approx u^\star$ computed by Algorithm 1.8 over the number of degrees of freedom (left) and over computation time in seconds (right).

*mogeneous Neumann part $\partial\Omega_N$ such that $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N = \partial\Omega \setminus \partial\Omega_D$. With the normal derivative $\partial_n$, we solve the Laplace problem*

$$-\Delta u^\star = 0 \quad \text{in } \Omega \quad \text{subject to} \quad u^\star = 0 \quad \text{on } \partial\Omega_D \quad \text{and} \quad \partial_n u^\star = g_N \quad \text{on } \partial\Omega_N.$$

*The exact solution reads $u^\star(x) = r^{2/3} \sin(2/3\,\varphi)$ for $x \in \Omega$ in polar coordinates $(r, \varphi) \in \mathbb{R}_{\geq 0} \times [0, 2\pi)$ and is used to determine $g_N$. Its derivative has the generic point singularity at the reentrant corner $(0,0)$. As refinement strategies, we compare uniform mesh refinement and an adaptive mesh-refining algorithm.*

*An empirical investigation of the convergence behavior is shown in Figure 1.3. On the left, we plot the energy error over the number of degrees of freedom. In practice, one is often more interested in measuring computation time in seconds. This is displayed in Figure 1.3 (right). The data points represent the iterates of the algebraic loop (more precisely: $\|u^\star - u_\ell^j\|_{H_0^1(\Omega)}$, where $\ell$ is the mesh-refinement index and $j$ the final algebraic solver index; the index set is defined in the spirit of Q from (1.20) below with a void linearization loop). In both cases, uniform refinement (circle, blue) converges with suboptimal rate $r = -1/3$ regardless of the polynomial degree $p \in \{1, 3\}$ of the ansatz space $\mathcal{S}^p_0(\Omega)$. Adaptive mesh refinement (diamond, red) restores the optimal rate $r = -p/2$ with respect to the number of degrees of freedom and computation time after a short preasymtotic phase.*

### 1.2.2 Module REFINE — Newest-vertex bisection

The *newest-vertex bisection* algorithm (NVB) is a local bisection method from [Sew72; Mit91; Ste08; DGS23] that preserves conformity and $\gamma$-shape regularity, which can be understood as a lower bound on the angles of all triangles. For the introduction, we restrict ourselves to the case $d = 2$, where NVB and the mesh closure can be presented as an

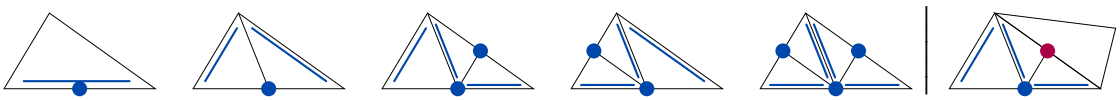



**Figure 1.4:** (1): Marked element $T$ with refinement edge $\texttt{refEdge}(T) \in \mathcal{M}_0$ (blue). (2): Refinement of (1) by NVB. Due to the mesh-closure step to avoid hanging nodes, all three edges might be marked for refinement; (1) for one marked edge, (3)–(4) for two marked edges after performed refinement, and (5) for three marked edges and performed refinement. (6): A hanging node in the neighboring triangle is highlighted in red (6).

**Figure 1.5:** Graphical illustration that NVB preserves $\gamma$-shape regularity. Left: Arbitrary triangle with refinement edge at the bottom. Center left to right: Successive NVB iterates of the children elements. The only possible similarity classes of triangles are highlighted in different colors.

inductive algorithm [KPP13].

For a triangle $T = \operatorname{conv}(z_0, z_1, z_2)$, where $\operatorname{conv}(z_0, z_1, z_2)$ denotes the convex hull of $\{z_0, z_1, z_2\} \in \Omega$, we use the convention that $\texttt{refEdge}(T) := \operatorname{conv}(z_1, z_2)$ is the edge opposite of $z_0$. To refine $T$ along $\texttt{refEdge}(T)$, we introduce the midpoint $m_T := \frac{z_1+z_2}{2}$. Then, $T$ is *refined by bisection* into $T = T_1 \cup T_2$, where $T_1 := \operatorname{conv}(m_T, z_1, z_0)$ and $T_2 := \operatorname{conv}(m_T, z_1, z_2)$ with $\frac{|T|}{2} = |T_1| = |T_2|$; see the two leftmost triangles in Figure 1.4 for the bisection of the triangle $T$ along the refinement edge and the midpoint $m_T$ (circle). The next $\texttt{refEdge}(T_{1,2})$ is opposite to $m_T$, justifying the name newest-vertex bisection.

Moreover, NVB also includes a procedure to avoid hanging nodes that may appear in the refinement process; see Figure 1.4(6), thus additionally refining elements that are not marked. This is denoted by $\mathcal{T}_{\ell+1} := \texttt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$, where $\mathcal{M}_\ell$ are the marked elements on level $\ell$.

We remark that $\gamma$-shape regularity of the input mesh $\mathcal{T}_\ell$, i.e., a uniform lower bound on the interior angles of all $T \in \mathcal{T}_\ell$, is respected by NVB, since at most four different similarity classes of children elements (and, thus, only finitely many interior angles) may occur in sequences of conforming triangulations generated by NVB refinement. This observation is depicted in Figure 1.5.

**Remark 1.6** (NVB in higher dimensions)**.** *We refer to [Mau95; Tra97; Ste08] for $d \geq 3$, where the NVB algorithm is formulated recursively. Until very recently, the termination of this recursive formulation was only ensured by the admissibility condition imposed on the initial triangulation $\mathcal{T}_0$ originating from [BDD04]. However, this can be circumvented by an initialization strategy proposed in the recent preprint [DGS23].*

We denote with $\mathbb{T}(\mathcal{T}_H)$ the set of all triangulations $\mathcal{T}_h$ that are the result of finitely many steps of newest-vertex bisection from $\mathcal{T}_H$, i.e., we write $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ if there exists $n \in \mathbb{N}$

such that $\mathcal{T}_n = \mathcal{T}_h$ and marked elements $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ with $\mathcal{T}_{\ell+1} = \texttt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ with for all $\ell = 1, \dots, n-1$ where $\mathcal{T}_1 = \mathcal{T}_H$. Moreover, we use $\mathcal{T}_0$ to denote a fixed initial mesh and we may abbreviate $\mathbb{T} = \mathbb{T}(\mathcal{T}_0)$.

NVB admits the following *fine properties of mesh refinement*, which will be applied to obtain optimal convergence rates. To this end, for $\mathcal{T}, \mathcal{T}' \in \mathbb{T} = \mathbb{T}(\mathcal{T}_0)$, we call the triangulation $\mathcal{T} \oplus \mathcal{T}' := \operatorname{argmin}_{\widetilde{\mathcal{T}} \in \mathbb{T}(\mathcal{T}) \cap \mathbb{T}(\mathcal{T}')} \#\widetilde{\mathcal{T}}$ the *coarsest common refinement* of $\mathcal{T}$ and $\mathcal{T}'$.

**(R1) children estimate:** For arbitrary $\mathcal{T}_H \in \mathbb{T}$ and arbitrary $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$, there exists $C_{\text{child}} \in \mathbb{N}$ such that

$$\#(\mathcal{T}_H \setminus \mathcal{T}_h) + \#\mathcal{T}_H \leq \#\mathcal{T}_h \leq C_{\text{child}} \#(\mathcal{T}_H \setminus \mathcal{T}_h) + \#(\mathcal{T}_H \cap \mathcal{T}_h). \tag{R1}$$

**(R2) overlay estimate:** For arbitrary $\mathcal{T}_H \in \mathbb{T}$ and a refinement $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$, we have that

$$\#(\mathcal{T}_H \oplus \mathcal{T}_h) \leq \#\mathcal{T}_H + \#\mathcal{T}_h - \#\mathcal{T}_0. \tag{R2}$$

**(R3) closure estimate:** For any sequence of triangulations $(\mathcal{T}_\ell)_{\ell \in \mathbb{N}_0}$ generated by subsets of marked elements $(\mathcal{M}_\ell)_{\ell \in \mathbb{N}_0}$ such that $\mathcal{T}_{\ell+1} := \texttt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ for $\ell \in \mathbb{N}_0$, it holds that

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 \leq C_{\text{mesh}} \sum_{\ell'=0}^{\ell-1} \#\mathcal{M}_{\ell'} \quad \text{for all } \ell \in \mathbb{N}_0, \tag{R3}$$

where $C_{\text{mesh}} \geq 1$ is independent of the sequences $(\mathcal{T}_\ell)_{\ell \in \mathbb{N}_0}$ and $(\mathcal{M}_\ell)_{\ell \in \mathbb{N}_0}$, but depends only on $\mathcal{T}_0$.

The refinement is based on NVB and of constant cost for each element. Since there are at most finitely many children [GSS14] and there holds a mesh-closure estimate [BDD04; Ste08], the overall cost are thus of order $\mathcal{O}(\#\mathcal{T}_H)$ to generate $\mathcal{T}_h = \texttt{refine}(\mathcal{T}_H, \mathcal{M}_H)$.

### 1.2.3 Module ESTIMATE

As the local error measure for the model problem (1.1) (with homogeneous Dirichlet boundary conditions), we introduce the *residual a posteriori error estimators* following [AO00; Ver13]. Residual error estimators are motivated by the fact that the (computationally not available) error $u^\star - u_H^\star$ can be estimated via the residual in the dual space $H^{-1}(\Omega)$ by a discrete, computable quantity $\eta_H(u_H^\star)$.

Suppose that $\boldsymbol{A}|_T \in [W^{1,\infty}(T)]^{d\times d}_{\text{sym}}$ and $\boldsymbol{f}|_T \in [W^{1,\infty}(T)]^d$ for all $T \in \mathcal{T}_0$. With $h_T = |T|^{1/d}$, elementwise integration by parts of the residual $F - \mathcal{A}(v_H)$ with $v_H \in \mathcal{X}_H$ give rise to the elementwise contribution

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

Moreover, we abbreviate, for $\mathcal{U}_H \subseteq \mathcal{T}_H$,

$$\eta_H(\mathcal{U}_H, v_H)^2 := \sum_{T \in \mathcal{U}_H} \eta_H(T, v_H)^2 \tag{1.11}$$

and the global contribution by $\eta_H(v_H)^2 := \eta_H(\mathcal{T}_H, v_H)^2$.

To ensure optimal convergence rates, four *abstract conditions of the error estimator* are required. The analysis in its axiomatic form has been introduced by [CFPP14]. For nonlinear problems, the proof of stability (A1) below requires new ideas and, in particular, the stability constant inherits the local Lipschitz continuity (LIP) of our abstract framework, whereas (A2)–(A4) follow from standard arguments in [CFPP14].

**Proposition 1.7** ([①GOA, Proposition 2.15 below]). *Let $\mathcal{T}_H \in \mathbb{T}$ and $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$. The standard residual a posteriori error estimator (1.11) for the semilinear model problem satisfies the following properties:*

**(A1) stability:** For all $\vartheta > 0$, there exists $C_{\mathrm{stab}}[\vartheta] > 0$ such that for all $v_h \in \mathcal{X}_h$ and $v_H \in \mathcal{X}_H$ with $\max\{|||v_h|||, |||v_H|||\} \leq \vartheta$, it holds that

$$\big|\eta_h(\mathcal{T}_h \cap \mathcal{T}_H, v_h) - \eta_H(\mathcal{T}_h \cap \mathcal{T}_H, v_H)\big| \leq C_{\mathrm{stab}}[\vartheta]\, |||v_h - v_H|||. \tag{A1}$$

**(A2) reduction:** With $0 < q_{\mathrm{red}} := 2^{-1/(2d)} < 1$ and provided that simplices are refined by NVB, there holds, for all $v_H \in \mathcal{X}_H$ and all $w \in H^1_0(\Omega)$, that

$$\eta_h(\mathcal{T}_h \backslash \mathcal{T}_H, v_H) \leq q_{\mathrm{red}}\, \eta_H(\mathcal{T}_H \backslash \mathcal{T}_h, v_H). \tag{A2}$$

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

$$|||u^\star - u_H^\star||| \leq C_{\mathrm{rel}}\, \eta_H(u_H^\star). \tag{A3}$$

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

$$|||u_h^\star - u_H^\star||| \leq C_{\mathrm{drel}}\, \eta_H(\mathcal{T}_H \backslash \mathcal{T}_h, u_H^\star). \tag{A4}$$

We remark that the constants $C_{\mathrm{stab}}[\vartheta]$, $C_{\mathrm{rel}}$, and $C_{\mathrm{drel}}$ depend on $\gamma$-shape regularity of the mesh and that $C_{\mathrm{stab}}[\vartheta]$ and $C_{\mathrm{drel}}$ depend additionally on the polynomial degree $p$. We also highlight that (A3)–(A4) hold only for the exact solution $u^\star$ and Galerkin solutions $u_H^\star$, $u_h^\star$, respectively.

### 1.2.4 Module `MARK`

The marking procedure determines elements with large estimator contributions for refinement. In adaptive finite element methods — in particular in the context of rate-optimality — the Dörfler marking from [Dör96] is frequently used and reads as follows. For an adaptivity parameter $0 < \theta \leq 1$, we seek a (possibly nonunique) set of marked elements $\mathcal{M}_H \subseteq \mathcal{T}_H$

such that

$$\theta\,\eta_H^2 \le \eta_H(\mathcal{M}_H)^2. \tag{1.12}$$

Among the many sets $\mathcal{M}_\ell$ that satisfy (1.12), we choose $\mathcal{M}_\ell$ with *quasi-minimal* cardinality, i.e., for a fixed constant $C_{\mathrm{mark}} \ge 1$ that does not depend on the mesh-refinement index such that

$$\#\mathcal{M}_H \le C_{\mathrm{mark}} \min_{\mathcal{U}_H^\star \in \mathbb{M}_H} \mathcal{U}_H^\star \quad \text{with} \quad \mathbb{M}_H := \{\mathcal{U}_H \subseteq \mathcal{T}_H \mid \theta\,\eta_H^2 \le \eta_H(\mathcal{U}_H)^2\}. \tag{1.13}$$

Then, the choice of $\theta = 1$ results in uniform refinement, whereas a small adaptivity parameter $0 < \theta \ll 1$ marks very few elements with largest error indicators.

The standard approach to determine a set of marked elements is to sort the error contributions [Dör96]. However, this would lead to the minimal cardinality in loglinear complexity. For the price of marking slightly too many elements, [Ste07] singles out a set of quasi-minimal cardinality $C_{\mathrm{mark}} = 2$ in linear complexity based on a bin-sort approach. An adaptation of quickselect in [PP20] realizes Dörfer marking at linear cost with $C_{\mathrm{mark}} = 1$.

Concerning the proof of optimal convergence rates, the quasi-minimality of Dörfler marking (1.13) is sufficient [Dör96] and even necessary [Ste07] (cf. Proposition 1.11 below).

### 1.2.5 Literature on adaptive FEM

Improving the computational accuracy via a feedback loop as in Figure 1.1 already appears in early works such as, e.g., [BR79; BV84; ZZ87]. The proof of (plain and linear) convergence of an adaptive FEM is significantly more challenging than for uniform refinement, since the maximal mesh size $h_T = |T|^{1/d}$ may not tend to zero uniformly (cf. (1.10) and Figure 1.2 for an illustration of that). While a first convergence result was presented in [BV84] for $d = 1$, its proof remained open for $d \ge 2$ until [Dör96; MNS00]. At about the same time, an overview on available error estimators for AFEM appeared in [AO00].

After the introduction of suitable approximation classes in the context of adaptive wavelet methods [CDD01; CDD03] and by means of an additional coarsening step of the mesh, [BDD04] proved rate-optimality of an AFEM for the 2D Poisson model problem for the first time. The coarsening procedure was circumvented in [Ste07] for the 2D Poisson model problem. In the publication [MSV08], other marking strategies such as the maximum marking criterion and their implications on convergence of AFEM are analyzed.

The paper [DK08] frees the proposed AFEM of the additional separate marking for the oscillations $\mathrm{osc}_H(u_H^\star)$. This strategy is refined in [CKNS08], which observed that

$$|||u^\star - u_H^\star||| + \mathrm{osc}_H(u_H^\star)^2 \approx \eta_H(u_H^\star)^2 \approx |||u^\star - u_H^\star||| + \mu\,\eta_H(u_H^\star)^2 \quad \text{with } \mu > 0.$$

The term on the left-hand side is the *total error*. It consists of the approximation quality of $u^\star$ and the data approximation properties encoded in $\mathrm{osc}_H(u_H^\star)$. The total error was also used to define the approximation class of $u^\star$; cf. (1.32) below. The term on the right-hand side is the *quasi-error* that is contracted by the AFEM for a suitable choice of $\mu > 0$. Thus,

no inner loop for data approximation is necessary in the setting of symmetric second-order elliptic PDEs (if the initial triangulation resolves the geometry). Fostered by the presented breakthroughs, many papers on particular schemes or applications appeared. For instance, the paper [KS11] investigated and proved convergence for other (locally) equivalent estimators. Moreover, extracting the essential (estimator) properties (A1)–(A4) and (QO) gave rise to an axiomatic framework that relies solely on upper bounds of the error estimator [CFPP14].

The extension of the framework to nonlinear problems goes back to, e.g., [Vee02; DK08; BDK12] for the p-Laplacian. Moreover, we refer to [GMZ12] for strongly-monotone and globally Lipschitz continuous quasilinear PDEs, and to [FFP14] for second-order nonsymmetric PDEs and some nonlinear problems.

### 1.2.6 AFEM with linearization and linear solver

The solution of the nonlinear discrete problem (1.2) is an involved task. One way to solve nonlinear problems is by using a fixed-point iteration to linearize the nonlinear equation. We extend the AFEM setting which usually supposes an exact solution of the discrete problem to Algorithm 1.8 below. This Algorithm 1.8 has an inner loop which employs the Zarantonello iteration to linearize the nonlinear equation and steers adaptively the number of linearization steps. Moreover, we employ an algebraic solver to solve the linearized system, we refer to this extended AFEM as *adaptive iteratively linearized FEM* (AILFEM).

The linearization and the algebraic solver that is used to efficiently solve the linearized discrete problem can also be stopped adaptively with *a posteriori* information. Schematically, this leads to nested loops that are depicted in Figure 1.6, where the *nested loops and their hierarchy* are indicated by the boxes — discretization (with index $\ell$), linearization (with index $k$), and algebraic solver (with index $j$).

In all presented algorithms in this thesis, discretization ($\ell$, blue) incorporates the modules `MARK` and `REFINE`. When including additional errors that stem from linearization ($k$, red), this is realized as a nested loop that leads to a symmetric and positive definite (SPD) problem. The expensive SPD problem is solved by an algebraic solver. This constitutes yet another nested loop ($j$, green).

The Zarantonello iteration is used as a *linearization method*. It is particularly attractive for two reasons: On each level, only a Laplace-type problem has to be solved. Moreover, the assembly of the Laplace system can be done only once at each mesh level $\ell$ and does not depend on the computed iterates. For a damping parameter $\delta > 0$ and given $w_H \in \mathcal{X}_H$, the Zarantonello update $\Phi_H(\delta; w_H) \in \mathcal{X}_H$ solves

$$\langle\!\langle \Phi_H(\delta; u_H)\,,\, v_H \rangle\!\rangle = \langle\!\langle u_H\,,\, v_H \rangle\!\rangle + \delta \left[ F(v_H) - \langle \mathcal{A}u_H\,,\, v_H \rangle \right] \quad \text{for all } v_H \in \mathcal{X}_H. \tag{1.14}$$

The Lax–Milgram theorem proves existence and uniqueness of $\Phi_H(\delta; u_H)$, i.e., the Zarantonello operator $\Phi_H(\delta;\cdot)\colon \mathcal{X}_H \to \mathcal{X}_H$ is well-defined. In particular, $u_H^\star = \Phi(\delta; u_H^\star)$ is the unique fixed point of $\Phi_H(\delta;\cdot)$ for any damping parameter $\delta > 0$. For a sufficiently small damping parameter $\delta > 0$, the Zarantonello iteration is norm-contractive; cf. [Zei90, Section 25.4]. This is the main ingredient to show that the energy norm of two succes-

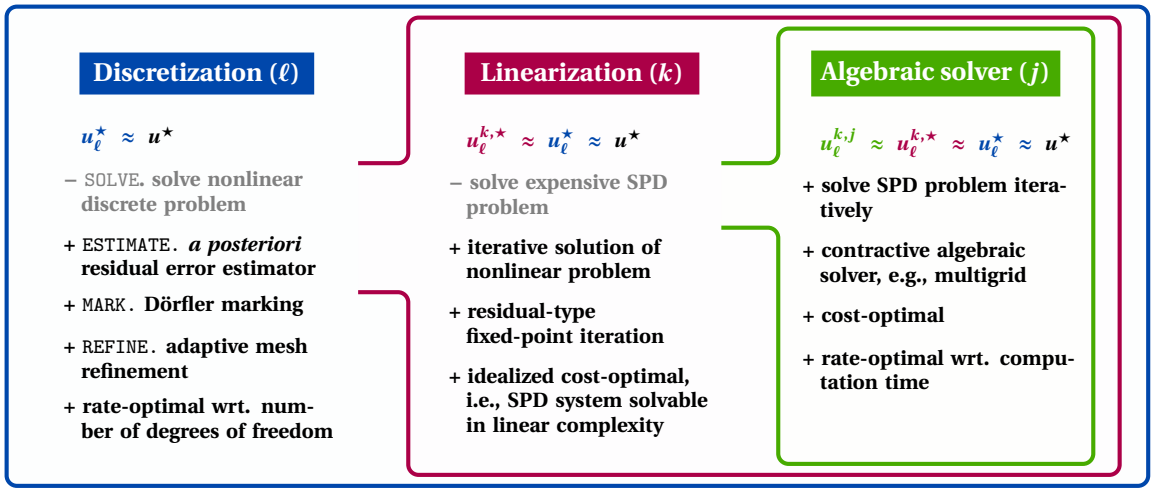



**Figure 1.6:** Illustration of a two-fold nested loop in adaptive mesh-refinement algorithm with linearization and algebraic solver.

sive iterates is bounded. For a detailed discussion of the Zarantonello iteration, we refer to [②AIL1, Section 3.2.2–3.2.4 below].

The Zarantonello linearization leads to an SPD problem (1.14). Solving large SPD problems in linear complexity requires an advanced solving procedure. To this end, we employ an *iterative algebraic solver* with process function $\Psi_H : \mathcal{X}' \times \mathcal{X}_H \to \mathcal{X}_H$ to solve the linearized system (1.14). More precisely, given a linear functional $\varphi \in \mathcal{X}'$ and an approximation $w_H \in \mathcal{X}_H$ of the exact solution $w_H^\star \in \mathcal{X}_H$ to

$$\langle\!\langle w_H^\star , v_H \rangle\!\rangle = \varphi(v_H) \quad \text{for all } v_H \in \mathcal{X}_H, \tag{1.15}$$

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

$$|||w_H^\star - \Psi_H(\varphi; w_H)||| \le q_{\mathrm{alg}} |||w_H^\star - w_H||| \quad \text{for all } w_H \in \mathcal{X}_H. \tag{1.16}$$

To simplify notation in case of a complicated right-hand side $\varphi$ (as for the Zarantonello iteration (1.14)), we shall write $\Psi_H(w_H^\star; \cdot)$ instead of $\Psi_H(\varphi; \cdot)$, even though $w_H^\star$ is unknown and is never computed.

Examples of norm-contractive solvers include optimally preconditioned conjugate gradient methods [CNX12] or optimal geometric multigrid methods; see, e.g., [WZ17] for fixed $p \in \mathbb{N}$ or [IMPS23] for an $hp$-robust multigrid method, where the latter will be employed for some of the numerical experiments.

Define $j := k := \ell := 0$, where $\ell$ is the counter for mesh refinement, $k = k[\ell]$ is the counter for linearization, and $j = j[\ell, k]$ is the counter for the algebraic solver. Algorithm 1.8 presents the quasi-optimal AILFEM strategy.

**Algorithm 1.8: adaptive iteratively linearized FEM (AILFEM)**

**Input:** $\mathcal{T}_0$ conforming mesh, initial guess $u_0^{0,0} \in \mathcal{X}_H$, marking parameters $0 < \theta \le 1$ and $C_{\mathrm{mark}} \ge 1$ for Dörfler marking, and Zarantonello damping parameter $\delta > 0$.
**Adaptive loop:** For all $\ell = 0, 1, 2, \dots$, repeat the following steps (I)–(III):

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

(a) Set $u_\ell^{k,0} := u_\ell^{k,0}$ and define, for theoretical reasons, the exact solution of the linearization iterate $u_\ell^{k,\star} := \Phi_\ell(\delta; u_\ell^{k,0})$ from (1.14).

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

(i) Compute $u_\ell^{k,j} := \Psi(u_\ell^{k,\star}; u_\ell^{k,j-1}) \approx u_\ell^{k,\star}$ from (1.15) and $\eta_\ell(u_\ell^{k,j})$.

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

$$\text{the algebraic error} \quad |||u_\ell^{k,\star} - u_\ell^{k,j}||| \quad \text{is sufficiently small.} \tag{1.17}$$

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

$$\text{the linearization error} \quad |||u_\ell^{\star} - u_\ell^{k,\underline{j}}||| \quad \text{is sufficiently small.} \tag{1.18}$$

(II) `MARK`. With $\mathbb{M}_\ell[\theta, u_\ell^{k,\underline{j}}] := \{\mathcal{U}_\ell \subseteq \mathcal{T}_\ell \mid \theta\, \eta_\ell(u_\ell^{k,\underline{j}})^2 \le \eta_\ell(\mathcal{U}_\ell, u_\ell^{k,\underline{j}})^2\}$, determine a set $\mathcal{M}_\ell \in \mathbb{M}_\ell[\theta, u_\ell^{k,\underline{j}}]$ from (1.13) with quasi-minimal cardinality

$$\#\mathcal{M}_\ell \le C_{\mathrm{mark}} \min_{\mathcal{U}_\ell \in \mathbb{M}_\ell[\theta, u_\ell^{k,\underline{j}}]} \#\mathcal{U}_\ell. \tag{1.19}$$

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

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

For the analysis of Algorithm 1.8, we define the index set

$$\mathcal{Q} := \{(\ell, k, j) \in \mathbb{N}_0^3 \mid u_\ell^{k,j} \text{ is used in Algorithm 1.8}\}, \tag{1.20}$$

where, for any $(\ell, 0, 0) \in \mathcal{Q}$, the stopping indices are defined in coincidence with Algorithm 1.8 as

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

We introduce the total step counter i.e., for two indices $(\ell,k,j),(\ell',k',j')\in\mathcal{Q}$, it holds that

$$|\ell,k,j|\le|\ell',k',j'| \quad :\Longleftrightarrow \quad (\ell,k,j) \text{ appears not later than } (\ell',k',j') \text{ in Algorithm 1.8.}$$

This provides indeed a lexicographic ordering with respect to the total step counter on $\mathcal{Q}$,

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

**Remark 1.9** (Linear complexity)**.** *Each module of Algorithm 1.8 is realizable in linear complexity:*

- `SOLVE & ESTIMATE`. *The employed Zarantonello linearization produces a linear system that is solved by means of the geometric multigrid method from [IMPS23] as a solver with linear complexity, i.e., each iterate $u_\ell^{k,j}$ can be obtained with $\mathcal{O}(\#\mathcal{T}_\ell)$ operations. The computation of the refinement indicators $\eta_\ell(T,u_\ell^{k,j})$ for all $T\in\mathcal{T}_\ell$ can be parallelized and done at the cost of $\mathcal{O}(\#\mathcal{T}_\ell)$.*
- `MARK`. *The employed Dörfler marking (and the involved determination of the marked elements $\mathcal{M}_\ell$) is indeed a linear complexity problem; see [Ste07] for quasi-minimality with $C_{\mathrm{mark}}=2$ and [PP20] for minimal cardinality.*
- `REFINE`. *The refinement of $\mathcal{T}_\ell$ is based on NVB and hence is of linear cost $\mathcal{O}(\#\mathcal{T}_\ell)$.*

The cumulative nature of the AILFEM suggests to consider the *computational cost* as a more restrictive measure than the degrees of freedom of the underlying space $\mathcal{X}_\ell$. The computational cost for obtaining $u_\ell^{k,j}$ depends on the whole history and, since each step is of linear complexity, the cost is proportional to the sum of the number of elements in each iteration. More formally, it holds that

$$\mathtt{cost}(\ell,k,j) := \sum_{\substack{(\ell',k',j'),(\ell,k,j)\in\mathcal{Q}\\ |\ell',k',j'|\le|\ell,k,j|}}\#\mathcal{T}_{\ell'} = \sum_{\ell'=0}^{\ell-1}\sum_{k'=1}^{\underline{k}[\ell']}\sum_{j'=1}^{\underline{j}[\ell',k']}\#\mathcal{T}_{\ell'}+\sum_{k'=1}^{k-1}\sum_{j'=1}^{\underline{j}[\ell,k']}\#\mathcal{T}_\ell+\sum_{j'=1}^{j-1}\#\mathcal{T}_\ell. \tag{1.22}$$

This subsection concludes with a discussion of possible stopping criteria in Algorithm 1.8. Recall the discrete exact solution $u_H^\star\approx u^\star$ from (1.3), the exact solution of the Zarantonello iteration $u_H^{k,\star}=\Phi_H(\delta;u_H^{k,0})\approx u_H^\star$ from (1.14), and the linear solver iterate $u_H^{k,j}=\Psi(u_H^{k,\star},u_H^{k,j-1})\approx u_H^{k,\star}$ from (1.15). By stability (A1) and reliability (A3), we have that

$$\begin{aligned}|||u^\star-u_\ell^{k,j}|||&\le|||u^\star-u_\ell^\star|||+|||u_\ell^\star-u_\ell^{k,j}|||\\&\lesssim\eta_\ell(u_\ell^{k,j})+|||u_\ell^\star-u_\ell^{k,\star}|||+|||u_\ell^{k,\star}-u_\ell^{k,j}|||,\end{aligned} \tag{1.23}$$

which controls the overall error from above by splitting it into discretization error, linearization error, and algebraic error. This splitting is used to derive the stopping criteria (1.17)–(1.18) in Algorithm 1.8. However, the exact solutions $u_\ell^\star$ and $u_\ell^{k,\star}$ are not available. A computable variant of the stopping criteria relies on the following heuristics: The lin-

earization error shall be dominated by the discretization error, and the algebraic error shall be dominated by the discretization and linearization errors.

The contraction of the algebraic solver (1.16) yields the *a posteriori* error estimate

$$|||u_\ell^{k,\star} - u_\ell^{k,j}||| \le \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} |||u_\ell^{k,j} - u_\ell^{k,j-1}||| \quad \text{for all } 1 \le j \le \underline{j}[\ell,k].$$

To bound the algebraic error by the discretization and linearization error, we ask for

$$\frac{1-q_{\mathrm{alg}}}{q_{\mathrm{alg}}} |||u_\ell^{k,\star} - u_\ell^{k,j}||| \le |||u_\ell^{k,j} - u_\ell^{k,j-1}||| \overset{!}{\le} \lambda_{\mathrm{alg}} \big[\lambda_{\mathrm{lin}}\, \eta_\ell(u_\ell^{k,j}) + |||u_\ell^{k,\underline{j}} - u_\ell^{k,0}|||\big], \tag{1.24}$$

where the second estimate can be checked in practice and is used in (1.17). The Zarantonello linearization is contractive for the exact solution [BIM+23, Equation (4.1)], i.e.,

$$|||u_\ell^\star - u_\ell^{k,\star}||| \le q_{\mathrm{Zar}}^\star |||u_\ell^\star - u_\ell^{k,0}||| \le q_{\mathrm{Zar}}^\star \big[|||u_\ell^\star - u_\ell^{k,\star}||| + |||u_\ell^{k,\star} - u_\ell^{k,\underline{j}}||| + |||u_\ell^{k,\underline{j}} - u_\ell^{k,0}|||\big].$$

We also have the *a posteriori* error estimate

$$\begin{aligned}(1-q_{\mathrm{Zar}}^\star)\, |||u_\ell^\star - u_\ell^{k,\star}||| &\le |||u_\ell^{k,\star} - u_\ell^{k,\underline{j}}||| + |||u_\ell^{k,\underline{j}} - u_\ell^{k,0}||| \\ &\overset{(1.24)}{\le} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} \lambda_{\mathrm{alg}} \big[\lambda_{\mathrm{lin}}\, \eta_\ell(u_\ell^{k,j}) + |||u_\ell^{k,\underline{j}} - u_\ell^{k,0}|||\big] + |||u_\ell^{k,\underline{j}} - u_\ell^{k,0}|||,\end{aligned}$$

where $|||u_\ell^{k,\underline{j}} - u_\ell^{k,0}||| \overset{!}{\le} \lambda_{\mathrm{lin}}\, \eta_\ell(u_\ell^{k,\underline{j}})$ can also be checked in practice for the final iterates of the $j$-loop. This is used to stop the $k$-loop. With a hidden constant that includes the constants from the previous estimates and also relies on stability (A1) and reliability (A3), this overall ensures that

$$|||u^\star - u_\ell^{\underline{k},\underline{j}}||| \lesssim (1 + \lambda_{\mathrm{alg}}\, \lambda_{\mathrm{lin}})\, \eta_\ell(u_\ell^{\underline{k},\underline{j}})$$

for the final iterate once both, the linearization and the algebraic solver have terminated.

**Remark 1.10.** *The stopping criteria* (1.17)–(1.18) *are adapted in the proposed AILFEMs to enforce algorithmically that enough linearization as well as algebraic solver steps are made. This is a crucial ingredient to uniform boundedness* (UB).

### 1.2.7 Literature on AILFEM

The inclusion of iterative solvers into AFEMs already goes back to [Ste07]. Under realistic assumptions on a generic iterative solver, AFEM with optimal complexity was first proven in [Ste07] for the Poisson model problem and [CG12] for the Poisson eigenvalue problem. Other contributions to the development of AILFEMs with either linearization or an inexact solver are, e.g., found in [BMS10; EEV11; AGL13; EV13; AW15; CW17]. To use the Zarantonello iteration as a numerical linearization strategy seems to go back to [CW17], and is also used for globally Lipschitz continuous nonlinearities in [GHPS18; HW20a; HW20b; GHPS21; HPW21; HPSV21].

We point out that [HW20a; HW20b; HPW21] also consider other common linearization strategies, namely the Kačanov iteration and a damped Newton method. These, however, are (so far) hard to use in the semilinear setting, since the bilinear forms associated with the linearization depend on the previous iterate and norm contraction may not hold. This prevents a main ingredient to uniform boundedness (UB) of the iterates.

The coupling of the Zarantonello linearization with an algebraic loop is analyzed in the own work [BIM+23] for nonsymmetric second-order linear elliptic PDEs and for strongly monotone (and globally Lipschitz continuous) model problems in [HPSV21; BFM+23].

## 1.3 Convergence with optimal rates

The ultimate goal of any numerical scheme is to drive down the error with the least computational effort possible. In this section, we sketch the interplay of model problem properties such as (quasi-) orthogonality (QO)/(O) and uniform boundedness (UB) with estimator properties (A1)–(A4), results on Dörfler marking, and fine properties of the mesh refinement (R1)–(R3).

### 1.3.1 Dörfler marking: sufficient and necessary

The Dörfler marking in the MARK procedure is sufficient to ensure convergence (and also optimal rates) of the finite element method. In some sense, [Ste07] observed that Dörfler marking is even necessary. Since these results follow from standard reasoning with minor modifications due to the local Lipschitz continuity, we include a short proof as these will not be proven in the main chapters.

**Proposition 1.11** (see, e.g., [CFPP14, Lemma 4.7, Lemma 4.12])**.** *Let $\ell \in \mathbb{N}_0$ be such that $\ell < \underline{\ell}$. Let $(\mathcal{T}_\ell)_\ell$ be the sequence of meshes generated by Algorithm 1.8. Let $\mathcal{A}$ satisfy (SM), (LIP), and (POT). Then, the following implications hold:*

(i) *Suppose (UB) with $C_{\mathrm{bnd}} = C_{\mathrm{bnd}}[M]$ for all final iterates of the two inner loops of Algorithm 1.8. Under the estimator properties (A1)–(A2) and for $u_\ell^{\underline{k},j} \in \mathcal{X}_\ell$ and $u_{\ell+1}^{\underline{k},j} \in \mathcal{X}_{\ell+1}$, the Dörfler marking (1.19) implies that*

$$\eta_{\ell+1}(u_{\ell+1}^{\underline{k},j}) \le q_\theta\, \eta_\ell(u_\ell^{\underline{k},j}) + C_{\mathrm{stab}}[C_{\mathrm{bnd}}]\,|||u_{\ell+1}^{\underline{k},j} - u_\ell^{\underline{k},j}||| \quad \text{with } 0 < q_\theta < 1, \tag{1.25}$$

*where $q_\theta := \big[1 - (1 - q_{\mathrm{red}}^2)\,\theta\big]^{1/2}$. Note that $q_\theta \to 1$ if $\theta \to 0$.*

(ii) *Suppose (A1) and (A4) as well as (UB) for the exact solution $u_H^\star$ with $C_{\mathrm{bnd}} = M$ from (1.7). Let $0 < \theta < \theta_{\mathrm{opt}} := (1 + C_{\mathrm{stab}}[2M]^2\, C_{\mathrm{drel}}^2)^{-1}$. Then, there exists $0 < q_{\mathrm{opt}} < 1$ such that*

$$\eta_{\ell+n}(u_{\ell+n}^\star)^2 \le q_{\mathrm{opt}}\, \eta_\ell(u_\ell^\star)^2 \implies \theta\, \eta_\ell(u_\ell^\star)^2 \le \eta_\ell(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+n}; u_\ell^\star)^2 \quad \text{for } n \in \mathbb{N}_0 \tag{1.26}$$

**Remark 1.12.** (i) *The term $|||u_{\ell+1}^{\underline{k},j} - u_\ell^{\underline{k},j}|||$ vanishes for $|\ell, \underline{k}, j| \to \infty$, either by a priori convergence in the discrete limit space $\mathcal{X}_{\underline{\ell}} := \bigcup_{\ell=0}^\infty \mathcal{X}_\ell$ if $\underline{\ell} = \infty$ or by the contraction of the Zaran-*

*tonello iteration if* $\underline{k}[\underline{\ell}] = \infty$. *The case* $\underline{j}[\underline{\ell}, \underline{k}] = \infty$ *is analytically not possible (cf. [③AIL2, Lemma 4.7 below]).*

(ii) *Dörfler marking is used for the final iterates of the* `SOLVE & ESTIMATE` *module. In case of an AILFEM with an exact solution of the linearization as in [②AIL1], we have* $u_\ell^{k,\star}$ *and* $u_{\ell+1}^{k,\star}$ *as the final iterates. In case of discretization only, i.e., no linearization and no algebraic solver, the final approximations are* $u_\ell^\star$ *and* $u_{\ell+1}^\star$.

*Proof of Proposition 1.11.* The first statement, in essence, is presented in [CKNS08]. The second statement, formulated for the error, goes back to [Ste07], while the formulation through the error estimator is first found in [CFPP14].

(i) Stability (A1) and reduction (A2) prove that

$$\begin{aligned}\eta_{\ell+1}(u_\ell^{\underline{k},\underline{j}})^2 &= \eta_{\ell+1}(\mathcal{T}_{\ell+1}\cap\mathcal{T}_\ell, u_\ell^{\underline{k},\underline{j}})^2 + \eta_{\ell+1}(\mathcal{T}_{\ell+1}\setminus\mathcal{T}_\ell, u_\ell^{\underline{k},\underline{j}})^2\\ &\le \eta_\ell(\mathcal{T}_{\ell+1}\cap\mathcal{T}_\ell, u_\ell^{\underline{k},\underline{j}})^2 + q_{\mathrm{red}}^2\,\eta_\ell(\mathcal{T}_\ell\setminus\mathcal{T}_{\ell+1}, u_\ell^{\underline{k},\underline{j}})^2\\ &= \eta_\ell(u_\ell^{\underline{k},\underline{j}})^2 - (1-q_{\mathrm{red}}^2)\,\eta_\ell(\mathcal{T}_\ell\setminus\mathcal{T}_{\ell+1}, u_\ell^{\underline{k},\underline{j}})^2.\end{aligned}$$

Dörfler marking (1.19) and refinement of (at least) all marked elements lead to

$$\theta\,\eta_\ell(u_\ell^{\underline{k},\underline{j}})^2 \le \eta_\ell(\mathcal{M}_\ell, u_\ell^{\underline{k},\underline{j}})^2 \le \eta_\ell(\mathcal{T}_\ell\setminus\mathcal{T}_{\ell+1}, u_\ell^{\underline{k},\underline{j}})^2.$$

Combining the last two estimates leads to

$$\eta_{\ell+1}(u_\ell^{\underline{k},\underline{j}}) \le q_\theta\,\eta_\ell(u_\ell^{\underline{k},\underline{j}}) \quad\text{with}\quad 0 < q_\theta := \big[1-(1-q_{\mathrm{red}}^2)\,\theta\big]^{1/2} < 1.$$

Combined with $\eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{j}}) \le \eta_{\ell+1}(u_\ell^{\underline{k},\underline{j}}) + C_{\mathrm{stab}}[C_{\mathrm{bnd}}]\,|||u_{\ell+1}^{\underline{k},\underline{j}} - u_\ell^{\underline{k},\underline{j}}|||$ due to stability (A1) and (UB) for the final iterates $u_\ell^{\underline{k},\underline{j}}$ and $u_{\ell+1}^{\underline{k},\underline{j}}$ yields (1.25).

(ii) Since $|||u_{\ell+n}^\star - u_\ell^\star||| \le 2M$, the Young inequality with $\delta > 0$ shows that

$$\begin{aligned}\eta_\ell(u_\ell^\star)^2 &= \eta_\ell(\mathcal{T}_\ell\setminus\mathcal{T}_{\ell+n}; u_\ell^\star)^2 + \eta_\ell(\mathcal{T}_\ell\cap\mathcal{T}_{\ell+n}; u_\ell^\star)^2\\ &\overset{(A1)}{\le} \eta_\ell(\mathcal{T}_\ell\setminus\mathcal{T}_{\ell+n}; u_\ell^\star)^2 + (1+\delta)\,\eta_{\ell+n}(u_{\ell+n}^\star)^2 + (1+\delta^{-1})\,C_{\mathrm{stab}}[2M]^2\,|||u_\ell^\star - u_{\ell+n}^\star|||^2\\ &\overset{(A4)}{\le} \eta_\ell(\mathcal{T}_\ell\setminus\mathcal{T}_{\ell+n}; u_\ell^\star)^2 + (1+\delta)\,q_\theta^2\,\eta_\ell(u_\ell^\star)^2 + (1+\delta^{-1})\,C_{\mathrm{stab}}[2M]^2\,C_{\mathrm{drel}}^2\,\eta_\ell(\mathcal{T}_\ell\setminus\mathcal{T}_{\ell+n}; u_\ell^\star)^2.\end{aligned}$$

Rearrangement yields

$$\frac{1-(1+\delta)\,q_{\mathrm{opt}}^2}{1+(1+\delta^{-1})\,C_{\mathrm{stab}}[2M]^2\,C_{\mathrm{drel}}^2}\,\eta_\ell(u_\ell^\star)^2 \le \eta_\ell(\mathcal{T}_\ell\setminus\mathcal{T}_{\ell+n}; u_\ell^\star)^2. \tag{1.27}$$

Choosing $\delta > 0$ and afterwards $0 < q_{\mathrm{opt}} < 1$ in a way that ensures

$$\theta \le \frac{1-(1+\delta)\,q_{\mathrm{opt}}^2}{1+(1+\delta^{-1})\,C_{\mathrm{stab}}[2M]^2\,C_{\mathrm{drel}}^2} < \frac{1}{1+C_{\mathrm{stab}}[2M]^2\,C_{\mathrm{drel}}^2} =: \theta_{\mathrm{opt}}$$

concludes the proof. □

For the sake of completeness, we also include the monotonicity of the estimators.

**Lemma 1.13.** *Let* (UB) *hold for the Galerkin solutions* $u_\ell^\star$ *to* (1.3) *with* $C_{\mathrm{bnd}} = M$ *associated to the meshes* $\mathcal{X}_\ell$ *that appear in Algorithm 1.8. Suppose* (A1), (A3), *and a Céa-type estimate* (1.7). *Then, there holds quasi-monotonicity of the estimator, i.e., there exists* $C_{\mathrm{mon}} > 0$ *such that, for all* $\mathcal{T}_\ell \in \mathbb{T}$ *and* $\mathcal{T}_{\ell'} \in \mathbb{T}(\mathcal{T}_\ell)$ *with* $0 \le \ell < \ell' < \underline{\ell}$,

$$\eta_{\ell'}(u_{\ell'}^\star) \le C_{\mathrm{mon}}\, \eta_\ell(u_\ell^\star) \quad \text{where} \quad C_{\mathrm{mon}} = 1 + C_{\mathrm{stab}}[2M]\, C_{\mathrm{rel}}\, (1 + C_{\mathrm{Céa}}[2M]). \tag{1.28}$$

*Proof.* Since $|||u_{\ell'}^\star - u_\ell^\star||| \le 2M$ by (UB), it holds that

$$\eta_{\ell'}(u_{\ell'}^\star) \overset{\text{(A1)}}{\le} \eta_\ell(u_\ell^\star) + C_{\mathrm{stab}}[2M]\, |||u_\ell^\star - u_{\ell'}^\star||| \overset{(1.7)}{\le} \eta_\ell(u_\ell^\star) + C_{\mathrm{stab}}[2M]\, (1 + C_{\mathrm{Céa}}[2M])\, |||u^\star - u_\ell^\star|||$$

and reliability (A3) concludes the proof. □

### 1.3.2 Full linear convergence

A cornerstone to prove optimal convergence rates is full R-linear convergence. Regardless of whether the mesh is refined, another linearization step is made, or an additional linear solver step is performed, the algorithm contracts the quasi-error from the right-hand side of (1.23).

Theorem 1.14: Full R-linear convergence; [③AIL2, Theorem 4.13 below]

*Let* $\mathcal{A}$ *satisfy* (SM), (LIP), *and* (POT). *Suppose* (UB) *for all iterates and the estimator properties* (A1)–(A3). *Then, for arbitrary adaptivity parameters* $\lambda_{\mathrm{lin}}, \lambda_{\mathrm{alg}} > 0$, *and* $0 < \theta \le 1$, *the quasi-error from* (1.23)

$$\mathrm{H}_\ell^{k,j} := \eta_\ell(u_\ell^{k,j}) + |||u_\ell^\star - u_\ell^{k,\star}||| + |||u_\ell^{k,\star} - u_\ell^{k,j}||| \tag{1.29}$$

*is R-linear convergent, i.e., there exists* $C_{\mathrm{lin}} > 0$ *and* $0 < q_{\mathrm{lin}} < 1$ *such that*

$$\mathrm{H}_\ell^{k,j} \le C_{\mathrm{lin}} q_{\mathrm{lin}}^{|\ell,k,j|-|\ell',k',j'|} \mathrm{H}_{\ell'}^{k',j'} \text{ for all } (\ell,k,j), (\ell',k',j') \in Q \text{ with } |\ell',k',j'| \le |\ell,k,j|. \tag{1.30}$$

*In particular, it follows that*

$$|||u^\star - u_\ell^{k,j}||| \to 0 \quad \text{as} \quad |\ell,k,j| \to \infty. \tag{1.31}$$

*Proof idea.* The gist of the proof of (1.30) lies in the estimator reduction (1.25) (index $\ell$), the contraction of the Zarantonello iteration (index $k$) and the contraction of the algebraic solver (index $j$). Also the stopping criteria and quasi-monotonicity arguments of the noncontracted error components are used. A detailed proof of (1.30) is given in [③AIL2, Theorem 4.13 below].

To see the convergence (1.31), note that reliability (A3), uniform boundedness (UB) in (A1), and R-linear convergence (1.30) yield that

$$\begin{aligned}|||u^\star - u_\ell^{k,j}||| &\le |||u^\star - u_\ell^\star||| + |||u_\ell^\star - u_\ell^{k,i}||| \overset{(A3)}{\lesssim} \eta_\ell(u_\ell^\star) + |||u_\ell^\star - u_\ell^{k,j}||| \\ &\overset{(A1)}{\lesssim} \eta_\ell(u_\ell^{k,j}) + |||u_\ell^\star - u_\ell^{k,j}||| \overset{(1.23)}{\lesssim} \mathrm{H}_\ell^{k,j} \overset{(1.30)}{\lesssim} q_{\mathrm{lin}}^{|\ell,k,j|}\,\mathrm{H}_0^{0,0} \to 0 \quad \text{as} \quad |\ell,k,j| \to \infty.\end{aligned}$$

This concludes the proof. □

**Remark 1.15.** (i) *In [②AIL1, Lemma 3.12 below], the $k$-loop stopping criterion and [③AIL2, Theorem 4.8 below] also the $j$-loop stopping criterion need to be adjusted to ensure* (UB). *This adaptation, however, also allows for arbitrary adaptivity parameters $0 < \theta \le 1, 0 < \lambda_{\mathrm{lin}}$, and $0 < \lambda_{\mathrm{alg}}$ in the statement of R-linear convergence.*

(ii) *We remark that the estimator contraction* (1.25) *in Proposition 1.11 holds only for the final iterates, while full R-linear convergence proves the statement for any two indices $(\ell,k,j), (\ell',k',j') \in \mathcal{Q}$ with $|\ell',k',j'| \le |\ell,k,j|$.*

Moreover, with full R-linear convergence (1.30), we conclude that convergences rates with respect to the number of degrees of freedom coincide with the rates with respect to overall computational cost. This will become apparent with the notation of approximability below.

**Corollary 1.16** (rates ≙ complexity; [BFM+23, Corollary 14])**.** *Let $r > 0$. Under the assumptions of Theorem 1.14 and with* $\mathtt{cost}(\ell,k,j)$ *from* (1.22)*, there holds that*

$$\sup_{(\ell,k,j)\in\mathcal{Q}} (\#\mathcal{T}_\ell)^r\,\mathrm{H}_\ell^{k,j} < \infty \quad\Longleftrightarrow\quad \sup_{(\ell,k,j)\in\mathcal{Q}} \mathtt{cost}(\ell,k,j)^r\,\mathrm{H}_\ell^{k,j} < \infty. \qquad \square$$

### 1.3.3 Optimal convergence rates and the notion of approximability

We introduce the *approximation class* of rate $r > 0$ along the lines of [CFPP14], which were introduced in the context of AFEM in [BDD04; Ste07; CKNS08]. Let $\mathcal{T}_0$ be the initial triangulation. We define

$$\|u^\star\|_{\mathbb{A}_r} := \|u^\star\|_{\mathbb{A}_r(\mathcal{T}_0)} := \sup_{N\in\mathbb{N}_0} (N+1)^r \min_{\mathcal{T}_{\mathrm{opt}}(N)\in\mathbb{T}(N)} \eta_{\mathrm{opt}}(u_{\mathrm{opt}}^\star) \in [0,\infty], \tag{1.32}$$

where the minimum is taken over the finite set $\mathbb{T}(N) = \{\mathcal{T}\in\mathbb{T} \mid \#\mathcal{T} - \#\mathcal{T}_0 \le N\}$ with the error estimator $\eta$ from (1.11). The index $\mathrm{opt} = \mathrm{opt}(N)$ is used for quantities that depend on functions in the finite element space $\mathcal{X}_{\mathrm{opt}}$ associated with a minimizing (optimal) mesh in $\mathbb{T}(N)$.

If $\|u^\star\|_{\mathbb{A}_r} < \infty$, we say that the exact solution $u^\star$ of the model problem (1.2) is *in the approximation class of rate $r > 0$*. This is ensured if and only if

$$\min_{\mathcal{T}_{\mathrm{opt}}\in\mathbb{T}(N)} \eta_{\mathrm{opt}}(u_{\mathrm{opt}}^\star) = O(N^{-r}) \quad \text{for} \quad N \to \infty.$$

With other words, starting from an initial triangulation $\mathcal{T}_0$, the sequence of the error measure $(\eta_{\mathrm{opt}(N)})_{N\in\mathbb{N}_0}$ on the corresponding theoretically optimal (but too costly to compute)

meshes $(\mathcal{T}_{\mathrm{opt}(N)})_{N\in\mathbb{N}_0}$ decays at least with algebraic rate $r>0$ with respect to the number $N$ of additional triangles. To see the equivalence, recall that elementary calculations [BHP17, Lemma 22] show that

$$\#\mathcal{T}_{\mathrm{opt}} - \#\mathcal{T}_0 + 1 \simeq \#\mathcal{T}_{\mathrm{opt}},$$

where the hidden constant depends only on $\#\mathcal{T}_0$. From this, we infer the claimed equivalence

$$\|u^\star\|_{\mathbb{A}_r} < \infty \quad \Longleftrightarrow \quad \min_{\mathcal{T}_{\mathrm{opt}}\in\mathbb{T}(N)} \eta_{\mathrm{opt}}(u^\star_{\mathrm{opt}}) \le C(r)\,(\#\mathcal{T}_{\mathrm{opt}})^{-r} < \infty \quad \text{for all } N\in\mathbb{N}_0.$$

**Remark 1.17.** *In general, the sequence of optimal meshes $(\mathcal{T}_{\mathrm{opt}(N)})_{N\in\mathbb{N}_0}$ may not be nested and does not necessarily stem from successive refinement.*

### 1.3.4 Comparison lemma

To connect the sequences of the meshes generated by Algorithm 1.8 with the optimal meshes from the definition of the (1.32), we recall the comparison lemma. It states that by adding a certain number of elements to the triangulation from Algorithm 1.8 there holds contraction of the estimator.

**Proposition 1.18** (comparison lemma)**.** *Let quasi-monotonicity of the estimators from Lemma 1.13 and the overlay estimate (R2) hold. Let $0\le\ell<\underline{\ell}$ with $\mathcal{T}_\ell\in\mathbb{T}$ that satisfies $\eta_\ell(u^\star_\ell)>0$. Moreover, let $0<q<1$ and let $r>0$ such that $\|u^\star\|_{\mathbb{A}_r}<\infty$. Then, for every $\ell$, there exists a refinement $\mathcal{T}_{\ell'}\in\mathbb{T}(\mathcal{T}_\ell)$ that satisfies*

$$\#\mathcal{T}_{\ell'} - \#\mathcal{T}_\ell \le \Big(\frac{C_{\mathrm{mon}}\,\|u^\star\|_{\mathbb{A}_r}}{q\,\eta_\ell(u^\star_\ell)}\Big)^{1/r}, \tag{1.33}$$

$$\eta_{\ell'}(u^\star_{\ell'}) \le q\,\eta_\ell(u^\star_\ell). \tag{1.34}$$

*Proof.* First, pick the minimal $N\in\mathbb{N}_0$ such that

$$C_{\mathrm{mon}}\,\|u^\star\|_{\mathbb{A}_r} \le q\,(N+1)^r\,\eta_\ell(u^\star_\ell). \tag{1.35}$$

Note that $N=0$ would give $C_{\mathrm{mon}}\,\|u^\star\|_{\mathbb{A}_r}\le q\,\eta_\ell(u^\star_\ell)$. This is not possible since $q\,\eta_\ell(u^\star_\ell)<\eta_\ell(u^\star_\ell)\le C_{\mathrm{mon}}\,\eta_0(u^\star_0)\le C_{\mathrm{mon}}\,\|u^\star\|_{\mathbb{A}_r}$. We conclude that $N\in\mathbb{N}$.

Next, we determine

$$\mathcal{T}_{\mathrm{opt}^\star} := \operatorname*{argmin}_{\mathcal{T}_{\mathrm{opt}}\in\mathbb{T}(N)} \eta_{\mathrm{opt}}(u^\star_{\mathrm{opt}}) \quad \text{and define the overlay} \quad \mathcal{T}_{\ell'} := \mathcal{T}_\ell\oplus\mathcal{T}_{\mathrm{opt}^\star}.$$

The overlay estimate (R2) gives

$$\#\mathcal{T}_{\ell'} - \#\mathcal{T}_\ell \overset{\text{(R2)}}{\le} \big(\#\mathcal{T}_{\mathrm{opt}^\star} - \#\mathcal{T}_0 + \#\mathcal{T}_\ell\big) - \#\mathcal{T}_\ell \le N.$$

The (not met) minimality for $N-1\in\mathbb{N}_0$ in (1.35) ensures that $q\,N^r\,\eta_\ell(u^\star_\ell) < C_{\mathrm{mon}}\,\|u^\star\|_{\mathbb{A}_r}$. Rearrangement together with the previous estimate results in (1.33). The quasi-monoto-

nicity of the estimators and the definition of the approximation class (1.32) yield

$$\eta_{\ell'}(u^\star_{\ell'}) \le C_{\mathrm{mon}}\,\eta_{\mathrm{opt}^\star}(u^\star_{\mathrm{opt}^\star}) \le \frac{C_{\mathrm{mon}}\,\|u^\star\|_{\mathbb{A}_r}}{(N+1)^r} \overset{(1.35)}{\le} q\,\eta_\ell(u^\star_\ell). \tag{1.36}$$

This concludes the proof of (1.34). □

### 1.3.5 Main theorem on optimal rates with respect to cost

One main result in this thesis, namely quasi-optimal convergence of the proposed AILFEM strategy, is the content of the following theorem. This guaranteed cost-optimal steering of discretization, linearization, and the algebraic solver hold for general locally Lipschitz continuous problems and, thus, extends known results on strongly-monotone and Lipschitz continuous problems [GHPS21; HPSV21; HPW21].

**Theorem 1.19: optimal complexity; [②AIL1, Theorem 3.17 below]**

*Suppose* (SM), (LIP), *and* (POT) *as well as* (UB). *Under the assumptions of* (A1)–(A4) *and the fine properties of mesh refinement* (R1)–(R3), *let* $r > 0$. *Then, for arbitrary adaptivity parameter* $\lambda_{\mathrm{alg}} > 0$ *and sufficiently small* $\lambda_{\mathrm{lin}} > 0$ *and* $\theta > 0$, *Algorithm 1.8 reproduces optimal rates with respect to the cost and computation time. Formally, with the quasi-error* $\mathrm{H}^{k,j}_\ell$ *from* (1.29) *and* $\mathtt{cost}(\ell,k,j)$ *from* (1.22), *it holds that*

$$\sup_{N\in\mathbb{N}_0} (N+1)^r \min_{\mathcal{T}_{\mathrm{opt}}(N)\in\mathbb{T}(N)} \eta_{\mathrm{opt}}(u^\star_{\mathrm{opt}}) < \infty \quad\Longrightarrow\quad \sup_{(\ell,k,j)\in Q} \mathtt{cost}(\ell,k,j)^r\,\mathrm{H}^{k,j}_\ell < \infty. \tag{1.37}$$

*With other words, the quasi-error* $\mathrm{H}^{k,j}_\ell$ *decays with the best possible rate* $r > 0$ *over the computational cost, if* $u^\star$ *satisfies* $\|u^\star\|_{\mathbb{A}_r} < \infty$, *i.e.,* $u^\star$ *can theoretically be approximated at rate* $r > 0$ *on a sequence of error estimators of Galerkin solutions* $u^\star_{\mathrm{opt}}$ *on optimally chosen meshes* $\mathcal{T}_{\mathrm{opt}}$. □

**Remark 1.20.** (i) *In case of [②AIL1], where an exact algebraic solver of the linear procedure is assumed, the theorem holds with a modified quasi-error without algebraic contribution.*

(ii) *The proof of optimal convergence rates uses a perturbation argument that* $0 < \theta$ *and* $0 < \lambda_{\mathrm{lin}}$ *are sufficiently small to relate the Dörfler marking on the final iterates in Algorithm 1.8 to Dörfler marking on exact Galerkin solutions* $u^\star_\ell$. *The comparison lemma then connects the optimal meshes in the approximation class to the meshes generated by the algorithm.*

(iii) *The stopping criterion of the* $j$*-loop is tailored to ensure* (UB) *for all iterates. This implicitly, but not explicitly enforces that also* $\lambda_{\mathrm{alg}}$ *is sufficiently small.*

We conclude the section on AILFEMs with a schematic connection of the presented results; see Figure 1.7.

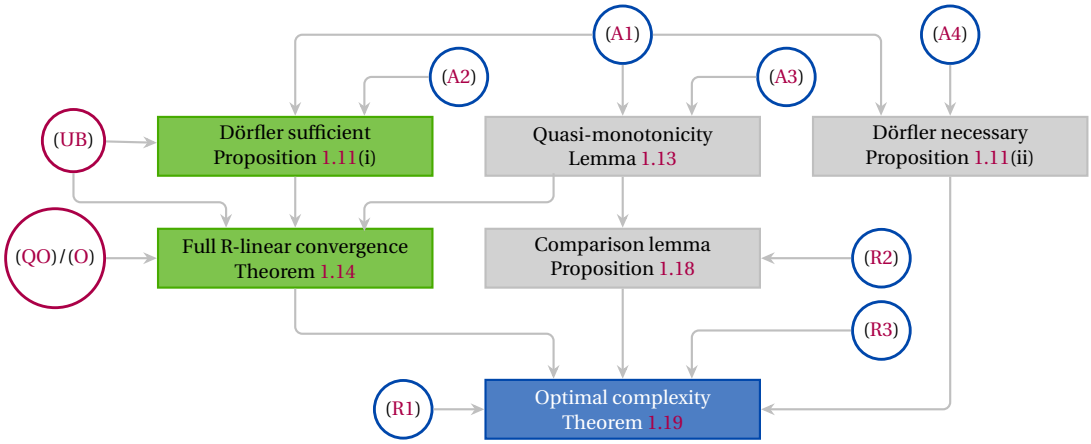



**Figure 1.7:** Overview of the proof strategy to obtain optimal complexity. The blocks highlighted in green concern the approximate solutions from the algorithm, whereas the results in gray are for the exact discrete solutions. The comparison lemma (Proposition 1.18) is used to connect R-linear convergence (Theorem 1.14) on the algorithmic approximations with the definition of the approximation class that relies on the exact solution to obtain optimal complexity (Theorem 1.19). The properties (UB) and (QO)/(O) need to hold only for the meshes that are generated by the proposed Algorithm 1.8.

## 1.4 Goal-oriented AFEM

The next section is devoted to goal-oriented AFEMs for semilinear PDEs. Usually, standard adaptive FEM aims to compute the exact solution $u^\star \in H^1_0(\Omega)$ of the given model problem (1.2). In applications, it is oftentimes more important to compute a scalar *goal functional* $G \in H^{-1}(\Omega)$, e.g., an energy, evaluated at the solution $u^\star$. Goal-oriented adaptive FEMs (GOAFEMs) thus seek to approximate $G(u^\star) \approx G(u^\star_\ell)$, where, in our case, the *quantity of interest* is assumed to be of the linear form

$$G(v) = \int_\Omega g\, v \,\mathrm{d}x + \int_\Omega \boldsymbol{g} \cdot \nabla v \,\mathrm{d}x, \tag{1.38}$$

which models a general $H^{-1}(\Omega)$ right-hand side.

Unlike the naive approach based on continuity of $G$

$$|G(u^\star) - G(u^\star_H)| = |G(u^\star - u^\star_H)| \le \|G\|_{H^{-1}(\Omega)}\, \|u^\star - u^\star_H\|_{H^1_0(\Omega)}, \tag{1.39}$$

a more sophisticated approach relies on duality techniques from [GS02] and leads to a (potential) doubling of convergence rates. The increased difficulty of *goal-oriented* methods lies in the fact that the algorithm needs to balance singularities of the model problem (1.2) with singularities that only appear in the goal quantity $G \in H^{-1}(\Omega)$ to overall minimize the goal error $G(u^\star) - G(u^\star_H)$ with the best convergence rate possible.

### 1.4.1 Dual problems and a goal-error estimate

We first consider a linear model problem to gain insights on how to define the dual problem and derive a goal-error identity as well as a goal-error estimate. Afterwards, we discuss the changes needed to cover the semilinear case as well.

**The linear case.** For vanishing nonlinearity $b(\cdot) = 0$ and $F, G \in H^{-1}(\Omega)$, the primal and the dual problem reads: Find $u^\star \in H^1_0(\Omega)$ and $z^\star \in H^1_0(\Omega)$, respectively, that satisfies

$$\langle\!\langle u^\star , v\rangle\!\rangle = \langle F , v\rangle \quad \text{for all } v \in H^1_0(\Omega) \quad \text{resp.} \quad \langle\!\langle v , z^\star\rangle\!\rangle = \langle G , v\rangle \quad \text{for all } v \in H^1_0(\Omega). \tag{1.40}$$

For the finite element space $\mathcal{X}_H$ with corresponding Galerkin solutions $u^\star_H \in \mathcal{X}_H$ and $z^\star_H \in \mathcal{X}_H$. The linear goal quantity $G$ together with Galerkin orthogonality establish

$$G(u^\star) - G(u^\star_H) = G(u^\star - u^\star_H) \overset{(1.40)}{=} \langle\!\langle u^\star - u^\star_H , z^\star\rangle\!\rangle \tag{1.41}$$

$$\overset{(1.40)}{=} \langle\!\langle u^\star - u^\star_H , z^\star - z^\star_H\rangle\!\rangle \lesssim |||u^\star - u^\star_H||| \, |||z^\star - z^\star_H|||. \tag{1.42}$$

This already yields two important observations: First, in (1.41), we see that the goal error is the dual problem tested with $u^\star - u^\star_H$. Second, in (1.42) we already see a (nonlinear) product structure that allows for a doubling of convergence rates (under the premise that both problems can be approximated with the same rate; cf. (1.39)). We also remark that in linear problems, the primal and dual error contribute equally to the error product.

**Theoretical dual problem and goal-error identity.** We shift the focus to the semilinear setting by translating (1.41) to the semilinear model problem (1.2). For a linear goal $G \in H^{-1}(\Omega)$ and the semilinear model problem (1.2), the (symbolic) dual solution $\tilde{z}^\star[u^\star_H] \in H^1_0(\Omega)$ satisfies

$$\begin{aligned} G(u^\star) - G(u^\star_H) = G(u^\star - u^\star_H) &= \langle \mathcal{A}(u^\star) - \mathcal{A}(u^\star_H) , \tilde{z}^\star[u^\star_H]\rangle \\ &= \langle\!\langle u^\star - u^\star_H , \tilde{z}^\star[u^\star_H]\rangle\!\rangle + \langle b(u^\star) - b(u^\star_H) , \tilde{z}^\star[u^\star_H]\rangle_\Omega, \end{aligned}$$

where we use the notation $[\,\cdot\,]$ for arguments in $H^1_0(\Omega)$ to signify the dependence on the *linearization point*. Under the assumption of a differentiable $b$ (cf. [②AIL1, Assumption (CAR) below]), the difference $b(u^\star) - b(u^\star_H)$ can be rewritten with the main theorem of calculus as

$$b(u^\star) - b(u^\star_H) = \Big( \int_0^1 b'(u^\star + \tau\,(u^\star_H - u^\star))\,\mathrm{d}\tau \Big)\,(u^\star - u^\star_H). \tag{1.43}$$

By defining

$$\boldsymbol{B}(u^\star, u^\star_H)\,w \coloneqq \Big( \int_0^1 b'(u^\star + \tau\,(u^\star_H - u^\star))\,\mathrm{d}\tau \Big)\,w \quad \text{for } w \in H^1_0(\Omega). \tag{1.44}$$

and motivated by the linear case, this gives rise to the *theoretical* dual problem: Find $\tilde{z}^\star[u_H^\star] \in H_0^1(\Omega)$ such that

$$\langle\!\langle v\,,\,\tilde{z}^\star[u_H^\star]\rangle\!\rangle + \langle v\,,\,\boldsymbol{B}(u^\star,u_H^\star)\,\tilde{z}^\star[u_H^\star]\rangle = G(v) \quad \text{for all } v \in H_0^1(\Omega). \tag{1.45}$$

Overall, for $z_H \in \mathcal{X}_H$, the Galerkin orthogonality for the primal problem (1.2) yields a goal-error identity that is similar to (1.41)

$$\begin{aligned} G(u^\star) - G(u_H^\star) &\overset{(1.43)}{=} \langle\!\langle u^\star - u_H^\star\,,\,\tilde{z}^\star[u_H^\star]\rangle\!\rangle + \langle b(u^\star) - b(u_H^\star)\,,\,\tilde{z}^\star[u_H^\star]\rangle \\ &\overset{(1.2)}{=} \langle\!\langle u^\star - u_H^\star\,,\,\tilde{z}^\star[u_H^\star] - z_H\rangle\!\rangle + \langle b(u^\star) - b(u_H^\star)\,,\,\tilde{z}^\star[u_H^\star] - z_H\rangle. \end{aligned} \tag{1.46}$$

**Practical dual problem and goal-error estimate.** Though similar to the linear case, the goal-error identity (1.46) faces two major differences when compared to the linear problem: First, the dual problem depends on the linearization point $u_H^\star$ and thus changes on each mesh level. Second, the theoretical dual problem (1.45) is not computable in practice, since the operator $\boldsymbol{B}(u^\star,u_H^\star)$ involves the unavailable exact solution $u^\star$. A remedy to the second issue comes from the observation that $\boldsymbol{B}(u^\star,u_H^\star) \to b'(u_H^\star)$ as $u_H^\star \to u^\star$. This motivates the so-called *practical* dual problem: Seek $z^\star[u_H^\star] \in H_0^1(\Omega)$ such that

$$\langle\!\langle v\,,\,z^\star[u_H^\star]\rangle\!\rangle + \langle v\,,\,b'(u_H^\star)\,z^\star[u_H^\star]\rangle = G(v) \quad \text{for all } v \in H_0^1(\Omega). \tag{1.47}$$

We include the practical problem into (1.46) and arrive at the *goal-error identity*

$$\begin{aligned} G(u^\star)-G(u_H^\star) &\overset{(1.46)}{=} \langle\!\langle u^\star-u_H^\star\,,\,\tilde{z}^\star[u_H^\star]-z^\star[u_H^\star]\rangle\!\rangle + \langle b(u^\star)-b(u_H^\star)\,,\,\tilde{z}^\star[u_H^\star]-z^\star[u_H^\star]\rangle \\ &\quad + \langle\!\langle u^\star-u_H^\star\,,\,z^\star[u_H^\star]-z_H^\star[u_H^\star]\rangle\!\rangle + \langle b(u^\star)-b(u_H^\star)\,,\,z^\star[u_H^\star]-z_H^\star[u_H^\star]\rangle. \end{aligned} \tag{1.48}$$

The dual problems (1.45) and (1.47) are well-posed due to the Lax–Milgram lemma, relying on the monotonicity of $b$ (and $\boldsymbol{B}(u^\star,u_H^\star)$) (see [②AIL1, Assumption (MON) below]), the growth condition [①GOA, Section 2.7 below], and the ellipticity of the diffusion part (see [②AIL1, Assumption (ELL) below]).

To reliably control the goal error $|G(u^\star) - G(u_H^\star)|$, we rigorously derive two stability results ([①GOA, Lemma 2.9 and 2.10 below]).

**Proposition 1.21.** *Suppose* (SM) *and* (LIP). *Then, it holds that*

$$\|b(u^\star) - b(u_H^\star)\|_{H^{-1}(\Omega)} \lesssim |||u^\star - u_H^\star||| \quad \textit{and} \quad |||\tilde{z}^\star[u_H^\star] - z^\star[u_H^\star]||| \lesssim |||u^\star - u_H^\star|||. \tag{1.49}$$

*Applied to* (1.48), *this proves the* goal-error estimate

$$|G(u^\star) - G(u_H^\star)| \overset{(1.48)}{\lesssim} |||u^\star - u_H^\star||| \left[|||z^\star[u_H^\star] - z_H^\star[u_H^\star]|||^2 + |||u^\star - u_H^\star|||^2\right]^{1/2}. \tag{1.50}$$

*for semilinear PDEs; compare with* (1.42). □

We remark that the first stability estimate (1.49) is immediate if we suppose global Lipschitz continuity of $b$. The second stability estimate (1.49) follows if $b'$ is Lipschitz continuous (since $\boldsymbol{B}(u^\star,u_H^\star) = \boldsymbol{B}(u_H^\star,u^\star)$).

**Goal-oriented adaptivity.** The goal-error estimate (1.50) is the starting point for a reliable goal-error control. To this end, we introduce the error estimator related to the practical dual problem. For a linearization point $w_H \in \mathcal{X}_H$, the strong form of the practical dual problem reads

$$\begin{aligned} -\operatorname{div}(\boldsymbol{A}\nabla z^\star[w_H]) + b'(w_H)z^\star[w_H] &= g - \operatorname{div}(\boldsymbol{g}) && \text{in } \Omega, \\ z^\star[w_H] &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{1.51}$$

The elementwise contribution of the *a posteriori* estimator related to the practical dual problem (1.47) read

$$\zeta_H(w_H;T,v_H) := h_T^2\,\|g + \operatorname{div}(\boldsymbol{A}\nabla v_H - \boldsymbol{g}) - b'(w_H)\,v_H\|_{L^2(T)}^2 + h_T\,\|[\![\boldsymbol{A}\nabla v_H - \boldsymbol{g}]\!]\|_{L^2(\partial T\cap\Omega)},$$

which depends on the linearization point $w_H \in \mathcal{X}_H$. Moreover, for $\mathcal{U}_H \subseteq \mathcal{T}_H$ and $w_H \in \mathcal{X}_H$, the practical and computable dual estimator reads

$$\zeta_H(w_H;\mathcal{U}_H,v_H)^2 := \sum_{T\in\mathcal{U}_H} \zeta_H(w_H;T,v_H)^2 \quad \text{and} \quad \zeta_H(w_H;v_H)^2 := \zeta_H(w_H;\mathcal{T}_H,v_H)^2.$$

The dual estimator $\zeta_H(u_H^\star;z_H^\star[u_H^\star])$ also satisfies (A1)–(A4); see [①GOA, Proposition 2.15 below]. Therefore, reliability (A3) and the goal error estimate (1.50) show

$$\begin{aligned} |G(u^\star)-G(u_H^\star)| &\overset{(1.50)}{\lesssim} |||u^\star-u_H^\star|||\,\big[|||z^\star[u_H^\star]-z_H^\star[u_H^\star]|||^2 + |||u^\star-u_H^\star|||^2\big]^{1/2} \\ &\lesssim \eta_H(u_H^\star)\,\big[\zeta_H(u_H^\star;z_H^\star[u_H^\star])^2 + \eta_H(u_H^\star)^2\big]^{1/2} =: \eta_H(u_H^\star)\,\rho_H(u_H^\star,z_H^\star[u_H^\star]), \end{aligned} \tag{1.52}$$

where we call the estimator $\rho_H(\cdot,\cdot)$ *combined estimator*. The combined estimator also satisfies the axioms of adaptivity (A1)–(A4); see [BIP21, Proposition 7], where we note that [BIP21] considers a linear PDE with a nonlinear goal functional, but ultimately obtains the same goal-error estimate (1.50).

### 1.4.2 Goal-oriented AFEM algorithm and the MARK module

The product structure in (1.52) requires a suitable marking step to preserve quasi-minimality of the marked elements. An abstract algorithmic formulation of a goal-oriented adaptive FEM is given in Algorithm 1.22.

**Algorithm 1.22: schematic goal-oriented adaptive FEM**

**Input:** $\mathcal{T}_0$ conforming mesh, marking parameters $0<\theta\le 1$, $C_{\mathrm{mark}}\ge 1$ for MARK module.
**Adaptive loop:** For all $\ell = 0,1,2,\dots$, repeat the following steps (I)–(V):

(I) SOLVE & ESTIMATE (primal). Compute $u_\ell^\star$ from the discrete primal problem (1.3) and the primal estimator $\eta_\ell(u_\ell^\star)$.

(II) SOLVE & ESTIMATE (dual). Compute $z_\ell^\star[u_\ell^\star]$ from the practical dual problem (1.47) and the practical dual estimator $\zeta_\ell(u_\ell^\star;z_\ell^\star[u_\ell^\star])$.

(III) Compute the combined estimator $\rho_\ell(u_\ell^\star, z_\ell^\star[u_\ell^\star])$ from (1.52).

(IV) `MARK`. $\mathcal{M}_\ell \leftarrow \texttt{mark}(\mathcal{T}_\ell, \eta_\ell, \rho_\ell)$; % respects product structure (1.52).

(V) `REFINE`. $\mathcal{T}_{\ell+1} \leftarrow \texttt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$. % newest-vertex bisection

**Output:** Sequence of successively refined conforming triangulations $\mathcal{T}_\ell$, discrete solutions $u_\ell^\star$ and $z_\ell^\star[u_\ell^\star]$, and corresponding error estimators $\eta_\ell(u_\ell^\star)$, $\rho_\ell(u_\ell^\star, z_\ell^\star[u_\ell^\star])$.

We remark that the dual problem depends on the previously computed discrete solution $u_\ell^\star$ and thus (unlike linear PDEs [BBPS23]), the `SOLVE & ESTIMATE`-modules cannot be parallelized, but have to be solved sequentially.

In Proposition 1.11(i), quasi-minimal cardinality in the Dörfler marking (1.19) is a crucial ingredient. The naive approach for primal and dual problems is separate Dörfer marking (1.19) for the primal estimator $\eta_\ell$ with set $\mathcal{M}_\ell^\eta$ and the second estimator $\rho_\ell$ with set $\mathcal{M}_\ell^\rho$ and then taking the union $\mathcal{M}_\ell^\eta \cup \mathcal{M}_\ell^\rho$ as the set of overall marked elements [HPZ15]. However, this may lead to a suboptimal allocation of resources that may cause rates to deteriorate. A problematic case is if one estimator marks only a few elements with very large contributions and the other estimator marks a large number of elements, but with comparably very small indicators compared to the first estimator. A remedy to this problem was introduced in [MS09] for the Poisson problem, where only the set of lesser cardinality constitutes the set of marked elements. The algorithm is presented in Algorithm 1.23.

**Algorithm 1.23: `mark` — `MARK` module from [MS09]**

**Input:** triangulation $\mathcal{T}_\ell$, estimators $\eta_\ell$, $\rho_\ell$, marking parameters $0 < \theta \le 1$, $C_{\mathrm{mark}} \ge 1$.

(i) Find a quasi-minimal set that satisfies the Dörfler marking for the estimator $\eta_\ell$

$$\#\mathcal{M}_\ell^\eta \le C_{\mathrm{mark}} \min_{\mathcal{U}_\ell^\star \in \mathbb{M}_\ell^\eta} \#\mathcal{U}_\ell^\star \quad \text{with} \quad \mathbb{M}_\ell^\eta := \{\mathcal{U}_\ell \subseteq \mathcal{T}_\ell \mid \theta\, \eta_\ell^2 \le \eta_\ell(\mathcal{U}_\ell)^2\}.$$

(ii) Find a quasi-minimal set that satisfies the Dörfler marking for the estimator $\rho_\ell$

$$\#\mathcal{M}_\ell^\rho \le C_{\mathrm{mark}} \min_{\mathcal{U}_\ell^\star \in \mathbb{M}_\ell^\rho} \#\mathcal{U}_\ell^\star \quad \text{with} \quad \mathbb{M}_\ell^\rho := \{\mathcal{U}_\ell \subseteq \mathcal{T}_\ell \mid \theta\, \rho_\ell^2 \le \rho_\ell(\mathcal{U}_\ell)^2\}.$$

(iii) Choose $\mathcal{M}_\ell \in \{\#\mathcal{M}_\ell^\eta, \#\mathcal{M}_\ell^\rho\}$ such that $\#\mathcal{M}_\ell = \min\{\#\mathcal{M}_\ell^\eta, \#\mathcal{M}_\ell^\rho\}$.

**Output:** marked elements $\mathcal{M}_\ell$.

In our case, due to the goal error estimate (1.52), the `MARK` module takes the primal and combined estimator from (1.52). Marking only the set with lesser cardinality ensures that at each level $\ell$, Dörfler marking holds either for the primal or the combined estimator.

**Remark 1.24** (alternative marking strategies)**.** *The marking in [MS09] is an approach where marking is performed separately as a first step and then the primal and dual a posteriori information is combined. A refined version of the [MS09] marking is proposed in [FPZ16], where Step* (iii) *is replaced by*

(iii') *Define* $N := \min\{\#\mathcal{M}_\ell^\eta, \#\mathcal{M}_\ell^\rho\}$.

(iv) *Pick* $\widetilde{\mathcal{M}}_\ell^\eta \subseteq \mathcal{M}_\ell^\eta$ *and* $\widetilde{\mathcal{M}}_\ell^\rho \subseteq \mathcal{M}_\ell^\rho$, *where* $\#\widetilde{\mathcal{M}}_\ell^\eta = \#\widetilde{\mathcal{M}}_\ell^\rho = N$.

(v) *Define* $\mathcal{M}_\ell := \widetilde{\mathcal{M}}_\ell^\eta \cup \widetilde{\mathcal{M}}_\ell^\rho$.

*These two marking strategies are in the spirit of* mark first, combine later.

*A different but equivalent marking in the sense of estimator equivalence is proposed in [BET11]. For given estimators* $\eta_\ell$ *and* $\rho_\ell$, *the elementwise contributions of the weighted estimator* $\varrho_\ell$ *read*

$$\varrho_\ell(T) := \eta_\ell(T)\rho_\ell + \eta_\ell\rho_\ell(T). \tag{1.53}$$

*Then, Dörfler marking is performed for the weighted estimator* $\varrho_\ell$. *Thus, this approach can be summarized as* combine first, mark later.

*For GOAFEM for semilinear PDEs, all three marking strategies involve the primal estimator* $\eta_\ell = \eta_\ell(u_\ell^\star)$ *as well as the combined estimator* $\rho_\ell = \rho_\ell(u_\ell^\star, z_H^\star[u_H^\star])$ *as motivated by the goal-error estimate* (1.52).

### 1.4.3 Literature on goal-oriented adaptive FEM

Despite the high relevance of goal-oriented adaptive FEM (GOAFEM) in practice, literature is comparably scarce compared to standard AFEM. GOAFEMs are related to dual and adjoint methods that were developed in [EEHJ95; BR01; BR03; GS02] to improve computational performance in the presence of a goal functional.. Roughly speaking, GOAFEMs are divided into two schools of thought: First, *dual weighted residual methods* (DWR), where the primal and dual estimators are weighted elementwise. These DWR methods are computationally very performant but convergence of the related adaptive strategies is hard to analyze rigorously and appears to be still open. We refer to [ELW19; ELW20; DBR21; AESW22] for some recent contributions. The second large class are AFEM methods that are based on (merely) global error estimation for the primal and dual problem, which allows for guaranteed convergence rates. We shall discuss the latter in greater detail.

The first result on optimal convergence rates for a GOAFEM is found in [MS09] for the Poisson model problem. A computable and less local variant of DWR was used to single out elements in the `MARK`-module in [BET11] (see (1.53) above) with an empirically improved performance, yielding a connection between both schools of thought.

Linear convergence of GOAFEM for the semilinear model problem (1.2) was shown in [HPZ15; XHYM21], however, relying on the global Lipschitz continuity of $b$ and discrete $L^\infty(\Omega)$-bound on the discrete solutions. The mere convergence of GOAFEM for general nonsymmetric second-order linear elliptic problems was proven in [HP16], while [FFGHP16] published results on optimal rates for symmetric second-order linear elliptic PDEs. The paper [FPZ16] generalized those results on optimal rates to general second-order linear elliptic PDEs with marking more elements compared to [MS09]. In [BIP21], optimal convergence rates were proven for a nonsymmetric model problem with quadratic goal functional, where the key goal error estimate has a similar product structure as in (1.52) (for different reasons).

GOAFEMs that also include an inexact solver are scarce. The aforementioned contributions, except for [MS09] with a generic contractive solver, are under the assumption of an exact solve procedure. Moreover, the solve procedure of [MS09] does not comment on the practical steering of the contractive solver. By including (and also adaptively steering)

an inexact solver, a cost-optimal GOAFEM for symmetric second-order linear PDEs is presented in [BGIP23]. For a linear and nonsymmetric problem with a coupled loop of the form of Algorithm 1.8 with symmetrization and an algebraic solver, we refer to the recent own work [BBPS23].

This thesis presents the following major extensions to the existing literature: In [①GOA], a thorough analysis replaces global Lipschitz continuity with growth conditions on the nonlinear reaction $b$. Moreover, any discrete $L^\infty(\Omega)$-assumptions are avoided. By introducing the practical dual problem, we are able to prove the stability estimates (1.49) as well as the goal-error estimate (1.50). Moreover, since the primal and dual problem do not decouple as in the linear case [MS09; BIP21; BGIP23], stability estimates with respect to the linearization point $w \in H^1_0(\Omega)$ (Lemma 2.25 below) of the form

$$|||z^\star[u^\star] - z^\star[w]||| + |||z^\star_H[u^\star] - z^\star_H[w]||| \lesssim |||u^\star - w||| \tag{1.54}$$

are required and pose a significant additional challenge.

In addition, we rigorously prove R-linear convergence (Theorem 2.19 below) and optimal convergence rates (Theorem 2.20 below) with respect to the number of degrees of freedom for the semilinear model problem in the goal-oriented setting. This result is obtained under a supposed exact solve procedure. Overall, this is the first mathematically rigorous result on optimal convergence rates of GOAFEM for a nonlinear PDEs.

### 1.4.4 Main results: linear convergence and optimal convergence rates

For a sufficiently large mesh-refinement index $\ell_0 \in \mathbb{N}_0$, we establish the quasi-orthogonality (QO) in the energy norm for the primal problem [①GOA, Lemma 2.29 below] and the exact dual problem [①GOA, Lemma 2.30 below] as an intermediate step. From this and motivated by (1.52), we prove the technically demanding quasi-orthogonality for the combined quantity ([①GOA, Lemma 2.31 below])

$$|||z^\star[u^\star] - z^\star_H[u^\star_H]|||^2 + |||u^\star - u^\star_H|||^2. \tag{1.55}$$

With this, we are able to present R-linear convergence regardless of the whether Dörfler marking is performed for the primal or the combined estimator.

**Theorem 1.25: R-linear convergence [①GOA, Theorem 2.19 below]**

*Suppose* (SM) *and* (LIP). *From* (1.52), *recall the combined estimator* $\rho_\ell(u^\star_\ell, z^\star_\ell[u^\star_\ell]) = \big[\zeta_\ell(u^\star_\ell; z^\star_\ell[u^\star_\ell])^2 + \eta_\ell(u^\star_\ell)^2\big]^{1/2}$. *Let* $\eta_\ell$ *and* $\rho_\ell$ *satisfy* (A1)–(A3) *and let* (QO) *hold for*

$$|||u^\star - u^\star_H|||^2 \quad \text{and} \quad |||z^\star[u^\star] - z^\star_H[u^\star_H]|||^2 + |||u^\star - u^\star_H|||^2.$$

*Then, for arbitrary* $0 < \theta \le 1$ *and arbitrary* $1 \le C_{\mathrm{mark}} \le \infty$, *there exists* $0 < q_{\mathrm{lin}} < 1$, $C_{\mathrm{lin}} > 0$, *and* $\ell_0 \in \mathbb{N}_0$, *such that, for all* $m \ge \ell \ge \ell_0$, *it holds that*

$$\eta_m(u^\star_m)\,\rho_m(u^\star_m, z^\star_m[u^\star_m]) \le C_{\mathrm{lin}}\, q_{\mathrm{lin}}^{m-\ell}\, \eta_\ell(u^\star_\ell)\,\rho_\ell(u^\star_\ell, z^\star_\ell[u^\star_\ell]). \qquad \square \tag{1.56}$$

For the product setting, we state a suitable comparison lemma from the literature customized for the case of the primal and combined estimators.

**Proposition 1.26** (variant of comparison lemma [FPZ16, Lemma 15])**.** *Suppose quasi-monotonicity of the estimators (Lemma 1.13) separately for $\eta_\ell$ and $\rho_\ell$ and the overlay estimate (R2). Let $0 \le \ell < \infty$ with $\mathcal{T}_\ell \in \mathbb{T}$ that satisfies $\eta_\ell(u_\ell^\star) > 0$ and $\rho_\ell(z_\ell^\star[u_\ell^\star]) > 0$. Moreover, let $r > 0$ and $s > 0$ such that $\|u^\star\|_{\mathbb{A}_r} + \|z^\star[u^\star]\|_{\mathbb{A}_s} < \infty$. Then, for every $0 < q < 1$, there exists a refinement $\mathcal{T}_{\ell'} \in \mathbb{T}(\mathcal{T}_\ell)$ that satisfies*

$$\#\mathcal{T}_{\ell'} - \#\mathcal{T}_\ell \le \left( \frac{C_{\mathrm{mon}} \|u^\star\|_{\mathbb{A}_r} \left[ \|z^\star[u^\star]\|_{\mathbb{A}_s} + \|u^\star\|_{\mathbb{A}_r} \right]}{q\, \eta_\ell(u_\ell^\star)\, \rho_\ell(z_\ell^\star[u_\ell^\star])} \right)^{-\min\{2r, s+r\}}, \tag{1.57}$$

$$\eta_{\ell'}(u_{\ell'}^\star)\, \rho_{\ell'}(z_{\ell'}^\star[u_{\ell'}^\star]) \le q\, \eta_\ell(u_\ell^\star)\, \rho_\ell(z_\ell^\star[u_\ell^\star]), \tag{1.58}$$

*where $C_{\mathrm{mon}}$ depends only on the primal and dual quasi-monotonicity constant.* □

We remark that the proof constructs two overlays (and uses the overlay estimate (R2) twice), namely $\mathcal{T}_\ell \oplus \mathcal{T}_\eta \oplus \mathcal{T}_\zeta$, where $\mathcal{T}_\eta$ and $\mathcal{T}_\zeta$ denote the optimal meshes from the primal and dual approximation classes, respectively.

We are able to present the main result on optimal convergence rates with respect to the number of degrees of freedom.

Theorem 1.27: Optimal rates [ⓘGOA, Theorem 2.20 below]

*Let $\mathcal{A}$ fulfill (SM) and (LIP). Suppose that $\eta_\ell$ and $\zeta_\ell$ satisfy (A1)–(A4) and that (R1)–(R3) holds. Recall the combined estimator $\rho_\ell(u_\ell^\star, z_\ell^\star[u_\ell^\star]) = \left[\zeta_\ell(u_\ell^\star; z_\ell^\star[u_\ell^\star])^2 + \eta_\ell(u_\ell^\star)^2\right]^{1/2}$. Let (QO) hold for a sufficiently large mesh-refinement index $\ell_0 \in \mathbb{N}_0$ and*

$$|||u^\star - u_H^\star|||^2 \quad \text{and} \quad |||z^\star[u^\star] - z_H^\star[u_H^\star]|||^2 + |||u^\star - u_H^\star|||^2.$$

*Suppose that $\|u^\star\|_{\mathbb{A}_r} < \infty$ for $r > 0$ and that $\|z^\star[u^\star]\|_{\mathbb{A}_s} < \infty$ for $s > 0$. Then, for sufficiently small $\theta > 0$ and $1 \le C_{\mathrm{mark}} < \infty$, and for all $\ell \ge \ell_0$, it holds that*

$$\eta_\ell(u_\ell^\star)\, \rho_\ell(u_\ell^\star, z_\ell^\star[u_\ell^\star]) \lesssim \|u^\star\|_{\mathbb{A}_r} \left[ \|u^\star\|_{\mathbb{A}_r} + \|z^\star[u^\star]\|_{\mathbb{A}_s} \right] (\#\mathcal{T}_\ell - \#\mathcal{T}_0)^{-\min\{2r, r+s\}}. \tag{1.59}$$

*In particular, the rate of convergence is $\min\{2r, r+s\}$.* □

This concludes the introduction.

## 1.5 Outline of the thesis

The following section gives an overview of the contributions presented in this thesis. The remainder of this thesis presents results that I, together with collaborators, have established during my PhD studies. These contributions are subdivided into three additional chapters — one for each research question, which shall be motivated in the following.

### 1.5.1 Chapter 2: goal-oriented adaptive finite element method (GOAFEM) with exact solver for semilinear PDEs

[①GOA]: 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.*, 118:18–35, 2022. DOI: 10.1016/j.camwa.2022.05.008

In this publication, we consider the semilinear primal problem (1.2) and a linear goal quantity $G \in H^{-1}(\Omega)$ under the assumption of an exact solve procedure. Semilinear problems in a goal-oriented setting have been investigated in [HPZ15; XHYM21], however, without proven optimal convergence rates. Closing this gap is the main achievement of [①GOA], where optimal convergence rates are understood with respect to the number of degrees of freedom. This achievement relies on the following main observations:

First, we replace the theoretical dual problem by a practical dual problem along the lines of [HPZ15]. By clarifying the assumptions on the problem setting, the growth argument [BHSZ11] can be used to obtain the main stability estimates (1.49).

Second, existing literature on rate-optimal GOAFEMs focuses mainly on linear problems and linear goals [MS09; BET11; FFGHP16; FPZ16]. The publication [BIP21] considers a linear model problem but a quadratic goal and derives a structurally similar goal error estimate (1.52) (for different reasons). The marking procedure therein is motivated by [MS09] and ensures quasi-minimal cardinality of the marked elements and is used to drive the mesh refinement in the proposed GOAFEM.

Third, for nonlinear problems, the dual problem depends on the linearization point. By proving a stability result (1.54) with respect to the linearization point, we are able to verify the *combined* quasi-orthogonality (1.55), which then gives rise to the full R-linear convergence (Theorem 1.25) and, eventually, optimal convergence rates (Theorem 1.27).

### 1.5.2 Chapter 3: adaptive iteratively linearized finite element method (AILFEM) with exact linearization for semilinear PDEs

[②AIL1]: R. Becker, M. Brunner, M. Innerberger, J. M. Melenk, and D. Praetorius. Cost-optimal adaptive iterative linearized FEM for semilinear elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 57(4):2193–2225, 2023. DOI: 10.1051/m2an/2023036

Numerical methods to linearize nonlinear equations such as the discrete problem (1.3) are analyzed in depth and widely applied, yet, their application in the context of adaptive FEM poses an exciting research question. The linearization method produces cost in the sense of (1.22) and the question arises when to adaptively stop the linearization without spoiling optimal convergence rates.

In principle, the seminal work [Ste07] addresses this question for a generic iterative solver in GALSOLVE that contracts the energy error of an initial guess with contraction factor $0 < q < 1$ at the cost of $\mathcal{O}(|\log(q)|\,\#\mathcal{T}_H)$ for the Poisson model problem. R-linear convergence and optimal convergence rates then hold for sufficiently small adaptivity parameter $\theta > 0$ and solver parameter $\lambda_{\mathrm{lin}} > 0$ and only for the final iterates of the iterative solver. Following [CKNS08], the convergence analysis can be generalized to arbitrary $0 < \theta \leq 1$ (yet sufficiently small $\lambda_{\mathrm{lin}} > 0$) by a perturbation argument from [CFPP14]. Importantly, [Ste07] does not explicitly state on how the inexact solver is stopped in practice.

By recasting the semilinear problem into an abstract operator framework from [CW17; HW20b; HPW21; GHPS21], we propose an AILFEM strategy in the spirit of Algorithm 1.8 with a void algebraic solver loop, i.e., by supposing an exact algebraic solver for the linearized problem. The damped Zarantonello operator $\Phi(\delta;\cdot)\colon \mathcal{X}_H \to \mathcal{X}_H$ with damping parameter $\delta > 0$ is employed as a means of linearization. For a sufficiently small $\delta > 0$, the Zarantonello operator is contractive in the energy norm in the sense that there exists a $0 < \widetilde{q} < 1$ such that

$$|||\Phi_H(\delta; v_H) - \Phi_H(\delta; w_H)||| \leq \widetilde{q}\,|||v_H - w_H||| \quad \text{for all } v_H, w_H \in \mathcal{X}_H,$$

where it is important to note that $\delta$ and $\widetilde{q}$ depend on $\max\{|||v_H|||, |||w_H|||\}$ due to the locally Lipschitz continuous setting. The AILFEM strategy steers and equibalances errors that come from discretization and linearization, respectively. For two successive iterates $u_H^k \in \mathcal{X}_H$ and $u_H^{k+1} \in \mathcal{X}_H$, the proposed algorithm stops the inner loop for linearization given that

$$|||u_H^{k+1} - u_H^k||| \leq \lambda_{\mathrm{lin}}\,\eta_H(u_H^{k+1}) \quad \text{and} \quad |||u_H^{k+1}||| \leq \widetilde{C}_{\mathrm{bnd}}[M] \quad \text{for some } 0 < \widetilde{C}_{\mathrm{bnd}}[M] \tag{1.60}$$

with $M$ from (1.7). This extends the algorithmic stopping criterion found in, e.g., [BIM+23; HPSV21; GHPS21] for globally Lipschitz continuous operators.

In practice, the norm criterion on the right-hand side in (1.60) further constrains (UB) in the sense that $|||u_H^k||| \leq \widetilde{C}_{\mathrm{bnd}}[M] < C_{\mathrm{bnd}}[M]$ for the final iterates of the linearization loop. This *nested iteration* criterion (cf. (3.28) below) follows essentially from the norm contraction of the Zarantonello iteration under the premise of a suitable $\delta > 0$ and will be met after $k_0$-many linearization steps for $k_0 \in \mathbb{N}_0$ (cf. Corollary 3.11).

The algorithmic adaption allows us to prove full R-linear convergence, i.e., contraction of a suitable quasi-error regardless of the algorithmic decision to either refine the mesh or perform another linearization step (cf. Theorem 1.14) also for arbitrary $\theta > 0$ and $\lambda_{\mathrm{lin}} > 0$. With full R-linear convergence, we infer that convergence rates with respect to the number of degrees of freedom and with respect to computational cost (1.22) coincide under the assumption of linear complexity for the solution of the linearized system.

Finally, we prove optimal convergence rates with respect to the degrees of freedom for $\theta > 0$ and $\lambda_{\mathrm{lin}} > 0$ sufficiently small. Since the Pythagorean identity (O) holds for a setting with energy $\mathcal{E}$, the results hold for all mesh-refinement levels $\ell \in \mathbb{N}_0$.

We conclude the paper with a practical algorithm that asymptotically ensures that a suitable damping parameter is determined. In future research, this may be extended to a fully-adaptive damping strategy in the spirit of [AW15], where optimal convergence rates are observed experimentally for an employed Newton method.

### 1.5.3 Chapter 4: adaptive iteratively linearized finite element method (AILFEM) with linearization and algebraic solver for semilinear PDEs

> [③AIL2]: M. Brunner, D. Praetorius, and J. Streitberger. Cost-optimal adaptive FEM with linearization and algebraic solver for semilinear elliptic PDEs, 2024. arXiv: 2401.06486

The exact solution of a sparse SPD system such as the Zarantonello update (1.14) is

experimentally of loglinear complexity; cf. [BIM+23, Figure 3]. This can be avoided by including a contractive solver as an inner loop. Examples are an optimally preconditioned conjugate gradient method [CNX12] or an optimal geometric multigrid method [WZ17; IMPS23].

The additional inner solver loop with iterates $u_\ell^{k,j}$ approximates the (exact and in practice unavailable) Zarantonello solutions $u_\ell^{k,\star}$. Heuristically, it is clear that sufficiently many algebraic solver steps will cause the perturbation to be negligible. Since the linear solver loop produces also cost in the sense of (1.22), we propose an AILFEM (Algorithm 1.8) that equibalances discretization, linearization, and algebraic solver errors. This motivates the research question of [③AIL2]: Is the proposed AILFEM strategy cost-optimal with respect to this perturbation? This question is more delicate than it seems at first glance for the following two reasons:

First, the perturbed Zarantonello iteration is contractive only for $1 \le k < \underline{k}[\ell]$ (cf. [BIM+23, Lemma 5.1]), i.e., there exists $0 < q < 1$ such that

$$|||u_\ell^\star - u_\ell^{k+1,j}||| \le q\,|||u_\ell^\star - u_\ell^{k,j}||| \quad \text{for all } 1 \le k+1 < \underline{k}[\ell], \tag{1.61}$$

unless there are sufficiently many algebraic solver steps. We prove that there exists an $j_{\min} \in \mathbb{N}_0$ that is independent of the mesh-refinement index $\ell$ and the linearization index $k$ and enforce algorithmically that at least $j_{\min}$ steps are performed.

Second, there holds a Pythagorean identity (O) holds for all $\ell \in \mathbb{N}_0$ in case of the energy. Thus, it is desirable to formulate the linearization error as an energy difference instead of a difference in norm. However, the algebraic solver contracts in norm, while the (exact) linearization contracts in energy. A link of energy and energy norm for the final iterates of the $j$-loop is established in the spirit of [HPW21, Property (F4)] (cf. Lemma 4.9 below), which also depends on sufficiently many linear solver steps. This equivalence is important, since it consists only of computable quantities.

As a first result, we prove uniform boundedness (UB) for all iterates (see Theorem 4.8 below) and establish R-linear convergence based on a new proof strategy from [BFM+23] for arbitrary adaptivity parameters $\theta > 0$, $\lambda_{\mathrm{lin}} > 0$, and $\lambda_{\mathrm{alg}} > 0$. Consequently, for $\theta > 0$ and $\lambda_{\mathrm{lin}} > 0$ sufficiently small, we prove optimal convergence rates understood with respect to the number of degrees of freedom and with respect to computational cost (1.22). Moreover, since all steps in the AILFEM strategy are rigorously of linear complexity, we also infer optimal rates with respect to computation time.

## 1.6 Other contributions on nonsymmetric elliptic PDEs

### 1.6.1 Adaptive iteratively symmetrized FEM (AISFEM) with symmetrization and linear solver

[BIM+23]: M. Brunner, M. Innerberger, A. Miraçi, D. Praetorius, J. Streitberger, and P. Heid. Adaptive FEM with quasi-optimal overall cost for nonsymmetric linear elliptic PDEs. *IMA J. Numer. Anal.*, 2023. DOI: 10.1093/imanum/drad039. Corrigendum to: Adaptive FEM with quasi-optimal overall cost for nonsymmetric linear elliptic PDEs. *IMA J. Numer. Anal.*, 2024. DOI: 10.1093/imanum/drad103

In this publication, we consider a general nonsymmetric second-order linear elliptic PDE in the framework of the Lax–Milgram lemma.

The usual approach to apply generalized minimal residual methods (GMRES) to nonsymmetric problems is replaced by an adaptive algorithm of the form of Algorithm 1.8. This is motivated by the fact that the link of an abstract contraction property of (optimally preconditioned) GMRES methods in vector norms lack a connection to the functional analytical setting of the finite element formulation until now. In addition, the Zarantonello iteration (1.14) does not only linearize but also *symmetrize* the underlying problem. This is combined with a linear solver for the arising symmetric and positive definite Zarantonello system.

Since nonsymmetric problems do not possess an energy, only quasi-orthogonality results in the norm (as in (O) in Proposition 1.4(i)) can be exploited. Thus, there exists a mesh refinement index $\ell_0 \in \mathbb{N}_0$ such that full R-linear convergence holds for all $\ell \geq \ell_0$. Regardless of the preasyptotic phase, we prove optimal convergence rates with respect to the overall computational cost, i.e., the total computation time for the proposed adaptive iteratively symmetrized finite element method (AISFEM).

### 1.6.2 Goal-oriented adaptive iteratively symmetrized FEM (GOAISFEM) with symmetrization and linear solver

> [BBPS23]: P. Bringmann, M. Brunner, D. Praetorius, and J. Streitberger. Optimal complexity of goal-oriented adaptive FEM for nonsymmetric linear elliptic PDEs, 2023. arXiv: 2312.00489

In this preprint, we extend the given framework from [BIM+23] to the setting of goal-oriented FEM. The analytical challenge is the nonlinear product structure for the quasi-error (similar to (1.52)), where the marking leads only to reduction of one of the factors.

Since the nested loops (symmetrization and algebraic solver) admit only the contraction of the inexact Zarantonello iteration for all but the last indices [BIM+23, Lemma 5.1] (cf. (1.61)), the proof of full R-linear convergence requires a novel approach based on a tail-summability criterion from [BFM+23]. The proof exploits a relaxed quasi-orthogonality condition from [Fei22] that, in addition to [BIM+23], enables us to prove full R-linear convergence for all $\ell \geq \ell_0 = 0$, which holds for linear second-order elliptic PDEs. With full R-linear convergence, we are able to prove optimal complexity of the proposed goal-oriented adaptive iteratively symmetrized finite element method (GOAISFEM).

## 1.7 Additional remarks on notation

**Sobolev spaces.** For a smooth $v \in C^\infty(\overline{\Omega})$, we define the Sobolev scalar product according to [Alt16, Section 3.27–3.29]

$$\langle v\,,\, w\rangle_{H^1(\Omega)} := \int_\Omega \big(v(x)\,w(x) + \nabla v(x)\cdot\nabla w(x)\big)\,\mathrm{d}x \quad \text{for all } v, w \in C^\infty(\overline{\Omega})$$

with the induced norm

$$\|v\|^2_{H^1(\Omega)} \coloneqq \langle v\,,\, v\rangle_{H^1(\Omega)}.$$

The Sobolev space $H^1(\Omega)$ is defined by the closure cl with respect to the $\|\cdot\|_{H^1(\Omega)}$-norm as $H^1(\Omega) \coloneqq \mathrm{cl}\{v \in C^\infty(\overline{\Omega}) \mid \|v\|_{H^1(\Omega)} < \infty\}$. Analogously, the Sobolev space $H^1_0(\Omega)$ is defined as the closure of $C^\infty_0(\overline{\Omega})$ with respect to the $\|\cdot\|_{H^1(\Omega)}$-norm.

**Jump term.** The normal *jumps* across a face $F$ of a triangulation $\mathcal{T}_H$ that is shared by two neighboring elements $T_1$ and $T_2$ is defined as

$$[[v]] \coloneqq [[v \cdot \boldsymbol{n}]]_F \coloneqq v|_{T_1}\,\boldsymbol{n}_1 + v|_{T_2}\,\boldsymbol{n}_2,$$

where $\boldsymbol{n}_i$ denotes the outer normal on $\partial T_i$ for $i = 1, 2$.

**Inequality and equality up to constants.** Frequently, we will make use of the notation $a \lesssim b$ for $a, b \in \mathbb{R}$, if there exists a constant $C > 0$ that is clear from the context such that $a \leq Cb$. Moreover, if $a \lesssim b$ and $b \lesssim a$, we write $a \simeq b$.

**Approximation.** We use $u^\star \approx u_H \in \mathcal{X}_H$ as an abbreviation for the phrase that $u_H \in \mathcal{X}_H$ is an approximation of $u^\star$. This relation is not symmetric and context-sensitive.

# 2 Rate-optimal goal-oriented adaptive FEM for semilinear elliptic PDEs

This chapter is taken from:

> [①GOA]: 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.*, 118:18–35, 2022. DOI: 10.1016/j.camwa.2022.05.008

The reference [BGIP23] was updated from the preprint to the publication.

## 2.1 Introduction

### 2.1.1 Goal-oriented adaptive FEM

While standard adaptivity aims to approximate the exact solution $u^\star \in H^1_0(\Omega)$ of a suitable PDE at optimal rate in the energy norm (see, e.g., [Dör96; MNS00; BDD04; Ste07; CKNS08] for some seminal contributions and [FFP14] for the present model problem), goal-oriented adaptivity aims to approximate, at optimal rate, only the functional value $G(u^\star) \in \mathbb{R}$ (also called *quantity of interest* in the literature). Usually, goal-oriented adaptivity is more important in practice than standard adaptivity and, therefore, has attracted much interest also in the mathematical literature; see, e.g., [BR03; EEHJ95; GS02; BR01] for some prominent works and [KVD19; ELW19; ELW20; DBR21; BGIP23; BMZ21] for some recent contributions. Often, dual-weighted residual (DWR) estimators are used for goal-oriented adaptivity [BR03; GS02; ELW19; ELW20]. One drawback of such an approach, however, is that it requires an approximation of the dual solution to make the DWR estimator computable. Instead, the present work takes a different route following the seminal paper [MS09] and only employs computable error estimators via a suitably modified dual problem.

Unlike standard adaptivity, there are only few works that aim for a thorough mathematical understanding of optimal rates for goal-oriented adaptivity; see [MS09; BET11; FFGHP16; FPZ16] for linear problems with linear goal functional and [BIP21] for a linear problem, but nonlinear goal functional. The works [HPZ15; XHYM21] consider semilinear PDEs and linear goal functionals, but only prove convergence, while optimal convergence rates remain open (and can hardly be proved for the proposed algorithms). The present work proves, for the first time, optimal convergence rates for goal-oriented adaptivity for a *nonlinear* problem. To this end, we see, in particular, that the marking strategy used in [HPZ15; XHYM21] must be modified along the ideas of [BIP21].

### 2.1.2 Model problem

For $d \in \{1,2,3\}$, let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain. Given $f, g \in L^2(\Omega)$ and $\boldsymbol{f}, \boldsymbol{g} \in [L^2(\Omega)]^d$, we aim to approximate the linear goal quantity

$$G(u^\star) := \int_\Omega g u^\star \,\mathrm{d}x + \int_\Omega \boldsymbol{g} \cdot \nabla u^\star \,\mathrm{d}x, \tag{2.1}$$

where $u^\star \in H^1_0(\Omega)$ is the weak solution of the semilinear elliptic PDE

$$-\operatorname{div}(\boldsymbol{A}\nabla u^\star) + b(u^\star) = f - \operatorname{div}\boldsymbol{f} \text{ in } \Omega \quad \text{subject to} \quad u^\star = 0 \text{ on } \Gamma := \partial\Omega. \tag{2.2}$$

While the precise assumptions on the coefficients $\boldsymbol{A}\colon \Omega \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ and $b\colon \Omega \times \mathbb{R} \to \mathbb{R}$ are given in Section 2.2.1–2.2.2, we note that, here and below, we abbreviate $\boldsymbol{A}\nabla u^\star \equiv \boldsymbol{A}(\cdot)\nabla u^\star(\cdot)\colon \Omega \to \mathbb{R}^d$ and $b(u^\star) \equiv b(\cdot, u^\star(\cdot))\colon \Omega \to \mathbb{R}$.

The weak formulation of the so-called *primal problem* (2.2) reads as follows: Find $u^\star \in H^1_0(\Omega)$ such that

$$\langle\!\langle u^\star, v\rangle\!\rangle + \langle b(u^\star), v\rangle = F(v) := \langle f, v\rangle + \langle \boldsymbol{f}, \nabla v\rangle \quad \text{for all } v \in H^1_0(\Omega), \tag{2.3}$$

where $\langle v, w\rangle := \int_\Omega vw \,\mathrm{d}x$ denotes the $L^2(\Omega)$-scalar product and $\langle\!\langle v, w\rangle\!\rangle := \langle \boldsymbol{A}\nabla v, \nabla w\rangle$ is the $\boldsymbol{A}$-induced energy scalar product on $H^1_0(\Omega)$. We stress that existence and uniqueness of the solution $u^\star \in H^1_0(\Omega)$ of (2.3) follow from the Browder–Minty theorem on monotone operators (see Section 2.2.4).

Based on conforming triangulations $\mathcal{T}_H$ of $\Omega$ and a fixed polynomial degree $m \in \mathbb{N}$, let $\mathcal{X}_H := \{v_H \in H^1_0(\Omega) \mid \forall T \in \mathcal{T}_H\colon v_H|_T \text{ is a polynomial of degree } \le m\}$. Then, the FEM discretization of the primal problem (2.3) reads: Find $u^\star_H \in \mathcal{X}_H$ such that

$$\langle\!\langle u^\star_H, v_H\rangle\!\rangle + \langle b(u^\star_H), v_H\rangle = F(v_H) \quad \text{for all } v_H \in \mathcal{X}_H. \tag{2.4}$$

This allows to approximate the sought goal quantity $G(u^\star)$ by means of the computable quantity $G(u^\star_H)$.

### 2.1.3 Error control and GOAFEM algorithm

The optimal error control of the goal error $G(u^\star) - G(u^\star_H)$ involves the so-called *(practical) dual problem*: Find $z^\star[u^\star_H] \in H^1_0(\Omega)$ such that

$$\langle\!\langle z^\star[u^\star_H], v\rangle\!\rangle + \langle b'(u^\star_H) z^\star[u^\star_H], v\rangle = G(v) \quad \text{for all } v \in H^1_0(\Omega), \tag{2.5}$$

where $b'(x,t) := \partial_t b(x,t)$. Existence and uniqueness of $z^\star[u^\star_H]$ follow from the Lax–Milgram lemma (see Section 2.2.5). With the same FEM spaces as for the primal problem, the FEM discretization of the dual problem (2.5) reads: Find $z^\star_H[u^\star_H] \in \mathcal{X}_H$ such that

$$\langle\!\langle z^\star_H[u^\star_H], v_H\rangle\!\rangle + \langle b'(u^\star_H) z^\star_H[u^\star_H], v_H\rangle = G(v_H) \quad \text{for all } v_H \in \mathcal{X}_H. \tag{2.6}$$

The notation $z^\star[u_H^\star]$ emphasizes that the dual solution depends on the (exact) discrete primal solution $u_H^\star$ (instead of the practically unavailable exact primal solution $u^\star$); the same holds for the discrete dual solution $z_H^\star[u_H^\star]$.

For this setting, we derive below (see Theorem 2.7) the goal error estimate

$$|G(u^\star) - G(u_H^\star)| \lesssim \|u^\star - u_H^\star\|_{H^1(\Omega)} \|z^\star[u_H^\star] - z_H^\star[u_H^\star]\|_{H^1(\Omega)} + \|u^\star - u_H^\star\|_{H^1(\Omega)}^2 \tag{2.7}$$

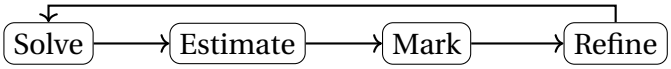
where $\lesssim$ denotes $\leq$ up to some generic multiplicative constant $C > 0$. While the product of primal and dual error is also present for goal error estimates for linear PDEs (see, e.g., [GS02; MS09; FPZ16]), the last summand on the right-hand side of (2.7) controls the linearization error of the (practical) dual problem. The arising error terms are controlled by standard residual *a posteriori* error estimates (see Section 2.3.2), i.e.,

$$\|u^\star - u_H^\star\|_{H^1(\Omega)} \lesssim \eta_H(u_H^\star) \quad \text{and} \quad \|z^\star[u_H^\star] - z_H^\star[u_H^\star]\|_{H^1(\Omega)} \lesssim \zeta_H(z_H^\star[u_H^\star]).$$

Hence, (2.7) gives rise to the fully computable error bound

$$|G(u^\star) - G(u_H^\star)| \lesssim \eta_H(u_H^\star) \, [\eta_H(u_H^\star)^2 + \zeta_H(z_H^\star[u_H^\star])^2]^{1/2} \tag{2.8}$$

that, following [MS09; FPZ16; BIP21], is used to steer an adaptive loop of the type




(2.9)

### 2.1.4 Outline

This work is organized as follows: In Section 2.2, the analytical preliminaries for the semilinear setting and its linearizations are presented. This includes the precise assumptions on the PDE and the right-hand sides as well as well-posedness of the arising continuous and discrete problems. In Section 2.2.7, the key estimate (2.7) is proved; cf. Theorem 2.7. In Section 2.3, we formulate the GOAFEM algorithm (cf. Algorithm 2.17), which employs a marking strategy that respects the product structure found in (2.8). We proceed with stating the main results. First, Theorem 2.19 shows linear convergence of the proposed algorithm. Second, Theorem 2.20 shows optimal convergence rates. Section 2.4 is devoted to the proofs of the aforementioned results, which contain the axioms of adaptivity [CFPP14] for the semilinear setting (Section 2.4.1), a stability result for the linearized dual problem (Section 2.4.2),which turns out to be important, and the necessary quasi-orthogonalities (Section 2.4.5). Numerical experiments in 1D and 2D underline our theoretical findings in Section 2.5. Finally, some conclusions are drawn in Section 2.6.

### 2.1.5 General notation

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 $\omega \subseteq \overline{\Omega}$, depending on the respective context. Furthermore, $\#\mathcal{U}$ denotes the cardinality of a finite set $\mathcal{U}$.

## 2.2 Model problem

### 2.2.1 Assumptions on diffusion coefficient

The diffusion coefficient $\boldsymbol{A}\colon \Omega \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ satisfies the following standard assumptions:

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

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

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

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

### 2.2.2 Assumptions on the nonlinear reaction coefficient

The nonlinearity $b\colon \Omega\times\mathbb{R} \to \mathbb{R}$ satisfies the following assumptions, which follow [BHSZ11, (A1)–(A3)]:

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

- for any $\xi \in \mathbb{R}$, the function $x \mapsto \partial^n_\xi b(x,\xi)$ is measurable on $\Omega$,
- for any $x \in \Omega$, the function $\xi \mapsto \partial^n_\xi b(x,\xi)$ is smooth.

(MON) We assume monotonicity in the second argument, i.e., $b'(x,\xi) := \partial_\xi b(x,\xi) \ge 0$ for all $x \in \Omega$ and $\xi \in \mathbb{R}$. In order to avoid technicalities[1], we assume that $b(x,0) = 0$.

To establish continuity of $\langle b(v)\,,\, w\rangle_\Omega$ resp. $\langle b'(v)\varphi\,,\, w\rangle_\Omega$, we impose the following growth condition on $b(v)$; see, e.g., [FK80, Chapter III, (12)] or [BHSZ11, (A4)]:

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

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

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 $N \in \{2,3\}$.

[1] The assumption $b(0) = 0$ is without loss of generality, since we could consider $\tilde b(v) := b(v) - b(0)$ and $\tilde f := f - b(0)$ instead.

**Remark 2.1.** (i) *Let $v, w \in H^1_0(\Omega)$. To establish continuity of $(v,w) \mapsto \langle b(v)\,,\, w\rangle$, we apply the Hölder inequality with $1 \le s, s' := s/(s-1) \le \infty$ to obtain that*

$$\langle b(v), w\rangle \le \|b(v)\|_{L^{s'}(\Omega)}\|w\|_{L^s(\Omega)} \overset{\text{(GC)}}{\lesssim} (1+\|v^n\|_{L^{s'}(\Omega)})\|w\|_{L^s(\Omega)} = (1+\|v\|^n_{L^{ns'}(\Omega)})\|w\|_{L^s(\Omega)}. \tag{2.10}$$

*To guarantee that $\langle b(v)\,,\, w\rangle < \infty$, condition (GC) has to ensure that the embedding*

$$H^1_0(\Omega) \hookrightarrow L^r(\Omega) \quad \text{is continuous} \quad \text{for} \quad r = s \quad \text{and} \quad r = ns'. \tag{2.11}$$

*If $d \in \{1,2\}$, then (2.11) holds true for arbitrary $1 \le r < \infty$ and hence arbitrary $1 < s < \infty$ and $n \in \mathbb{N}$. If $d = 3$, then $r = s = 6$ is the maximal index in (2.11) and $s' = 6/5$. Hence, it follows necessarily that $n \le 6/s' = 5$. Furthermore, if $d = 3$, note that it suffices to consider $n = 5$ since, for $n < 5$, we can estimate $(1+|\xi|^{n-k}) \lesssim (1+|\xi|^{5-k})$ for all $\xi \in \mathbb{R}$, and all $0 \le k \le n < 5$. Altogether, we conclude continuity of $(v,w) \mapsto \langle b(v)\,,\, w\rangle$ for all $n \in \mathbb{N}$ if $d \in \{1,2\}$, and $n \le 5$ if $d = 3$.*
(ii) *Let $v, w, \varphi \in H^1_0(\Omega)$. In the same spirit, we establish continuity of $(v,w,\varphi) \mapsto \langle b'(v)\,\varphi, w\rangle$. If $d \in \{1,2\}$, for arbitrary $1 < t < \infty$, we use the generalized Hölder inequality; see, e.g., [KJF77, Section 2.2]. To this end, define $t''$ by $1 = 1/t'' + 1/t + 1/t$ and observe that*

$$\langle b'(v)\,\varphi\,,\, w\rangle \le \|b'(v)\|_{L^{t''}(\Omega)}\|\varphi\|_{L^t(\Omega)}\|w\|_{L^t(\Omega)} \overset{\text{(GC)}}{\lesssim} (1+\|v^{n-1}\|_{L^{t''}(\Omega)})\|\varphi\|_{L^t(\Omega)}\|w\|_{L^t(\Omega)}. \tag{2.12}$$

*Using $\|v^{n-1}\|_{L^{t''}(\Omega)} = \|v\|^{n-1}_{L^{(n-1)t''}(\Omega)}$, the (GC) needs to ensure that the Sobolev embedding $H^1_0(\Omega) \hookrightarrow L^r(\Omega)$ is continuous for both $r = (n-1)t''$ and $r = t$. If $d \in \{1,2\}$, this holds for arbitrary $1 < t < \infty$ and $n \in \mathbb{N}$. If $d = 3$, then $r = t = 6$ is the maximal index in (2.11) and hence $t'' = 3/2$. The upper bound $(n-1) \le 6/t'' = 4$ thus guarantees continuity.*
(iii) *Let $v, \varphi \in H^1_0(\Omega)$ and $w \in L^\infty(\Omega)$. Then, the reasoning of* (ii) *reduces to the Hölder conjugates from* (i).
(iv) *The additional constraints on the upper bounds of $n$ in (CGC) will become apparent later; see Remark 2.32.*
(v) *The lower bound $2 \le n$ imposed for $d \in \{1,2,3\}$ stems from the necessity of a Taylor expansion of the dual problem; cf. (2.45).* □

### 2.2.3 Assumptions on the right-hand sides

For $d = 1$, the exact solution $u^\star$ from (2.3) and the dual solutions $\tilde z^\star[w]$ and $z^\star[w]$ with arbitrary $w \in H^1_0(\Omega)$ from (2.15) and (2.18) below satisfy $L^\infty$-bounds, 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 fulfills that

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

To guarantee later that the residual *a posteriori* error estimators from (2.54)–(2.55) are well-defined, we additionally require that $\boldsymbol{f}|_T, \boldsymbol{g}|_T \in H(\mathrm{div}, T)$ with traces $\boldsymbol{f}|_T\cdot\boldsymbol{n}$, $\boldsymbol{g}|_T\cdot\boldsymbol{n} \in L^2(\partial T)$ for all $T \in \mathcal{T}_0$, where $\mathcal{T}_0$ is the initial triangulation of the adaptive algorithm.

### 2.2.4 Well-posedness of primal problem

First, we deal with the continuous primal problem (2.3). With the dual space $H^{-1}(\Omega) := H^1_0(\Omega)^*$, we consider the operator

$$\mathcal{A}\colon H^1_0(\Omega) \to H^{-1}(\Omega), \quad \mathcal{A}w := \langle\!\langle w\,,\,\cdot\,\rangle\!\rangle + \langle b(w)\,,\,\cdot\,\rangle. \tag{2.13}$$

The assumption (GC) implies that $\mathcal{A}$ is well-defined, (ELL) and (MON) yield that $\mathcal{A}$ is strongly monotone, and (CAR) is used to show that $\mathcal{A}$ is hemi-continuous. Overall, the Browder–Minty theorem (see, e.g., [Zei90, Theorem 26.A (a)–(c)]) applies and proves that the primal problem (2.3) admits a unique solution $u^\star \in H^1_0(\Omega)$. The same argument shows that the discrete primal problem (2.4) admits a unique solution $u^\star_H \in \mathcal{X}_H$. Details are provided in Appendix 2.7.

### 2.2.5 Well-posedness of dual problem and goal error identity

For $v, w \in H^1_0(\Omega)$, define

$$\boldsymbol{B}(w,v) := \int_0^1 b'\big(w + (v-w)\tau\big)\,\mathrm{d}\tau \geq 0 \quad \text{a.e. in } \Omega. \tag{2.14}$$

Note that $\boldsymbol{B}(w,v)\colon \Omega \to \mathbb{R}_{\geq 0}$. If $v = u^\star$ is the exact primal solution, we introduce the shorthand $\boldsymbol{B}^\star(w) := \boldsymbol{B}(w,u^\star)$. With this notation, the *theoretical* dual problem reads as follows: Find $\tilde z^\star[w] \in H^1_0(\Omega)$ and $\tilde z^\star_H[w] \in \mathcal{X}_H$ such that

$$\langle\!\langle \tilde z^\star[w]\,,\,v\rangle\!\rangle + \langle \boldsymbol{B}^\star(w)\tilde z^\star[w]\,,\,v\rangle = G(v) \quad \text{for all } v \in H^1_0(\Omega), \tag{2.15a}$$

$$\langle\!\langle \tilde z^\star_H[w]\,,\,v_H\rangle\!\rangle + \langle \boldsymbol{B}^\star(w)\tilde z^\star_H[w]\,,\,v_H\rangle = G(v_H) \quad \text{for all } v_H \in \mathcal{X}_H. \tag{2.15b}$$

Under the assumptions (ELL), (MON), and (GC), the Lax–Milgram lemma proves existence and uniqueness of $\tilde z^\star[w] \in H^1_0(\Omega)$ and $\tilde z^\star_H[w] \in \mathcal{X}_H$. Details are found in Appendix 2.7.

According to the Taylor theorem, it holds that

$$b(u^\star) - b(w) = (u^\star - w)\,\boldsymbol{B}^\star(w) \quad \text{in } \Omega. \tag{2.16}$$

For any approximation $\tilde z^\star[u^\star_H] \approx z_H \in \mathcal{X}_H$, this yields the error identity

$$\begin{aligned} G(u^\star) - G(u^\star_H) &\overset{(2.15a)}{=} \langle\!\langle \tilde z^\star[u^\star_H]\,,\,u^\star - u^\star_H\rangle\!\rangle + \langle \boldsymbol{B}^\star(u^\star_H)\tilde z^\star[u^\star_H]\,,\,u^\star - u^\star_H\rangle \\ &\overset{(2.16)}{=} \langle\!\langle u^\star - u^\star_H\,,\,\tilde z^\star[u^\star_H]\rangle\!\rangle + \langle b(u^\star) - b(u^\star_H)\,,\,\tilde z^\star[u^\star_H]\rangle \\ &\overset{(2.3),(2.4)}{=} \langle\!\langle u^\star - u^\star_H\,,\,\tilde z^\star[u^\star_H] - z_H\rangle\!\rangle + \langle b(u^\star) - b(u^\star_H)\,,\,\tilde z^\star[u^\star_H] - z_H\rangle. \end{aligned} \tag{2.17}$$

While this error identity looks similar to the one for linear problems (see, e.g., [MS09; BET11; FPZ16] as well as [BGIP23] in the presence of inexact solvers), we stress that it suffers from one essential shortcoming: The theoretical dual problem (2.15) involves $\boldsymbol{B}^\star(u^\star_H) = \boldsymbol{B}(u^\star_H, u^\star)$ which depends on the unknown exact solution $u^\star$. Consequently, the corresponding bilinear form cannot be implemented in practice and, hence, $\tilde z^\star[u^\star_H]$

cannot be approximated by its FEM solution $\tilde{z}_H^\star[u_H^\star]$.

However, it follows formally that $\boldsymbol{B}^\star(u_H^\star) - b'(u_H^\star) \to 0$ as $u_H^\star \to u^\star$. Hence, we introduce the *practical* dual problem (2.5) and its discretization (2.6), now considered for a general argument: Given $w \in H_0^1(\Omega)$, find $z^\star[w] \in H_0^1(\Omega)$ and $z_H^\star[w] \in \mathcal{X}_H$ such that

$$\langle\!\langle z^\star[w]\,,\,v\rangle\!\rangle + \langle b'(w)z^\star[w]\,,\,v\rangle = G(v) \quad \text{for all } v \in H_0^1(\Omega), \tag{2.18a}$$
$$\langle\!\langle z_H^\star[w]\,,\,v_H\rangle\!\rangle + \langle b'(w)z_H^\star[w]\,,\,v_H\rangle = G(v_H) \quad \text{for all } v_H \in \mathcal{X}_H. \tag{2.18b}$$

The same arguments as for the theoretical problem (2.15) apply and prove existence and uniqueness of $z^\star[w] \in H_0^1(\Omega)$ and $z_H^\star[w] \in \mathcal{X}_H$. Details are found in Appendix 2.7.

Overall, the error identity (2.17) for $z_H = z_H^\star[u_H^\star]$ then takes the following form

$$\begin{aligned} G(u^\star) - G(u_H^\star) \overset{(2.17)}{=} &\langle\!\langle u^\star - u_H^\star, z^\star[u_H^\star] - z_H^\star[u_H^\star]\rangle\!\rangle + \langle b(u^\star) - b(u_H^\star), z^\star[u_H^\star] - z_H^\star[u_H^\star]\rangle \\ &+ \langle\!\langle u^\star - u_H^\star, \tilde{z}^\star[u_H^\star] - z^\star[u_H^\star]\rangle\!\rangle + \langle b(u^\star) - b(u_H^\star), \tilde{z}^\star[u_H^\star] - z^\star[u_H^\star]\rangle. \end{aligned} \tag{2.19}$$

This identity will be the starting point for proving the goal error estimate (2.7); see Theorem 2.7 below for the formal statement.

### 2.2.6 Pointwise boundedness of primal and dual solutions

In this section, we prove that imposing regularity assumptions on the right-hand side yields that the exact solution $u^\star$ and the dual solutions $\tilde{z}^\star[w]$ and $z^\star[w]$ are bounded in $L^\infty(\Omega)$. For $d = 1$, this is immediate, since $H^1(\Omega) \hookrightarrow C(\overline{\Omega})$. For $d \in \{2,3\}$ and $\boldsymbol{f} = 0 = \boldsymbol{g}$, we refer to, e.g., [BHSZ11, Theorem 2.2]. These $L^\infty$-bounds turn out to be crucial for the goal error estimate (Theorem 2.7) as well as for the numerical analysis of the proposed adaptive goal-oriented strategy (Algorithm 2.17). In particular, they also allow one to derive Céa-type estimates for the discrete primal and dual solutions (Proposition 2.11, 2.12).

**Proposition 2.2.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Then, the weak solution* $u^\star \in H_0^1(\Omega)$ *of* (2.3) *is bounded in* $L^\infty(\Omega)$. *In particular, for* $d \in \{2,3\}$, *there holds*

$$\|u^\star\|_{L^\infty(\Omega)} \le C\,\mu_0^{-1}\,|\Omega|^{(1/d-1/p)}\,(\|f\|_{L^q(\Omega)} + \|\boldsymbol{f}\|_{L^p(\Omega)}) \tag{2.20}$$

*with a constant* $C = C(d,p) > 0$.

**Remark 2.3.** *In this remark, we consider special choices of p and q from* (RHS). *If* $d = 2$ *and* $p = \infty$, *then* $q = 2$. *If* $d = 3$ *and* $p = 6$, *then also* $q = 2$. *In [BHSZ11, Theorem 2.2], the following statement is proven with a slightly simplified proof: Suppose* $f \in L^2(\Omega)$ *and* $\boldsymbol{f} = 0$ *as well as* (ELL), (CAR), *and* (MON). *Then, the weak solution* $u^\star \in H_0^1(\Omega)$ *of* (2.3) *satisfies* $\|u^\star\|_{L^\infty(\Omega)} \le C(\Omega,d)\,\mu_0^{-1}\,\|f\|_{L^2(\Omega)}$. □

The proof of Proposition 2.2 requires the following elementary result from [WYW06, Lemma 4.1.1]:

**Lemma 2.4.** *With positive constants $C, \kappa_0 > 0$ and $\kappa_1 > 1$, let $\phi\colon \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ satisfy*

$$0 \leq \phi(\Lambda) \leq \phi(\lambda) \quad \text{and} \quad \phi(\Lambda) \leq \left(\frac{C}{\Lambda - \lambda}\right)^{\kappa_0} \phi(\lambda)^{\kappa_1} \quad \text{for all } 0 \leq \lambda < \Lambda. \tag{2.21}$$

*Then, there holds that $\phi(\Lambda) = 0$ for all $\Lambda \geq C\,\phi(0)^{(\kappa_1-1)/\kappa_0}\,2^{\kappa_1/(\kappa_1-1)}$.* □

Furthermore, the proof of Proposition 2.2 requires the Gagliardo–Nirenberg–Sobolev inequality; see, e.g., [FK80, Theorem 16.6] or [Chi09, Theorem 12.7]:

**Lemma 2.5** (Gagliardo–Nirenberg–Sobolev inequality)**.** *Let $\Omega \subseteq \mathbb{R}^d$ be open and bounded and suppose that $1 \leq p < \infty$. If $1 \leq p < d$, let $1 \leq r \leq p^* := dp/(d-p) < \infty$. If $d \leq p < \infty$, let $1 \leq r < \infty$. Then, there exists a constant $C'_{\mathrm{GNS}} = C'_{\mathrm{GNS}}(d,p,r)$ such that*

$$\|v\|_{L^r(\Omega)} \leq C_{\mathrm{GNS}}\,\|\nabla v\|_{L^p(\Omega)} \quad \text{for all } v \in W^{1,p}_0(\Omega), \tag{2.22}$$

*where $C_{\mathrm{GNS}} := C'_{\mathrm{GNS}}\,|\Omega|^{1/d-1/p+1/r}$. If $1 \leq p < d$ and $r = p^*$, then $C_{\mathrm{GNS}} = p(d-1)/(d-p)$ depends only on $d$ and $p$.* □

*Proof of Proposition 2.2.* If $d = 1$, $u^\star \in C(\overline{\Omega}) \subset L^\infty(\Omega)$ holds due to the Sobolev embedding. If $d \in \{2,3\}$ and for $\lambda \geq 0$, we define the test function

$$\varphi^+_\lambda(x) := \max\{u^\star(x) - \lambda, 0\}$$

and recall from [Chi09, Theorem 12.4] that $\varphi^+_\lambda \in H^1_0(\Omega)$ with

$$\nabla \varphi^+_\lambda(x) = \nabla u^\star(x) \quad \text{for almost all } x \in \Omega(\lambda) := \{x \in \Omega \mid u^\star(x) > \lambda\}. \tag{2.23}$$

The mean value theorem and (MON) prove for some $\min\{\xi_1,\xi_2\} < \zeta < \max\{\xi_1,\xi_2\}$ that

$$\big(b(\xi_2) - b(\xi_1)\big)(\xi_2 - \xi_1) = b'(\zeta)(\xi_2 - \xi_1)^2 \geq 0 \quad \text{for all } \xi_1, \xi_2 \in \mathbb{R}. \tag{2.24}$$

Hence, it follows that $\big\langle b(u^\star) - b(\lambda)\,,\, u^\star - \lambda\big\rangle_{\Omega(\lambda)} \geq 0$. Using (MON), we see that $b(\lambda) \geq b(0) = 0$ and hence $\big\langle b(\lambda)\,,\, u^\star - \lambda\big\rangle_{\Omega(\lambda)} \geq 0$. Using the coercivity assumption (ELL) and testing the weak formulation (2.3) with $\varphi^+_\lambda$, we observe that

$$\begin{aligned}
\mu_0\|\nabla \varphi^+_\lambda\|^2_{L^2(\Omega)} &\overset{\text{(ELL)}}{\leq} \langle\!\langle \varphi^+_\lambda\,,\, \varphi^+_\lambda\rangle\!\rangle = \langle\!\langle u^\star\,,\, \varphi^+_\lambda\rangle\!\rangle \overset{(2.3)}{=} \langle f\,,\, \varphi^+_\lambda\rangle + \langle \boldsymbol{f}\,,\, \nabla\varphi^+_\lambda\rangle - \langle b(u^\star)\,,\, \varphi^+_\lambda\rangle \\
&= \langle f\,,\, \varphi^+_\lambda\rangle + \langle \boldsymbol{f}\,,\, \nabla\varphi^+_\lambda\rangle - \langle b(u^\star)\,,\, u^\star - \lambda\,\rangle_{\Omega(\lambda)} \\
&= \langle f\,,\, \varphi^+_\lambda\rangle + \langle \boldsymbol{f}\,,\, \nabla\varphi^+_\lambda\rangle - \langle b(u^\star) - b(\lambda)\,,\, u^\star - \lambda\,\rangle_{\Omega(\lambda)} - \langle b(\lambda)\,,\, u^\star - \lambda\,\rangle_{\Omega(\lambda)} \\
&\leq \langle f\,,\, \varphi^+_\lambda\rangle + \langle \boldsymbol{f}\,,\, \nabla\varphi^+_\lambda\rangle.
\end{aligned} \tag{2.25}$$

With the Hölder inequality, we arrive at

$$\mu_0\|\nabla \varphi^+_\lambda\|^2_{L^2(\Omega)} \leq \|f\|_{L^q(\Omega)}\,\|\varphi^+_\lambda\|_{L^{q'}(\Omega)} + \|\boldsymbol{f}\|_{L^p(\Omega)}\,\|\nabla\varphi^+_\lambda\|_{L^{p'}(\Omega)}. \tag{2.26}$$

Moreover, (RHS) yields that $1/q' = 1 - 1/q = 1 - 1/p - 1/d = 1/p' - 1/d = 1/p'^*$, where $p' < 2 \leq d$. Since $|\Omega| < \infty$, we have that $H^1_0(\Omega) \hookrightarrow W^{1,p'}_0(\Omega)$. Therefore, Lemma 2.5 (applied

to $1 \le p' < d$ and $r = p'^* = q'$) yields that

$$\|\varphi_\lambda^+\|_{L^{q'}(\Omega)} \le C_{\mathrm{GNS}} \|\nabla \varphi_\lambda^+\|_{L^{p'}(\Omega)}, \tag{2.27}$$

where $C_{\mathrm{GNS}}$ depends only on $d, p'$. Collecting (2.25)–(2.27), we obtain that

$$\|\nabla \varphi_\lambda^+\|_{L^2(\Omega)}^2 \le C_1 \|\nabla \varphi_\lambda^+\|_{L^{p'}(\Omega)} \quad \text{with} \quad C_1 = \frac{\max\{C_{\mathrm{GNS}}, 1\}}{\mu_0} \left(\|f\|_{L^q(\Omega)} + \|\boldsymbol{f}\|_{L^p(\Omega)}\right). \tag{2.28}$$

Now, we aim at a lower bound for the left-hand side of (2.28). Recall the definition of $\Omega(\lambda)$ from (2.23). Applying the Hölder inequality, we observe that

$$\|\nabla \varphi_\lambda^+\|_{L^{p'}(\Omega)}^{p'} = \int_{\Omega(\lambda)} |\nabla \varphi_\lambda^+|^{p'} \,\mathrm{d}x \le \left(\int_{\Omega(\lambda)} |\nabla \varphi_\lambda^+|^2 \,\mathrm{d}x\right)^{p'/2} |\Omega(\lambda)|^{1-p'/2}.$$

Taking the last equation to the power of $2/p' > 1$, we show that

$$\|\nabla \varphi_\lambda^+\|_{L^{p'}(\Omega)}^2 \le \|\nabla \varphi_\lambda^+\|_{L^2(\Omega)}^2 |\Omega(\lambda)|^{2/p'-1} \overset{(2.28)}{\le} C_1 C_{\mathrm{GNS}} \|\nabla \varphi_\lambda^+\|_{L^{p'}(\Omega)} |\Omega(\lambda)|^{2/p'-1}. \tag{2.29}$$

In combination with (2.27), we arrive at

$$\|\varphi_\lambda^+\|_{L^{q'}(\Omega)} \le C_{\mathrm{GNS}} \|\nabla \varphi_\lambda^+\|_{L^{p'}(\Omega)} \le C_1 C_{\mathrm{GNS}}^2 |\Omega(\lambda)|^{2/p'-1}. \tag{2.30}$$

For $0 < \lambda < \Lambda$, we observe that

$$\Omega(\lambda) \supseteq \Omega(\Lambda).$$

This observation and $u^\star > \Lambda$ on $\Omega(\Lambda)$ provide a lower bound for the left-hand side of (2.30):

$$\|\varphi_\lambda^+\|_{L^{q'}(\Omega)} \ge \left(\int_{\Omega(\Lambda)} \left(\max\{u^\star(x) - \lambda, 0\}\right)^{q'} \,\mathrm{d}x\right)^{1/q'} \ge (\Lambda - \lambda)|\Omega(\Lambda)|^{1/q'}.$$

Combining this estimate with (2.30), we see that

$$|\Omega(\Lambda)| \le \left(\frac{C_2}{\Lambda - \lambda}\right)^{q'} |\Omega(\lambda)|^{q'(2/p'-1)} \quad \text{with} \quad C_2 = C_1 C_{\mathrm{GNS}}. \tag{2.31}$$

Recall that $1/\kappa_0 := 1/q' = 1 - 1/p - 1/d$. Together with $p > d \ge 2$, we thus observe that

$$\kappa_1 := q'(2/p' - 1) = (2 - 2/p - 1)/(1 - 1/p - 1/d) = (1 - 2/p)/(1 - 1/p - 1/d) > 1.$$

Therefore, we are able to apply Lemma 2.4 to (2.31). This yields that $|\Omega(\Lambda)| = 0$ for $\Lambda \ge C_2 |\Omega(0)|^{(\kappa_1-1)/\kappa_0} 2^{\kappa_1/(\kappa_1-1)}$. By definition of $\Omega(\Lambda)$, this proves that

$$u^\star(x) \le C_2 |\Omega|^{(1/d-1/p)} 2^{(1/d-1/p)/(1-2/p)} \quad \text{for almost all } x \in \Omega.$$

To see that $-u^\star$ satisfies the same bound, we argue analogously. For $\lambda \ge 0$, we define the test function $\varphi_\lambda^- := \min\{u^\star(x) - \lambda, 0\} \le 0$ and observe that $\varphi_\lambda^- \in H_0^1(\Omega)$. With $\Omega(\lambda) := \{x \in \Omega \mid u(x) < -\lambda\}$ and the above arguments, we then conclude the proof. □

In the same spirit as in Proposition 2.2, we are able to establish $L^\infty$-bounds for the solutions of the theoretical and practical dual problems (2.15a) and (2.18a).

**Proposition 2.6.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Let* $w \in H^1_0(\Omega)$. *Then, the weak solutions* $\tilde z^\star[w] \in H^1_0(\Omega)$ *of the theoretical dual problem* (2.15a) *and* $z^\star[w] \in H^1_0(\Omega)$ *of the practical dual problem* (2.18a) *are bounded in* $L^\infty(\Omega)$. *In particular, for* $d \in \{2,3\}$, *there holds*

$$\|\tilde z^\star[w]\|_{L^\infty(\Omega)} + \|z^\star[w]\|_{L^\infty(\Omega)} \le C\,\mu_0^{-1}\,|\Omega|^{(1/d-1/p)}\,(\|g\|_{L^q(\Omega)} + \|\boldsymbol{g}\|_{L^p(\Omega)}) \tag{2.32}$$

*with a constant* $C = C(d,p) > 0$, *which is, in particular, independent of* $w$.

*Proof.* We argue as for Proposition 2.2. The case $d = 1$ follows from the Sobolev embedding. For $d \in \{2,3\}$ and for $\lambda \ge 0$, we define the test function

$$\varphi_\lambda^+(x) := \max\{\tilde z^\star[w](x) - \lambda, 0\}$$

and recall that $\varphi_\lambda^+ \in H^1_0(\Omega)$ with

$$\nabla\varphi_\lambda^+(x) = \nabla\tilde z^\star[w](x) \quad \text{for almost all } x \in \Omega(\lambda) := \{x \in \Omega \mid \tilde z^\star[w] > \lambda\}.$$

From (2.14), recall that $\boldsymbol{B}^\star(w) = \boldsymbol{B}(w,u^\star) \ge 0$. In particular, it follows that $\langle \boldsymbol{B}^\star(w)\tilde z^\star[w]\,,\,\tilde z^\star[w] - \lambda\rangle_{\Omega(\lambda)} \ge 0$. Using the coercivity assumption (ELL) and testing the weak formulation (2.15a) with $\varphi_\lambda^+$, we observe that

$$\begin{aligned}\mu_0\|\nabla\varphi_\lambda^+\|^2_{L^2(\Omega)} &\overset{\text{(ELL)}}{\le} \langle\!\langle \varphi_\lambda^+, \varphi_\lambda^+\rangle\!\rangle = \langle\!\langle \tilde z^\star[w], \varphi_\lambda^+\rangle\!\rangle \overset{(2.15a)}{=} \langle g, \varphi_\lambda^+\rangle + \langle \boldsymbol{g}, \nabla\varphi_\lambda^+\rangle - \langle \boldsymbol{B}^\star(w)\tilde z^\star[w], \varphi_\lambda^+\rangle \\ &= \langle g\,,\,\varphi_\lambda^+\rangle + \langle \boldsymbol{g}\,,\,\nabla\varphi_\lambda^+\rangle - \langle \boldsymbol{B}^\star(w)\tilde z^\star[w]\,,\,\tilde z^\star[w] - \lambda\,\rangle_{\Omega(\lambda)} \le \langle g\,,\,\varphi_\lambda^+\rangle + \langle \boldsymbol{g}\,,\,\nabla\varphi_\lambda^+\rangle.\end{aligned}$$

Following the steps of the proof of Proposition 2.2 (where the latter estimate corresponds to (2.25)), we conclude the proof for $\tilde z[w]$. The same argument applies for the practical dual problem, where $\boldsymbol{B}^\star(w)$ is replaced by $b'(w) \ge 0$. This concludes the proof. □

### 2.2.7 Goal error estimate

The following theorem provides, up to norm equivalence, the formal statement of the goal error estimate (2.7).

**Theorem 2.7**

*Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Let* $u^\star \in H^1_0(\Omega)$ *solve* (2.3) *and* $u_H^\star \in \mathcal{X}_H$ *be its approximation* (2.4). *Then, it holds that*

$$|G(u^\star) - G(u_H^\star)| \le C_{\mathrm{est}}\,\big[|||u^\star - u_H^\star|||\,|||z^\star[u_H^\star] - z_H^\star[u_H^\star]||| + |||u^\star - u_H^\star|||^2\big]\,, \tag{2.33}$$

*where* $C_{\mathrm{est}} = C_{\mathrm{est}}(|\Omega|, d, \|u^\star\|_{L^\infty(\Omega)}, n, R, p, f, \boldsymbol{f}, g, \boldsymbol{g}, \mu_0)$.

The proof of Theorem 2.7 requires some preparations. We start with the following lemma which extends [BHSZ11, Lemma 3.1] to $\boldsymbol{f} \neq 0$.

**Lemma 2.8.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Let* $w \in H^1_0(\Omega)$. *Then, it holds that*

$$|||u^\star||| + |||u_H^\star||| \le C_{\mathrm{bnd}}, \tag{2.34a}$$

$$|||\tilde z^\star[w]||| + |||\tilde z_H^\star[w]||| + |||z^\star[w]||| + |||z_H^\star[w]||| \le C_{\mathrm{bnd}}, \tag{2.34b}$$

*where* $C_{\mathrm{bnd}} = C_{\mathrm{bnd}}(|\Omega|, d, p, f, \boldsymbol{f}, \mu_0)$ *for* (2.34a) *and* $C_{\mathrm{bnd}} = C_{\mathrm{bnd}}(|\Omega|, d, p, g, \boldsymbol{g}, \mu_0)$ *for* (2.34b). *The constant* $C_{\mathrm{bnd}}$ *is independent of* $w \in H^1_0(\Omega)$.

*Proof.* In the case $d = 1$, (2.34) follows from the Sobolev embedding and (ELL). Moreover, note that $b(0) = 0$ and (2.24) prove that $\langle b(u^\star)\,,\, u^\star\rangle \ge 0$. Using (ELL), (MON), and the Hölder inequality, we obtain that

$$\begin{aligned} |||u^\star|||^2 = \langle\!\langle u^\star\,,\, u^\star\rangle\!\rangle &\overset{(2.3)}{=} \langle f\,,\, u^\star\rangle + \langle \boldsymbol{f}\,,\, \nabla u^\star\rangle - \langle b(u^\star)\,,\, u^\star\rangle \\ &\le \|f\|_{L^q(\Omega)}\, \|u^\star\|_{L^{q'}(\Omega)} + \|\boldsymbol{f}\|_{L^p(\Omega)}\, \|\nabla u^\star\|_{L^{p'}(\Omega)}. \end{aligned} \tag{2.35}$$

Arguing as for (2.27) and applying the Hölder inequality, we see that

$$\begin{aligned} |||u^\star|||^2 &\le \max\{C_{\mathrm{GNS}}, 1\}\, \big(\|f\|_{L^q(\Omega)} + \|\boldsymbol{f}\|_{L^p(\Omega)}\big)\, \|\nabla u^\star\|_{L^{p'}(\Omega)} \\ &\le \max\{C_{\mathrm{GNS}}, 1\}\, \big(\|f\|_{L^q(\Omega)} + \|\boldsymbol{f}\|_{L^p(\Omega)}\big)\, |\Omega|^{1/p'-1/2}\, \|\nabla u^\star\|_{L^2(\Omega)}, \end{aligned} \tag{2.36}$$

where $C_{\mathrm{GNS}}$ depends only on $d$ and $p'$. With $\|\nabla u^\star\|_{L^2(\Omega)} \le \mu_0^{-1/2} |||u^\star|||$, this concludes the proof for $u^\star$. The same argument (based on (2.4) instead of (2.3)) applies for $u_H^\star$. Furthermore, the same argument applies also for the dual problems (based on (2.15) and (2.18) instead of (2.3) for the theoretical and practical dual problem, respectively). For $w, v \in H^1_0(\Omega)$, the monotonicity $\langle b(v)\,,\, v\rangle \ge 0$ is substituted in case of the dual problems by $\langle b'(w)v\,,\, v\rangle \ge 0$ and $\langle \boldsymbol{B}^\star(w)v\,,\, v\rangle \ge 0$, respectively. This concludes the proof. □

The following lemma is one of the two main ingredients for the proof of Theorem 2.7.

**Lemma 2.9.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Let* $w \in H^1_0(\Omega)$ *with* $|||w||| \le M < \infty$. *Then, it holds that*

$$\langle b(u^\star) - b(w)\,,\, v\rangle \le C_{\mathrm{Lip}}\, |||u^\star - w|||\,|||v||| \quad \text{for all } v \in H^1_0(\Omega) \tag{2.37}$$

*with* $C_{\mathrm{Lip}} = C_{\mathrm{Lip}}(|\Omega|, d, \|u^\star\|_{L^\infty(\Omega)}, M, n, R, p, f, \boldsymbol{f}, \mu_0)$.

*Proof.* We argue as in the proof of [BHSZ11, Theorem 3.4]. With respect to Remark 2.1, choose $s > 1$ arbitrarily for $d \in \{1, 2\}$ and $s = 6$ for $d = 3$. In any case, we see that

$$\langle b(u^\star) - b(w)\,,\, v\rangle \le \|b(u^\star) - b(w)\|_{L^{s'}(\Omega)} \|v\|_{L^s(\Omega)} \le C\, \|b(u^\star) - b(w)\|_{L^{s'}(\Omega)} |||v|||, \tag{2.38}$$

where $C := \mu_0^{-1}\, C_{\mathrm{GNS}}$. It remains to prove that

$$\|b(u^\star) - b(w)\|_{L^{s'}(\Omega)} \lesssim |||u^\star - w|||.$$

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

$$b(w) = \sum_{k=0}^{n-1} b^{(k)}(u^\star)\frac{(w-u^\star)^k}{k!} + \frac{(w-u^\star)^n}{(n-1)!}\int_0^1 (1-\tau)^{n-1}\, b^{(n)}\big(u^\star + (w-u^\star)\,\tau\big)\,\mathrm{d}\tau. \tag{2.39}$$

Since $b$ is smooth and $u^\star \in L^\infty(\Omega)$, we obtain that

$$\|b^{(k)}(u^\star)\|_{L^\infty(\Omega)} \le C \quad \text{for all } k = 1, \dots, n-1,$$

where $C$ depends only on $\|u^\star\|_{L^\infty(\Omega)}$, and $n$. Moreover, (GC) allows to bound the remainder term, i.e., for any $0 \le \tau \le 1$, it holds that

$$\|b^{(n)}(u^\star + (w-u^\star)\tau)\|_{L^\infty(\Omega)} \le C,$$

where $C$ depends only on $|\Omega|$, $n$, and $R$. The triangle inequality yields that

$$\|b(u^\star) - b(w)\|_{L^{s'}(\Omega)} \lesssim \sum_{k=1}^{n} \|(u^\star - w)^k\|_{L^{s'}(\Omega)} = \sum_{k=1}^{n} \|u^\star - w\|^k_{L^{ks'}(\Omega)}. \tag{2.40}$$

Recall from Remark 2.1 that $H^1_0(\Omega) \hookrightarrow L^{ks'}(\Omega)$ for all $1 \le k \le n$ by choice of $s$ and $n$. Therefore, the Gagliardo–Nirenberg–Sobolev inequality proves that

$$\|b(u^\star) - b(w)\|_{L^{s'}(\Omega)} \lesssim \sum_{k=1}^{n} \|\nabla(u^\star - w)\|^k_{L^2(\Omega)}, \tag{2.41}$$

where the hidden constant depends only on $|\Omega|$, $d$, $\|u^\star\|_{L^\infty(\Omega)}$, $n$, and $R$. With $\|\nabla(u^\star - w)\|_{L^2(\Omega)} \simeq |||u^\star - w||| \le C_{\mathrm{bnd}} + M$, this leads to

$$\|b(u^\star) - b(w)\|_{L^{s'}(\Omega)} \lesssim \left(\sum_{k=1}^{n} \|\nabla(u^\star - w)\|^{k-1}_{L^2(\Omega)}\right) \|\nabla(u^\star - w)\|_{L^2(\Omega)} \lesssim |||u^\star - w|||, \tag{2.42}$$

with hidden constants $C = C(|\Omega|, d, \|u^\star\|_{L^\infty(\Omega)}, M, n, R, p, f, \boldsymbol{f}, \mu_0) > 0$. Together with (2.38), this concludes the proof of (2.37). □

The following lemma is the last missing part for establishing Theorem 2.7.

**Lemma 2.10.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Let* $w \in H^1_0(\Omega)$ *with* $|||w||| \le M < \infty$. *Then, it holds that*

$$|||\tilde z^\star[w] - z^\star[w]||| \le C_{\mathrm{dual}}\, |||u^\star - w|||, \tag{2.43}$$

*where* $C_{\mathrm{dual}} = C_{\mathrm{dual}}(|\Omega|, d, \|u^\star\|_{L^\infty(\Omega)}, M, n, R, p, f, \boldsymbol{f}, g, \boldsymbol{g}, \mu_0)$.

*Proof.* Define $\delta := \tilde z^\star[w] - z^\star[w] \in H^1_0(\Omega)$. For the exact primal solution $u^\star$, we observe that the theoretical dual problem and the practical dual problem coincide, as $\boldsymbol{B}^\star(u^\star) = \boldsymbol{B}(u^\star, u^\star) = \int_0^1 b'(u^\star)\,\mathrm{d}\tau = b'(u^\star)$ and hence $z[u^\star] = \tilde z[u^\star]$. Using monotonicity and the

definition of the theoretical as well as practical dual problem, we obtain that

$$\begin{aligned}
|||\delta|||^2 = \langle\!\langle \delta\,,\,\delta\rangle\!\rangle &\overset{\text{(MON)}}{\leq} \langle\!\langle \delta\,,\,\delta\rangle\!\rangle + \langle b'(w)\delta\,,\,\delta\rangle \\
&\overset{(2.18a)}{=} [\langle\!\langle \tilde{z}^\star[w]\,,\,\delta\rangle\!\rangle + \langle b'(w)\tilde{z}^\star[w]\,,\,\delta\rangle] - G(\delta) \\
&= [\langle\!\langle \tilde{z}^\star[w]\,,\,\delta\rangle\!\rangle + \langle \boldsymbol{B}^\star(w)\tilde{z}^\star[w]\,,\,\delta\rangle] - G(\delta) + \langle [b'(w) - \boldsymbol{B}^\star(w)]\,\tilde{z}^\star[w]\,,\,\delta\rangle \\
&\overset{(2.15a)}{=} \langle [b'(w) - b'(u^\star) + \boldsymbol{B}^\star(u^\star) - \boldsymbol{B}^\star(w)]\,\tilde{z}^\star[w]\,,\,\delta\rangle.
\end{aligned}$$

Since Proposition 2.6 yields that $\tilde{z}^\star[w] \in L^\infty(\Omega)$ independently of $w$, we can proceed as in Remark 2.1(i). To this end, we choose $s > 1$ arbitrarily for $d \in \{1,2\}$ and $s = 6$ for $d = 3$. Assumption (GC) then yields that

$$|||\delta|||^2 \lesssim \Big[\|b'(u^\star) - b'(w)\|_{L^{s'}(\Omega)} + \|\boldsymbol{B}^\star(u^\star) - \boldsymbol{B}^\star(w)\|_{L^{s'}(\Omega)}\Big]\,\|\tilde{z}^\star[w]\|_{L^\infty(\Omega)}. \tag{2.44}$$

It remains to prove that

$$\|b'(u^\star) - b'(w)\|_{L^{s'}(\Omega)} \lesssim |||u^\star - w||| \text{ and } \|\boldsymbol{B}^\star(u^\star) - \boldsymbol{B}^\star(w)\|_{L^{s'}(\Omega)} \lesssim |||u^\star - w|||. \tag{2.45}$$

We observe that the change of variables $\tau \mapsto 1 - \tau$ leads to

$$\boldsymbol{B}^\star(w) = \boldsymbol{B}(w,u^\star) = \int_0^1 b'\big(w + (u^\star - w)\,\tau\big)\,\mathrm{d}\tau = \int_0^1 b'\big(u^\star + (w - u^\star)\,\tau\big)\,\mathrm{d}\tau = \boldsymbol{B}(u^\star,w),$$

and, hence,

$$\boldsymbol{B}^\star(u^\star) - \boldsymbol{B}^\star(w) = \int_0^1 \big[b'\big(u^\star\big) - b'\big(u^\star + (w - u^\star)\,\tau\big)\big]\,\mathrm{d}\tau.$$

We only prove the second inequality of (2.45), but note that the first estimate follows for $\tau = 1$ by the subsequent arguments: Due to the smoothness assumption (CAR), we may consider the Taylor expansion of the integrand $b'(u^\star + (w - u^\star)\,\tau)$ for $0 \leq \tau \leq 1$ to see that

$$\begin{aligned}
b'(u^\star + (w - u^\star)\,\tau) = &\sum_{k=1}^{n-1} b^{(k)}(u^\star)\,\frac{(w - u^\star)^{k-1}\,\tau^{k-1}}{(k-1)!} \\
&+ \frac{(w - u^\star)^{n-1}\,\tau^{n-1}}{(n-2)!}\int_0^1 (1-\sigma)^{n-2}\,b^{(n)}\big(u^\star + (w - u^\star)\,\tau\,\sigma\big)\,\mathrm{d}\sigma.
\end{aligned} \tag{2.46}$$

Since $b$ is smooth and $u^\star \in L^\infty(\Omega)$, we obtain that

$$\|b^{(k)}(u^\star)\|_{L^\infty(\Omega)} \leq C \quad \text{for all } k = 2,\dots,n-1,$$

where $C$ depends only on $\|u^\star\|_{L^\infty(\Omega)}$, and $n$. Moreover, (GC) allows us to bound the remainder term, i.e., for any $0 \leq \tau\,\sigma \leq 1$, it holds that

$$\|b^{(n)}\big(u^\star + (w - u^\star)\,\tau\,\sigma\big)\|_{L^\infty(\Omega)} \leq C,$$

where $C$ depends only on $|\Omega|$, $n$, and $R$. If $d \in \{1,2\}$, note that $(n-1)s' < \infty$. If $d = 3$, it holds that $(n-1)s' < 6$. Hence, we obtain for all $2 \le k \le n-1$ that

$$\|(u^\star - w)^{k-1}\|_{L^{s'}(\Omega)} = \|u^\star - w\|_{L^{(k-1)s'}(\Omega)}^{k-1} \lesssim (C_{\mathrm{bnd}} + M)^{n-2} \, |||u^\star - w|||,$$

where the hidden constant depends only on norm equivalence $|||\cdot||| \simeq \|\nabla(\cdot)\|_{L^2(\Omega)}$. Arguing as for (2.40)–(2.42) above, we infer that

$$\|b'(u^\star) - b'(w)\|_{L^{s'}(\Omega)} + \|\boldsymbol{B}^\star(u^\star) - \boldsymbol{B}^\star(w)\|_{L^{s'}(\Omega)} \le C'_{\mathrm{dual}} \, |||u^\star - w|||, \tag{2.47}$$

where $C'_{\mathrm{dual}} = C'_{\mathrm{dual}}(|\Omega|, d, \|u^\star\|_{L^\infty(\Omega)}, M, n, R, p, f, \boldsymbol{f}, g, \boldsymbol{g}, \mu_0) > 0$. This shows the inequalities in (2.45). The estimate (2.44) together with (2.45) yields (2.43), where $C_{\mathrm{dual}} = C_{\mathrm{dual}}(|\Omega|, d, \|u^\star\|_{L^\infty(\Omega)}, M, n, R, p, f, \boldsymbol{f}, g, \boldsymbol{g}, \mu_0) > 0$. This concludes the proof. □

*Proof of Theorem 2.7.* Since Lemma 2.8 guarantees $|||u_H^\star||| \le C_{\mathrm{bnd}}$, we can apply Lemma 2.9 and Lemma 2.10 to $w = u_H^\star$ to obtain that

$$\langle b(u^\star) - b(u_H^\star) \,,\, v\rangle \le C_{\mathrm{Lip}} \, |||u^\star - u_H^\star||| \, |||v||| \quad \text{for all } v \in H^1_0(\Omega) \tag{2.48a}$$

as well as

$$|||\tilde z^\star[u_H^\star] - z^\star[u_H^\star]||| \le C_{\mathrm{dual}} \, |||u^\star - u_H^\star|||. \tag{2.48b}$$

Combining these estimates with the error identity (2.19), we prove the error estimate

$$\begin{aligned} |G(u^\star) - G(u_H^\star)| &= |\langle\!\langle u^\star - u_H^\star \,,\, z^\star[u_H^\star] - z_H^\star[u_H^\star]\rangle\!\rangle + \langle b(u^\star) - b(u_H^\star) \,,\, z^\star[u_H^\star] - z_H^\star[u_H^\star]\rangle \\ &\qquad + \langle\!\langle u^\star - u_H^\star \,,\, \tilde z^\star[u_H^\star] - z^\star[u_H^\star]\rangle\!\rangle + \langle b(u^\star) - b(u_H^\star) \,,\, \tilde z^\star[u_H^\star] - z^\star[u_H^\star]\rangle| \\ &\overset{(2.48)}{\le} C_{\mathrm{est}} \, \big[ |||u^\star - u_H^\star||| \, |||z^\star[u_H^\star] - z_H^\star[u_H^\star]||| + |||u^\star - u_H^\star|||^2 \big] \,, \end{aligned}$$

where $C_{\mathrm{est}} = (1 + C_{\mathrm{Lip}}) \max\{1, C_{\mathrm{dual}}\}$. This concludes the proof. □

The assumptions of Lemma 2.9 (resp. Lemma 2.10) also yield the validity of a Céa-type best approximation property for the discrete primal solution $u_H^\star \in \mathcal{X}_H$ (resp. for the discrete dual solutions $\tilde z_H^\star[w]$, $z_H^\star[w]$ for any $w \in H^1_0(\Omega)$ with $|||w||| \le M < \infty$), even though the PDE operator $\mathcal{A}$ from (2.13) is *not* Lipschitz continuous.

**Proposition 2.11** (Céa lemma for primal problem)**.** *Under the assumptions of Lemma 2.9, it holds that*

$$|||u^\star - u_H^\star||| \le C_{\mathrm{Céa}} \min_{v_H \in \mathcal{X}_H} |||u^\star - v_H|||, \tag{2.49}$$

*where* $C_{\mathrm{Céa}} = C_{\mathrm{Céa}}(|\Omega|, d, n, R, p, f, \boldsymbol{f}, \mu_0)$.

*Proof.* The Galerkin orthogonality reads

$$\langle\!\langle u^\star - u_H^\star \,,\, v_H\rangle\!\rangle + \langle b(u^\star) - b(u_H^\star) \,,\, v_H\rangle = 0, \quad \text{for all } v_H \in \mathcal{X}_H. \tag{2.50}$$

Using (MON) and the Galerkin orthogonality, we observe that

$$\begin{aligned}|||u^\star - u_H^\star|||^2 &\overset{(2.24)}{\le} \langle\!\langle u^\star - u_H^\star\,,\, u^\star - u_H^\star\rangle\!\rangle + \langle b(u^\star) - b(u_H^\star)\,,\, u^\star - u_H^\star\rangle \\ &\overset{(2.50)}{=} \langle\!\langle u^\star - u_H^\star\,,\, u^\star - v_H\rangle\!\rangle + \langle b(u^\star) - b(u_H^\star)\,,\, u^\star - v_H\rangle \\ &\overset{(2.37)}{\le} C_{\text{Céa}}\,|||u^\star - u_H^\star|||\,|||u^\star - v_H|||,\end{aligned}$$

where $C_{\text{Céa}} := 1 + C_{\text{Lip}}$. This proves (2.49), where the minimum is attained due to finite dimension of $\mathcal{X}_H$. □

**Proposition 2.12** (Céa lemma for dual problems)**.** *Let $w \in H^1_0(\Omega)$ with $|||w||| \le M < \infty$. Under the assumptions of Lemma 2.10, it holds that*

$$|||\tilde z^\star[w] - \tilde z_H^\star[w]||| \le C_{\text{Céa}} \min_{v_H \in \mathcal{X}_H} |||\tilde z^\star[w] - v_H|||, \tag{2.51}$$

$$|||z^\star[w] - z_H^\star[w]||| \le C_{\text{Céa}} \min_{v_H \in \mathcal{X}_H} |||z^\star[w] - v_H|||, \tag{2.52}$$

*where $C_{\text{Céa}} = C_{\text{Céa}}(|\Omega|, d, \|u^\star\|_{L^\infty(\Omega)}, M, n, R, p, f, \boldsymbol{f}, \mu_0)$.*

*Proof.* We prove the statement for the practical dual problem. With minor modifications, the same argument also applies for the theoretical dual problem. We only need to show that the bilinear form of the practical dual problem is continuous and elliptic. Then, by standard theory for Lax–Milgram-type problems, this proves the Céa lemma (2.52). To this end, we exploit (MON) and obtain that

$$|||v|||^2 = \langle\!\langle v\,,\, v\rangle\!\rangle \le \langle\!\langle v\,,\, v\rangle\!\rangle + \langle b'(w)v\,,\, v\rangle \quad \text{for all } v \in H^1_0(\Omega),$$

i.e., the bilinear form is elliptic with constant 1. In view of Remark 2.1, choose $t > 1$ arbitrarily for $d \in \{1, 2\}$ and $t = 6$ and, hence, $t'' = 3/2$ for $d = 3$. With (ELL) and (GC), it follows that

$$\langle\!\langle z\,,\, v\rangle\!\rangle + \langle b'(w)z\,,\, v\rangle \le (1 + C\|b'(w)\|_{L^{t''}(\Omega)})|||z|||||v||| \quad \text{for all } v, z \in H^1_0(\Omega).$$

With (2.45) and $|||u^\star - w||| \le C_{\text{bnd}} + M$, we can finally bound

$$\|b'(w)\|_{L^{t''}(\Omega)} \le \|b'(u^\star)\|_{L^{t''}(\Omega)} + C|||u^\star - w||| \le \|b'(u^\star)\|_{L^{t''}(\Omega)} + C(C_{\text{bnd}} + M).$$

Combining the last two estimates, we prove continuity of the bilinear form with $C_{\text{Céa}} = C_{\text{Céa}}(|\Omega|, d, \|u^\star\|_{L^\infty(\Omega)}, M, n, R, p, f, \boldsymbol{f}, \mu_0)$. This concludes the proof. □

**Remark 2.13.** *If it is a priori guaranteed that $\|u_H^\star\|_{L^\infty(\Omega)} \le C < \infty$, then the proofs of Section 2.2.7 simplify considerably and the use of (GC) can be avoided. By Proposition 2.2, we infer*

$$\|u_H^\star - \tau(u^\star - u_H^\star)\|_{L^\infty(\Omega)} < \infty \quad \textit{for all} \quad 0 \le \tau \le 1. \tag{2.53}$$

*To establish Lemma 2.9, recall* $\boldsymbol{B}^\star(u_H^\star)$ *from* (2.16)*. The observation* (2.53) *together with the smoothness assumption* (CAR) *yields that* $\|\boldsymbol{B}^\star(u_H^\star)\|_{L^\infty(\Omega)} < \infty$*. Altogether, we obtain that*

$$\langle b(u^\star) - b(u_H^\star)\,,\, v\rangle \le \|\boldsymbol{B}^\star(u_H^\star)\|_{L^\infty(\Omega)}\|u^\star - u_H^\star\|_{L^2(\Omega)}\|v\|_{L^2(\Omega)}.$$

*Note that* (2.53) *also establishes the crucial estimate* (2.45) *from Lemma 2.10 due to the local Lipschitz continuity from* (CAR)*; see [HPZ15, Proposition 1]. However, we stress that already for lowest-order FEM, the validity of a discrete maximum principle requires assumptions on the triangulation which are not imposed for* (GC) *and usually not met for adaptive mesh refinement.* □

**Remark 2.14.** *Note that* (CAR) *implies only that* $b(x,\cdot)$ *is locally Lipschitz. If we additionally assume global Lipschitz continuity, i.e.,* $L' := \sup_{x\in\Omega}\|b'(x,\cdot)\|_{L^\infty(\mathbb{R})} < \infty$*, then the strongly monotone operator* $\mathcal{A}\colon H^1_0(\Omega) \to H^{-1}(\Omega)$ *from* (2.13) *is also Lipschitz continuous with* $L := \max\{\mu_1, L'\}$*. In particular, the problem* (2.3) *fits into the framework of the main theorem on strongly monotone operators, and the proof of Lemma 2.9 becomes trivial. The same applies to the proof of Lemma 2.10, if* $b'$ *is globally Lipschitz continuous.* □

## 2.3 Goal-oriented adaptive algorithm and main results

### 2.3.1 Mesh refinement

From now on, let $\mathcal{T}_0$ be a given conforming triangulation of $\Omega$. For mesh refinement, we employ newest vertex bisection (NVB); see [Ste08]. For each triangulation $\mathcal{T}_H$ and marked elements $\mathcal{M}_H \subseteq \mathcal{T}_H$, let $\mathcal{T}_h := \texttt{refine}(\mathcal{T}_H, \mathcal{M}_H)$ be the coarsest triangulation where all $T \in \mathcal{M}_H$ have been refined, i.e., $\mathcal{M}_H \subseteq \mathcal{T}_H \backslash \mathcal{T}_h$. We write $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$, if $\mathcal{T}_h$ 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 the finite-dimensional FEM space $\mathcal{X}_H \subseteq H^1_0(\Omega)$ from the introduction, and, since we employ NVB, $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ implies nestedness $\mathcal{X}_H \subseteq \mathcal{X}_h$.

### 2.3.2 *A posteriori* error estimators

For $\mathcal{T}_H \in \mathbb{T}$, $v_H \in \mathcal{X}_H$, and $w \in H^1_0(\Omega)$, let

$$\begin{aligned}\eta_H(T, v_H)^2 := h_T^2\,\|f + \operatorname{div}(\boldsymbol{A}\nabla v_H - \boldsymbol{f}) - b(v_H)\|^2_{L^2(T)} \\ + h_T\,\|[\![(\boldsymbol{A}\nabla v_H - \boldsymbol{f})\cdot\boldsymbol{n}]\!]\|^2_{L^2(\partial T\cap\Omega)},\end{aligned} \tag{2.54}$$

$$\begin{aligned}\zeta_H(w; T, v_H)^2 := h_T^2\,\|g + \operatorname{div}(\boldsymbol{A}\nabla v_H - \boldsymbol{g}) - b'(w)(v_H)\|^2_{L^2(T)} \\ + h_T\,\|[\![(\boldsymbol{A}\nabla v_H - \boldsymbol{g})\cdot\boldsymbol{n}]\!]\|^2_{L^2(\partial T\cap\Omega)}\end{aligned} \tag{2.55}$$

be the local contributions of the standard residual error estimators, where $[\![\cdot]\!]$ denotes the jump across edges (for $d = 2$) resp. faces (for $d = 3$) and $\boldsymbol{n}$ denotes the outer unit normal

vector. For $d = 1$, these jumps vanish, i.e., $[\![ \cdot ]\!] = 0$. For $\mathcal{U}_H \subseteq \mathcal{T}_H$, let

$$\eta_H(\mathcal{U}_H, v_H) := \Big( \sum_{T \in \mathcal{U}_H} \eta_H(T, v_H)^2 \Big)^{1/2} \quad \text{and} \quad \zeta_H(w; \mathcal{U}_H, v_H) := \Big( \sum_{T \in \mathcal{U}_H} \zeta_H(w; T, v_H)^2 \Big)^{1/2}.$$

To abbreviate notation, let $\eta_H(v_H) := \eta_H(\mathcal{T}_H, v_H)$ and $\zeta_H(w; v_H) := \zeta_H(w; \mathcal{T}_H, v_H)$. Furthermore, we write, e.g., $\zeta_H(\mathcal{U}_H, z_H^\star[w]) := \zeta_H(w; \mathcal{U}_H, z_H^\star[w])$, since $w$ is clear from the context.

The next result establishes that the error estimators (2.54)–(2.55) satisfy the following slightly relaxed axioms of adaptivity from [CFPP14]. Compared to [CFPP14], stability (A1) is slightly relaxed and reduction (A2) is simplified due to the nestedness of the discrete spaces. Furthermore, we note that well-posedness of (2.54)–(2.55) requires additional regularity assumptions on $\boldsymbol{A}$, $\boldsymbol{f}$, and $\boldsymbol{g}$ (as stated in Section 2.2.1 and 2.2.3) so that the jump terms are well-defined. The proof is postponed to Section 2.4.1.

**Proposition 2.15.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (CGC). *Let* $\mathcal{T}_H \in \mathbb{T}$ *and* $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$. *Then, there hold the following properties:*

**(A1) stability:** For all $M > 0$, there exists $C_{\mathrm{stab}}[M] > 0$ such that for all $w \in H^1_0(\Omega)$, $v_h \in \mathcal{X}_h$, and $v_H \in \mathcal{X}_H$ with $\max\{|||w|||, |||v_h|||, |||v_H|||\} \le M$, it holds that

$$\begin{aligned} \big|\eta_h(\mathcal{T}_h \cap \mathcal{T}_H, v_h) - \eta_H(\mathcal{T}_h \cap \mathcal{T}_H, v_H)\big| &\le C_{\mathrm{stab}}[M]\, |||v_h - v_H|||, \\ \big|\zeta_h(w; \mathcal{T}_h \cap \mathcal{T}_H, v_h) - \zeta_H(w; \mathcal{T}_h \cap \mathcal{T}_H, v_H)\big| &\le C_{\mathrm{stab}}[M]\, |||v_h - v_H|||. \end{aligned}$$

**(A2) reduction:** With $0 < q_{\mathrm{red}} := 2^{-1/(2d)} < 1$, there holds that, for all $v_H \in \mathcal{X}_H$ and all $w \in H^1_0(\Omega)$,

$$\eta_h(\mathcal{T}_h \backslash \mathcal{T}_H, v_H) \le q_{\mathrm{red}}\, \eta_H(\mathcal{T}_H \backslash \mathcal{T}_h, v_H) \text{ and } \zeta_h(w; \mathcal{T}_h \backslash \mathcal{T}_H, v_H) \le q_{\mathrm{red}}\, \zeta_H(w; \mathcal{T}_H \backslash \mathcal{T}_h, v_H).$$

**(A3) reliability:** For all $w \in H^1_0(\Omega)$, there exists $C_{\mathrm{rel}} > 0$ such that

$$|||u^\star - u_H^\star||| \le C_{\mathrm{rel}}\, \eta_H(u_H^\star) \text{ and } |||z^\star[w] - z_H^\star[w]||| \le C_{\mathrm{rel}}\, \zeta_H(z_H^\star[w]).$$

**(A4) discrete reliability:** For all $w \in H^1_0(\Omega)$, there exists $C_{\mathrm{drel}} > 0$ such that

$$|||u_h^\star - u_H^\star||| \le C_{\mathrm{drel}}\, \eta_H(\mathcal{T}_H \backslash \mathcal{T}_h, u_H^\star) \text{ and } |||z_h^\star[w] - z_H^\star[w]||| \le C_{\mathrm{drel}}\, \zeta_H(\mathcal{T}_H \backslash \mathcal{T}_h, z_H^\star[w]).$$

*The constant* $C_{\mathrm{rel}}$ *depends only on* $d$, $\mu_0$, *and uniform shape regularity of the meshes* $\mathcal{T}_H \in \mathbb{T}$. $C_{\mathrm{drel}}$ *depends additionally on the polynomial degree* $m$, *and* $C_{\mathrm{stab}}[M]$ *depends furthermore on* $|\Omega|$, $M$, $n$, $R$, *and* $\boldsymbol{A}$.

**Remark 2.16.** *As far as the* axioms of adaptivity (A1)–(A4) *are concerned, we stress that only the constant* $C_{\mathrm{stab}}[M]$ *depends on* $M > 0$. *From Lemma 2.8, we know that* $|||v||| \le C_{\mathrm{bnd}}$ *for all* $v \in \{u^\star, u_h^\star, u_H^\star, z^\star[w], z_h^\star[w], z_H^\star[w]\}$. *Hence, for* $w \in \{u^\star, u_h^\star, u_H^\star\}$, $v_h \in \{u_h^\star, z_h^\star[w]\}$, *and* $v_H \in \{u_H^\star, z_H^\star[w]\}$, *also the constant* $C_{\mathrm{stab}} = C_{\mathrm{stab}}[C_{\mathrm{bnd}}]$ *in* (A1) *becomes generic.*

### 2.3.3 Goal-oriented adaptive algorithm

The following algorithm essentially coincides with that of [HPZ15]. Following [BIP21], we adapt the marking strategy to mathematically guarantee optimal convergence rates.

**Algorithm 2.17: Goal-oriented adaptive FEM**

**Input:** Adaptivity parameters $0<\theta\le 1$ and $C_{\mathrm{mark}}\ge 1$, initial mesh $\mathcal{T}_0$.
**Loop:** For all $\ell = 0, 1, 2, \dots$, perform the following steps (i)–(v):

(i) Compute the discrete solutions $u_\ell^\star, z_\ell^\star[u_\ell^\star] \in \mathcal{X}_\ell$ to (2.4) resp. (2.6).

(ii) Compute the refinement indicators $\eta_\ell(T, u_\ell^\star)$ and $\zeta_\ell(T, z_\ell^\star[u_\ell^\star])$ for all $T \in \mathcal{T}_\ell$.

(iii) Determine sets $\overline{\mathcal{M}}_\ell^u, \overline{\mathcal{M}}_\ell^{uz} \subseteq \mathcal{T}_\ell$ of up to the multiplicative constant $C_{\mathrm{mark}}$ minimal cardinality such that

$$\theta\,\eta_\ell(u_\ell^\star)^2 \le \eta_\ell(\overline{\mathcal{M}}_\ell^u, u_\ell^\star)^2, \tag{2.56a}$$

$$\theta\,\big[\,\eta_\ell(u_\ell^\star)^2 + \zeta_\ell(u_\ell^\star; z_\ell^\star[u_\ell^\star])^2\,\big] \le \big[\,\eta_\ell(\overline{\mathcal{M}}_\ell^{uz}, u_\ell^\star)^2 + \zeta_\ell(u_\ell^\star; \overline{\mathcal{M}}_\ell^{uz}, z_\ell^\star[u_\ell^\star])^2\,\big]. \tag{2.56b}$$

(iv) Select $\mathcal{M}_\ell^u \subseteq \overline{\mathcal{M}}_\ell^u$ and $\mathcal{M}_\ell^{uz} \subseteq \overline{\mathcal{M}}_\ell^{uz}$ with $\#\mathcal{M}_\ell^u = \#\mathcal{M}_\ell^{uz} = \min\{\#\overline{\mathcal{M}}_\ell^u, \#\overline{\mathcal{M}}_\ell^{uz}\}$.

(v) Define $\mathcal{M}_\ell := \mathcal{M}_\ell^u \cup \mathcal{M}_\ell^{uz}$ and generate $\mathcal{T}_{\ell+1} := \mathtt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$.

**Output:** Sequence of triangulations $\mathcal{T}_\ell$ with corresponding discrete solutions $u_\ell^\star$ and $z_\ell^\star[u_\ell^\star]$ as well as error estimators $\eta_\ell(u_\ell^\star)$ and $\zeta_\ell(u_\ell^\star; z_\ell^\star[u_\ell^\star])$.

### 2.3.4 Main results

In the following, we give formal statements of our main results on Algorithm 2.17. The proofs are postponed to Section 2.4 below. Our first result states that Algorithm 2.17 indeed relies on reliable *a posteriori* error control for the goal error and guarantees plain convergence.

**Proposition 2.18.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Then, there hold the following statements* (i)–(ii)*:*

(i) *There exists a constant* $C'_{\mathrm{rel}} > 0$ *such that*

$$\big|G(u^\star) - G(u_H^\star)\big| \le C'_{\mathrm{rel}}\,\eta_H(u_H^\star)\,\big[\,\eta_H(u_H^\star)^2 + \zeta_H(z_H^\star[u_H^\star])^2\,\big]^{1/2} \quad \text{for all } \mathcal{T}_H \in \mathbb{T}. \tag{2.57}$$

(ii) *For all* $0 < \theta \le 1$ *and* $1 < C_{\mathrm{mark}} \le \infty$, *Algorithm 2.17 leads to convergence*

$$|G(u^\star) - G(u_\ell^\star)| \le C'_{\mathrm{rel}}\eta_\ell(u_\ell^\star)\,\big[\,\eta_\ell(u_\ell^\star)^2 + \zeta_\ell(z_\ell^\star[u_\ell^\star])^2\,\big]^{1/2} \longrightarrow 0 \quad \text{as } \ell \to \infty \tag{2.58}$$

*where* $C'_{\mathrm{rel}} = C'_{\mathrm{rel}}(|\Omega|, \mathbb{T}, d, n, R, p, f, \boldsymbol{f}, g, \boldsymbol{g}, \mu_0)$.

We stress that (2.57) is an immediate consequence of the goal error estimate (2.33) from

Theorem 2.7 and reliability (A3), i.e.,

$$\begin{aligned} C_{\mathrm{est}}^{-1}\,|G(u^\star)-G(u_H^\star)| &\le \big[|||u^\star-u_H^\star|||\,|||z^\star[u_H^\star]-z_H^\star[u_H^\star]||| + |||u^\star-u_H^\star|||^2\big] \\ &\le C_{\mathrm{rel}}^2\big[\eta_H(u_H^\star)\zeta_H(z_H^\star[u_H^\star]) + \eta_H(u_H^\star)^2\big] \\ &\le \sqrt{2}\,C_{\mathrm{rel}}^2\eta_H(u_H^\star)\big[\zeta_H(z_H^\star[u_H^\star])^2 + \eta_H(u_H^\star)^2\big]^{1/2}. \end{aligned}$$

Consequently, only the convergence (2.58) of Proposition 2.18(ii) has to be proven. Replacing the assumption (GC) on the nonlinearity by the stronger assumption (CGC), we even get linear convergence, which improves Proposition 2.18(ii).

**Theorem 2.19**

*Suppose* (RHS), (ELL), (CAR), (MON), *and* (CGC). *Then, for all* $0<\theta\le 1$ *and* $1\le C_{\mathrm{mark}}\le\infty$, *there exists* $\ell_0\in\mathbb{N}_0$, $C_{\mathrm{lin}}>0$, *and* $0<q_{\mathrm{lin}}<1$ *such that Algorithm 2.17 guarantees that, for all* $\ell, k\in\mathbb{N}_0$ *with* $k\ge\ell\ge\ell_0$,

$$\eta_k(u_k^\star)\,\big[\eta_k(u_k^\star)^2+\zeta_k(z_k^\star[u_k^\star])^2\,\big]^{1/2} \le C_{\mathrm{lin}}q_{\mathrm{lin}}^{k-\ell}\,\eta_\ell(u_\ell^\star)\,\big[\eta_\ell(u_\ell^\star)^2+\zeta_\ell(z_\ell^\star[u_\ell^\star])^2\,\big]^{1/2}. \tag{2.59}$$

*The constants* $C_{\mathrm{lin}}$ *and* $q_{\mathrm{lin}}$ *as well as the index* $\ell_0$ *depend only on* $|\Omega|$, $\mathbb{T}$, $d$, $m$, $\theta$, $n$, $R$, $p$, $f$, $\boldsymbol{f}$, $g$, $\boldsymbol{g}$, $\mu_0$, *and* $\boldsymbol{A}$.

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

$$\begin{aligned} \|u^\star\|_{\mathbb{A}_s} &:= \sup_{N\in\mathbb{N}_0}\Big((N+1)^s\min_{\mathcal{T}_H\in\mathbb{T}_N}\eta_H(u_H^\star)\Big)\in\mathbb{R}_{\ge0}\cup\{\infty\},\\ \|z^\star[u^\star]\|_{\mathbb{A}_t} &:= \sup_{N\in\mathbb{N}_0}\Big((N+1)^t\min_{\mathcal{T}_H\in\mathbb{T}_N}\zeta_H(z_H^\star[u^\star])\Big)\in\mathbb{R}_{\ge0}\cup\{\infty\}. \end{aligned}$$

In explicit terms, e.g., $\|u^\star\|_{\mathbb{A}_s}<\infty$ means that an algebraic convergence rate $O(N^{-s})$ for the error estimator $\eta_\ell$ is possible, if the optimal triangulations are chosen.

In comparison to [HPZ15] or [XHYM21], our proof of Theorem 2.19 avoids any $L^\infty$-bounds on the discrete solutions as well as the assumption that the initial mesh is sufficiently fine. Moreover, in contrast to [XHYM21], which proves linear convergence for the marking strategy suggested in [HPZ15] (and a multilevel correction step), we even prove optimal convergence rates without assuming global Lipschitz continuity for the primal and dual operators.

**Theorem 2.20**

*Suppose* (RHS), (ELL), (CAR), (MON), *and* (CGC). *Let* $s,t>0$ *with* $\|u^\star\|_{\mathbb{A}_s}+\|z^\star[u^\star]\|_{\mathbb{A}_t}<\infty$. *Then, for all* $0<\theta<\theta_{\mathrm{opt}} := (1+C_{\mathrm{stab}}^2C_{\mathrm{drel}}^2)^{-1}$ *and* $1\le C_{\mathrm{mark}}<\infty$, *there holds the following: With the index* $\ell_0\in\mathbb{N}_0$ *from Theorem 2.19, there exists* $C_{\mathrm{opt}}>0$ *such that*

*Algorithm 2.17 guarantees that, for all $\ell \in \mathbb{N}_0$ with $\ell \geq \ell_0$,*

$$\eta_\ell(u_\ell^\star)\big[\,\eta_\ell(u_\ell^\star)^2 + \zeta_\ell(z_\ell^\star[u_\ell^\star])^2\,\big]^{1/2} \leq C_{\mathrm{opt}}\,\|u^\star\|_{\mathbb{A}_s}(\|u^\star\|_{\mathbb{A}_s} + \|z^\star[u^\star]\|_{\mathbb{A}_t})\,(\#\mathcal{T}_\ell - \#\mathcal{T}_0)^{-\alpha}, \tag{2.60}$$

*where $\alpha := \min\{2s, s+t\}$. The constant $C_{\mathrm{opt}}$ depends only on $|\Omega|$, $\mathbb{T}$, $d$, $m$, $C_{\mathrm{nvb}}$, $C_{\mathrm{mark}}$, $\ell_0$, $\theta$, $n$, $R$, $p$, $f$, $\boldsymbol{f}$, $g$, $\boldsymbol{g}$, $\mu_0$, and $\boldsymbol{A}$.*

**Remark 2.21.** *Compared to the treatment of linear problems in [FPZ16], the crucial change in the marking strategy considers the combined error estimator due to the structure from (2.7). This allows us to prove convergence rates in contrast to [HPZ15; XHYM21] by using key ingredients from [BIP21]. As a trade-off, the proofs of the essential quasi-orthogonalities are much more involved both in the semilinear primal setting as well as for the combined error estimator.*

**Remark 2.22.** *With the estimate $\eta_\ell(u_\ell^\star)[\eta_\ell(u_\ell^\star)^2 + \zeta_\ell(z_\ell^\star[u_\ell^\star])^2]^{1/2} \leq \eta_\ell(u_\ell^\star)^2 + \zeta_\ell(z_\ell^\star[u_\ell^\star])^2$, one can also consider Algorithm 2.17 with $\mathcal{M}_\ell := \overline{\mathcal{M}}_\ell^{uz}$, which then takes the form of the standard AFEM algorithm (see, e.g., [CFPP14]) for the product space estimator. Then, Theorem 2.19 and 2.20 hold accordingly with the product replaced by the square sum and $\alpha = \min\{2s, 2t\}$, which is slightly worse than the rate $\alpha$ from Theorem 2.20. We refer to [BIP21] for details (in a different, but structurally similar setting).*

**Remark 2.23.** *The marking strategy proposed in [BET11] uses Dörfler marking*

$$\theta \sum_{T\in\mathcal{T}_\ell} \rho_\ell(T, u_\ell^\star, z_\ell[u_\ell^\star])^2 \leq \sum_{T\in\mathcal{M}_\ell} \rho_\ell(T, u_\ell^\star, z_\ell[u_\ell^\star])^2 \tag{2.61}$$

*for the weighted estimator product*

$$\rho_\ell(T, u_\ell^\star, z_\ell[u_\ell^\star])^2 := \eta_\ell(T, u_\ell^\star)^2\zeta_\ell(z_\ell^\star[u_\ell^\star])^2 + \eta_\ell(u_\ell^\star)^2\zeta_\ell(T, z_\ell^\star[u_\ell^\star])^2.$$

*This combined estimator can be interpreted as a computable (but less local) upper bound for the dual weighted residual estimator. Beyond linear PDEs [FPZ16], however, convergence cannot be guaranteed, since the linearization error of the dual problem will not tend to zero in general. As a possible remedy, [BIP21, Remark 3(iii)] proposes to consider*

$$\rho_\ell(T, u_\ell^\star, z_\ell[u_\ell^\star])^2 := \eta_\ell(T, u_\ell^\star)^2\big[\eta_\ell(u_\ell^\star)^2 + \zeta_\ell(z_\ell^\star[u_\ell^\star])^2\big] + \eta_\ell(u_\ell^\star)^2\big[\eta_\ell(T, u_\ell^\star)^2 + \zeta_\ell(T, z_\ell^\star[u_\ell^\star])^2\big]$$

*for the Dörfler marking (2.61). The present results in combination with the analysis from [FPZ16] show that this strategy implies convergence with optimal rate $\min\{2s, s+t\}$. Details are left to the reader.*

## 2.4 Proofs

In this section, we give the proofs of Proposition 2.15 and 2.18 as well as Theorem 2.19 and 2.20.

### 2.4.1 Axioms of Adaptivity

In this section, we sketch the proof of Proposition 2.15 and verify that the residual error estimators from Section 2.3.2 satisfy the (relaxed) axioms of adaptivity (A1)–(A4) from [CFPP14]. As usual for nonlinear problems, only the verification of stability (A1) requires new ideas, while (A2)–(A4) follow from standard arguments. For a triangulation $\mathcal{T}_h \in \mathbb{T}$ and an element $T \in \mathcal{T}_h$, let $\mathcal{E}(T)$ be the set of its facets (i.e., nodes for $d = 1$, edges for $d = 2$, and faces for $d = 3$, respectively). Moreover, let

$$\Omega_h[T] := \bigcup \{T' \in \mathcal{T}_h \mid T \cap T' \neq \emptyset\} \tag{2.62}$$

denote the usual element patch. Recall that (RHS) ensures that the error estimators (2.54)–(2.55) are well-defined. To abbreviate notation, we define the primal and dual residuals

$$\mathfrak{R}(v_H) := f + \operatorname{div}(\boldsymbol{A}\,\nabla v_H - \boldsymbol{f}) - b(v_H), \tag{2.63a}$$
$$\mathfrak{R}^*(w; v_H) := g + \operatorname{div}(\boldsymbol{A}\nabla v_H - \boldsymbol{g}) - b'(w)v_H \tag{2.63b}$$

for all $v_H \in \mathcal{X}_H$ and $w \in H^1_0(\Omega)$. We stress that we do not explicitly state the dependence of the constants on the $\gamma$-shape regularity constant.

To prove stability (A1), we need the following auxiliary result:

**Lemma 2.24.** *Suppose* (ELL), (CAR), *and* (CGC). *Let* $M > 0$ *and* $v, w \in H^1_0(\Omega)$ *with* $\max\{|||v|||, |||w|||\} \leq M$. *Then, it holds that*

$$\|b(v) - b(w)\|_{L^2(\Omega)} \leq C[M]\,|||v - w||| \tag{2.64}$$

*with* $C[M] = C(|\Omega|, d, M, n, R, \mu_0)$.

*Proof.* Similarly to the Taylor expansion in (2.39), it holds that

$$b(v) = \sum_{k=0}^{n-1} b^{(k)}(w)\,\frac{(v-w)^k}{k!} + \frac{(v-w)^n}{(n-1)!}\int_0^1 (1-\tau)^{n-1}\,b^{(n)}(w + (v-w)\,\tau)\,\mathrm{d}\tau. \tag{2.65}$$

This yields that

$$\|b(v)-b(w)\|_{L^2(\Omega)} \lesssim \Big\| \sum_{k=1}^{n-1} b^{(k)}(w)(v-w)^k + (v-w)^n \int_0^1 (1-\tau)^{n-1}\,b^{(n)}(w+(v-w)\tau)\,\mathrm{d}\tau \Big\|_{L^2(\Omega)}.$$

Recall the generalized Hölder inequality

$$\|\varphi\psi\|_{L^2(\Omega)} \leq \|\varphi\|_{L^{2\rho'}(\Omega)}\,\|\psi\|_{L^{2\rho}(\Omega)}, \quad \text{where} \quad 1/2 = 1/2\,(1/\rho + 1/\rho').$$

Recall that $\|v^k\|_{L^\rho(\Omega)} = \|v\|^k_{L^{k\rho}(\Omega)}$. For any $k = 1, \dots, n-1$ and $1 < \rho < \infty$, it holds that

$$\|b^{(k)}(w)(v-w)^k\|_{L^2(\Omega)} \leq \|b^{(k)}(w)\|_{L^{2\rho'}(\Omega)}\|v-w\|^k_{L^{2k\rho}(\Omega)} \overset{\text{(CGC)}}{\lesssim} (1+\|w\|^{n-k}_{L^{2(n-k)\rho'}(\Omega)})\|v-w\|^k_{L^{2k\rho}(\Omega)}.$$

For $d = 1, 2$, both norms can be estimated by the corresponding energy norm. For $d = 3$, let $\rho = n/k$ and hence $\rho' = n/(n-k) > 1$. Note that $2(n-k)\rho' = 2n \le 6 = 2^*$ as well as $2k\rho = 2n \le 6 = 2^*$ by virtue of (CGC). For the remainder term in (2.65), (CGC) yields that $|b^{(n)}(\widetilde{w})| \lesssim 1$ for all $\widetilde{w} \in H^1_0(\Omega)$ and thus the integral is bounded by a constant. Altogether, this guarantees that

$$\|b(v)-b(w)\|_{L^2(\Omega)} \lesssim \sum_{k=1}^{n-1} \|b^{(k)}(w)(v-w)^k\|_{L^2(\Omega)} + \|(v-w)^n\|_{L^2(\Omega)} \lesssim \sum_{k=1}^{n} (1+|||w|||^{n-k})\, |||v-w|||^k,$$

where the hidden constant depends only on $\Omega$, $d$, $M$, $n$, and $R$ from (CGC), and $\mu_0$ from (ELL). Note that $|||v-w|||^k \lesssim |||v-w|||$, where the hidden constant depends only on $M$. This concludes the proof. □

With Lemma 2.24 at hand, stability (A1) follows as for a linear model problem [CKNS08].

*Proof of stability (A1) for primal problem.* With the primal residual $\mathfrak{R}(v_H)$ from (2.63a), the refinement indicators read

$$\eta_H(T, v_H)^2 = h_T^2 \, \|\mathfrak{R}(v_H)\|^2_{L^2(T)} + h_T \, \|[\![ (\boldsymbol{A}\nabla v_H + \boldsymbol{f}) \cdot \boldsymbol{n} ]\!]\|^2_{L^2(\partial T \cap \Omega)}.$$

Define $\delta_h := v_h - v_H \in \mathcal{X}_h$ and $\mathfrak{D}(\delta_h) := \operatorname{div}(\boldsymbol{A}\nabla\delta_h) + b(v_H) - b(v_h)$. Observe that

$$\mathfrak{R}(v_h) = \big[f + \operatorname{div}(\boldsymbol{A}\nabla v_H - \boldsymbol{f}) - b(v_H)\big] + [\operatorname{div}(\boldsymbol{A}\nabla\delta_h) + b(v_H) - b(v_h)] = \mathfrak{R}(v_H) + \mathfrak{D}(\delta_h).$$

Elementary calculus proves that

$$\begin{aligned} \eta_h(T, v_h) &= \Big( h_T^2 \|\mathfrak{R}(v_H) + \mathfrak{D}(\delta_h)\|^2_{L^2(T)} + h_T \|[\![ (\boldsymbol{A}\nabla(v_H+\delta_h) + \boldsymbol{f}) \cdot \boldsymbol{n} ]\!]\|^2_{L^2(\partial T\cap\Omega)} \Big)^{1/2} \\ &\le \eta_h(T, v_H) + h_T \|\mathfrak{D}(\delta_h)\|_{L^2(T)} + h_T^{1/2} \, \|[\![ \boldsymbol{A}\nabla\delta_h \cdot \boldsymbol{n} ]\!]\|_{L^2(\partial T\cap\Omega)}. \end{aligned} \tag{2.66}$$

Recalling the definition of $\mathfrak{D}(\delta_h)$, we see that

$$\|\mathfrak{D}(\delta_h)\|_{L^2(T)} \le \|\operatorname{div}(\boldsymbol{A}\nabla\delta_h)\|_{L^2(T)} + \|b(v_H) - b(v_h)\|_{L^2(T)}. \tag{2.67}$$

For the first term in (2.67), we use the product rule and an inverse inequality to see that

$$\begin{aligned} \|\operatorname{div}(\boldsymbol{A}\nabla\delta_h)\|_{L^2(T)} &\le \|(\operatorname{div}\boldsymbol{A})\cdot\nabla\delta_h\|_{L^2(T)} + \|\boldsymbol{A} : \mathrm{D}^2\delta_h\|_{L^2(T)} \\ &\lesssim \Big( \|\operatorname{div}\boldsymbol{A}\|_{L^\infty(T)} + h_T^{-1}\|\boldsymbol{A}\|_{L^\infty(T)} \Big) \, \|\nabla\delta_h\|_{L^2(T)}, \end{aligned} \tag{2.68}$$

where $:$ denotes the Frobenius scalar product on $\mathbb{R}^{d\times d}$ and $\mathrm{D}^2\delta_h$ is the Hessian of $\delta_h$. The jump term in (2.66) can be estimated by a discrete trace inequality:

$$\|[\![ \boldsymbol{A}\nabla\delta_h \cdot \boldsymbol{n} ]\!]\|_{L^2(\partial T\cap\Omega)} \lesssim h_T^{-1/2} \, \|\boldsymbol{A}\|_{L^\infty(\Omega_h[T])} \, \|\nabla\delta_h\|_{L^2(\Omega_h[T])}. \tag{2.69}$$

Collecting (2.66)–(2.69), we obtain that

$$\begin{aligned}
&|\eta_h(T,v_h) - \eta_h(T,v_H)| \\
&\quad\lesssim h_T \left[\|\operatorname{div} \boldsymbol{A}\|_{L^\infty(T)} + h_T^{-1}\|\boldsymbol{A}\|_{L^\infty(\Omega_h[T])}\right] \|\nabla\delta_h\|_{L^2(\Omega_h[T])} + h_T\,\|b(v_H) - b(v_h)\|_{L^2(T)} \\
&\quad\lesssim \left[|\Omega|^{1/d} \max_{T'\in\mathcal{T}_0}\|\operatorname{div} \boldsymbol{A}\|_{L^\infty(T')} + \|\boldsymbol{A}\|_{L^\infty(\Omega)}\right]\|\nabla\delta_h\|_{L^2(\Omega_h[T])} + |\Omega|^{1/d}\|b(v_H) - b(v_h)\|_{L^2(T)},
\end{aligned}$$

where the hidden constant depends only on the shape regularity of $\mathcal{T}_h$, and the polynomial degree $m$ of the ansatz spaces. Together with Lemma 2.24, this yields that

$$\begin{aligned}
|\eta_h(\mathcal{T}_h\cap\mathcal{T}_H, v_h) - \eta_h(\mathcal{T}_h\cap\mathcal{T}_H, v_H)| &\le \Big(\sum_{T\in\mathcal{T}_h\cap\mathcal{T}_H} |\eta_h(T,v_h) - \eta_h(T,v_H)|^2\Big)^{1/2} \\
&\lesssim \Big(\sum_{T\in\mathcal{T}_h\cap\mathcal{T}_H} \big(\|\nabla\delta_h\|^2_{L^2(\Omega_h[T])} + \|b(v_H) - b(v_h)\|^2_{L^2(T)}\big)\Big)^{1/2} \\
&\lesssim \big(\|\nabla\delta_h\|^2_{L^2(\Omega)} + \|b(v_H) - b(v_h)\|^2_{L^2(\Omega)}\big)^{1/2} \lesssim |||v_h - v_H|||.
\end{aligned}$$

The hidden constant depends only on $|\Omega|$, the shape regularity of $\mathcal{T}_h$, $d$, $m$, $M$, $n$, $R$, $\mu_0$, and $\boldsymbol{A}$. Note that for any non-refined element $T\in\mathcal{T}_h\cap\mathcal{T}_H$, it holds that $\eta_h(T,v_H) = \eta_H(T,v_H)$. This concludes the proof. □

*Proof of stability* (A1) *for dual problem.* With the dual residual $\mathfrak{R}^*(w;v_H)$ from (2.63b), the refinement indicators read

$$\zeta_H(w;T,v_H)^2 = h_T^2\,\|\mathfrak{R}^*(w;v_H)\|^2_{L^2(T)} + h_T\,\|[\![(\boldsymbol{A}\nabla v_H - \boldsymbol{g})\cdot\boldsymbol{n}]\!]\|^2_{L^2(\partial T\cap\Omega)}.$$

We define $\delta_h := v_h - v_H \in \mathcal{X}_h$ and $\mathfrak{D}^*(\delta_h) := \operatorname{div}(\boldsymbol{A}\nabla\delta_h) - b'(w)\delta_h$. Observe that similar arguments as for the proof of stability (A1) of the primal problem lead to

$$\mathfrak{R}^*(w;v_h) = \mathfrak{R}^*(w;v_H) + \mathfrak{D}^*(\delta_h)$$

and, hence,

$$\zeta_h(w;T,v_h) \le \zeta_h(w;T,v_H) + h_T\|\mathfrak{D}^*(\delta_h)\|_{L^2(T)} + h_T^{1/2}\,\|[\![\boldsymbol{A}\nabla\delta_h\cdot\boldsymbol{n}]\!]\|_{L^2(\partial T\cap\Omega)}.$$

Here, we only estimate the term $\|b'(w)\delta_h\|_{L^2(\Omega)}$, since the other terms follow from the arguments provided for the primal problem. To this end, choose $2<\rho<\infty$ arbitrarily if $d\in\{1,2\}$. If $d=3$, let $\rho=3$ and, hence, $\rho'=3/2$. Assumption (CGC) guarantees that the Sobolev embedding (2.11) holds with $r=2\rho$ and $r=2(n-1)\rho'$ simultaneously. Therefore, we obtain that

$$\begin{aligned}
\|b'(w)\delta_h\|_{L^2(\Omega)} &\le \|b'(w)\|_{L^{2\rho'}(\Omega)}\|\delta_h\|_{L^{2\rho}(\Omega)} \lesssim (1 + \|w^{n-1}\|_{L^{2\rho'}(\Omega)})\|\delta_h\|_{L^{2\rho}(\Omega)} \\
&= (1 + \|w\|^{n-1}_{L^{2(n-1)\rho'}(\Omega)})\|\delta_h\|_{L^{2\rho}(\Omega)} \lesssim (1 + |||w|||^{n-1})|||\delta_h||| \lesssim |||\delta_h|||.
\end{aligned}$$

Arguing as for the primal problem, we see that

$$\begin{aligned}&|\zeta_h(w;\,\mathcal{T}_h\cap\mathcal{T}_H,v_h)-\zeta_h(w;\,\mathcal{T}_h\cap\mathcal{T}_H,v_H)|\\&\qquad\lesssim\Big(\sum_{T\in\mathcal{T}_h\cap\mathcal{T}_H}\|\nabla\delta_h\|^2_{L^2(\Omega_h[T])}+\|b'(w)(v_H-v_h)\|^2_{L^2(\Omega)}\Big)^{1/2}\lesssim|||v_h-v_H|||.\end{aligned}$$

The hidden constant depends only on $|\Omega|$, the shape regularity of $\mathcal{T}_h$, $d$, the polynomial degree $m$ of the ansatz spaces, $M$, $n$, $R$, $\mu_0$, and $\boldsymbol{A}$. This concludes the proof. □

The proof of (A2) follows as for linear PDEs; see, e.g., [CKNS08]. The verification of (A3)–(A4) is based on standard arguments found in, e.g., [Ver13]. Therefore, the proofs of (A2)–(A4) are only sketched in Appendix 2.8.

### 2.4.2 Stability of dual problem

The next result transfers [BIP21, Lemma 6] to the present setting of semilinear PDEs. It shows that the norm difference of dual solutions can be estimated by that of the corresponding primal solutions.

**Lemma 2.25.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Let* $M>0$ *and* $w\in H^1_0(\Omega)$ *with* $|||w|||\le M$. *Then, it holds that*

$$|||z^\star[u^\star]-z^\star[w]|||+|||z^\star_H[u^\star]-z^\star_H[w]|||\le C_{\mathrm{diff}}|||u^\star-w|||, \tag{2.70}$$

*where* $C_{\mathrm{diff}}=C_{\mathrm{diff}}(|\Omega|,d,M,n,R,p,f,\boldsymbol{f},g,\boldsymbol{g},\mu_0)$.

*Proof.* First, note that

$$\begin{aligned}\langle\!\langle z^\star[u^\star]-z^\star[w]\,,\,v\rangle\!\rangle+\langle b'(u^\star)z^\star[u^\star]-b'(w)z^\star[w]\,,\,v\rangle=0\quad&\text{for all } v\in H^1_0(\Omega),\\ \langle\!\langle z^\star_H[u^\star]-z^\star_H[w]\,,\,v_H\rangle\!\rangle+\langle b'(u^\star)z^\star_H[u^\star]-b'(w)z^\star_H[w]\,,\,v_H\rangle=0\quad&\text{for all } v_H\in\mathcal{X}_H.\end{aligned} \tag{2.71}$$

We aim to prove that

$$|||z^\star[u^\star]-z^\star[w]|||\le C_{\mathrm{diff}}|||u^\star-w|||.$$

To this end, note that the strategy in the proof of Proposition 2.10 provides a similar estimate to (2.47) by choosing $t$ from Remark 2.1(ii) instead of $s$ from Remark 2.1(i), i.e.,

$$\|b'(u^\star)-b'(w)\|_{L^{t'}(\Omega)}\le C'_{\mathrm{dual}}|||u^\star-w|||, \tag{2.72}$$

with $C'_{\mathrm{dual}}=C'_{\mathrm{dual}}(|\Omega|,d,\|u^\star\|_{L^\infty(\Omega)},M,n,R,p,f,\boldsymbol{f},g,\boldsymbol{g},\mu_0)>0$. The Hölder inequality

leads us to

$$\begin{aligned}
|||z^\star[u^\star] - z^\star[w]|||^2 &= \langle\!\langle z^\star[u^\star] - z^\star[w]\,,\, z^\star[u^\star] - z^\star[w]\rangle\!\rangle \\
&\overset{(2.71)}{=} -\langle b'(u^\star)z^\star[u^\star] - b'(w)z^\star[w]\,,\, z^\star[u^\star] - z^\star[w]\rangle \\
&= -\langle (b'(u^\star)-b'(w))z^\star[u^\star], z^\star[u^\star]-z^\star[w]\rangle - \langle b'(w)(z^\star[u^\star]-z^\star[w]), z^\star[u^\star]-z^\star[w]\rangle \\
&\overset{(\mathrm{MON})}{\leq} -\langle (b'(u^\star) - b'(w))z^\star[u^\star]\,,\, z^\star[u^\star] - z^\star[w]\rangle \\
&\leq \|b'(u^\star) - b'(w)\|_{L^{t''}(\Omega)}\, \|z^\star[u^\star]\|_{L^t(\Omega)}\, \|z^\star[u^\star] - z^\star[w]\|_{L^{t''}(\Omega)} \\
&\overset{(2.72)}{\lesssim} |||u^\star - w|||\, |||z^\star[u^\star]|||\, |||z^\star[u^\star] - z^\star[w]|||,
\end{aligned} \tag{2.73}$$

where the hidden constant depends only on $C'_{\mathrm{dual}}$ from (2.72) and norm equivalence. Finally, recall that $|||z^\star[u^\star]||| \leq C_{\mathrm{bnd}}$ from Lemma 2.8. The same reasoning applies for $|||z_H^\star[u^\star] - z_H^\star[w]|||$. This concludes the proof. □

### 2.4.3 Proof of Proposition 2.18

The proof of Proposition 2.18 builds on the following lemma, which adapts [BIP21, Proposition 14] to the present setting.

> **Lemma 2.26.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Then, for any choice of the marking parameters* $0 < \theta \leq 1$ *and* $1 \leq C_{\mathrm{mark}} \leq \infty$, *Algorithm 2.17 guarantees that*
>
> - $|||u^\star - u_\ell^\star||| + \eta_\ell(u_\ell^\star) \to 0$ *if* $\#\{k \in \mathbb{N}_0 \mid \mathcal{M}_k$ *satisfies* (2.56a)$\} = \infty$,
> - $|||u^\star - u_\ell^\star||| + \eta_\ell(u_\ell^\star) + |||z^\star[u^\star] - z_\ell^\star[u_\ell^\star]||| + |||z^\star[u^\star] - z_\ell^\star[u^\star]||| + \zeta_\ell(z_\ell^\star[u_\ell^\star]) \to 0$ *if* $\#\{k \in \mathbb{N}_0 \mid \mathcal{M}_k$ *satisfies* (2.56b)$\} = \infty$,
>
> *as* $\ell \to \infty$. *Moreover, at least one of these two cases is met.*

*Sketch of proof.* The proof is essentially verbatim to that of [BIP21, Proposition 14] and therefore only sketched. From the Céa lemma (2.49) for the primal problem (resp. (2.52) for the dual problem), the nestedness $\mathcal{X}_\ell \subseteq \mathcal{X}_{\ell+1}$ of the discrete spaces for all $\ell \in \mathbb{N}_0$, and the stability of the dual problem (Lemma 2.25), it follows that there exist *a priori* limits $u_\infty^\star, z_\infty^\star[u_\infty^\star] \in H_0^1(\Omega)$ such that

$$|||u_\infty^\star - u_\ell^\star||| + |||z_\infty^\star[u_\infty^\star] - z_\ell^\star[u_\ell^\star]||| \xrightarrow{\ell\to\infty} 0.$$

Together with stability (A1) and reduction (A2), the *estimator reduction principle* proves that

$$\begin{aligned}
\eta_\ell(u_\ell^\star) &\xrightarrow{\ell\to\infty} 0 \quad \text{if } \#\{k \in \mathbb{N}_0 \mid \mathcal{M}_k \text{ satisfies (2.56a)}\} = \infty, \\
\eta_\ell(u_\ell^\star) + \zeta_\ell(z_\ell^\star[u_\ell^\star]) &\xrightarrow{\ell\to\infty} 0 \quad \text{if } \#\{k \in \mathbb{N}_0 \mid \mathcal{M}_k \text{ satisfies (2.56b)}\} = \infty.
\end{aligned}$$

Clearly, at least one of these two cases is met. With reliability (A3), it follows that $u^\star = u_\infty^\star$, while $z^\star[u^\star] = z_\infty^\star[u_\infty^\star]$ requires that $\#\{k \in \mathbb{N}_0 \mid \mathcal{M}_k$ satisfies (2.56b)$\} = \infty$; see [BIP21, Proposition 14] for details. □

***Proof of Proposition 2.18.*** The proof is verbatim that of [BIP21, Proposition 1] and therefore only sketched. From (A1)–(A3), the Céa lemma (2.49) for the primal problem (resp. (2.52) for the practical dual problem), and the nestedness of the discrete spaces, there follows boundedness

$$\eta_\ell(u_\ell^\star) + \zeta_\ell(z_\ell^\star[u_\ell^\star]) \lesssim \eta_0(u_0^\star) + \zeta_0(z_0^\star[u_0^\star]) < \infty \quad \text{for all } \ell \in \mathbb{N}_0;$$

see [BIP21, Section 4.1] for details. Together with the convergence results of Lemma 2.26, this yields convergence

$$\eta_\ell(u_\ell^\star)\big[\eta_\ell(u_\ell^\star)^2 + \zeta_\ell(z_\ell^\star[u_\ell^\star])^2\big]^{1/2} \xrightarrow{\ell\to\infty} 0.$$

This concludes the proof. □

### 2.4.4 Auxiliary results

We continue with some preliminary results, which are needed for proving the quasi-orthogonalities and which are, hence, crucial to prove linear convergence. To this end, consider the Fréchet derivative of $\mathcal{A}$ at $w \in H^1_0(\Omega)$, i.e.,

$$\mathcal{A}'[w](\cdot)\colon H^1_0(\Omega) \to H^{-1}(\Omega), \quad \mathcal{A}'[w](z) := \langle\!\langle z\,,\,\cdot\,\rangle\!\rangle + \langle b'(w)z\,,\,\cdot\,\rangle. \tag{2.74}$$

**Lemma 2.27.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Then, there exists a constant* $C = C(|\Omega|, d, \|u^\star\|_{L^\infty(\Omega)}, n, R, p, f, \boldsymbol{f}, \mu_0)$ *such that*

$$\langle \mathcal{A}(u^\star) - \mathcal{A}(u_\ell^\star) - \mathcal{A}'[u^\star](u^\star - u_\ell^\star)\,,\,v\rangle \le C\,|||u^\star - u_\ell^\star|||^2 |||v||| \quad \textit{for all } v \in H^1_0(\Omega). \tag{2.75}$$

*Proof.* Due to the linearity of $\langle\!\langle\,\cdot\,,\,\cdot\,\rangle\!\rangle$ in the left-hand argument, we conclude that the only contribution is due to $b$, i.e., for all $v \in H^1_0(\Omega)$, it holds that

$$\langle \mathcal{A}(u^\star) - \mathcal{A}(u_\ell^\star) - \mathcal{A}'[u^\star](u^\star - u_\ell^\star)\,,\,v\rangle = \langle b(u^\star) - b(u_\ell^\star) - b'(u^\star)(u^\star - u_\ell^\star)\,,\,v\rangle.$$

For $v \in H^1_0(\Omega)$, the Hölder inequality with arbitrary $1 < s < \infty$ if $d \in \{1,2\}$ and $s = 2^*$ if $d = 3$ proves that

$$\langle b(u^\star) - b(u_\ell^\star) - b'(u^\star)(u^\star - u_\ell^\star)\,,\,v\rangle \lesssim \|b(u^\star) - b(u_\ell^\star) - b'(u^\star)(u^\star - u_\ell^\star)\|_{L^{s'}(\Omega)} |||v|||.$$

From the Taylor expansion (2.39), note that

$$\begin{aligned} b(u^\star) - b(u_\ell^\star) - b'(u^\star)(u^\star - u_\ell^\star) = &- \sum_{k=2}^{n-1} b^{(k)}(u^\star)\,\frac{(u_\ell^\star - u^\star)^k}{k!} \\ &- \frac{(u_\ell^\star - u^\star)^n}{(n-1)!} \int_0^1 (1-\tau)^{n-1}\, b^{(n)}\big(u^\star + (u_\ell^\star - u^\star)\,\tau\big)\,\mathrm{d}\tau. \end{aligned}$$

Together with Lemma 2.8 and $u^\star \in L^\infty(\Omega)$, the assumption (GC) yields that

$$\|b(u^\star) - b(u_\ell^\star) - b'(u^\star)(u^\star - u_\ell^\star)\|_{L^{s'}(\Omega)} \lesssim \sum_{k=2}^{n} \|(u^\star - u_\ell^\star)^k\|_{L^{s'}(\Omega)}$$
$$= \sum_{k=2}^{n} \|u^\star - u_\ell^\star\|_{L^{ks'}(\Omega)}^k \lesssim \|\nabla(u^\star - u_\ell^\star)\|_{L^2(\Omega)}^2 \simeq |||u^\star - u_\ell^\star|||^2,$$

where the hidden constants depend only on $C_{\mathrm{bnd}}$ from Lemma 2.8, $n, R$ from (GC) and norm equivalence. This concludes the proof. □

The next lemma is an auxiliary result for establishing quasi-orthogonality. Our proof combines arguments from the linear setting [BHP17, Lemma 17] with ideas from [FFP14, Lemma 6.10]. We stress that the proof exploits the *a priori* convergence $|||u^\star - u_\ell^\star||| \to 0$ from Lemma 2.26.

**Lemma 2.28.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Then, the normalized sequences*

$$e_\ell := \begin{cases} \dfrac{u^\star - u_\ell^\star}{|||u^\star - u_\ell^\star|||}, & \text{for } u^\star \neq u_\ell^\star \\ 0, & \text{otherwise} \end{cases} \quad \text{and} \quad E_\ell := \begin{cases} \dfrac{u_{\ell+1}^\star - u_\ell^\star}{|||u_{\ell+1}^\star - u_\ell^\star|||}, & \text{for } u_{\ell+1}^\star \neq u_\ell^\star \\ 0, & \text{otherwise} \end{cases} \tag{2.76}$$

*converge weakly to* $0$ *in* $H_0^1(\Omega)$.

*Proof.* We only show the statement for $e_\ell$. The proof for $E_\ell$ follows by similar arguments. To prove that $e_\ell \rightharpoonup 0$ in $H_0^1(\Omega)$, we show that each subsequence $(e_{\ell_k})_{k\in\mathbb{N}_0}$ admits a further subsequence $(e_{\ell_{k_j}})_{j\in\mathbb{N}_0}$ such that $e_{\ell_{k_j}} \rightharpoonup 0$ as $j \to \infty$. To this end, consider a subsequence $(e_{\ell_k})_{k\in\mathbb{N}_0}$ of $(e_\ell)_{\ell\in\mathbb{N}_0}$. Without loss of generality, we may assume that $e_{\ell_k} \neq 0$ for all $k \in \mathbb{N}_0$. Note that $|||e_{\ell_k}||| \leq 1$. Hence, the Banach–Alaoglu theorem yields a further subsequence $(e_{\ell_{k_j}})_{j\in\mathbb{N}_0}$ satisfying weak convergence $e_{\ell_{k_j}} \rightharpoonup w_\infty \in H_0^1(\Omega)$ as $j \to \infty$. It remains to show that $w_\infty = 0$. Lemma 2.26 implies that $u^\star \in \mathcal{X}_\infty$ and, hence, $e_\ell \in \mathcal{X}_\infty$ for all $\ell \in \mathbb{N}_0$. Mazur's lemma (see, e.g., [FK80, Theorem 25.2]) yields that $w_\infty \in \mathcal{X}_\infty$.

First, the Galerkin orthogonality shows that

$$\langle \mathcal{A}(u^\star) - \mathcal{A}(u_{\ell_{k_j}}^\star), v_i \rangle = 0 \quad \text{for all } i \leq \ell_{k_j} \text{ and } v_i \in \mathcal{X}_i.$$

Letting $j \to \infty$, we infer that

$$\lim_{j\to\infty} \frac{\langle \mathcal{A}(u^\star) - \mathcal{A}(u_{\ell_{k_j}}^\star), v_i \rangle}{|||u^\star - u_{\ell_{k_j}}^\star|||} = 0 \quad \text{for all } i \in \mathbb{N}_0 \text{ and } v_i \in \mathcal{X}_i.$$

Let $v_\infty \in \mathcal{X}_\infty$. By definition of $\mathcal{X}_\infty$, there exists a sequence $(v_i)_{i\in\mathbb{N}_0}$ with $v_i \in \mathcal{X}_i$ and $|||v_\infty - v_i||| \to 0$ as $i \to \infty$. Given $\varepsilon > 0$, there exists $i_0 \in \mathbb{N}_0$ such that $|||v_i - v_\infty||| \leq \varepsilon$ for all $i_0 \leq i \in \mathbb{N}_0$.

Estimate (2.37) yields that

$$\lim_{j\to\infty} \frac{\langle \mathcal{A}(u^\star) - \mathcal{A}(u^\star_{\ell_{k_j}}) \,,\, v_i - v_\infty\rangle}{|||u^\star - u^\star_{\ell_{k_j}}|||} \lesssim |||v_i - v_\infty||| \le \varepsilon,$$

where the hidden constant depends only on $C_{\mathrm{Lip}}$. Hence, we get that

$$\lim_{j\to\infty} \frac{\langle \mathcal{A}(u^\star) - \mathcal{A}(u^\star_{\ell_{k_j}}) \,,\, v_\infty\rangle}{|||u^\star - u^\star_{\ell_{k_j}}|||} = 0 \quad \text{for all } v_\infty \in \mathcal{X}_\infty.$$

Moreover, Lemma 2.27 and the triangle inequality lead to

$$\frac{|\langle \mathcal{A}'[u^\star](u^\star - u^\star_{\ell_{k_j}}) \,,\, v_\infty\rangle|}{|||u^\star - u^\star_{\ell_{k_j}}|||} \le \frac{|\langle \mathcal{A}(u^\star) - \mathcal{A}(u^\star_{\ell_{k_j}}) \,,\, v_\infty\rangle|}{|||u^\star - u^\star_{\ell_{k_j}}|||} + C\,|||u^\star - u^\star_{\ell_{k_j}}|||\,|||v_\infty|||.$$

Together with *a priori* convergence $|||u^\star - u^\star_{\ell_{k_j}}||| \to 0$, we thus obtain that

$$\lim_{j\to\infty} \frac{|\langle \mathcal{A}'[u^\star](u^\star - u^\star_{\ell_{k_j}}) \,,\, v_\infty\rangle|}{|||u^\star - u^\star_{\ell_{k_j}}|||} = 0. \tag{2.77}$$

Note that due to (ELL) and (MON), $\mathcal{A}'[u^\star](\,\cdot\,)$ is bounded from below, i.e.,

$$|||v|||^2 \le \langle \mathcal{A}'[u^\star](v) \,,\, v\rangle \lesssim \|\mathcal{A}'[u^\star](v)\|_{H^{-1}(\Omega)}\,|||v||| \quad \text{for all } v \in H^1_0(\Omega).$$

Due to the smoothness of $\xi \mapsto b(t,\xi)$ and the $L^\infty$-bound for $u^\star$ from Proposition 2.2, we infer that $0 \le b'(u^\star) \le C$. Hence, $\mathcal{A}'[u^\star](\,\cdot\,)$ is a bounded linear operator and the restriction $\mathcal{A}'[u^\star](\,\cdot\,)|_{\mathcal{X}_\infty} \colon \mathcal{X}_\infty \to \mathcal{X}_\infty^*$ is an isomorphism. Consequently, also the adjoint $(\mathcal{A}'[u^\star]|_{\mathcal{X}_\infty})^* \colon \mathcal{X}_\infty^* \to \mathcal{X}_\infty$ is an isomorphism, where we note that $\mathcal{X}_\infty$ is a closed subspace of the Hilbert space $H^1_0(\Omega)$ and, hence, reflexive. Hence, for every $\widetilde{v}_\infty \in \mathcal{X}_\infty$, there exists $v_\infty \in \mathcal{X}_\infty$ such that

$$\begin{aligned} 0 = \lim_{j\to\infty} \frac{|\langle \mathcal{A}'[u^\star](u^\star - u^\star_{\ell_{k_j}}) \,,\, v_\infty\rangle|}{|||u^\star - u^\star_{\ell_{k_j}}|||} &= \lim_{j\to\infty} \frac{|\langle \mathcal{A}'[u^\star]^*(v_\infty) \,,\, u^\star - u^\star_{\ell_{k_j}}\rangle|}{|||u^\star - u^\star_{\ell_{k_j}}|||} \\ &= \lim_{j\to\infty} \frac{|\langle\!\langle u^\star - u^\star_{\ell_{k_j}} \,,\, \widetilde{v}_\infty\rangle\!\rangle|}{|||u^\star - u^\star_{\ell_{k_j}}|||} = \lim_{j\to\infty} \langle\!\langle e_{\ell_{k_j}} \,,\, \widetilde{v}_\infty\rangle\!\rangle. \end{aligned}$$

This shows that $w_\infty = 0$ and concludes the proof. □

### 2.4.5 Quasi-orthogonalities

Our proof of the crucial quasi-orthogonalities adapts that of [BHP17, Lemma 17, 18] from the linear setting in the Lax–Milgram framework to the present nonlinear setting. However,

we stress that the following results all need the stronger growth condition (CGC), while our earlier results only require (GC).

**Lemma 2.29** (quasi-orthogonality for primal problem)**.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (CGC). *Then, for all* $0 < \varepsilon < 1$, *there exists* $\ell_0 \in \mathbb{N}$ *such that for all* $\ell \geq \ell_0$ *and all* $k \in \mathbb{N}_0$, *it holds that*

$$|||u^\star - u^\star_{\ell+k}|||^2 + |||u^\star_{\ell+k} - u^\star_\ell|||^2 \leq \frac{1}{1-\varepsilon}\,|||u^\star - u^\star_\ell|||^2. \tag{2.78}$$

*Proof.* Together with the Rellich–Kondrachov compactness theorem (see, e.g., [KJF77, Theorem 5.8.2]), Lemma 2.28 yields strong convergence

$$\|e_\ell\|_{L^\sigma(\Omega)}, \|E_\ell\|_{L^\sigma(\Omega)} \xrightarrow{\ell\to\infty} 0 \quad \text{where} \quad \begin{cases} \sigma \in [1,\infty), & \text{if } d \in \{1,2\},\\ \sigma \in [1,2^*), & \text{if } d = 3. \end{cases}$$

If $d = 1, 2$, let $1 < \sigma < \infty$ be arbitrary with Hölder conjugate $\sigma' > 1$. If $d = 3$, let $\sigma = 5 = 2^* - 1$ and hence $\sigma' = 5/4 = (2^* - 1)/(2^* - 2)$. Note that (CGC) yields that $n\sigma' \leq \sigma$ and hence $\|e_\ell\|_{L^{n\sigma'}(\Omega)} \lesssim \|e_\ell\|_{L^\sigma(\Omega)} \to 0$ as $\ell \to \infty$. We argue as for (2.40): By the Taylor expansion, $n\sigma' < 2^*$, and with Lemma 2.8, we obtain that

$$\begin{aligned} \|b(u^\star) - b(u^\star_\ell)\|_{L^{\sigma'}(\Omega)} &\lesssim \sum_{j=1}^{n} \|u^\star - u^\star_\ell\|^j_{L^{j\sigma'}(\Omega)} \lesssim \|u^\star - u^\star_\ell\|_{L^{n\sigma'}(\Omega)} \\ &= |||u^\star - u^\star_\ell|||\,\|e_\ell\|_{L^{n\sigma'}(\Omega)} \lesssim |||u^\star - u^\star_\ell|||\,\|e_\ell\|_{L^\sigma(\Omega)}. \end{aligned} \tag{2.79}$$

Furthermore, for $k, \ell \in \mathbb{N}$, recall the Galerkin orthogonality

$$\langle\!\langle u^\star - u^\star_{\ell+k}\,,\, v_{\ell+k}\rangle\!\rangle + \langle b(u^\star) - b(u^\star_{\ell+k})\,,\, v_{\ell+k}\rangle = 0 \quad \text{for all } v_{\ell+k} \in \mathcal{X}_{\ell+k}. \tag{2.80}$$

Due to the bilinearity and symmetry of $\langle\!\langle \cdot\,,\,\cdot \rangle\!\rangle$, we have that

$$|||u^\star - u^\star_\ell|||^2 = |||u^\star - u^\star_{\ell+k}|||^2 + |||u^\star_{\ell+k} - u^\star_\ell|||^2 + 2\langle\!\langle u^\star - u^\star_{\ell+k}\,,\, u^\star_{\ell+k} - u^\star_\ell\rangle\!\rangle. \tag{2.81}$$

Let $0 < \varepsilon < 1$. Note that $u^\star_{\ell+k} - u^\star_\ell \in \mathcal{X}_{\ell+k}$ due to nestedness of the discrete spaces. Exploiting the Galerkin orthogonality (2.80) and the Young inequality, we thus obtain that there exists $\ell_0$ such that, for all $\ell \geq \ell_0$ and all $k \geq 0$,

$$\begin{aligned} 2\langle\!\langle u^\star - u^\star_{\ell+k}\,,\, u^\star_{\ell+k} - u^\star_\ell\rangle\!\rangle &\overset{(2.80)}{\geq} -2|\langle b(u^\star) - b(u^\star_{\ell+k})\,,\, u^\star_{\ell+k} - u^\star_\ell\rangle| \\ &\geq -2\|b(u^\star) - b(u^\star_{\ell+k})\|_{L^{\sigma'}(\Omega)}\|u^\star_{\ell+k} - u^\star_\ell\|_{L^\sigma(\Omega)} \\ &\overset{(2.79)}{\geq} -2\varepsilon|||u^\star - u^\star_{\ell+k}|||\,|||u^\star_{\ell+k} - u^\star_\ell||| \\ &\gtrsim -\varepsilon[|||u^\star - u^\star_{\ell+k}|||^2 + |||u^\star_{\ell+k} - u^\star_\ell|||^2] \end{aligned}$$

The combination with (2.81) proves that

$$\frac{1}{1-\varepsilon}|||u^\star - u^\star_\ell|||^2 \geq |||u^\star - u^\star_{\ell+k}|||^2 + |||u^\star_{\ell+k} - u^\star_\ell|||^2.$$

This concludes the proof. □

While Lemma 2.26 guarantees *a priori* convergence $|||u^\star - u_\ell^\star||| \to 0$ of the primal problem, *a priori* convergence of the dual problem has to be assumed (and depends on the marking steps).

**Lemma 2.30** (quasi-orthogonality for *exact* practical dual problem)**.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (CGC). *Suppose that* $|||z^\star[u^\star] - z_\ell^\star[u^\star]||| \to 0$ *as* $\ell \to \infty$. *Then, for all* $0 < \varepsilon < 1$, *there exists* $\ell_0 \in \mathbb{N}$ *such that for all* $\ell \geq \ell_0$ *and all* $k \in \mathbb{N}_0$, *it holds that*

$$|||z^\star[u^\star] - z_{\ell+k}^\star[u^\star]|||^2 + |||z_{\ell+k}^\star[u^\star] - z_\ell^\star[u^\star]|||^2 \leq \frac{1}{1-\varepsilon}\,|||z^\star[u^\star] - z_\ell^\star[u^\star]|||^2. \tag{2.82}$$

*Proof.* Note that the dual problem reads

$$a(z^\star[u^\star], v) + \langle \mathcal{K}(z^\star[u^\star])\,,\, v\rangle = G(v) \quad \text{for all } v \in H^1_0(\Omega),$$

where $\mathcal{K}(w) := b'(u^\star)w \in L^2(\Omega)$ defines a compact operator $\mathcal{K}\colon H^1_0(\Omega) \to H^{-1}(\Omega)$. The claim thus follows from [BHP17, Lemma 17, 18]. □

**Lemma 2.31** (combined quasi-orthogonality for *inexact* practical dual problem)**.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (CGC). *Suppose that* $|||z^\star[u^\star] - z_\ell^\star[u_\ell^\star]||| \to 0$ *as* $\ell \to \infty$. *Then, for all* $0 < \delta < 1$, *there exists* $\ell_0 \in \mathbb{N}$ *such that for all* $\ell \geq \ell_0$ *and all* $k \in \mathbb{N}_0$, *it holds that*

$$\begin{aligned}\big[\,|||u^\star - u_{\ell+k}^\star|||^2 + |||z^\star[u^\star] - z_{\ell+k}^\star[u_{\ell+k}^\star]|||^2\,\big] + \big[\,|||u_{\ell+k}^\star - u_\ell^\star|||^2 + |||z_{\ell+k}^\star[u_{\ell+k}^\star] - z_\ell^\star[u_\ell^\star]|||^2\,\big]\\ \leq \frac{1}{1-\delta}\,\big[\,|||u^\star - u_\ell^\star|||^2 + |||z^\star[u^\star] - z_\ell^\star[u_\ell^\star]|||^2\,\big].\end{aligned} \tag{2.83}$$

*Proof.* According to Lemma 2.25, it holds that

$$\begin{aligned}|||z^\star[u^\star] - z_\ell^\star[u^\star]||| &\leq |||z^\star[u^\star] - z_\ell^\star[u_\ell^\star]||| + |||z_\ell^\star[u^\star] - z_\ell^\star[u_\ell^\star]|||\\ &\overset{(2.70)}{\lesssim} |||z^\star[u^\star] - z_\ell^\star[u_\ell^\star]||| + |||u^\star - u_\ell^\star||| \xrightarrow{\ell\to\infty} 0.\end{aligned}$$

Hence, we may exploit the conclusions of Lemma 2.29 and Lemma 2.30. For arbitrary $\alpha > 0$, the Young inequality guarantees that

$$\begin{aligned}|||z^\star[u^\star] - z_{\ell+k}^\star[u_{\ell+k}^\star]|||^2 &\leq (1+\alpha)\,|||z^\star[u^\star] - z_{\ell+k}^\star[u^\star]|||^2 + (1+\alpha^{-1})\,|||z_{\ell+k}^\star[u^\star] - z_{\ell+k}^\star[u_{\ell+k}^\star]|||^2,\\ |||z_{\ell+k}^\star[u_{\ell+k}^\star] - z_\ell^\star[u_\ell^\star]|||^2 &\leq (1+\alpha)\,|||z_{\ell+k}^\star[u^\star] - z_\ell^\star[u^\star]|||^2 + (1+\alpha^{-1})^2\,|||z_\ell^\star[u^\star] - z_\ell^\star[u_\ell^\star]|||^2\\ &\quad + (1+\alpha)(1+\alpha^{-1})\,|||z_{\ell+k}^\star[u^\star] - z_{\ell+k}^\star[u_{\ell+k}^\star]|||^2,\\ |||z^\star[u^\star] - z_\ell^\star[u^\star]|||^2 &\leq (1+\alpha)\,|||z^\star[u^\star] - z_\ell^\star[u_\ell^\star]|||^2 + (1+\alpha^{-1})\,|||z_\ell^\star[u^\star] - z_\ell^\star[u_\ell^\star]|||^2.\end{aligned}$$

Together with Lemma 2.30, this leads to

$$\begin{aligned}
&|||z^\star[u^\star] - z^\star_{\ell+k}[u^\star_{\ell+k}]|||^2 + |||z^\star_{\ell+k}[u^\star_{\ell+k}] - z^\star_\ell[u^\star_\ell]|||^2 \\
&\quad\le (1+\alpha)\,\big[\,|||z^\star[u^\star] - z^\star_{\ell+k}[u^\star]|||^2 + |||z^\star_{\ell+k}[u^\star] - z^\star_\ell[u^\star]|||^2\,\big] \\
&\qquad + (2+\alpha)(1+\alpha^{-1})\,|||z^\star_{\ell+k}[u^\star] - z^\star_{\ell+k}[u^\star_{\ell+k}]|||^2 + (1+\alpha^{-1})^2\,|||z^\star_\ell[u^\star] - z^\star_\ell[u^\star_\ell]|||^2 \\
&\quad\overset{(2.82)}{\le} \frac{1+\alpha}{1-\varepsilon}\,|||z^\star[u^\star] - z^\star_\ell[u^\star]|||^2 + (1+\alpha^{-1})^2\,|||z^\star_\ell[u^\star] - z^\star_\ell[u^\star_\ell]|||^2 \\
&\qquad + (2+\alpha)(1+\alpha^{-1})\,|||z^\star_{\ell+k}[u^\star] - z^\star_{\ell+k}[u^\star_{\ell+k}]|||^2 \\
&\quad\le \frac{(1+\alpha)^2}{1-\varepsilon}\,|||z^\star[u^\star] - z^\star_\ell[u^\star_\ell]|||^2 + \Big[(1+\alpha^{-1})^2 + \frac{(1+\alpha^{-1})(1+\alpha)}{1-\varepsilon}\Big]\,|||z^\star_\ell[u^\star] - z^\star_\ell[u^\star_\ell]|||^2 \\
&\qquad + (2+\alpha)(1+\alpha^{-1})\,|||z^\star_{\ell+k}[u^\star] - z^\star_{\ell+k}[u^\star_{\ell+k}]|||^2
\end{aligned} \tag{2.84}$$

for all $0<\varepsilon<1$ and all $\ell\ge\ell_0$, where $\ell_0\in\mathbb{N}_0$ depends only on $\varepsilon$. If $d\in\{1,2\}$, let $1<t<\infty$ be arbitrary. If $d=3$, let $t=2^*$ and, hence, $t''=3/2$; cf. Remark 2.1. We argue as for (2.40): By the Taylor expansion, $\sigma := (n-1)t'' < 2^*$, and with Lemma 2.8, we obtain that

$$\|b'(u^\star)-b'(u^\star_\ell)\|_{L^{t''}(\Omega)} \lesssim \sum_{j=1}^{n-1} \|u^\star-u^\star_\ell\|^j_{L^{jt''}(\Omega)} \lesssim \|u^\star-u^\star_\ell\|_{L^{(n-1)t''}(\Omega)} \lesssim |||u^\star-u^\star_\ell|||\,\|e_\ell\|_{L^\sigma(\Omega)}, \tag{2.85}$$

where $\|e_\ell\|_{L^\sigma(\Omega)}\to 0$ as $\ell\to\infty$. Recall that the inequality (2.73) in the proof of Lemma 2.25 does not rely on any $L^\infty(\Omega)$-bounds; hence, we may exploit the discrete analogue of (2.73) in combination with the Hölder inequality to obtain that

$$\begin{aligned}
|||z^\star_\ell[u^\star] - z^\star_\ell[u^\star_\ell]|||^2 &\overset{(2.73)}{\le} -\langle [b'(u^\star)-b'(u^\star_\ell)]\,z^\star_\ell[u^\star]\,,\, z^\star_\ell[u^\star]-z^\star_\ell[u^\star_\ell]\rangle \\
&\lesssim \|b'(u^\star)-b'(u^\star_\ell)\|_{L^{t''}(\Omega)}\,\|z^\star_\ell[u^\star]\|_{L^t(\Omega)}\,\|z^\star_\ell[u^\star]-z^\star_\ell[u^\star_\ell]\|_{L^t(\Omega)} \\
&\overset{(2.70)}{\lesssim} |||z^\star_\ell[u^\star]|||\,\|b'(u^\star)-b'(u^\star_\ell)\|_{L^{t''}(\Omega)}\,|||u^\star-u^\star_\ell||| \overset{(2.85)}{\lesssim} |||z^\star_\ell[u^\star]|||\,\|e_\ell\|_{L^\sigma(\Omega)}\,|||u^\star-u^\star_\ell|||^2.
\end{aligned}$$

Since $|||z^\star_\ell[z^\star]|||\le C_{\mathrm{bnd}}$ due to Lemma 2.8, this proves that

$$|||z^\star_\ell[u^\star] - z^\star_\ell[u^\star_\ell]|||^2 \;\le\; \kappa_\ell\,|||u^\star-u^\star_\ell|||^2 \quad\text{with}\quad 0\le\kappa_\ell\xrightarrow{\ell\to\infty}0. \tag{2.86}$$

Plugging (2.86) into (2.84), we thus have shown that

$$\begin{aligned}
&|||z^\star[u^\star] - z^\star_{\ell+k}[u^\star_{\ell+k}]|||^2 + |||z^\star_{\ell+k}[u^\star_{\ell+k}] - z^\star_\ell[u^\star_\ell]|||^2 \\
&\quad\le \frac{(1+\alpha)^2}{1-\varepsilon}\,|||z^\star[u^\star] - z^\star_\ell[u^\star_\ell]|||^2 + \Big[(1+\alpha^{-1})^2 + \frac{(1+\alpha^{-1})(1+\alpha)}{1-\varepsilon}\Big]\,\kappa_\ell\,|||u^\star-u^\star_\ell|||^2 \\
&\qquad + (2+\alpha)(1+\alpha^{-1})\,\kappa_{\ell+k}\,|||u^\star-u^\star_{\ell+k}|||^2
\end{aligned}$$

for all $0<\varepsilon<1$, all $\alpha>0$, and all $\ell\ge\ell_0$, where $\ell_0\in\mathbb{N}_0$ depends only on $\varepsilon$. We combine

this estimate with that of Lemma 2.29. This leads to

$$\begin{aligned}&\big[\,|||u^\star - u^\star_{\ell+k}|||^2 + |||z^\star[u^\star] - z^\star_{\ell+k}[u^\star_{\ell+k}]|||^2\,\big] + \big[\,|||u^\star_{\ell+k} - u^\star_\ell|||^2 + |||z^\star_{\ell+n}[u^\star_{\ell+k}] - z^\star_\ell[u^\star_\ell]|||^2\,\big]\\ &\quad\le C(\alpha,\varepsilon,\ell)\,\big[\,|||u^\star - u^\star_\ell|||^2 + |||z^\star[u^\star] - z^\star_\ell[u^\star_\ell]|||^2\,\big] + (2+\alpha)(1+\alpha^{-1})\,\kappa_{\ell+k}\,|||u^\star - u^\star_{\ell+k}|||^2,\end{aligned}$$

where, since $1/(1-\varepsilon) \le (1+\alpha)^2/(1-\varepsilon)$,

$$C(\alpha,\varepsilon,\ell) := \max\left\{\frac{(1+\alpha)^2}{1-\varepsilon},\,\left[(1+\alpha^{-1})^2 + \frac{(1+\alpha^{-1})(1+\alpha)}{1-\varepsilon}\right]\kappa_\ell\right\}$$

for all $0<\varepsilon<1$, all $\alpha>0$, and all $\ell \ge \ell_0$, where $\ell_0 \in \mathbb{N}_0$ depends only on $\varepsilon$. For arbitrary $0<\alpha,\beta,\varepsilon<1$, there exists $\ell_0' \in \mathbb{N}_0$ such that for all $\ell \ge \ell_0'$, it holds that

$$(2+\alpha)(1+\alpha^{-1})\kappa_{\ell+k} \le \beta$$

as well as

$$\left[(1+\alpha^{-1})^2 + \frac{(1+\alpha^{-1})(1+\alpha)}{1-\varepsilon}\right]\kappa_\ell \le \frac{(1+\alpha)^2}{1-\varepsilon}.$$

Hence, we are led to

$$\begin{aligned}&\big[\,|||u^\star - u^\star_{\ell+k}|||^2 + |||z^\star[u^\star] - z^\star_{\ell+k}[u^\star_{\ell+k}]|||^2\,\big] + \big[\,|||u^\star_{\ell+k} - u^\star_\ell|||^2 + |||z^\star_{\ell+k}[u^\star_{\ell+k}] - z^\star_\ell[u^\star_\ell]|||^2\,\big]\\ &\quad\le \frac{(1+\alpha)^2}{(1-\varepsilon)(1-\beta)}\,\big[\,|||u^\star - u^\star_\ell|||^2 + |||z^\star[u^\star] - z^\star_\ell[u^\star_\ell]|||^2\,\big].\end{aligned} \tag{2.87}$$

Given $0<\delta<1$, we first fix $\alpha>0$ such that $(1+\alpha)^2 < \frac{1}{1-\delta}$. Then, we choose $0<\varepsilon,\beta<1$ such that $\frac{(1+\alpha)^2}{(1-\varepsilon)(1-\beta)} \le \frac{1}{1-\delta}$. The choices of $\varepsilon$ and $\beta$ also provide some index $\ell_0 \in \mathbb{N}_0$ such that estimate (2.87) holds for all $\ell \ge \ell_0$. This concludes the proof. □

**Remark 2.32.** *For $d = 3$, assumption (CGC) requires $n \in \{2,3\}$. We note that, while well-posedness of the residual error estimator relies on this assumption, the quasi-orthogonalities (2.78) and (2.83) only require $n \in \{2,3,4\}$ for $d = 3$.*

**Remark 2.33.** *If $d > 3$, the same reasoning using the Hölder inequality still holds true, though the polynomial degree $n$ in (CGC) becomes more constrained.*

### 2.4.6 Proof of Theorem 2.19 and Theorem 2.20

It is a key observation in the analysis of [BIP21] that it suffices to prove

- stability of the (practical) dual problem (see Lemma 2.25 resp. [BIP21, Lemma 6]),
- quasi-orthogonality of the primal problem (see Lemma 2.29 resp. [BIP21, Lemma 11]),
- combined quasi-orthogonality for the practical dual problem (see Lemma 2.31 resp. [BIP21, Lemma 13]).

Then, the estimator axioms (A1)–(A4) already prove linear convergence (2.59) in the sense of Theorem 2.19 (see [BIP21, Theorem 2(i)] and [BIP21, Section 6.1]) with optimal convergence rates (2.60) in the sense of Theorem 2.20 (see [BIP21, Theorem 2(ii)] and [BIP21, Section 6.2]).

## 2.5 Numerical experiments

In this section, we test and illustrate Algorithm 2.17 with numerical experiments for $d = 1$ and $d = 2$. We consider equation (2.2), where $\boldsymbol{A} = \left(\begin{smallmatrix} 1 & 0 \\ 0 & 1 \end{smallmatrix}\right)$. The adaptivity parameter is set to $\theta = 0.5$. We compare the proposed GOAFEM (Algorithm 2.17) with standard AFEM (adapted from, e.g., [CFPP14; CKNS08]), where mesh-refinement is driven by the primal estimator (i.e., Algorithm 2.17 with $\mathcal{M}_\ell := \overline{\mathcal{M}}_\ell^u$ in step (v)) and standard AFEM driven by the product space estimator (see Remark 2.22).

**Example 2.34** (boundary value problem in 1D)**.** *For $d = 1$ and $\Omega = (0,1)$, consider*

$$-(u^\star)'' + \arctan(u^\star) = f \quad \text{in } \Omega \quad \text{subject to} \quad u^\star(0) = u^\star(1) = 0, \tag{2.88}$$

*with semilinearity $b(v) = \arctan(v)$ and hence $b'(v) = 1/(1+v^2)$. We set $\boldsymbol{f} = 0$ and choose $f$ in such a way that*

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

*The implementation of conforming finite elements of order $m \in \{1,2,3,4\}$ is done using Legendre polynomials and Gauss–Legendre quadrature and Gauss–Jacobi quadrature for the interval containing the left interval endpoint. For mesh refinement, 1D bisection is used. Moreover, we employ the (damped) Newton method from [AW15, Section 3] for step* (i) *in Algorithm 2.17 to approximate the nonlinear primal problem. Let $g = x^{-9/20} \in L^2(\Omega)$ and $\boldsymbol{g} = 0$ serve as the goal functions. As a reference, we use the value of the integral which reads*

$$G(u^\star) = \int_0^1 \frac{\sin(\pi x)}{x^{9/20}} \,\mathrm{d}x \approx 0.95925303932778833\ldots. \tag{2.89}$$

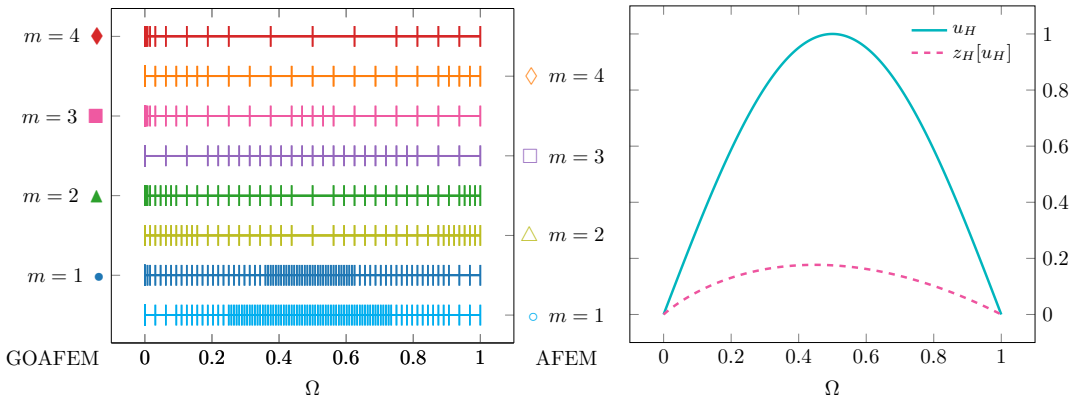
*The uniform initial mesh is given by $\mathcal{T}_0 = \{[\frac{k-1}{2}, \frac{k}{2}] \mid k = 1,2\}$. Figure 2.1 shows meshes of GOAFEM and AFEM for $m \in \{1,2,3,4\}$ as well as discrete solutions $u_H^\star$ and $z_H^\star[u_H^\star]$.*

*The numerical convergence results are depicted in Figure 2.2. We observe that the estimator as well as the goal error achieve the expected rate $\#\mathcal{T}_H^{-2m}$ if computed with Algorithm 2.17. In contrast, standard AFEM leads to a slower convergence for $m \geq 2$, since singularities induced by the goal functional might not be resolved properly.*

**Example 2.35.** *For $\Omega = (0,1)^2$, we test Algorithm 2.17 with a semilinear variant of [MS09, Example 7.3]: The weak formulation of the primal problem reads: Find $u^\star \in H^1_0(\Omega)$ such that*

$$\langle\!\langle u^\star, v \rangle\!\rangle + \langle b(u^\star), v \rangle = \int_\Omega \boldsymbol{f} \cdot \nabla v \,\mathrm{d}x, \quad \text{for all } v \in H^1_0(\Omega), \tag{2.90}$$

*where $b(v) = v^3$ and $\boldsymbol{f} = \chi_{\Omega_{\boldsymbol{f}}}(-1,0)$ with the characteristic function $\chi_{\Omega_{\boldsymbol{f}}}$ of $\Omega_{\boldsymbol{f}} = \{x \in \Omega \mid$*




**(a)** Mesh plot for $m \in \{1,2,3,4\}$, where #DOF $\in \{80, 77, 79, 85\}$ for GOAFEM (left) and #DOF $\in \{90, 79, 79, 85\}$ for AFEM (right).

**(b)** Primal solution $u_H^\star$ (solid) and dual solution $z_H^\star[u_H^\star]$ (dashed), where $\#\mathcal{T}_H = 1236$ and $m = 1$.

**Figure 2.1:** Mesh plots (left) for ansatz spaces with $m \in \{1,2,3,4\}$ for GOAFEM (Algorithm 2.17) and standard AFEM and plots of the solutions $u_H^\star$ and $z_H[u_H^\star]$ (right) for for Example 2.34.

$x_1 + x_2 \leq \frac{1}{2}\}$. *The weak formulation of the practical dual problem for* $w \in H^1_0(\Omega)$ *reads: Find* $z^\star[w] \in H^1_0(\Omega)$ *such that*

$$\langle\!\langle z^\star[w]\,,\, v\rangle\!\rangle + \langle b'(w)z^\star[w]\,,\, v\rangle = \int_\Omega \boldsymbol{g}\cdot\nabla v\,\mathrm{d}x, \qquad \text{for all } v \in H^1_0(\Omega),$$

*where* $b'(v) = 3v^2$ *and* $\boldsymbol{g} = \chi_{\Omega_{\boldsymbol{g}}}(-1,0)$ *with* $\Omega_{\boldsymbol{g}} = \{x \in \Omega \mid x_1 + x_2 \geq \frac{3}{2}\}$. *Our implementation employs the Matlab code package* MooAFEM *[IP23] for 2D AFEM.*

*For various polynomial degrees* $m \in \{1,2,3,4\}$, *Figure 2.4 shows the goal error calculated with the proposed GOAFEM algorithm, the standard AFEM driven by the primal estimator* $\eta_\ell(u_\ell)^2$, *and AFEM driven by* $\eta_\ell(u_\ell)^2 + \zeta_\ell(z_\ell[u_\ell])^2$ *for the marking (AFEM+). Following [HPW21], we solve the discrete primal problems by an energy-based Newton iteration, where the energy reads*

$$\mathcal{E}(u^\star) = \frac{1}{2}\int_\Omega |\nabla u^\star|^2\,\mathrm{d}x + \int_\Omega \int_0^{u(x)} b(s)\,\mathrm{d}s\,\mathrm{d}x - \int_\Omega \boldsymbol{f}\cdot\nabla u^\star\,\mathrm{d}x.$$

*The reference goal value* $G(u^\star) = -0.0015849518088245$ *is obtained from the calculated goal values using GOAFEM with* $m = 4$. *For* $m = 1$, *an example of the meshes generated by GOAFEM (Algorithm 2.17) is shown in Figure 2.5a, by the standard AFEM algorithm in Figure 2.5b, and AFEM+ in Figure 2.5c. For GOAFEM and AFEM+, the singularities for both the primal and the dual problem are resolved, whereas for standard AFEM only those of the primal problem are taken into account. The meshes for* $m \in \{2,3,4\}$ *look similar with an increasing focus of the refinement on the singular points for increasing* $m$ *(not displayed). In particular, GOAFEM and AFEM+ lead to similar results, although in practice AFEM+ is slightly inferior from the point of theory (see Remark 2.22).*

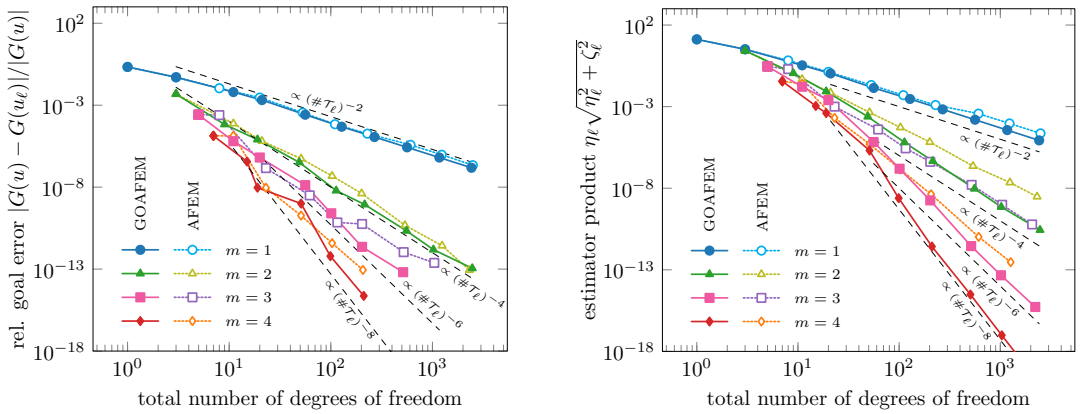



**Figure 2.2:** Relative goal error $|G(u^\star)-G(u_\ell^\star)|/|G(u^\star)|$ (left) and the estimator product $\eta_\ell\sqrt{\eta_\ell^2+\zeta_\ell^2}$ (right) over the total number of degrees of freedom in Example 2.34 for ansatz spaces with $m \in \{1,2,3,4\}$ for Algorithm 2.17 (solid) and standard AFEM (dotted).

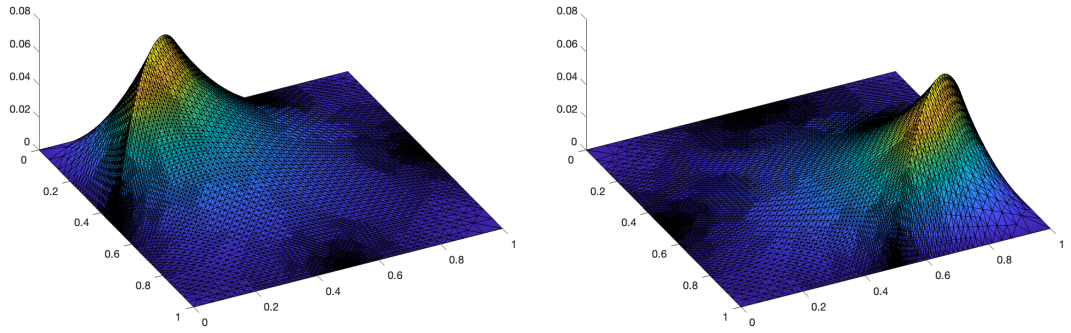


**Figure 2.3:** Plot of $u_H^\star$ (left) and $z_H^\star[u_H^\star]$ (right) generated by Algorithm 2.17, where $m=2$ and #DOF = 54653.

## 2.6 Contributions and conclusion

Let $(\mathcal{T}_\ell)_{\ell\in\mathbb{N}_0}$ be the sequence of meshes generated by the adaptive loop (2.9) of Algorithm 2.17. Let $\eta_\ell := \eta_\ell(u_\ell)$ and $\zeta_\ell := \zeta_\ell(z_\ell[u_\ell])$ be the corresponding computable error estimators, where $u_\ell$ and $z_\ell[u_\ell]$ are conforming piecewise polynomials of degree $\le m$ on $\mathcal{T}_\ell$, which solve the discrete primal and dual problem (2.4) and (2.6), respectively. We prove that the proposed adaptive strategy leads to linear convergence

$$\eta_{\ell+n}\,[\eta_{\ell+n}^2+\zeta_{\ell+n}^2]^{1/2} \le C_{\mathrm{lin}}\,q_{\mathrm{lin}}^n\,\eta_\ell\,[\eta_\ell^2+\zeta_\ell^2]^{1/2} \quad \text{for all } \ell, n\in\mathbb{N}_0, \tag{2.91}$$

where $C_{\mathrm{lin}} > 0$ and $0 < q_{\mathrm{lin}} < 1$ are generic constants. This guarantees that

$$\|u-u_\ell\|_{H^1(\Omega)}\|z[u_\ell]-z_\ell[u_\ell]\|_{H^1(\Omega)} + \|u-u_\ell\|_{H^1(\Omega)}^2 \xrightarrow{\ell\to\infty} 0.$$

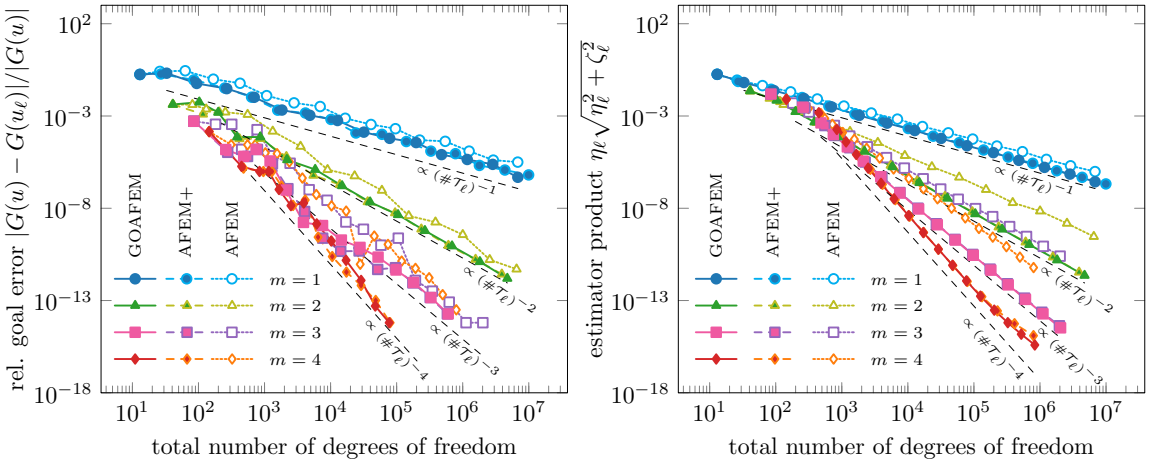
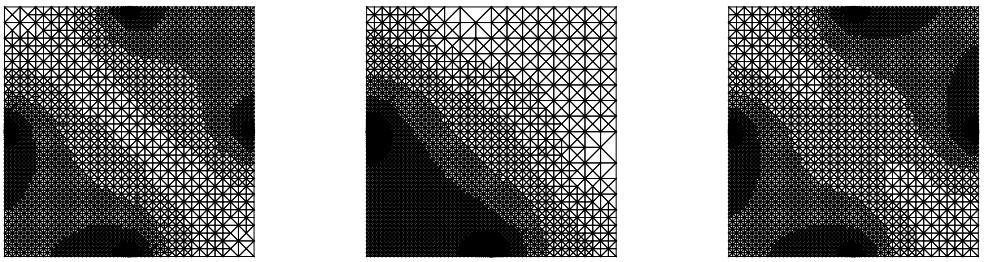



**Figure 2.4:** Relative goal error $|G(u^\star) - G(u_\ell^\star)|/|G(u)|$ (left) and estimator product $\eta_\ell\sqrt{\eta_\ell^2+\zeta_\ell^2}$ (right) for $m \in \{1,2,3,4\}$ with adaptive refinement according to Algorithm 2.17 (solid) compared to standard AFEM (dotted), and AFEM+ (dashed).



**(a)** GOAFEM mesh with $\#\mathcal{T}_\ell = 3602$.

**(b)** Standard AFEM mesh with $\#\mathcal{T}_\ell = 3727$.

**(c)** AFEM+ mesh with $\#\mathcal{T}_\ell = 4161$.

**Figure 2.5:** Visualization of adaptive meshes for Example 2.35 generated by Algorithm 2.17 (left), standard AFEM (center), and AFEM+ (right) for $m = 1$.

According to the goal-error estimate (2.7), this also yields convergence of the goal quantity $G(u_\ell) \to G(u)$ as $\ell \to \infty$.

Furthermore, we prove that the estimator product leads to convergence

$$\eta_\ell\,[\eta_\ell^2 + \zeta_\ell^2]^{1/2} = O((\#\mathcal{T}_\ell)^\alpha), \tag{2.92}$$

where the rate $\alpha = \min\{2s, s+t\}$ is optimal in the sense that $s > 0$ is any possible rate for $\eta_\ell$ and $t > 0$ is any possible rate for $\zeta_\ell$ (with respect to the usual approximation classes [CFPP14]). In particular, this is the first optimality result on GOAFEM for a nonlinear model problem. While the optimal rate would be $\alpha = s+t$ for linear model problems [MS09; FPZ16], the slightly worse rate $\alpha = \min\{2s, s+t\}$ stems from the fact that the adaptive algorithm must also control the linearization of the dual problem. Technical key results include Pythagoras-type quasi-orthogonalities for the semilinear model problem (2.2)

and the linearized dual problem (2.5). Finally, we note that our analysis allows to modify the marking strategies of [HPZ15; XHYM21] to ensure linear convergence of $\eta_\ell^2 + \zeta_\ell^2 = \mathcal{O}((\#\mathcal{T}_\ell)^{-\alpha})$ with rate $\alpha = \min\{2s, 2t\}$.

Finally, while prior results in the literature usually assumed global Lipschitz continuity of the semilinearity $b(u)$, our analysis relies only on growth conditions on $b(u)$ that imply local Lipschitz continuity. Furthermore, our analysis avoids any $L^\infty$-boundedness assumption on the discrete solutions as well as the necessity of a sufficiently fine initial mesh $\mathcal{T}_0$. Under such (usually unrealistic) assumptions, the present analysis could be simplified significantly.

## 2.7 Appendix: Well-posedness of primal and dual problems

Recall the operator

$$\mathcal{A}\colon H^1_0(\Omega) \to H^{-1}(\Omega), \quad \mathcal{A}w := \langle\!\langle w\,,\,\cdot\,\rangle\!\rangle + \langle b(w)\,,\,\cdot\,\rangle.$$

Assumption (GC) and the resulting estimate (2.10) yield that

$$\langle b(v)\,,\,w\rangle \overset{(2.10)}{\lesssim} \|b(v)\|_{L^{s'}(\Omega)}\,|||w||| < \infty.$$

Together with the continuity of $\langle\!\langle\,\cdot\,,\,\cdot\,\rangle\!\rangle$, we infer that $\mathcal{A}$ is well-defined.

The estimate (2.24) in combination with (ELL) leads us to

$$\begin{aligned}\langle \mathcal{A}w - \mathcal{A}v\,,\,w-v\rangle &= \langle\!\langle w-v\,,\,w-v\rangle\!\rangle + \langle b(w)-b(v)\,,\,w-v\rangle\\ &\geq |||w-v|||^2 \eqsim \|\nabla(w-v)\|^2_{L^2(\Omega)} \quad \text{for all } v,w\in H^1_0(\Omega),\end{aligned} \tag{2.93}$$

where the hidden constant depends only on $\mu_0$ from (ELL). This proves that $\mathcal{A}$ is strongly monotone and hence, in particular, monotone and coercive. Moreover, the solution $u^\star \in H^1_0(\Omega)$ of (2.3) is necessarily unique. Finally, recall from (CAR) that $b$ is smooth in $\xi$. Therefore, the mapping

$$\tau \mapsto \int_\Omega b(v+\tau w)\,\varphi\,\mathrm{d}x \in \mathbb{R} \quad \text{for } \tau\in[0,1] \text{ and } v,w,\varphi\in H^1_0(\Omega)$$

is continuous, i.e., $\mathcal{A}$ is hemi-continuous. Therefore, the Browder–Minty theorem applies and yields existence and uniqueness.

To address well-posedness of the theoretical dual problem (2.15), we show that (GC) implies that $\int_\Omega |\boldsymbol{B}^\star(w)zv|\,\mathrm{d}x < \infty$ for all $v,w,z\in H^1_0(\Omega)$. The cases $d\in\{1,2\}$ are covered, e.g., in [AW15, Lemma A.1]. If $d=3$, we exploit (GC) and apply the same reasoning as for

the estimate (2.12) to obtain that, with $t = 6$ and $t'' = 3/2$,

$$\begin{aligned}\langle \boldsymbol{B}(w)z\,,\,v\rangle &\overset{(2.12)}{\leq} \|\boldsymbol{B}(w)\|_{L^{t''}(\Omega)}\|z\|_{L^t(\Omega)}\|v\|_{L^t(\Omega)} \lesssim \left\|\int_0^1 b'(u+\tau(w-u))\,\mathrm{d}\tau\right\|_{L^{t''}(\Omega)}|||z|||\,|||v||| \\ &\overset{(GC)}{\lesssim} \left\|\int_0^1 (1+|u+\tau(w-u)|^{n-1}\,\mathrm{d}\tau\right\|_{L^{t''}(\Omega)}|||z|||\,|||v||| \\ &\lesssim \left(1+\int_0^1 \|(u+\tau(w-u))^{n-1}\|_{L^{t''}(\Omega)}\,\mathrm{d}\tau\right)|||z|||\,|||v||| < \infty,\end{aligned} \tag{2.94}$$

where the last step uses that $\|(u+\tau(w-u))^{n-1}\|_{L^{t''}(\Omega)} = \|u+\tau(w-u)\|^{n-1}_{L^{(n-1)t''}(\Omega)}$ with $(n-1)t'' \leq 4\cdot 3/2 = 6$ so that the bracket is uniformly bounded in terms of $|||u||| + |||w|||$; see Remark 2.1. Using (ELL) and (MON) for coercivity (see, e.g., (2.93) above), the Lax–Milgram lemma proves existence and uniqueness of $\tilde{z}^\star[w] \in H^1_0(\Omega)$ and $\tilde{z}^\star_H[w] \in \mathcal{X}_H$.

## 2.8 Appendix: Proof of Axioms of Adaptivity (A2)–(A4)

This section contains the standard arguments of (A2)–(A4), which carry over to the semilinear setting.

*Proof of reduction* (A2). For $T \in \mathcal{T}_H\backslash\mathcal{T}_h$, let $\mathcal{T}_h|_T := \{T' \in \mathcal{T}_h \mid T' \subseteq T\}$ denote the set of its children. Note that NVB guarantees that

$$h_{T'} \leq 2^{-1/d} h_T = 2^{-1/d}|T|^{1/d} \quad \text{for all } T' \in \mathcal{T}_h|_T\,. \tag{2.95}$$

Recall that

$$\eta_h(T', v_H)^2 = h_{T'}^2\,\|\Re(v_H)\|^2_{L^2(T')} + h_{T'}\,\|[\![(\boldsymbol{A}\nabla v_H + \boldsymbol{f})\cdot\boldsymbol{n}]\!]\|^2_{L^2(\partial T'\cap\Omega)}.$$

Applying the bisection estimate (2.95), we obtain that

$$\begin{aligned}\eta_h(\mathcal{T}_h\backslash\mathcal{T}_H, v_H)^2 &= \sum_{T'\in\mathcal{T}_h\backslash\mathcal{T}_H}\eta_h(T', v_H)^2 = \sum_{T\in\mathcal{T}_H\backslash\mathcal{T}_h}\sum_{T'\in\mathcal{T}_h|_T}\eta_h(T', v_H)^2 \\ &= \sum_{T\in\mathcal{T}_H\backslash\mathcal{T}_h}\sum_{T'\in\mathcal{T}_h|_T}\Big(h_{T'}^2\,\|\Re(v_H)\|^2_{L^2(T')} + h_{T'}\,\|[\![(\boldsymbol{A}\nabla v_H + \boldsymbol{f})\cdot\boldsymbol{n}]\!]\|^2_{L^2(\partial T'\cap\Omega)}\Big).\end{aligned}$$

For the first term, it holds that

$$\sum_{T'\in\mathcal{T}_h|_T} h_{T'}^2\,\|\Re(v_H)\|^2_{L^2(T')} \leq 2^{-2/d}\,h_T^2\,\|\Re(v_H)\|^2_{L^2(T)}.$$

For the second term, note that $v_H \in \mathcal{X}_H$ is a coarse-mesh function and, hence, smooth in

the interior of $T \in \mathcal{T}_H$. Hence, all jumps in the interior of $T \in \mathcal{T}_H$ vanish. This leads to

$$\sum_{T' \in \mathcal{T}_h|_T} h_{T'} \, \|[\![ (\boldsymbol{A}\nabla v_H + \boldsymbol{f}) \cdot \boldsymbol{n} ]\!]\|^2_{L^2(\partial T' \cap \Omega)} = \sum_{T' \in \mathcal{T}_h|_T} h_{T'} \, \|[\![ (\boldsymbol{A}\nabla v_H + \boldsymbol{f}) \cdot \boldsymbol{n} ]\!]\|^2_{L^2(\partial T' \cap \partial T \cap \Omega)}$$
$$\le 2^{-1/d} \, h_T \sum_{T' \in \mathcal{T}_h|_T} \|[\![ (\boldsymbol{A}\nabla v_H + \boldsymbol{f}) \cdot \boldsymbol{n} ]\!]\|^2_{L^2(\partial T' \cap \partial T \cap \Omega)} = 2^{-1/d} \, h_T \|[\![ (\boldsymbol{A}\nabla v_H + \boldsymbol{f}) \cdot \boldsymbol{n} ]\!]\|^2_{L^2(\partial T \cap \Omega)}.$$

Altogether, we conclude reduction (A2) for the primal estimator

$$\begin{aligned}\eta_h(\mathcal{T}_h \backslash \mathcal{T}_H, v_H)^2 &\le 2^{-2/d} \sum_{T \in \mathcal{T}_H \backslash \mathcal{T}_h} h_T^2 \|\Re(v_H)\|^2_{L^2(T)} + 2^{-1/d} \sum_{T \in \mathcal{T}_H \backslash \mathcal{T}_h} h_T \, \|[\![ (\boldsymbol{A}\nabla v_H + \boldsymbol{f}) \cdot \boldsymbol{n} ]\!]\|^2_{L^2(\partial T \cap \Omega)} \\ &\le 2^{-1/d} \, \eta_H(\mathcal{T}_H \backslash \mathcal{T}_h, v_H)^2.\end{aligned}$$

The same arguments apply for the dual estimator. □

*Sketch of proof of reliability* (A3). Assumptions (ELL) and (MON) yield that, for all $u, w, z \in H^1_0(\Omega)$,

$$|||u - u_H^\star|||^2 \overset{(2.93)}{\le} \langle \mathcal{A}(u) - \mathcal{A}(u_H^\star) \, , \, u - u_H^\star \rangle, \tag{2.96a}$$
$$|||z - z_H^\star[w]|||^2 \le \langle \mathcal{A}'[w](z - z_H^\star[w]) \, , \, z - z_H^\star[w] \rangle. \tag{2.96b}$$

For all $v_H \in \mathcal{X}_H$, the Galerkin orthogonalities for the primal and dual setting read

$$\langle \mathcal{A}(u^\star) - \mathcal{A}(u_H^\star) \, , \, v_H \rangle = 0 = \langle \mathcal{A}(u_h^\star) - \mathcal{A}(u_H^\star) \, , \, v_H \rangle, \tag{2.97a}$$
$$\langle \mathcal{A}'[w](z^\star[w]) - \mathcal{A}'[w](z_H^\star[w]) \, , \, v_H \rangle = 0 = \langle \mathcal{A}'[w](z_h^\star[w]) - \mathcal{A}'[w](z_H^\star[w]) \, , \, v_H \rangle. \tag{2.97b}$$

For $d \in \{2, 3\}$, let $_H : H^1_0(\Omega) \to \mathcal{X}_H$ be a Clément-type quasi-interpolation operator, while $_H$ is the nodal interpolation operator for $d = 1$. For the primal setting, let $u \in \{u^\star, u_h^\star\}$ and choose $\mathcal{X} \in \{H^1_0(\Omega), \mathcal{X}_h\}$ accordingly. Then, (2.96)–(2.97) and (2.3) or (2.4) (according to $u$) lead to

$$\begin{aligned}|||u - u_H^\star||| &\le \sup_{0 \ne v \in \mathcal{X}} |||v|||^{-1} \langle \mathcal{A}(u) - \mathcal{A}(u_H^\star) \, , \, v \rangle = \sup_{0 \ne v \in \mathcal{X}} |||v|||^{-1} \langle \mathcal{A}(u) - \mathcal{A}(u_H^\star) \, , \, v - {}_H v \rangle \\ &= \sup_{0 \ne v \in \mathcal{X}} |||v|||^{-1} \big[ \langle f \, , \, v - {}_H v \rangle + \langle \boldsymbol{f} \, , \, \nabla(v - {}_H v) \rangle - \langle \mathcal{A}(u_H^\star) \, , \, v - {}_H v \rangle \big].\end{aligned} \tag{2.98}$$

For the dual setting, let $z \in \{z^\star[w], z_h^\star[w]\}$ and choose $\mathcal{X} \in \{H^1_0(\Omega), \mathcal{X}_h\}$ accordingly. Using (2.96)–(2.97), and (2.5) or (2.6) (according to $z$), the same arguments as above yield that

$$|||z - z_H^\star[w]||| \le \sup_{0 \ne v \in \mathcal{X}} |||v|||^{-1} \big[ \langle g, v - {}_H v \rangle + \langle \boldsymbol{g}, \nabla(v - {}_H v) \rangle - \langle \mathcal{A}'[w](z_H^\star[w]), v - {}_H v \rangle \big]. \tag{2.99}$$

Based on (2.98)–(2.99), standard arguments employing elementwise integration by parts

and fine properties of Clément-type operators conclude reliability (A3), i.e.,

$$|||u^\star - u_H^\star||| \lesssim \eta_H(u_H^\star) \quad \text{and} \quad |||z^\star[w] - z_H^\star[w]||| \le C_{\mathrm{rel}}\, \zeta_H(z_H^\star[w]).$$

The hidden constants depend only on $_H$ and, hence, only on $d$ and $\mu_0$. □

*Sketch of proof of discrete reliability* (A4). To prove discrete reliability (A4), we choose $_H$ as the Scott–Zhang projector [SZ90] for $d \in \{2,3\}$, which is a Clément-type quasi-interpolation operator, and note that $_H$ can be chosen in such a way that $(v_h -_H v_h)|_T = 0$ for all $T \in \mathcal{T}_H \cap \mathcal{T}_h$ and $v_h \in \mathcal{X}_h$; see [CKNS08]. Standard arguments then show that

$$|||u_h^\star - u_H^\star||| \lesssim \eta_H(\mathcal{T}_H \backslash \mathcal{T}_h, u_H^\star) \quad \text{and} \quad |||z_h^\star[w] - z_H^\star[w]||| \lesssim \zeta_H(\mathcal{T}_H \backslash \mathcal{T}_h, z_H^\star[w]).$$

The hidden constants depend only on the dimension $d$, the polynomial degree $m$, and norm equivalence. This concludes the proof of discrete reliability (A4). □

# 3 Cost-optimal adaptive linearized adaptive FEM for semilinear elliptic PDEs

This chapter is taken from:

> [②AIL1]: R. Becker, M. Brunner, M. Innerberger, J. M. Melenk, and D. Praetorius. Cost-optimal adaptive iterative linearized FEM for semilinear elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 57(4):2193–2225, 2023. DOI: 10.1051/m2an/2023036

## 3.1 Introduction

### 3.1.1 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 [BV84; Dör96; MNS00; BDD04; Ste07; MSV08; CKNS08; KS11; CN12; FFP14] for linear elliptic equations, to [Vee02; DK08; BDK12; GMZ12; GHPS18] for certain quasi-linear PDEs, and to [CFPP14] 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 [Ste07; BMS10; AGL13; ALMS13] for algebraic solvers for linear PDEs, to [EEV11; GMZ11; AW15; HW18; GHPS18; HW20a; HW20b] for the iterative linearization of nonlinear PDEs, and to [EV13; HPSV21] for fully adaptive schemes including linearization and algebraic solver. For the latter works, the consideration is usually restricted to the class of strongly monotone and globally Lipschitz continuous nonlinearities; see [GMZ11] for the first plain convergence result, [HW20a] for an abstract framework for plain convergence of adaptive iteratively linearized finite element methods (AILFEM), [GHPS18; GHPS21] for rate-optimality of AILFEM based on the Zarantonello iteration (as proposed in [CW17]), and [HPW21] for rate-optimality for other linearization strategies including the Kačanov iteration as well as the damped Newton method. In particular, we note that [GHPS21; HPW21; HPSV21] prove optimal convergence rates with respect to the overall computational cost. For more general nonlinear operators, optimal convergences rates are empirically observed (e.g., [EV13]), but the quest for a sound mathematical analysis is still ongoing.

### 3.1.2 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 $u^\star \in H^1_0(\Omega)$ to the (scalar) semilinear elliptic PDE

$$-\operatorname{div}(\boldsymbol{A}\nabla u^\star) + b(u^\star) = f - \operatorname{div} \boldsymbol{f} \text{ in } \Omega \quad \text{subject to} \quad u^\star = 0 \text{ on } \partial\Omega, \tag{3.1}$$

where we refer to Section 3.3 for a discussion of the precise assumptions on the diffusion matrix $\boldsymbol{A}$, the semilinearity $b$, and the given data $f$ and $\boldsymbol{f}$. 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 [HW20a; GHPS21; HPW21] 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 $\delta$ is chosen appropriately (Corollary 3.11). For arbitrary adaptivity parameters ($\theta$ for marking and $\lambda$ for stopping the Zarantonello iteration), we then prove *full* linear convergence (Theorem 3.14), 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* (Theorem 3.17) and *cost-optimality*, i.e., rate-optimality with respect to the overall computational cost (Corollary 3.19).

### 3.1.3 Outline

This work is organized as follows: In Section 3.2, we present our adaptive iterative linearized finite element method (Algorithm 3.10) 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 (Theorem 3.14) and optimal rates both with respect to the degrees of freedom (Theorem 3.17) as well as the overall computational cost (Corollary 3.19). In Section 3.3, we introduce and discuss semilinear

elliptic PDEs, which fit into the abstract framework of Section 3.2. Section 3.4 presents a practical extension of our AILFEM strategy (Algorithm 3.23), which includes the adaptive choice of the Zarantonello damping parameter $\delta$. In Section 3.5, we support our theoretical findings with numerical experiments. Finally, Appendix 3.6 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.

### 3.1.4 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 $\mathcal{U}$.

## 3.2 Strongly monotone operators

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

### 3.2.1 Abstract model problem

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

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

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

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

**Remark 3.1.** *[Zei90, p. 565] defines local Lipschitz continuity as follows: For all $\Theta > 0$, there exists $L'[\Theta] > 0$ such that*

$$\langle \mathcal{A}v-\mathcal{A}w,\,\varphi\rangle \le L'[\Theta]\,|||v-w|||\,|||\varphi||| \text{ for all } v,w,\varphi \in \mathcal{X} \text{ with } \max\{|||v|||,|||w|||\} \le \Theta. \tag{3.2}$$

*Conditions* (LIP) *and* (3.2) *are indeed equivalent in the sense that*

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

*However,* (LIP) *is better suited for the inductive structure in the proof of Corollary 3.5.*

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

$$\mathcal{A}u^\star = F. \tag{3.3}$$

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

$$\langle \mathcal{A}u_H^\star \,,\, v_H\rangle = \langle F \,,\, v_H\rangle \quad \text{for all } v_H \in \mathcal{X}_H. \tag{3.4}$$

We observe that the setting of strongly monotone and locally Lipschitz operators yields existence and uniqueness of the solutions to (3.3)–(3.4) as well as a Céa-type estimate.

**Proposition 3.2.** *Suppose that $\mathcal{A}$ satisfies (SM) and (LIP). Then, (3.3)–(3.4) admit unique solutions $u^\star \in \mathcal{X}$ and $u_H^\star \in \mathcal{X}_H$, respectively, and it holds that*

$$\max\big\{|||u^\star|||, |||u_H^\star|||\big\} \leq M := \frac{1}{\alpha}\,\|F - \mathcal{A}0\|_{\mathcal{X}'} \neq 0 \tag{3.5}$$

*as well as*

$$|||u^\star - u_H^\star||| \leq C_{\text{Céa}} \min_{v_H \in \mathcal{X}_H} |||u^\star - v_H||| \quad \textit{with} \quad C_{\text{Céa}} = L[2M]/\alpha. \tag{3.6}$$

*Proof.* Since $\mathcal{A}$ is (even locally Lipschitz) continuous, existence of $u_H^\star$ follows from the Browder–Minty theorem on monotone operators [Zei90, Theorem 26.A]. Uniqueness of $u_H^\star$ follows from strong monotonicity, since any two solutions $u_H^\star, u_H \in \mathcal{X}_H$ to (3.4) satisfy

$$\alpha\,|||u_H^\star - u_H|||^2 \overset{\text{(SM)}}{\leq} \langle \mathcal{A}u_H^\star - \mathcal{A}u_H \,,\, u_H^\star - u_H\rangle \overset{(3.4)}{=} 0$$

and hence $u_H^\star = u_H$. Boundedness (3.5) follows from

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

Since (3.3) is equivalent to (3.4) with $\mathcal{X} = \mathcal{X}_H$, the foregoing results also cover $u^\star \in \mathcal{X}$. This concludes the proof of (3.5). To see the Céa-type estimate (3.6), recall the Galerkin orthogonality

$$\langle \mathcal{A}u^\star - \mathcal{A}u_H^\star \,,\, v_H\rangle = 0 \quad \text{for all } v_H \in \mathcal{X}_H. \tag{3.7}$$

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

$$\begin{aligned}\alpha\,|||u^\star - u_H^\star|||^2 &\overset{\text{(SM)}}{\leq} \langle \mathcal{A}u^\star - \mathcal{A}u_H^\star \,,\, u^\star - u_H^\star\rangle \overset{(3.7)}{=} \langle \mathcal{A}u^\star - \mathcal{A}u_H^\star \,,\, u^\star - v_H\rangle \\ &\overset{\text{(LIP)}}{\leq} L[2M]\,|||u^\star - u_H^\star|||\,|||u^\star - v_H|||.\end{aligned}$$

Rearranging the last estimate, we prove (3.6), where the minimum is attained since $\mathcal{X}_H$ is closed. This concludes the proof. □

Finally, we suppose that the operator $\mathcal{A}$ possesses a potential $\mathcal{P}$: there exists a Gâteaux

differentiable function $\mathcal{P}\colon X \to \mathbb{R}$ such that its derivative $\mathrm{d}\mathcal{P}\colon X \to X'$ coincides with $\mathcal{A}$, i.e., it holds that

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

We define the energy $\mathcal{E}(v) := (\mathcal{P}-F)v$, where $F$ is the right-hand side from (3.3).

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

$$\mathcal{E}(v_H)-\mathcal{E}(u^\star) = \big[\mathcal{E}(v_H)-\mathcal{E}(u_H^\star)\big] + \big[\mathcal{E}(u_H^\star)-\mathcal{E}(u^\star)\big] \quad \text{for all } v_H\in X_H \tag{3.8}$$

and all these energy differences are non-negative; see (3.10).

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

**Lemma 3.3** (see, e.g., [GHPS18, Lemma 5.1])**.** *Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and (POT). Let $\vartheta \geq M$. Let $v_H\in X_H$ with $|||v_H-u_H^\star||| \leq \vartheta$. Then, it holds that*

$$\frac{\alpha}{2}\,|||v_H-u_H^\star|||^2 \leq \mathcal{E}(v_H)-\mathcal{E}(u_H^\star) \leq \frac{L[\vartheta]}{2}\,|||v_H-u_H^\star|||^2. \tag{3.9}$$

*In particular, the solution $u_H^\star$ of (3.4) is indeed the unique minimizer of $\mathcal{E}$ in $X_H$, i.e.,*

$$\mathcal{E}(u_H^\star) \leq \mathcal{E}(v_H) \quad \textit{for all } v_H\in X_H, \tag{3.10}$$

*and, therefore, (3.4) can equivalently be reformulated as an energy minimization problem:*

$$\textit{Find} \quad u_H^\star\in X_H \quad \textit{such that} \quad \mathcal{E}(u_H^\star) = \min_{v_H\in X_H}\mathcal{E}(v_H). \qquad \square$$

### 3.2.2 Zarantonello iteration

Let $X_H \subseteq X$ be a closed subspace. For given damping parameter $\delta > 0$, we define the Zarantonello mapping $\Phi_H(\delta;\cdot)\colon X_H \to X_H$ by

$$\langle\!\langle \Phi_H(\delta;w_H)\,,\,v_H\rangle\!\rangle = \langle\!\langle w_H\,,\,v_H\rangle\!\rangle + \delta\,\langle F-\mathcal{A}w_H\,,\,v_H\rangle \quad \text{for all } v_H\in X_H. \tag{3.11}$$

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

$$|||\Phi_H(\delta;w_H)-w_H||| \leq \delta\,\|F-\mathcal{A}w_H\|_{X'} = \delta \sup_{v\in X\setminus\{0\}} \frac{\langle F-\mathcal{A}w_H\,,\,v\rangle}{|||v|||} \text{ for all } w_H\in X_H; \tag{3.12}$$

second,

$$|||\Phi_H(\delta;v_H)-\Phi_H(\delta;w_H)||| \leq |||v_H-w_H||| + \delta\,\|\mathcal{A}v_H-\mathcal{A}w_H\|_{X'} \text{ for all } v_H,w_H\in X_H. \tag{3.13}$$

Due to the local Lipschitz continuity (LIP) of $\mathcal{A}$, this proves that also $\Phi_H(\delta;\cdot)$ is locally Lipschitz continuous. By definition, $u_H^\star \in \mathcal{X}_H$ solves (3.4) if and only it is a fixed point of $\Phi_H(\delta;\cdot)$, i.e., $u_H^\star = \Phi_H(\delta; u_H^\star)$.

### 3.2.3 Zarantonello iteration and norm contraction

Let $\mathcal{X}_H \subseteq \mathcal{X}$ be a closed subspace. The next proposition [Zei90, Section 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 3.4** (norm contraction)**.** *Suppose that $\mathcal{A}$ satisfies (SM) and (LIP). Let $\vartheta > 0$ and $v_H, w_H \in \mathcal{X}_H$ with $\max\{|||v_H|||, |||v_H - w_H|||\} \le \vartheta$. 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*

$$|||\Phi_H(\delta; v_H) - \Phi_H(\delta; w_H)||| \le q_N[\delta]\, |||v_H - w_H|||. \tag{3.14}$$

*We note that $q_N[\delta] \to 1$ as $\delta \to 0$. Moreover, for known $\alpha$ 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

$$I_H\colon \mathcal{X}_H \to \mathcal{X}_H', \quad v_H \mapsto I_H(v_H) := \langle\!\langle \cdot\,,\, v_H\rangle\!\rangle \quad \text{for all } v_H \in \mathcal{X}_H \tag{3.15}$$

is an isometric isomorphism; cf., e.g., [Yos95, Chapter 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 $\max\{|||v_H|||, |||v_H - w_H|||\} \le \vartheta$, 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

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

The isometry property of $I_H$ implies that

$$|||I_H^{-1}(\mathcal{A}v_H - \mathcal{A}w_H)|||^2 \overset{(3.15)}{=} \|\mathcal{A}v_H - \mathcal{A}w_H\|_{\mathcal{X}'}^2 \overset{(LIP)}{\le} L[\vartheta]^2\, |||v_H - w_H|||^2.$$

Moreover, it holds that

$$\langle\!\langle v_H - w_H\,,\, I_H^{-1}(\mathcal{A}v_H - \mathcal{A}w_H)\rangle\!\rangle \overset{(3.15)}{=} \langle \mathcal{A}v_H - \mathcal{A}w_H\,,\, v_H - w_H\rangle \overset{(SM)}{\ge} \alpha\, |||v_H - w_H|||^2.$$

Combining these observations, we see that

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

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 $q_{\mathrm{N}}[\delta]$ if $\alpha$ and $L[\vartheta]^2$ are fixed. This concludes the proof. □

**Corollary 3.5.** *Suppose that $\mathcal{A}$ satisfies (SM) and (LIP). Let $u_H^0 \in \mathcal{X}_H$ with $|||u_H^0||| \le 2M$. Let $0 < \delta < 2\alpha/L[3M]^2$ and let $0 < q_{\mathrm{N}}[\delta] < 1$ be chosen according to Proposition 3.4, where $\vartheta = 3M$. Define*

$$u_H^{k+1} := \Phi_H(\delta; u_H^k) \quad \text{for all } k \in \mathbb{N}_0. \tag{3.16}$$

*Then, it holds that*

$$(1 - q_{\mathrm{N}}[\delta])\,|||u_H^\star - u_H^k||| \le |||u_H^{k+1} - u_H^k||| \le (1 + q_{\mathrm{N}}[\delta])\,|||u_H^\star - u_H^k||| \tag{3.17}$$

*and*

$$|||u_H^\star - u_H^{k+1}||| \le q_{\mathrm{N}}[\delta]\,|||u_H^\star - u_H^k||| \le q_{\mathrm{N}}[\delta]^{k+1}\,|||u_H^\star - u_H^0||| \le 3M \quad \text{for all } k \in \mathbb{N}_0. \tag{3.18}$$

*In particular, it follows that*

$$|||u_H^k||| \le 4M \quad \text{for all } k \in \mathbb{N}_0. \tag{3.19}$$

*Proof.* The claim (3.18) is proved by induction on $k$. By recalling (3.5), it holds that $|||u_H^\star||| \le M$ as well as $|||u_H^\star - u_H^0||| \le |||u_H^\star||| + |||u_H^0||| \le 3M$. Therefore, Proposition 3.4 proves that

$$|||u_H^\star - u_H^1||| = |||\Phi_H(\delta; u_H^\star) - \Phi_H(\delta; u_H^0)||| \overset{(3.14)}{\le} q_{\mathrm{N}}[\delta]\,|||u_H^\star - u_H^0||| \le 3M.$$

This proves (3.18) for $k = 0$. In the induction step, we know that $|||u_H^\star - u_H^k||| \le 3M$. As before, (3.14) from Proposition 3.4 and the induction hypothesis prove that

$$\begin{aligned} |||u_H^\star - u_H^{k+1}||| = |||\Phi_H(\delta; u_H^\star) - \Phi_H(\delta; u_H^k)||| &\overset{(3.14)}{\le} q_{\mathrm{N}}[\delta]\,|||u_H^\star - u_H^k||| \\ &\le q_{\mathrm{N}}[\delta]^{k+1}\,|||u_H^\star - u_H^0||| \le 3M. \end{aligned}$$

This proves (3.18) for general $k \in \mathbb{N}_0$, and the inequalities (3.17) follow from (3.14) and the triangle inequality. Moreover, the triangle inequality yields that

$$|||u_H^k||| \le |||u_H^\star||| + |||u_H^\star - u_H^k||| \le 4M.$$

This concludes the proof. □

**Corollary 3.6.** *Suppose that $\mathcal{A}$ satisfies (SM) and (LIP). Let $u_H^0 \in \mathcal{X}_H$ with $|||u_H^0||| \le 2M$. Let $0 < \delta < 2\alpha/L[6M]^2$ and let $0 < q_{\mathrm{N}}[\delta] < 1$ be chosen according to Proposition 3.4, where $\vartheta = 6M$. Then, the Zarantonello iterates from (3.16) satisfy (3.17)–(3.19) as well as*

$$|||u_H^{k+1} - u_H^k||| \le q_{\mathrm{N}}[\delta]\,|||u_H^k - u_H^{k-1}||| \le q_{\mathrm{N}}[\delta]^k\,|||u_H^1 - u_H^0||| \le 6M \text{ for all } k \in \mathbb{N}. \tag{3.20}$$

*Proof.* Since $L[3M] \le L[6M]$, it only remains to prove (3.20). We argue by induction and note that

$$|||u_H^1 - u_H^0||| \le |||u_H^1||| + |||u_H^0||| \overset{(3.19)}{\le} 6M.$$

Therefore, Proposition 3.4 proves that

$$|||u_H^2 - u_H^1||| = |||\Phi_H(\delta; u_H^1) - \Phi_H(\delta; u_H^0)||| \overset{(3.14)}{\le} q_{\mathrm{N}}[\delta]\, |||u_H^1 - u_H^0||| \le 6M.$$

This proves (3.20) for $k = 1$. In the induction step, we know that $|||u_H^{k+1} - u_H^k||| \le 6M$. Therefore, Proposition 3.4 and the induction hypothesis prove that

$$|||u_H^{k+2} - u_H^{k+1}||| = |||\Phi_H(\delta; u_H^{k+1}) - \Phi_H(\delta; u_H^k)||| \overset{(3.14)}{\le} q_{\mathrm{N}}[\delta]\, |||u_H^{k+1} - u_H^k||| \le 6M.$$

This proves (3.20) for general $k \in \mathbb{N}$ and concludes the proof. □

### 3.2.4 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 [HW20a, Proposition 1] to the Zarantonello iteration in the locally Lipschitz continuous setting.

**Lemma 3.7.** *Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and (POT). Let $u_H^0 \in \mathcal{X}_H$ with $|||u_H^0||| \le 2M$. Then, for $0 < \delta < 2\alpha/L[6M]^2$, the Zarantonello iteration (3.11) yields that*

$$0 \le \kappa[\delta]\, |||u_H^{k+1} - u_H^k|||^2 \le \mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1}) \le K[\delta]\, |||u_H^{k+1} - u_H^k|||^2, \tag{3.21}$$

*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

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

For $0 < \delta < 2\alpha/L[6M]^2$, Corollary 3.6 together with the boundedness $|||u_H^k||| \le 4M$ from (3.19) and the convexity of the norm show that

$$\max\{|||e_H^{k+1}|||, |||u_H^k - t\, e_H^{k+1}|||\} \le 6M \quad \text{for all } k \in \mathbb{N}_0. \tag{3.22}$$

With the fundamental theorem of calculus and the Zarantonello iteration (3.11), we see

that

$$\begin{aligned}
\mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1}) &= -\int_0^1 \langle \mathcal{A}(u_H^k + t\, e_H^{k+1}) - F\,,\, e_H^{k+1}\rangle \,\mathrm{d}t \\
&= -\int_0^1 \langle \mathcal{A}(u_H^k + t\, e_H^{k+1}) - \mathcal{A}u_H^k\,,\, e_H^{k+1}\rangle \,\mathrm{d}t - \langle \mathcal{A}u_H^k - F\,,\, e_H^{k+1}\rangle \\
&\overset{(3.11)}{=} -\int_0^1 \langle \mathcal{A}(u_H^k + t\, e_H^{k+1}) - \mathcal{A}u_H^k\,,\, e_H^{k+1}\rangle \,\mathrm{d}t + \frac{1}{\delta}\langle\!\langle e_H^{k+1}\,,\, e_H^{k+1}\rangle\!\rangle \\
&\overset{(LIP)}{\geq} \Big(\frac{1}{\delta} - \int_0^1 t L[6M]\,\mathrm{d}t\Big)|||u_H^{k+1} - u_H^k|||^2 = \Big(\frac{1}{\delta} - \frac{L[6M]}{2}\Big)|||u_H^{k+1} - u_H^k|||^2.
\end{aligned}$$

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 (3.21). Moreover, the same argument also yields that

$$\begin{aligned}
\mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1}) &\overset{(3.11)}{=} -\int_0^1 \langle \mathcal{A}(u_H^k + t\, e_H^{k+1}) - \mathcal{A}u_H^k\,,\, e_H^{k+1}\rangle \,\mathrm{d}t + \frac{1}{\delta}\langle\!\langle e_H^{k+1}\,,\, e_H^{k+1}\rangle\!\rangle \\
&\overset{(SM)}{\leq} \Big(\frac{1}{\delta} - \int_0^1 \alpha\, t\,\mathrm{d}t\Big)|||u_H^{k+1} - u_H^k|||^2 = \Big(\frac{1}{\delta} - \frac{\alpha}{2}\Big)|||u_H^{k+1} - u_H^k|||^2.
\end{aligned}$$

This concludes the proof. □

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

**Proposition 3.8** (energy contraction)**.** *Suppose that $\mathcal{A}$ satisfies* (SM), (LIP), *and* (POT). *Then, for $0 < \delta < 2\alpha/L[6M]^2$, it holds that*

$$0 \leq \mathcal{E}(u_H^{k+1}) - \mathcal{E}(u_H^\star) \leq q_{\mathrm{E}}[\delta]^2\,[\mathcal{E}(u_H^k) - \mathcal{E}(u_H^\star)] \quad \text{for all } k \in \mathbb{N}_0 \tag{3.23a}$$

*with contraction constant*

$$0 \leq q_{\mathrm{E}}[\delta]^2 := 1 - \Big(1 - \frac{\delta L[6M]}{2}\Big)\frac{2\delta\alpha^2}{L[3M]} < 1. \tag{3.23b}$$

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

$$(1 - q_{\mathrm{E}}[\delta]^2)\,\big[\mathcal{E}(u_H^k) - \mathcal{E}(u_H^\star)\big] \leq \mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1}) \leq (1 + q_{\mathrm{E}}[\delta]^2)\,\big[\mathcal{E}(u_H^k) - \mathcal{E}(u_H^\star)\big]. \tag{3.24}$$

*Proof.* First, we observe that

$$\begin{aligned}
\alpha\,|||u_H^\star - u_H^k|||^2 &\leq \langle \mathcal{A}u_H^\star - \mathcal{A}u_H^k\,,\, u_H^\star - u_H^k\rangle \overset{(3.4)}{=} \langle F - \mathcal{A}u_H^k\,,\, u_H^\star - u_H^k\rangle \\
&\overset{(3.11)}{=} \frac{1}{\delta}\langle\!\langle u_H^{k+1} - u_H^k\,,\, u_H^\star - u_H^k\rangle\!\rangle \leq \frac{1}{\delta}|||u_H^{k+1} - u_H^k|||\,|||u_H^\star - u_H^k|||.
\end{aligned} \tag{3.25}$$

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

$$\begin{aligned}
0 &\overset{(3.9)}{\leq} \mathcal{E}(u_H^{k+1}) - \mathcal{E}(u_H^\star) = \mathcal{E}(u_H^k) - \mathcal{E}(u_H^\star) - \big[\mathcal{E}(u_H^k) - \mathcal{E}(u_H^{k+1})\big] \\
&\overset{(3.21)}{\leq} \mathcal{E}(u_H^k) - \mathcal{E}(u_H^\star) - \Big(\frac{1}{\delta} - \frac{L[6M]}{2}\Big) |||u_H^{k+1} - u_H^k|||^2 \\
&\overset{(3.25)}{\leq} \mathcal{E}(u_H^k) - \mathcal{E}(u_H^\star) - \Big(\frac{1}{\delta} - \frac{L[6M]}{2}\Big) \delta^2\alpha^2 |||u_H^\star - u_H^k|||^2 \\
&\overset{(3.9)}{\leq} \Big[1 - \Big(1 - \frac{\delta L[6M]}{2}\Big) \frac{2\delta\alpha^2}{L[3M]}\Big] \big[\mathcal{E}(u_H^k) - \mathcal{E}(u_H^\star)\big],
\end{aligned}$$

where (3.9) holds due to (3.18) from Corollary 3.5. This proves (3.23). The inequalities (3.24) follow from the triangle inequality. This concludes the proof. □

**Remark 3.9.** *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 $q_{\mathrm{E}}[\delta]^2 = 1 - \frac{\alpha^2}{L^2}$. In contrast, the optimal norm contraction factor $q_{\mathrm{N}}[\delta]^2 = 1 - \frac{\alpha^2}{L^2}$ is obtained for $\delta = \frac{\alpha}{L^2}$; cf. Proposition 3.4. To allow a larger damping parameter $\delta > 0$, energy contraction is preferred.*

### 3.2.5 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., [Ste08]), or the 1D bisection from [AFF+13] for $d = 1$. For each triangulation $\mathcal{T}_H$ and a set of marked elements $\mathcal{M}_H \subseteq \mathcal{T}_H$, let $\mathcal{T}_h := \texttt{refine}(\mathcal{T}_H, \mathcal{M}_H)$ be the coarsest triangulation such that all $T \in \mathcal{M}_H$ have been refined, i.e., $\mathcal{M}_H \subseteq \mathcal{T}_H \backslash \mathcal{T}_h$. 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 $\mathcal{X}_H \subseteq \mathcal{X}_h \subset \mathcal{X}$.

### 3.2.6 Axioms of adaptivity and *a posteriori* error estimator

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

$$\eta_H(T, \cdot)\colon \mathcal{X}_H \to \mathbb{R}_{\geq 0} \quad \text{for all } T \in \mathcal{T}_H \tag{3.26}$$

be the local contributions of an *a posteriori* error estimator

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

We suppose that the error estimator $\eta_H$ satisfies the following axioms of adaptivity from [CFPP14] with a slightly relaxed variant of stability (A1) from [①GOA].

**(A1) stability:** For all $\vartheta > 0$ and all[1] $\mathcal{U}_H \subseteq \mathcal{T}_h \cap \mathcal{T}_H$, there exists $C_{\mathrm{stab}}[\vartheta] > 0$ 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

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

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

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

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

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

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

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

### 3.2.7 Idealized adaptive algorithm

In the following, we formulate and analyze an AILFEM algorithm in the spirit of [GHPS21], but with an extended stopping criterion in Algorithm 3.10(i.b), i.e.,

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

Clearly, if the stopping criterion from Algorithm 3.10(i.b) holds, then also the simpler stopping criterion $|\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \leq \lambda^2\, \eta_\ell(u_\ell^k)$ from [GHPS21, Algorithm 2] holds.

The proposed algorithm is idealized in the sense that an appropriate parameter $\delta > 0$ is chosen *a priori*; see Theorem 3.17 below.

**Algorithm 3.10: idealized AILFEM with energy contraction**

**Input:** initial triangulation $\mathcal{T}_0$, initial guess $u_0^0 := 0$ with $M = \frac{1}{a}\|F - \mathcal{A}0\|_{X'} < \infty$ according to (3.5), marking parameters $0 < \theta \leq 1$ and $1 \leq C_{\mathrm{mark}} < \infty$, damping parameter $\delta > 0$, solver parameter $\lambda > 0$.

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

(i) For all $k = 1, 2, 3, \ldots$, 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 $\big(|\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \leq \lambda^2\, \eta_\ell(u_\ell^k)^2 \quad \wedge \quad |||u_\ell^k||| \leq 2M\big)$.

(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 $C_{\mathrm{mark}}$ minimal cardinality such that $\theta\, \eta_\ell(u_\ell^{\underline{k}(\ell)})^2 \leq \sum_{T \in \mathcal{M}_\ell} \eta_\ell(T, u_\ell^{\underline{k}(\ell)})^2$.

[1] While [①GOA, Proposition 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.

(iv) Generate $\mathcal{T}_{\ell+1} := \texttt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ and define $u_{\ell+1}^0 := u_\ell^{\underline{k}(\ell)}$.

Following [GHPS21], the analysis of Algorithm 3.10 requires the ordered index set

$$\mathcal{Q} := \{(\ell, k) \in \mathbb{N}_0^2 \mid \text{index pair } (\ell, k) \text{ occurs in Algorithm 3.10 and } k < \underline{k}(\ell)\}, \tag{3.27}$$

where $\underline{k}(\ell) \geq 1$ counts the number of solver steps for each $\ell$. 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 3.10 is sequential, the index set $\mathcal{Q}$ is lexicographically ordered: For $(\ell, k)$ and $(\ell', k') \in \mathcal{Q}$, we write $(\ell', k') < (\ell, k)$ if and only if $(\ell', k')$ appears earlier in Algorithm 3.10 than $(\ell, k)$. 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 $u_\ell^k$.

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 $(\ell, k) \in \mathbb{N}_0 \times \mathbb{N}_0$,

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

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 3.11.** *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 3.10 such that there exist $0 < q_{\mathrm{N}} < 1$ and $0 < q_{\mathrm{E}} < 1$ such that the following properties hold:*

- **nested iteration:** $|||u_\ell^0||| \leq 2M$ for all $(\ell, 0) \in \mathcal{Q}$; (3.28)
- **boundedness:** $|||u_\ell^k||| \leq 4M$ for all $(\ell, k) \in \mathcal{Q}$; (3.29)
- **norm contraction:** $|||u_\ell^\star - u_\ell^{k+1}||| \leq q_{\mathrm{N}} |||u_\ell^\star - u_\ell^k|||$ for all $(\ell, k) \in \mathcal{Q}$; (3.30)
- **energy contraction:** $\mathcal{E}(u_\ell^{k+1}) - \mathcal{E}(u_\ell^\star) \leq q_{\mathrm{E}}^2 \big[\mathcal{E}(u_\ell^k) - \mathcal{E}(u_\ell^\star)\big]$ for all $(\ell, k) \in \mathcal{Q}$. (3.31)

*Moreover, this guarantees (3.17)–(3.18) for all $(\ell, k) \in \mathcal{Q}$ with $q_{\mathrm{N}}[\delta]$ replaced by $q_{\mathrm{N}}$. 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 3.10 and $u_0^0 := 0$, we have that $|||u_\ell^0||| \leq 2M$. Then, $|||u_\ell^\star - u_\ell^0||| \leq 3M$. Choose $0 < q_{\mathrm{N}} := q_{\mathrm{N}}[\delta] < 1$ according to

Proposition 3.4, where $\vartheta = 3M$ as well as $0 < q_{\mathrm{E}} := q_{\mathrm{E}}[\delta] < 1$ according to Proposition 3.8. This proves norm contraction (3.30) as well as energy contraction (3.31) for all $(\ell, k) \in \mathcal{Q}$. Furthermore, for all $(\ell, k) \in \mathcal{Q}$, it follows that

$$|||u_\ell^k||| \le |||u_\ell^\star||| + |||u_\ell^\star - u_\ell^k||| \overset{(3.18)}{\le} M + q_{\mathrm{N}}^k\, |||u_\ell^\star - u_\ell^0||| \le M + q_{\mathrm{N}}^k\, 3M \le 4M, \tag{3.32}$$

which proves boundedness (3.29). Moreover, (3.32) together with $0 < q_{\mathrm{N}} < 1$ from (3.30) proves that there exists $k_0 \in \mathbb{N}_0$, which is independent of $\ell$, such that, for all $k \ge k_0$, it holds that

$$|||u_\ell^k||| \overset{(3.32)}{\le} M + q_{\mathrm{N}}^k\, 3M \overset{!}{\le} 2M.$$

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

### 3.2.8 AILFEM under the assumption of energy contraction (3.31)

Norm contraction (3.30) is the critical ingredient in the proof of Corollary 3.11 — leading to boundedness (Corollary 3.5), which is key to the proof of energy contraction (3.31) (cf. (3.22)). Thus, norm contraction (3.30) is sufficient for obtaining nested iteration (3.28), boundedness (3.29), and energy contraction (3.31). However, supposing (3.31) 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 (3.31) holds for all $(\ell, k) \in \mathcal{Q}$.

**Lemma 3.12.** *Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and (POT). Suppose that the choice of $\delta > 0$ guarantees that Algorithm 3.10 satisfies energy contraction (3.31). Then, it holds that*

$$|||u_\ell^k||| \le M + 3M\,\frac{L[3M]}{\alpha} =: \frac{\tau}{2} \quad \text{for all } (\ell, k) \in \mathcal{Q}. \tag{3.33a}$$

*Moreover, it holds that*

$$|||u_\ell^k - u_\ell^{k'}||| \le \tau \quad \text{for all } (\ell, k), (\ell, k') \in \mathcal{Q}. \tag{3.33b}$$

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

$$|||u_\ell^k||| \le 2M \quad \text{for all } (\ell, k) \in \mathcal{Q} \text{ with } k \ge k_0. \tag{3.34}$$

*Proof.* From Algorithm 3.10 and $u_0^0 := 0$, we have that $|||u_\ell^0||| \le 2M$. With $|||u_\ell^\star||| \le M$ from (3.5), it holds that $|||u_\ell^\star - u_\ell^0||| \le 3M$. For all $(\ell, k) \in \mathcal{Q}$, it follows that

$$\begin{aligned}
|||u_\ell^k||| &\le |||u_\ell^\star||| + |||u_\ell^\star - u_\ell^k||| \overset{(3.9)}{\le} M + \Big(\frac{2}{\alpha}\Big)^{1/2} \big(\mathcal{E}(u_\ell^k) - \mathcal{E}(u_\ell^\star)\big)^{1/2} \\
&\overset{(3.31)}{\le} M + q_{\mathrm{E}}^k \Big(\frac{2}{\alpha}\Big)^{1/2} \big(\mathcal{E}(u_\ell^0) - \mathcal{E}(u_\ell^\star)\big)^{1/2} \overset{(3.9)}{\le} M + q_{\mathrm{E}}^k\, 3M \Big(\frac{L[3M]}{\alpha}\Big)^{1/2} \qquad (3.35) \\
&\overset{(3.31)}{\le} M + 3M \Big(\frac{L[3M]}{\alpha}\Big)^{1/2} =: \frac{\tau}{2}. \qquad (3.36)
\end{aligned}$$

This and the triangle inequality prove (3.33b). Moreover, inequality (3.35) together with $0 < q_{\mathrm{E}} < 1$ from energy contraction (3.31) proves that there exists $k_0 \in \mathbb{N}_0$, which is independent of $\ell$, such that

$$|||u_\ell^k||| \overset{(3.35)}{\leq} M + q_{\mathrm{E}}^k\, 3M \left(\frac{L[3M]}{\alpha}\right)^{1/2} \overset{!}{\leq} 2M \quad \text{for all } (\ell,k) \in \mathcal{Q} \text{ with } k \geq k_0. \tag{3.37}$$

This concludes the proof. □

**Remark 3.13.** (i) *From Lemma 3.12, 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 (3.31), we note that (3.33b) shows that* $\tau$ *provides a uniform upper bound for the involved stability and Lipschitz constants* $C_{\mathrm{stab}}[\tau]$ *and* $L[\tau]$*, respectively. Indeed, it will become apparent later that stability and local Lipschitz continuity will only be exploited for the differences* $|||u_H^k - u_H^{k-1}|||$, $|||u_H^\star - u_H^{k+1}|||$, *or* $|||u^\star - u_H^\star|||$ *in (A1), (3.9), and (3.21).*

### 3.2.9 Main results

Given the Pythagoras identity (3.8) and energy contraction (3.31), the first main theorem states full linear convergence of the quasi-error

$$\Delta_\ell^k := |||u^\star - u_\ell^k||| + \eta_\ell(u_\ell^k). \tag{3.38}$$

**Theorem 3.14: full linear convergence**

*Suppose that* $\mathcal{A}$ *satisfies (SM), (LIP), and (POT). Suppose the axioms of adaptivity (A1)–(A3) and orthogonality (3.8), where* $\mathcal{X}_H$ *is understood as* $\mathcal{X}_\ell$ *for* $(\ell,k) \in \mathcal{Q}$*. Let* $0 < \theta \leq 1$, $1 \leq C_{\mathrm{mark}} \leq \infty$, *and* $\lambda > 0$*. Suppose that the choice of* $\delta > 0$ *guarantees that Algorithm 3.10 satisfies energy contraction (3.31). Then, there exist* $C_{\mathrm{lin}} > 0$ *and* $0 < q_{\mathrm{lin}} < 1$ *such that Algorithm 3.10 leads to*

$$\Delta_\ell^k \leq C_{\mathrm{lin}} q_{\mathrm{lin}}^{|(\ell,k)|-|(\ell',k')|}\, \Delta_{\ell'}^{k'} \quad \text{for all } (\ell,k),(\ell',k') \in \mathcal{Q} \text{ with } (\ell',k') < (\ell,k). \tag{3.39}$$

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

The proof of Theorem 3.14 extends that of [GHPS21, Theorem 4], since the stopping criterion from Algorithm 3.10(i.b) requires further analysis to cover all cases. To ease notation, we introduce the shorthand

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

The following lemma provides the essential step in the proof of Theorem 3.14.

**Lemma 3.15.** *Under the assumptions of Theorem 3.14, there exist constants $\mu > 0$ and $0 < q_{\mathrm{lin}} < 1$ such that*

$$\Lambda_\ell^k := \mathrm{d}(u^\star, u_\ell^k)^2 + \mu\, \eta_\ell(u_\ell^k)^2 \quad \text{for all } (\ell, k) \in \mathcal{Q} \tag{3.40}$$

*satisfies the following statements* (i)–(ii)*:*

(i) $\Lambda_\ell^{k+1} \le q_{\mathrm{lin}}^2\, \Lambda_\ell^k$ *for all* $(\ell, k+1) \in \mathcal{Q}$.

(ii) $\Lambda_{\ell+1}^0 \le q_{\mathrm{lin}}^2\, \Lambda_\ell^{\underline{k}-1}$ *for all* $(\ell+1, 0) \in \mathcal{Q}$.

*The constants $\mu$ and $q_{\mathrm{lin}}$ depend only on $M$, $L[2M]$, $\alpha$, $C_{\mathrm{stab}}[\tau]$, $q_{\mathrm{red}}$, $C_{\mathrm{rel}}$, and $q_{\mathrm{E}}$ as well as on the adaptivity parameters $0 < \theta \le 1$ and $\lambda > 0$.*

*Proof.* For $k \in \mathbb{N}$ such that $1 \le k \le \underline{k}(\ell)$, the stopping criterion of Algorithm 3.10(i.b), i.e.,

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

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

$$|||u^\star - u_\ell^\star|||^2 \overset{(A3)}{\le} C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^\star)^2 \overset{(A1),(3.33a)}{\le} 2\, C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^{k+1})^2 + 2\, C_{\mathrm{rel}}^2\, C_{\mathrm{stab}}^2[\tau]\, |||u_\ell^\star - u_\ell^{k+1}|||^2.$$

Together with (3.9), this leads us to

$$\mathrm{d}(u^\star, u_\ell^\star)^2 \overset{(3.9)}{\le} \frac{L[2M]}{2} |||u^\star - u_\ell^\star|||^2 \overset{(3.9)}{\le} C_1\, \eta_\ell(u_\ell^{k+1})^2 + C_2\, \mathrm{d}(u_\ell^\star, u_\ell^{k+1})^2,$$

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

$$\begin{aligned}
\mathrm{d}(u^\star, u_\ell^{k+1})^2 &\overset{(3.8)}{=} (1-\varepsilon)\, \mathrm{d}(u^\star, u_\ell^\star)^2 + \varepsilon\, \mathrm{d}(u^\star, u_\ell^\star)^2 + \mathrm{d}(u_\ell^\star, u_\ell^{k+1})^2 \\
&\le (1-\varepsilon)\, \mathrm{d}(u^\star, u_\ell^\star)^2 + \varepsilon\, C_1\, \eta_\ell(u_\ell^{k+1})^2 + (1 + \varepsilon\, C_2)\, \mathrm{d}(u_\ell^\star, u_\ell^{k+1})^2 \\
&\overset{(3.31)}{\le} (1-\varepsilon)\, \mathrm{d}(u^\star, u_\ell^\star)^2 + \varepsilon\, C_1\, \eta_\ell(u_\ell^{k+1})^2 + (1 + \varepsilon\, C_2)\, q_{\mathrm{E}}^2\, \mathrm{d}(u_\ell^\star, u_\ell^k)^2.
\end{aligned}$$

We use the last inequality for the quasi-error $\Lambda_\ell^{k+1}$ to obtain that

$$\begin{aligned}
\Lambda_\ell^{k+1} &= \mathrm{d}(u^\star, u_\ell^{k+1})^2 + \mu\, \eta_\ell(u_\ell^{k+1})^2 \\
&\le (1-\varepsilon)\, \mathrm{d}(u^\star, u_\ell^\star)^2 + (\mu + \varepsilon\, C_1)\, \eta_\ell(u_\ell^{k+1})^2 + (1 + \varepsilon\, C_2)\, q_{\mathrm{E}}^2\, \mathrm{d}(u_\ell^\star, u_\ell^k)^2.
\end{aligned} \tag{3.41}$$

We need four auxiliary estimates:

First, since $|||u_\ell^\star||| \le M$ and $|||u_\ell^\star - u_0^\star||| \le 2M$ hold independently of $\ell$, the axioms (A1)–(A3) and Proposition 3.2 imply quasi-monotonicity of the estimators, i.e.,

$$\eta_\ell(u_\ell^\star) \le C_{\mathrm{mon}} \eta_0(u_0^\star) \quad \text{with} \quad C_{\mathrm{mon}} = \big(2 + 8 C_{\mathrm{stab}}[2M]^2 (1 + C_{\text{Céa}}^2) C_{\mathrm{rel}}^2\big)^{1/2}; \tag{3.42}$$

cf. [CFPP14, Lemma 3.6]. With $C_0 \coloneqq C_{\mathrm{mon}} \max\{1, C_{\mathrm{stab}}[M]M\}$, we infer that

$$\eta_\ell(u_\ell^\star) \overset{(3.42)}{\le} C_{\mathrm{mon}}\eta_0(u_0^\star) \overset{(A1)}{\le} C_{\mathrm{mon}}\eta_0(0) + C_{\mathrm{mon}}C_{\mathrm{stab}}[M]|||u_0^\star||| \le C_0(\eta_0(0)+1). \tag{3.43}$$

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

$$\begin{aligned}\eta_\ell(u_\ell^{k+1})^2 &\overset{(A1)}{\le} 2\,\eta_\ell(u_\ell^\star)^2 + 2\,C_{\mathrm{stab}}[\tau]^2\,|||u_\ell^\star - u_\ell^{k+1}|||^2 \overset{(3.9)}{\le} 2\,\eta_\ell(u_\ell^\star)^2 + \frac{4}{\alpha}\,C_{\mathrm{stab}}[\tau]^2\,\mathrm{d}(u_\ell^\star, u_\ell^{k+1})^2\\ &\overset{(3.31)}{\le} 2\,\eta_\ell(u_\ell^\star)^2 + \frac{4}{\alpha}\,C_{\mathrm{stab}}[\tau]^2\,q_{\mathrm{E}}^2\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2 \overset{(3.43)}{\le} C_3 + C_4\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2.\end{aligned} \tag{3.44}$$

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

$$\begin{aligned}\eta_\ell(u_\ell^{k+1})^2 &\overset{(A1)}{\le} (1+\gamma)\,\eta_\ell(u_\ell^k)^2 + (1+\gamma^{-1})\,C_{\mathrm{stab}}[\tau]^2\,|||u_\ell^{k+1} - u_\ell^k|||^2\\ &\le (1+\gamma)\,\eta_\ell(u_\ell^k)^2 + 2\,(1+\gamma^{-1})\,C_{\mathrm{stab}}[\tau]^2\,\big[|||u_\ell^\star - u_\ell^{k+1}|||^2 + |||u_\ell^\star - u_\ell^k|||^2\big]\\ &\overset{(3.9)}{\le} (1+\gamma)\,\eta_\ell(u_\ell^k)^2 + \frac{4}{\alpha}\,(1+\gamma^{-1})\,C_{\mathrm{stab}}[\tau]^2\,\big[\mathrm{d}(u_\ell^\star, u_\ell^{k+1})^2 + \mathrm{d}(u_\ell^\star, u_\ell^k)^2\big]\\ &\overset{(3.31)}{\le} (1+\gamma)\,\eta_\ell(u_\ell^k)^2 + C_5\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2,\end{aligned} \tag{3.45}$$

where $C_5 \coloneqq 4\,\alpha^{-1}(1+\gamma^{-1})\,C_{\mathrm{stab}}[\tau]^2\,(1+q_{\mathrm{E}}^2)$.

Fourth, we observe that the case $\texttt{true} \wedge \texttt{false}$ yields that

$$2M < |||u_\ell^{k+1}||| \le |||u_\ell^\star||| + |||u_\ell^\star - u_\ell^{k+1}||| \le M + |||u_\ell^\star - u_\ell^{k+1}|||$$

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

$$1 < \frac{|||u_\ell^\star - u_\ell^{k+1}|||^2}{M^2} \overset{(3.9)}{\le} 2\,\alpha^{-1}M^{-2}\,\mathrm{d}(u_\ell^\star, u_\ell^{k+1})^2 \overset{(3.31)}{\le} C_6\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2. \tag{3.46}$$

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

$$(1-\sigma)\,(1+\gamma) = 1 - \varepsilon. \tag{3.47}$$

We apply these observations to the term $(\mu + \varepsilon\,C_1)\,\eta_\ell(u_\ell^{k+1})^2$ of (3.41) to arrive at

$$\begin{aligned}&(\mu+\varepsilon\,C_1)\,\eta_\ell(u_\ell^{k+1})^2 = (1-\sigma)\,\mu\,\eta_\ell(u_\ell^{k+1})^2 + (\sigma\,\mu + \varepsilon\,C_1)\,\eta_\ell(u_\ell^{k+1})^2\\ &\overset{(3.45)}{\le} (1-\sigma)(1+\gamma)\,\mu\,\eta_\ell(u_\ell^k)^2 + (1-\sigma)\,\mu\,C_5\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2 + (\sigma\,\mu+\varepsilon\,C_1)\,\eta_\ell(u_\ell^{k+1})^2\\ &\overset{(3.44)}{\le} (1-\sigma)(1+\gamma)\,\mu\,\eta_\ell(u_\ell^k)^2 + (1-\sigma)\,\mu\,C_5\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2 + (\sigma\,\mu+\varepsilon\,C_1)\big[C_3 + C_4\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2\big]\\ &\overset{(3.46)}{<} (1-\sigma)(1+\gamma)\,\mu\,\eta_\ell(u_\ell^k)^2 + (1-\sigma)\,\mu\,C_5\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2 + (\sigma\,\mu+\varepsilon\,C_1)(C_3C_6 + C_4)\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2\\ &\overset{(3.47)}{=} (1-\varepsilon)\,\mu\,\eta_\ell(u_\ell^k)^2 + \big[(1-\sigma)\,C_5 + \sigma\,C_7\big]\,\mu\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2 + \varepsilon\,C_1\,C_7\,\mathrm{d}(u_\ell^\star, u_\ell^k)^2,\end{aligned} \tag{3.48}$$

where $C_7 := C_3 C_6 + C_4$. Together with (3.41), we obtain that

$$\begin{aligned}
\Lambda_\ell^{k+1} &\overset{(3.41)}{\leq} (1-\varepsilon)\,\mathbb{d}(u^\star, u_\ell^\star)^2 + (\mu + \varepsilon\,C_1)\,\eta_\ell(u_\ell^{k+1})^2 + (1+\varepsilon\,C_2)\,q_{\mathrm{E}}^2\,\mathbb{d}(u_\ell^\star, u_\ell^k)^2 \\
&\overset{(3.48)}{\leq} (1-\varepsilon)\,\mathbb{d}(u^\star, u_\ell^\star)^2 + (1-\varepsilon)\,\mu\,\eta_\ell(u_\ell^k)^2 \\
&\qquad + \big\{\big[(1-\sigma)\,C_5 + \sigma\,C_7\big]\mu + \varepsilon\,C_1\,C_7 + (1+\varepsilon\,C_2)\,q_{\mathrm{E}}^2\big\}\,\mathbb{d}(u_\ell^\star, u_\ell^k)^2 \\
&\leq (1-\varepsilon)\,\mathbb{d}(u^\star, u_\ell^\star)^2 + (1-\varepsilon)\,\mu\,\eta_\ell(u_\ell^k)^2 \\
&\qquad + \big\{\mu\,\max\{C_5, C_7\} + \varepsilon\,C_1\,C_7 + (1+\varepsilon\,C_2)\,q_{\mathrm{E}}^2\big\}\,\mathbb{d}(u_\ell^\star, u_\ell^k)^2.
\end{aligned}$$

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

$$\mu\,\max\{C_5, C_7\} + \varepsilon\,C_1\,C_7 + (1+\varepsilon\,C_2)\,q_{\mathrm{E}}^2 \leq 1-\varepsilon, \tag{3.49}$$

we conclude that

$$\begin{aligned}
\Lambda_\ell^{k+1} &\leq (1-\varepsilon)\,\big[\mathbb{d}(u^\star, u_\ell^\star)^2 + \mathbb{d}(u_\ell^\star, u_\ell^k)^2 + \mu\,\eta_\ell(u_\ell^k)^2\big] \\
&\overset{(3.8)}{=} (1-\varepsilon)\,\big[\mathbb{d}(u^\star, u_\ell^k)^2 + \mu\,\eta_\ell(u_\ell^k)^2\big] = (1-\varepsilon)\,\Lambda_\ell^k.
\end{aligned}$$

**Case 2: Evaluation of (i.b) returns `false` ∧ `false` or `false` ∧ `true`.** These cases follow from the arguments found in [GHPS21, Lemma 10(i)]. There, the proof is based in essence on the estimate

$$\eta_\ell(u_\ell^{k+1})^2 < \lambda^{-2}\,\mathbb{d}(u_\ell^{k+1}, u_\ell^k)^2 \overset{(3.31)}{\leq} \lambda^{-2}\,(1+q_{\mathrm{E}}^2)\,\mathbb{d}(u_\ell^\star, u_\ell^k)^2,$$

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

$$(\mu + \varepsilon\,C_1)\,C_8 + (1+\varepsilon\,C_2)\,q_{\mathrm{E}}^2 \leq 1-\varepsilon, \tag{3.50}$$

[GHPS21, 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` ∧ `true`.** The case `true` ∧ `true` is analyzed in [GHPS21, Lemma 10(ii)] and is based on the contractivity of the error estimator given that the Dörfler marking is employed.

Define $q_\theta := \big(1 - (1-q_{\mathrm{red}}^2)\,\theta\big)$ and $C_9 := 4\,\alpha^{-1}(1+q_{\mathrm{E}}^2)\,C_{\mathrm{stab}}[\tau]^2$. 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

$$\varepsilon\,C_1\,\mu^{-1} + q_\theta\,(1+\delta) \leq 1-\varepsilon \quad \text{and} \quad \varepsilon\,C_2 + q_{\mathrm{E}}^2 + \mu\,q_\theta\,(1+\omega^{-1})\,C_9 \leq 1-\varepsilon, \tag{3.51}$$

we obtain from [GHPS21, 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 3.15(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 $q_{\mathrm{E}}^2 + \mu \max\{C_5, C_7\} < 1$, $q_{\mathrm{E}}^2 + \mu C_8 < 1$, and $q_{\mathrm{E}}^2 + \mu\, q_\theta (1+\omega)^{-1} C_9 < 1$.
3. Finally, choose $\varepsilon > 0$ sufficiently small such that (3.49)–(3.51) are satisfied.

This concludes the proof of Lemma 3.15 with $q_{\mathrm{lin}}^2 := (1-\varepsilon)$. □

*Proof of Theorem 3.14.* According to (3.9), it holds that $\Delta_\ell^k \simeq (\Lambda_\ell^k)^{1/2}$, where the hidden constants depend only on $\mu$, $\alpha$, and $L[\tau/2]$. We use (3.9) for the term $|||u_\ell^\star - u_\ell^k|||$, and hence the dependency $L[\tau/2]$ is justified by (3.36). Then, linear convergence (3.39) follows from Lemma 3.15 and induction, since the set $Q$ is linearly ordered with respect to the total step counter $|(\cdot,\cdot)|$. □

**Remark 3.16.** (i) *Provided that energy contraction* (3.31) *holds and that the adaptivity parameter $\lambda > 0$ is sufficiently small, the stopping criterion*

$$|\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \le \lambda^2\, \eta_\ell(u_\ell^k)^2 \tag{i.b$'$}$$

*from [GHPS21] is a viable alternative to the stopping criterion of Algorithm 3.10*(i.b)*. The main difficulty is to ensure nested iteration* (3.28)*. This relies, in essence, on the estimate*

$$\begin{aligned}\frac{\alpha}{2}\, |||u_\ell^\star - u_\ell^k|||^2 &\overset{(3.9)}{\le} \mathcal{E}(u_\ell^k) - \mathcal{E}(u_\ell^\star) \overset{(3.31),\,(3.24)}{\le} \frac{q_{\mathrm{E}}[\delta]^2}{1-q_{\mathrm{E}}[\delta]^2}\, \big[\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)\big] \\ &\overset{(i.b')}{\le} \frac{q_{\mathrm{E}}[\delta]^2}{1-q_{\mathrm{E}}[\delta]^2}\, \lambda^2\, \eta_\ell(u_\ell^k)^2 \overset{(A1)}{\le} 2\, \frac{q_{\mathrm{E}}[\delta]^2}{1-q_{\mathrm{E}}[\delta]^2}\, \lambda^2 \big[\eta_\ell(u_\ell^\star)^2 + C_{\mathrm{stab}}[\tau/2]^2 |||u_\ell^\star - u_\ell^k|||^2\big],\end{aligned}$$

*where $|||u_\ell^\star - u_\ell^k||| \le \tau/2$ stems from* (3.36)*. Using a uniform estimate for the error estimator as in* (3.43)*, the last estimate, and the observation that $|||u_\ell^k||| \le M + |||u_\ell^\star - u_\ell^k|||$ lead us to*

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

*where a sufficiently small $\lambda$ such that $\lambda^2\, r[\delta]^2\, C_{\mathrm{stab}}[\tau/2]^2 < 1$ is required and where $C_0 := C_{\mathrm{mon}} \max\{1, C_{\mathrm{stab}}[M]\, M\}$. We see that a sufficiently small $\lambda > 0$ ensures nested iteration* (3.28)*. In contrast,* (i.b) *leads to full linear convergence for arbitrary $\lambda > 0$.*

(ii) *Theorem 3.14 proves linear convergence, and hence in particular plain convergence $\Delta_\ell^k \to 0$ as $|(\ell, k)| \to \infty$. In Appendix 3.6, it is shown that plain convergence also holds for Algorithm 3.10 with the modified stopping criterion*

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

*(instead of Algorithm 3.10*(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* (3.30) *instead of energy contraction* (3.31)*.*

To formulate our main result on optimal convergence rates, we need some additional notation. For $N \in \mathbb{N}_0$, let $\mathbb{T}_N := \{\mathcal{T} \in \mathbb{T} \mid \#\mathcal{T} - \#\mathcal{T}_0 \le N\}$ 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

$$\|u^\star\|_{\mathbb{A}_s} := \sup_{N\in\mathbb{N}_0} \Big( (N+1)^s \min_{\mathcal{T}_{\rm opt}\in\mathbb{T}_N} \big[ |||u^\star - u^\star_{\rm opt}||| + \eta_{\rm opt}(u^\star_{\rm opt}) \big] \Big) \in \mathbb{R}_{\ge 0} \cup \{\infty\}. \tag{3.52}$$

Here, $u^\star_{\rm opt} \in \mathcal{X}_{\rm opt}$ denotes the exact Galerkin solution (3.4) with respect to the optimal mesh $\mathcal{T}_{\rm opt}$, where optimality is understood with respect to the quasi-error $\Delta^\star_{\rm opt}$ from (3.38) (consisting of the energy norm error plus error estimator). In explicit terms, $\|u^\star\|_{\mathbb{A}_s} < \infty$ means that an algebraic convergence rate $O(N^{-s})$ for the quasi-error $\Delta^\star_{\rm opt}$ is possible, if the optimal triangulations are chosen.

The second main theorem states optimal convergence rates of the quasi-error (3.38) with respect to the number of degrees of freedom. As usual in this context (see, e.g., [CFPP14]), the result requires that the adaptivity parameters $0 < \theta \le 1$ and $\lambda > 0$ are sufficiently small. The proof is found in, e.g., [GHPS21, Theorem 8]. A careful inspection of the proof reveals that it requires only estimates of the form

$$\mathrm{d}(u^k_\ell, u^{k-1}_\ell) \le \lambda\, \eta_\ell(u^k_\ell),$$

as well as linear convergence (3.39), which are satisfied for Algorithm 3.10. The results from [GHPS21] are proven for a uniform Lipschitz and stability constant; in the present setting, this follows from Remark 3.13(ii).

**Theorem 3.17: 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 3.10 satisfies energy contraction (3.31). Define*

$$\lambda_{\rm opt} := \frac{1 - q_{\rm E}}{q_{\rm E}\, C_{\rm stab}[\tau]} \Big(\frac{\alpha}{2}\Big)^{1/2}, \tag{3.53}$$

*with $\tau$ from (3.33). Let $0 < \theta \le 1$ and $0 < \lambda < \lambda_{\rm opt}\theta$ such that*

$$0 < \theta' := \frac{\theta + \lambda/\lambda_{\rm opt}}{1 - \lambda/\lambda_{\rm opt}} < (1 + C_{\rm stab}[\tau]^2 C_{\rm rel}^2)^{-1/2}. \tag{3.54}$$

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

$$c_{\rm opt}^{-1}\, \|u^\star\|_{\mathbb{A}_s} \le \sup_{(\ell,k)\in\mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s\, \Delta^k_\ell \le C_{\rm opt}\, \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta^0_0\}, \tag{3.55}$$

*where $\|u^\star\|_{\mathbb{A}_s}$ is defined in (3.52). The constant $c_{\rm opt} > 0$ depends only on $C_{\rm Céa} = L[2M]/\alpha$, fine properties of NVB refinement, $C_{\rm stab}[\tau]$, $C_{\rm rel}$, $\#\mathcal{T}_0$, and $s$, and additionally on $\underline{\ell}$ or $\ell_0$, if $\underline{\ell} < \infty$ or $\eta_{\ell_0}(u^k_{\ell_0}) = 0$ for some $(\ell_0, 0) \in \mathcal{Q}$, respectively. The constant $C_{\rm opt} > 0$ depends only on fine properties of NVB refinement, $\alpha$, $C_{\rm stab}[\tau]$, $q_{\rm red}$, $C_{\rm rel}$, $C_{\rm drel}$, $1 - \lambda/\lambda_{\rm opt}$ (and hence on*

*energy contraction $q_{\mathrm{E}}$), $C_{\mathrm{mark}}$, $C_{\mathrm{lin}}$, $q_{\mathrm{lin}}$, 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 3.10(iii) can be performed at linear cost $\mathcal{O}(\#\mathcal{T}_\ell)$ (cf. [Ste07], or the algorithm from [PP20] providing $\mathcal{M}_\ell$ with minimal cardinality);
- We have linear cost $\mathcal{O}(\#\mathcal{T}_\ell)$ to generate the new mesh $\mathcal{T}_{\ell+1}$ in Algorithm 3.10(iv).

In addition, we make the following "idealized" assumption, but refer to Remark 3.18(ii):

- The solutions $u_\ell^k \in \mathcal{X}_\ell$ of the linearized problems in Algorithm 3.10(i.a) can be computed in linear complexity $\mathcal{O}(\#\mathcal{T}_\ell)$.

Since a step $(\ell, k) \in \mathcal{Q}$ of Algorithm 3.10 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

$$\mathtt{work}(\ell, k) := \sum_{\substack{(\ell', k') \in \mathcal{Q} \\ (\ell', k') \le (\ell, k)}} \#\mathcal{T}_{\ell'} \quad \text{for all } (\ell, k) \in \mathcal{Q}. \tag{3.56}$$

**Remark 3.18.** (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 3.10*(i.a) *is linear, can be avoided with an extended algorithm (and refined analysis) in the spirit of [HPSV21]. 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 [HPSV21] 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 3.19** (rate-optimality w.r.t. computational cost)**.** *Let $(\mathcal{T}_\ell)_{\ell \in \mathbb{N}_0}$ be the sequence generated by Algorithm 3.10. Suppose full linear convergence* (3.39) *with respect to the quasi-error $\Delta_\ell^k$ from* (3.38)*. Then, for all $s > 0$, it holds that*

$$C_{\mathrm{rate}} := \sup_{(\ell,k) \in \mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \, \Delta_\ell^k \le \sup_{(\ell,k) \in \mathcal{Q}} \mathtt{work}(\ell, k)^s \, \Delta_\ell^k \le \frac{(\#\mathcal{T}_0)^s \, C_{\mathrm{lin}}}{\big(1 - q_{\mathrm{lin}}^{1/s}\big)^s} \, C_{\mathrm{rate}}. \tag{3.57}$$

*Consequently, rate-optimality with respect to the number of elements* (3.55) *yields that*

$$c_{\mathrm{opt}}^{-1}\,\|u^\star\|_{\mathbb{A}_s} \le \sup_{(\ell,k)\in\mathcal{Q}} \mathtt{work}(\ell,k)^s\,\Delta_\ell^k \le C_{\mathrm{opt}}\,\frac{(\#\mathcal{T}_0)^s\,C_{\mathrm{lin}}}{\big(1-q_{\mathrm{lin}}^{1/s}\big)^s}\,\max\{\|u^\star\|_{\mathbb{A}_s},\Delta_0^0\}. \tag{3.58}$$

*Proof.* The first inequality in (3.57) is obvious. To obtain the upper bound, let $(\ell,k)\in\mathcal{Q}$. Elementary calculus (see [BHP17, Lemma 22]) proves that

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

Moreover, linear convergence (3.39) and the geometric series lead us to

$$\sum_{\substack{(\ell',k')\in\mathcal{Q}\\(\ell',k')\le(\ell,k)}} (\Delta_{\ell'}^{k'})^{-1/s} \overset{(3.39)}{\le} C_{\mathrm{lin}}^{1/s}\,(\Delta_\ell^k)^{-1/s} \sum_{\substack{(\ell',k')\in\mathcal{Q}\\(\ell',k')\le(\ell,k)}} (q_{\mathrm{lin}}^{1/s})^{|(\ell,k)|-|(\ell',k')|} \le \frac{C_{\mathrm{lin}}^{1/s}\,(\Delta_\ell^k)^{-1/s}}{1-q_{\mathrm{lin}}^{1/s}}.$$

Combining the last two inequalities, we obtain that

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

Rearranging this estimate, we obtain the upper bound in (3.57). □

## 3.3 Semilinear model problem

### 3.3.1 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 $\boldsymbol{f}\in[L^2(\Omega)]^d$, we aim to approximate the weak solution $u^\star\in\mathcal{X}:=H_0^1(\Omega)$ of the semilinear elliptic PDE

$$-\operatorname{div}(\boldsymbol{A}\nabla u^\star) + b(u^\star) = f - \operatorname{div}\boldsymbol{f} \text{ in } \Omega \quad \text{subject to} \quad u^\star = 0 \text{ on } \partial\Omega. \tag{3.59}$$

While the precise assumptions on the coefficients $\boldsymbol{A}\colon\Omega\to\mathbb{R}^{d\times d}_{\mathrm{sym}}$ and $b\colon\Omega\times\mathbb{R}\to\mathbb{R}$ are given in Section 3.3.3–3.3.4, we note that, here and below, we abbreviate $\boldsymbol{A}\nabla u^\star \equiv \boldsymbol{A}(\cdot)\nabla u^\star(\cdot)\colon\Omega\to\mathbb{R}^d$ and $b(u^\star)\equiv b(\cdot,u^\star(\cdot))\colon\Omega\to\mathbb{R}$.

Let $\langle\cdot\,,\cdot\rangle_\Omega$ denote the $L^2(\Omega)$-scalar product $\langle v\,,w\rangle_\Omega := \int_\Omega vw\,\mathrm{d}x$ and let $\langle\!\langle v\,,w\rangle\!\rangle := \langle\boldsymbol{A}\nabla v\,,\nabla w\rangle_\Omega$ be the $\boldsymbol{A}$-induced energy scalar product on $H_0^1(\Omega)$. Then, the weak formulation of (3.59) reads as follows: Find $u^\star\in H_0^1(\Omega)$ such that

$$\langle\mathcal{A}u^\star\,,v\rangle := \langle\!\langle u^\star\,,v\rangle\!\rangle + \langle b(u^\star)\,,v\rangle_\Omega = \langle f\,,v\rangle_\Omega + \langle\boldsymbol{f}\,,\nabla v\rangle_\Omega =: \langle F\,,v\rangle \text{ for all } v\in H_0^1(\Omega). \tag{3.60}$$

Existence and uniqueness of the solution $u^\star \in H^1_0(\Omega)$ of (3.60) follow from the Browder–Minty theorem on monotone operators (see Section 3.3.6 for details).

Based on conforming triangulations $\mathcal{T}_H$ of $\Omega$ and fixed polynomial degree $m \in \mathbb{N}$, let $\mathcal{X}_H := \{v_H \in H^1_0(\Omega) \mid \forall T \in \mathcal{T}_H\colon\ v_H|_T \text{ is a polynomial of degree } \le m\}$. Then, the FEM discretization of (3.60) reads: Find $u^\star_H \in \mathcal{X}_H$ such that

$$\langle\!\langle u^\star_H\,,\,v_H\rangle\!\rangle + \langle b(u^\star_H)\,,\,v_H\rangle_\Omega = \langle F\,,\,v_H\rangle \quad \text{for all } v_H \in \mathcal{X}_H. \tag{3.61}$$

The FEM solution $u^\star_H$ approximates the sought exact solution $u^\star$.

### 3.3.2 General notation

For $1 \le p \le \infty$, let $1 \le p' \le \infty$ be the conjugate Hölder index which ensures that $\|\phi\,\psi\|_{L^1(\Omega)} \le \|\phi\|_{L^p(\Omega)}\|\psi\|_{L^{p'}(\Omega)}$ 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 $1 \le p < d$, let $1 \le p^* := dp/(d-p) < \infty$ denote the critical Sobolev exponent of $p$ in dimension $d \in \mathbb{N}$. We recall the Gagliardo–Nirenberg–Sobolev inequality (see, e.g., [FK80, Theorem 16.6])

$$\|v\|_{L^r(\Omega)} \le C_{\mathrm{GNS}}\,\|\nabla v\|_{L^p(\Omega)} \quad \text{for all } v \in W^{1,p}_0(\Omega) \tag{3.62}$$

with a constant $C_{\mathrm{GNS}} = C_{\mathrm{GNS}}(|\Omega|, d, p, r)$. With $\mathcal{X} = H^1_0(\Omega)$, we restrict to $p = 2$. If $d \in \{1,2\}$, (3.62) holds for any $1 \le r < \infty$. If $d = 3$, (3.62) holds for all $1 \le r \le p^* = 6$, where $r = p^*$ is the largest possible exponent such that the embedding $W^{1,p}(\Omega) \hookrightarrow L^r(\Omega)$ is continuous.

### 3.3.3 Assumptions on diffusion coefficient

The diffusion coefficient $\boldsymbol{A}\colon \Omega \to \mathbb{R}^{d\times d}_{\mathrm{sym}}$ satisfies the following standard assumptions:

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

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

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

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.

### 3.3.4 Assumptions on the nonlinear reaction coefficient

The nonlinearity $b\colon \Omega\times\mathbb{R} \to \mathbb{R}$ satisfies the following assumptions, which follow [BHSZ11, (A1)–(A3)]:

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

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

(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 generality[2], 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., [FK80, Chapter III, (12)] or [BHSZ11, (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

$$|b^{(N)}(x,\xi)| \leq R \quad \text{for a.e. } x \in \Omega \text{ and all } \xi \in \mathbb{R}. \tag{3.63}$$

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 3.20.** (i) Let $v, w \in H^1_0(\Omega)$. 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

$$|\langle b(v)\,,\, w\rangle_\Omega| \leq \|b(v)\|_{L^{s'}(\Omega)} \|w\|_{L^s(\Omega)}. \tag{3.64}$$

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

$$b(v) \overset{\text{(MON)}}{=} \sum_{n=1}^{N-1} \frac{b^{(n)}(0)}{n!}\, v^n + \left( \int_0^1 \frac{(1-\xi)^{N-1}}{(N-1)!}\, b^{(N)}(\xi v)\, \mathrm{d}\xi \right) v^N. \tag{3.65}$$

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

$$\begin{aligned} \|b(v)\|_{L^{s'}(\Omega)} &\overset{\text{(GC)}}{\lesssim} \sum_{n=1}^{N-1} \|v^n\|_{L^{s'}(\Omega)} + \|v^N\|_{L^{s'}(\Omega)} = \sum_{n=1}^{N-1} \|v\|^n_{L^{ns'}(\Omega)} + \|v\|^N_{L^{Ns'}(\Omega)} \\ &\lesssim \sum_{n=1}^{N} \|v\|^n_{L^{Ns'}(\Omega)} \leq N \max\{1, \|v\|^{N-1}_{L^{Ns'}(\Omega)}\}\, \|v\|_{L^{Ns'}(\Omega)}, \end{aligned}$$

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

$$H^1_0(\Omega) \hookrightarrow L^r(\Omega) \quad \text{is continuous} \quad \text{for} \quad r = s \quad \text{and} \quad r = Ns'. \tag{3.66}$$

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

If $d \in \{1,2\}$, (3.66) follows if $1 \le r < \infty$ and hence arbitrary $1 < s < \infty$ and $N \in \mathbb{N}$. If $d = 3$, $r = s = 2^* = 6$ is the maximal index in (3.66). Hence, it follows that $N \le 2^*/s' = 2^*/2^{*\prime} = 2^* - 1 = 5$. 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 $N \le 5$ if $d = 3$.

(ii) The definition of [①GOA, (GC)] uses

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

instead of (3.63). However, the following observation replaces the estimates for all $b^{(n)}$ with $0 \le n < N$. Due to the smoothness assumption (CAR), we may apply a Taylor expansion for an admissible $\sigma$ such that $(N-n)\,\sigma < \infty$ if $d = 1,2$ and $(N-n)\,\sigma \le 6$ if $d = 3$. Together with $\|v^n\|_{L^\sigma(\Omega)} = \|v\|^n_{L^{n\sigma}(\Omega)}$, this leads us to

$$\begin{aligned}
\|b^{(n)}(v)\|_{L^\sigma(\Omega)} &\le \sum_{j=n}^{N-1} \frac{b^{(j)}(0)}{(j-n)!}\,\|v^{j-n}\|_{L^\sigma(\Omega)} + \left(\int_0^1 \frac{(1-\xi)^{N-1-n}}{(N-1-n)!}\,b^{(N)}(\xi v)\,\mathrm{d}\xi\right)\|v^{N-n}\|_{L^\sigma(\Omega)} \\
&\overset{\text{(GC)}}{\lesssim} \sum_{j=n}^{N-1} \|v\|^{j-n}_{L^{(j-n)\sigma}(\Omega)} + \|v\|^{N-n}_{L^{(N-n)\sigma}(\Omega)} \lesssim \sum_{j=n}^{N} \|v\|^{j-n}_{L^{(N-n)\sigma}(\Omega)} \\
&\le (N-n)\,\big(1 + \|v\|^{N-n}_{L^{(N-n)\sigma}(\Omega)}\big) \lesssim (N-n)\,(1+|||v|||^{N-n}),
\end{aligned} \tag{3.67}$$

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 3.21 below and the discussion thereafter. □

### 3.3.5 Assumptions on the right-hand sides

For $d = 1$, the exact solution $u^\star$ from (3.60) below satisfies an $L^\infty$-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 fulfills that

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

To guarantee later that the residual *a posteriori* error estimator from (3.74) is well-defined, we additionally require that $\boldsymbol{f}|_T \in H(\mathrm{div},T)$ and $\boldsymbol{f}|_T \cdot \boldsymbol{n} \in L^2(\partial T)$ for all $T \in \mathcal{T}_0$, where $\mathcal{T}_0$ is the initial triangulation of the adaptive algorithm.

### 3.3.6 Well-posedness and applicability of abstract framework

Let $v, w \in H^1_0(\Omega)$. We consider the operator $\mathcal{A}$, where $H^{-1}(\Omega) := H^1_0(\Omega)'$ is used to denote the dual space of $H^1_0(\Omega)$,

$$\mathcal{A}\colon H^1_0(\Omega) \to H^{-1}(\Omega), \quad \mathcal{A}w := \langle\!\langle w\,,\,\cdot\rangle\!\rangle + \langle b(w)\,,\,\cdot\rangle_\Omega. \tag{3.68}$$

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

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

Together with (ELL) and for $v, w \in H^1_0(\Omega)$, we thus see that

$$\langle \mathcal{A}w - \mathcal{A}v \,,\, w - v\rangle = \langle\!\langle w - v \,,\, w - v\rangle\!\rangle + \langle b(w) - b(v) \,,\, w - v\rangle_\Omega \geq |||w - v|||^2. \tag{3.69}$$

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 3.21.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (GC). *Let* $\vartheta > 0$ *and let* $v, w \in H^1_0(\Omega)$ *with* $\max\{|||w|||, |||w - v|||\} \leq \vartheta < \infty$. *Then, it holds that*

$$\langle b(w) - b(v) \,,\, z\rangle_\Omega \leq \widetilde{L}[\vartheta]\, |||w - v|||\,|||z||| \quad \textit{for all } z \in H^1_0(\Omega) \tag{3.70}$$

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

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

$$b(v) = \sum_{n=0}^{N-1} b^{(n)}(w)\frac{(v-w)^n}{n!} + \frac{(v-w)^N}{(N-1)!}\int_0^1 (1-\xi)^{N-1}\, b^{(N)}\big(w + (v-w)\,\xi\big)\,\mathrm{d}\xi. \tag{3.71}$$

In order to apply the generalized Hölder inequality for three terms $\phi, \varphi, \psi \in H^1_0(\Omega)$

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

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

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

where the hidden constant depends on $R$ from (GC). Since $H^1_0(\Omega) \hookrightarrow L^{(N-1)t''}(\Omega)$ for $d \in \{1,2,3\}$, it remains to prove that

$$\|b^{(n)}(w)(w-v)^{n-1}\|_{L^{t''}(\Omega)} \leq C[\vartheta] \quad \text{for all } n = 1, \dots, N-1. \tag{3.72}$$

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

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

Using the Hölder inequality, we arrive at

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

Since $\|\varphi^j\|_{L^\sigma(\Omega)} = \|\varphi\|^j_{L^{j\sigma}(\Omega)}$ and $(N-1)t'' < \infty$ if $d \in \{1,2\}$ and $(N-1)t'' \le 6$ if $d = 3$ guarantee admissibility as in Remark 3.20(ii), we apply the Sobolev embedding to obtain that

$$\|b^{(n)}(w)\|_{L^{t_1}(\Omega)} \overset{(3.67)}{\lesssim} 1 + \|w\|^{N-n}_{L^{(N-n)t_1}(\Omega)} = 1 + \|w\|^{N-n}_{L^{(N-1)t''}(\Omega)} \lesssim 1 + |||w|||^{N-n}$$

and

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

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

To see the local Lipschitz continuity of $\mathcal{A}$, let $v, w, \psi \in H^1_0(\Omega)$ and observe that

$$\langle \mathcal{A}w - \mathcal{A}v\,,\,\psi\rangle = \langle\!\langle w-v\,,\,\psi\rangle\!\rangle + \langle b(w)-b(v)\,,\,\psi\rangle_\Omega \overset{(3.70)}{\le} (1+\widetilde{L}[\vartheta])\,|||w-v|||\,|||\psi|||,$$

provided that $|||w||| \le \vartheta$ and $|||w-v||| \le \vartheta$. This shows that $\mathcal{A}$ is locally Lipschitz continuous with Lipschitz constant $L[\vartheta] := 1 + \widetilde{L}[\vartheta]$. Hence, $\mathcal{A}$ fits into the abstract setting of Section 3.2.

Furthermore, following [AHW23], we note that the energy for the semilinear model problem (3.59) of Section 3.3 for $v \in H^1_0(\Omega)$ is given by

$$\mathcal{E}(v) = \frac{1}{2}\int_\Omega |\boldsymbol{A}^{1/2}\nabla v|^2\,\mathrm{d}x + \int_\Omega \int_0^{v(x)} b(s)\,\mathrm{d}s\,\mathrm{d}x - \int_\Omega fv\,\mathrm{d}x - \int_\Omega \boldsymbol{f}\cdot\nabla v\,\mathrm{d}x. \tag{3.73}$$

To see that the second integral is well-defined, note that the integration of the Taylor expansion (3.65) gives rise to a term $s^{N+1}$ evaluated at $s = v(x)$ and $s = 0$. Its integrability $\|v^{N+1}\|_{L^1(\Omega)} = \|v\|^{N+1}_{L^{(N+1)}(\Omega)} < \infty$ is ensured by (CGC).

### 3.3.7 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 (3.60) read

$$\begin{aligned}\eta_H(T, v_H)^2 := h_T^2\,\|f + \operatorname{div}(\boldsymbol{A}\nabla v_H - \boldsymbol{f}) - b(v_H)\|^2_{L^2(T)} \\ + h_T\,\|[\![(\boldsymbol{A}\nabla v_H - \boldsymbol{f})\cdot\boldsymbol{n}]\!]\|^2_{L^2(\partial T\cap\Omega)},\end{aligned} \tag{3.74}$$

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

**Proposition 3.22** ([①GOA, Proposition 15]). *Suppose* (RHS), (ELL), (CAR), (MON), *and* (CGC). *Then, the residual error estimator from* (3.74) *satisfies* (A1)–(A4) *from Section 3.2.6. The constant* $C_{\mathrm{rel}}$ *depends only on* $d$, $\mu_0$, *and uniform shape regularity of the meshes* $\mathcal{T}_H \in \mathbb{T}$. *The constant* $C_{\mathrm{drel}}$ *depends, in addition, on the polynomial degree* $m$, *and* $C_{\mathrm{stab}}[\vartheta]$ *depends furthermore on* $|\Omega|$, $\vartheta$, $n$, $R$, *and* $\boldsymbol{A}$. □

## 3.4 Practical algorithm

For the semilinear problem (3.59) of Section 3.3, it holds that $\alpha = 1$ according to (3.69). 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.

### 3.4.1 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 3.9.

**Algorithm 3.23: practical AILFEM**

**Input:** initial triangulation $\mathcal{T}_0$, initial guess $u_0^0 := 0$ and $M = \|F - \mathcal{A}0\|_{X'} < \infty$ according to (3.5), marking parameters $0 < \theta \le 1$ and $C_{\mathrm{mark}} \ge 1$, 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 $\delta_\ell \leftarrow\!\shortmid 1/L_\ell$ and $q_\ell^2 \leftarrow\!\shortmid 1 - \delta_\ell^2$.

(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 $\big(|\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^k)| \le \lambda^2\, \eta_\ell(u_\ell^k)^2 \quad \wedge \quad |||u_\ell^k||| \le 2M\big)$.

(c) If $\big(\mathcal{E}(u_\ell^k) > q_\ell^2\, \mathcal{E}(u_\ell^{k-1})\big)$, then

(c1) Discard the computed $u_\ell^k$ and set $k \leftarrow\!\shortmid k - 1$.

(c2) Increase $L_\ell \leftarrow\!\shortmid \beta\, L_\ell$.

(c3) Update $\delta_\ell \leftarrow\!\shortmid 1/L_\ell$ and $q_\ell^2 \leftarrow\!\shortmid 1 - \delta_\ell^2$.

(iii) Upon termination of the $k$-loop, define $\underline{k}(\ell) := k$.

(iv) Determine $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ with $\theta\, \eta_\ell(u_\ell^{\underline{k}(\ell)})^2 \le \sum_{T \in \mathcal{M}_\ell} \eta_\ell(T, u_\ell^{\underline{k}(\ell)})^2$.

(v) Generate $\mathcal{T}_{\ell+1} := \mathtt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ and define $u_{\ell+1}^0 := u_\ell^{\underline{k}(\ell)}$.

**Remark 3.24.** *The motivation of the criterion in Algorithm 3.23*(ii.c) *is based on the equivalence*

$$\mathcal{E}(u_\ell^k) - \mathcal{E}(u_\ell^\star) \le q_\ell^2 \big[\mathcal{E}(u_\ell^{k-1}) - \mathcal{E}(u_\ell^\star)\big] \quad \Longleftrightarrow \quad \mathcal{E}(u_\ell^k) - q_\ell^2 \mathcal{E}(u_\ell^{k-1}) \le (1 - q_\ell^2)\, \mathcal{E}(u_\ell^\star). \tag{3.75}$$

*The energy minimization property from Lemma 3.3 and $b(0)=0$ from (MON) show that $\mathcal{E}(u_\ell^\star)\le\mathcal{E}(0)=0$; cf. (3.73). As a necessary criterion for energy contraction (3.31), we thus obtain $\mathcal{E}(u_\ell^k)\le q_\ell^2\mathcal{E}(u_\ell^{k-1})$, which is enforced by Algorithm 3.23(ii.c).*

**Remark 3.25.** *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 $\eta_\ell(u_\ell^\star)$ and hence approximately $\eta_\ell(u_\ell^k)$ controls the discretization error, while $|||u_\ell^\star-u_\ell^k|||^2 \overset{(3.9)}{\simeq} \mathcal{E}(u_\ell^k)-\mathcal{E}(u_\ell^\star) \overset{(3.24)}{\lesssim} \mathcal{E}(u_\ell^{k-1})-\mathcal{E}(u_\ell^k) \overset{(3.21)}{\simeq} |||u_\ell^k-u_\ell^{k-1}|||^2$ controls the linearization error — at least if $\delta_H$ is sufficiently small. Hence, $\mathcal{E}(u_\ell^{k-1})-\mathcal{E}(u_\ell^k)\le 0.1^2\,\eta_\ell(u_\ell^k)^2$ heuristically aims at limiting the linearization error to be at most 10% of the current discretization error.*

The next result states that Algorithm 3.23(ii.c) will not lead to an infinite loop.

**Proposition 3.26.** *Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and (POT). Let $u_H^0\in\mathcal{X}_H$ with $|||u_H^0|||\le 2M$. 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 $k\ge 1$, compute $u_H^{k+1}:=\Phi_H(\delta_H;u_H^k)\in\mathcal{X}_H$ and check if*

$$\mathcal{E}(u_H^{k+1})\le q_H^2\,\mathcal{E}(u_H^k). \tag{3.76}$$

- *If (3.76) holds, then increase $k\leftarrow k+1$.*
- *If (3.76) 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 (3.76) 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 $q_{\mathrm{E}}[\delta_H]$ from (3.23b) and observe that

$$q_{\mathrm{E}}[\delta_H(k)]^2=1-\big(1-\frac{\delta_H(k)L[6M]}{2}\big)\,\frac{2\delta_H(k)\alpha^2}{L[3M]}\simeq 1-\delta_H(k)+\delta_H(k)^2 \quad\text{for } \delta_H(k)\to 0.$$

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

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

This implies that Proposition 3.8 holds for the theoretical sequence $\widetilde{u}_H^0:=u_H^{k_0}$ and $\widetilde{u}_H^{k+1}:=\Phi_H(\delta_H;\widetilde{u}_H^k)$. In particular, we conclude that energy contraction (3.31) holds with $q_H^2=1-\delta_H^2$. Moreover, Remark 3.24 shows that the necessary criterion (3.76) is guaranteed to hold for the iterates $(\widetilde{u}_H^k)_{k\in\mathbb{N}_0}$ as soon as (3.31) holds.

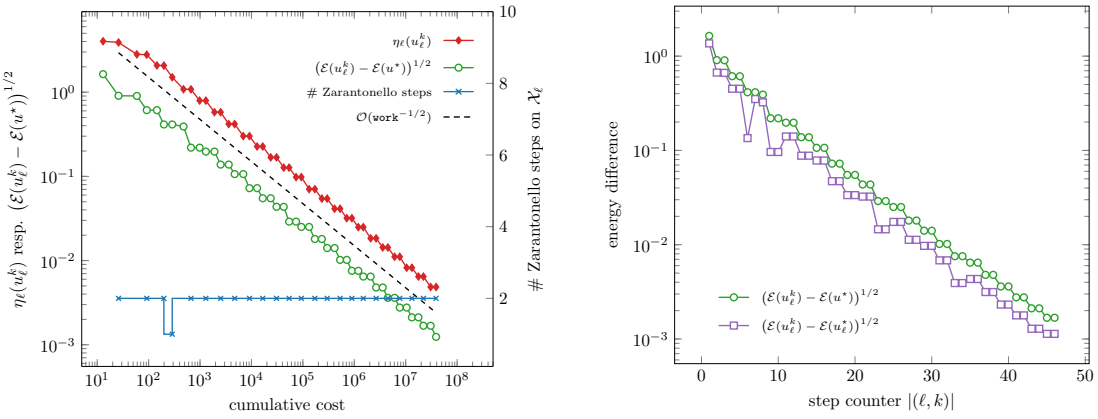



**Figure 3.1:** Results of Experiment 3.28 with polynomial degree $m = 1$. Left: Error estimator $\eta_\ell(u_\ell^k)$ (diamond, left ordinate) and energy difference of iterative solutions $\big(\mathcal{E}(u_\ell^k)-\mathcal{E}(u^\star)\big)^{1/2}$ (circle, left ordinate) against $\mathtt{work}(\ell,k)$ and the number of Zarantonello steps on $\mathcal{X}_\ell$ (cross, right ordinate). Right: Energy difference of $\mathcal{E}(u_\ell^k)$ to $\mathcal{E}(u^\star)$ (circle) and to $\mathcal{E}(u_\ell^\star)$ (square) over the total step counter $|(\ell,k)|$. Throughout, $\mathcal{E}(u^\star)$ is obtained by Aitken extrapolation and $\mathcal{E}(u_\ell^\star)$ by sufficient Zarantonello steps on each level $\ell$.

**Step 2.** Since the failure of (3.76) increases the current value of $L$ to $\beta L$, it follows from Step 1 that (3.76) can fail only finitely often, until the recomputed sequence $(u_H^k)_{k\in\mathbb{N}_0}$ satisfies (3.76) for all $k \in \mathbb{N}_0$ with $k \geq k_0$. □

**Remark 3.27.** *The optimality results for Algorithm 3.10 are expected to carry over — at least asymptotically — to Algorithm 3.23; see Proposition 3.26. The major difficulty lies in algorithmically determining whether the correct estimate of the Lipschitz constant (and thus $\delta_H$) 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 $\delta_H$ remains uniformly bounded from below.*

## 3.5 Numerical experiments

In this section, we test and illustrate Algorithm 3.23 with numerical experiments. All experiments were implemented using the Matlab code *MooAFEM* [IP23]. 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 (3.59) with isotropic diffusion $\boldsymbol{A} = \left(\begin{smallmatrix}\varepsilon & 0\\ 0 & \varepsilon\end{smallmatrix}\right)$ with $0 < \varepsilon \leq 1$. The adaptivity parameter is set to $\theta = 0.5$ and $C_{\mathrm{mark}} = 1$. Moreover, recall the definition of the overall computational cost from (3.56), which reads

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

**Experiment 3.28** (nonlinear variant of the sine-Gordon equation [AHW23, Experiment 5.1])**.** *For $\Omega = (0,1)^2$, let $\mathcal{X} = H^1_0(\Omega)$ with $|||\cdot|||^2 = \langle \nabla\cdot, \nabla\cdot\rangle$ (i.e., $\varepsilon = 1$) and consider*

$$-\Delta u^\star + (u^\star)^3 + \sin(u^\star) = f \quad \text{in } \Omega \quad \text{subject to} \quad u^\star = 0 \text{ on } \partial\Omega, \tag{3.77}$$

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

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

*which satisfies* (RHS)*. In Figure 3.1, we plot the a posteriori estimator $\eta_\ell(u_\ell^k)$ and the energy difference of the iterative solutions $\big(\mathcal{E}(u_\ell^k) - \mathcal{E}(u^\star)\big)^{1/2}$ against the $\mathtt{work}(\ell,k)$ for lowest order FEM $m = 1$, where we approximate $\mathcal{E}(u^\star)$ by means of Aitken convergence acceleration on uniform meshes with up to $\#\mathcal{T}_{\mathrm{final}} = 67108864$ degrees of freedom on the finest mesh. The decay rate is of (expected) optimal order $O(\mathtt{work}(\ell,k)^{-1/2})$ as $|(\ell,k)| \to \infty$. Moreover, the experimentally observed number of sufficient linearization steps $\underline{k}(\ell)$ is two. Furthermore, in Figure 3.1, we plot the difference of $\mathcal{E}(u_\ell^k)$ to the approximated reference energy $\mathcal{E}(u^\star)$ using Aitken's acceleration and to the energy $\mathcal{E}(u_\ell^\star)$ on $\mathcal{X}_\ell$ over the step counter $|(\ell,k)|$. The reference energy $\mathcal{E}(u_\ell^\star)$ is calculated by a sufficient number of Zarantonello iterations on each level $\ell$ until the energy difference of successive iterates is below the tolerance $\mathtt{tol} < 10^{-15}$.*

**Experiment 3.29** (singularly perturbed sine-Gordon equation)**.** *This example is a variant of [AHW23, Experiment 5.2]. For $d = 2$ and $\Omega = (0,1)^2$, let $\varepsilon = 10^{-5}$ and consider*

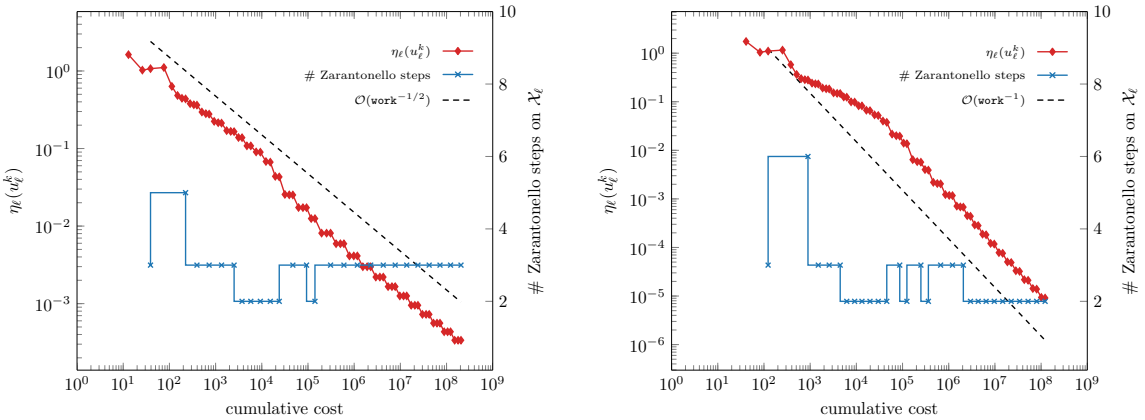
$$-\varepsilon\Delta u^\star + 2u^\star + \sin(u^\star) = 1 \quad \text{in } \Omega \quad \text{subject to} \quad u^\star = 0 \text{ on } \partial\Omega,$$

*with the monotone semilinearity $b(v) = v + \sin(v)$. In this case, the exact solution $u^\star$ is unknown. The used $\mathcal{X}$-norm is given by $|||\cdot|||^2 = \varepsilon\langle\nabla\cdot, \nabla\cdot\rangle + \langle\cdot,\cdot\rangle$. 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 $\boldsymbol{f} = 0$ satisfy* (RHS)*. In this experiment, we employ a slight modification of the error estimator* (3.74) *following [Ver13, Remark 4.14]*

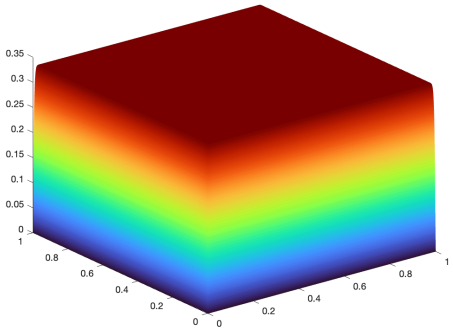
$$\eta_H(T, v_H)^2 := \hbar_T^2 \, \|f + \varepsilon\Delta v_H - b(v_H)\|^2_{L^2(T)} + \hbar_T \, \|[\![\varepsilon \nabla v_H \cdot \boldsymbol{n}]\!]\|^2_{L^2(\partial T\cap\Omega)},$$

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

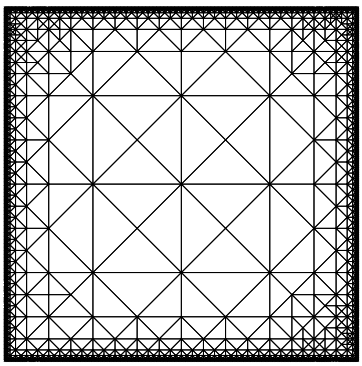
*In Figure 3.2A, we plot the error estimator $\eta_\ell(u_\ell^k)$ for all $(\ell,k) \in Q$ against the $\mathtt{work}(\ell,k)$ for polynomial degrees $m \in \{1,2\}$. The decay rate is of (expected) optimal order $O(\mathtt{work}(\ell,k)^{-m/2})$ as $|(\ell,k)| \to \infty$. The number of Zarantonello steps on each mesh refinement level $\ell$ stabilizes for $m \in \{1,2\}$ at three ($m = 1$) and two ($m = 2$) after an initial phase. For $m = 2$, Figure 3.2B shows the approximate solution $u_\ell^{\underline{k}}$, where $\ell = 28$ and $\underline{k}(28) = 2$. Figure 3.2C depicts a mesh plot for $\#\mathcal{T}_\ell = 4295$ for $\ell = 11$ and $m = 1$. In particular, this experiment shows that Algorithm 3.23 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, $\delta_\ell$ is decreased twice, i.e., $\delta_\ell$ decreases from 1 to $0.5 = 1/L$,*




**(A)** Error estimator $\eta_\ell(u_\ell^k)$ over `work` (diamond, left ordinate) and number of Zarantonello iteration steps on $\mathcal{X}_\ell$ over `work` (cross, right ordinate) for $m = 1$ (left) and $m = 2$ (right).

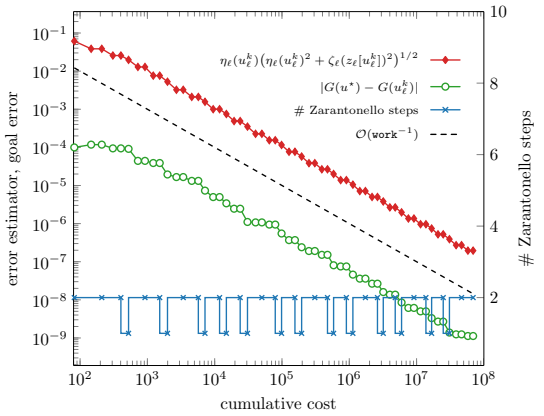


**(B)** Approximate solutions $u_{\underline{\ell}}^{\underline{k}}$, where $\underline{\ell} = 28$, $\underline{k}(28) = 2$, and $m = 2$.

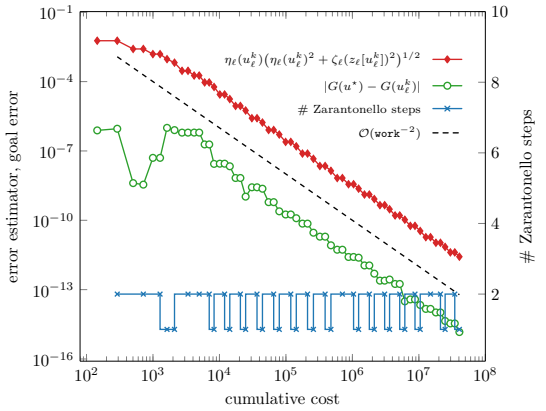


**(C)** Mesh with $\#\mathcal{T}_\ell = 4295$, where $\ell = 11$ and $m = 1$.

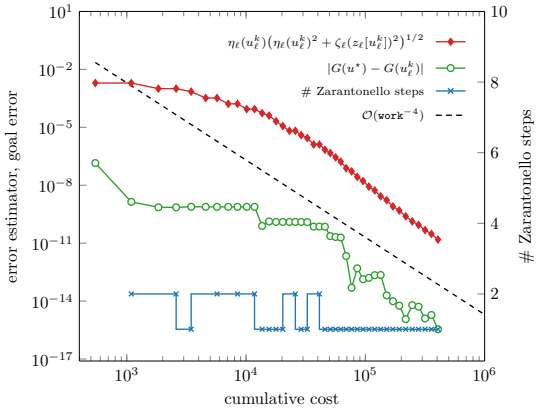
**Figure 3.2:** Using the norm $|||\cdot|||^2 = \varepsilon\,\langle\nabla\cdot\,,\,\nabla\cdot\rangle + \langle\cdot\,,\,\cdot\rangle$ in Experiment 3.29. Top: Convergence plot of the error estimator $\eta_\ell(u_\ell^k)$ over $\mathtt{work}(\ell, k)$ and number of Zarantonello iterations on $\mathcal{X}_\ell$ over `work` for $m = 1$ (top, left) and $m = 2$ (top, right). Bottom: Plot of the approximate solution $u_\ell^k$ (bottom, left) and plot of a sample mesh (bottom, right).



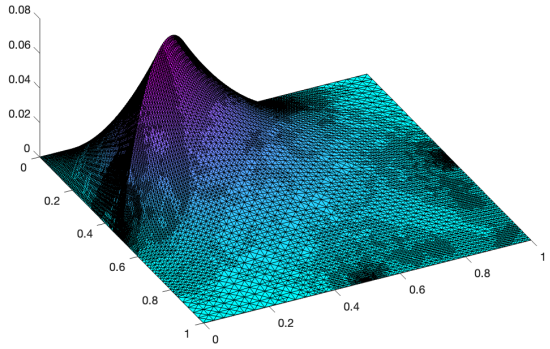

**(A)** Results for $m = 1$.




**(B)** Results for $m = 2$.




**(C)** Results for $m = 4$.



**(D)** Plot of iterative solution $u_\ell^k$, where $\ell = 16$, $\underline{k}(16) = 2$, $\dim(\mathcal{X}_\ell) = 14599$, and $m = 1$.

**Figure 3.3:** 3.3A–3.3C: Product error estimator $\eta_\ell(u_\ell^k)\big[\eta_\ell(u_\ell^k)^2+\zeta_\ell(z_\ell[u_\ell^k])^2\big]^{1/2}$ (diamond, left ordinate), absolute goal error $|G(u^\star)-G(u_\ell^k)|$ (circle, left ordinate), and number of Zarantonello steps on $\mathcal{X}_\ell$ over work (cross, right ordinate) for $m = 1$ (top, left), $m = 2$ (top, right), and $m = 4$ (bottom, left). 3.3D: Plot of an iterative solution $u_\ell^k$. (bottom, right).

*which is optimal according to Remark 3.9 and remains uniformly bounded from below; cf. Remark 3.27.*

**Experiment 3.30** (Goal-oriented AILFEM (GAILFEM))**.** *We also test a canonical extension of Algorithm 3.23 in a goal-oriented setting similar to that of [MS09, Example 7.3]. A thorough treatment of this problem (and the assumptions thereof) is found in [ⓘGOA, Example 35]. We use the proposed practical Algorithm 3.23 as the solve module for the semilinear primal problem in the GOAFEM algorithm [ⓘGOA, Algorithm 17]. Let* $\Omega = (0,1)^2$ *and* $\varepsilon = 1$*. The weak formulation of the primal problem reads: Find* $u^\star \in H^1_0(\Omega)$ *such that*

$$\langle \nabla u^\star\,,\,\nabla v\rangle + \langle b(u^\star)\,,\,v\rangle = \int_\Omega \boldsymbol{f}\cdot\nabla v\,\mathrm{d}x, \qquad \text{for all } v\in H^1_0(\Omega), \tag{3.78}$$

*where* $b(v) = v^3$ *and* $\boldsymbol{f} = \chi_{\Omega_f}(-1,0)$ *with the characteristic function* $\chi_{\Omega_f}$ *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* $w \in H^1_0(\Omega)$ *reads: Find* $z^\star[w] \in H^1_0(\Omega)$ *such that*

$$\langle \nabla z^\star[w], \nabla v\rangle + \langle b'(w) z^\star[w], v\rangle = \int_\Omega \boldsymbol{g}\cdot\nabla v \,\mathrm{d}x, \quad \text{for all } v \in H^1_0(\Omega),$$

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

$$G(v) := -\int_{\Omega_{\boldsymbol{g}}} \frac{\partial v}{\partial x_1}\,\mathrm{d}x \quad \text{for all } v \in H^1_0(\Omega).$$

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

$$\zeta_H(w;T,v_H)^2 := h_T^2 \,\|\Delta v_H - b'(w)(v_H)\|^2_{L^2(T)} + h_T\,\|[\![(\nabla v_H - \boldsymbol{g})\cdot\boldsymbol{n}]\!]\|^2_{L^2(\partial T\cap\Omega)}. \tag{3.79}$$

*We used* $|||\cdot|||^2 = \langle\!\langle \cdot\,,\,\cdot\rangle\!\rangle$ *as the* $\mathcal{X}$*-norm. For various polynomial degrees* $m \in \{1,2,4\}$, *Figure 3.3A–3.3C shows the results of the proposed GAILFEM algorithm driven by the product estimator* $\eta_\ell(u_\ell^k)\big[\eta_\ell(u_\ell^k)^2 + \zeta_\ell(z_\ell[u_\ell^k])^2\big]^{1/2}$, *which is an upper bound to the goal error difference* $G(u^\star) - G(u_\ell^\star)$ *and a viable way to recover optimal convergence rates; cf. [①GOA]. We plot the estimator product* $\eta_\ell(u_\ell^k)\big[\eta_\ell(u_\ell^k)^2 + \zeta_\ell(z_\ell[u_\ell^k])^2\big]^{1/2}$, *the number of Zarantonello steps, and the absolute goal error difference* $|G(u^\star) - G(u_\ell^k)|$ *over the* $\mathtt{work}(\ell,k)$, *where* $G(u^\star) = -0.0015849518088245$ *serves as a reference value; see [①GOA, Example 35]. In Figure 3.3D, we plot the sample solution* $u_\ell^{\underline{k}}$, *where* $\ell = 13$, $\underline{k}(13) = 2$, *and* $m = 1$.

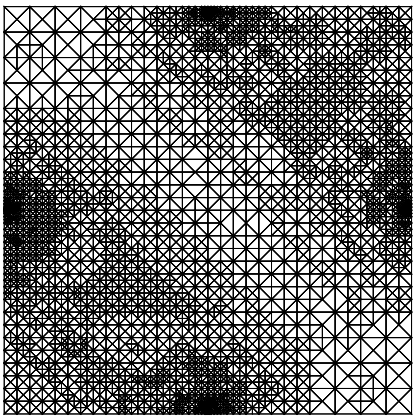
*The decay rate is of (expected) optimal order* $O(\mathtt{work}(\ell,k)^{-m})$ *for* $|(\ell,k)| \to \infty$, *where* $m \in \{1,2,4\}$ *is the polynomial degree of the FEM space* $\mathcal{X}_\ell$. *The number of Zarantonello steps does not exceed two for* $m = \{1,2,4\}$ *and stabilizes after an initial phase at one for* $m = 4$, *respectively. Figure 3.4 depicts two meshes for* $m = 1$ *and* $m = 4$.

## 3.6 Appendix: Convergence for vector-valued semilinear PDEs

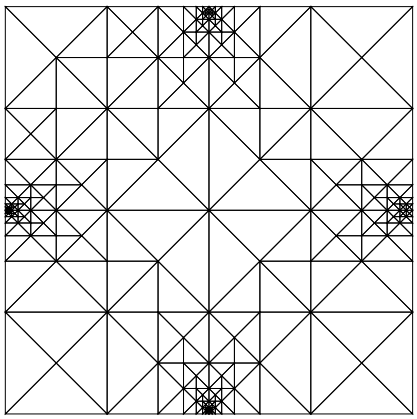
This appendix aims to extend the analysis from Section 3.2 to problems where the monotone operator does not have a potential, e.g., vector-valued semilinear PDEs. We prove plain convergence of Algorithm 3.10 without the assumption (POT) and with the modified stopping criterion

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

replacing Algorithm 3.10(i.b). The proof requires some preliminary observations: First, the convergence of the exact discrete solutions $u_\ell^\star$ towards the exact solution $u_\infty^\star$ in the so-called discrete limit space, which dates back to the seminal work [BV84]. Second, we need to show that the approximate discrete solutions $u_\ell^k$ converge to the same limit.



**(A)** Mesh generated for $m = 1$, where $\dim \mathcal{X}_\ell = 3092$ and $\ell = 12$.



**(B)** Mesh generated for $m = 4$, where $\dim \mathcal{X}_\ell = 3081$ and $\ell = 12$.

**Figure 3.4:** Generated GAILFEM meshes for $m = 1$ (Figure 3.4A) and $m = 4$ (Figure 3.4B).

**Lemma 3.31.** *Suppose that $\mathcal{A}$ satisfies (SM) and (LIP). With the discrete subspaces $\mathcal{X}_\ell \subset \mathcal{X}$ from Algorithm 3.10 (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 $\underline{\ell} = \sup\{\ell \in \mathbb{N}_0 \mid (\ell, 0) \in Q\}$. Then, there exists a unique $u_\infty^\star \in \mathcal{X}_\infty$ which solves*

$$\langle \mathcal{A}u_\infty^\star, v_\infty \rangle = \langle F, v_\infty \rangle \quad \text{for all } v_\infty \in \mathcal{X}_\infty. \tag{3.80}$$

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

$$|||u_\infty^\star - u_\ell^\star||| \to 0 \quad \text{as} \quad \ell \to \underline{\ell}. \tag{3.81}$$

*Additionally, suppose (A1)–(A3) and suppose that the choice of $\delta > 0$ in Algorithm 3.10 ensures norm contraction (3.30). Then, the approximations $u_\ell^k$ computed in Algorithm 3.10 fulfill that*

$$|||u_\infty^\star - u_\ell^k||| \to 0 \quad \text{as} \quad (\ell, k) \in Q \quad \text{with} \quad |(\ell, k)| \to \infty. \tag{3.82}$$

*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 3.2 proves the existence of a unique $u_\infty^\star \in \mathcal{X}_\infty$ satisfying (3.80). The Galerkin solutions $u_\ell^\star$ from (3.4) are also Galerkin approximations of $u_\infty^\star$. Hence, there holds the Céa-type estimate

$$|||u_\infty^\star - u_\ell^\star||| \overset{(3.6)}{\leq} C_{\text{Céa}} \min_{v_\ell \in \mathcal{X}_\ell} |||u_\infty^\star - v_\ell||| \xrightarrow{\ell \to \underline{\ell}} 0, \tag{3.83}$$

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

**Step 2 (approximate solutions for $\underline{\ell} = \infty$).** The norm contraction (3.30) and $u_{\ell+1}^0 = u_\ell^{\underline{k}}$ reveal that

$$0 \le |||u_{\ell+1}^\star - u_{\ell+1}^{\underline{k}(\ell+1)}||| \overset{(3.30)}{\le} q_{\mathrm{N}}^{\underline{k}(\ell+1)} |||u_{\ell+1}^\star - u_{\ell+1}^0||| \le q_{\mathrm{N}} \big[ |||u_\ell^\star - u_\ell^{\underline{k}(\ell)}||| + |||u_{\ell+1}^\star - u_\ell^\star||| \big].$$

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

$$0 \le a_{\ell+1} \le q_{\mathrm{N}}\, a_\ell + b_\ell, \quad \text{where} \quad \lim_{\ell \to \infty} b_\ell = 0.$$

It follows from elementary calculus (cf. [CFPP14, Corollary 4.8]) that

$$0 = \lim_{\ell \to \infty} a_\ell = \lim_{\ell \to \infty} |||u_\ell^\star - u_\ell^{\underline{k}}|||.$$

Altogether, we obtain that

$$\begin{aligned} |||u_\infty^\star - u_\ell^{\underline{k}}||| &\le |||u_\infty^\star - u_\ell^\star||| + |||u_\ell^\star - u_\ell^{\underline{k}}||| \overset{(3.30)}{\le} |||u_\infty^\star - u_\ell^\star||| + |||u_\ell^\star - u_\ell^0||| \\ &\le |||u_\infty^\star - u_\ell^\star||| + |||u_\ell^\star - u_{\ell-1}^\star||| + |||u_{\ell-1}^\star - u_{\ell-1}^{\underline{k}}||| \to 0 \quad \text{as} \quad \ell \to \infty. \end{aligned}$$

**Step 3 (approximate solutions for $\underline{\ell} < \infty$ and $\underline{k}(\ell) = \infty$).** It holds that $u_\infty^\star = u_{\underline{\ell}}^\star$ and hence, due to (3.30),

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

This concludes the proof. □

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

**Theorem 3.32: 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 3.10 ensures (3.30). Then, for any choice of the marking parameters $0 < \theta \le 1$, $\lambda > 0$, and $1 \le C_{\mathrm{mark}} \le \infty$, Algorithm 3.10 with modified stopping criterion (i.b'') guarantees convergence of the quasi-error from (3.38), i.e.,*

$$\Delta_\ell^k = |||u^\star - u_\ell^k||| + \eta_\ell(u_\ell^k) \to 0 \quad \text{as } (\ell, k) \in \mathcal{Q} \text{ with } |(\ell, k)| \to \infty. \tag{3.84}$$

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

**Case 1 ($\underline{\ell} = \infty$).** Recall the generalized estimator reduction [CFPP14, Lemma 4.7]: Let $\omega > 0$. Given the Dörfler marking in Algorithm 3.10(iii), it follows that

$$\eta_{\ell+1}(u_{\ell+1}^{\underline{k}})^2 \le q_{\mathrm{est}}\, \eta_\ell(u_\ell^{\underline{k}})^2 + C_{\mathrm{est}} |||u_{\ell+1}^{\underline{k}} - u_\ell^{\underline{k}}|||^2, \tag{3.85}$$

where $0 < q_{\mathrm{est}} := (1+\omega)\big[1-(1-q_{\mathrm{red}}^2)\,\theta\big] < 1$ and $C_{\mathrm{est}} := (1+\omega^{-1})\,C_{\mathrm{stab}}[4M]^2$ with $\omega > 0$ being sufficiently small and where $4M$ stems from nested iteration (3.28). From Lemma 3.31, we infer that $|||u_{\ell+1}^k - u_\ell^k||| \to 0$ as $\ell \to \infty$. Hence, it follows from elementary calculus (cf. [CFPP14, Corollary 4.8]) that $\eta_\ell(u_\ell^k) \to 0$ as $\ell \to \infty$. Moreover, this and Lemma 3.31 prove that

$$\begin{aligned}|||u^\star - u_\ell^k||| &\overset{(A3)}{\le} C_{\mathrm{rel}}\eta_\ell(u_\ell^\star) + |||u_\ell^\star - u_\ell^k||| \overset{(A1)}{\le} C_{\mathrm{rel}}\eta_\ell(u_\ell^k) + (1+C_{\mathrm{rel}}C_{\mathrm{stab}}[3M])\,|||u_\ell^\star - u_\ell^k||| \\ &\le C_{\mathrm{rel}}\eta_\ell(u_\ell^k) + (1+C_{\mathrm{rel}}C_{\mathrm{stab}}[3M])\big[|||u_\ell^\star - u_\infty^\star||| + |||u_\infty^\star - u_\ell^k|||\big] \xrightarrow{\ell\to\infty} 0.\end{aligned}$$

We conclude that $|||u^\star - u_\ell^k||| + \eta_\ell(u_\ell^k) + \eta_\ell(u_\ell^\star) \to 0$ as $\ell \to \infty$. Due to (3.18) together with Lemma 3.31 and for $C'_{\mathrm{rel}} := 1 + C_{\mathrm{rel}}$, this yields for all $(\ell, k) \in \mathcal{Q}$ that

$$\begin{aligned}\Delta_\ell^k &\le C'_{\mathrm{rel}}\eta_\ell(u_\ell^\star) + \big[1+C_{\mathrm{stab}}[3M]\big]\,|||u_\ell^\star - u_\ell^k||| \overset{(3.30)}{\le} C'_{\mathrm{rel}}\eta_\ell(u_\ell^\star) + \big[1+C_{\mathrm{stab}}[3M]\big]\,|||u_\ell^\star - u_\ell^0||| \\ &\le C'_{\mathrm{rel}}\eta_\ell(u_\ell^\star) + \big[1+C_{\mathrm{stab}}[3M]\big]\,\big[|||u_\ell^\star - u_{\ell-1}^\star||| + |||u_{\ell-1}^\star - u_{\ell-1}^k|||\big] \xrightarrow{\ell\to\infty} 0.\end{aligned}$$

This concludes the proof of the first case.

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

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

Since norm contraction (3.30) holds, the arguments to obtain (3.32) prove the existence of $k_0 \in \mathbb{N}$ such that, for all $k \ge k_0$, it holds that

$$|||u_{\underline{\ell}}^k||| \le 2M.$$

We deduce from the (not met) stopping criterion in Algorithm 3.10(i.b'') and (3.30) that

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

With contraction (3.30), we see that

$$|||u^\star - u_{\underline{\ell}}^k||| \overset{(A3)}{\le} C_{\mathrm{rel}}\eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) + |||u_{\underline{\ell}}^\star - u_{\underline{\ell}}^k||| \overset{(A1)}{\le} C_{\mathrm{rel}}\eta_{\underline{\ell}}(u_{\underline{\ell}}^k) + (1+C_{\mathrm{stab}}[3M])\,|||u_{\underline{\ell}}^\star - u_{\underline{\ell}}^k||| \xrightarrow{k\to\infty} 0.$$

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

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

**Corollary 3.33.** *Under the assumptions of Theorem 3.32, there hold the following implications:*

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

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

*Proof.* **(i).** According to Theorem 3.32, it holds that

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

Norm contraction (3.30) proves that

$$|||u_{\underline{\ell}}^\star - u_{\underline{\ell}}^k||| \le q_{\mathrm{N}}^k \, |||u_{\underline{\ell}}^\star - u_{\underline{\ell}}^0||| \to 0 \quad \text{as} \quad k \to \infty.$$

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

$$0 \le \eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) \le \eta_{\underline{\ell}}(u_{\underline{\ell}}^k) + C_{\mathrm{stab}}[3M] \, |||u_{\underline{\ell}}^\star - u_{\underline{\ell}}^k||| \to 0 \quad \text{as} \quad k \to \infty.$$

This concludes the proof of (i).

**(ii).** Note that the stopping criterion in Algorithm 3.10(i.b'') implies that $|||u_\ell^k - u_\ell^{k-1}||| \le \lambda \, \eta_\ell(u_\ell^k) = 0$ by assumption. Thus, $u_\ell^k = u_\ell^{k-1}$. This implies that $u_\ell^{k-1}$ is a fixed point of $\Phi_\ell(\delta;\cdot)$. Since the fixed point is unique, we infer that $u_\ell^k = u_\ell^{k-1} = u_\ell^\star$. With reliability (A3), we thus obtain that

$$|||u^\star - u_\ell^\star||| \overset{\text{(A3)}}{\le} C_{\mathrm{rel}} \, \eta_\ell(u_\ell^\star) = C_{\mathrm{rel}} \, \eta_\ell(u_\ell^k) = 0.$$

This concludes the proof. □

Plain convergence is required to obtain results proving weak convergence in the spirit of [①GOA, Lemma 28]. This is pivotal for achieving quasi-orthogonality along the lines of [①GOA, Lemma 29], which can substitute (3.8) in the proof of full linear convergence. Details are omitted.

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

This chapter is taken from:

> [③AIL2]: M. Brunner, D. Praetorius, and J. Streitberger. Cost-optimal adaptive FEM with linearization and algebraic solver for semilinear elliptic PDEs, 2024. arXiv: 2401.06486

## 4.1 Introduction

### 4.1.1 Problem setting and main results

Undoubtedly, adaptive finite element methods (AFEMs) are in the canon of reliable numerical methods for the solution of partial differential equations (PDEs). Some of the seminal contributions in this still very active area are [BV84; Dör96; MNS00; BDD04; Ste07; CKNS08; KS11; CN12; FFP14] for linear problems, [Vee02; DK08; BDK12; GMZ12; GHPS21] for nonlinear problems, and [CFPP14] for an abstract framework.

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

$$-\operatorname{div}(\boldsymbol{A}\nabla u^\star) + b(u^\star) = F \quad \text{in } \Omega \quad \text{subject to} \quad u^\star = 0 \quad \text{on } \partial\Omega, \tag{4.1}$$

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

Moreover, the model problem (4.1) can be recast into the framework of strongly monotone and locally Lipschitz continuous operators such that the abstract model problem reads: For $\mathcal{X} = H^1_0(\Omega)$ with topological dual space $\mathcal{X}' = H^{-1}(\Omega)$ and duality bracket $\langle\cdot\,,\,\cdot\rangle$, a nonlinear operator $\mathcal{A}\colon \mathcal{X} \to \mathcal{X}'$, and given data $F \in \mathcal{X}'$, we aim to approximate the solution $u^\star \in \mathcal{X}$ to

$$\langle \mathcal{A}u^\star\,,\, v\rangle = \langle F\,,\, v\rangle \quad \text{for all } v \in \mathcal{X}. \tag{4.2}$$

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

$$\langle \mathcal{A}u^\star_H\,,\, v_H\rangle = \langle F\,,\, v_H\rangle \quad \text{for all } v_H \in \mathcal{X}_H, \tag{4.3}$$

which, however, can hardly be computed exactly, since (4.3) is still a discrete nonlinear system of equations.

The major difficulty of such problems is that the Lipschitz constant of $\mathcal{A}$ depends on

the considered functions $v$ and $w$ in the sense that for $\vartheta > 0$, it holds that

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

Moreover, this dependence also appears in the stability constant of the residual-based *a posteriori* error estimator [Ver13; ①GOA].

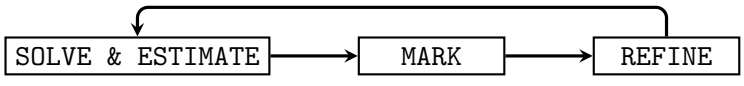
Hence, for such a problem class, any approximate numerical scheme must ensure uniform boundedness of all computed approximations $u_H^\star \approx u_H \in \mathcal{X}_H$ throughout the algorithm. This constitutes the first main result: The developed *adaptive iteratively linearized FEM* (AILFEM) algorithm (more detailed in Algorithm 4.6 below) guarantees a uniform upper bound on all iterates (see Theorem 4.8 below). In particular, the algorithm steers the decision whether it is more preferable to refine the mesh adaptively or to do an additional step of linearization or a further algebraic solver step instead.

Once uniform boundedness is established, we prove full R-linear convergence (Theorem 4.13 below) as the second main result. Full R-linear convergence establishes contraction in each step of the algorithm regardless of the algorithmic decision. At the expense of a more challenging analysis that links energy arguments with the energy norm of the algebraic solver, full R-linear convergence is guaranteed for all mesh levels $\ell \geq \ell_0 = 0$ while prior works [①GOA; BIM+23] used compactness arguments which only guaranteed the existence of the index $\ell_0 \in \mathbb{N}_0$ (and not necessarily $\ell_0 = 0$). As a consequence of uniform boundedness and full R-linear convergence, the third main result proves optimal rates both understood with respect to the degrees of freedom and with respect to the overall computational cost (Corollary 4.14 and Theorem 4.15) of the proposed algorithm.

Compared to existing results in the literature [GHPS21; HPSV21; HPW21; BFM+23], all three main results require a suitable adaptation of the stopping criteria of the linearization loop as well as sufficiently many iterations in the algebra loop, together with subtle technical challenges, in particular, for the proof of full R-linear convergence.

### 4.1.2 From AFEM to AILFEM

On each mesh level (with mesh index $\ell$), the arising discrete nonlinear problems cannot be solved exactly in practice as supposed in classical AFEM [Vee02; DK08; BDK12; GMZ12]. To deal with this issue, we follow [CW17; GHPS18; HW20b] and consider the so-called *Zarantonello iteration* from [Zar60] as a linearization method (with index $k$). The Zarantonello iteration is a Richardson-type iteration where only a Laplace-type problem has to be solved in each iteration. Since the arising large SPD systems are still expensive to solve exactly, we employ a contractive algebraic solver as a nested loop to solve the Zarantonello system inexactly (with iteration index $i$). The loops thus come with a natural nestedness (see Figure 4.1), where the overall schematic loop of the algorithm reads




Since the proposed adaptive loop depends on all previous computations, optimal convergence rates should be understood with respect to the overall computational cost. This

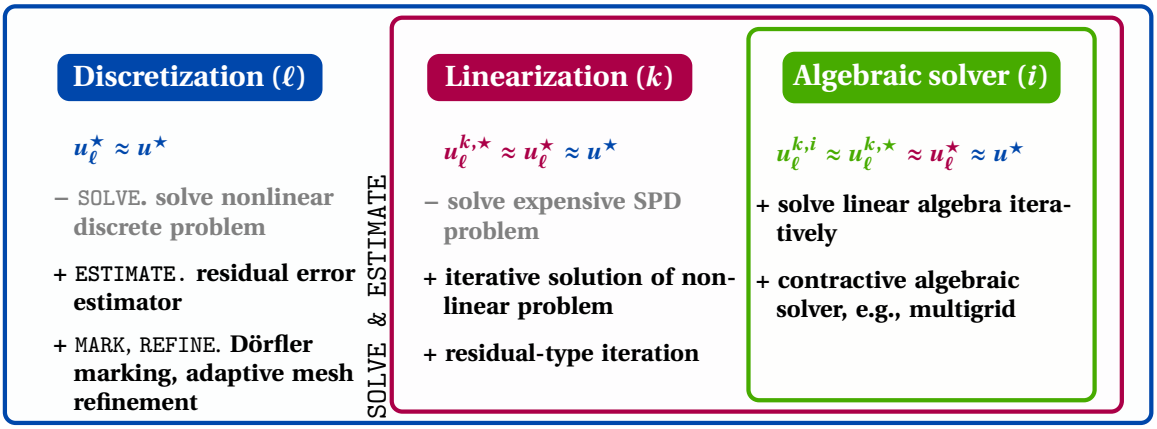



**Figure 4.1:** Depiction of the nested loops of the AILFEM algorithm 4.6 below.

idea of optimal complexity originates from the wavelet community [CDD01; CDD03] and was later used in the context of AFEM in [Ste07] for the Poisson model problem and [CG12] for the Poisson eigenvalue problem, both under realistic assumptions on generic iterative solvers.

AILFEMs with iterative and/or inexact solver with *a posteriori* error estimators are found in, e.g., [BMS10; EEV11; AGL13; EV13; AW15; CW17] and references therein. Besides the Zarantonello iteration, for globally Lipschitz continuous nonlinearities, the works [HW20a; HW20b; HPW21] analyze also other linearizations such as the Kačanov iteration or damped Newton schemes. Optimal complexity of the Zarantonello loop that is coupled with an algebraic loop is analyzed in [BIM+23] for nonsymmetric second-order linear elliptic PDEs and for strongly monotone (and globally Lipschitz continuous) model problems in [GHPS18; GHPS21; HPSV21; HPW21; BFM+23].

The literature on AILFEMs for locally Lipschitz continuous problems is scarce and closing this gap is the aim of this work. The semilinear model problem is treated in, e.g., [AW15] by a damped Newton iteration and in [AHW23] by an energy-based approach with experimentally observed optimal rates. We also refer to the own work [②AIL1] for an AILFEM with optimal rates with respect to the the overall computational cost under the assumption that the algebraic solver can be performed at linear cost.

### 4.1.3 Outline

This paper is structured as follows: Section 4.2 introduces the abstract framework on locally Lipschitz continuous operators. In Section 4.3, we formulate the (idealized) AILFEM algorithm (Algorithm 4.6). We prove uniform boundedness for the final iterates of the algebraic solver (Theorem 4.8). Section 4.4 presents the second main result: Full R-linear convergence (Theorem 4.13). In particular, rates with respect to the degrees of freedom coincide with rates with respect to the computational cost (Corollary 4.14). In Section 4.5, we prove the main result on optimal complexity of the proposed AILFEM

algorithm (Theorem 4.15). In Section 4.6, we present numerical experiments of the proposed AILFEM strategy and investigate its optimal complexity for various choices of the adaptivity parameters.

## 4.2 Strongly monotone operators

This section introduces an abstract framework of strongly monotone and locally Lipschitz continuous operators. This class of operators covers the model problem (4.1) of semilinear elliptic PDEs with monotone semilinearity.

### 4.2.1 Abstract model problem

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

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

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

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

**Remark 4.1.** *We remark that local Lipschitz continuity is often defined differently in the existing literature, cf. [Zei90, p. 565]: For all $\Theta > 0$, there exists $L'[\Theta] > 0$ such that*

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

*We note that the conditions* (LIP) *and* (LIP$'$) *are indeed equivalent in the sense that* (LIP) *yields* (LIP$'$) *with $\Theta = 2\vartheta$, and, conversely,* (LIP$'$) *yields* (LIP) *with $\vartheta = 2\,\Theta$. However, condition* (LIP) *is better suited for the inductive proof of Proposition 4.4 which is the main ingredient to guarantee uniform boundedness in Theorem 4.8.*

Without loss of generality, we may suppose that $\mathcal{A}0 \neq F \in \mathcal{X}'$. We consider the operator equation: Seek $u^\star \in \mathcal{X}$ that solves (4.2). For any closed subspace $\mathcal{X}_H \subseteq \mathcal{X}$, we consider the corresponding Galerkin discretization (4.3). We note existence and uniqueness of the solutions to (4.2)–(4.3) and a Céa-type estimate.

**Proposition 4.2** ([②AIL1, Proposition 2])**.** *Suppose that $\mathcal{A}$ satisfies* (SM) *and* (LIP). *Then,* (4.2)–(4.3) *admit unique solutions $u^\star \in \mathcal{X}$ and $u_H^\star \in \mathcal{X}_H$, respectively, and*

$$\max\{|||u^\star|||,|||u_H^\star|||\} \le M := \frac{1}{\alpha}\,\|F-\mathcal{A}0\|_{\mathcal{X}'} > 0 \tag{4.4}$$

*as well as*

$$|||u^\star - u_H^\star||| \le C_{\text{Céa}} \min_{v_H \in \mathcal{X}_H} |||u^\star - v_H||| \quad \text{with} \quad C_{\text{Céa}} = L[2M]/\alpha. \qquad \square \tag{4.5}$$

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

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

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

**Lemma 4.3** (see, e.g., [GHPS18, Lemma 5.1])**.** *Let $\mathcal{X}_H \subseteq \mathcal{X}$ be a closed subspace (where also $\mathcal{X}_H$ replaced by $\mathcal{X}_H$ is admissible). Suppose that $\mathcal{A}$ satisfies* (SM)*,* (LIP)*, and* (POT)*. Let $\vartheta \ge M$. Let $v_H \in \mathcal{X}_H$ with $|||v_H - u_H^\star||| \le \vartheta$. Then, it holds that*

$$\frac{\alpha}{2}\,|||v_H - u_H^\star|||^2 \le \mathcal{E}(v_H) - \mathcal{E}(u_H^\star) \le \frac{L[\vartheta]}{2}\,|||v_H - u_H^\star|||^2. \tag{4.6}$$

*In particular, the solution $u_H^\star$ of* (4.3) *is indeed the unique minimizer of $\mathcal{E}$ in $\mathcal{X}_H$, i.e.,*

$$\mathcal{E}(u_H^\star) \le \mathcal{E}(v_H) \quad \text{for all } v_H \in \mathcal{X}_H, \tag{4.7}$$

*and, therefore,* (4.3) *can equivalently be reformulated as an energy minimization problem:*

$$\text{Find} \quad u_H^\star \in \mathcal{X}_H \quad \text{such that} \quad \mathcal{E}(u_H^\star) = \min_{v_H \in \mathcal{X}_H} \mathcal{E}(v_H). \tag{4.8}$$

*In particular, it holds that*

$$\mathcal{E}(v_H) - \mathcal{E}(u^\star) = \big[\mathcal{E}(v_H) - \mathcal{E}(u_H^\star)\big] + \big[\mathcal{E}(u_H^\star) - \mathcal{E}(u^\star)\big] \quad \text{for all } v_H \in \mathcal{X}_H \tag{4.9}$$

*and all these energy differences are nonnegative.* $\square$

### 4.2.2 Iterative linearization and algebraic solver

Let $\mathcal{X}_H \subset \mathcal{X}$ be a finite-dimensional (and hence closed) subspace of $\mathcal{X}$. In order to solve the arising nonlinear discrete problems (4.3), we will incorporate a linearization method as well as an algebraic solver into the proposed algorithm.

**Linearization by Zarantonello iteration.** For a detailed discussion of the Zarantonello iteration, we refer to [②AIL1, Section 2.2–2.4]. For a damping parameter $\delta > 0$ and $w_H \in \mathcal{X}_H$, let $\Phi_H(\delta; w_H) \in \mathcal{X}_H$ solve

$$\langle \Phi_H(\delta; u_H)\,,\, v_H\rangle = \langle\!\langle u_H\,,\, v_H\rangle\!\rangle + \delta\,\big[F(v_H) - \langle \mathcal{A}(u_H)\,,\, v_H\rangle\big] \quad \text{for all } v_H \in \mathcal{X}_H. \tag{4.10}$$

The Lax–Milgram lemma proves existence and uniqueness of $\Phi_H(\delta; u_H)$, i.e., the Zarantonello operator $\Phi_H(\delta;\cdot)\colon \mathcal{X}_H \to \mathcal{X}_H$ is well-defined. In particular, $u_H^\star = \Phi(\delta; u_H^\star)$ is the

unique fixed point of $\Phi_H(\delta;\cdot)$ for any damping parameter $\delta > 0$. Moreover, for sufficiently small $\delta > 0$, the Zarantonello operator is norm-contractive.

**Proposition 4.4** (see, e.g., [②AIL1, Proposition 3.4])**.** *Suppose that $\mathcal{A}$ satisfies (SM) and (LIP). Let $\vartheta > 0$ and $v_H, w_H \in \mathcal{X}_H$ with $\max\{|||v_H|||, |||v_H - w_H|||\} \le \vartheta$. Then, for all $0 < \delta < 2\alpha/L[\vartheta]^2$ and $0 < q^\star_{\mathrm{Zar}}[\delta,\vartheta]^2 := 1 - \delta\,(2\alpha - \delta L[\vartheta]^2) < 1$, it holds that*

$$|||\Phi_H(\delta; v_H) - \Phi_H(\delta; w_H)||| \le q^\star_{\mathrm{Zar}}[\delta,\vartheta]\,|||v_H - w_H|||. \tag{4.11}$$

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

**Algebraic solver.** The Zarantonello system (4.10) leads to an SPD system of equations to compute $\Phi_H(\delta; u_H)$. Since large SPD problems are still computationally expensive, we employ an iterative algebraic solver with process function $\Psi_H\colon \mathcal{X}' \times \mathcal{X}_H \to \mathcal{X}_H$ to solve the arising system (4.10). More precisely, given a linear functional $\varphi \in \mathcal{X}'$ and an approximation $w_H \in \mathcal{X}_H$ of the exact solutions $w_H^\star \in \mathcal{X}_H$ to

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

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

$$|||w_H^\star - \Psi_H(\varphi; w_H)||| \le q_{\mathrm{alg}}\,|||w_H^\star - w_H||| \quad \text{for all } w_H \in \mathcal{X}_H. \tag{4.12}$$

To simplify notation when the right-hand side $\varphi$ is complicated or lengthy (as for the Zarantonello iteration (4.10)), we shall write $\Psi_H(w_H^\star;\cdot)$ instead of $\Psi_H(\varphi;\cdot)$, even though $w_H^\star$ is unknown and will never be computed.

### 4.2.3 Mesh refinement

Henceforth, let $\mathcal{T}_0$ be an initial triangulation of $\Omega$ into compact triangles. For mesh refinement, we use newest vertex bisection (NVB); cf. [Ste08] for $d \ge 2$ with admissible $\mathcal{T}_0$ as well as [KPP13] for $d = 2$ and [DGS23] for $d \ge 2$ with nonadmissible $\mathcal{T}_0$. For $d = 1$, we refer to [AFF+13]. For each triangulation $\mathcal{T}_H$ and marked elements $\mathcal{M}_H \subseteq \mathcal{T}_H$, let $\mathcal{T}_h := \texttt{refine}(\mathcal{T}_H, \mathcal{M}_H)$ be the coarsest refinement of $\mathcal{T}_H$ such that at least all elements $T \in \mathcal{M}_H$ have been refined, i.e., $\mathcal{M}_H \subseteq \mathcal{T}_H \setminus \mathcal{T}_h$. We write $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ if $\mathcal{T}_h$ can be obtained from $\mathcal{T}_H$ by finitely many steps of NVB, and, for $N \in \mathbb{N}_0$, we write $\mathcal{T}_h \in \mathbb{T}_N(\mathcal{T}_H)$ if $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ and $\#\mathcal{T}_h - \#\mathcal{T}_H \le N$. To abbreviate notation, let $\mathbb{T} := \mathbb{T}(\mathcal{T}_0)$. Throughout, any $\mathcal{T}_H \in \mathbb{T}$ is associated with a finite-dimensional space $\mathcal{X}_H \subset \mathcal{X}$ such that nestedness of meshes $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ implies nestedness of the associated spaces $\mathcal{X}_H \subseteq \mathcal{X}_h$.

### 4.2.4 Axioms of adaptivity and *a posteriori* error estimator

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

$$\eta_H(v_H) \coloneqq \eta_H(\mathcal{T}_H, v_H), \text{ where } \eta_H(\mathcal{U}_H, v_H) \coloneqq \Big( \sum_{T \in \mathcal{U}_H} \eta_H(T, v_H)^2 \Big)^{1/2} \text{for all } \mathcal{U}_H \subseteq \mathcal{T}_H. \tag{4.13}$$

We suppose that the error estimator $\eta_H$ satisfies the following axioms of adaptivity from [CFPP14] with a slightly relaxed variant of stability (A1) in the spirit of [①GOA].

**(A1) stability:** For all $\vartheta > 0$ and all $\mathcal{U}_H \subseteq \mathcal{T}_h \cap \mathcal{T}_H$, there exists $C_{\mathrm{stab}}[\vartheta] > 0$ 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

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

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

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

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

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

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

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

### 4.2.5 Application of abstract framework (4.2) to semilinear PDEs (4.1)

In the following, we comment on how the semilinear PDE (4.1) fits into the abstract framework in Section 4.2.1–4.2.4. Let $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\,|\,\}$, be a bounded Lipschitz domain with polygonal boundary. The weak formulation of the semilinear model problem (4.1) reads: Given $F \in H^{-1}(\Omega)$, find $u^\star \in \mathcal{X} \coloneqq H^1_0(\Omega)$ such that

$$\langle \boldsymbol{A}\, \nabla u^\star\,,\, \nabla v \rangle_\Omega + \langle b(u^\star)\,,\, v \rangle_\Omega = \langle F\,,\, v \rangle \quad \text{for all } v \in H^1_0(\Omega), \tag{4.14}$$

where $\langle \cdot\,,\, \cdot \rangle_\Omega$ denotes the $L^2(\Omega)$-scalar product. Note that (4.14) coincides with (4.2), where $\mathcal{A}u \coloneqq \langle \boldsymbol{A}\, \nabla u\,,\, \nabla \cdot \rangle_\Omega + \langle b(u)\,,\, \cdot \rangle_\Omega$ with $u \in \mathcal{X}$. The precise assumptions on the model problem are given as follows.

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

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

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

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

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

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

**Assumptions on the nonlinear reaction coefficient.** The nonlinearity $b(\cdot)$ satisfies the following assumptions from [BHSZ11, (A1)–(A3)]:

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

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

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

To establish continuity of $v\mapsto\langle b(v)\,,\,w\rangle_\Omega$, we impose the following **growth condition** on $b(v)$; see, e.g., [FK80, Chapter III, (12)] or [BHSZ11, (A4)]:

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

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

These assumptions suffice to prove that the operator $\mathcal{A} := \mathcal{X}\to\mathcal{X}' = H^{-1}(\Omega)$ associated with the model problem (4.14) is strongly monotone (SM) and locally Lipschitz continuous (LIP) in the sense of Section 4.2.1; see [②AIL1, Lemma 3.21].

**Energy minimization.** Associated with the semilinear model problem (4.14), we consider the **energy**

$$\mathcal{E}(v) = \frac12\int_\Omega |\boldsymbol{A}^{1/2}\nabla v|^2\,\mathrm{d}x + \int_\Omega\int_0^{v(x)} b(s)\,\mathrm{d}s\,\mathrm{d}x - \int_\Omega fv\,\mathrm{d}x - \int_\Omega \boldsymbol{f}\cdot\nabla v\,\mathrm{d}x \quad \text{for } v\in H^1_0(\Omega).$$

To ensure the well-posedness of integrals, we require the following stronger growth condition (guaranteeing **compactness** of the nonlinear reaction term). Indeed, the same assumption is also required for stability (A1) of the residual error estimator (4.15) below.

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

**Residual error estimator.** To guarantee well-posedness, we additionally require that $\boldsymbol{A}|_T\in[W^{1,\infty}(T)]^{d\times d}$ and $\boldsymbol{f}|_T\in[W^{1,\infty}(T)]^d$ for all $T\in\mathcal{T}_0$, where $\mathcal{T}_0$ is the initial triangulation

of the adaptive algorithm. Then, for $\mathcal{T}_H \in \mathbb{T}$ and $v_H \in \mathcal{X}_H$, the local contributions of the standard residual error estimator (4.13) for the semilinear model problem (4.14) read

$$\begin{aligned}\eta_H(T, v_H)^2 &:= h_T^2 \, \|f + \operatorname{div}(\boldsymbol{A}\nabla v_H - \boldsymbol{f}) - b(v_H)\|_{L^2(T)}^2 \\ &\quad + h_T \, \|[\![(\boldsymbol{A}\nabla v_H - \boldsymbol{f}) \cdot \boldsymbol{n}]\!]\|_{L^2(\partial T\cap\Omega)}^2,\end{aligned} \tag{4.15}$$

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

**Proposition 4.5** ([①GOA, Proposition 2.15])**.** *Suppose* (RHS), (ELL), (CAR), (MON), *and* (CGC). *Suppose that NVB is employed as a refinement strategy. Then, the residual error estimator from* (4.15) *satisfies* (A1)–(A4) *from Section 4.2.4. The constant* $C_{\mathrm{rel}}$ *depends only on* $d$, $\mu_0$, *and uniform shape regularity of the initial mesh* $\mathcal{T}_0$. *The constant* $C_{\mathrm{drel}}$ *depends, in addition, on the polynomial degree* $p$, *and* $C_{\mathrm{stab}}[\vartheta]$ *depends furthermore on* $|\Omega|$, $\vartheta$, $N$, $R$, *and* $\boldsymbol{A}$. □

**Algebraic solver.** As an algebraic solver, we employ a norm-contractive solver to solve the Zarantonello system (4.10). Possible choices are, e.g., an optimally preconditioned conjugate gradient method [CNX12] or an optimal geometric multigrid [WZ17; IMPS23]. More precisely, the numerical experiments below employ the $hp$-robust multigrid method from [IMPS23], which is well-defined owing to ellipticity (ELL).

## 4.3 Fully adaptive algorithm

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

### 4.3.1 Fully adaptive algorithm

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

**Algorithm 4.6: adaptive iterative linearized FEM (AILFEM)**

**Input:** Initial mesh $\mathcal{T}_0$, marking parameters $0 < \theta \le 1$, $C_{\mathrm{mark}} \ge 1$, solver parameters $\lambda_{\mathrm{lin}}, \lambda_{\mathrm{alg}} > 0$, minimal number of algebraic solver steps $i_{\min} \in \mathbb{N}$, initial guess $u_0^{0,0} := u_0^{0,\star} := u_0^{0,\underline{i}} \in \mathcal{X}_0$ with $|||u_0^{0,0}||| \le 2M$, and Zarantonello damping parameter $\delta > 0$.
**Adaptive loop:** For all $\ell = 0, 1, 2, \dots$, repeat the following steps (I)–(III):

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

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

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

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

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

$$|||u_\ell^{k,i-1} - u_\ell^{k,i}||| \le \lambda_{\mathrm{alg}} \left[\lambda_{\mathrm{lin}}\, \eta_\ell(u_\ell^{k,i}) + |||u_\ell^{k,i} - u_\ell^{k,0}|||\right] \ \texttt{AND}\ i_{\min} \le i. \tag{4.16}$$

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

$$\mathcal{E}(u_\ell^{k,0}) - \mathcal{E}(u_\ell^{k,\underline{i}}) \le \lambda_{\mathrm{lin}}^2\, \eta_\ell(u_\ell^{k,\underline{i}})^2 \quad \texttt{AND} \quad |||u_\ell^{k,\underline{i}}||| \le 2M. \tag{4.17}$$

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

$$\#\mathcal{M}_\ell \le C_{\mathrm{mark}} \min_{\mathcal{U}_\ell \in \mathbb{M}_\ell[\theta, u_\ell^{k,\underline{i}}]} \#\mathcal{U}_\ell. \tag{4.18}$$

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

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

For the analysis of Algorithm 4.6, we define the countably infinite index set

$$\mathcal{Q} := \{(\ell, k, i) \in \mathbb{N}_0^3 \colon u_\ell^{k,i} \text{ is used in Algorithm 4.6 } |, \}$$

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

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

We note, first, that these definitions are consistent with those of Algorithm 4.6, second, that Lemma 4.7 below proves that $\underline{i}[\ell, k] < \infty$, and, third, that hence either $\underline{\ell} = \infty$ or $\underline{\ell} < \infty$ with $\underline{k}[\underline{\ell}] = \infty$. For all $(\ell, k, i) \in \mathcal{Q}$, we introduce the total step counter $|\cdot, \cdot, \cdot|$ defined by

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

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

In the later application to AILFEM for semilinear elliptic PDEs, every step of Algorithm 4.6 can be performed in linear complexity as the following arguments show.

- `SOLVE`. The employed algebraic solver is an $hp$-robust multigrid [IMPS23] and hence each algebraic solver step requires only $\mathcal{O}(\#\mathcal{T}_\ell)$ operations.

- ▹ `ESTIMATE`. The simultaneous computation of the standard error indicators $\eta_\ell(T, u_\ell^{k,i})$ for all $T \in \mathcal{T}_\ell$ can be done at the cost of $\mathcal{O}(\#\mathcal{T}_\ell)$.
- ▹ `MARK`. The employed Dörfler marking (and the involved determination of $\mathcal{M}_\ell$) is indeed a linear complexity problem; see [Ste07] for $C_{\mathrm{mark}} = 2$ and [PP20] for $C_{\mathrm{mark}} = 1$.
- ▹ `REFINE`. The refinement of $\mathcal{T}_\ell$ is based on NVB and, owing to the mesh-closure estimate [BDD04; Ste08], requires only linear cost $\mathcal{O}(\#\mathcal{T}_\ell)$.

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

$$\mathtt{cost}(\ell,k,i) := \sum_{\substack{(\ell',k',i')\in\mathcal{Q}\\ |\ell',k',i'|\le|\ell,k,i|}} \#\mathcal{T}_{\ell'} = \sum_{\ell'=0}^{\ell-1}\sum_{k'=1}^{\underline{k}[\ell']}\sum_{i'=1}^{\underline{i}[\ell',k']} \#\mathcal{T}_{\ell'} + \sum_{k'=1}^{k-1}\sum_{i'=1}^{\underline{i}[\ell,k']} \#\mathcal{T}_\ell + \sum_{i'=1}^{i} \#\mathcal{T}_\ell. \tag{4.19}$$

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

**Lemma 4.7.** *Independently of the adaptivity parameters $\theta$, $\lambda_{\mathrm{lin}}$, and $\lambda_{\mathrm{alg}}$, the $i$-loop of Algorithm 4.6 always terminates, i.e., $\underline{i}[\ell,k] < \infty$ for all $(\ell,k,0) \in \mathcal{Q}$.*

*Proof.* We argue as in [BIM+23, Lemma 3.2]. Let $(\ell,k,0) \in \mathcal{Q}$. We argue by contradiction and assume that the $i$-loop stopping criterion (4.16) in Algorithm 4.6(I.b.ii) always fails and hence $\underline{i}[\ell,k] = \infty$. By assumption (4.12), the algebraic solver $\Psi_\ell(u_\ell^{k,\star};\cdot)$ is contractive and hence convergent with limit $u_\ell^{k,\star} := \Phi_\ell(\delta; u_\ell^{k-1,\underline{i}})$ from Algorithm 4.6(I.a). Moreover, by failure of the stopping criterion (4.16) in Algorithm 4.6(I.b.ii), we thus obtain that

$$\eta_\ell(u_\ell^{k,i}) + |||u_\ell^{k,i} - u_\ell^{k,0}||| \overset{(4.16)}{\lesssim} |||u_\ell^{k,i} - u_\ell^{k,i-1}||| \xrightarrow{i\to\infty} 0.$$

This yields $|||u_\ell^{k,\star} - u_\ell^{k,0}||| = 0$ and hence $u_\ell^{k,\star} = u_\ell^{k,i}$ for all $i \in \mathbb{N}_0$m since the algebraic solver is contractive. Consequently, the $i$-loop stopping criterion (4.16) in Algorithm 4.6(I.b.ii) will be satisfied for $i = i_{\mathrm{min}}$. This contradicts our assumption, and hence we conclude that $\underline{i}[\ell,k] < \infty$. □

### 4.3.2 Energy contraction for the inexact Zarantonello iteration

In this section, we prove uniform boundedness of the iterates $u_\ell^{k,i}$ from Algorithm 4.6: Note that the algorithm does not compute the Zarantonello iterate $u_\ell^{k,\star} := \Phi_\ell(\delta; u_\ell^{k-1,\underline{i}})$ exactly, but relies on an approximation $u_\ell^{k,\underline{i}} \approx u_\ell^{k,\star}$. We prove that this inexact Zarantonello iteration is contractive with respect to the energy, which is the case if at least $i_{\mathrm{min}} \in \mathbb{N}$ steps of the contractive algebraic solver are performed, i.e., $\underline{i}[\ell,k] \ge i_{\mathrm{min}}$. In particular, a suitable choice of the damping parameter $\delta > 0$ and the index $i_{\mathrm{min}}$ are derived in the following.

**Theorem 4.8**

*Suppose that $\mathcal{A}$ satisfies (SM), (LIP), and (POT). With $M$ from (4.4), define $\tau := M + 3M\big(\frac{L[3M]}{\alpha}\big)^{1/2} \ge 4M$. Let $\lambda_{\mathrm{lin}}, \lambda_{\mathrm{alg}} > 0$ and $0 < \theta \le 1$ be arbitrary. Suppose that $|||u_\ell^{0,0}||| =$*

$|||u_\ell^{0,\underline{i}}||| \le 2M$ *with* $M > 0$ *from* (4.4). *Choose* $i_{\min} \in \mathbb{N}$ *such that*

$$q_{\mathrm{alg}}^{i_{\min}} \le 1/3. \tag{4.20}$$

*Then, for any choice of* $\delta > 0$ *satisfying* $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$, *there exists a uniform energy contraction constant* $0 < q_{\mathrm{E}} = q_{\mathrm{E}}[\delta,\tau] < 1$ *(see* (4.33b) *below) such that the following holds.*

- **nested iteration:** $|||u_\ell^{k,\underline{i}}||| \le 2M$ *if* $(\ell,\underline{k},\underline{i}) \in Q$; (4.21)
- $\underline{i}$**-uniform bound:** $|||u_\ell^{k,\underline{i}}||| \le \tau$ *if* $(\ell,k,\underline{i}) \in Q$; (4.22)
- $\mathcal{E}$**-contraction:** $\mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star) \le q_{\mathrm{E}}^2 \big(\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big)$ *if* $(\ell,k+1,\underline{i}) \in Q$. (4.23)

*With* (4.21)–(4.23), *we obtain for all iterates the*

- **uniform bound:** $|||u_\ell^{k,i}||| \le 5\tau$ *if* $(\ell,k,i) \in Q$. (4.24)

*Moreover, there exists an index* $k_0 = k_0[\delta,\tau,\alpha,L[3M],M] \in \mathbb{N}$ *independently of the mesh refinement index* $\ell$ *such that, for all* $k' \ge k_0$, *the nested iteration condition* $|||u_\ell^{k',\underline{i}}||| \le 2M$ *in the* $k$*-loop stopping criterion* (4.17) *is always met.*

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

**Lemma 4.9.** *Suppose that* $\mathcal{A}$ *satisfies* (SM), (LIP), *and* (POT). *Let* $\lambda_{\mathrm{lin}}, \lambda_{\mathrm{alg}} > 0$ *be arbitrary and define* $\tau := M + 3M\big(\frac{L[3M]}{\alpha}\big)^{1/2} \ge 4M$. *Let* $k \in \mathbb{N}_0$ *with* $0 \le k < \underline{k}[\ell]$ *and*

$$|||u_\ell^{k,\underline{i}}||| \le \tau. \tag{4.25}$$

*Then, for* $i_{\min} \in \mathbb{N}$ *satisfying* (4.20) *and for any* $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$, *it holds that*

$$\begin{aligned} 0 \le \Big(\frac{1}{2\delta} - \frac{L[5\tau]}{2}\Big) |||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}|||^2 &\le \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) \\ &\le \Big(\frac{1}{\delta\,(1-q_{\mathrm{alg}}^{i_{\min}})} - \frac{\alpha}{2}\Big) |||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}|||^2 \text{ for all } (\ell,k+1,\underline{i}) \in Q. \end{aligned} \tag{4.26}$$

*Proof.* The proof is subdivided into five steps.

**Step 1 (choice of $i_{\min}$).** We note that for any $i_{\min} \in \mathbb{N}$, the property (4.20) is indeed equivalent to

$$\frac{1}{2} \overset{!}{\le} \frac{1-2\,q_{\mathrm{alg}}^i}{1-q_{\mathrm{alg}}^i} \quad \text{for all } i \ge i_{\min}. \tag{4.27}$$

**Step 2 (boundedness).** Define $e_\ell^{k+1} := u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}$. Recall that for $0 < \delta < 2\alpha/L[2\tau]^2$, the Zarantonello iteration satisfies contraction (4.11). Hence, the contraction of the algebraic

solver (4.12), the triangle inequality, nested iteration $u_\ell^{k+1,0} = u_\ell^{k,\underline{i}}$, assumption (4.25), and $4M \le \tau$ show that

$$\begin{aligned}
|||e_\ell^{k+1}||| &\le |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| + |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \overset{(4.12)}{\le} q_{\mathrm{alg}}^{\underline{i}[\ell,k+1]}\, |||u_\ell^{k+1,\star} - u_\ell^{k+1,0}||| + |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \\
&\le 2\, |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \le 2\left[|||u_\ell^\star - u_\ell^{k,\underline{i}}||| + |||u_\ell^\star - u_\ell^{k+1,\star}|||\right] \\
&\overset{(4.11)}{\le} 2\,(1 + q_{\mathrm{Zar}}^\star[\delta, 2\tau])\, |||u_\ell^\star - u_\ell^{k,\underline{i}}||| \overset{(4.25)}{\le} 4(M+\tau) \le 5\tau.
\end{aligned}$$

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

$$|||u_\ell^{k+1,\underline{i}}||| \le \max_{0\le t\le 1} |||u_\ell^{k,\underline{i}} - t\, e_\ell^{k+1}||| \le 5\tau. \tag{4.28}$$

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

$$\varphi'(t) = \langle \mathrm{d}\mathcal{E}(u_\ell^{k,\underline{i}} + t\, e_\ell^{k+1}),\, e_\ell^{k+1}\rangle = \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\, e_\ell^{k+1}) - F,\, e_\ell^{k+1}\rangle. \tag{4.29}$$

The fundamental theorem of calculus and the exact Zarantonello iteration (4.10) show that

$$\begin{aligned}
\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) &= \varphi(0) - \varphi(1) = -\int_0^1 \varphi'(t)\,\mathrm{d}t \overset{(4.29)}{=} -\int_0^1 \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\, e_\ell^{k+1}) - F,\, e_\ell^{k+1}\rangle\,\mathrm{d}t \\
&= -\int_0^1 \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\, e_\ell^{k+1}) - \mathcal{A}(u_\ell^{k,\underline{i}}),\, e_\ell^{k+1}\rangle\,\mathrm{d}t - \langle \mathcal{A}(u_\ell^{k,\underline{i}}) - F,\, e_\ell^{k+1}\rangle \\
&\overset{(4.10)}{=} -\int_0^1 \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\, e_\ell^{k+1}) - \mathcal{A}(u_\ell^{k,\underline{i}}),\, e_\ell^{k+1}\rangle\,\mathrm{d}t + \frac{1}{\delta}\,\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}},\, e_\ell^{k+1}\rangle\!\rangle.
\end{aligned} \tag{4.30}$$

**Step 4 (proof of lower bound in (4.26)).** For any $i \in \mathbb{N}$ with $i \le \underline{i}[\ell, k]$, the contraction (4.12) of the algebraic solver and nested iteration $u_\ell^{k,\underline{i}} = u_\ell^{k+1,0}$ prove that

$$|||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| \overset{(4.12)}{\le} q_{\mathrm{alg}}^{\underline{i}[\ell,k+1]}\, |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \le q_{\mathrm{alg}}^i\, |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| + q_{\mathrm{alg}}^i\, |||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}|||.$$

This gives rise to the *a posteriori* estimate

$$|||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| \le \frac{q_{\mathrm{alg}}^i}{1 - q_{\mathrm{alg}}^i}\, |||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}||| = \frac{q_{\mathrm{alg}}^i}{1 - q_{\mathrm{alg}}^i}\, |||e_\ell^{k+1}|||. \tag{4.31}$$

With (4.31), $i_{\min} \le i \le \underline{i}[\ell, k+1]$, and (4.27), we derive

$$\begin{aligned}
&\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}},\, e_\ell^{k+1}\rangle\!\rangle = \langle\!\langle u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}},\, e_\ell^{k+1}\rangle\!\rangle + \langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}},\, e_\ell^{k+1}\rangle\!\rangle \\
&= |||e_\ell^{k+1}|||^2 + \langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}},\, e_\ell^{k+1}\rangle\!\rangle \ge |||e_\ell^{k+1}|||^2 - |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}|||\, |||e_\ell^{k+1}||| \\
&\overset{(4.31)}{\ge} |||e_\ell^{k+1}|||\left[|||e_\ell^{k+1}||| - \frac{q_{\mathrm{alg}}^i}{1-q_{\mathrm{alg}}^i}\, |||e_\ell^{k+1}|||\right] = \left(\frac{1-2\,q_{\mathrm{alg}}^i}{1-q_{\mathrm{alg}}^i}\right)|||e_\ell^{k+1}|||^2 \overset{(4.27)}{\ge} \frac{1}{2}\,|||e_\ell^{k+1}|||^2 \ge 0. \qquad (4.32)
\end{aligned}$$

With the local Lipschitz continuity (LIP) and (4.28), it follows from (4.30) that

$$\begin{aligned}
\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) &\overset{(\mathrm{LIP})}{\ge} -\Big(\int_0^1 tL[5\tau]\,\mathrm{d}t\Big)\,|||e_\ell^{k+1}|||^2 + \frac{1}{\delta}\,\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}},\, e_\ell^{k+1}\rangle\!\rangle \\
&\overset{(4.32)}{\ge} \left[\frac{1}{2\,\delta} - \frac{L[5\tau]}{2}\right]|||e_\ell^{k+1}|||^2.
\end{aligned}$$

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

**Step 5 (proof of upper bound in (4.26)).** To derive the upper equivalence constant, we infer from Step 4 that

$$\begin{aligned}
\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}},\, e_\ell^{k+1}\rangle\!\rangle &\le |||e_\ell^{k+1}|||^2 + |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}|||\, |||e_\ell^{k+1}||| \\
&\overset{(4.31)}{\le} |||e_\ell^{k+1}|||\left[|||e_\ell^{k+1}||| + \frac{q_{\mathrm{alg}}^i}{1-q_{\mathrm{alg}}^i}\, |||e_\ell^{k+1}|||\right] = \left(\frac{1}{1-q_{\mathrm{alg}}^i}\right)|||e_\ell^{k+1}|||^2.
\end{aligned}$$

Combined with Step 3, we obtain that

$$\begin{aligned}
\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) &\overset{(4.30)}{=} -\int_0^1 \langle \mathcal{A}(u_\ell^{k,\underline{i}} + t\,e_\ell^{k+1}) - \mathcal{A}(u_\ell^{k,\underline{i}}), e_\ell^{k+1}\rangle\,\mathrm{d}t + \frac{1}{\delta}\,\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}, e_\ell^{k+1}\rangle\!\rangle \\
&\overset{(\mathrm{SM})}{\le} -\Big(\int_0^1 t\alpha\,\mathrm{d}t\Big)\,|||e_\ell^{k+1}|||^2 + \frac{1}{\delta}\,\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}},\, e_\ell^{k+1}\rangle\!\rangle \le \left(\frac{1}{\delta\,(1-q_{\mathrm{alg}}^{i_{\min}})} - \frac{\alpha}{2}\right)|||e_\ell^{k+1}|||^2.
\end{aligned}$$

This concludes the proof. □

**Lemma 4.10** (energy contraction)**.** *Suppose the assumptions of Lemma 4.9. Recall $i_{\min} \in \mathbb{N}$ from (4.20). Then, for $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$, it holds that*

$$0 \le \mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star) \le q_{\mathrm{E}}[\delta,\tau]^2\,[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)] \qquad (4.33a)$$

*with the contraction constant*

$$0 \le q_{\mathrm{E}}[\delta,\tau]^2 := 1 - \left(\frac{1}{\delta} - L[5\tau]\right)\frac{(1-q_{\mathrm{alg}}^{i_{\min}})^2\,\delta^2\alpha^2}{L[2\tau]} < 1. \qquad (4.33b)$$

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

$$(1 - q_{\mathrm{E}}[\delta,\tau]^2)\,\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big] \le \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) \le \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star). \tag{4.34}$$

*Proof.* First, we observe that

$$\begin{aligned}\alpha\,|||u_\ell^\star - u_\ell^{k,\underline{i}}|||^2 &\overset{(\mathrm{SM})}{\le} \langle \mathcal{A}u_\ell^\star - \mathcal{A}u_\ell^{k,\underline{i}},\, u_\ell^\star - u_\ell^{k,\underline{i}}\rangle \overset{(4.3)}{=} \langle F - \mathcal{A}u_\ell^{k,\underline{i}},\, u_\ell^\star - u_\ell^{k,\underline{i}}\rangle \\ &\overset{(4.10)}{=} \frac{1}{\delta}\,\langle\!\langle u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}},\, u_\ell^\star - u_\ell^{k,\underline{i}}\rangle\!\rangle \le \frac{1}{\delta}\,|||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}|||\,|||u_\ell^\star - u_\ell^{k,\underline{i}}|||.\end{aligned} \tag{4.35}$$

The inverse triangle inequality and contraction (4.12) of the algebraic solver prove that

$$\begin{aligned}|||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}||| &\ge |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| - |||u_\ell^{k+1,\star} - u_\ell^{k+1,\underline{i}}||| \\ &\overset{(4.12)}{\ge} (1 - q_{\mathrm{alg}}^{i_{\min}})\,|||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}||| \overset{(4.35)}{\ge} (1 - q_{\mathrm{alg}}^{i_{\min}})\,\delta\,\alpha\,|||u_\ell^\star - u_\ell^{k,\underline{i}}|||.\end{aligned} \tag{4.36}$$

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

$$\begin{aligned}0 &\overset{(4.6)}{\le} \mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star) = \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star) - \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}})\big] \\ &\overset{(4.26)}{\le} \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star) - \Big(\frac{1}{2\delta} - \frac{L[5\tau]}{2}\Big)\,|||u_\ell^{k+1,\underline{i}} - u_\ell^{k,\underline{i}}|||^2 \\ &\overset{(4.36)}{\le} \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star) - \Big(\frac{1}{2\delta} - \frac{L[5\tau]}{2}\Big)\,(1 - q_{\mathrm{alg}}^{i_{\min}})^2\,\delta^2\,\alpha^2\,|||u_\ell^\star - u_\ell^{k,\underline{i}}|||^2 \\ &\overset{(4.6)}{\le} \Big(1 - \big[1 - \delta\,L[5\tau]\big]\,\frac{(1 - q_{\mathrm{alg}}^{i_{\min}})^2\,\alpha^2\,\delta}{L[2\tau]}\Big)\,\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big] \\ &=: q_{\mathrm{E}}[\delta,\tau]^2\,\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big].\end{aligned}$$

We may rewrite $q_{\mathrm{E}}[\delta,\tau]^2 = 1 - C\delta + C\,L[5\tau]\,\delta^2$ with $C = \frac{(1-q_{\mathrm{alg}}^{i_{\min}})^2\,\alpha^2}{L[2\tau]}$. Since $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\} \le \frac{1}{L[5\tau]}$, we obtain that $0 < q_{\mathrm{E}}[\delta,\tau] < 1$. This proves (4.33). The lower inequality in (4.34) follows from the triangle inequality. The upper inequality in (4.34) holds due to $0 \le \mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star)$ and hence $\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) = \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star) + \mathcal{E}(u_\ell^\star) - \mathcal{E}(u_\ell^{k+1,\underline{i}}) \le \mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)$. This concludes the proof. □

***Proof of Theorem 4.8.*** The proof consists of four steps.

**Step 1 (proof of (4.22)–(4.23) for $k = 0$ and all $\ell \in \mathbb{N}_0$)** Let $\ell \in \mathbb{N}_0$ with $\ell \le \underline{\ell}$ be arbitrary, but fixed. From the initial guess $u_0^{0,0}$ or Algorithm 4.6(I.c) and $u_\ell^{0,\underline{i}} = u_\ell^{0,0} = u_{\ell-1}^{\underline{k},\underline{i}}$ for any $\ell \in \mathbb{N}$, we have that $|||u_\ell^{0,0}||| \le 2M$ and *a fortiori* $|||u_\ell^{0,0}||| \le \tau$. This proves (4.22) for $k = 0$ and all $\ell \in \mathbb{N}_0$ with $\ell \le \underline{\ell}$ (even with the stronger bound $2M \le \tau$).

In particular, we may apply Lemma 4.10 to obtain that $\mathcal{E}(u_\ell^{1,\underline{i}}) - \mathcal{E}(u_\ell^\star) \le q_{\mathrm{E}}[\delta,\tau]^2\,\big[\mathcal{E}(u_\ell^{0,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]$, which proves (4.23) for $k = 0$ and $\ell \in \mathbb{N}_0$.

**Step 2 (proof of (4.22)–(4.23) for $k \ge 0$ and all $\ell \in \mathbb{N}_0$)** Let $\ell \in \mathbb{N}_0$ with $\ell \le \underline{\ell}$. We argue by induction on $k$, where Step 1 proves the base case $k = 0$. Hence, we may assume that boundedness (4.22) holds for all $0 \le k' \le k$. Lemma 4.10 applied separately for all

$0 \le k' \le k$ yields energy contraction (4.33) for the indices $0 \le k' \le k$. Overall, we obtain that

$$\mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star) \overset{(4.33)}{\le} q_{\mathrm{E}}[\delta,\tau]^2 \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big] \overset{(4.33)}{\le} q_{\mathrm{E}}[\delta,\tau]^{2(k+1)} \big[\mathcal{E}(u_\ell^{0,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big], \tag{4.37}$$

where we only used energy contraction (4.33) for $0 \le k' \le k$, i.e., for indices that are covered by the induction hypothesis. From (4.37), $|||u_\ell^\star||| \le M$ from (4.4), and $|||u_\ell^{0,\underline{i}}||| \le 2M$ and $u_\ell^{0,\underline{i}} = u_\ell^{0,0}$ from Step 1, we obtain that

$$\begin{aligned}
|||u_\ell^{k+1,\underline{i}}||| &\le |||u_\ell^\star||| + |||u_\ell^\star - u_\ell^{k+1,\underline{i}}||| \\
&\overset{(4.6)}{\le} M + \left(\frac{2}{\alpha}\right)^{1/2} \big[\mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \\
&\overset{(4.37)}{\le} M + q_{\mathrm{E}}^{k+1} \left(\frac{2}{\alpha}\right)^{1/2} \big[\mathcal{E}(u_\ell^{0,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \\
&\overset{(4.6)}{\le} M + q_{\mathrm{E}}^{k+1} \left(\frac{L[3M]}{\alpha}\right)^{1/2} |||u_\ell^\star - u_\ell^{0,\underline{i}}||| \le M + q_{\mathrm{E}}^{k+1} \left(\frac{L[3M]}{\alpha}\right)^{1/2} 3M \le \tau.
\end{aligned} \tag{4.38}$$

Thus, boundedness (4.22) is satisfied for $0 \le k' \le k+1$. Again, Lemma 8 yields energy contraction for $0 \le k' \le k+1$. This completes the induction argument and concludes that (4.22)–(4.23) hold for all $\ell \in \mathbb{N}_0$ and all $k \in \mathbb{N}_0$.

**Step 3 (uniform boundedness).** Contraction of the algebraic solver (4.12), the straightforward estimate from the exact Zarantonello iteration (4.10), $|||u^\star||| \le M \le \tau$ from (4.4), $|||u_\ell^{k,0}||| \le \tau$ from (4.22), and the constraint $\delta < \min\{1/L[5\tau], 2\alpha/L[2\tau]^2\}$ which ensures that $\delta L[2\tau] \le \delta L[5\tau] < 1$, yield that

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

With $|||u_\ell^{k,\star}||| \le |||u_\ell^{k,0}||| + |||u_\ell^{k,\star} - u_\ell^{k,0}||| \le 3\tau$ owing to (4.21), it follows that

$$|||u_\ell^{k,i}||| \overset{(4.12)}{\le} |||u_\ell^{k,\star}||| + q_{\mathrm{alg}}^i \, |||u_\ell^{k,\star} - u_\ell^{k,0}||| \le 5\tau \quad \text{for all } (\ell,k,i) \in Q.$$

**Step 4 (existence of $k_0$)** Let $\ell \in \mathbb{N}_0$ with $\ell \le \underline{\ell}$. As in (4.38) from Step 2, we obtain

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

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

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

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

**Remark 4.11.** (i) *According to uniform boundedness* (4.24)*, all involved Lipschitz constants or stability constants are uniformly bounded by* $L[10\tau]$ *and* $C_{\mathrm{stab}}[10\tau]$*, respectively.*

(ii) *Under the assumption that* $0 < \delta < \min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$*, energy contraction* (4.23) *and the lower bound in the norm-energy equivalence* (4.26) *are even equivalent, i.e.,*

$$(4.23) \quad \Longleftrightarrow \quad (4.26).$$

*To see this, recall that the proof of energy contraction* (4.23) *in Lemma 4.10 exploits* (4.26)*. The converse implication is obtained as follows: First, energy contraction yields*

$$\mathcal{E}(u_\ell^{k+1,\star}) - \mathcal{E}(u_\ell^\star) \le q_{\mathrm{E}}^2 \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big] = q_{\mathrm{E}}^2 \big\{\big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\star})\big] + \big[\mathcal{E}(u_\ell^{k+1,\star}) - \mathcal{E}(u_\ell^\star)\big]\big\} \tag{4.39}$$

*which gives rise to the a posteriori estimate*

$$0 \le \mathcal{E}(u_\ell^{k+1,\star}) - \mathcal{E}(u_\ell^\star) \le \frac{q_{\mathrm{E}}^2}{1-q_{\mathrm{E}}^2} \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\star})\big]. \tag{4.40}$$

*In particular, we note that the energy difference on the right-hand side is nonnegative. Exploiting uniform boundedness* (4.22)*, the last inequality yields that*

$$\begin{aligned} |||u_\ell^{k+1,\star} - u_\ell^{k,\underline{i}}|||^2 &\lesssim |||u_\ell^\star - u_\ell^{k+1,\star}|||^2 + |||u_\ell^\star - u_\ell^{k,\underline{i}}|||^2 \overset{(4.10)}{\le} \big(1 + (q_{\mathrm{Zar}}^\star[\delta, 2\tau])^2\big)\, |||u_\ell^\star - u_\ell^{k,\underline{i}}|||^2 \\ &\overset{(4.6)}{\lesssim} \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\star})\big] + \big[\mathcal{E}(u_\ell^{k+1,\star}) - \mathcal{E}(u_\ell^\star)\big] \overset{(4.40)}{\le} \frac{1}{1-q_{\mathrm{E}}^2} \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\star})\big]. \end{aligned}$$

*This concludes the argument.*

**Remark 4.12.** (i) *The stopping criteria* (4.16) *and* (4.17) *read schematically*

$$\texttt{[accuracy criterion]} \quad \texttt{AND} \quad \texttt{[iteration criterion]}.$$

(ii) *The accuracy criterion in* (4.17) *is heuristically motivated by the fact that the discretization error (estimated by* $\eta_\ell(\cdot)$*) shall dominate the linearization error*

$$\frac{\alpha}{2} |||u_\ell^\star - u_\ell^{k+1,\underline{i}}|||^2 \overset{(4.6)}{\le} \mathcal{E}(u_\ell^{k+1,\underline{i}}) - \mathcal{E}(u_\ell^\star) \overset{(4.23)}{\le} \frac{q_{\mathrm{E}}^2}{1-q_{\mathrm{E}}^2} \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^{k+1,\underline{i}})\big] \overset{(4.17)}{\lesssim} \lambda_{\mathrm{lin}}^2\, \eta_\ell(u_\ell^{k+1,\underline{i}})^2. \tag{4.41}$$

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

(iii) *The accuracy criterion* (4.16) *is satisfied given that the discretization and linearization error dominate the algebraic error in the sense of*

$$|||u_\ell^{k,\star} - u_\ell^{k,i}||| \overset{(4.12)}{\le} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} |||u_\ell^{k,i} - u_\ell^{k,i-1}||| \overset{(4.16)}{\le} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} \lambda_{\mathrm{alg}} \big[\lambda_{\mathrm{lin}}\, \eta_\ell(u_\ell^{k,i}) + |||u_\ell^{k,i} - u_\ell^{k,0}|||\big]. \tag{4.42}$$

*Once the* $i$*-loop is stopped, the equivalence* (4.26) *and nested iteration* $u_\ell^{k,0} = u_\ell^{k-1,\underline{i}}$ *yield* $|||u_\ell^{k,\underline{i}} - u_\ell^{k,0}|||^2 = |||u_\ell^{k,\underline{i}} - u_\ell^{k-1,\underline{i}}|||^2 \simeq \mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^{k,\underline{i}})$.

## 4.4 Full R-linear convergence

We prove full R-linear convergence of Algorithm 4.6 by adapting the analysis of [HPSV21; BFM+23]. The new result extends [②AIL1, Theorem 13], where an exact solve for the Zarantonello iteration (4.10) is supposed. The new proof is built on a summability argument, but the stopping criteria (4.16)–(4.17) with iteration count criteria require further analysis to prove full R-linear convergence even (and unlike [HPSV21; BFM+23]) for arbitrary adaptivity parameters $0 < \theta \le 1$, $\lambda_{\mathrm{lin}} > 0$ and $\lambda_{\mathrm{alg}} > 0$.

**Theorem 4.13: full R-linear convergence of Algorithm 4.6**

*Suppose the assumptions of Theorem 4.8. Suppose the axioms of adaptivity (A1)–(A3). Let $\lambda_{\mathrm{lin}}, \lambda_{\mathrm{alg}} > 0$, $0 < \theta \le 1$, $C_{\mathrm{mark}} \ge 1$, and $u_0^{0,0} \in \mathcal{X}_0$ with $|||u_0^{0,0}||| \le 2M$. Then, Algorithm 4.6 guarantees full R-linear convergence of the quasi-error*

$$\mathrm{H}_\ell^{k,i} := |||u_\ell^\star - u_\ell^{k,i}||| + |||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i}), \tag{4.43}$$

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

$$\mathrm{H}_\ell^{k,i} \le C_{\mathrm{lin}} q_{\mathrm{lin}}^{|\ell,k,i|-|\ell',k',i'|} \mathrm{H}_{\ell'}^{k',i'} \text{ for all } (\ell',k',i'), (\ell,k,i) \in \mathcal{Q} \text{ with } |\ell',k',i'| < |\ell,k,i|. \tag{4.44}$$

*The constant $q_{\mathrm{lin}}$ depends only on $\theta$, $q_{\mathrm{red}}$ from (A2), $q_{\mathrm{Zar}}^\star[\delta, 2\tau]$ from Proposition 4.4, $q_{\mathrm{E}}$ from Theorem 4.8, and $q_{\mathrm{alg}}$ from (4.12). The constant $C_{\mathrm{lin}}$ depends only on $M$, $\alpha$, $C_{\mathrm{Céa}}[2M]$, $q_{\mathrm{Zar}}^\star[\delta; 2\tau]$, $\lambda_{\mathrm{lin}}$, $q_{\mathrm{alg}}$, $\lambda_{\mathrm{alg}}$, $C_{\mathrm{rel}}$, $C_{\mathrm{stab}}[10\tau]$, and $i_{\mathrm{min}}$.*

***Proof of Theorem 4.13.*** The proof is split into seven steps.

**Step 1 (equivalences of quasi-error quantities).** Throughout the proof, we approach $\mathrm{H}_\ell^{k,i}$ from (4.43) after introducing auxiliary quantities such as

$$\mathrm{H}_\ell^k := [\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)]^{1/2} + \eta_\ell(u_\ell^{k,\underline{i}}) \quad \text{for all } (\ell,k,\underline{i}) \in \mathcal{Q} \tag{4.45}$$

and

$$\mathrm{H}_\ell := [\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_\ell^\star)]^{1/2} + \gamma\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}) \overset{(4.45)}{\simeq} \mathrm{H}_\ell^{\underline{k}} \quad \text{for all } (\ell,\underline{k},\underline{i}) \in \mathcal{Q}, \tag{4.46}$$

where $0 < \gamma < 1$ is a free parameter to be fixed later in (4.51) below. In the following, we show that $\mathrm{H}_\ell^{\underline{k},\underline{i}} \simeq \mathrm{H}_\ell^{\underline{k}} \overset{(4.46)}{\simeq} \mathrm{H}_\ell$. First, note that the equivalence of energy and norm from (4.6) (with $L[2\tau]$ from boundedness (4.22) and (4.4)) yields that

$$\mathrm{H}_\ell^k \le \mathrm{H}_\ell^k + |||u_\ell^{k,\star} - u_\ell^{k,\underline{i}}||| \overset{(4.6)}{\simeq} \mathrm{H}_\ell^{k,\underline{i}} \quad \text{for all } (\ell,k,\underline{i}) \in \mathcal{Q}. \tag{4.47}$$

The *a posteriori* estimate (4.42) for the algebraic solver from Remark 4.12(iii), norm-energy

equivalence (4.26), and the stopping criterion (4.17) show that

$$\begin{aligned}|||u_\ell^{\underline{k},\star} - u_\ell^{\underline{k},\underline{i}}||| &\overset{(4.42)}{\leq} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}}\,\lambda_{\mathrm{alg}}\,\big[\lambda_{\mathrm{lin}}\,\eta_\ell(u_\ell^{\underline{k},\underline{i}}) + |||u_\ell^{\underline{k},\underline{i}} - u_\ell^{\underline{k},0}|||\big]\\ &\overset{(4.26)}{\lesssim} \eta_\ell(u_\ell^{\underline{k},\underline{i}}) + \big[\mathcal{E}(u_\ell^{\underline{k},0}) - \mathcal{E}(u_\ell^{\underline{k},\underline{i}})\big]^{1/2} \overset{(4.17)}{\lesssim} \eta_\ell(u_\ell^{\underline{k},\underline{i}}) \leq \mathrm{H}_\ell^{\underline{k}}.\end{aligned}$$

With (4.47), we conclude that $\mathrm{H}_\ell \simeq \mathrm{H}_\ell^{\underline{k}} \simeq \mathrm{H}_\ell^{\underline{k},\underline{i}}$.

**Step 2 (estimator reduction).** The axioms (A1)–(A2) and Dörfler marking (4.18) prove the estimator reduction estimate (cf., e.g., [GHPS21, Equation (52)])

$$\eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{i}}) \leq q_\theta\,\eta_\ell(u_\ell^{\underline{k},\underline{i}}) + C_{\mathrm{stab}}[4M]\,|||u_{\ell+1}^{\underline{k},\underline{i}} - u_\ell^{\underline{k},\underline{i}}||| \quad \text{for all } \ell \in \mathbb{N}_0, \tag{4.48}$$

where $4M$ stems from nested iteration (4.21) from Theorem 4.8. Moreover, the triangle inequality, the equivalence (4.6), and energy contraction (4.23) give that

$$\begin{aligned}|||u_{\ell+1}^{\underline{k},\underline{i}} - u_\ell^{\underline{k},\underline{i}}||| &\leq |||u_{\ell+1}^{\star} - u_{\ell+1}^{\underline{k},\underline{i}}||| + |||u_{\ell+1}^{\star} - u_\ell^{\underline{k},\underline{i}}|||\\ &\overset{(4.6)}{\leq} \Big(\frac{2}{\alpha}\Big)^{1/2}\big[\mathcal{E}(u_{\ell+1}^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + \Big(\frac{2}{\alpha}\Big)^{1/2}\big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2}\\ &\overset{(4.23)}{\leq} (1 + q_{\mathrm{E}}^{\underline{k}[\ell+1]})\Big(\frac{2}{\alpha}\Big)^{1/2}\big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2}.\end{aligned}$$

Combined with the estimator reduction estimate (4.48) and with $1 + q_{\mathrm{E}} < 2$, we obtain with $C_1 := 2\,(2/\alpha)^{1/2}\,C_{\mathrm{stab}}[4M]$ that

$$\eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{i}}) \leq q_\theta\,\eta_\ell(u_\ell^{\underline{k},\underline{i}}) + C_1\,\big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} \quad \text{for all } 0 \leq \ell < \underline{\ell}. \tag{4.49}$$

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

$$\begin{aligned}&\mathrm{H}_{\ell+1} \overset{(4.46)}{=} \big[\mathcal{E}(u_{\ell+1}^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + \gamma\,\eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{i}}) \overset{(4.23)}{\leq} q_{\mathrm{E}}\,\big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + \gamma\,\eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{i}})\\ &\overset{(4.49)}{\leq} (q_{\mathrm{E}} + C_1\,\gamma)\,\big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + q_\theta\,\gamma\,\eta_\ell(u_\ell^{\underline{k},\underline{i}})\\ &\leq \max\{q_{\mathrm{E}}+C_1\,\gamma,\,q_\theta\}\,\big(\big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + \gamma\,\eta_\ell(u_\ell^{\underline{k},\underline{i}})\big) \quad \text{for all } (\ell+1,\underline{k},\underline{i}) \in \mathcal{Q}.\end{aligned} \tag{4.50}$$

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

$$0 < \widetilde{q} := \max\{q_{\mathrm{E}} + C_1\,\gamma,\,q_\theta\} < 1. \tag{4.51}$$

With the triangle inequality, (4.50) leads us to

$$\begin{aligned}a_{\ell+1} &:= \big[\mathcal{E}(u_{\ell+1}^{\underline{k},\underline{i}}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2} + \gamma\,\eta_{\ell+1}(u_{\ell+1}^{\underline{k},\underline{i}})\\ &\overset{(4.50)}{\leq} \widetilde{q}\,\big(\big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_\ell^{\star})\big]^{1/2} + \gamma\,\eta_\ell(u_\ell^{\underline{k},\underline{i}})\big) + \widetilde{q}\,\big[\mathcal{E}(u_\ell^{\star}) - \mathcal{E}(u_{\ell+1}^{\star})\big]^{1/2}\\ &=: \widetilde{q}\,a_\ell + b_\ell \quad \text{for all } (\ell,\underline{k},\underline{i}) \in \mathcal{Q}.\end{aligned} \tag{4.52}$$

By exploiting the equivalence (4.6) and stability (A1) (since all $u_\ell^{\underline{k},\underline{i}}$ are uniformly bounded by nested iteration (4.21)), the Céa lemma (4.5), and reliability (A3) prove that

$$\begin{aligned}\big[\mathcal{E}(u_{\ell'}^\star) - \mathcal{E}(u_{\ell''}^\star)\big]^{1/2} &\overset{(4.6)}{\simeq} |||u_{\ell''}^\star - u_{\ell'}^\star||| \overset{(4.5)}{\lesssim} |||u^\star - u_{\ell'}^\star||| \overset{(A3)}{\lesssim} \eta_{\ell'}(u_{\ell'}^\star) \overset{(A1)}{\lesssim} |||u_{\ell'}^\star - u_\ell^{\underline{k},\underline{i}}||| + \eta_\ell(u_\ell^{\underline{k},\underline{i}}) \\ &\overset{(4.6)}{\simeq} \big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \eta_\ell(u_\ell^{\underline{k},\underline{i}}) \simeq a_\ell \text{ for all } \ell \le \ell' \le \ell'' \le \underline{\ell} \text{ with } (\ell,\underline{k},\underline{i}) \in \mathcal{Q}.\end{aligned} \tag{4.53}$$

Hence, we infer that $b_{\ell+N} \lesssim a_\ell$ for all $0 \le \ell \le \ell+N \le \underline{\ell}$ with $(\ell,\underline{k},\underline{i}) \in \mathcal{Q}$, where the hidden stability constant $C_{\mathrm{stab}}[3M]$ depends on $3M$ due to (4.4) and nested iteration (4.21).

The energy $\mathcal{E}$ from (POT) (and its Pythagorean identity that leads to a telescoping sum) as well as the minimization property (4.7) for $\mathcal{X}_H = \mathcal{X}$ allow for the estimate

$$\begin{aligned}\sum_{\ell'=\ell}^{\ell+N-1} b_{\ell'}^2 \simeq \sum_{\ell'=\ell}^{\underline{\ell}-1} \big[\mathcal{E}(u_{\ell'}^\star) - \mathcal{E}(u_{\ell'+1}^\star)\big] \le \mathcal{E}(u_\ell^\star) - \mathcal{E}(u_{\underline{\ell}}^\star) &\overset{(4.7)}{\le} \mathcal{E}(u_\ell^\star) - \mathcal{E}(u^\star) \\ \overset{(4.6)}{\le} \frac{L[2M]}{2} |||u^\star - u_\ell^\star|||^2 \overset{(A3)}{\le} C_{\mathrm{rel}}^2 \frac{L[2M]}{2} \eta_\ell(u_\ell^\star)^2 &\overset{(4.53)}{\lesssim} a_\ell^2 \text{ for all } 0 \le \ell < \ell+N \le \underline{\ell},\end{aligned} \tag{4.54}$$

where the hidden stability constant $C_{\mathrm{stab}}$ depends on $3M$ due to (4.4) and nested iteration (4.21).

With (4.52)–(4.54), the assumptions for the tail summability criterion from [BFM+23, Lemma 6] are met. We thus conclude tail summability of $\mathrm{H}_{\ell+1} \simeq \mathrm{H}_\ell^{\underline{k}} \simeq a_\ell$, i.e.,

$$\boxed{\sum_{\ell'=\ell+1}^{\underline{\ell}-1} \mathrm{H}_{\ell'}^{\underline{k}} \lesssim \mathrm{H}_\ell^{\underline{k}} \quad \text{for all } 0 \le \ell < \underline{\ell}.} \tag{4.55}$$

**Step 4 (quasi-contraction in $k$).** We distinguish three cases.

**Case 4.1: Evaluation of (4.17) yields** `TRUE` $\wedge$ `FALSE`**.** This gives rise to

$$2M \overset{(4.17)}{<} |||u_\ell^{\underline{k},\underline{i}}||| \le |||u_\ell^\star||| + |||u_\ell^\star - u_\ell^{\underline{k},\underline{i}}||| \overset{(4.4)}{\le} M + |||u_\ell^\star - u_\ell^{\underline{k},\underline{i}}|||$$

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

$$\begin{aligned}1 = \frac{M}{M} < \frac{|||u_\ell^\star - u_\ell^{\underline{k},\underline{i}}|||}{M} &\overset{(4.6)}{\le} \frac{1}{M}\Big(\frac{2}{\alpha}\Big)^{1/2} \big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \\ &\overset{(4.23)}{\le} \frac{q_{\mathrm{E}}}{M}\Big(\frac{2}{\alpha}\Big)^{1/2} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2}.\end{aligned} \tag{4.56}$$

We recall from (4.4) that $|||u_\ell^\star||| \le M$ and $|||u_\ell^\star - u_0^\star||| \le 2M$ independently of $\ell$. Moreover, there holds quasi-monotonicity of the estimators in the sense that

$$\eta_\ell(u_\ell^\star) \le C_{\mathrm{mon}}\, \eta_0(u_0^\star) \quad \text{with } C_{\mathrm{mon}} = \big[2 + 8\, C_{\mathrm{stab}}[2M]^2 (1 + C_{\mathrm{Céa}}[2M]^2)\, C_{\mathrm{rel}}^2\big]^{1/2}; \tag{4.57}$$

cf. [CFPP14, Lemma 3.6] or [②AIL1, Equation 3.42] for the locally Lipschitz continuous set-

ting. In particular, estimate (4.57) holds also for the discrete limit space $\mathcal{X}_{\underline{\ell}} := \operatorname{closure}\big(\bigcup_{\ell=0}^{\underline{\ell}} \mathcal{X}_\ell\big)$. Additionally, we note that the estimate (4.57) admits

$$\eta_\ell(u_\ell^\star) \overset{(4.57)}{\leq} C_{\mathrm{mon}}\,\eta_0(u_0^\star) \overset{(A1)}{\leq} C_{\mathrm{mon}}\,\eta_0(0) + C_{\mathrm{mon}}\,C_{\mathrm{stab}}[M]\,|||u_0^\star||| \overset{(4.56)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2}. \tag{4.58}$$

The estimate (4.58), stability (A1) with stability constant $C_{\mathrm{stab}}[2\tau]$ due to (4.22) and (4.4), and energy contraction (4.23) yield that

$$\begin{aligned}\eta_\ell(u_\ell^{k,\underline{i}}) &\overset{(A1)}{\leq} \eta_\ell(u_\ell^\star) + C_{\mathrm{stab}}[2\tau]\,|||u_\ell^\star - u_\ell^{k,\underline{i}}||| \overset{(4.6)}{\lesssim} \eta_\ell(u_\ell^\star) + \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2}\\ &\overset{(4.58)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(4.23)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2}.\end{aligned} \tag{4.59}$$

For $0 \leq k' < k < \underline{k}[\ell]$, the definition (4.45), energy contraction (4.23), and (4.59) prove

$$\begin{aligned}\mathrm{H}_\ell^k &\overset{(4.23)}{\lesssim} q_{\mathrm{E}}\big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \eta_\ell(u_\ell^{k,\underline{i}}) \overset{(4.59)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2}\\ &\overset{(4.23)}{\lesssim} q_{\mathrm{E}}^{(k-1)-k'}\big[\mathcal{E}(u_\ell^{k',\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(4.45)}{\lesssim} q_{\mathrm{E}}^{k-k'}\,\mathrm{H}_\ell^{k'}.\end{aligned} \tag{4.60}$$

This concludes Case 4.1. ◇

**Case 4.2: Evaluation of (4.17) yields** `FALSE` ∧ `FALSE` **or** `FALSE` ∧ `TRUE`**.** For $0 \leq k' < k < \underline{k}[\ell]$, the definition (4.45), the failure of the accuracy condition in the stopping criterion for the inexact Zarantonello linearization (4.17), energy minimization (4.7), and energy contraction (4.23) prove that

$$\begin{aligned}\mathrm{H}_\ell^k &\overset{(4.17)}{<} \big[\mathcal{E}(u_\ell^{k,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + \lambda_{\mathrm{lin}}^{-1}\big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^{k,\underline{i}})\big]^{1/2}\\ &\overset{(4.7),\,(4.23)}{\lesssim} \big[\mathcal{E}(u_\ell^{k-1,\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(4.23)}{\lesssim} q_{\mathrm{E}}^{(k-1)-k'}\big[\mathcal{E}(u_\ell^{k',\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \overset{(4.45)}{\lesssim} q_{\mathrm{E}}^{k-k'}\,\mathrm{H}_\ell^{k'}.\end{aligned} \tag{4.61}$$

This concludes Case 4.2. ◇

**Case 4.3: Evaluation of (4.17) yields** `TRUE` ∧ `TRUE`**.** The equivalence (4.26), boundedness (4.22), and energy minimization (4.7) prove that

$$\begin{aligned}\mathrm{H}_\ell^{\underline{k}} &\overset{(A1)}{\lesssim} \big[\mathcal{E}(u_\ell^{\underline{k},\underline{i}}) - \mathcal{E}(u_\ell^\star)\big]^{1/2} + |||u_\ell^{\underline{k},\underline{i}} - u_\ell^{\underline{k}-1,\underline{i}}||| + \eta_\ell(u_\ell^{\underline{k}-1,\underline{i}})\\ &\overset{(4.26)}{\lesssim} \mathrm{H}_\ell^{\underline{k}-1} + \big[\mathcal{E}(u_\ell^{\underline{k}-1,\underline{i}}) - \mathcal{E}(u_\ell^{\underline{k},\underline{i}})\big]^{1/2} \overset{(4.7)}{\leq} 2\,\mathrm{H}_\ell^{\underline{k}-1} \quad \text{for all } (\ell,\underline{k},\underline{i}) \in \mathcal{Q}.\end{aligned} \tag{4.62}$$

Since $k = \underline{k}[\ell] - 1$ is covered by Case 4.1 or Case 4.2, estimate (4.62) leads to

$$\mathrm{H}_\ell^{\underline{k}} \overset{(4.62)}{\lesssim} \frac{q_{\mathrm{E}}}{q_{\mathrm{E}}}\,\mathrm{H}_\ell^{\underline{k}-1} \lesssim q_{\mathrm{E}}\,\mathrm{H}_\ell^{\underline{k}-1} \overset{(4.60),\,(4.61)}{\lesssim} q_{\mathrm{E}}^{\underline{k}[\ell]-k'}\,\mathrm{H}_\ell^{k',\underline{i}}. \tag{4.63}$$

This concludes Case 4.3. ◇

Overall, the estimates (4.60)–(4.61) and (4.63) result in

$$\boxed{\mathrm{H}_\ell^k \lesssim q_{\mathrm{E}}^{k-k'} \mathrm{H}_\ell^{k'} \quad \text{for all } (\ell,k,j) \in \mathcal{Q} \text{ with } 0 \le k' \le k \le \underline{k}[\ell],} \tag{4.64}$$

where the hidden constant depends only on $M$, $C_{\mathrm{stab}}[2\tau]$, $\alpha$, $L[2M]$, $C_{\mathrm{Céa}}[2M]$, $C_{\mathrm{rel}}$, $\lambda_{\mathrm{lin}}$, and $q_{\mathrm{E}}$. Furthermore, we recall from (4.53) that $\big[\mathcal{E}(u_{\ell-1}^\star) - \mathcal{E}(u_\ell^\star)\big]^{1/2} \lesssim \mathrm{H}_{\ell-1}^k$. Together with nested iteration $u_{\ell-1}^{\underline{k},\underline{i}} = u_\ell^{0,\underline{i}} = u_\ell^{0,\star}$, this yields that

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

and thus

$$\boxed{\mathrm{H}_\ell^0 \lesssim \mathrm{H}_{\ell-1}^{\underline{k}} \quad \text{for all } (\ell,0,0) \in \mathcal{Q} \text{ with } \ell \ge 1.} \tag{4.65}$$

**Step 5 (tail summability with respect to $\ell$ and $k$).** The estimates (4.64)–(4.65) from Step 4 as well as (4.55) from Step 3 and the geometric series prove that

$$\begin{aligned}\sum_{\substack{(\ell',k',\underline{i})\in\mathcal{Q}\\ |\ell',k',\underline{i}|>|\ell,k,\underline{i}|}} \mathrm{H}_{\ell'}^{k'} &= \sum_{k'=k+1}^{\underline{k}[\ell]} \mathrm{H}_\ell^{k'} + \sum_{\ell'=\ell+1}^{\underline{\ell}} \sum_{k'=0}^{\underline{k}[\ell]} \mathrm{H}_{\ell'}^{k'} \overset{(4.64)}{\lesssim} \mathrm{H}_\ell^k + \sum_{\ell'=\ell+1}^{\underline{\ell}} \mathrm{H}_{\ell'}^0 \\ &\overset{(4.65)}{\lesssim} \mathrm{H}_\ell^k + \sum_{\ell'=\ell}^{\underline{\ell}-1} \mathrm{H}_{\ell'}^{\underline{k}} \overset{(4.55)}{\lesssim} \mathrm{H}_\ell^k + \mathrm{H}_\ell^{\underline{k}} \overset{(4.64)}{\lesssim} \mathrm{H}_\ell^k \quad \text{for all } (\ell,k,\underline{i}) \in \mathcal{Q}.\end{aligned} \tag{4.66}$$

**Step 6 (contraction in $i$).** For $i = 0$ and $k = 0$, we recall that $u_\ell^{0,0} = u_\ell^{0,\underline{i}} = u_\ell^{0,\star}$ by definition and hence $\mathrm{H}_\ell^{0,0} \overset{(4.6)}{\simeq} \mathrm{H}_\ell^0$. For $k \ge 1$, nested iteration $u_\ell^{k,0} = u_\ell^{k-1,\underline{i}}$, contraction of the exact Zarantonello iteration (4.11), and energy equivalence (4.6) imply that

$$|||u_\ell^{k,\star} - u_\ell^{k,0}||| \le |||u_\ell^\star - u_\ell^{k,\star}||| + |||u_\ell^\star - u_\ell^{k-1,\underline{i}}||| \overset{(4.11)}{\le} (q_{\mathrm{Zar}}^\star[\delta;3M]+1)\,|||u_\ell^\star - u_\ell^{k-1,\underline{i}}||| \overset{(4.6)}{\lesssim} 2\,\mathrm{H}_\ell^{k-1}.$$

Therefore, by using the equivalence (4.6) once more, we obtain that

$$\boxed{\mathrm{H}_\ell^{k,0} \lesssim \mathrm{H}_\ell^{(k-1)_+} \quad \text{for all } (\ell,k,0) \in \mathcal{Q}, \quad \text{where } (k-1)_+ := \max\{0,k-1\}.} \tag{4.67}$$

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

$$\begin{aligned}\mathrm{H}_\ell^{k,i} &\overset{(4.43)}{=} |||u_\ell^\star - u_\ell^{k,i}||| + |||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i}) \\ &\overset{(A1)}{\le} \mathrm{H}_\ell^{k,i-1} + (2 + C_{\mathrm{stab}}[10\tau])\,|||u_\ell^{k,i} - u_\ell^{k,i-1}||| \\ &\overset{(4.12)}{\le} \mathrm{H}_\ell^{k,i-1} + (2 + C_{\mathrm{stab}}[10\tau])(q_{\mathrm{alg}}+1)\,|||u_\ell^{k,\star} - u_\ell^{k,i-1}||| \overset{(4.43)}{\lesssim} \mathrm{H}_\ell^{k,i-1},\end{aligned} \tag{4.68}$$

where $C_{\mathrm{stab}}[10\tau]$ stems from the uniform bound (4.24) from Theorem 4.8. Hence, we

obtain

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

For all $0\le i'<i_{\min}\le i<\underline{i}[\ell,k]$, we obtain with an *a posteriori* estimate based on the contraction of the Zarantonello iteration (4.11) (where $q_{\mathrm{Zar}}^\star = q_{\mathrm{Zar}}^\star[\delta,2\tau]$ depends on $\tau$ from (4.22)), the *a posteriori* estimate (4.42) for the algebraic solver, the failure of the accuracy criterion of (4.16), and the contraction of the algebraic solver (4.12) that

$$\begin{aligned}
\mathrm{H}_\ell^{k,i} &\overset{(4.43)}{=} |||u_\ell^\star - u_\ell^{k,i}||| + |||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i}) \le |||u_\ell^\star - u_\ell^{k,\star}||| + 2\,|||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i})\\
&\le \frac{q_{\mathrm{Zar}}^\star[\delta;2\tau]}{1-q_{\mathrm{Zar}}^\star[\delta;2\tau]}\,|||u_\ell^{k,i} - u_\ell^{k-1,\underline{i}}||| + \Big(2+\frac{q_{\mathrm{Zar}}^\star[\delta;2\tau]}{1-q_{\mathrm{Zar}}^\star[\delta;2\tau]}\Big)\,|||u_\ell^{k,\star} - u_\ell^{k,i}||| + \eta_\ell(u_\ell^{k,i})\\
&\overset{(4.42)}{\lesssim} |||u_\ell^{k,i} - u_\ell^{k-1,\underline{i}}||| + |||u_\ell^{k,i} - u_\ell^{k,i-1}||| + \eta_\ell(u_\ell^{k,i})\\
&\overset{(4.16)}{\lesssim} |||u_\ell^{k,i} - u_\ell^{k,i-1}||| \overset{(4.12)}{\lesssim} |||u_\ell^{k,\star} - u_\ell^{k,i-1}||| \overset{(4.12)}{\lesssim} q_{\mathrm{alg}}^{i-i'}\,|||u_\ell^{k,\star} - u_\ell^{k,i'}||| \le q_{\mathrm{alg}}^{i-i'}\,\mathrm{H}_\ell^{k,i'},
\end{aligned} \tag{4.69}$$

Altogether, the combination of (4.68)–(4.69) proves that

$$\boxed{\mathrm{H}_\ell^{k,i} \lesssim q_{\mathrm{alg}}^{i-i'}\,\mathrm{H}_\ell^{k,i'} \quad \text{for all } (\ell,k,i)\in\mathcal{Q} \quad \text{with} \quad 0\le i'\le i\le \underline{i}[\ell,k],} \tag{4.70}$$

where the hidden constant depends only on $q_{\mathrm{Zar}}^\star[\delta;2\tau]$, $q_{\mathrm{alg}}$, $\lambda_{\mathrm{alg}}$, $C_{\mathrm{stab}}[10\tau]$, and $i_{\min}$.

**Step 7 (tail summability with respect to $\ell$, $k$, and $i$).** Finally, we observe that

$$\begin{aligned}
\sum_{\substack{(\ell',k',i')\in\mathcal{Q}\\ |\ell',k',i'|>|\ell,k,i|}} \mathrm{H}_{\ell'}^{k',i'} &= \sum_{i'=i+1}^{\underline{i}[\ell,k]} \mathrm{H}_\ell^{k,i'} + \sum_{k'=k+1}^{\underline{k}[\ell]}\sum_{i'=0}^{\underline{i}[\ell,k']} \mathrm{H}_\ell^{k',i'} + \sum_{\ell'=\ell+1}^{\underline{\ell}}\sum_{k'=0}^{\underline{k}[\ell']}\sum_{i'=0}^{\underline{i}[\ell',k']} \mathrm{H}_{\ell'}^{k',i'}\\
&\overset{(4.70)}{\lesssim} \mathrm{H}_\ell^{k,i} + \sum_{k'=k+1}^{\underline{k}[\ell]} \mathrm{H}_\ell^{k',0} + \sum_{\ell'=\ell+1}^{\underline{\ell}}\sum_{k'=0}^{\underline{k}[\ell]} \mathrm{H}_{\ell'}^{k',0} \overset{(4.67)}{\lesssim} \mathrm{H}_\ell^{k,i} + \sum_{\substack{(\ell',k',\underline{i})\in\mathcal{Q}\\ |\ell',k',\underline{i}|>|\ell,k,\underline{i}|}} \mathrm{H}_{\ell'}^{k'}\\
&\overset{(4.66)}{\lesssim} \mathrm{H}_\ell^{k,i} + \mathrm{H}_\ell^{k} \overset{(4.47)}{\lesssim} \mathrm{H}_\ell^{k,i} + \mathrm{H}_\ell^{k,\underline{i}} \overset{(4.70)}{\lesssim} \mathrm{H}_\ell^{k,i} \quad \text{for all } (\ell,k,i)\in\mathcal{Q}.
\end{aligned}$$

Since $\mathcal{Q}$ is countable and linearly ordered, [CFPP14, Lemma 4.9] applies and proves R-linear convergence (4.44) of $\mathrm{H}_\ell^{k,i}$. This concludes the proof. □

Given full R-linear convergence from Theorem 4.13, then convergence rates with respect to the degrees of freedom coincide with rates with respect to the overall computational cost, where we recall $\mathtt{cost}(\ell,k,i)$ from (4.19). Since all essential arguments are provided, the proof follows verbatim from [BFM+23, Corollary 16].

**Corollary 4.14** (rates ≅ complexity)**.** *Suppose full R-linear convergence* (4.44)*. Recall*

$\texttt{cost}(\ell,k,i)$ *from* (4.19). *Then, for any* $s>0$, *it holds that*

$$M(s) := \sup_{(\ell,k,i)\in\mathcal{Q}} (\#\mathcal{T}_\ell)^s\, \mathrm{H}_\ell^{k,i} \le \sup_{(\ell,k,i)\in\mathcal{Q}} \texttt{cost}(\ell,k,i)^s\, \mathrm{H}_\ell^{k,i} \le C_{\mathrm{cost}}\, M(s), \tag{4.71}$$

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

## 4.5 Optimal complexity

A formal approach to optimal complexity relies on the notion of approximation classes [BDD04; Ste07; CKNS08; CFPP14], which reads as follows: For $s>0$, define

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

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

Finally, we are in the position to present the third main result of this paper, namely optimal complexity of Algorithm 4.6. Its proof relies, in essence, on perturbation arguments. More precisely, sufficiently small $\theta$ and $\lambda_{\mathrm{lin}}$ are required to ensure that Algorithm 4.6 guarantees convergence rate $s$ with respect to the overall computational cost (and time) if the solution $u^\star$ of (4.2) can be approximated at rate $s$ in the sense of $\|u^\star\|_{\mathbb{A}_s}<\infty$.

**Theorem 4.15: optimal complexity**

*Define* $\tau := M + 3M\big(\frac{L[3M]}{\alpha}\big)^{1/2} \ge 4M$ *with* $M$ *from* (4.4). *Let* $0<\delta<\min\{\frac{1}{L[5\tau]}, \frac{2\alpha}{L[2\tau]^2}\}$ *to ensure validity of Theorem 4.8. Define*

$$\lambda_{\mathrm{lin}}^\star := \min\Big\{1, \Big(\frac{\alpha\,(1-q_{\mathrm{E}}^2)}{2\,q_{\mathrm{E}}^2}\Big)^{1/2} / C_{\mathrm{stab}}[3M]\Big\}. \tag{4.72}$$

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

$$0<\theta_{\mathrm{mark}} := \frac{(\theta^{1/2}+\lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)^2}{(1-\lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)^2} < \theta^\star := (1+C_{\mathrm{stab}}[2M]^2\,C_{\mathrm{rel}}^2)^{-1} < 1. \tag{4.73}$$

*Then, Algorithm 4.6 guarantees, for all* $s>0$, *that*

$$\sup_{(\ell,k,j)\in\mathcal{Q}} \texttt{cost}(\ell,k,i)^s\, \mathrm{H}_\ell^{k,j} \le C_{\mathrm{opt}}\, \max\{\|u^\star\|_{\mathbb{A}_s}, \mathrm{H}_0^{0,0}\}. \tag{4.74}$$

*The constant* $C_{\mathrm{opt}}>0$ *depends only on* $q_{\mathrm{E}}$, $\alpha$, $C_{\mathrm{stab}}[10\tau]$, $C_{\mathrm{rel}}$, $C_{\mathrm{drel}}$, $C_{\mathrm{mark}}$, $C_{\mathrm{mesh}}$, $C_{\mathrm{lin}}$, $q_{\mathrm{lin}}$, $\#\mathcal{T}_0$, *and* $s$. *In particular, there holds optimal complexity of Algorithm 4.6.*

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

**Lemma 4.16** (estimator equivalence)**.** *Suppose the assumptions of Theorem 4.8. Recall* $\lambda_{\mathrm{lin}}^\star$ *from* (4.72)*. Then, for all* $(\ell, \underline{k}, \underline{i}) \in \mathcal{Q}$ *with* $\underline{k}[\ell] < \infty$*, it holds that*

$$\eta_\ell(u_\ell^\star) \le (1 + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}), \tag{4.75a}$$

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

$$(1 - \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}) \le \eta_\ell(u_\ell^\star). \tag{4.75b}$$

*For* $0 < \lambda_{\mathrm{lin}} < \lambda_{\mathrm{lin}}^\star$*, Dörfler marking for* $u_\ell^\star$ *with parameter* $\theta_{\mathrm{mark}}$ *from* (4.73) *implies Dörfler marking for* $u_\ell^{\underline{k},\underline{i}}$ *with parameter* $\theta$*, i.e., for any* $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$*, there holds the implication*

$$\theta_{\mathrm{mark}}\, \eta_\ell(u_\ell^\star)^2 \le \eta_\ell(\mathcal{R}_\ell; u_\ell^\star)^2 \quad \Longrightarrow \quad \theta\, \eta_\ell(u_\ell^{\underline{k},\underline{i}})^2 \le \eta_\ell(\mathcal{R}_\ell; u_\ell^{\underline{k},\underline{i}})^2. \tag{4.76}$$

*Proof.* The proof consists of two steps.

**Step 1.** First, we obtain from Remark 4.12(ii) that

$$\frac{\alpha}{2}\, |||u_\ell^\star - u_\ell^{\underline{k},\underline{i}}|||^2 \overset{(4.41)}{\le} \frac{\lambda_{\mathrm{lin}}^2\, q_{\mathrm{E}}^2}{1 - q_{\mathrm{E}}^2}\, \eta_\ell(u_\ell^{\underline{k},\underline{i}})^2.$$

Exploiting this together with stability (A1), nested iteration (4.23), and boundedness of the exact discrete solution (4.4), we obtain for any $\mathcal{U}_\ell \subseteq \mathcal{T}_\ell$ that

$$\begin{aligned} \eta_\ell(\mathcal{U}_\ell; u_\ell^\star) &\overset{(A1)}{\le} \eta_\ell(\mathcal{U}_\ell; u_\ell^{\underline{k},\underline{i}}) + C_{\mathrm{stab}}[3M]\, |||u_\ell^\star - u_\ell^{\underline{k},\underline{i}}||| \\ &\overset{(4.41)}{\le} \eta_\ell(\mathcal{U}_\ell; u_\ell^{\underline{k},\underline{i}}) + \lambda_{\mathrm{lin}}\, C_{\mathrm{stab}}[3M] \Big(\frac{2\, q_{\mathrm{E}}^2}{\alpha\, (1 - q_{\mathrm{E}}^2)}\Big)^{1/2} \eta_\ell(u_\ell^{\underline{k},\underline{i}}) \\ &= \eta_\ell(\mathcal{U}_\ell; u_\ell^{\underline{k},\underline{i}}) + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}). \end{aligned} \tag{4.77}$$

The choice $\mathcal{U}_\ell = \mathcal{T}_\ell$ yields (4.75a). The same arguments prove that

$$\eta_\ell(\mathcal{U}_\ell; u_\ell^{\underline{k},\underline{i}}) \le \eta_\ell(\mathcal{U}_\ell; u_\ell^\star) + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}). \tag{4.78}$$

For $0 < \lambda_{\mathrm{lin}} < \lambda_{\mathrm{lin}}^\star$ and $\mathcal{U}_\ell = \mathcal{T}_\ell$, the rearrangement of (4.78) proves (4.75b).

**Step 2.** Let $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$ satisfy $\theta_{\mathrm{mark}}^{1/2}\, \eta_\ell(u_\ell^\star) \le \eta_\ell(\mathcal{R}_\ell; u_\ell^\star)$. Then, (4.77)–(4.78) prove

$$\begin{aligned} &\big[1 - \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star\big]\, \theta_{\mathrm{mark}}^{1/2}\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}) \overset{(4.75b)}{\le} \theta_{\mathrm{mark}}^{1/2}\, \eta_\ell(u_\ell^\star) \le \eta_\ell(\mathcal{R}_\ell; u_\ell^\star) \\ &\overset{(4.77)}{\le} \eta_\ell(\mathcal{R}_\ell; u_\ell^{\underline{k},\underline{i}}) + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}) \overset{(4.73)}{=} \eta_\ell(\mathcal{R}_\ell; u_\ell^{\underline{k},\underline{i}}) + \big[\theta_{\mathrm{mark}}^{1/2}\, (1 - \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star) - \theta^{1/2}\big]\, \eta_\ell(u_\ell^{\underline{k},\underline{i}}). \end{aligned}$$

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

***Proof of Theorem 4.15.*** By Corollary 4.14, it is enough to show

$$\sup_{(\ell,k,i) \in \mathcal{Q}} \big(\#\mathcal{T}_\ell\big)^s\, \mathrm{H}_\ell^{k,i} \lesssim \max\{\|u^\star\|_{\mathbb{A}_s}, \mathrm{H}_0^{0,0}\}. \tag{4.79}$$

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

**Step 1.** Let $0 < \theta_{\mathrm{mark}} := (\theta^{1/2} + \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)^2\,(1 - \lambda_{\mathrm{lin}}/\lambda_{\mathrm{lin}}^\star)^{-2} < \theta^\star := (1 + C_{\mathrm{stab}}[2M]^2\,C_{\mathrm{rel}}^2)^{-1}$ and fix any $0 \le \ell' \le \underline{\ell} - 1$. The validity of (A4) and [CFPP14, Lemma 4.14] guarantee the existence of a set $\mathcal{R}_{\ell'} \subseteq \mathcal{T}_{\ell'}$ with $0 \le \ell' \le \underline{\ell} - 1$ such that

$$\begin{aligned} \#\mathcal{R}_{\ell'} &\lesssim \|u^\star\|_{\mathbb{A}_s}^{1/s}\,[\eta_{\ell'}(u_{\ell'}^\star)]^{-1/s}, \\ \theta_{\mathrm{mark}}\,\eta_{\ell'}(u_{\ell'}^\star) &\le \eta_{\ell'}(\mathcal{R}_{\ell'}, u_{\ell'}^\star), \end{aligned} \tag{4.80}$$

where the hidden constant depends only on (A1)–(A4). By means of (4.76) in Lemma 4.16, we infer that $\mathcal{R}_{\ell'}$ satisfies the Dörfler marking (4.18) in Algorithm 4.6 with $\theta$, i.e., $\theta\,\eta_{\ell'}(u_{\ell'}^{k,i})^2 \le \eta_{\ell'}(\mathcal{R}_{\ell'}; u_{\ell'}^{k,i})^2$. Hence, since $0 < \theta < \theta_{\mathrm{mark}} < \theta^\star$, the optimality of Dörfler marking proves

$$\#\mathcal{M}_{\ell'} \le C_{\mathrm{mark}}\,\#\mathcal{R}_{\ell'} \overset{(4.80)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\big[\eta_{\ell'}(u_{\ell'}^\star)\big]^{-1/s}. \tag{4.81}$$

Moreover, full R-linear convergence (4.44) together with *a posteriori* error estimates for the final iterates (4.41) and (4.42) which use the stopping criteria (4.16)–(4.17), norm-energy equivalence (4.26)m and estimator equivalence (4.75) prove

$$\begin{aligned} \mathrm{H}_{\ell'+1}^{0,\underline{i}} &\overset{(4.44)}{\lesssim} \mathrm{H}_{\ell'}^{\underline{k},\underline{i}} \overset{(4.43)}{=} |||u_{\ell'}^\star - u_{\ell'}^{\underline{k},\underline{i}}||| + |||u_{\ell'}^{\underline{k},\star} - u_{\ell'}^{\underline{k},\underline{i}}||| + \eta_{\ell'}(u_{\ell'}^{\underline{k},\underline{i}}) \\ &\overset{(4.42),\,(4.26)}{\lesssim} |||u_{\ell'}^\star - u_{\ell'}^{\underline{k},\underline{i}}||| + [\mathcal{E}(u_{\ell'}^{\underline{k},0}) - \mathcal{E}(u_{\ell'}^{\underline{k},\underline{i}})]^{1/2} + \eta_{\ell'}(u_{\ell'}^{\underline{k},\underline{i}}) \\ &\overset{(4.41),\,(4.17)}{\lesssim} \eta_{\ell'}(u_{\ell'}^{\underline{k},\underline{i}}) \overset{(4.75)}{\lesssim} \eta_{\ell'}(u_{\ell'}^\star). \end{aligned} \tag{4.82}$$

Consequently, a combination of (4.81) and (4.82) concludes that

$$\#\mathcal{M}_{\ell'} \overset{(4.81)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\big[\eta_{\ell'}(u_{\ell'}^\star)\big]^{-1/s} \overset{(4.82)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\big[\mathrm{H}_{\ell'+1}^{0,\underline{i}}\big]^{-1/s}. \tag{4.83}$$

**Step 2.** For $(\ell, k, i) \in \mathcal{Q}$, full R-linear convergence (4.44) and the geometric series prove

$$\sum_{\substack{(\ell',k',i')\in\mathcal{Q}\\ |\ell',k',i'|\le|\ell,k,i|}} (\mathrm{H}_{\ell'}^{k',i'})^{-1/s} \overset{(4.44)}{\lesssim} (\mathrm{H}_\ell^{k,i})^{-1/s} \sum_{\substack{(\ell',k',i')\in\mathcal{Q}\\ |\ell',k',i'|\le|\ell,k,i|}} (q_{\mathrm{lin}}^{1/s})^{|\ell,k,i|-|\ell',k',i'|} \lesssim (\mathrm{H}_\ell^{k,i})^{-1/s}. \tag{4.84}$$

We recall the mesh-closure estimate [BDD04; Ste08; KPP13; DGS23]

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 \le C_{\mathrm{mesh}} \sum_{\ell'=0}^{\ell-1} \#\mathcal{M}_{\ell'} \quad \text{for all } \ell \ge 0, \tag{4.85}$$

where $C_{\mathrm{mesh}} > 1$ depends only on $\mathcal{T}_0$ and hence in particular on the dimension $d$. For

$(\ell,k,i)\in\mathcal{Q}$, the preceding estimates show that

$$\#\mathcal{T}_\ell-\#\mathcal{T}_0 \overset{(4.85)}{\lesssim} \sum_{\ell'=0}^{\ell-1}\#\mathcal{M}_{\ell'}$$
$$\overset{(4.83)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}\sum_{\ell'=0}^{\ell-1}\big(\mathrm{H}^{0,\underline{i}}_{\ell'+1}\big)^{-1/s} \le \|u^\star\|_{\mathbb{A}_s}^{1/s}\sum_{\substack{(\ell',k',i')\in\mathcal{Q}\\ |\ell',k',i'|\le|\ell,k,i|}}(\mathrm{H}^{k',i'}_{\ell'})^{-1/s} \overset{(4.84)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}(\mathrm{H}^{k,i}_\ell)^{-1/s}.$$

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

$$(\#\mathcal{T}_\ell-\#\mathcal{T}_0+1)^s\,\mathrm{H}^{k,i}_\ell\lesssim\|u^\star\|_{\mathbb{A}_s}\quad\text{for all }(\ell,k,i)\in\mathcal{Q}\text{ with }\ell\ge1. \tag{4.86a}$$

Theorem 4.13 proves that

$$(\#\mathcal{T}_\ell-\#\mathcal{T}_\ell+1)^s\,\mathrm{H}^{k,i}_\ell=\mathrm{H}^{k,i}_0\overset{(4.44)}{\lesssim}\mathrm{H}^{0,0}_0\quad\text{for all }(\ell,k,i)\in\mathcal{Q}\text{ with }\ell=0. \tag{4.86b}$$

For all $\mathcal{T}_\ell\in\mathbb{T}$, elementary calculation [BHP17, Lemma 22] shows that

$$\#\mathcal{T}_\ell-\#\mathcal{T}_0+1\le\#\mathcal{T}_\ell\le\#\mathcal{T}_0\,(\#\mathcal{T}_\ell-\#\mathcal{T}_0+1). \tag{4.87}$$

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

$$(\#\mathcal{T}_\ell)^s\,\mathrm{H}^{k,i}_\ell\overset{(4.87)}{\lesssim}(\#\mathcal{T}_\ell-\#\mathcal{T}_0+1)^s\,\mathrm{H}^{k,i}_\ell\overset{(4.86)}{\lesssim}\max\{\|u^\star\|_{\mathbb{A}_s},\mathrm{H}^{0,0}_0\mid.\}$$

This concludes the proof of (4.79). □

## 4.6 Numerical experiments

The experiments are performed with the open-source software package MooAFEM [IP23]. In the following, Algorithm 4.6 employs the optimal local *hp*-robust multigrid method [IMPS23] as algebraic solver. We remark that in our implementation the condition (4.20) is slightly relaxed to $|\mathcal{E}(u^{k,0}_\ell)-\mathcal{E}(u^{k,i}_\ell)|<10^{-12}=:\mathtt{tol}$.

**Experiment 4.17** (modified sine-Gordon equation [AHW23, Experiment 5.1])**.** *For $\Omega=(0,1)^2$, we consider*

$$-\Delta u^\star+(u^\star)^3+\sin(u^\star)=f\quad\text{in }\Omega\quad\text{subject to}\quad u^\star=0\text{ on }\partial\Omega \tag{4.88}$$

*with the monotone semilinearity $b(v)=v^3+\sin(v)$, which fits into the locally Lipschitz continuous framework (cf. [②AIL1, Experiment 3.28]). We choose $f$ such that*

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

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

$$\eta_H(T,v_H)^2:=h_T^2\,\|f+\Delta v_H-b(v_H)\|^2_{L^2(T)}+h_T\,\|[\![\nabla v_H\cdot\boldsymbol{n}]\!]\|^2_{L^2(\partial T\cap\Omega)}. \tag{4.89}$$

| $\delta = 0.3$ | $\theta = 0.1$ | | | | | $\theta = 0.2$ | | | | | $\theta = 0.3$ | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| $\lambda_{\text{alg}}$ \ $\lambda_{\text{lin}}$ | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 | 0.1 | 0.3 | 0.5 | 0.7 | 0.9 |
| 0.1 | 1306 | 650 | 660 | 660 | 660 | 735 | 639 | 347 | 347 | 347 | 724 | 659 | 373 | 373 | 373 |
| 0.3 | 928 | 660 | 660 | 660 | 660 | 545 | 269 | 269 | 269 | 269 | 505 | 333 | 241 | 241 | 241 |
| 0.5 | 654 | 654 | 654 | 654 | 654 | 534 | 274 | 274 | 273 | 273 | 462 | 278 | 262 | 262 | 262 |
| 0.7 | 649 | 617 | 617 | 617 | 617 | 293 | 262 | 262 | 262 | 262 | 420 | 298 | 259 | 259 | 259 |
| 0.9 | 676 | 646 | 646 | 646 | 646 | 268 | 269 | 269 | 269 | 269 | 422 | 321 | 247 | 247 | 247 |
| | $\theta = 0.4$ | | | | | $\theta = 0.5$ | | | | | $\theta = 0.6$ | | | | |
| 0.1 | 807 | 643 | 357 | 357 | 357 | 816 | 658 | 337 | 350 | 350 | 882 | 600 | 332 | 361 | 361 |
| 0.3 | 533 | 375 | 252 | 252 | 252 | 532 | 448 | 266 | 266 | 266 | 663 | 466 | 293 | 293 | 293 |
| 0.5 | 464 | 346 | 253 | 253 | 253 | 572 | 399 | 278 | 278 | 278 | 643 | 389 | 292 | 292 | 292 |
| 0.7 | 487 | 377 | 247 | 247 | 247 | 573 | 427 | 293 | 293 | 293 | 606 | 402 | 296 | 296 | 296 |
| 0.9 | 502 | 390 | 264 | 264 | 264 | 520 | 417 | 288 | 288 | 288 | 563 | 512 | 288 | 288 | 288 |
| | $\theta = 0.7$ | | | | | $\theta = 0.8$ | | | | | $\theta = 0.9$ | | | | |
| 0.1 | 856 | 634 | 361 | 337 | 337 | 985 | 741 | 413 | 375 | 375 | 1028 | 710 | 466 | 344 | 344 |
| 0.3 | 663 | 457 | 321 | 321 | 321 | 673 | 471 | 328 | 328 | 328 | 735 | 551 | 349 | 349 | 349 |
| 0.5 | 705 | 446 | 299 | 299 | 299 | 638 | 452 | 340 | 340 | 340 | 700 | 542 | 374 | 374 | 374 |
| 0.7 | 630 | 541 | 338 | 338 | 338 | 752 | 518 | 343 | 343 | 343 | 680 | 586 | 352 | 352 | 352 |
| 0.9 | 639 | 518 | 347 | 347 | 347 | 770 | 579 | 373 | 373 | 373 | 722 | 667 | 367 | 367 | 367 |

**Table 4.1:** The weighted cost (4.90) of the sine-Gordon problem (4.88) for different adaptivity parameters $\lambda_{\text{lin}}, \lambda_{\text{alg}}, \theta \in \{0.1, 0.2, \dots, 0.9\}$ and fixed damping parameter $\delta = 0.3$, where the mesh refinement is stopped if $\eta_\ell(u_\ell^{\underline{k},\underline{i}}) < 10^{-4}$, where the $\theta$-blockwise minimal values are highlighted in green and the overall minimal value in red.

*For $p = 2$, damping parameter $\delta = 0.3$, and $i_{\min} = 1$, we stop the computation as soon as $\eta_\ell(u_\ell^{\underline{k},\underline{i}}) < 10^{-4}$. Table 4.1 depicts the values of the weighted cost*

$$\eta_\ell(u_\ell^{\underline{k},\underline{i}})\,\mathtt{cost}(\ell,\underline{k},\underline{i})^{p/2} \tag{4.90}$$

*to determine the best parameter choice. We observe that the parameters $\theta \in \{0.3, 0.4\}$ and $\lambda_{\text{lin}} \geq 0.5$ perform comparably well. The parameter $\lambda_{\text{alg}}$ may be used for fine-tuning, but for moderate $\theta \in \{0.2, 0.3, 0.4, 0.5, 0.6\}$ and as soon as $\lambda_{\text{lin}}$ is set, the influence is comparably low.*

*For the following experiments, we set $\delta = 0.3$, $\theta = 0.3$, $\lambda_{\text{lin}} = 0.7$, and $\lambda_{\text{alg}} = 0.3$. Figure 4.2 depicts the error $|||u^\star - u_\ell^{\underline{k},\underline{i}}|||$ and the estimator $\eta_\ell(u_\ell^{\underline{k},\underline{i}})$ over $\mathtt{cost}(\ell,\underline{k},\underline{i})$ (left) and over the cumulative time in seconds (right) for the displayed polynomial degrees $p \in \{1, 2, 3\}$. In both plots, the decay rate is of (expected) optimal order $p/2$ for $p \in \{1, 2, 3\}$.*

**Experiment 4.18.** *We consider a globally Lipschitz continuous example from [HPW21,*

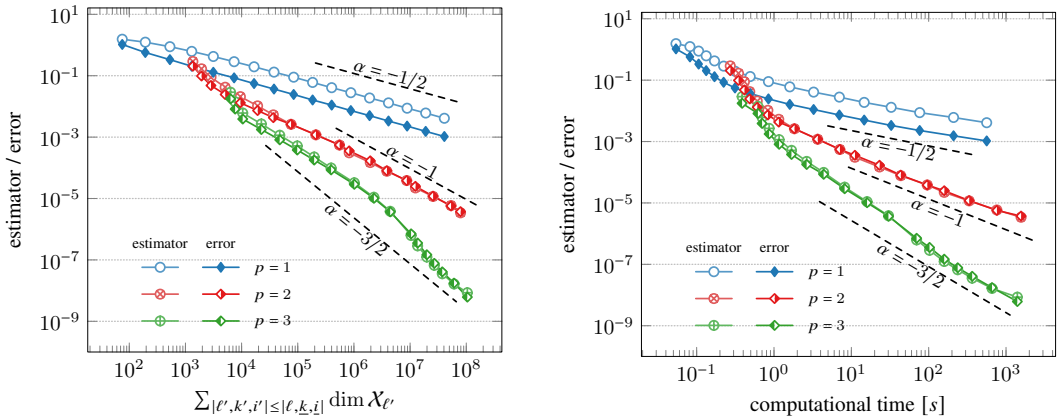



**Figure 4.2:** Experiment 4.17: Convergence plots of the error $|||u^\star - u_\ell^{\underline{k},\underline{i}}|||$ (diamond) and the error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{i}})$ (circle) over $\mathtt{cost}(\ell, \underline{k}, \underline{i})$ (left) and over computational time in seconds (right).

*Section 5.3] with Lipschitz constant $L = 2$ and monotonicity constant $\alpha = 1 - 2\exp(-\frac{3}{2})$ and hence $\delta = \alpha/L^2 \approx 0.138434919925785$ is a viable choice. For $d = 2$ and the L-shaped domain $(-1,1)^2 \setminus ([0,1] \times [-1,0]) \subset \mathbb{R}^2$, we seek $u^\star \in H^1_0(\Omega)$ such that*

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

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

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

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

*In Figure 4.3, we plot a sample solution (right) as well as convergence results of various error components (left) over the degrees of freedom. We observe that after a preasymptotic phase, optimal convergence rate $-1/2$ is restored for the exact error (diamond), the quasi-error $\mathrm{H}_\ell^{\underline{k},\underline{i}}$, the linearization error $\mathcal{E}(u_\ell^{\underline{k},0}) - \mathcal{E}(u_\ell^{\underline{k},\underline{i}})$ (triangle), and the error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{i}})$ (circle).*

**Experiment 4.19** (singularly perturbed sine-Gordon equation)**.** *This example is a variant of [AHW23, Experiment 5.2]. For $d = 2$ and the L-shaped domain $(-1,1)^2 \setminus ([0,1]\times[-1,0]) \subset \mathbb{R}^2$, let $\varepsilon = 10^{-5}$ and consider*

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

*with the monotone semilinearity $b(v) = v^3 + \sin(v)$. In this case, the exact solution $u^\star$ is unknown. We use the energy norm $|||\cdot|||^2 = \varepsilon\langle\nabla\cdot, \nabla\cdot\rangle + \langle\cdot,\cdot\rangle$. The experiment is conducted with damping parameter $\delta = 0.1$, $\lambda_{\mathrm{alg}} = 0.7$, $\theta = 0.3$, and $i_{\min} = 1$. The refinement*

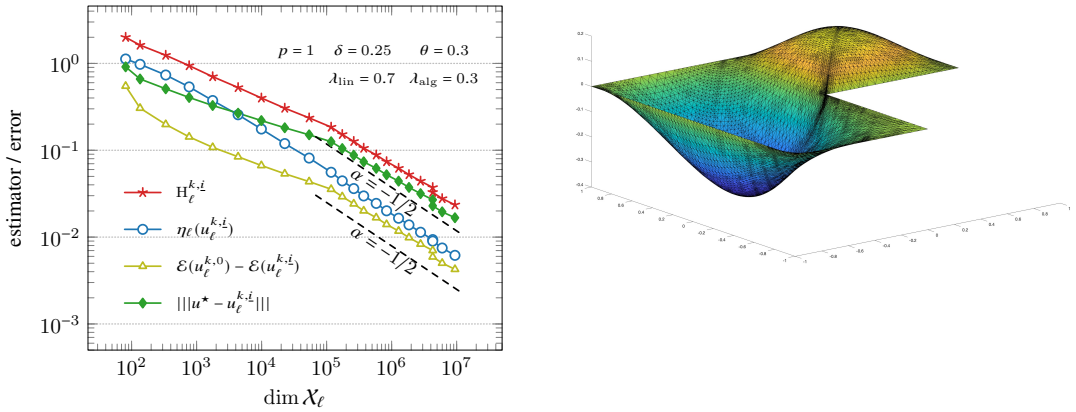



**Figure 4.3:** Experiment 4.18: Convergence plots of various error components over the degrees of freedom (left). Right: Plot of the approximate solution $u_{13}^{1,1}$ on $\mathcal{X}_{13}$ with $\#\mathcal{X}_{13} = 10209$.

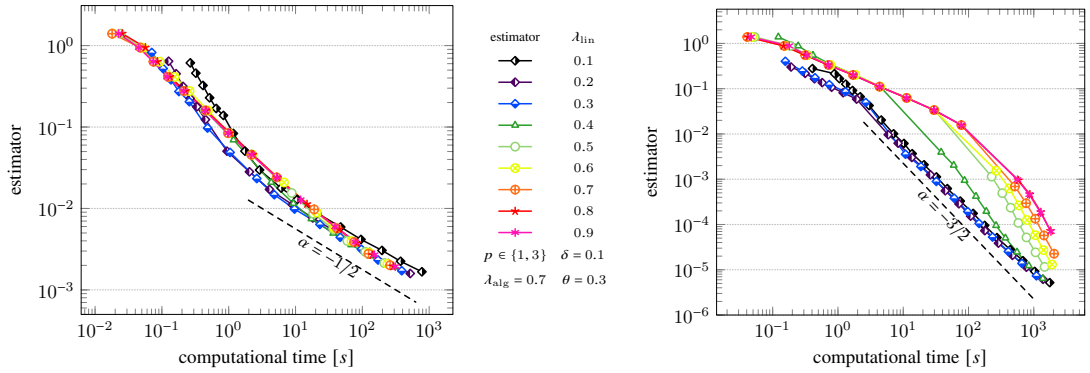



**Figure 4.4:** Convergence plots of the error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{i}})$ over computational time of Experiment 4.19. Left: Convergence plot for $p = 1$ Right: Convergence plot for $p = 3$.

*indicator* (4.89) *is modified along the lines of [Ver13, Remark 4.14] to*

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

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

*In Figure 4.4, we plot the error estimator* $\eta_\ell(u_\ell^{\underline{k},\underline{i}})$ *for all* $(\ell, \underline{k}, \underline{i}) \in \mathcal{Q}$ *against the computational time for* $\lambda_{\mathrm{lin}} \in \{0.1, 0.2, ..., 0.9\}$ *and polynomial degrees* $p \in \{1, 3\}$*. The decay rate is of (expected) optimal order* $p/2$*. The choice of* $\lambda_{\mathrm{lin}}$ *does not play a major role in Figure 4.4 (left) for* $p = 1$*, but significantly prolongs the preasymptotic phase for* $p = 3$*; see Figure 4.4 (right). Figure 4.5 shows meshes with* $\#\mathtt{nDof} = 12475$ *for* $\lambda_{\mathrm{lin}} = 0.2$ *and* $\#\,\mathtt{nDof} = 12152$ *for* $\lambda_{\mathrm{lin}} = 0.7$*. We see that* $\lambda_{\mathrm{lin}} = 0.7$ *causes refinement in the interior, since less local smoothing steps are performed. This experiment shows that Algorithm 4.6 is suitable for a setting with dominating reaction given that a suitable norm on* $\mathcal{X}$ *is chosen. A large choice of* $\lambda_{\mathrm{lin}}$ *seems*

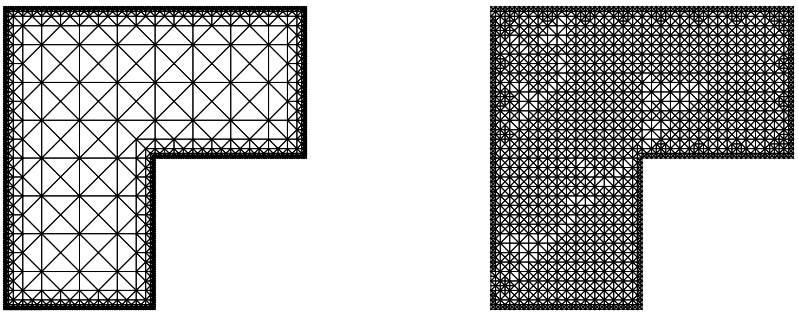


**Figure 4.5:** Mesh plot of Experiment 4.19 for $p = 3$. Left: Adaptivity parameter $\lambda_{\text{lin}} = 0.2$. Right: Adaptivity parameter $\lambda_{\text{lin}} = 0.7$.

*possible, but pays off only after a long preasymptotic phase.*

# Bibliography


[①GOA] 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.*, 118:18–35, 2022. DOI: 10.1016/j.camwa.2022.05.008.

[②AIL1] R. Becker, M. Brunner, M. Innerberger, J. M. Melenk, and D. Praetorius. Cost-optimal adaptive iterative linearized FEM for semilinear elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 57(4):2193–2225, 2023. DOI: 10.1051/m2an/2023036.

[③AIL2] M. Brunner, D. Praetorius, and J. Streitberger. Cost-optimal adaptive FEM with linearization and algebraic solver for semilinear elliptic PDEs, 2024. arXiv: 2401.06486.

[AESW22] K. Ahuja, B. Endtmayer, M. C. Steinbach, and T. Wick. Multigoal-oriented error estimation and mesh adaptivity for fluid-structure interaction. *J. Comput. Appl. Math.*, 412:Paper No. 114315, 18, 2022. DOI: 10.1016/j.cam.2022.114315.

[AFF+13] 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(3):305–332, 2013. DOI: 10.1515/cmam-2013-0010.

[AGL13] M. Arioli, E. H. Georgoulis, and D. Loghin. Stopping criteria for adaptive finite element solvers. *SIAM J. Sci. Comput.*, 35(3):A1537–A1559, 2013. DOI: 10.1137/120867421.

[AHW23] M. Amrein, P. Heid, and T. P. Wihler. A numerical energy reduction approach for semilinear diffusion-reaction boundary value problems based on steady-state iterations. *SIAM J. Numer. Anal.*, 61(2):755–783, 2023. DOI: 10.1137/22M1478586.

[ALMS13] M. Arioli, J. Liesen, A. Miçdlar, and Z. Strakoš. Interplay between discretization and algebraic computation in adaptive numerical solution of elliptic PDE problems. *GAMM-Mitt.*, 36(1):102–129, 2013. DOI: 10.1002/gamm.201310006.

[Alt16] H. W. Alt. *Linear functional analysis. Springer London.* first edition, 2016.

[AO00] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*. Wiley-Interscience [John Wiley & Sons], New York, 2000. DOI: 10.1002/9781118032824.

[AW15] M. Amrein and T. P. Wihler. Fully adaptive Newton-Galerkin methods for semilinear elliptic partial differential equations. *SIAM J. Sci. Comput.*, 37(4):A1637–A1657, 2015. DOI: 10.1137/140983537.

[BBPS23] P. Bringmann, M. Brunner, D. Praetorius, and J. Streitberger. Optimal complexity of goal-oriented adaptive FEM for nonsymmetric linear elliptic PDEs, 2023. arXiv: 2312.00489.

[BDD04] P. Binev, W. Dahmen, and R. DeVore. Adaptive finite element methods with convergence rates. *Numer. Math.*, 97(2):219–268, 2004. DOI: 10.1007/s00211-003-0492-7.

[BDK12] L. Belenki, L. Diening, and C. Kreuzer. Optimality of an adaptive finite element method for the $p$-Laplacian equation. *IMA J. Numer. Anal.*, 32(2):484–510, 2012. DOI: 10.1093/imanum/drr016.

[BET11] R. Becker, E. Estecahandy, and D. Trujillo. Weighted marking for goal-oriented adaptive finite element methods. *SIAM J. Numer. Anal.*, 49(6):2451–2469, 2011. DOI: 10.1137/100794298.

[BFM+23] P. Bringmann, M. Feischl, A. Miraci, D. Praetorius, and J. Streitberger. On full linear convergence and optimal complexity of adaptive FEM with inexact solver, 2023. arXiv: 2311.15738.

[BGIP23] R. Becker, G. Gantner, M. Innerberger, and D. Praetorius. Goal-oriented adaptive finite element methods with optimal computational complexity. *Numer. Math.*, 153(1):111–140, 2023. DOI: 10.1007/s00211-022-01334-8.

[BHP17] 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. Engrg.*, 317:318–340, 2017. DOI: 10.1016/j.cma.2016.12.014.

[BHSZ11] 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*, 2011. arXiv: 1108.3661 [math.NA].

[BIM+23] M. Brunner, M. Innerberger, A. Miraçi, D. Praetorius, J. Streitberger, and P. Heid. Adaptive FEM with quasi-optimal overall cost for nonsymmetric linear elliptic PDEs. *IMA J. Numer. Anal.*, 2023. DOI: 10.1093/imanum/drad039. Corrigendum to: Adaptive FEM with quasi-optimal overall cost for nonsymmetric linear elliptic PDEs. *IMA J. Numer. Anal.*, 2024. DOI: 10.1093/imanum/drad103.

[BIP21] R. Becker, M. Innerberger, and D. Praetorius. Optimal convergence rates for goal-oriented FEM with quadratic goal functional. *Comput. Meth. Appl. Math.*, 21(2):267–288, 2021.

[BMS10] R. Becker, S. Mao, and Z. Shi. A convergent nonconforming adaptive finite element method with quasi-optimal complexity. *SIAM J. Numer. Anal.*, 47(6):4639–4659, 2010. DOI: 10.1137/070701479.

[BMZ21] I. Brevis, I. Muga, and K. G. van der Zee. A machine-learning minimal-residual (ML-MRes) framework for goal-oriented finite element discretizations. *Comput. Math. Appl.*, 95:186–199, 2021. DOI: 10.1016/j.camwa.2020.08.012.

[BR01] R. Becker and R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta Numer.*, 10:1–102, 2001. DOI: 10.1017/S0962492901000010.

[BR03] W. Bangerth and R. Rannacher. *Adaptive finite element methods for differential equations*. Birkhäuser, Basel, 2003. DOI: 10.1007/978-3-0348-7605-6. URL: https://doi.org/10.1007/978-3-0348-7605-6.

[BR79] I. Babuška and W. C. Rheinboldt. Adaptive approaches and reliability estimations in finite element analysis. *Comput. Methods Appl. Mech. Engrg.*, 17/18(part):519–540, 1979. DOI: 10.1016/0045-7825(79)90042-2.

[BV84] I. Bubuška and M. Vogelius. Feedback and adaptive finite element solution of one-dimensional boundary value problems. *Numer. Math.*, 44(1):75–102, 1984.

[CDD01] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comp.*, 70(233):27–75, 2001. DOI: 10.1090/S0025-5718-00-01252-7.

[CDD03] A. Cohen, W. Dahmen, and R. Devore. Adaptive wavelet schemes for nonlinear variational problems. *SIAM J. Numer. Anal.*, 41(5):1785–1823, 2003. DOI: 10.1137/S0036142902412269.

[CFPP14] C. Carstensen, M. Feischl, M. Page, and D. Praetorius. Axioms of adaptivity. *Comput. Math. Appl.*, 67(6):1195–1253, 2014. DOI: 10.1016/j.camwa.2013.12.003.

[CG12] C. Carstensen and J. Gedicke. An adaptive finite element eigenvalue solver of asymptotic quasi-optimal computational complexity. *SIAM J. Numer. Anal.*, 50(3):1029–1057, 2012. DOI: 10.1137/090769430.

[Chi09] M. Chipot. *Elliptic equations: an introductory course*. Birkhäuser, Basel, 2009.

[CKNS08] 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(5):2524–2550, 2008. DOI: 10.1137/07069047X.

[CN12] J. M. Cascón and R. H. Nochetto. Quasioptimal cardinality of AFEM driven by nonresidual estimators. *IMA J. Numer. Anal.*, 32(1):1–29, 2012. DOI: 10.1093/imanum/drr014.

[CNX12] L. Chen, R. H. Nochetto, and J. Xu. Optimal multilevel methods for graded bisection grids. *Numer. Math.*, 120(1):1–34, 2012. DOI: 10.1007/s00211-011-0401-4.

[CW17] S. Congreve and T. P. Wihler. Iterative Galerkin Discretizations for Strongly Monotone Problems. *J. Comput. Appl. Math.*, 311:457–472, 2017. DOI: 10.1016/j.cam.2016.08.014.

[DBR21] V. Dolejší, O. Bartoš, and F. Roskovec. Goal-oriented mesh adaptation method for nonlinear problems including algebraic errors. *Comput. Math. Appl.*, 93:178–198, 2021. DOI: 10.1016/j.camwa.2021.04.004.

[DGS23] L. Diening, L. Gehring, and J. Storn. Adaptive Mesh Refinement for arbitrary initial Triangulations, 2023. arXiv: 2306.02674.

[DK08] L. Diening and C. Kreuzer. Linear convergence of an adaptive finite element method for the $p$-Laplacian equation. *SIAM J. Numer. Anal.*, 46(2):614–638, 2008. DOI: 10.1137/070681508.

[Dör96] W. Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996. DOI: 10.1137/0733054.

[EEHJ95] K. Eriksson, D. Estep, P. Hansbo, and C. Johnson. Introduction to adaptive methods for differential equations. In Acta Numer. Pages 105–158. Cambridge Univ. Press, Cambridge, 1995. DOI: 10.1017/S0962492900002531.

[EEV11] 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. Engrg.*, 200(37-40):2782–2795, 2011. DOI: 10.1016/j.cma.2010.03.024.

[EG04] A. Ern and J.-L. Guermond. *Theory and practice of finite elements*. Springer-Verlag, New York, 2004. DOI: 10.1007/978-1-4757-4355-5.

[ELW19] B. Endtmayer, U. Langer, and T. Wick. Multigoal-oriented error estimates for non-linear problems. *J. Numer. Math.*, 27(4):215–236, 2019. DOI: 10.1515/jnma-2018-0038.

[ELW20] B. Endtmayer, U. Langer, and T. Wick. Two-side a posteriori error estimates for the dual-weighted residual method. *SIAM J. Sci. Comput.*, 42(1):A371–A394, 2020. DOI: 10.1137/18M1227275.

[EV13] A. Ern and M. Vohralík. Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. *SIAM J. Sci. Comput.*, 35(4):A1761–A1791, 2013. DOI: 10.1137/120896918.

[Fei22] M. Feischl. Inf-sup stability implies quasi-orthogonality. *Math. Comp.*, 91 (337):2059–2094, 2022. DOI: 10.1090/mcom/3748.

[FFGHP16] M. Feischl, T. Führer, G. Gantner, A. Haberl, and D. Praetorius. Adaptive boundary element methods for optimal convergence of point errors. *Numer. Math.*, 132(3):541–567, 2016. DOI: 10.1007/s00211-015-0727-4.

[FFP14] 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(2):601–625, 2014. DOI: 10.1137/120897225.

[FK80] S. Fučík and A. Kufner. *Nonlinear differential equations*. Elsevier, Amsterdam, 1980.

[FPZ16] M. Feischl, D. Praetorius, and K. G. van der Zee. An abstract analysis of optimal goal-oriented adaptivity. *SIAM J. Numer. Anal.*, 54(3):1423–1448, 2016. DOI: 10.1137/15M1021982.

[GHPS18] G. Gantner, A. Haberl, D. Praetorius, and B. Stiftner. Rate optimal adaptive FEM with inexact solver for nonlinear operators. *IMA J. Numer. Anal.*, 38(4):1797–1831, 2018. DOI: 10.1093/imanum/drx050.

[GHPS21] G. Gantner, A. Haberl, D. Praetorius, and S. Schimanko. Rate optimality of adaptive finite element methods with respect to overall computational costs. *Math. Comp.*, 90(331):2011–2040, 2021. DOI: 10.1090/mcom/3654.

[GMZ11] E. M. Garau, P. Morin, and C. Zuppa. Convergence of an adaptive Kačanov FEM for quasi-linear problems. *Appl. Numer. Math.*, 61(4):512–529, 2011. DOI: 10.1016/j.apnum.2010.12.001.

[GMZ12] 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(2):131–156, 2012. DOI: 10.4208/nmtma.2012.m1023.

[Gri11] P. Grisvard. *Elliptic problems in nonsmooth domains*. SIAM, Philadelphia, Pa, 2011. DOI: 10.1137/1.9781611972030.ch1.

[GS02] M. B. Giles and E. Süli. Adjoint methods for PDEs: a posteriori error analysis and postprocessing by duality. *Acta Numer.*, 11:145–236, 2002. DOI: 10.1017/S096249290200003X.

[GSS14] D. Gallistl, M. Schedensack, and R. P. Stevenson. A remark on newest vertex bisection in any space dimension. *Comput. Methods Appl. Math.*, 14(3):317–320, 2014. DOI: 10.1515/cmam-2014-0013.

[HP16] M. Holst and S. Pollock. Convergence of goal-oriented adaptive finite element methods for nonsymmetric problems. *Numer. Methods Partial Differential Equations*, 32(2):479–509, 2016. DOI: 10.1002/num.22002.

[HPSV21] 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(3):679–725, 2021. DOI: 10.1007/s00211-021-01176-w.

[HPW21] 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(2):407–422, 2021. DOI: 10.1515/cmam-2021-0025.

[HPZ15] M. Holst, S. Pollock, and Y. Zhu. Convergence of goal-oriented adaptive finite element methods for semilinear problems. *Comput. Vis. Sci.*, 17(1):43–63, 2015. DOI: 10.1007/s00791-015-0243-1.

[HW18] P. Houston and T. P. Wihler. An *hp*-adaptive Newton-discontinuous-Galerkin finite element approach for semilinear elliptic boundary value problems. *Math. Comp.*, 87(314):2641–2674, 2018. DOI: 10.1090/mcom/3308.

[HW20a] P. Heid and T. P. Wihler. Adaptive iterative linearization Galerkin methods for nonlinear problems. *Math. Comp.*, 89(326):2707–2734, 2020. DOI: 10.1090/mcom/3545.

[HW20b] P. Heid and T. P. Wihler. On the convergence of adaptive iterative linearized Galerkin methods. *Calcolo*, 57(3):Paper No. 24, 23, 2020. DOI: 10.1007/s10092-020-00368-4.

[IMPS23] M. Innerberger, A. Miraçi, D. Praetorius, and J. Streitberger. *hp*-robust multigrid solver on locally refined meshes for FEM discretizations of symmetric elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 2023. DOI: 10.1051/m2an/2023104.

[IP23] 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. DOI: 10.1016/j.amc.2022.127731.

[KJF77] A. Kufner, O. John, and S. Fučík. *Function spaces*. Noordhoff, Academia, Leyden, Prague, 1977.

[KPP13] M. Karkulik, D. Pavlicek, and D. Praetorius. On 2D newest vertex bisection: optimality of mesh-closure and $H^1$-stability of $L_2$-projection. *Constr. Approx.*, 38(2):213–234, 2013. DOI: 10.1007/s00365-013-9192-4.

[KS11] C. Kreuzer and K. G. Siebert. Decay rates of adaptive finite elements with Dörfler marking. *Numer. Math.*, 117(4):679–716, 2011. DOI: 10.1007/s00211-010-0324-5.

[KVD19] B. Keith, A. Vaziri Astaneh, and L. F. Demkowicz. Goal-oriented adaptive mesh refinement for discontinuous Petrov-Galerkin methods. *SIAM J. Numer. Anal.*, 57(4):1649–1676, 2019. DOI: 10.1137/18M1181754.

[Mau95] J. M. Maubach. Local bisection refinement for $n$-simplicial grids generated by reflection. *SIAM J. Sci. Comput.*, 16(1):210–227, 1995. DOI: 10.1137/0916014.

[Mit91] W. F. Mitchell. Adaptive refinement for arbitrary finite-element spaces with hierarchical bases. *J. Comput. Appl. Math.*, 36(1):65–78, 1991. DOI: 10.1016/0377-0427(91)90226-A.

[MNS00] P. Morin, R. H. Nochetto, and K. G. Siebert. Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.*, 38(2):466–488, 2000. DOI: 10.1137/S0036142999360044.

[MS09] M. S. Mommer and R. Stevenson. A goal-oriented adaptive finite element method with convergence rates. *SIAM J. Numer. Anal.*, 47(2):861–886, 2009. DOI: 10.1137/060675666.

[MSV08] P. Morin, K. G. Siebert, and A. Veeser. A basic convergence result for conforming adaptive finite elements. *Math. Models Methods Appl. Sci.*, 18(5):707–737, 2008. DOI: 10.1142/S0218202508002838.

[PP20] C.-M. Pfeiler and D. Praetorius. Dörfler marking with minimal cardinality is a linear complexity problem. *Math. Comp.*, 89(326):2735–2752, 2020. DOI: 10.1090/mcom/3553.

[Sew72] E. G. Sewell. *Automatic generation of triangulations for piecewise polynomial approximation. PhD thesis*. Purdue University, 1972.

[Ste07] R. Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007. DOI: 10.1007/s10208-005-0183-0.

[Ste08] R. Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comp.*, 77(261):227–241, 2008. DOI: 10.1090/S0025-5718-07-01959-X.

[SZ90] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990. DOI: 10.2307/2008497.

[Tra97] C. T. Traxler. An algorithm for adaptive mesh refinement in $n$ dimensions. *Computing*, 59(2):115–137, 1997. DOI: 10.1007/BF02684475.

[Vee02] A. Veeser. Convergent adaptive finite elements for the nonlinear Laplacian. *Numer. Math.*, 92(4):743–770, 2002. DOI: 10.1007/s002110100377.

[Ver13] R. Verfürth. *A posteriori error estimation techniques for finite element methods*. Oxford University Press, Oxford, 2013. DOI: 10.1093/acprof:oso/9780199679423.001.0001.

[WYW06] Z. Wu, J. Yin, and C. Wang. *Elliptic & parabolic equations*. World Scientific, New Jersey, 2006.

[WZ17] J. Wu and H. Zheng. Uniform convergence of multigrid methods for adaptive meshes. *Appl. Numer. Math.*, 113:109–123, 2017. DOI: 10.1016/j.apnum.2016.11.005.

[XHYM21] F. Xu, Q. Huang, H. Yang, and H. Ma. Multilevel correction goal-oriented adaptive finite element method for semilinear elliptic equations. *Appl. Numer. Math.*, 172:224–241, 2022. DOI: https://doi.org/10.1016/j.apnum.2021.10.001.

[Yos95] K. Yosida. *Functional analysis*. Springer, Berlin, 1995. DOI: 10.1007/978-3-642-61859-8.

[Zar60] E. H. Zarantonello. Solving functional equations by contractive averaging, 1960.

[Zei90] E. Zeidler. *Nonlinear functional analysis and its applications. Part II/B*. Springer, New York, 1990.

[ZZ87] O. C. Zienkiewicz and J. Z. Zhu. A simple error estimator and adaptive procedure for practical engineering analysis. *Internat. J. Numer. Methods Engrg.*, 24(2):337–357, 1987. DOI: 10.1002/nme.1620240206.

# Academic curriculum vitae

## Personal data


| | |
|---|---|
| Name | Maximilian Brunner |
| Date of birth | 01.02.1993 |
| Citizenship | Austria |
| Email | maximilian.brunner@asc.tuwien.ac.at |
| Website | Homepage TU Wien |
| Google Scholar | Google Scholar |


## Own scientific publications and preprints

| | |
|---|---|
| 2024 | M. Brunner, D. Praetorius, and J. Streitberger. Cost-optimal adaptive FEM with linearization and algebraic solver for semilinear elliptic PDEs, 2024. arXiv: 2401.06486 |
| 2023 | P. Bringmann, M. Brunner, D. Praetorius, and J. Streitberger. Optimal complexity of goal-oriented adaptive FEM for nonsymmetric linear elliptic PDEs, 2023. arXiv: 2312.00489 |
| 2023 | M. Brunner, M. Innerberger, A. Miraçi, D. Praetorius, J. Streitberger, and P. Heid. Adaptive FEM with quasi-optimal overall cost for nonsymmetric linear elliptic PDEs. *IMA J. Numer. Anal.*, 2023. DOI: 10.1093/imanum/drad039. Corrigendum to: Adaptive FEM with quasi-optimal overall cost for nonsymmetric linear elliptic PDEs. *IMA J. Numer. Anal.*, 2024. DOI: 10.1093/imanum/drad103 |
| 2023 | R. Becker, M. Brunner, M. Innerberger, J. M. Melenk, and D. Praetorius. Cost-optimal adaptive iterative linearized FEM for semilinear elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 57(4):2193–2225, 2023. DOI: 10.1051/m2an/2023036 |
| 2022 | 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.*, 118:18–35, 2022. DOI: 10.1016/j.camwa.2022.05.008 |

## Scientific talks

| | |
|---|---|
| 09 2023 | *Rate-optimal goal-oriented adaptive FEM for semilinear elliptic PDEs*, European Conference on Numerical Mathematics and Advanced Applications (ENUMATH), Lisbon (invited to minisymposium on goal-oriented FEM) |
| 09 2022 | *Rate-optimal goal-oriented adaptive FEM for semilinear elliptic PDEs*, Computational Methods in Applied Mathematics (CMAM), Vienna |
| 05 2022 | *Rate-optimal goal-oriented adaptive FEM for semilinear elliptic PDEs*, Austrian Numerical Analysis Day, Linz |

## Scholarships

| | |
|---|---|
| 09 2023 | Conference scholarship of TU Wien for ENUMATH 2023, Lisbon |
| 06 2023 | Summer school on numerical analysis of nonlinear PDEs, Leipzig |

## Teaching

| | |
|---|---|
| 2021 | Exercise class on Numerical Mathematics WS2021 |
| 2017 | **Student assistant for Analysis I** for Technical Physics BSc, TU Wien |
| misc | Contributions in Introduction to Programming WS2020, Introduction to Scientific Programming SS2022 |

## Education

| | |
|---|---|
| since 11 2020 | **PhD student** at Institute of Analysis and Scientific Computing, TU Wien |
| thesis | On optimal adaptivity for semilinear PDEs |
| supervision | Prof. Dr. **Dirk Praetorius** |
| 10 2017 – 09 2020 | Technical Mathematics MSc, graduated with distinction, TU Wien |
| thesis | Local error analysis for generalised splitting methods |
| supervision | Prof. Dr. **Winfried Auzinger** |
| 10 2015 – 10 2017 | Technical Mathematics BSc, TU Wien |
| 03 2012 – 09 2015 | Mathematics BSc (112 credit points), ETH Zurich |
| 09 2011 – 03 2012 | Physics BSc, ETH Zurich |
| 09 2003 – 06 2011 | BG Dornbirn, graduated with distinction |