DISSERTATION

On optimality of adaptive FEM and BEM

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

Univ.-Prof. Dr. 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. Stefan Schimanko BSc.
[redacted]

Diese Dissertation haben begutachtet:

1. **Prof. Dr. Dirk Praetorius**
   Institut für Analysis und Scientific Computing, TU Wien
2. **Prof. Dr. Martin Vohralík**
   Inria, Paris
3. **Prof. Dr. Stefan Funken**
   Institut für Numerische Mathematik, Universität Ulm

Wien, am 23. Juni 2021

Kurzfassung

Im Rahmen elliptischer partieller Differentialgleichungen (PDE) betrachten wir die Finite Elemente Methode (FEM) und die Randelementmethode (BEM). Wir entwickeln sowie analysieren adaptive Algorithmen, die nicht nur die adaptive Netzverfeinerung steuern, sondern auch die Terminierung von geeigneten Lösern, d.h., die Linearisierung im Fall von nichtlinearen Differentialgleichungen und das iterative Lösen der sich ergebenden linearen Gleichungssysteme.

Zum einen betrachten wir elliptische PDEs zweiter Ordnung, bei denen die auftretenden diskreten Systeme nicht exakt gelöst werden. Für kontrahierende iterative Löser formulieren wir einen adaptiven Algorithmus, der die adaptive Netzverfeinerung sowie die inexakte Lösung der auftretenden nichtlinearen bzw. linearen Systeme überwacht und steuert. Wir beweisen, dass die vorgeschlagene Strategie zu linearer Konvergenz mit optimalen algebraischen Raten führt. Hierbei fokussieren wir uns auf Konvergenzraten in Bezug auf den gesamten Rechenaufwand. Unsere Analysis ist anwendbar auf lineare Probleme, bei denen die linearen Systeme mittels optimal vorkonditionierter CG-Verfahren (PCG) gelöst werden, sowie nichtlineare Probleme mit stark monotoner Nichtlinearität, die mittels der sogenannten Zarantonello-Iteration linearisiert werden.

Wir kombinieren die zuvor genannten Resultate im Rahmen elliptischer Randwertprobleme zweiter Ordnung mit stark monotoner und Lipschitz-stetiger Nichtlinearität. Wir präsentieren einen erweiterten adaptiven Algorithmus für die Berechnung der numerischen Approximation, der neben der adaptiven Gitterverfeinerung und der Zarantonello-Linearisierung auch einen kontrahierenden algebraischen Löser für die auftretenden linearen Gleichungssysteme steuert. Wir ermitteln Abbruchsbedingungen für den algebraischen Löser, die einerseits nicht zu einschränkend, aber andererseits ausreichend dafür sind, dass die inexakte Zarantonello-Linearisierung kontrahierend bleibt. In ähnlicher Weise ermitteln wir geeignete Abbruchsbedingungen für die Zarantonello-Iteration, sodass der Linearisierungsfehler sich nicht nachteilig auf den residualen *a posteriori* Fehlerschätzer auswirkt und die adaptive Netzverfeinerung zuverlässig gesteuert wird. Wir beweisen die Kontraktion der (geschachtelten) inexakten Iteration, die auf lineare Konvergenz des Gesamtverfahrens führt. Desweiteren beweisen wir, dass das Verfahren mit der optimalen Rate in Bezug auf die Freiheitsgrade konvergiert. Schließlich beweisen wir, dass es auch mit derselben optimalen Rate in Bezug auf den gesamten Rechenaufwand konvergiert.

Zum anderen betrachten wir Adaptivität und PCG im Rahmen von Randwertproblemen für elliptische Integralgleichungen erster Art. Ähnlich wie zuvor steuert der präsentierte adaptive Algorithmus die Terminierung von PCG sowie die lokale Netzverfeinerung. Neben Konvergenz mit optimalen algebraischen Raten beweisen wir, dass das Verfahren mit fast-optimaler Rate in Bezug auf den gesamten Rechenaufwand konvergiert.

Abstract

In the framework of elliptic partial differential equations (PDEs), we consider the finite element method (FEM) as well as the boundary element method (BEM). We design and analyze adaptive algorithms which do not only steer the adaptive mesh-refinement but also the termination of appropriate iterative solvers, namely, iterative linearization of nonlinear equations as well as iterative solvers for the arising linear systems.

On the one hand, we consider a general framework for treating linear and nonlinear second-order elliptic PDEs, where the arising discrete systems are not solved exactly. For contractive iterative solvers, we formulate an adaptive algorithm which monitors and steers the adaptive mesh-refinement as well as the inexact solution of the arising discrete systems. We prove that the proposed strategy leads to linear convergence with optimal algebraic rates, where we focus on convergence rates with respect to the overall computational cost. Our analysis covers linear PDEs where the linear systems are solved by an optimally preconditioned conjugate gradient method (PCG) as well as nonlinear PDEs with strongly monotone nonlinearity which are linearized by the so-called Zarantonello iteration.

Furthermore, we combine and extend the aforementioned results in the frame of second-order elliptic boundary value problems with strongly monotone and Lipschitz-continuous nonlinearity. We introduce an extended adaptive algorithm for the computation of the numerical approximation, which steers the adaptive mesh-refinement, the Zarantonello linearization, and a contractive algebraic solver to solve the arising linear systems. We identify stopping criteria for the algebraic solver that on the one hand do not request an overly tight tolerance, but on the other hand are sufficient for the inexact Zarantonello linearization to remain contractive. Similarly, we identify suitable stopping criteria for the Zarantonello iteration that leave an amount of linearization error that is not harmful for the residual *a posteriori* error estimator to steer the adaptive mesh-refinement reliably. We prove a contraction of the (nested) inexact iterations leading to linear convergence of the overall adaptive algorithm. Furthermore, we prove that the adaptive algorithm converges with optimal rates with respect to the number of degrees of freedom. Finally, we prove that the adaptive algorithm converges with the same optimal rate also with respect to the overall computational cost.

On the other hand, we consider the interplay of adaptive mesh-refinement and PCG in the frame of BEM for elliptic integral equations of the first kind. As before, the proposed algorithm steers the termination of PCG as well as the local mesh-refinement. Besides convergence with optimal algebraic rates with respect to the number of degrees of freedom, we also prove that the algorithm converges with almost optimal rates with respect to the overall computational cost.

Danksagung

Zuallererst gilt mein Dank meinem Betreuer Dirk Praetorius, den ich nach den gemeinsamen Jahren und allem, was er für mich getan hat, aber vielmehr als Freund ansehe. Mit größtem Einsatz und Ehrgeiz gibt er immer sein Bestes für jeden in unserer Arbeitsgruppe. Man kann sich wohl keinen engagierteren Betreuer wünschen.

Ebenfalls großer Dank gebührt Martin Vohralík und Stefan Funken für die Begutachtung dieser Arbeit, wobei ich mich bei ersterem zusätzlich noch für die spannende gemeinsame Arbeit bedanken möchte.

Besonders hervorheben möchte ich auch die Arbeitsgruppe, die es mir ermöglichte, nicht nur zum Arbeiten an die Universität zu kommen, sondern mich täglich mit Freunden zu treffen und Spaß zu haben. Vielen Dank an Maximilian Bernkopf, Maximilian Brunner, Giovanni Di Fratta, Markus Faustmann, Michael Feischl, Thomas Führer, Gregor Gantner, Alexander Haberl, Michael Innerberger, Ani Miraçi, Carl-Martin Pfeiler, Alexander Rieder, Michele Ruggeri, Andrea Scaglioni und Bernhard Stiftner. Ich werde euch in Zukunft definitiv vermissen. Einen ganz besonderen Platz nimmt dabei Ursula Schweigler ein, die für uns die Bürokratie bewältigt und uns immer mit Unmengen an Süßem versorgt.

Bedanken möchte ich mich auch bei meinen (Studien-)freunden für die Hilfe über all die Jahre und die wunderschöne Zeit mit euch. Danke an Anna Altreiter und Alexander Haberl, Claudia Mußnig-Wytrzens und Fabian Mußnig sowie Olivia Muthsam und Gregor Gantner. Auch abseits des universitären Umfelds gibt es einige Leute, denen ich danken möchte:

- meinen Mitbewohnern Wolfgang Scheuch und Thomas Wicht,
- der Pubquiz- sowie der CS-Gruppe,
- Christoph Gmeiner, Georg Grüßenberger, Oliver Heigl und Ulla Obereigner für so viele unvergessliche Abenteuer,
- und dem ganzen restlichen „Stammtisch“.

Von ganzem Herzen danke ich auch meiner Familie, ganz besonders Ulrike Schimanko und Wolfgang Kirchweger sowie Lisa und Mathias Mesaric. Ich weiß, ich kann immer auf euch zählen. Nicht unerwähnt bleiben darf auch die kulinarisch vorzügliche Verpflegung von Brigitte und Rudolf Kloiber. Vielen Dank für alles.

Abschließend möchte ich noch meiner Freundin Bianca Kloiber danken, für all die kleinen und großen Dinge, die sie immer für mich macht, und dafür, dass sie es mittlerweile schon so lange mit mir aushält.

Mein Dank gilt der TU Wien sowie dem Fonds zur Förderung der wissenschaftlichen Forschung (FWF), der mich über die Jahre im Zuge der Projekte *Optimal adaptivity for BEM and FEM-BEM coupling (grant P27005)*, *Optimal isogeometric boundary element method (grant P29096)*, und *Computational nonlinear PDEs (grant P33216)* finanziert hat.

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 23. Juni 2021

Stefan Schimanko

Contents

1 Introduction 1
1.1 Motivation . . . 1
1.2 Outline . . . 3

2 Basic notation and function spaces 11
2.1 Lebesgue spaces and basic notation . . . 11
2.2 Sobolev spaces on a domain Ω . . . 12
2.3 Sobolev spaces on the boundary $\partial\Omega$ . . . 13
2.4 Dual spaces . . . 13
2.5 Trace operators and normal derivatives . . . 14

3 Meshes 15
3.1 Triangulation of Ω . . . 15
3.2 Triangulations of $\partial\Omega$ . . . 16
3.3 Discrete function spaces . . . 18
3.4 Mesh-refinement . . . 18
3.5 Extended 1D bisection (EB) . . . 19
3.6 Newest vertex bisection (NVB) . . . 20
3.7 Other refinement strategies . . . 22

4 Adaptive FEM for second-order elliptic systems of partial differential equations 23
4.1 Introduction . . . 23
4.1.1 State of the art . . . 24
4.1.2 Outline . . . 25
4.2 Abstract model problem . . . 25
4.3 Error estimator . . . 27
4.4 Discrete iterative solver . . . 28
4.5 Adaptive algorithm . . . 29
4.6 Abstract main results . . . 30
4.6.1 Linear convergence of the quasi-error . . . 32
4.6.2 Proof of Theorem 17 (linear convergence) . . . 33
4.6.3 Optimal convergence rates of the quasi-error . . . 42
4.6.4 Proof of Theorem 23 (optimal convergence rates) . . . 44
4.7 AFEM for linear elliptic PDE with optimal PCG solver . . . 49
4.7.1 Optimal multilevel additive Schwarz preconditioner . . . 53
4.7.2 Auxiliary results . . . 55
4.7.3 Additive Schwarz operator . . . 61
4.7.4 Proof of Theorem 30 (optimal condition number) . . . 62

4.7.5 Proof of lower bound in Proposition 36 . . . 63
4.7.6 Proof of upper bound in Proposition 36 . . . 70
4.7.7 Numerical experiments . . . 72
4.8 AFEM for quasi-linear elliptic PDE with strongly monotone nonlinearity . . . 83
4.8.1 Numerical experiments . . . 86

5 Fully adaptive algorithm for AFEM for nonlinear operators 99
5.1 Introduction . . . 99
5.1.1 Finite element approximation and Banach–Picard iteration . . . 99
5.1.2 Fully adaptive algorithm . . . 100
5.1.3 State of the art . . . 101
5.1.4 Main results and outline . . . 102
5.2 Adaptive algorithm . . . 103
5.2.1 Abstract setting . . . 103
5.2.2 Mesh-refinement . . . 104
5.2.3 Error estimator . . . 104
5.2.4 Algebraic solver . . . 105
5.2.5 Adaptive algorithm . . . 105
5.2.6 Index set $\mathcal{Q}$ for the triple loop . . . 108
5.3 Main results . . . 108
5.3.1 Reliabilty estimates of Algorithm 41 . . . 108
5.3.2 Proof of Proposition 43 (reliability estimates) . . . 111
5.3.3 Linear convergence of the quasi-error . . . 112
5.3.4 Proof of Theorem 45 (linear convergence) . . . 114
5.3.5 Optimal convergence rates of the quasi-error . . . 125
5.3.6 Proof of Theorem 49 (optimal convergence rates) . . . 127
5.3.7 Optimal computational complexity . . . 132
5.4 Numerical experiments . . . 134
5.4.1 Weak formulation . . . 135
5.4.2 Discretization and *a posteriori* error estimator . . . 135
5.4.3 Experiment with known solution on Z-shaped domain . . . 137
5.4.4 Experiment with unknown solution . . . 141

6 Adaptive BEM for elliptic first-kind integral equations with optimal PCG solver 149
6.1 Introduction . . . 149
6.1.1 State of the art . . . 149
6.1.2 Outline . . . 150
6.2 Preliminaries and notation . . . 151
6.2.1 Boundary integral operators and functional analytic setting . . . 151
6.3 Model problem and boundary element method (BEM) . . . 153
6.3.1 Mesh-refinement . . . 154
6.3.2 *A posteriori* BEM error control . . . 155
6.3.3 Preconditioned conjugate gradient method (PCG) for the Galerkin system . . . 156
6.3.4 Optimal preconditioners . . . 157

6.4 Adaptive algorithm . . . 157
6.5 Main results . . . 158
6.5.1 Optimal additive Schwarz preconditioner . . . 158
6.5.2 Proof of Theorem 60 (optimality of additive Schwarz preconditioner) 160
6.5.3 Optimal convergence . . . 169
6.5.4 Proof of Theorem 68 (optimal convergence rates) . . . 170
6.5.5 Almost optimal computational complexity . . . 183
6.6 Hyper-singular integral equation . . . 185
6.7 Numerical experiments . . . 187
6.7.1 Slit problem in 2D . . . 187
6.7.2 Weakly-singular integral equation on Z-shaped domain in 2D . . . 196
6.7.3 Hyper-singular integral equation on L-shaped domain in 2D . . . 196
6.7.4 Weakly-singular integral equation on L-shaped domain in 3D . . . 200
6.7.5 Computational complexity . . . 203

Bibliography **207**

1 Introduction

1.1 Motivation

Two very important methods for numerically solving partial differential equations (PDEs) arising in engineering and natural sciences are the finite element method (FEM) and the boundary element method (BEM). While typical fields of application of FEM are, e.g., structural analysis, heat transfer, and fluid flow problems, BEM can be used to solve problems from, e.g., fluid mechanics, acoustics, or electromagnetics, where the PDEs on a possibly unbounded exterior domain have equivalently been formulated in terms of integral equations posed on the boundary.

This wide range of fields of application led to the development of various numerical schemes based on the principal ideas of finite elements. Most of these methods discretize the domain of interest by a mesh of polygons, thus leading to a reduction of the PDE to a finite dimensional system of equations, and consequently to a finite dimensional approximation of the in general unknown solution. The quality of this approximation can be controlled by the mesh-width of the discretization of the domain. As a result, a simple and widely used idea to decrease the error is to uniformly refine the corresponding mesh successively, which yields convergence of the error to zero. However, the order of convergence might be heavily spoiled by singularities of the unknown solution which can be induced by the given data, the differential operator, and/or the geometry. Hence, significantly more computational effort is needed to reach a required accuracy, since the convergence of the error can be arbitrarily slow. To circumvent this unnecessary computational effort, the mesh can be refined locally at these singularities. However, doing this beforehand would require *a priori* information of the unknown solution which, in general, is not available. This led to the development of adaptive algorithms which automatically steer the local refinement via *a posteriori* error estimators, i.e., adaptive finite element methods (AFEM). One particular focus in AFEM is on the numerical analysis of rate-optimal convergence, where one aims to prove that the adaptive strategy leads to convergence of order $\mathcal{O}\big((\#\mathcal{T}_\ell)^{-s}\big)$ along the sequence of generated triangulations, with $s > 0$ being maximal, where we plot the error estimator over the number of elements $\#\mathcal{T}_\ell$.

Concerning the rate-optimal convergence of AFEM, some seminal works for linear problems are, e.g., [Dör96, MNS00, BDD04, Ste07, CKNS08, CN12, FFP14]. For nonlinear problems, we refer to [Vee02, DK08, BDK12, GMZ12] as well as to [CFPP14] for a general framework of convergence of AFEM with optimal convergence rates. Some works also account for the approximate computation of the discrete solutions by iterative (and inexact) solvers, see, e.g., [BMS10, AGL13] for linear problems and [GMZ11, GHPS18, HW20a, HW20b] for nonlinear model problems. Moreover, there are many papers on a *posteriori* error estimation which also include the iterative and inexact solution for nonlinear problems, see, e.g., [EAEV11, EV13, AW15, HW18] and the references therein.

As far as optimal convergence rates are concerned, the mentioned works focus on rates with respect to the degrees of freedom. Contrary to this, in practice, one aims for the optimal rate of convergence with respect to the computational cost, i.e., the computational time, which is one of the main goals of the present thesis. In [Ste07], this is already addressed for the 2D Poisson model problem. However, this seminal work assumes that a sufficiently accurate discrete solution can be computed in linear complexity, e.g., by a multigrid solver. Under these so-called *realistic assumptions*, it is proved that the *total error*, which consists of the energy error plus data oscillations, converges also with optimal rate with respect to the computational cost.

One starting point of the present thesis is [GHPS18], where an elliptic PDE with strongly monotone nonlinearity is considered. There, the arising nonlinear FEM problems are linearized via the so-called *Zarantonello iteration*, which leads to a linear Poisson problem in each step. The adaptive algorithm presented therein drives the linearization strategy as well as the local mesh-refinement and *almost optimal* convergence rates with respect to the total computational cost are proved. In the present thesis, we prove *optimal* rates with respect to the overall computational cost based on an abstract analysis in the spirit of [CFPP14]. Besides the mentioned Zarantonello iteration for nonlinear model problems, this abstract setting also covers linear solvers like PCG with optimal preconditioner. In a next step, we then combine these two approaches in a fully adaptive algorithm and prove optimal convergence rates with respect to the overall computational cost. Here a key question is to identify suitable stopping criteria for the involved and nested iterative solvers.

For problems on unbounded domains, FEM often is not well applicable. In these situations, BEM can be the better option, since it does not consider and discretize the PDE itself but an equivalent boundary integral equation. Hence, a given problem on an unbounded domain can be reduced to a problem on its (possibly) bounded boundary. In a post-processing stage, the solution of this integral equation then gives rise to an approximation of the PDE solution on the whole space via a representation formula. Due to the dimension reduction and a potentially higher convergence order of BEM, this can lead to higher efficiency in terms of the computational cost.

We refer to [Gan13, FKMP13, FFK+14, FFK+15, AFF+17] for some milestones for adaptive BEM. These works assume that the arising Galerkin systems are solved exactly. However, we note that this is hardly possible in practice, where matrix compression techniques like the fast multipole method, panel clustering, or hierarchical matrix techniques are a must to deal with the dense BEM matrices. In particular, this prevents the use of direct solvers. Instead, we avoid the latter assumption and present an adaptive BEM algorithm to solve elliptic integral equations of the first kind. This algorithm uses a preconditioned conjugate gradient method (PCG) with optimal additive Schwarz preconditioner to approximately solve the arising linear discrete systems. Analogously to [GHPS18], we prove convergence with optimal rates with respect to the degrees of freedom. Due to an additional consistency error stemming from matrix compression techniques for the dense BEM matrices, this leads to *almost optimal* rates with respect to the computational complexity.

1.2 Outline

Chapter 2

First, in Chapter 2, we collect some preliminaries and basic notations which will be used throughout the whole thesis and introduce Lebesgue as well as Sobolev spaces on domains $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ and boundary $\partial\Omega$. We recall the most important results and properties from PDE theory and functional analysis which are needed for the analysis of the following chapters.

Chapter 3

In Chapter 3, we then introduce meshes $\mathcal{T}^\Omega$ of a domain $\Omega \subset \mathbb{R}^d$ as well as meshes $\mathcal{T}^\Gamma$ on subsets $\Gamma \subseteq \partial\Omega$ of the boundary $\partial\Omega$. Additionally, we recall structural properties (R1)–(R3) for the mesh-refinement from [CFPP14], which are essential for the abstract analysis concerning optimal convergence rates in the subsequent chapters. These assumptions are, e.g., fulfilled for the *extended 1D bisection* and the *newest vertex bisection*, which we recall in Section 3.5 and Section 3.6, respectively.

Abstract framework for Chapter 4–6

In the following chapters, we present and analyze adaptive algorithms, which take the form

$$\boxed{\text{Solve}} \longrightarrow \boxed{\text{Estimate}} \longrightarrow \boxed{\text{Mark}} \longrightarrow \boxed{\text{Refine}} \tag{1.1}$$

where $\boxed{\text{Mark}}$ is based on the Dörfler criterion from [Dör96] with (quasi-)minimal cardinality [Ste07, PP20]. These algorithms generate a sequence of discrete approximations $u_\ell^\star$ to the, generally not available, exact solution $u^\star$ of the given problem. Here, the index ℓ corresponds to the discretization of the given problem. However, since solving the arising discrete problems exactly is usually not possible or very costly, iterative solvers are employed. Therefore, we adapt the strategy (1.1) as follows:

$$\boxed{\text{Iteratively Solve \& Estimate}} \longrightarrow \boxed{\text{Mark}} \longrightarrow \boxed{\text{Refine}} \tag{1.2}$$

This gives rise to iterative approximations u_ℓ^k for the exact discrete solutions $u_\ell^\star$, where the index k corresponds to the iterative solver. The numerical analysis of (1.2) thus requires the index set

$$\mathcal{Q} := \big\{(\ell, k) \in \mathbb{N}_0^2 \,:\, \text{discrete approximation } u_\ell^k \text{ is computed by the algorithm}\big\} \tag{1.3}$$

together with an ordering

$$(\ell, k) < (\ell', k') \quad \overset{\text{def}}{\iff} \quad u_\ell^k \text{ is computed earlier than } u_{\ell'}^{k'}. \tag{1.4}$$

Additionally, we define the *total step counter* $|(\ell, k)|$ as

$$|(\ell', k')| := \#\big\{(\ell, k) \in \mathcal{Q} \,:\, (\ell, k) < (\ell', k')\big\}. \tag{1.5}$$

To prove convergence with optimal algebraic rates with respect to the number of degrees of freedom of the iterates u_ℓ^k to the exact solution $u^\star$, we consider a certain *quasi-error* $\Delta_\ell^k := |||u^\star - u_\ell^k||| + \eta_\ell(u_\ell^k)$ combining the error $|||u^\star - u_\ell^k|||$ as well as the error estimator $\eta_\ell(u_\ell^k)$. The key argument for the proof is the *full* linear convergence

$$\Delta_{\ell'}^{k'} \le C_{\rm lin}\, q_{\rm lin}^{|(\ell',k')|-|(\ell,k)|}\, \Delta_\ell^k \quad \text{for all } (\ell,k),(\ell',k') \in \mathcal{Q} \text{ with } |(\ell,k)| \le |(\ell',k')|, \tag{1.6}$$

where $C_{\rm lin} \ge 1$ and $0 < q_{\rm lin} < 1$ are generic constants.

Given $N \in \mathbb{N}_0$, let $\mathbb{T}(N)$ be the set of all refinements $\mathcal{T}$ of $\mathcal{T}_0$ with $\#\mathcal{T} - \#\mathcal{T}_0 \le N$. For $s > 0$, define

$$\|u^\star\|_{\mathbb{A}_s} := \sup_{N\in\mathbb{N}_0} (N+1)^s \inf_{\mathcal{T}_{\rm opt}\in\mathbb{T}(N)} \big(|||u^\star - u_{\rm opt}^\star||| + \eta_{\rm opt}(u_{\rm opt}^\star)\big) \in \mathbb{R}_{\ge 0} \cup \{\infty\}, \tag{1.7}$$

where $u_{\rm opt}^\star$ is the exact discrete solution associated to the mesh $\mathcal{T}_{\rm opt}$ and $\eta_{\rm opt}(u_{\rm opt}^\star)$ is the corresponding error estimator. It holds that $\|u^\star\|_{\mathbb{A}_s} < \infty$ if and only if the quasi-error $\Delta_{\rm opt}^\star := |||u^\star - u_{\rm opt}^\star||| + \eta_{\rm opt}(u_{\rm opt}^\star)$ for the exact discrete solutions decays at least with algebraic rate $s > 0$ along a sequence of optimal meshes. In usual applications, $\Delta_{\rm opt}^\star$ is equivalent to the so-called *total error* (i.e., error plus data oscillations) as well as to the estimator $\eta_{\rm opt}(u_{\rm opt}^\star)$ alone. Therefore, the approximability $\|u^\star\|_{\mathbb{A}_s}$ can equivalently be defined through the total error (see, e.g., [Ste07, CKNS08, CN12, FFP14]) or the estimator (see, e.g., [CFPP14]) instead of the quasi-error (used in (1.7)). The overall result will be the same. However, we stress that none of these equivalences hold for the solver iterates u_ℓ^k, since those lack the Galerkin orthogonality, in general.

Convergence of the adaptive loop (1.2) with optimal rates with respect to the degrees of freedom then means that, for all $s > 0$, there exists a constant $C(s) > 0$ such that

$$C(s)^{-1}\, \|u^\star\|_{\mathbb{A}_s} \le \sup_{(\ell,k)\in\mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s\, \Delta_\ell^k \le C(s)\, (\|u^\star\|_{\mathbb{A}_s} + 1). \tag{1.8}$$

Hence, the quasi-error Δ_ℓ^k for the computed discrete iterates u_ℓ^k decays with rate $s > 0$ if and only if rate s is possible for the exact discrete solutions on optimal meshes.

Finally, our main goal is to prove convergence with optimal rates with regard to the computational cost. Assuming that all steps of the adaptive loop (1.2) can be performed at linear cost $\mathcal{O}(\#\mathcal{T}_\ell)$, the sum

$$\sum_{\substack{(\ell',k')\in\mathcal{Q}\\(\ell',k')\le(\ell,k)}} \#\mathcal{T}_{\ell'}$$

is proportional to the overall computational work to compute the approximation u_ℓ^k, since it depends on the full adaptive history. Convergence with optimal rates with regard to the computational cost then means that, for all $s > 0$, there exists a constant $C'(s) > 0$ such that

$$C'(s)^{-1}\, \|u^\star\|_{\mathbb{A}_s} \le \sup_{(\ell,k)\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell',k')\in\mathcal{Q}\\(\ell',k')\le(\ell,k)}} \#\mathcal{T}_{\ell'} \Bigg)^s \Delta_\ell^k \le C'(s)\, (\|u^\star\|_{\mathbb{A}_s} + 1). \tag{1.9}$$

Thus, the quasi-error Δ_ℓ^k for the computed discrete solutions u_ℓ^k decays with rate $s > 0$ with respect to the overall computational cost if and only if rate s is possible with respect to the degrees of freedom for the exact discrete solutions on optimal meshes.

Chapter 4

This chapter is based on the recent own work [GHPS21].

> Gregor Gantner, Alexander Haberl, Dirk Praetorius, and Stefan Schimanko. Rate optimality of adaptive finite element methods with respect to the overall comput ational costs. *Math. Comp., accepted for publication*, 2021.

We consider the elliptic boundary value problem

$$\begin{aligned} -\operatorname{div} A(\nabla u^\star) &= f \quad \text{in } \Omega, \\ u^\star &= 0 \quad \text{on } \Gamma, \end{aligned} \tag{1.10}$$

where $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ is a bounded Lipschitz domain with boundary $\Gamma = \partial\Omega$ and $f \in L^2(\Omega)$ is a given load. We assume that the (possibly nonlinear) operator $A\colon L^2(\Omega)^d \to L^2(\Omega)^d$ is strongly monotone and Lipschitz continuous. From this, we get the equivalent variational formulation: Find $u^\star \in \mathcal{H} := H^1_0(\Omega)$ such that

$$\langle \mathcal{A}u^\star \,,\, v\rangle_{\mathcal{H}'\times\mathcal{H}} := \int_\Omega A(\nabla u^\star)\cdot\nabla v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x =: \langle F \,,\, v\rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{for all } v \in \mathcal{H}. \tag{1.11}$$

Due to the main theorem on monotone operators [Zei90, Section 25.4], there exists a unique solution $u^\star$ to this weak formulation. For a given discrete subspace $\mathcal{X}_\ell \subset \mathcal{H}$ related to a mesh $\mathcal{T}_\ell$ of Ω, the same holds for the discrete formulation

$$\langle \mathcal{A}u_\ell^\star \,,\, v_\ell\rangle_{\mathcal{H}'\times\mathcal{H}} = \langle F \,,\, v_\ell\rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{for all } v_\ell \in \mathcal{X}_\ell. \tag{1.12}$$

If A is nonlinear, the exact discrete solution $u_\ell^\star$ can hardly be computed exactly. Even if A is linear, usual FEM codes employ iterative solvers like PCG, GMRES, or multigrid. For the abstract analysis, we assume that we have an iterative solver which is contractive in each step with respect to the energy norm, i.e., it holds that

$$|||u_\ell^\star - u_\ell^k||| \le q\,|||u_\ell^\star - u_\ell^{k-1}||| \quad \text{for all } k \in \mathbb{N} \tag{1.13}$$

with a generic contraction constant $0 < q < 1$. Then, our adaptive algorithm takes the form (1.2). We note that (1.13) allows to control the solver error by means of

$$|||u^\star - u_\ell^k||| \le \frac{q}{1-q}|||u_\ell^k - u_\ell^{k-1}|||. \tag{1.14}$$

We terminate the solver if $|||u_\ell^k - u_\ell^{k-1}|||$ is small compared to $\eta_\ell(u_\ell^k)$ and employ nested iteration with $u_{\ell+1}^0 := u_\ell^k$ in this case. Under usual assumptions, we prove that the proposed adaptive strategy guarantees full linear convergence (1.6) of the *quasi-error* $\Delta_\ell^k := |||u^\star - u_\ell^k||| + \eta_\ell(u_\ell^k)$ consisting of error plus error estimator. Prior works, e.g., [Ste07, BMS10, CG12, GHPS18], proved linear convergence of the quasi-error only for those steps, where mesh-refinement takes place. Unlike this, full linear convergence (1.6) even holds for the full sequence of discrete approximations, i.e., independently of the algorithmic decision for mesh-refinement or one step of the discrete solver. Moreover, we prove convergence with

optimal rates with respect to the degrees of freedom (1.8) as well as the computational cost (1.9).

In Section 4.7, we consider the *linear* elliptic boundary value problem (1.10), where we assume that

$$A\colon L^2(\Omega)^d \to L^2(\Omega)^d \quad \text{has the form} \quad A(\mathbf{v}) = \big[x \mapsto \mathbf{A}(x)\mathbf{v}(x)\big], \tag{1.15}$$

where $\mathbf{A} \in W^{1,\infty}(\Omega)^{d\times d}$ is symmetric and uniformly positive definite. Then, the discrete formulation (1.12) is equivalent to the solution of a linear system

$$\mathbf{M}_\ell \mathbf{x}_\ell^\star = \mathbf{b}_\ell. \tag{1.16}$$

with a positive definite and symmetric matrix $\mathbf{M}_\ell \in \mathbb{R}^{N\times N}$. We note that the condition number of the Galerkin matrix $\mathbf{M}_\ell$ from (1.16) depends on the number of elements of $\mathcal{T}_\ell$, as well as the minimal and maximal diameter of its elements. Therefore, we use PCG in combination with an efficient preconditioner $\mathbf{P}_\ell \in \mathbb{R}^{N\times N}$ as an iterative solver. PCG formally applies the conjugate gradient method to the system matrix $\mathbf{P}_\ell^{-1/2}\mathbf{M}_\ell\mathbf{P}_\ell^{-1/2}$ of the preconditioned linear system

$$\mathbf{P}_\ell^{-1/2}\mathbf{M}_\ell\mathbf{P}_\ell^{-1/2}\widetilde{\mathbf{x}}_\ell^\star = \mathbf{P}_\ell^{-1/2}\mathbf{b}_\ell. \tag{1.17}$$

We assume that the matrix-vector products with $\mathbf{P}_\ell^{-1}$ can be computed at linear cost, and that $\mathbf{P}_\ell$ is optimal in the sense that the condition number of the preconditioned system is uniformly bounded, i.e,

$$\operatorname{cond}_2(\mathbf{P}_\ell^{-1/2}\mathbf{M}_\ell\mathbf{P}_\ell^{-1/2}) \le C, \tag{1.18}$$

where the constant $C \ge 1$ is independent of the mesh $\mathcal{T}_\ell$. This yields the contraction property (1.13) so that the abstract main results of Chapter 4 apply to this setting. In Sections 4.7.1–4.7.6, we formulate and analyze a multilevel diagonal scaling preconditioner $\mathbf{P}_\ell \in \mathbb{R}^{N\times N}$ in the frame of multilevel additive Schwarz methods and prove its optimality.

The abstract results of Chapter 4 also apply to AFEM for quasi-linear elliptic PDEs with strongly monotone nonlinearity (cf. Section 4.8), where we employ the Zarantonello iteration and assume that the arising linearized discrete equations are solved exactly at linear cost. The computation of one step of the Zarantonello iteration requires only the solution of one Poisson equation with homogeneous Dirichlet data, i.e., to compute u_ℓ^{k+1} from u_ℓ^k, we have to solve the linear problem

$$\langle\!\langle u_\ell^{k+1}, v_\ell \rangle\!\rangle = \langle\!\langle u_\ell^k, v_\ell \rangle\!\rangle - \frac{\alpha}{L^2}\langle \mathcal{A}u_\ell^k - F, v_\ell\rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{for all } v_\ell \in \mathcal{X}_\ell, \tag{1.19}$$

where $\langle\!\langle \cdot, \cdot \rangle\!\rangle = \langle \nabla\cdot, \nabla\cdot\rangle_{L^2(\Omega)}$. Again, the abstract main results apply to this setting.

To underpin the theoretical results, we present some numerical examples.

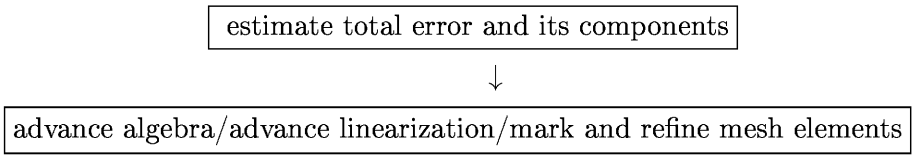
Chapter 5

As an extension of Chapter 4, the aim of Chapter 5 is to combine the two aforementioned approaches of Chapter 4, i.e., Section 4.7 as well as Section 4.8, into one fully adaptive

algorithm for elliptic PDEs with strongly monotone nonlinearity. As before, we consider the elliptic boundary value problem (1.10) where the nonlinearity $A\colon \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz-continuous and strongly monotone. The presented material is based on the recent own work [HPSV21]:

> Alexander Haberl, Dirk Praetorius, Stefan Schimanko, and Martin Vohralík. Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver. *Numer. Math.*, 2021.

We propose an adaptive algorithm of the type



which monitors and adequately stops the iterative linearization and the linear algebraic solver as well as steers the local mesh-refinement. We compute a sequence of discrete appoximations $u_\ell^{k,j}$ of the exact solution $u^\star$ that have an index ℓ for the mesh-refinement, an index k for the Zarantonello linearization (1.19), and an index j for the algebraic solver iteration approximating the exact solution $u_\ell^{k,\star}$ of (1.19) by $u_\ell^{k,j}$. First, we identify stopping criteria for the algebraic solver, e.g., PCG with optimal preconditioner, that on the one hand do not request an overly tight tolerance but on the other hand are sufficient for the inexact (perturbed) Zarantonello linearization to remain contractive. Similarly, we identify suitable stopping criteria for the Zarantonello iteration that leave an amount of linearization error that is not harmful for the residual *a posteriori* error estimate to steer the adaptive mesh-refinement reliably.

Analogously to Chapter 4, the sequential nature of the fully adaptive algorithm gives rise to the index set

$$\mathcal{Q} := \big\{(\ell,k,j) \in \mathbb{N}_0^3 \,:\, \text{discrete approximation } u_\ell^{k,j} \text{ is computed by the algorithm}\big\}$$

together with the ordering

$$(\ell,k,j) < (\ell',k',j') \quad \overset{\text{def}}{\iff} \quad u_\ell^{k,j} \text{ is computed earlier than } u_{\ell'}^{k',j'}.$$

Analogously to (1.5), we define the total step counter

$$|(\ell',k',j')| := \#\big\{(\ell,k,j) \in \mathcal{Q} \,:\, (\ell,k,j) < (\ell',k',j')\big\}, \tag{1.20}$$

as well as the quasi-error

$$\Delta_\ell^{k,j} := |\!|\!| u^\star - u_\ell^{k,j} |\!|\!| + |\!|\!| u_\ell^{k,\star} - u_\ell^{k,j} |\!|\!| + \eta_\ell(u_\ell^{k,j})$$

consisting, in order, of the overall error, the algebraic error, and the error estimator. Our first main result proves that the proposed adaptive strategy is *linearly convergent* in the sense of

$$\Delta_{\ell'}^{k',j'} \le C_{\mathrm{lin}}\, q_{\mathrm{lin}}^{|(\ell',k',j')|-|(\ell,k,j)|}\, \Delta_\ell^{k,j} \quad \text{for all } |(\ell,k,j)| \le |(\ell',k',j')|, \tag{1.21}$$

where $C_{\mathrm{lin}} \geq 1$ and $0 < q_{\mathrm{lin}} < 1$ are generic constants. Second, we prove the *optimal* error decay rate with respect to the number of degrees of freedom exceeding those of the initial mesh in the sense that there exists a constant $C(s) > 0$ such that

$$C(s)^{-1} \, |||u^\star|||_{\mathbb{A}_s} \leq \sup_{(\ell,k,j)\in\mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \, \Delta_\ell^{k,j} \leq C(s) \, (|||u^\star|||_{\mathbb{A}_s} + 1). \tag{1.22}$$

As before, estimate (1.21) is the key argument to prove *optimal* error decay rate with respect to the overall computational cost of the fully adaptive algorithm which steers the mesh-refinement, the perturbed Zarantonello linearization, and the algebraic solver, i.e., for all $s > 0$, there exists a constant $C'(s) > 0$ such that

$$C'(s)^{-1} \, |||u^\star|||_{\mathbb{A}_s} \leq \sup_{(\ell,k,j)\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell',k',j')\in\mathcal{Q} \\ (\ell',k',j')\leq(\ell,k,j)}} \#\mathcal{T}_{\ell'} \Bigg)^s \Delta_\ell^{k,j} \leq C'(s) \, (|||u^\star|||_{\mathbb{A}_s} + 1). \tag{1.23}$$

As above, we stress that under realistic assumptions the sum in (1.23) is indeed proportional to the overall computational cost invested into the fully adaptive numerical approximation of (1.10), if the cost of all procedures like matrix and right-hand-side assembly, one algebraic solver step, evaluation of the involved *a posteriori* error estimates, marking, and local adaptive mesh refinement is proportional to the number of mesh elements in $\mathcal{T}_\ell$, i.e., the number of degrees of freedom.

To underpin the theoretical results, we also present some numerical examples.

Chapter 6

Chapter 6 is based on the own work [FHPS19]:

> Thomas Führer, Alexander Haberl, Dirk Praetorius, and Stefan Schimanko. Adaptive BEM with inexact PCG solver yields almost optimal computational costs. *Numer. Math.*, 2019,

where we consider weakly-singular integral equations of first kind. We note that [FHPS19] was the first work in the context of adaptive FEM or BEM aiming for full linear convergence and corresponding optimal rates with respect to the computational cost. The core analysis was later improved by the analysis of [GHPS21] presented in Chapter 4 in such a way that the latter only needs a contractive iterative solver, whereas some of the results of [FHPS19] are tailored to the BEM setting with inexact PCG solver.

For a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ and polyhedral boundary $\partial\Omega$, let $\Gamma \subseteq \partial\Omega$ be a (relatively) open and connected subset. Given $f\colon \Gamma \to \mathbb{R}$, we seek the density $\phi^\star\colon \Gamma \to \mathbb{R}$ of the weakly-singular integral equation

$$(V\phi^\star)(x) := \int_\Gamma G(x-y)\phi^\star(y) \, \mathrm{d}y = f(x) \quad \text{for all } x \in \Gamma, \tag{1.24}$$

where $G(\cdot)$ denotes the fundamental solution of the Laplace operator in $\mathbb{R}^d$. Its lowest-order Galerkin formulation for a given triangulation $\mathcal{T}_\ell$ of Γ reads as follows: Find $\phi_\ell^\star \in \mathcal{P}^0(\mathcal{T}_\ell)$

such that

$$\int_\Gamma (V\phi_\ell^\star)(x)\,\psi_\ell(x)\,\mathrm{d}x = \int_\Gamma f(x)\,\psi_\ell(x)\,\mathrm{d}x \quad \text{for all } \psi_\ell \in \mathcal{P}^0(\mathcal{T}_\ell). \tag{1.25}$$

As for FEM for linear problems in Chapter 4, the discrete formulation (1.25) can be written as an equivalent linear system

$$\mathbf{M}_\ell \mathbf{x}_\ell^\star = \mathbf{b}_\ell \tag{1.26}$$

with a positive definite and symmetric matrix $\mathbf{M}_\ell \in \mathbb{R}^{N\times N}$ which, unlike FEM, is dense for BEM. For a given initial triangulation $\mathcal{T}_0$, we again consider an adaptive mesh-refinement strategy of the type (1.2), which generates a sequence of successively refined triangulations $\mathcal{T}_\ell$ for all $\ell \in \mathbb{N}_0$. As before in Chapter 4, the condition number of the Galerkin matrix $\mathbf{M}_\ell$ from (1.26) depends on the number of elements of $\mathcal{T}_\ell$, as well as the minimal and maximal diameter of the elements. Therefore, we require an efficient preconditioner as well as an appropriate iterative solver.

The available results for adaptive BEM [Gan13, FKMP13, FFK+14, FFK+15, AFF+17] assume that the Galerkin system (1.26) is solved exactly. Instead, our adaptive algorithm steers both the local mesh-refinement and the iterations of an iterative PCG solver for the Galerkin system (1.26). In principle, it is known [CFPP14, Section 7] that convergence and optimal convergence rates are preserved if the linear system is solved inexactly, but with sufficient accuracy. Analogously to Chapter 4, we guarantee this by incorporating an appropriate stopping criterion for the PCG solver into the adaptive algorithm. Moreover, to prove that the proposed algorithm does not only lead to optimal algebraic convergence rates, but also to (almost) optimal computational cost, we provide a preconditioner $\mathbf{P}_\ell \in \mathbb{R}^{N\times N}$ such that the evaluation of the matrix-vector product with $\mathbf{P}_\ell^{-1}$ can be done in $\mathcal{O}(\#\mathcal{T}_\ell)$ operations, and that $\mathbf{P}_\ell$ is optimal in the sense of (1.18), i.e., the system matrix $\mathbf{P}_\ell^{-1/2}\mathbf{M}_\ell\mathbf{P}_\ell^{-1/2}$ of the preconditioned linear system has a uniformly bounded condition number which is independent of $\mathcal{T}_\ell$.

As in Chapter 4, we prove that the quasi-error

$$\Delta_\ell^k := \big(|||\phi^\star - \phi_\ell^k|||^2 + \eta_\ell(\phi_\ell^k)^2\big)^{1/2}$$

consisting of energy error plus error estimator is linearly convergent in each step of the adaptive algorithm, independent of whether the algorithm locally refines the mesh or does one step of the PCG iteration, i.e., there holds (1.6). Furthermore, we also prove (1.8), i.e., the quasi-error decays with optimal rate with respect to the degrees of freedom.

Under realistic assumptions on the efficient treatment of the arising discrete integral operators, one step of the algorithm can be done in $\mathcal{O}\big((\#\mathcal{T}_\ell)\log^2(1+\#\mathcal{T}_\ell)\big)$ operations. Hence, the cumulative computational complexity for the adaptive step $(\ell,k)\in\mathcal{Q}$ is of order

$$\mathcal{O}\Bigg(\sum_{\substack{(\ell',k')\in\mathcal{Q}\\(\ell',k')\le(\ell,k)}} (\#\mathcal{T}_{\ell'})\log^2(1+\#\mathcal{T}_{\ell'})\Bigg). \tag{1.27}$$

As a consequence of the log-linear cost (1.27), we prove that the quasi-error converges at almost optimal rate with respect to the computational cost, i.e., with rate $s-\varepsilon$ for any

$\varepsilon > 0$ if rate $s > 0$ is possible for the exact Galerkin solution. This means that there holds the implication

$$\|\phi^\star\|_{\mathbb{A}_s} < \infty \quad \Longrightarrow \quad \sup_{(\ell,k)\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell',k')\in\mathcal{Q}\\(\ell',k')\le(\ell,k)}} (\#\mathcal{T}_{\ell'}) \log^2(1+\#\mathcal{T}_{\ell'}) \Bigg)^{s-\varepsilon} \Delta_\ell^k < \infty \quad \text{for all } \varepsilon > 0.$$

The difference to the abstract result (1.9) is the logarithmic term in the single-step complexity, which ultimately leads to the reduced order of convergence $s - \varepsilon$.

The final section underpins the theoretical findings by some 2D and 3D experiments.

2 Basic notation and function spaces

In this section, we introduce some basic notations which will be used throughout the whole thesis. Afterwards, we recall some definitions, notations, and results for the well-known Lebesgue and Sobolev spaces, cf., e.g., [McL00, Chapter 3] or [SS11, Chapter 2].

First, let $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ be a bounded Lipschitz domain with boundary $\partial\Omega$. Depending on the context, $|\cdot|$ denotes the absolute value of scalars as well as the Euclidian norm of vectors respectively. For measurable sets in Ω or in $\partial\Omega$, we use the same notation $|\cdot|$ for the corresponding Lebesgue measure as well as the surface measure, respectively.

In general, all constants as well as their dependencies are explicitly given for all statements. However, in proofs, we also abbreviate the notation, i.e., for real-valued quantities A, B, we write $A \lesssim B$ to abbreviate $A \leq c\,B$ with a generic constant $c > 0$ which is clear from the context. Analogously, $A \gtrsim B$ is the abbreviation of $A \geq c\,B$. Moreover, $A \simeq B$ states that both estimates $A \lesssim B$ and $A \gtrsim B$ hold true.

For the remaining part of this section, and in this section only, let Ω be any (Lebesgue) measurable subset of $\mathbb{R}^n$ with $n \geq 1$ and strictly positive measure.

2.1 Lebesgue spaces and basic notation

For $1 \leq p \leq \infty$, the ususal Lebesgue spaces on Ω are denoted by $L^p(\Omega)$ with corresponding norms

$$\|v\|_{L^p(\Omega)} := \Big(\int_\Omega |v(x)|^p \,\mathrm{d}x \Big)^{1/p} \qquad \text{for } 1 \leq p < \infty,$$

as well as $\|v\|_{L^\infty(\Omega)}$ being the essential supremum of u over Ω. Analogously, Lebesgue spaces on the boundary $\partial\Omega$ are denoted by $L^p(\partial\Omega)$ with corresponding norms $\|\cdot\|_{L^p(\partial\Omega)}$.

For all $p \geq 1$, it is well-known that $L^p(\Omega)$ is a Banach space. For $p = 2$, the corresponding Lebesgue space $L^2(\Omega)$ is also a Hilbert space. Hence, for all $u, v \in L^2(\Omega)$, we define the scalar product $\langle \cdot\,,\, \cdot \rangle_{L^2(\Omega)}$ by

$$\langle u\,,\, v \rangle_{L^2(\Omega)} := \int_\Omega u(x)\overline{v(x)} \,\mathrm{d}x.$$

Let $q \geq 1$ denote the conjugate exponent to p, i.e.,

$$\frac{1}{p} + \frac{1}{q} = 1.$$

Then, for all $u \in L^p(\Omega)$ and all $v \in L^q(\Omega)$, there holds the so-called *Hölder's inequality*

$$|\langle u\,,\, v \rangle_{L^2(\Omega)}| = \|u\,v\|_{L^1(\Omega)} \leq \|u\|_{L^p(\Omega)} \|v\|_{L^q(\Omega)}$$

2.2 Sobolev spaces on a domain Ω

Let $v\colon \Omega \to \mathbb{R}$, where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain with piecewise C^∞-boundary $\partial\Omega$, cf. [SS11, Definition 2.2.10]. For $n \in \mathbb{N}$ and a multi-index $\alpha = (\alpha_1, \dots, \alpha_n) \in \mathbb{N}_0^n$, i.e., an n-tuple of non-negative integers, we denote the partial derivatives of v by

$$\partial^\alpha v(x) = \Big(\frac{\partial}{\partial x_1}\Big)^{\alpha_1} \cdots \Big(\frac{\partial}{\partial x_n}\Big)^{\alpha_n} v(x),$$

if v is sufficiently smooth for them to exist. The order $|\alpha|$ of the partial derivative $\partial^\alpha v(x)$ is defined by

$$|\alpha| := \alpha_1 + \cdots + \alpha_n.$$

Definition 1. *Let $v \in L^2(\Omega)$. Then, v has a weak derivative $g := \partial^\alpha v \in L^2(\Omega)$ of order α if there holds that*

$$\int_\Omega g\, w \,\mathrm{d}x = (-1)^{|\alpha|} \int_\Omega v\, \partial^\alpha w \,\mathrm{d}x \quad \text{for all } w \in C_0^\infty(\Omega),$$

where $C_0^\infty(\Omega) := \big\{u \in C^\infty(\Omega) : u \text{ has compact support in } \Omega\big\}$ is the space of infinitely differentiable functions with compact support.

Note that if the weak derivative of $v \in L^2(\Omega)$ exists, it is unique and if v also has a classical derivative, the weak derivative coincides (almost everywhere) with the classical one.

Definition 2. *For $\ell \in \mathbb{N}_0$, the Sobolev space $H^\ell(\Omega)$ is defined by*

$$H^\ell(\Omega) := \big\{v \in L^2(\Omega) : \partial^\alpha v \in L^2(\Omega) \text{ exists in the weak sense for all } |\alpha| \le \ell\big\}.$$

The inner product $\langle \cdot\,, \cdot \rangle_{H^\ell(\Omega)}$ on $H^\ell(\Omega)$ is given by

$$\langle v\,, w\rangle_{H^\ell(\Omega)} := \sum_{|\alpha| \le \ell} \langle \partial^\alpha v\,, \partial^\alpha w\rangle_{L^2(\Omega)} \quad \text{for all } v, w \in H^\ell(\Omega),$$

and the corresponding norm $\|\cdot\|_{H^\ell(\Omega)}$ is given by

$$\|v\|^2_{H^\ell(\Omega)} := \langle v\,, v\rangle_{H^\ell(\Omega)} \quad \text{for all } v \in H^\ell(\Omega).$$

For $\ell = 1$, we hence get that

$$H^1(\Omega) = \big\{v \in L^2(\Omega) : \nabla v \in L^2(\Omega)^d \text{ exists in the weak sense}\big\}$$

with scalar product

$$\langle v\,, w\rangle_{H^1(\Omega)} = \int_\Omega v\, w \,\mathrm{d}x + \int_\Omega \nabla v \cdot \nabla w \,\mathrm{d}x,$$

and norm $\|v\|^2_{H^1(\Omega)} = \langle v\,, v\rangle_{H^1(\Omega)} = \|v\|^2_{L^2(\Omega)} + \|\nabla v\|^2_{L^2(\Omega)}$.

For a non-integer $\ell := k+s$ with $k \in \mathbb{N}_0$ and $0 < s < 1$, the Sobolev space $H^\ell(\Omega)$ is defined by interpolation via the K-method, i.e., $H^\ell(\Omega) := [H^k(\Omega), H^{k+1}(\Omega)]_{s,2}$, cf., e.g., [SS11, Tri95].

2.3 Sobolev spaces on the boundary $\partial\Omega$

Sobolev spaces on the boundary $\partial\Omega$ can be defined in various ways, cf. [HW08, McL00, SS11]. Let $H^0(\partial\Omega) := L^2(\partial\Omega)$ be the space of all square-integrable functions on $\partial\Omega$ with scalar product $\langle \cdot\,,\,\cdot \rangle_{\partial\Omega}$ and norm $\|\cdot\|_{L^2(\partial\Omega)}$. For $\boldsymbol{L}^2(\partial\Omega) := L^2(\partial\Omega)^d$, define the scalar product $\langle \boldsymbol{v}\,,\,\boldsymbol{w}\rangle_{\partial\Omega} := \sum_{j=1}^d \langle \boldsymbol{v}_j\,,\,\boldsymbol{w}_j\rangle_{\partial\Omega}$ and norm $\|\boldsymbol{v}\|^2_{\boldsymbol{L}^2(\partial\Omega)} := \langle \boldsymbol{v}\,,\,\boldsymbol{v}\rangle_{\partial\Omega}$. Then, the space $H^1(\partial\Omega)$ is defined as in [SS11, Section 2.4] with an equivalent norm on $H^1(\partial\Omega)$ given by

$$\|v\|_{L^2(\partial\Omega)} + \|\nabla_\Gamma v\|_{\boldsymbol{L}^2(\partial\Omega)},$$

where $\nabla_\Gamma\colon H^1(\partial\Omega) \to \boldsymbol{L}^2(\Gamma)$ denotes the surface gradient. For sufficiently smooth functions v on $\overline{\Omega}$, it holds that $\nabla_\Gamma v = \nabla v - (\nabla v \cdot \mathbf{n})\mathbf{n}$ with the normal vector $\mathbf{n}$ pointing from the domain Ω to the exterior domain $\Omega^{\mathrm{ext}} := \mathbb{R}^d \setminus \overline{\Omega}$.

For $s \in (0,1)$, the corresponding Sobolev space $H^s(\partial\Omega)$ is defined via interpolation techniques, cf. [SS11, Proposition 2.4.3].

Additionally, we also need Sobolev spaces on subsets Γ of the boundary $\partial\Omega$. Suppose that $\emptyset \neq \Gamma \subset \partial\Omega$ is a non-empty, relatively open set that stems from a Lipschitz dissection $\partial\Omega = \Gamma \cup \partial\Gamma \cup (\partial\Omega \setminus \Gamma)$, cf. [McL00, p. 99]. Define $E_{0,\Gamma}$ as the extension operator which extends a function on Γ to $\partial\Omega$ by zero. For $s \in \{-1/2, 0, 1/2\}$, the spaces $H^{1/2+s}(\Gamma)$ and $\widetilde{H}^{1/2+s}(\Gamma)$ are defined as in [AFF+17] by

$$\begin{aligned} H^{1/2+s}(\Gamma) &:= \big\{v|_\Gamma \,:\, v \in H^{1/2+s}(\partial\Omega)\big\} \\ \widetilde{H}^{1/2+s}(\Gamma) &:= \big\{v \,:\, E_{0,\Gamma} v \in H^{1/2+s}(\partial\Omega)\big\}, \end{aligned}$$

with corresponding norms

$$\begin{aligned} \|v\|_{H^{1/2+s}(\Gamma)} &:= \inf_{w \in H^{1/2+s}(\partial\Omega)} \{\|w\|_{H^{1/2+s}(\partial\Omega)} \,:\, w|_\Gamma = v\} \\ \|v\|_{\widetilde{H}^{1/2+s}(\Gamma)} &:= \|E_{0,\Gamma} v\|_{H^{1/2+s}(\partial\Omega)}. \end{aligned}$$

For $s = 1/2$, there hold the norm equivalences $\|v\|_{H^1(\partial\Omega)} \simeq \|v\|_{L^2(\partial\Omega)} + \|\nabla_\Gamma v\|_{\boldsymbol{L}^2(\partial\Omega)}$ as well as $\|v\|_{\widetilde{H}^1(\Gamma)} \simeq \|v\|_{L^2(\Gamma)} + \|\nabla_\Gamma v\|_{\boldsymbol{L}^2(\Gamma)}$, cf. [AFF+17, Facts 2.1] and [SS11, Section 2.4].

For ease of notation, if it is clear from the context, we identify a function $v \in \widetilde{H}^{1/2+s}(\Gamma)$ with its extension $E_{0,\Gamma} v \in H^{1/2+s}(\partial\Omega)$.

2.4 Dual spaces

For a normed space $\mathcal{X}$ with norm $\|\cdot\|_{\mathcal{X}}$, we denote the corresponding dual space by $\mathcal{X}'$ with the duality pairing

$$\langle v'\,,\,w\rangle_{\mathcal{X}'\times\mathcal{X}} := v'(w) \quad \text{for all } v' \in \mathcal{X}' \text{ and all } w \in \mathcal{X},$$

as well as the norm

$$\|v'\|_{\mathcal{X}'} = \sup_{0 \neq w \in \mathcal{X}} \frac{|\langle v'\,,\,w\rangle_{\mathcal{X}'\times\mathcal{X}}|}{\|w\|_{\mathcal{X}}} \quad \text{for all } v' \in \mathcal{X}'.$$

To simplify notation and if it is clear from the context, we write $\langle \cdot\,,\,\cdot \rangle$ for the duality pairing. If we now have a Hilbert space $\mathcal{X}$ with scalar product $\langle \cdot\,,\,\cdot \rangle_{\mathcal{X}}$ and a continuously embedded Hilbert space $\mathcal{H}$, the following lemma allows us to interpret the duality pairing $\langle \cdot\,,\,\cdot \rangle_{\mathcal{H}'\times\mathcal{H}}$ as a continuous extension of the scalar product $\langle \cdot\,,\,\cdot \rangle_{\mathcal{X}}$.

Lemma 3. *Let $\mathcal{H}$ and $\mathcal{X}$ be Hilbert spaces with continuous embedding $\mathcal{H} \to \mathcal{X}$. Then, the Riesz-isomorphism $J_{\mathcal{X}} \colon \mathcal{X} \to \mathcal{H}'$ is a well-defined, continuous, linear operator and $J_{\mathcal{X}}(\mathcal{X})$ is dense in $\mathcal{H}'$.* □

If we set $\mathcal{X} = L^2(\partial\Omega)$ and $\mathcal{H} = H^{1/2+s}(\partial\Omega)$ or $\mathcal{H} = \widetilde{H}^{1/2+s}(\partial\Omega)$, we get with the formal definition

$$\langle J_{\mathcal{X}} x\,,\,h\rangle_{\mathcal{H}'\times\mathcal{H}} := \langle J_{\mathcal{X}} x\,,\,h\rangle_{\mathcal{X}'\times\mathcal{X}} = \langle x\,,\,h\rangle_{\mathcal{X}} = \langle x\,,\,h\rangle_{L^2(\partial\Omega)} \quad \text{for all } x \in \mathcal{X}, h \in \mathcal{H}$$

so that it is legitimate to also write $\langle \cdot\,,\,\cdot \rangle_{\partial\Omega}$ (and analogously $\langle \cdot\,,\,\cdot \rangle_{\Gamma}$) for the duality pairing $\langle \cdot\,,\,\cdot \rangle_{\mathcal{H}'\times\mathcal{H}}$.

For $s \in \{-1/2, 0, 1/2\}$, the negative-order Sobolev spaces on the boundary are now defined by duality as

$$\begin{aligned} H^{-(1/2+s)}(\partial\Omega) &:= H^{1/2+s}(\partial\Omega)', \\ \widetilde{H}^{-(1/2+s)}(\Gamma) &:= H^{1/2+s}(\Gamma)', \\ H^{-(1/2+s)}(\Gamma) &:= \widetilde{H}^{1/2+s}(\Gamma)', \end{aligned}$$

with the extended L^2-scalar product on $\partial\Omega$ and Γ respectively, cf. [AFF+17]. For these spaces, the following continuous inclusions hold:

$$\begin{aligned} \widetilde{H}^{\pm(1/2+s)}(\Gamma) &\subseteq H^{\pm(1/2+s)}(\Gamma), \quad \text{as well as,} \\ \widetilde{H}^{\pm(1/2+s)}(\partial\Omega) &= H^{\pm(1/2+s)}(\partial\Omega). \end{aligned}$$

For $\psi \in L^2(\Gamma)$, the zero extension $E_{0,\Gamma}\psi$ satisfies

$$E_{0,\Gamma}\psi \in H^{-1/2}(\partial\Omega) \quad \text{as well as} \quad \|\psi\|_{\widetilde{H}^{-1/2}(\Gamma)} = \|E_{0,\Gamma}\psi\|_{H^{-1/2}(\partial\Omega)}.$$

2.5 Trace operators and normal derivatives

Let Ω be a bounded Lipschitz domain. Then, for $1/2 < s < 3/2$, there exists a linear and continuous interior trace operator

$$\gamma_0^{\text{int}} \colon H^s(\Omega) \to H^{s-1/2}(\partial\Omega) \quad \text{such that} \quad \gamma_0^{\text{int}} v = v|_{\partial\Omega} \text{ for all } v \in C^0(\overline{\Omega}),$$

cf., e.g., [SS11, Theorem 2.6.8]. We define $H^1_\Delta(\Omega) := \{v \in H^1(\Omega) \,:\, -\Delta v \in L^2(\Omega)\}$ as well as the interior conormal derivative operator $\gamma_1^{\text{int}} \colon H^1_\Delta(\Omega) \to H^{-1/2}(\partial\Omega)$ via the first Green's formula

$$\langle \gamma_1^{\text{int}} v\,,\,\gamma_0^{\text{int}} w\rangle_{\partial\Omega} = \langle \nabla v\,,\,\nabla w\rangle_{\Omega} - \langle -\Delta v\,,\,w\rangle_{\Omega} \quad \text{for all } w \in H^1(\Omega),$$

cf. [AFF+17]. Analogously, the exterior trace γ_0^{ext} and exterior conormal derivative operator γ_1^{ext} can be defined. Then, the interior as well as exterior traces and the conormal derivatives respectively give rise to jump terms, i.e., for a function v that admits both traces or conormal derivatives, we define the jumps $[v]_0 := \gamma_0^{\text{ext}} v - \gamma_0^{\text{int}} v$ and $[v]_1 := \gamma_1^{\text{ext}} v - \gamma_1^{\text{int}} v$ respectively.

3 Meshes

3.1 Triangulations of Ω

Throughout, let $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ be a polygonal or polyhedral Lipschitz domain and let $\mathrm{conv}(S)$ denote the convex hull of a set $S \subset \mathbb{R}^d$. With this, we define a triangulation $\mathcal{T}^\Omega$ on a domain Ω.

Definition 4. *A set $\mathcal{T}^\Omega$ is called a triangulation or mesh of Ω, if and only if:*

- *Each element $T \in \mathcal{T}^\Omega$ is a $(d+1)$-simplex, i.e., there exist $d+1$ affinely independent points $x_1, \cdots, x_{d+1} \in \overline{\Omega}$ such that*

$$T := \mathrm{conv}(\{x_1, \cdots, x_{d+1}\}).$$

  *We denote the **set of all vertices** of an element T by $\mathcal{N}(T) := \{x_1, \cdots, x_{d+1}\}$.*

- *The domain Ω is covered by $\mathcal{T}^\Omega$, i.e.,*

$$\overline{\Omega} = \bigcup_{T \in \mathcal{T}^\Omega} T.$$

- *Two distinct elements do not overlap, i.e., for all $T, T' \in \mathcal{T}^\Omega$ with $T \neq T'$, it holds that $|T \cap T'| = 0$, i.e., the overlap is a set of measure zero.*

Remark 5. *Usually, we do not want to allow so-called hanging nodes, i.e., no vertex of any element $T \in \mathcal{T}^\Omega$ lies in the interior of any edge or facet of another element $T' \in \mathcal{T}^\Omega$. Hence, we say that a triangulation $\mathcal{T}^\Omega$ is **conforming** or **regular** provided that the intersection of two elements $T, T' \in \mathcal{T}^\Omega$ with $T \neq T'$ is*

- *either empty,*
- *or a joint node,*
- *or a joint edge ($d \geq 2$),*
- *or a joint facet ($d = 3$),*

i.e., for two distinct elements $T, T' \in \mathcal{T}^\Omega$ with $T \neq T'$, it holds that

$$T \cap T' = \mathrm{conv}(\mathcal{N}(T) \cap \mathcal{N}(T')).$$

Further, we collect a couple more definitions. First, we define the **set of all nodes** $\mathcal{N}_{\mathcal{T}^\Omega}$ of a triangulation $\mathcal{T}^\Omega$ by

$$\mathcal{N}_{\mathcal{T}^\Omega} := \mathcal{N}(\mathcal{T}^\Omega) := \bigcup_{T\in\mathcal{T}^\Omega} \mathcal{N}(T).$$

The (local) **mesh-width function** $h_{\mathcal{T}^\Omega} \in L^\infty(\mathcal{T}^\Omega)$ of a triangulation $\mathcal{T}^\Omega$ is defined by

$$h_{\mathcal{T}^\Omega}|_T := h_{\mathcal{T}^\Omega}(T) := |T|^{1/d} \qquad \text{for all } T \in \mathcal{T}^\Omega,$$

where $|\cdot|$ denotes the volume (for $d = 3$) or the area (for $d = 2$) of an element, respectively. Moreover, we define the **element patch** $\omega_{\mathcal{T}^\Omega}(T)$ and $\omega_{\mathcal{T}^\Omega}(\mathcal{U})$ resp. for an element $T \in \mathcal{T}^\Omega$ as well as for a set of elements $\mathcal{U} \subseteq \mathcal{T}^\Omega$ by

$$\omega_{\mathcal{T}^\Omega}(T) := \bigcup \{T' \in \mathcal{T}^\Omega : T' \cap T \neq \emptyset\} \quad \text{and} \quad \omega_{\mathcal{T}^\Omega}(\mathcal{U}) := \bigcup_{T\in\mathcal{U}} \omega_{\mathcal{T}^\Omega}(T), \qquad \text{respectively.}$$

Next, the **shape-regularity constant** $\sigma(T)$ of an element $T \in \mathcal{T}^\Omega$ is denoted by

$$\sigma(T) := \frac{\operatorname{diam}(T)^d}{|T|} \quad \text{with} \quad \operatorname{diam}(T) := \sup_{x,y\in T} |x - y|.$$

Similarly, we define the **shape-regularity constant** $\sigma(\mathcal{T}^\Omega)$ of a mesh $\mathcal{T}^\Omega$ by

$$\sigma(\mathcal{T}^\Omega) := \max_{T\in\mathcal{T}^\Omega} \sigma(T),$$

and we say that a family $\mathbb{T}$ of meshes is **γ-shape regular** if there exists a constant $\gamma \geq 1$ such that

$$\sup_{\mathcal{T}^\Omega\in\mathbb{T}} \sigma(\mathcal{T}^\Omega) \leq \gamma.$$

3.2 Triangulations of $\partial\Omega$

Analogously to Section 2.3, we also need triangulations of the boundary $\partial\Omega$ for the boundary element method in Chapter 6. To this end, let $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ be a bounded Lipschitz domain with piecewise C^∞-boundary $\partial\Omega$, and we suppose that either Γ is the whole boundary, i.e., $\Gamma = \partial\Omega$, or Γ is a subset of the boundary, i.e., $\emptyset \neq \Gamma \subset \partial\Omega$, and relatively open such that $\partial\Omega = \Gamma \cup \partial\Gamma \cup (\partial\Omega \setminus \Gamma)$. Hence, Γ stems from a Lipschitz dissection, cf. [McL00, p. 99].

For the definition of a triangulation $\mathcal{T}^\Gamma$, we also need a reference element T_{ref} defined by

$$T_{\text{ref}} := \Big\{x \in \mathbb{R}^{d-1} : 0 \leq x_1, \dots, x_{d-1} \leq 1 \text{ and } \sum_{j=1}^{d-1} x_j \leq 1\Big\}.$$

Hence, we get that $T_{\text{ref}} = [0, 1] \subset \mathbb{R}$ is the closed unit interval for $d = 2$ as well as $T_{\text{ref}} = \operatorname{conv}\{(0, 0), (1, 0), (0, 1)\} \subset \mathbb{R}^2$ for $d = 3$.

Definition 6. *A set $\mathcal{T}^\Gamma$ is called a triangulation or mesh of Γ, if and only if:*

- *Every element $T \in \mathcal{T}^\Gamma$ is the image of the reference element T_{ref} under an affine, bijective element map $g_T \in C^\infty(T_{\mathrm{ref}}, T)$ with $g_T(T_{\mathrm{ref}}) = T$. The set of nodes is given by $\mathcal{N}(T) := g_T(\mathcal{N}(T_{\mathrm{ref}}))$, where $\mathcal{N}(T_{\mathrm{ref}}))$ is the set of all vertices of the reference element T_{ref}.*
- *The domain Γ is covered by $\mathcal{T}^\Gamma$, i.e.,*

$$\overline{\Gamma} = \bigcup_{T \in \mathcal{T}^\Gamma} T$$

Remark 7. *Analogously to Remark 5, we say that a triangulation $\mathcal{T}^\Gamma$ is **conforming** or **regular** provided that the intersection of two elements $T, T' \in \mathcal{T}^\Gamma$ with $T \neq T'$ is*

- *either empty,*
- *or a joint node ($d \geq 2$),*
- *or a joint facet ($d = 3$),*

and for $d = 3$, it holds that: If $T \cap T'$ is a facet for $T' \in \mathcal{T}^\Gamma$, there exist facets $f, f' \subseteq \partial T_{\mathrm{ref}}$ of T_{ref} such that $T \cap T' = g_T(f) = g_{T'}(f')$ and $g_T^{-1} \circ g_{T'} \colon f' \to f$ is affine.

The **set of nodes** as well as the **element patches** are defined as in Section 3.1, while the (local) **mesh-width function** $h_{\mathcal{T}^\Gamma} \in L^\infty(\mathcal{T})$ is given by

$$h_{\mathcal{T}^\Gamma}|_T := h_{\mathcal{T}^\Gamma}(T) := |T|^{1/(d-1)},$$

where $|\cdot|$ denotes the $(d-1)$-dimensional surface measure of an element.

Let $G_T(x) := Dg_T(x)^\intercal\, Dg_T(x) \in \mathbb{R}^{(d-1)\times(d-1)}$ be the symmetric Gramian matrix of g_T and $\lambda_{\min}(G_T(x))$ as well as $\lambda_{\max}(G_T(x))$ the corresponding extremal eigenvalues. Now, we call a regular triangulation $\mathcal{T}^\Gamma$ a γ-**shape regular** triangulation, if the element maps g_T satisfy the following:

- For all $T \in \mathcal{T}^\Gamma$, it holds that

$$\sigma(T) := \sup_{x \in T_{\mathrm{ref}}} \Big(\frac{h_{\mathcal{T}^\Gamma}(T)^2}{\lambda_{\min}(G_T(x))} + \frac{\lambda_{\max}(G_T(x))}{h_{\mathcal{T}^\Gamma}(T)^2} \Big) \leq \gamma.$$

- If $d = 2$, it is explicitly required that

$$\widetilde{\sigma}(\mathcal{T}^\Gamma) := \max_{\substack{T, T' \in \mathcal{T}^\Gamma \\ T \cap T' \neq \emptyset}} \frac{|T|}{|T'|} \leq \gamma.$$

Since the Gramian matrix $G_T(x)$ is symmetric and positive definite, it holds that $0 \leq \lambda_{\min}(G_T) \leq \lambda_{\max}(G_T)$. This implies that $\sigma(T) \geq 1$. For $d = 2$, the additional assumption ensures that the mesh-sizes of neighboring elements remain comparable.

3.3 Discrete function spaces

For the approximation of the exact solutions of the different problems, we need finite-dimensional spaces which we introduce in this section. To this end, let $\mathcal{T}_\bullet^\Omega$ be a regular triangulation of Ω and $p \geq 1$ a fixed polynomial order. We define the space of globally continuous piecewise polynomials $\mathcal{S}^p(\mathcal{T}_\bullet^\Omega)$ by

$$\mathcal{S}^p(\mathcal{T}_\bullet^\Omega) := \big\{ v_\bullet \in C(\Omega) \,:\, v_\bullet|_T \text{ is a polynomial of degree } \leq p \text{ for all } T \in \mathcal{T}_\bullet^\Omega \big\}$$

It holds that $\mathcal{S}^p(\mathcal{T}_\bullet^\Omega) \subset H^1(\Omega)$ and we define the corresponding conforming subspace $\mathcal{S}_0^p(\mathcal{T}_\bullet^\Omega)$ of $H_0^1(\Omega)$ by

$$\mathcal{S}_0^p(\mathcal{T}_\bullet^\Omega) := \mathcal{S}^p(\mathcal{T}_\bullet^\Omega) \cap H_0^1(\Omega).$$

3.4 Mesh-refinement

Suppose that $\mathcal{T}_\bullet \in \{\mathcal{T}^\Omega, \mathcal{T}^\Gamma\}$ is a given regular and γ-shape regular triangulation. Additionally, assume that $\texttt{refine}(\cdot)$ is a fixed mesh-refinement strategy, e.g., newest vertex bisection, cf. [Ste08]. We write $\mathcal{T}_\circ = \texttt{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$ for the coarsest one-level refinement of $\mathcal{T}_\bullet$, where all marked elements $\mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet$ have been refined, i.e., $\mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet \backslash \mathcal{T}_\circ$. We write $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$, if $\mathcal{T}_\circ$ can be obtained by finitely many steps of one-level refinement (with appropriate, yet arbitrary marked elements in each step). We define $\mathbb{T} := \texttt{refine}(\mathcal{T}_0)$ as the set of all meshes which can be generated from the fixed initial mesh $\mathcal{T}_0$ by use of $\texttt{refine}(\cdot)$.

Some important properties of γ-shape regular meshes are collected in the next lemma. For boundary meshes, a proof can be found, e.g., in [AFF+17, Lemma 2.6].

Lemma 8. *Let $\mathcal{T}_\bullet \in \{\mathcal{T}^\Omega, \mathcal{T}^\Gamma\}$ be a γ-shape regular triangulation. Then, there exists a constant $C > 0$ that depends only on γ and, in case of a boundary mesh, additionally on the Lipschitz parametrization of $\partial\Omega$, such that the following assertions hold:*

(i) *For all $T, T' \in \mathcal{T}_\bullet$ with $T \cap T' \neq \emptyset$, it holds that $h_{\mathcal{T}_\bullet}(T) \leq C\, h_{\mathcal{T}_\bullet}(T')$.*

(ii) *The number of elements in an element patch is bounded by C, i.e., $\#(\omega_\bullet(T)) \leq C$ for all $T \in \mathcal{T}_\bullet$.*

(iii) *It holds that $\max_{T \in \mathcal{T}_\bullet} \frac{\operatorname{diam}(T)}{h_{\mathcal{T}_\bullet}} \leq C$.* □

For our analysis, we only employ the following structural properties (R1)–(R3), where $C_{\text{son}} \geq 2$ and $C_{\text{mesh}} > 0$ are generic constants:

(R1) splitting property: Each refined element is split into finitely many sons, i.e., for all $\mathcal{T}_\bullet \in \mathbb{T}$ and all $\mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet$, the mesh $\mathcal{T}_\circ = \texttt{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$ satisfies that

$$\#(\mathcal{T}_\bullet \setminus \mathcal{T}_\circ) + \#\mathcal{T}_\bullet \leq \#\mathcal{T}_\circ \leq C_{\text{son}}\, \#(\mathcal{T}_\bullet \setminus \mathcal{T}_\circ) + \#(\mathcal{T}_\bullet \cap \mathcal{T}_\circ).$$

(R2) overlay estimate: For all meshes $\mathcal{T} \in \mathbb{T}$ and $\mathcal{T}_\bullet, \mathcal{T}_\circ \in \texttt{refine}(\mathcal{T})$, there exists a common refinement $\mathcal{T}_\bullet \oplus \mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet) \cap \texttt{refine}(\mathcal{T}_\circ) \subseteq \texttt{refine}(\mathcal{T})$ such that

$$\#(\mathcal{T}_\bullet \oplus \mathcal{T}_\circ) \leq \#\mathcal{T}_\bullet + \#\mathcal{T}_\circ - \#\mathcal{T}.$$

(R3) mesh-closure estimate: For each sequence $(\mathcal{T}_\ell)_{\ell\in\mathbb{N}_0}$ of successively refined meshes, i.e., $\mathcal{T}_{\ell+1} := \texttt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ with $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ for all $\ell \in \mathbb{N}_0$, it holds that

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 \leq C_{\mathrm{mesh}} \sum_{j=0}^{\ell-1} \#\mathcal{M}_j.$$

3.5 Extended 1D bisection (EB)

For refining meshes on a 1-dimensional boundary $\Gamma \subseteq \partial\Omega$ with $\Omega \subset \mathbb{R}^2$, we consider the extended bisection algorithm (EB) from [AFF+13].

Algorithm 9. ***Input:*** *Mesh $\mathcal{T}_\bullet \in \mathbb{T} := \texttt{refine}(\mathcal{T}_0)$, set of marked elements $\mathcal{M}_\bullet^{(0)} := \mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet$, counter $k := 0$.*
Refinement Loop:

(i) **Repeat** *the following steps* (a)–(c)*:*

(a) *Update the counter $k \mapsto k+1$.*

(b) *Define* $\mathcal{U}^{(k)} := \bigcup_{T\in\mathcal{M}_\bullet^{(k-1)}} \big\{T' \in \mathcal{T}_\bullet\backslash\mathcal{M}_\bullet^{(k-1)} : T'\cap T \neq \emptyset$ *and* $h_\bullet|_{T'} > \sigma(\mathcal{T}_0)\, h_\bullet|_T\big\}$.

(c) *Define* $\mathcal{M}_\bullet^{(k)} := \mathcal{M}_\bullet^{(k-1)} \cup \mathcal{U}^{(k)}$

Until $\mathcal{U}^{(k)} = \emptyset$.

(ii) *Bisect all elements $T \in \mathcal{M}_\bullet^{(k)}$ to obtain $\mathcal{T}_\circ := \texttt{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$.*

Output: *Refined mesh $\mathcal{T}_\circ = \texttt{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$.*

Let $\mathcal{T}_0$ be the initial mesh on a 1-dimensional boundary $\Gamma \subseteq \partial\Omega$ with $\Omega \subset \mathbb{R}^2$. Due to the bisection in Algorithm 9, i.e., Step (ii), EB yields a contraction of the local mesh-size on refined elements, i.e., $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$ implies that

$$h_\circ|_T \leq 2^{-1}\, h_\bullet|_T \quad \text{for all } T \in \mathcal{T}_\bullet \setminus \mathcal{T}_\circ. \tag{3.1}$$

Additionally, [AFF+13, Theorem 2.3 (i)] guarantees uniform γ-shape regularity with $\gamma := 2\,\sigma(\mathcal{T}_0)$, i.e., for all triangulations $\mathcal{T}_\bullet \in \mathbb{T}$, it holds that

$$\widetilde{\sigma}(\mathcal{T}_\bullet) \leq \gamma. \tag{3.2}$$

Splitting property (R1)

Since Step (ii) of Algorithm 9 uses bisection, there holds (R1) with $C_{\mathrm{son}} = 2$.

Overlay estimate (R2)

The overlay estimate (R2) is shown in [AFF+13, Theorem 2.3 (ii)].

Mesh-closure estimate (R3)

The mesh-closure estimate (R3) is shown in [AFF+13, Theorem 2.3 (iii)].

3.6 Newest vertex bisection (NVB)

One of the most popular mesh-refinement strategies is the so-called *newest vertex bisection* (NVB), cf. e.g., [Ste07] for $d = 2$ as well as [Ste08] for $d = 3$. We use NVB for $d = 2$ as $\texttt{refine}(\cdot)$ to refine triangulations of a given domain $\Omega \subset \mathbb{R}^2$ in Chapter 4 as well as Chapter 5. Additionally, we also use the same algorithm for refining surface triangulations on $\Gamma \subseteq \partial\Omega$ with $\Omega \subset \mathbb{R}^3$ in Chapter 6.

For the sake of completeness, we include the NVB algorithm for $d = 2$:

Algorithm 10. Initialization: ***Input:*** *Initial mesh $\mathcal{T}_0$.*

- *For each triangle $T \in \mathcal{T}_0$, define an arbitrary vertex as the* **newest vertex**.
- *For each triangle $T \in \mathcal{T}_0$, define the edge opposite to the newest vertex as the* **reference edge** E_T. *Let $\mathcal{E}_{\mathrm{ref},0} := \{E_T : T \in \mathcal{T}_0\}$ be the set of all reference edges of the initial mesh $\mathcal{T}_0$.*

Newest Vertex Bisection: ***Input:*** *Mesh $\mathcal{T}_\bullet \in \mathbb{T}$ with corresponding set of reference edges $\mathcal{E}_{\mathrm{ref},\bullet} := \{E_T : T \in \mathcal{T}_\bullet\}$, set of marked elements $\mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet$, counter $k := 0$.*
Refinement Loop:

(i) *Define the set of marked reference edges $\mathcal{M}_\bullet^{(0)} := \{E_T : T \in \mathcal{M}_\bullet\}$.*

(ii) **Repeat** *the following steps* (a)–(b)*:*

(a) *Update the counter $k \mapsto k + 1$.*

(b) *Define $\mathcal{M}_\bullet^{(k)} := \{E_T : T \in \mathcal{T}_\bullet$ s.t. there exists $E \in \mathcal{M}_\bullet^{(k-1)}$ with $E \subset T\}$.*

Until $\mathcal{M}_\bullet^{(k)} = \mathcal{M}_\bullet^{(k-1)}$.

(iii) *Refine all elements $T \in \mathcal{T}_\bullet$ which have at least one marked edge in the set $\mathcal{M}_\bullet^{(k)}$ according to the refinement rules depicted in Figure 3.1.*

Output: *Refined mesh $\mathcal{T}_\circ = \texttt{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$.*

Let $\mathcal{T}_0$ be the initial mesh on a domain $\Omega \subset \mathbb{R}^d$ with $d \geq 2$ and let $\mathcal{T}_\bullet \in \mathbb{T}$ be a refinement of $\mathcal{T}_0$. It holds that NVB reduces the local mesh-size on refined elements, i.e., $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$ implies that

$$h_\circ|_T \leq 2^{-1/d}\, h_\bullet|_T \quad \text{for all } T \in \mathcal{T}_\bullet \setminus \mathcal{T}_\circ. \tag{3.3}$$

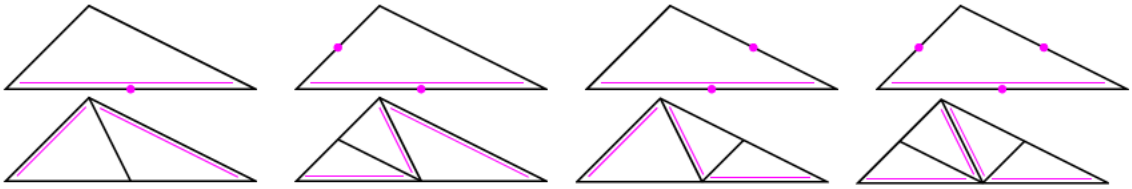


Figure 3.1: For each triangle $T \in \mathcal{T}_\bullet$, there is one fixed *reference edge* E_T, indicated by the extra pink line. If T is marked for refinement, we mark its reference edge, cf. Step (i) of Algorithm 10. Additionally, if $E_T \subset T'$ for a neigbouring element $T' \in \mathcal{T}_\bullet$, the edge reference edge $E_{T'}$ is marked to avoid hanging nodes, cf. Step (ii) of Algorithm 10. Hence, more than one edge of an element can be marked (pink dots). Then, refinement of T is done by bisecting the reference edge, where its midpoint becomes a new vertex of the refined triangulation $\mathcal{T}_\circ$. The reference edges of the son triangles are opposite to this newest vertex (bottom left). If more than one edge is marked (top), using iterated newest vertex bisection, the element is then split into 2, 3, or 4 son triangles (bottom).

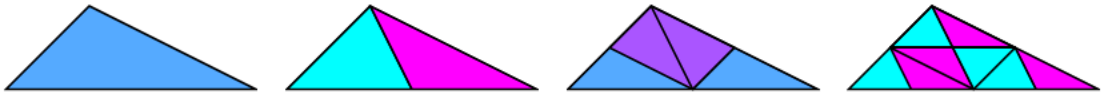


Figure 3.2: Newest vertex bisection does only lead (up to similarity) to a finite number of triangles. Above, the different colors represent similarity classes. Starting with one triangle (left), iterative use of NVB does only create (up to similarity) new triangles in the first two steps (mid left and mid right). Hence in following steps, no new similarity classes are generated.

A proof for (3.3) can be found, e.g., in [CKNS08, Ste07]. Additionally, NVB also preserves γ-shape regularity, i.e., there exists a constant $\gamma > 0$ such that for all triangulations $\mathcal{T}_\bullet \in \mathbb{T}$ it holds that

$$\sigma(\mathcal{T}_\bullet) = \max_{T \in \mathcal{T}_\bullet} \sigma(T) \leq \gamma, \tag{3.4}$$

which is proved in [Ste08]. The latter work also shows for $d = 3$ a similar result to Figure 3.2 which illustrates for $d = 2$ that (up to similarity) only a finite number of different triangles can be constructed from the initial mesh $\mathcal{T}_0$ using NVB, cf. [Ste08, Theorem 2.1].

Splitting property (R1)

There holds (R1) with $2 \leq C_{\mathrm{son}} < \infty$, which is proved in [GSS14]. The constant $C_{\mathrm{son}} > 0$ depends only on $\mathcal{T}_0$ and d. For $d = 2$, it holds that $C_{\mathrm{son}} = 4$, cf. Figure 3.1.

Overlay estimate (R2)

The proof of the overlay estimate (R2) can be found in [CKNS08, Ste07].

Mesh-closure estimate (R3)

First, the mesh-closure estimate (R3) has been proved for the case $d = 2$, cf. [BDD04]. Later, (R3) has been proved for $d \geq 2$ in [Ste08]. While both works [BDD04, Ste08] require a technical *admissibility condition* on $\mathcal{T}_0$ in order to prove the mesh-closure (R3), [KPP13] proved this admissibility condition to be unnecessary for $d = 2$.

3.7 Other refinement strategies

A different possible refinement strategy is red-refinement with first-order hanging nodes. We refer to [BN10], where the validity of (R1)–(R3) is shown. In the framework of isogeometric analysis, we mention the mesh-refinement techniques for analysis-suitable T-splines [MP15] and refer to [BGMP16] for truncated hierarchical B-splines as well as [GHP17] for hierarchical B-splines. For further details on mesh-refinement strategies which satisfy (R1)–(R3), we refer to [BN10, MP15, Fei15] and to the discussion in [CFPP14, Section 2.5].

4 Adaptive FEM for second-order elliptic systems of partial differential equations

4.1 Introduction

This chapter is based on the recent own work [GHPS21]. While the analytical main results are the same, we add an additional section on preconditioning and more in-depth numerical examples are provided. We consider and analyze adaptive finite element methods (AFEM) for second-order elliptic systems of partial differential equations (PDEs), where the arising discrete systems are not solved exactly. Our model problem reads as follows: Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with $d \in \{2, 3\}$ and boundary $\Gamma := \partial\Omega$. We assume that $A\colon L^2(\Omega)^d \to L^2(\Omega)^d$ is a strongly monotone and Lipschitz continuous operator, cf. Section 4.2 for the precise definition. We consider the following quasi-linear elliptic boundary value problem: Given a load $f \in L^2(\Omega)$, find $u^\star \in \mathcal{H} := H^1_0(\Omega)$ such that

$$\begin{aligned} -\operatorname{div} A(\nabla u^\star) &= f \quad \text{in } \Omega, \\ u^\star &= 0 \quad \text{on } \Gamma. \end{aligned} \tag{4.1}$$

Therefrom, we get the equivalent variational formulation: Given a load $f \in L^2(\Omega)$, find $u^\star \in \mathcal{H} := H^1_0(\Omega)$ such that

$$\langle \mathcal{A}u^\star , v\rangle_{\mathcal{H}'\times\mathcal{H}} := \int_\Omega A(\nabla u^\star)\cdot\nabla v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x =: \langle F , v\rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{for all } v \in \mathcal{H}. \tag{4.2}$$

The main theorem on monotone operators [Zei90, Section 25.4] admits a unique solution to the weak form (4.2). Given a discrete subspace $\mathcal{X}_\ell \subset \mathcal{H}$ related to some triangulation $\mathcal{T}_\ell$ of Ω, also the discrete formulation

$$\langle \mathcal{A}u_\ell^\star , v_\ell\rangle_{\mathcal{H}'\times\mathcal{H}} = \langle F , v_\ell\rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{for all } v_\ell \in \mathcal{X}_\ell \tag{4.3}$$

admits a unique solution $u_\ell^\star \in \mathcal{X}_\ell$, again due to the main theorem on monotone operators [Zei90, Section 25.4]. If A is nonlinear, then $u_\ell^\star$ can hardly be computed exactly. Even if A is linear, usual FEM codes employ iterative solvers like PCG, GMRES, or multigrid.

Given an initial guess $u_\ell^0 \in \mathcal{X}_\ell$, we assume that we can compute iterates $u_\ell^k := \Phi_\ell(u_\ell^{k-1}) \in \mathcal{X}_\ell$ which lead to a contraction in the energy norm on $\mathcal{H}$, i.e.,

$$|||u_\ell^\star - u_\ell^k||| \le q\,|||u_\ell^\star - u_\ell^{k-1}||| \quad \text{for all } k \in \mathbb{N} \tag{4.4}$$

with some $\mathcal{X}_\ell$-independent contraction constant $0 < q < 1$. In explicit terms, we assume that we have an iterative solver with iteration function $\Phi_\ell : \mathcal{X}_\ell \to \mathcal{X}_\ell$ which is uniformly contractive in each step. Additionally, we assume that we can control the discretization

error (for the exact, but never computed discrete solution $u_\ell^\star \in \mathcal{X}_\ell$ from (4.3)) by some reliable *a posteriori* error estimator

$$C_{\mathrm{rel}}^{-1} \, |||u^\star - u_\ell^\star||| \le \eta_\ell(u_\ell^\star) := \Big(\sum_{T \in \mathcal{T}_\ell} \eta_\ell(T, u_\ell^\star)^2 \Big)^{1/2}, \tag{4.5}$$

where the local indicators $\eta_\ell(T, \cdot)$ can also be evaluated for other discrete functions $v_\ell \in \mathcal{X}_\ell$ instead of the exact Galerkin solution $u_\ell^\star \in \mathcal{X}_\ell$.

Then, our adaptive algorithm takes the form

$$\boxed{\text{Iteratively Solve \& Estimate}} \longrightarrow \boxed{\text{Mark}} \longrightarrow \boxed{\text{Refine}} \tag{4.6}$$

where the first step may be understood (and stated) as an inner loop, and $\boxed{\text{Mark}}$ is based on the Dörfler criterion from [Dör96] with (quasi-) minimal cardinality [Ste07, PP20].

4.1.1 State of the art

The ultimate goal of any numerical scheme is to compute a discrete solution with error below a prescribed tolerance at, up to a multiplicative constant, the minimal computational cost. Since the convergence of numerical methods is usually spoiled by singularities of the (given) data as well as the (unknown) solution, *a posteriori* error estimation and related adaptive mesh-refinement strategies are indispensable tools for reliable numerical simulations. For many model problems, the mathematical understanding of rate-optimal convergence of adaptive FEM has matured. We refer to [Dör96, MNS00, BDD04, Ste07, CKNS08, CN12, FFP14] for some seminal works for linear problems, to [Vee02, DK08, BDK12, GMZ12] for nonlinear problems, and to [CFPP14] for a general framework of convergence of adaptive FEM with optimal convergence rates. Some works also account for the approximate computation of the discrete solutions by iterative (and inexact) solvers, see, e.g., [BMS10, AGL13] for linear problems and [GMZ11, GHPS18, HW20a, HW20b] for nonlinear model problems. Moreover, there are many papers on *a posteriori* error estimation which also include the iterative and inexact solution for nonlinear problems, see, e.g., [EAEV11, EV13, AW15, HW18] and the references therein.

As far as optimal convergence rates are concerned, the mentioned works focus on rates with respect to the degrees of freedom. However, in practice, one aims for the optimal rate of convergence with respect to the computational cost, i.e., the computational time. The issue of optimal computational cost is already addressed in the seminal work [Ste07] for the Poisson model problem. There, it is assumed that a sufficiently accurate discrete solution can be computed in linear complexity, e.g., by a multigrid solver. Under these so-called *realistic assumptions* on the solver, it is then proved that the *total error* (i.e., the sum of energy error plus data oscillations) will also converge with optimal rate with respect to the computational cost. A similar result is obtained in [CG12] for an adaptive Laplace eigenvalue computation.

In recent works, concrete solvers are included into the convergence analysis. In [GHPS18], adaptive FEM for an elliptic PDE with strongly monotone nonlinearity is adressed. The arising nonlinear FEM problems are linearized via the so-called *Zarantonello iteration* (or

Banach–Picard iteration), which leads to a *linear* Poisson problem in each step. The adaptive algorithm drives the linearization strategy as well as the local mesh-refinement. In [GHPS18], it is proved that the overall strategy leads to optimal convergence rates with respect to the degrees of freedom and to *almost optimal* convergence rates with respect to the total computational cost. The latter means that, if the total error converges with rate $s > 0$ with respect to the degrees of freedom, then it converges with rate $s - \varepsilon > 0$ with respect to the overall computational cost, for all $\varepsilon > 0$. Moreover, in [FHPS19] (cf. Chapter 6), we obtained analogous results for an adaptive boundary element method, where we employed a preconditioned conjugate gradient method (PCG) with optimal additive Schwarz preconditioner to approximately solve the arising linear discrete systems.

We now aim to prove *optimal rates with respect to the overall computational cost* for the algorithm from [GHPS18]. Moreover, we give an abstract analysis in the spirit of [CFPP14] and show that this also covers linear solvers like PCG.

4.1.2 Outline

First, we formulate the precise assumptions on the model problem, the mesh-refinement and the FEM spaces (Section 4.2), and the error estimator as well as the iterative solver (Section 4.3–4.4). Then, we formulate the adaptive algorithm in Section 4.5 and state the abstract main results in Section 4.6, namely linear convergence of the quasi-error in Section 4.6.1 and optimal convergence rates of the quasi-error in Section 4.6.3. Before we then apply the abstract setting to adaptive FEM with PCG solver for linear PDEs (Section 4.7) including numerical examples (Section 4.7.7), we construct an additive Schwarz preconditioner in Section 4.7.1 and prove its optimality in Section 4.7.3. Afterwards, we apply the abstract setting to the the adaptive algorithm from [GHPS18] for adaptive FEM for problems with strongly monotone nonlinearity (Section 4.8) including some numerical experiments in Section 4.8.1 to underline the theoretical findings.

4.2 Abstract model problem

Let $\mathcal{H}$ be a Hilbert space over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ with scalar product $\langle\!\langle \cdot\,, \cdot \rangle\!\rangle$ and corresponding norm $|\!|\!| \cdot |\!|\!|$. The usual dual space of $\mathcal{H}$ is denoted by $\mathcal{H}'$ with the corresponding norm $|\!|\!| \cdot |\!|\!|'$. We consider nonlinear elliptic equations in the following abstract setting with variational formulation: Given a linear and continuous functional $F \in \mathcal{H}'$, find $u^\star \in \mathcal{H}$ such that

$$\langle \mathcal{A}u^\star\,, v\rangle_{\mathcal{H}'\times\mathcal{H}} = \langle F\,, v\rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{for all } v \in \mathcal{H}. \tag{4.7}$$

To guarantee solvability, we suppose that the operator $\mathcal{A}\colon \mathcal{H} \to \mathcal{H}'$ satisfies the following conditions:

(O1) $\mathcal{A}$ is strongly monotone: There exists a constant $\alpha > 0$ such that

$$\alpha\,|\!|\!|w - v|\!|\!|^2 \le \operatorname{Re}\langle \mathcal{A}w - \mathcal{A}v\,, w - v\rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{for all } v, w \in \mathcal{H}.$$

(O2) $\mathcal{A}$ is Lipschitz continuous: There exists a constant $L > 0$ such that

$$|\!|\!|\mathcal{A}w - \mathcal{A}v|\!|\!|' \le L\,|\!|\!|w - v|\!|\!| \quad \text{for all } v, w \in \mathcal{H}.$$

(O3) $\mathcal{A}$ **has a potential:** There exists a Gâteaux differentiable function $P\colon \mathcal{H} \to \mathbb{K}$ such that its derivative $\mathrm{d}P\colon \mathcal{H} \to \mathcal{H}'$ coincides with $\mathcal{A}$, i.e., it holds that

$$\langle \mathcal{A}w\,,\, v\rangle_{\mathcal{H}'\times\mathcal{H}} = \langle \mathrm{d}P(w)\,,\, v\rangle_{\mathcal{H}'\times\mathcal{H}} = \lim_{\substack{t\to 0\\ t\in\mathbb{R}}} \frac{P(w+tv)-P(w)}{t} \quad \text{for all } v, w \in \mathcal{H}.$$

Let $\mathcal{T}_0$ be a given regular initial mesh and suppose that $\texttt{refine}(\cdot)$ is a fixed refinement strategy satisfying the axioms (R1)–(R3) from Section 3.4. To each $\mathcal{T}_\bullet \in \mathbb{T} := \texttt{refine}(\mathcal{T}_0)$, we associate the related finite-dimensional conforming subspace $\mathcal{X}_\bullet \subset \mathcal{H}$ of the given Hilbert space $\mathcal{H}$. We suppose that refinement $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$ leads to nestedness of the corresponding subspaces in the sense that $\mathcal{X}_\bullet \subseteq \mathcal{X}_\circ$.

Then, the discrete formulation of (4.7) reads as follows: Given a linear and continuous functional $F \in \mathcal{H}'$, find $u_\bullet^\star \in \mathcal{X}_\bullet$ such that

$$\langle \mathcal{A}u_\bullet^\star\,,\, v_\bullet\rangle_{\mathcal{H}'\times\mathcal{H}} = \langle F\,,\, v_\bullet\rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{for all } v_\bullet \in \mathcal{X}_\bullet. \tag{4.8}$$

The main theorem on monotone operators [Zei90, Section 25.4] yields existence and uniqueness of solutions $u^\star \in \mathcal{H}$ as well as $u_\bullet^\star \in \mathcal{X}_\bullet$ for both the model problem (4.7) and its discrete version (4.8), respectively.

Let $\mathcal{E} := \mathrm{Re}\,(P - F)$ be the energy functional. Then, it holds that

$$\frac{\alpha}{2}\,|||u_\bullet^\star - v_\bullet|||^2 \le \mathcal{E}(v_\bullet) - \mathcal{E}(u_\bullet^\star) \le \frac{L}{2}\,|||u_\bullet^\star - v_\bullet|||^2 \quad \text{for all } v_\bullet \in \mathcal{X}_\bullet, \tag{4.9}$$

which is proved, e.g., in [GHPS18, Lemma 5.1]. In particular, $u^\star \in \mathcal{H}$ is the unique minimizer of the minimization problem

$$\mathcal{E}(u^\star) = \min_{v\in\mathcal{H}} \mathcal{E}(v), \tag{4.10}$$

as well as $u_\bullet^\star \in \mathcal{X}_\bullet$ is the unique minimizer of the minimization problem

$$\mathcal{E}(u_\bullet^\star) = \min_{v_\bullet\in\mathcal{X}_\bullet} \mathcal{E}(v_\bullet). \tag{4.11}$$

As for linear elliptic problems, the present setting guarantees the following Céa lemma, where we include the proof for the sake of completeness.

Lemma 11. *Suppose that the operator $\mathcal{A}$ satisfies (O1)–(O2) with constants $0 < \alpha \le L$. Then, it holds with $C_{\text{Céa}} := L/\alpha$ that*

$$|||u^\star - u_\bullet^\star||| \le C_{\text{Céa}} \min_{v_\bullet\in\mathcal{X}_\bullet} |||u^\star - v_\bullet|||. \tag{4.12}$$

Proof. There holds the Galerkin orthogonality $\langle \mathcal{A}u^\star - \mathcal{A}u_\bullet^\star\,,\, v_\bullet\rangle_{\mathcal{H}'\times\mathcal{H}} = 0$ for all $v_\bullet \in \mathcal{X}_\bullet$. Let $w_\bullet \in \mathcal{X}_\bullet$ and $u^\star \neq u_\bullet^\star$. Then, it holds that

$$\begin{aligned}\alpha|||u^\star - u_\bullet^\star||| &\overset{\text{(O1)}}{\le} \frac{\mathrm{Re}\,\langle \mathcal{A}u^\star - \mathcal{A}u_\bullet^\star\,,\, u^\star - u_\bullet^\star\rangle_{\mathcal{H}'\times\mathcal{H}}}{|||u^\star - u_\bullet^\star|||}\\ &= \frac{\mathrm{Re}\,\langle \mathcal{A}u^\star - \mathcal{A}u_\bullet^\star\,,\, u^\star - w_\bullet\rangle_{\mathcal{H}'\times\mathcal{H}}}{|||u^\star - u_\bullet^\star|||} \overset{\text{(O2)}}{\le} L\,|||u^\star - w_\bullet|||.\end{aligned}$$

Hence, we take the infimum over all $w_\bullet \in \mathcal{X}_\bullet$. Since $\mathcal{X}_\bullet$ is finite-dimensional, the infimum is attained and is, in fact, a minimum. □

4.3 Error estimator

For each mesh $\mathcal{T}_\bullet \in \mathbb{T}$, suppose that we can compute refinement indicators

$$\eta_\bullet(T, v_\bullet) \geq 0 \quad \text{for all } T \in \mathcal{T}_\bullet \text{ and all } v_\bullet \in \mathcal{X}_\bullet. \tag{4.13}$$

To abbreviate notation, let $\eta_\bullet(v_\bullet) := \eta_\bullet(\mathcal{T}_\bullet, v_\bullet)$, where

$$\eta_\bullet(\mathcal{U}_\bullet, v_\bullet) := \Big(\sum_{T \in \mathcal{U}_\bullet} \eta_\bullet(T, v_\bullet)^2 \Big)^{1/2} \quad \text{for all } \mathcal{U}_\bullet \subseteq \mathcal{T}_\bullet. \tag{4.14}$$

We assume the following *axioms of adaptivity* from [CFPP14], where $C_{\mathrm{stab}}, C_{\mathrm{rel}} > 0$ and $0 < q_{\mathrm{red}} < 1$ are generic constants:

(A1) stability on non-refined element domains: For all triangulations $\mathcal{T}_\bullet \in \mathbb{T}$ and refinements $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$, arbitrary discrete functions $v_\circ \in \mathcal{X}_\circ$ and $w_\bullet \in \mathcal{X}_\bullet$, and an arbitrary set $\mathcal{U}_\bullet \subseteq \mathcal{T}_\bullet \cap \mathcal{T}_\circ$ of non-refined elements, it holds that

$$|\eta_\circ(\mathcal{U}_\bullet, v_\circ) - \eta_\bullet(\mathcal{U}_\bullet, w_\bullet)| \leq C_{\mathrm{stab}} |\!|\!| v_\circ - w_\bullet |\!|\!|.$$

(A2) reduction on refined elements: For all triangulations $\mathcal{T}_\bullet \in \mathbb{T}$ and refinements $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$, and arbitrary discrete functions $v_\bullet \in \mathcal{X}_\bullet$, it holds that

$$\eta_\circ(\mathcal{T}_\circ \backslash \mathcal{T}_\bullet, v_\bullet) \leq q_{\mathrm{red}}\, \eta_\bullet(\mathcal{T}_\bullet \backslash \mathcal{T}_\circ, v_\bullet).$$

(A3) reliability: For all triangulations $\mathcal{T}_\bullet \in \mathbb{T}$, the error of the exact discrete solution $u_\bullet^\star \in \mathcal{X}_\bullet$ of (4.8) can be bound by the error estimator, i.e.,

$$|\!|\!| u^\star - u_\bullet^\star |\!|\!| \leq C_{\mathrm{rel}}\, \eta_\bullet(u_\bullet^\star).$$

(A4) discrete reliability: For all triangulations $\mathcal{T}_\bullet \in \mathbb{T}$ and refinements $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$, the difference of the exact solutions $u_\bullet^\star \in \mathcal{X}_\bullet$ and $u_\circ^\star \in \mathcal{X}_\circ$ can be bounded by

$$|\!|\!| u_\circ^\star - u_\bullet^\star |\!|\!| \leq C_{\mathrm{rel}}\, \eta_\bullet(\mathcal{T}_\bullet \backslash \mathcal{T}_\circ, u_\bullet^\star).$$

We stress that the exact discrete solutions $u_\bullet^\star \in \mathcal{X}_\bullet$ and $u_\circ^\star \in \mathcal{X}_\circ$ in (A3)–(A4) will never be computed but are only auxiliary quantities for the analysis.

Remark 12. *The verification of (A1)–(A4) in Section 4.7 and 4.8 relies on scaling arguments and implicitly uses that all meshes $\mathcal{T}_\bullet \in \mathbb{T}$ are uniformly shape regular. Moreover, we note that the analysis is implicitly tailored to weighted-residual error estimators, since the usual verification of (A2) relies on exploiting the contraction of the mesh-size on refined elements.*

4.4 Discrete iterative solver

For all triangulations $\mathcal{T}_\bullet \in \mathbb{T}$, let $\Phi_\bullet\colon \mathcal{X}_\bullet \to \mathcal{X}_\bullet$ be the iteration function of one step of the iterative solver, i.e., for a given initial guess $u_\bullet^0 \in \mathcal{X}_\bullet$, we can compute iterates $u_\bullet^k := \Phi_\bullet(u_\bullet^{k-1}) \in \mathcal{X}_\bullet$. We require one of the following two contraction properties with some uniform constant $0 < q_{\mathrm{ctr}} < 1$, which is independent of $\mathcal{T}_\bullet$:

(C1) energy contraction: For all triangulations $\mathcal{T}_\bullet \in \mathbb{T}$ and an arbitrary discrete function $v_\bullet \in \mathcal{X}_\bullet$, it holds that

$$\mathcal{E}(\Phi_\bullet(v_\bullet)) - \mathcal{E}(u_\bullet^\star) \leq q_{\mathrm{ctr}}^2 \big(\mathcal{E}(v_\bullet) - \mathcal{E}(u_\bullet^\star)\big).$$

(C2) norm contraction: For all triangulations $\mathcal{T}_\bullet \in \mathbb{T}$ and an arbitrary discrete function $v_\bullet \in \mathcal{X}_\bullet$, it holds that

$$|||u_\bullet^\star - \Phi_\bullet(v_\bullet)||| \leq q_{\mathrm{ctr}} |||u_\bullet^\star - v_\bullet|||.$$

Remark 13. *For linear symmetric problems, one usually has that* $\mathcal{E}(v_\bullet) - \mathcal{E}(u_\bullet^\star) = \frac{1}{2} |||v_\bullet - u_\bullet^\star|||^2$ *for* $v_\bullet \in \mathcal{X}_\bullet$*, and hence* (C1) *and* (C2) *are equivalent.*

To formulate the stopping criterion for the iterative solver of the adaptive algorithm, we need an additional auxiliary quantity. Let

$$\mathrm{dl}(w, v) := \begin{cases} |\mathcal{E}(v) - \mathcal{E}(w)|^{1/2} & \text{in case of (C1),} \\ |||w - v||| & \text{in case of (C2).} \end{cases} \tag{4.15}$$

Then, the following lemma provides the means to stop the iterative solver.

Lemma 14. *Let* $\mathcal{T}_\bullet \in \mathbb{T}$ *and* $v_\bullet \in \mathcal{X}_\bullet$*. Then, both* (C1) *and* (C2)*, respectively, imply the following estimates:*

(i) $\mathrm{dl}(u_\bullet^\star, \Phi(v_\bullet)) \leq q_{\mathrm{ctr}}\, \mathrm{dl}(u_\bullet^\star, v_\bullet)$,

(ii) $\mathrm{dl}(v_\bullet, \Phi(v_\bullet)) \leq (1 + q_{\mathrm{ctr}})\, \mathrm{dl}(u_\bullet^\star, v_\bullet)$,

(iii) $\mathrm{dl}(u_\bullet^\star, v_\bullet) \leq (1 - q_{\mathrm{ctr}})^{-1}\, \mathrm{dl}(v_\bullet, \Phi(v_\bullet))$.

Proof. First, let assumption (C1) hold true. From the definition of $\mathrm{dl}(\cdot,\cdot)$ follows that

$$\mathrm{dl}(u_\bullet^\star, \Phi(v_\bullet)) \overset{(4.15)}{=} |\mathcal{E}(\Phi(v_\bullet)) - \mathcal{E}(u_\bullet^\star)|^{1/2} \overset{(C1)}{\leq} q_{\mathrm{ctr}} |\mathcal{E}(v_\bullet) - \mathcal{E}(u_\bullet^\star)|^{1/2} = q_{\mathrm{ctr}}\, \mathrm{dl}(u_\bullet^\star, v_\bullet).$$

Hence, claim (i) holds true. Note that $\mathrm{dl}(\cdot,\cdot)$ is a quasi-metric, i.e., it holds for all $v_\bullet, w_\bullet, z_\bullet \in \mathcal{X}_\bullet$ that

- $\mathrm{dl}(v_\bullet, v_\bullet) = 0$,
- $\mathrm{dl}(v_\bullet, w_\bullet) = \mathrm{dl}(w_\bullet, v_\bullet)$, and,

- $\mathrm{dl}(v_\bullet, z_\bullet) \leq \mathrm{dl}(v_\bullet, w_\bullet) + \mathrm{dl}(w_\bullet, z_\bullet)$,

where the triangle inequality follows from the fact that $(a+b)^{1/2} \leq a^{1/2} + b^{1/2}$ for $a, b > 0$. Therefrom, we get with claim (i) that

$$\mathrm{dl}(v_\bullet, \Phi(v_\bullet)) \leq \mathrm{dl}(v_\bullet, u_\bullet^\star) + \mathrm{dl}(u_\bullet^\star, \Phi(v_\bullet)) \leq (1 + q_{\mathrm{ctr}})\, \mathrm{dl}(u_\bullet^\star, v_\bullet),$$

which proves claim (ii). Claim (iii) also follows from the triangle inequality combined witch claim (i). It holds that

$$\mathrm{dl}(u_\bullet^\star, v_\bullet) \leq \mathrm{dl}(u_\bullet^\star, \Phi(v_\bullet)) + \mathrm{dl}(\Phi(v_\bullet), v_\bullet) \leq q_{\mathrm{ctr}}\, \mathrm{dl}(u_\bullet^\star, v_\bullet) + \mathrm{dl}(v_\bullet, \Phi(v_\bullet)),$$

which is equivalent to claim (iii).

Now, let assumption (C2) hold true. Then, claim (i) is simply the norm contraction (C2) and claim (ii)–(iii) follow from the triangle inequality of the energy norm. □

4.5 Adaptive algorithm

Now, we propose our adaptive algorithm. We will employ a lower index ℓ for the adaptive mesh-refinement as well as an upper index k for the respective steps of the iterative solver.

Algorithm 15. ***Input:*** *Initial mesh $\mathcal{T}_0$ and initial guess $u_0^0 \in \mathcal{X}_0$, adaptivity parameters $0 < \theta \leq 1$, $\lambda_{\mathrm{ctr}} > 0$, and $C_{\mathrm{mark}} \geq 1$, counters $\ell := 0 =: k$.*
Adaptive Loop: *Iterate the following Steps* (i)–(v)*:*

(i) **Repeat** *the following steps* (a)–(c)*:*

(a) *Update the counter $(\ell, k) \mapsto (\ell, k+1)$.*

(b) *Do one step of the iterative solver to obtain $u_\ell^k := \Phi_\ell(u_\ell^{k-1})$.*

(c) *Compute the local contributions $\eta_\ell(T, u_\ell^k)$ of the error estimator for all $T \in \mathcal{T}_\ell$.*

Until $\mathrm{dl}(u_\ell^k, u_\ell^{k-1}) \leq \lambda_{\mathrm{ctr}}\, \eta_\ell(u_\ell^k)$. (4.16)

(ii) *Define $\underline{k}(\ell) := k$.*

(iii) *Determine a set $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ with up to the multiplicative constant C_{mark} minimal cardinality such that*

$$\theta\, \eta_\ell(u_\ell^{\underline{k}}) \leq \eta_\ell(\mathcal{M}_\ell, u_\ell^{\underline{k}}). \tag{4.17}$$

(iv) *Generate $\mathcal{T}_{\ell+1} := \mathtt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ and define $u_{\ell+1}^0 := u_\ell^{\underline{k}(\ell)}$.*

(v) *Update the counter $(\ell, k) \mapsto (\ell+1, 0)$ and continue with* (i)*.*

Output: *Sequences of successively refined triangulations $\mathcal{T}_\ell$, discrete solutions u_ℓ^k, and corresponding error estimators $\eta_\ell(u_\ell^k)$, for all $\ell \geq 0$ and $k \geq 0$.*

We define the following set of indices $\mathcal{Q}$ by

$$\mathcal{Q} := \big\{(\ell,k) \in \mathbb{N}_0^2 \,:\, \text{index pair } (\ell,k) \text{ is used in Algorithm 15 and } k < \underline{k}(\ell)\big\}.$$

Since $u_{\ell+1}^0 = u_\ell^{\underline{k}(\ell)}$, we exclude $(\ell,\underline{k}(\ell))$ from the index set $\mathcal{Q}$, if $(\ell+1,0) \in \mathcal{Q}$. Since Algorithm 15 is sequential, the index set $\mathcal{Q}$ is naturally ordered. For $(\ell,k),(\ell',k') \in \mathcal{Q}$, we write

$$(\ell',k') < (\ell,k) \quad \overset{\text{def}}{\iff} \quad (\ell',k') \text{ appears earlier in Algorithm 15 than } (\ell,k). \tag{4.18}$$

With this order, we can define the *total step counter*

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

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

To abbreviate notation, we make the convention that if the mesh index $\ell \in \mathbb{N}_0$ is clear from the context, we simply write $\underline{k} := \underline{k}(\ell)$, e.g., $u_\ell^{\underline{k}} := u_\ell^{\underline{k}(\ell)}$. In addition, we introduce some further notation. Define

$$\underline{\ell} := \sup\big\{\ell \in \mathbb{N}_0 \,:\, (\ell,0) \in \mathcal{Q}\big\}.$$

Generically, it holds that $\underline{\ell} = \infty$, i.e., infinitely many steps of mesh-refinement occur. Moreover, for $(\ell,0) \in \mathcal{Q}$, define $\underline{k}(\ell) := \sup\big\{k \in \mathbb{N}_0 \,:\, (\ell,k) \in \mathcal{Q}\big\} + 1$. We note that the latter definition is consistent with that of Algorithm 15, but additionally defines $\underline{k}(\underline{\ell}) = \infty$ if $\underline{\ell} < \infty$.

4.6 Abstract main results

In this section, we state the main results in the abstract framework of Section 4.2. The analysis relies only on the assumptions (R1)–(R3) on the mesh-refinement, (A1)–(A4) on the error estimator, and (C1) as well as (C2) on the iterative solver respectively. Hence, for concrete model problems, only these assumptions have to be verified, cf. Section 4.7 and Section 4.8.

First, due to the contraction property (C1) and (C2) respectively, we have *a posteriori* error control of the error.

Proposition 16. *Suppose (C1) or (C2) as well as (A1)–(A3). Then, the quasi-error Δ_ℓ^k (consisting of error plus error estimator), which is defined via*

$$\Delta_\ell^k := |||u^\star - u_\ell^k||| + \eta_\ell(u_\ell^k) \quad \text{for all } (\ell,k) \in \overline{\mathcal{Q}} := \mathcal{Q} \cup \big\{(\ell,\underline{k}) \,:\, \underline{k}(\ell) < \infty\big\}, \tag{4.19}$$

satisfies that

$$\Delta_\ell^k \le C'_{\mathrm{rel}} \begin{cases} \eta_\ell(u_\ell^k) + \mathrm{dl}(u_\ell^k, u_\ell^{k-1}) & \text{if } 0 < k \le \underline{k}(\ell), \\ \eta_\ell(u_\ell^{\underline{k}}) & \text{if } k = \underline{k}(\ell), \\ \eta_{\ell-1}(u_\ell^0) & \text{if } k = 0 \text{ and } \ell > 0. \end{cases} \tag{4.20}$$

The constant $C'_{\mathrm{rel}} > 0$ depends only on C_{stab}, C_{rel}, q_{ctr}, and λ_{ctr} under (C2), while it additionally depends on α under (C1).

Proof. Let $(\ell, k) \in \overline{\mathcal{Q}}$ and $k > 0$. Then, it holds that

$$\begin{aligned} \|\!| u^\star - u_\ell^k |\!\| &\le \|\!| u^\star - u_\ell^\star |\!\| + \|\!| u_\ell^\star - u_\ell^k |\!\| \\ &\overset{(A3)}{\le} C_{\mathrm{rel}}\, \eta_\ell(u_\ell^\star) + \|\!| u_\ell^\star - u_\ell^k |\!\| \\ &\le C_{\mathrm{rel}} \big(|\eta_\ell(u_\ell^\star) - \eta_\ell(u_\ell^k)| + \eta_\ell(u_\ell^k) \big) + \|\!| u_\ell^\star - u_\ell^k |\!\| \\ &\overset{(A1)}{\le} C_{\mathrm{rel}}\, \eta_\ell(u_\ell^k) + (C_{\mathrm{rel}} C_{\mathrm{stab}} + 1)\, \|\!| u_\ell^\star - u_\ell^k |\!\|. \end{aligned}$$

Now, we distinguish between the different contraction properties. First, suppose (C1). With (4.9) and Lemma 14(i)&(iii), it then follows that

$$\begin{aligned} \|\!| u_\ell^\star - u_\ell^k |\!\| &\overset{(4.9)}{\le} \sqrt{2/\alpha}\, \mathrm{dl}(u_\ell^\star, u_\ell^k) \\ &= \sqrt{2/\alpha}\, \mathrm{dl}(u_\ell^\star, \Phi(u_\ell^{k-1})) \\ &\le \sqrt{2/\alpha}\, q_{\mathrm{ctr}}\, \mathrm{dl}(u_\ell^\star, u_\ell^{k-1}) \\ &\le \sqrt{2/\alpha}\, \frac{q_{\mathrm{ctr}}}{1 - q_{\mathrm{ctr}}}\, \mathrm{dl}(u_\ell^k, u_\ell^{k-1}). \end{aligned}$$

Next, suppose (C2). With Lemma 14 (i)&(iii), it then follows that

$$\begin{aligned} \|\!| u_\ell^\star - u_\ell^k |\!\| &= \mathrm{dl}(u_\ell^\star, \Phi(u_\ell^{k-1})) \\ &\le q_{\mathrm{ctr}}\, \mathrm{dl}(u_\ell^\star, u_\ell^{k-1}) \\ &\le \frac{q_{\mathrm{ctr}}}{1 - q_{\mathrm{ctr}}}\, \mathrm{dl}(u_\ell^k, u_\ell^{k-1}). \end{aligned}$$

Since $\Delta_\ell^k = \|\!| u^\star - u_\ell^k |\!\| + \eta_\ell(u_\ell^k)$, this proves (4.20) for the case that $0 < k \le \underline{k}(\ell)$. If $k = \underline{k}(\ell)$, the stopping criterion (4.16) in Algorithm 15(i) yields that

$$\mathrm{dl}(u_\ell^{\underline{k}}, u_\ell^{\underline{k}-1}) \le \lambda_{\mathrm{ctr}}\, \eta_\ell(u_\ell^{\underline{k}}).$$

This proves (4.20) for $k = \underline{k}(\ell)$. If $k = 0$ and $\ell > 0$, it holds that $u_\ell^0 = u_{\ell-1}^{\underline{k}}$. Hence, it follows from the previous step that

$$\|\!| u^\star - u_\ell^0 |\!\| = \|\!| u^\star - u_{\ell-1}^{\underline{k}} |\!\| \lesssim \eta_{\ell-1}(u_{\ell-1}^{\underline{k}}) = \eta_{\ell-1}(u_\ell^0). \tag{4.21}$$

Moreover, the equality $u_\ell^0 = u_{\ell-1}^k$ implies that $u_\ell^0 \in \mathcal{X}_{\ell-1}$. Therefrom, (A1)–(A2) yield that

$$\begin{aligned}\eta_\ell(u_\ell^0) &= \big(\eta_\ell(\mathcal{T}_\ell \cap \mathcal{T}_{\ell-1}, u_\ell^0)^2 + \eta_\ell(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell-1}, u_\ell^0)^2\big)^{1/2} \\ &\overset{\text{(A1)}}{=} \big(\eta_{\ell-1}(\mathcal{T}_\ell \cap \mathcal{T}_{\ell-1}, u_\ell^0)^2 + \eta_\ell(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell-1}, u_\ell^0)^2\big)^{1/2} \\ &\overset{\text{(A2)}}{\leq} \big(\eta_{\ell-1}(\mathcal{T}_\ell \cap \mathcal{T}_{\ell-1}, u_\ell^0)^2 + \eta_{\ell-1}(\mathcal{T}_{\ell-1} \setminus \mathcal{T}_\ell, u_\ell^0)^2\big)^{1/2} \\ &= \eta_{\ell-1}(u_\ell^0).\end{aligned} \tag{4.22}$$

Since $\Delta_\ell^0 = |||u^\star - u_\ell^0||| + \eta_\ell(u_\ell^0)$, combining (4.21)–(4.22) concludes the proof. □

4.6.1 Linear convergence of the quasi-error

The first main theorem states linear convergence of the quasi-error. We note that under certain assumptions, linear convergence holds for arbitrary parameters $0 < \theta \leq 1$ and $\lambda_{\mathrm{ctr}} > 0$.

Theorem 17. *Suppose (C1) or (C2) as well as (A1)–(A3). Define*

$$\lambda_{\mathrm{conv}} := \begin{cases} \infty & \text{if (C1) is valid,} \\ \frac{1-q_{\mathrm{ctr}}}{C_{\mathrm{stab}} q_{\mathrm{ctr}}} & \text{otherwise.} \end{cases} \tag{4.23}$$

Then, for all $0 < \theta \leq 1$ and $0 < \lambda_{\mathrm{ctr}} < \lambda_{\mathrm{conv}}\,\theta$, there exist constants $C_{\mathrm{lin}} \geq 1$ and $0 < q_{\mathrm{lin}} < 1$ such that the quasi-error (4.19) is linearly convergent in the sense of

$$\Delta_\ell^k \leq C_{\mathrm{lin}}\, q_{\mathrm{lin}}^{|(\ell,k)|-|(\ell',k')|}\, \Delta_{\ell'}^{k'} \qquad \text{for all } (\ell,k),(\ell',k') \in \mathcal{Q} \text{ with } (\ell',k') < (\ell,k). \tag{4.24}$$

The constants C_{lin} and q_{lin} depend only on $C_{\mathrm{Céa}} = L/\alpha$, C_{stab}, q_{red}, C_{rel}, q_{ctr}, and the adaptivity parameters θ and λ_{ctr}, while it additionally depends on L in case of (C1).

The following corollary states that the exact solution $u^\star$ is discrete if $\underline{\ell} < \infty$, i.e., if the number of mesh refinements is bounded.

Corollary 18. *Suppose the assumptions of Theorem 17. Then, $\underline{\ell} < \infty$ implies that $u^\star = u_{\underline{\ell}}^\star$ and $\eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) = 0$.*

Proof. According to Theorem 17, it holds that

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

Moreover, contraction (C1) or (C2) (together with (4.9) in case of (C1)) prove that

$$|||u_{\underline{\ell}}^\star - u_{\underline{\ell}}^k||| \simeq \mathrm{dl}(u_{\underline{\ell}}^\star, u_{\underline{\ell}}^k) \leq q_{\mathrm{ctr}}^k\, \mathrm{dl}(u_{\underline{\ell}}^\star, u_{\underline{\ell}}^0) \to 0 \quad \text{as } k \to \infty.$$

Uniqueness of the limit yields that $u_{\underline{\ell}}^\star = u^\star$. Moreover, it follows that

$$0 \leq \eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) \overset{\text{(A1)}}{\leq} \eta_{\underline{\ell}}(u_{\underline{\ell}}^k) + |||u_{\underline{\ell}}^\star - u_{\underline{\ell}}^k||| \to 0 \quad \text{as } k \to \infty.$$

This concludes the proof. □

4.6.2 Proof of Theorem 17 (linear convergence)

Recall the definition of $\mathrm{dl}(\cdot,\cdot)$ from (4.15). According to Algorithm 15, the contractive solver stops for the minimal $k = \underline{k}(\ell) \geq 1$ such that

$$\mathrm{dl}(u_\ell^{\underline{k}}, u_\ell^{\underline{k}-1}) \leq \lambda_{\mathrm{ctr}}\, \eta_\ell(u_\ell^{\underline{k}}). \tag{4.25}$$

In particular, since we exluded $\underline{k}$ from the index set $\mathcal{Q}$, this implies that

$$\eta_\ell(u_\ell^k) < \lambda_{\mathrm{ctr}}^{-1}\, \mathrm{dl}(u_\ell^k, u_\ell^{k-1}) \quad \text{for all } (\ell,k) \in \mathcal{Q} \text{ with } k > 0. \tag{4.26}$$

Proof of Theorem 17 under assumption (C1)

In this section, we give a proof of Theorem 17 under the assumption (C1), i.e., that the iterative solver Φ_ℓ leads to a uniform contraction of the discrete energy. Therefore, we first recall that the solution $u^\star \in \mathcal{H}$ minimizes the energy $\mathcal{E}$ in $\mathcal{H}$, i.e.,

$$\mathcal{E}(u^\star) = \min_{v \in \mathcal{H}} \mathcal{E}(v)$$

as well as that the discrete Galerkin solution $u_\bullet^\star \in \mathcal{X}_\bullet$ minimizes the energy $\mathcal{E}$ in $\mathcal{X}_\bullet$, i.e.,

$$\mathcal{E}(u_\bullet^\star) = \min_{v_\bullet \in \mathcal{X}_\bullet} \mathcal{E}(v_\bullet),$$

cf. Section 4.2. Hence, for $v_\bullet \in \mathcal{X}_\bullet$ the energy differences $\mathcal{E}(v_\bullet) - \mathcal{E}(u^\star)$, $\mathcal{E}(u_\bullet^\star) - \mathcal{E}(u^\star)$, and $\mathcal{E}(v_\bullet) - \mathcal{E}(u_\bullet^\star)$ are all non-negative. Therefrom, the absolute values in the definition of $\mathrm{dl}(\cdot,\cdot)$ can be omitted which yields the Pythagoras-type identity

$$\mathrm{dl}(u^\star, v_\bullet)^2 = \mathrm{dl}(u^\star, u_\bullet^\star)^2 + \mathrm{dl}(u_\bullet^\star, v_\bullet)^2 \quad \text{for all } v_\bullet \in \mathcal{X}_\bullet. \tag{4.27}$$

The core of the proof of Theorem 17 is the following lemma, where $0 < \theta \leq 1$ and $\lambda_{\mathrm{ctr}} > 0$ are, in fact, arbitrary parameters.

Lemma 19. *Suppose (A1)–(A3) and (C1). Let $0 < \theta \leq 1$ and $\lambda_{\mathrm{ctr}} > 0$. Then, there exist constants $\mu > 0$ and $0 < q_{\mathrm{lin}} < 1$ such that*

$$\Lambda_\ell^k := \mathrm{dl}(u^\star, u_\ell^k)^2 + \mu\, \eta_\ell(u_\ell^k)^2 \quad \textit{for all } (\ell,k) \in \mathcal{Q} \tag{4.28}$$

satisfies the following statements (i)–(ii)*:*

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

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

The constants μ and q_{lin} depend only on L, α, C_{stab}, q_{red}, C_{rel}, and q_{ctr} as well as on the adaptivity parameters $0 < \theta \leq 1$ and $\lambda_{\mathrm{ctr}} > 0$.

Proof of Lemma 19(i). Let $\mu, \varepsilon > 0$ be free parameters, which will be fixed below. First, we note that reliability (A3) and stability (A1) yield

$$\begin{aligned}\|\!|u^\star - u_\ell^\star|\!\|^2 &\overset{\text{(A3)}}{\leq} C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^\star)^2 \\ &\overset{\text{(A1)}}{\leq} 2\, C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^{k+1})^2 + 2\, C_{\mathrm{rel}}^2\, C_{\mathrm{stab}}^2\, \|\!|u_\ell^\star - u_\ell^{k+1}|\!\|^2.\end{aligned}$$

Together with the equivalence (4.9), this leads to

$$\begin{aligned}\mathrm{dl}(u^\star, u_\ell^\star)^2 &\overset{\text{(4.9)}}{\leq} \frac{L}{2}\, \|\!|u^\star - u_\ell^\star|\!\|^2 \\ &\leq L\, C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^{k+1})^2 + L\, C_{\mathrm{rel}}^2\, C_{\mathrm{stab}}^2\, \|\!|u_\ell^\star - u_\ell^{k+1}|\!\|^2 \\ &\overset{\text{(4.9)}}{\leq} L\, C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^{k+1})^2 + 2\, L\, \alpha^{-1}\, C_{\mathrm{rel}}^2\, C_{\mathrm{stab}}^2\, \mathrm{dl}(u_\ell^\star, u_\ell^{k+1})^2.\end{aligned}$$

Let $C_1 := L\, C_{\mathrm{rel}}^2$ and $C_2 := 2\, L\, \alpha^{-1}\, C_{\mathrm{rel}}^2\, C_{\mathrm{stab}}^2$. With this, combining the last inequality and the energy contraction (C1), we obtain that

$$\begin{aligned}\mathrm{dl}(u^\star, u_\ell^{k+1})^2 &\overset{\text{(4.27)}}{=} (1-\varepsilon)\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + \varepsilon\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + \mathrm{dl}(u_\ell^\star, u_\ell^{k+1})^2 \\ &\leq (1-\varepsilon)\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + \varepsilon\, C_1\, \eta_\ell(u_\ell^{k+1})^2 + (1+\varepsilon\, C_2)\, \mathrm{dl}(u_\ell^\star, u_\ell^{k+1})^2 \\ &\overset{\text{(C1)}}{\leq} (1-\varepsilon)\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + \varepsilon\, C_1\, \eta_\ell(u_\ell^{k+1})^2 + (1+\varepsilon\, C_2)\, q_{\mathrm{ctr}}^2\, \mathrm{dl}(u_\ell^\star, u_\ell^k)^2\end{aligned}$$

Since $(\ell, k+1) \in \mathcal{Q}$ and according to the definition of $\mathcal{Q}$, it holds that $k+1 < \underline{k}(\ell)$. Hence, inequality (4.26) and Lemma 14(ii) yield that

$$\begin{aligned}\eta_\ell(u_\ell^{k+1})^2 &\overset{\text{(4.26)}}{<} \lambda_{\mathrm{ctr}}^{-2}\, \mathrm{dl}(u_\ell^{k+1}, u_\ell^k)^2 \\ &\overset{\text{Lemma 14(ii)}}{\leq} \lambda_{\mathrm{ctr}}^{-2}\, (1+q_{\mathrm{ctr}})^2\, \mathrm{dl}(u_\ell^\star, u_\ell^k)^2.\end{aligned}$$

Let $C_3 := \lambda_{\mathrm{ctr}}^{-2}\, (1+q_{\mathrm{ctr}})^2$. Combining the latter two estimates, we see that

$$\begin{aligned}\Lambda_\ell^{k+1} &= \mathrm{dl}(u^\star, u_\ell^{k+1})^2 + \mu\, \eta_\ell(u_\ell^{k+1})^2 \\ &\leq (1-\varepsilon)\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + (\mu + \varepsilon\, C_1)\, \eta_\ell(u_\ell^{k+1})^2 + (1+\varepsilon\, C_2)\, q_{\mathrm{ctr}}^2\, \mathrm{dl}(u_\ell^\star, u_\ell^k)^2 \\ &\leq (1-\varepsilon)\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + \big\{(\mu + \varepsilon\, C_1)\, C_3 + (1+\varepsilon\, C_2)\, q_{\mathrm{ctr}}^2\big\}\, \mathrm{dl}(u_\ell^\star, u_\ell^k)^2\end{aligned}$$

Note that C_1, C_2, C_3 depend only on the problem setting. Provided that

$$(\mu + \varepsilon\, C_1)\, C_3 + (1+\varepsilon\, C_2)\, q_{\mathrm{ctr}}^2 \leq 1 - \varepsilon, \tag{4.29}$$

we are thus led to

$$\begin{aligned}\Lambda_\ell^{k+1} &\leq (1-\varepsilon)\, \big(\mathrm{dl}(u^\star, u_\ell^\star)^2 + \mathrm{dl}(u_\ell^\star, u_\ell^k)^2\big) \\ &\overset{\text{(4.27)}}{=} (1-\varepsilon)\, \mathrm{dl}(u^\star, u_\ell^k)^2 \\ &\leq (1-\varepsilon)\, \Lambda_\ell^k.\end{aligned}$$

Up to the final choice of $\mu, \varepsilon > 0$ (see below), this concludes the proof of Lemma 19(i). □

Proof of Lemma 19(ii). Let $\mu, \delta, \varepsilon > 0$ be free parameters, which will be fixed below. First, we note that

$$\begin{aligned}
|||u^\star - u_\ell^\star|||^2 &\overset{(A3)}{\leq} C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^\star)^2 \\
&\overset{(A1)}{\leq} 2\, C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^{k-1})^2 + 2\, C_{\mathrm{rel}}^2\, C_{\mathrm{stab}}^2\, |||u_\ell^\star - u_\ell^{k-1}|||^2.
\end{aligned}$$

Together with the equivalence (4.9), this leads to

$$\begin{aligned}
\mathrm{dl}(u^\star, u_\ell^\star)^2 &\overset{(4.9)}{\leq} \frac{L}{2}\, |||u^\star - u_\ell^\star|||^2 \\
&\leq L\, C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^{k-1})^2 + L\, C_{\mathrm{rel}}^2\, C_{\mathrm{stab}}^2\, |||u_\ell^\star - u_\ell^{k-1}|||^2 \\
&\overset{(4.9)}{\leq} L\, C_{\mathrm{rel}}^2\, \eta_\ell(u_\ell^{k-1})^2 + 2\, L\, \alpha^{-1}\, C_{\mathrm{rel}}^2\, C_{\mathrm{stab}}^2\, \mathrm{dl}(u_\ell^\star, u_\ell^{k-1})^2
\end{aligned}$$

Recall that $C_1 = L\, C_{\mathrm{rel}}^2$ and $C_2 = 2L\alpha^{-1}\, C_{\mathrm{rel}}^2\, C_{\mathrm{stab}}^2$. With this, we obtain that

$$\begin{aligned}
\mathrm{dl}(u^\star, u_\ell^k)^2 &\overset{(4.27)}{=} (1-\varepsilon)\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + \varepsilon\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + \mathrm{dl}(u_\ell^\star, u_\ell^k)^2 \\
&\leq (1-\varepsilon)\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + \varepsilon\, C_1\, \eta_\ell(u_\ell^{k-1})^2 + \varepsilon\, C_2\, \mathrm{dl}(u_\ell^\star, u_\ell^{k-1})^2 + \mathrm{dl}(u_\ell^\star, u_\ell^k)^2 \qquad (4.30) \\
&\overset{(C1)}{\leq} (1-\varepsilon)\, \mathrm{dl}(u^\star, u_\ell^\star)^2 + \varepsilon\, C_1\, \eta_\ell(u_\ell^{k-1})^2 + (\varepsilon\, C_2 + q_{\mathrm{ctr}}^2)\, \mathrm{dl}(u_\ell^\star, u_\ell^{k-1})^2.
\end{aligned}$$

Next, stability (A1) and reduction (A2) show that

$$\begin{aligned}
\eta_{\ell+1}(u_\ell^k)^2 &= \eta_{\ell+1}(\mathcal{T}_\ell \cap \mathcal{T}_{\ell+1}, u_\ell^k)^2 + \eta_{\ell+1}(\mathcal{T}_{\ell+1} \backslash \mathcal{T}_\ell, u_\ell^k)^2 \\
&\overset{(A1)}{=} \eta_\ell(\mathcal{T}_\ell \cap \mathcal{T}_{\ell+1}, u_\ell^k)^2 + \eta_{\ell+1}(\mathcal{T}_{\ell+1} \backslash \mathcal{T}_\ell, u_\ell^k)^2 \\
&\overset{(A2)}{\leq} \eta_\ell(\mathcal{T}_\ell \cap \mathcal{T}_{\ell+1}, u_\ell^k)^2 + q_{\mathrm{red}}^2\, \eta_\ell(\mathcal{T}_\ell \backslash \mathcal{T}_{\ell+1}, u_\ell^k)^2 \\
&= \eta_\ell(u_\ell^k)^2 - (1 - q_{\mathrm{red}}^2)\, \eta_\ell(\mathcal{T}_\ell \backslash \mathcal{T}_{\ell+1}, u_\ell^k)^2.
\end{aligned}$$

According to the Dörfler marking criterion (4.17) in Algorithm 15(iii), we are led to

$$\eta_{\ell+1}(u_\ell^k)^2 \leq \big(1 - (1 - q_{\mathrm{red}}^2)\, \theta^2\big)\, \eta_\ell(u_\ell^k)^2 =: q_\theta\, \eta_\ell(u_\ell^k)^2. \qquad (4.31)$$

Note that

$$\begin{aligned}
|||u_\ell^k - u_\ell^{k-1}|||^2 &\leq 2\, \big(|||u_\ell^\star - u_\ell^k|||^2 + |||u_\ell^\star - u_\ell^{k-1}|||^2\big) \\
&\overset{(4.9)}{\leq} \frac{4}{\alpha}\, \big(\mathrm{dl}(u_\ell^\star, u_\ell^k)^2 + \mathrm{dl}(u_\ell^\star, u_\ell^{k-1})^2\big) \\
&\overset{(C1)}{\leq} \frac{4}{\alpha}\, (q_{\mathrm{ctr}}^2 + 1)\, \mathrm{dl}(u_\ell^\star, u_\ell^{k-1})^2.
\end{aligned}$$

Next, with $\delta > 0$ which we specify further on, we use the following variant of Young's inequality

$$(a+b)^2 \leq (1+\delta)\, a^2 + (1+\delta^{-1})\, b^2 \quad \text{for all } a, b \in \mathbb{R}.$$

This leads to

$$\begin{aligned}\eta_\ell(u_\ell^k)^2 &\overset{(A1)}{\leq} \big(\eta_\ell(u_\ell^{k-1}) + C_{\mathrm{stab}}\,|||u_\ell^k - u_\ell^{k-1}|||\big)^2\\ &\leq (1+\delta)\,\eta_\ell(u_\ell^{k-1})^2 + (1+\delta^{-1})\,C_{\mathrm{stab}}^2\,|||u_\ell^k - u_\ell^{k-1}|||^2\\ &\leq (1+\delta)\,\eta_\ell(u_\ell^{k-1})^2 + (1+\delta^{-1})\,\frac{4}{\alpha}\,(q_{\mathrm{ctr}}^2+1)\,C_{\mathrm{stab}}^2\,\mathrm{dl}(u_\ell^\star, u_\ell^{k-1})^2.\end{aligned} \tag{4.32}$$

Let $C_4 := 4\,\alpha^{-1}\,(q_{\mathrm{ctr}}^2+1)\,C_{\mathrm{stab}}^2$. Note that Algorithm 15 guarantees that $u_{\ell+1}^0 = u_\ell^k$. Combining the latter estimates, we see that

$$\begin{aligned}\Lambda_{\ell+1}^0 &= \mathrm{dl}(u^\star, u_{\ell+1}^0)^2 + \mu\,\eta_{\ell+1}(u_{\ell+1}^0)^2\\ &\overset{(4.31)}{\leq} \mathrm{dl}(u^\star, u_\ell^k)^2 + \mu\,q_\theta\,\eta_\ell(u_\ell^k)^2\\ &\overset{(4.30)}{\leq} (1-\varepsilon)\,\mathrm{dl}(u^\star, u_\ell^\star)^2 + \varepsilon\,C_1\,\eta_\ell(u_\ell^{k-1})^2 + (\varepsilon\,C_2 + q_{\mathrm{ctr}}^2)\,\mathrm{dl}(u_\ell^\star, u_\ell^{k-1})^2 + \mu\,q_\theta\,\eta_\ell(u_\ell^k)^2\\ &\overset{(4.32)}{\leq} (1-\varepsilon)\,\mathrm{dl}(u^\star, u_\ell^\star)^2 + \big\{\varepsilon\,C_1\,\mu^{-1} + q_\theta\,(1+\delta)\big\}\,\mu\,\eta_\ell(u_\ell^{k-1})^2\\ &\qquad + \big\{\varepsilon\,C_2 + q_{\mathrm{ctr}}^2 + \mu\,q_\theta\,(1+\delta^{-1})\,C_4\big\}\,\mathrm{dl}(u_\ell^\star, u_\ell^{k-1})^2.\end{aligned}$$

Note that C_1, C_2, C_4 and $0 < q_\theta < 1$ depend only on the problem setting. Provided that

$$\varepsilon\,C_1\,\mu^{-1} + q_\theta\,(1+\delta) \leq 1-\varepsilon \quad\text{and}\quad \varepsilon\,C_2 + q_{\mathrm{ctr}}^2 + \mu\,q_\theta\,(1+\delta^{-1})\,C_4 \leq 1-\varepsilon, \tag{4.33}$$

we are thus led to

$$\begin{aligned}\Lambda_{\ell+1}^0 &\leq (1-\varepsilon)\,\big(\mathrm{dl}(u^\star, u_\ell^\star)^2 + \mathrm{dl}(u_\ell^\star, u_\ell^{k-1})^2 + \mu\,\eta_\ell(u_\ell^{k-1})^2\big)\\ &\overset{(4.27)}{=} (1-\varepsilon)\,\big(\mathrm{dl}(u^\star, u_\ell^{k-1}) + \mu\,\eta_\ell(u_\ell^{k-1})^2\big)\\ &= (1-\varepsilon)\,\Lambda_\ell^{k-1}.\end{aligned}$$

Up to the final choice of $\delta, \mu, \varepsilon > 0$, this concludes the proof of Lemma 19(ii). □

Proof of Lemma 19 (fixing the free parameters). To fix all the free parameters and to show that there exists a choice such that all the necessary assumptions are fulfilled, we proceed as follows:

- Choose $\delta > 0$ such that $(1+\delta)\,q_\theta < 1$.
- Choose $\mu > 0$ such that $q_{\mathrm{ctr}}^2 + \mu\,q_\theta(1+\delta)^{-1}\,C_4 < 1$ and $\mu\,C_3 + q_{\mathrm{ctr}}^2 < 1$.
- Finally, choose $\varepsilon > 0$ sufficiently small such that (4.29) and (4.33) are satisfied.

This concludes the proof of Lemma 19 with $(1-\varepsilon) = q_{\mathrm{lin}}^2$. □

Proof of Theorem 17 under assumption (C1). According to (4.9), it holds that $\mathrm{dl}(u^\star, u_\ell^k) \eqsim |||u^\star - u_\ell^k|||$ and as a consequence that $\Delta_\ell^k \eqsim (\Lambda_\ell^k)^{1/2}$, where the hidden constants depend only on μ, α, and L.

Since the index set $\mathcal{Q}$ is linearly ordered with respect to the total step counter $|(\cdot,\cdot)|$, linear convergence (4.24) now follows directly from Lemma 19 via induction on the index pair. □

Proof of Theorem 17 under assumption (C2)

In order to prove Theorem 17 under assumption (C2), we first have to recall the following main result from [GHPS18] whose proof is based on a perturbation argument.

Lemma 20 ([GHPS18, Lemma 4.9, Theorem 5.3]). *Suppose (A1)–(A3) and (C2). Let $0 < \theta \le 1$ and $0 < \lambda_{\mathrm{ctr}} < \lambda_{\mathrm{conv}}\,\theta$, where $\lambda_{\mathrm{conv}} = \frac{1-q_{\mathrm{ctr}}}{C_{\mathrm{stab}} q_{\mathrm{ctr}}}$. Then, it holds that*

$$\vertiii{u_\ell^\star - u_\ell^k} \le \lambda_{\mathrm{ctr}} \frac{q_{\mathrm{ctr}}}{1-q_{\mathrm{ctr}}} \min\Big\{\eta_\ell(u_\ell^k), \frac{1}{1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}}}\eta_\ell(u_\ell^\star)\Big\} \tag{4.34}$$

as well as

$$(1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}})\,\eta_\ell(u_\ell^k) \le \eta_\ell(u_\ell^\star) \le (1+\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}})\,\eta_\ell(u_\ell^k). \tag{4.35}$$

Moreover, there exist $C_{\mathrm{GHPS}} > 0$ and $0 < q_{\mathrm{GHPS}} < 1$ such that

$$\eta_{\ell+n}(u_{\ell+n}^k) \le C_{\mathrm{GHPS}}\, q_{\mathrm{GHPS}}^n\, \eta_\ell(u_\ell^k) \quad \text{for all } (\ell+n+1,0) \in \mathcal{Q}. \tag{4.36}$$

The constants C_{GHPS} and q_{GHPS} depend only on $C_{\mathsf{Céa}} = L/\alpha$, C_{rel}, C_{stab}, q_{red}, and q_{ctr}, as well as on the adaptivity parameters θ and λ_{ctr}. □

Lemma 20 shows that the given constraint on λ_{ctr} guarantees estimator equivalence $\eta_\ell(u_\ell^\star) \simeq \eta_\ell(u_\ell^k)$. Assume Dörfler marking for $\eta_\ell(u_\ell^k)$ and θ, cf. Algorithm 15(iii), then there holds with stability (A1) that

$$\begin{aligned}
\frac{\theta-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}}}{1+\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}}}\,\eta_\ell(u_\ell^\star) &\overset{(4.35)}{\le} (\theta-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}})\,\eta_\ell(u_\ell^k)\\
&\overset{(4.17)}{\le} \eta_\ell(\mathcal{M}_\ell, u_\ell^k) - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}}\,\eta_\ell(u_\ell^k)\\
&\overset{(A1)}{\le} \eta_\ell(\mathcal{M}_\ell, u_\ell^\star) + C_{\mathrm{stab}}\,\vertiii{u_\ell^\star - u_\ell^k} - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}}\,\eta_\ell(u_\ell^k)\\
&\overset{(4.34)}{\le} \eta_\ell(\mathcal{M}_\ell, u_\ell^\star).
\end{aligned} \tag{4.37}$$

In other words, Dörfler marking for $\eta_\ell(u_\ell^k)$ and θ implies Dörfler marking for $\eta_\ell(u_\ell^\star)$ and $\theta^\star := (\theta-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}})/(1+\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{conv}}) > 0$.

In the present case, the core of the proof of Theorem 17 is the following summability result.

Lemma 21. *Suppose (A1)–(A3) and (C2). Let $0 < \theta \le 1$ and $0 < \lambda_{\mathrm{ctr}} < \lambda_{\mathrm{conv}}\,\theta$, where again $\lambda_{\mathrm{conv}} = \frac{1-q_{\mathrm{ctr}}}{C_{\mathrm{stab}} q_{\mathrm{ctr}}}$. Then, there exists $C_{\mathrm{sum}} > 0$ such that*

$$\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)>(\ell',k')}} \Delta_\ell^k \le C_{\mathrm{sum}}\,\Delta_{\ell'}^{k'} \quad \text{for all } (\ell',k') \in \mathcal{Q}. \tag{4.38}$$

The constant $C_{\mathrm{sum}} > 0$ depends only on L, α, C_{rel}, C_{stab}, q_{red}, and q_{ctr}, as well as on the adaptivity parameters θ and λ_{ctr}.

Proof. The proof is split into six steps.

Step 1. This step provides an equivalent quasi-error quantity. First, note that

$$\begin{aligned}\|\!|u^\star - u_\ell^k|\!\| &\le \|\!|u^\star - u_\ell^\star|\!\| + \|\!|u_\ell^\star - u_\ell^k|\!\| \\ &\overset{(A3)}{\lesssim} \eta_\ell(u_\ell^\star) + \|\!|u_\ell^\star - u_\ell^k|\!\| \\ &\overset{(A1)}{\lesssim} \eta_\ell(u_\ell^k) + \|\!|u_\ell^\star - u_\ell^k|\!\| =: \mathrm{A}_\ell^k.\end{aligned}$$

This proves that $\Delta_\ell^k = \|\!|u^\star - u_\ell^k|\!\| + \eta_\ell(u_\ell^k) \lesssim \mathrm{A}_\ell^k$. Second, the Céa lemma (4.12) proves that

$$\|\!|u_\ell^\star - u_\ell^k|\!\| \le \|\!|u^\star - u_\ell^\star|\!\| + \|\!|u^\star - u_\ell^k|\!\| \overset{(4.12)}{\lesssim} \|\!|u^\star - u_\ell^k|\!\|.$$

This concludes that

$$\boxed{\mathrm{A}_\ell^k = \|\!|u_\ell^\star - u_\ell^k|\!\| + \eta_\ell(u_\ell^k) \simeq \Delta_\ell^k.} \tag{4.39}$$

Step 2. This step collects some auxiliary estimates. We start with

$$\boxed{\mathrm{A}_\ell^0 \lesssim \eta_{\ell-1}(u_{\ell-1}^k) \le \mathrm{A}_{\ell-1}^k \quad \text{for all } (\ell,0) \in \mathcal{Q} \text{ with } \ell > 0.} \tag{4.40}$$

With the Céa lemma (4.12) and reliability (4.20), it follows that

$$\begin{aligned}\|\!|u_\ell^\star - u_{\ell-1}^k|\!\| &\le \|\!|u^\star - u_\ell^\star|\!\| + \|\!|u^\star - u_{\ell-1}^k|\!\| \\ &\overset{(4.12)}{\lesssim} \|\!|u^\star - u_{\ell-1}^k|\!\| \\ &\overset{(4.20)}{\lesssim} \eta_{\ell-1}(u_{\ell-1}^k)\end{aligned}$$

With nested iteration $u_\ell^0 = u_{\ell-1}^k$ and (A1)–(A2), we thus obtain that

$$\begin{aligned}\mathrm{A}_\ell^0 &= \|\!|u_\ell^\star - u_\ell^0|\!\| + \eta_\ell(u_\ell^0) \\ &= \|\!|u_\ell^\star - u_{\ell-1}^k|\!\| + \eta_\ell(u_{\ell-1}^k) \\ &\lesssim \eta_{\ell-1}(u_{\ell-1}^k) \\ &\le \mathrm{A}_{\ell-1}^k\end{aligned}$$

This proves (4.40). Next, we prove that

$$\boxed{\mathrm{A}_\ell^k \lesssim \mathrm{A}_\ell^k \quad \text{for all } (\ell+1,0) \in \mathcal{Q} \text{ and } 0 \le k \le \underline{k}(\ell).} \tag{4.41}$$

To see this, note that

$$\|\!|u_\ell^k - u_\ell^k|\!\| \le \|\!|u_\ell^\star - u_\ell^k|\!\| + \|\!|u_\ell^\star - u_\ell^k|\!\| \overset{(C2)}{\le} (q_{\mathrm{ctr}}^{k-k} + 1)\, \|\!|u_\ell^\star - u_\ell^k|\!\|.$$

Hence, it follows that

$$\begin{aligned}\mathrm{A}_\ell^{\underline{k}} &= |||u_\ell^\star - u_\ell^{\underline{k}}||| + \eta_\ell(u_\ell^{\underline{k}}) \\ &\overset{\text{(A1)}}{\lesssim} |||u_\ell^\star - u_\ell^k||| + |||u_\ell^{\underline{k}} - u_\ell^k||| + \eta_\ell(u_\ell^k) \\ &\lesssim |||u_\ell^\star - u_\ell^k||| + \eta_\ell(u_\ell^k) \\ &= \mathrm{A}_\ell^k.\end{aligned}$$

This proves (4.41). Finally, we prove that

$$\boxed{\mathrm{A}_\ell^k \lesssim |||u_\ell^\star - u_\ell^{k-1}||| \quad \text{for all } (\ell,k) \in \mathcal{Q} \text{ with } k > 0.} \tag{4.42}$$

With the inequality (4.26), which stems from the stopping criterion (4.16) of Algorithm 15(i), and Lemma 14(ii), we get that

$$\eta_\ell(u_\ell^k) \overset{(4.26)}{\lesssim} |||u_\ell^k - u_\ell^{k-1}||| \overset{\text{Lemma 14(ii)}}{\lesssim} |||u_\ell^\star - u_\ell^{k-1}|||.$$

This leads to

$$\begin{aligned}\mathrm{A}_\ell^k &= |||u_\ell^\star - u_\ell^k||| + \eta_\ell(u_\ell^k) \\ &\overset{\text{(C2)}}{\lesssim} |||u_\ell^\star - u_\ell^{k-1}||| + \eta_\ell(u_\ell^k) \\ &\lesssim |||u_\ell^\star - u_\ell^{k-1}|||\end{aligned}$$

and thus proves (4.42).

Step 3. Suppose that $\underline{\ell} = \infty$ and hence $\underline{k}(\ell) < \infty$ for all $\ell \in \mathbb{N}_0$. Note that

$$\begin{aligned}\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)>(\ell',k')}} \mathrm{A}_\ell^k &= \sum_{\ell=\ell'+1}^{\infty} \sum_{k=0}^{\underline{k}(\ell)-1} \mathrm{A}_\ell^k + \sum_{k=k'+1}^{\underline{k}(\ell')-1} \mathrm{A}_{\ell'}^k \\ &\overset{(4.40)}{\lesssim} \sum_{\ell=\ell'+1}^{\infty} \sum_{k=1}^{\underline{k}(\ell)} \mathrm{A}_\ell^k + \sum_{k=k'+1}^{\underline{k}(\ell')} \mathrm{A}_{\ell'}^k.\end{aligned}$$

With contraction (C2), the geometric series proves for all $(\ell,i) \in \mathcal{Q}$ that

$$\begin{aligned}\sum_{k=i+1}^{\underline{k}(\ell)-1} \mathrm{A}_\ell^k &\overset{(4.42)}{\lesssim} \sum_{k=i+1}^{\underline{k}(\ell)-1} |||u_\ell^\star - u_\ell^{k-1}||| \\ &\overset{\text{(C2)}}{\lesssim} |||u_\ell^\star - u_\ell^i||| \sum_{k=i}^{\infty} q_{\mathrm{ctr}}^{k-i} \\ &\lesssim \mathrm{A}_\ell^i.\end{aligned} \tag{4.43}$$

Hence, it follows that

$$\sum_{k=1}^{\underline{k}(\ell)} A_\ell^k \begin{cases} = \mathrm{A}_\ell^1 \overset{(4.41)}{\lesssim} A_\ell^0 & \text{if } \underline{k}(\ell) = 1, \\ \overset{(4.41)}{\lesssim} \sum_{k=1}^{\underline{k}(\ell)-1} \mathrm{A}_\ell^k \overset{(4.43)}{\lesssim} A_\ell^0 & \text{if } \underline{k}(\ell) > 1. \end{cases}$$

Moreover, it follows that

$$\sum_{k=k'+1}^{\underline{k}(\ell')} A_{\ell'}^k \begin{cases} = A_{\ell'}^{k'+1} \overset{(4.41)}{\lesssim} A_{\ell'}^{k'} & \text{if } \underline{k}(\ell') = k'+1, \\ \overset{(4.41)}{\lesssim} \sum_{k=k'+1}^{\underline{k}(\ell')-1} \mathrm{A}_{\ell'}^k \overset{(4.43)}{\lesssim} A_{\ell'}^{k'} & \text{if } \underline{k}(\ell') > k'+1. \end{cases}$$

So far, this proves that

$$\sum_{\substack{(\ell,k)\in\mathcal{Q} \\ (\ell,k)>(\ell',k')}} \mathrm{A}_\ell^k \lesssim \mathrm{A}_{\ell'}^{k'} + \sum_{\ell=\ell'+1}^{\infty} \mathrm{A}_\ell^0.$$

Exploiting the linear convergence (4.36) together with the geometric series, we prove that

$$\begin{aligned} \sum_{\ell=\ell'+1}^{\infty} \mathrm{A}_\ell^0 &\overset{(4.40)}{\lesssim} \sum_{\ell=\ell'+1}^{\infty} \eta_{\ell-1}(u_{\ell-1}^{\underline{k}}) \\ &= \sum_{\ell=\ell'}^{\infty} \eta_\ell(u_\ell^{\underline{k}}) \\ &\overset{(4.36)}{\lesssim} \eta_{\ell'}(u_{\ell'}^{\underline{k}}) \sum_{\ell=\ell'}^{\infty} q_{\mathrm{GHPS}}^{\ell-\ell'} \\ &\simeq \eta_{\ell'}(u_{\ell'}^{\underline{k}}) \\ &\le \mathrm{A}_{\ell'}^{\underline{k}}. \end{aligned}$$

Overall, this proves that

$$\sum_{\substack{(\ell,k)\in\mathcal{Q} \\ (\ell,k)>(\ell',k')}} \mathrm{A}_\ell^k \lesssim \mathrm{A}_{\ell'}^{k'} + \mathrm{A}_{\ell'}^{\underline{k}} \overset{(4.41)}{\simeq} \mathrm{A}_{\ell'}^{k'} \quad \text{provided that } \underline{\ell} = \infty. \tag{4.44}$$

Step 4. Suppose that $\ell' = \underline{\ell} < \infty$ and hence $\underline{k}(\ell') = \underline{k}(\underline{\ell}) = \infty$. Then, the geometric series proves that

$$\sum_{\substack{(\ell,k)\in\mathcal{Q} \\ (\ell,k)>(\ell',k')}} \mathrm{A}_\ell^k = \sum_{k=k'+1}^{\infty} \mathrm{A}_{\ell'}^k \overset{(4.43)}{\lesssim} \mathrm{A}_{\ell'}^{k'}. \tag{4.45}$$

Step 5. Suppose that $\ell' < \underline{\ell} < \infty$ and hence $\underline{k}(\underline{\ell}) = \infty$. Then, it holds that

$$\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)>(\ell',k')}} \mathrm{A}_\ell^k = \sum_{\ell=\ell'+1}^{\underline{\ell}-1} \sum_{k=0}^{\underline{k}(\ell)-1} \mathrm{A}_\ell^k + \sum_{k=k'+1}^{\underline{k}(\ell')-1} \mathrm{A}_{\ell'}^k + \sum_{k=0}^{\infty} \mathrm{A}_{\underline{\ell}}^k.$$

First, note that

$$\sum_{k=0}^{\infty} \mathrm{A}_{\underline{\ell}}^k = \mathrm{A}_{\underline{\ell}}^0 + \sum_{k=1}^{\infty} \mathrm{A}_{\underline{\ell}}^k \overset{(4.43)}{\leq} \mathrm{A}_{\underline{\ell}}^0 \overset{(4.40)}{\lesssim} \mathrm{A}_{\underline{\ell}-1}^{\underline{k}}.$$

Provided that $\ell' < \underline{\ell} < \infty$, it hence holds that

$$\begin{aligned}\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)>(\ell',k')}} \mathrm{A}_\ell^k &\lesssim \sum_{\ell=\ell'+1}^{\underline{\ell}-1} \sum_{k=0}^{\underline{k}(\ell)} \mathrm{A}_\ell^k + \sum_{k=k'+1}^{\underline{k}(\ell')-1} \mathrm{A}_{\ell'}^k \\ &\overset{(4.40)}{\lesssim} \sum_{\ell=\ell'+1}^{\underline{\ell}-1} \sum_{k=1}^{\underline{k}(\ell)} \mathrm{A}_\ell^k + \sum_{k=k'+1}^{\underline{k}(\ell')} \mathrm{A}_{\ell'}^k.\end{aligned}$$

Along the lines of Step 3, one concludes that

$$\sum_{\ell=\ell'+1}^{\underline{\ell}-1} \sum_{k=1}^{\underline{k}(\ell)} \mathrm{A}_\ell^k + \sum_{k=k'+1}^{\underline{k}(\ell')} \mathrm{A}_{\ell'}^k \lesssim \mathrm{A}_{\ell'}^{k'}. \tag{4.46}$$

Step 6. In any case, (4.44)–(4.46) prove for all $(\ell', k') \in \mathcal{Q}$ that

$$\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)>(\ell',k')}} \Delta_\ell^k \simeq \sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)>(\ell',k')}} \mathrm{A}_\ell^k \lesssim \mathrm{A}_{\ell'}^{k'} \simeq \Delta_{\ell'}^{k'}.$$

This concludes the proof of (4.38). □

Proof of Theorem 17 under the assumption (C2). The proof is split into two steps.

Step 1. From [CFPP14, Lemma 4.9], we recall the following implication for sequences $(\alpha_n)_{n\in\mathbb{N}_0}$ in $\mathbb{R}_{\geq 0}$ and constants $C > 0$: Assume that

$$\sum_{n=N+1}^{\infty} \alpha_n \leq C\,\alpha_N \quad \text{for all } N \in \mathbb{N}_0.$$

Then, for $N \in \mathbb{N}_0$, it holds that

$$(1 + C^{-1}) \sum_{n=N+1}^{\infty} \alpha_n \leq \sum_{n=N+1}^{\infty} \alpha_n + \alpha_N = \sum_{n=N}^{\infty} \alpha_n.$$

Inductively, it follows that

$$(1+C^{-1})^m \sum_{n=N+m}^{\infty} \alpha_n \le \sum_{n=N+1}^{\infty} \alpha_n + \alpha_N = \sum_{n=N}^{\infty} \alpha_n \quad \text{for all } N, m \in \mathbb{N}_0.$$

We thus conclude that

$$\alpha_{N+m} \le (1+C^{-1})^{-m} \sum_{n=N}^{\infty} \alpha_n \le (1+C)\,(1+C^{-1})^{-m} \alpha_N \quad \text{for all } N, m \in \mathbb{N}_0.$$

Step 2. Since the index set $\mathcal{Q}$ is linearly ordered with respect to the total step counter $|(\cdot,\cdot)|$, Lemma 21 and Step 1 imply that

$$\Delta_{\ell'}^{k'} \le C_{\mathrm{lin}}\, q_{\mathrm{lin}}^{|(\ell',k')|-|(\ell,k)|}\, \Delta_{\ell}^{k} \quad \text{for all } (\ell,k),(\ell',k') \in \mathcal{Q} \text{ with } (\ell',k') > (\ell,k),$$

where $C_{\mathrm{lin}} = 1 + C_{\mathrm{sum}}$ and $q_{\mathrm{lin}} = 1/(1+C_{\mathrm{sum}}^{-1})$. This concludes the proof.

□

4.6.3 Optimal convergence rates of the quasi-error

The second main theorem states optimal convergence rates of the quasi-error (4.19) with respect to the overall computational costs. As usual in this context (see, e.g., [CFPP14]), the result requires that the adaptivity parameters $0 < \theta \le 1$ and $\lambda_{\mathrm{ctr}} > 0$ are sufficiently small. With the following definition, we then get Theorem 23.

Definition 22. *For $N \in \mathbb{N}_0$, let $\mathbb{T}(N)$ be the set of all refinements $\mathcal{T}$ of $\mathcal{T}_0$ with*

$$\#\mathcal{T} - \#\mathcal{T}_0 \le N.$$

Then, for given $s > 0$, define

$$\|u^\star\|_{\mathbb{A}_s} := \sup_{N\in\mathbb{N}_0} (N+1)^s \inf_{\mathcal{T}_{\mathrm{opt}}\in\mathbb{T}(N)} \big(|||u^\star - u_{\mathrm{opt}}^\star||| + \eta_{\mathrm{opt}}(u_{\mathrm{opt}}^\star)\big) \in \mathbb{R}_{\ge 0} \cup \{\infty\}. \tag{4.47}$$

Theorem 23. *Suppose (C1) or (C2) as well as (R1)–(R3) and (A1)–(A4). Define*

$$\lambda_{\mathrm{opt}} := \begin{cases} \frac{1-q_{\mathrm{ctr}}}{q_{\mathrm{ctr}} C_{\mathrm{stab}}} & \text{if (C2) is valid,} \\ \frac{1-q_{\mathrm{ctr}}}{q_{\mathrm{ctr}} C_{\mathrm{stab}}} \sqrt{\alpha/2} & \text{otherwise.} \end{cases} \tag{4.48}$$

Let $0 < \theta \le 1$ and $0 < \lambda_{\mathrm{ctr}} < \lambda_{\mathrm{opt}}\,\theta$ such that

$$0 < \theta' := \frac{\theta + \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}}{1 - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}} < (1 + C_{\mathrm{stab}}^2\, C_{\mathrm{rel}}^2)^{-1/2}. \tag{4.49}$$

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

$$c_{\rm opt}^{-1} \, \|u^\star\|_{\mathbb{A}_s} \le \sup_{(\ell',k')\in\mathcal{Q}} (\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1)^s \, \Delta_{\ell'}^{k'}$$
$$\le \sup_{(\ell',k')\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)\le(\ell',k')}} \#\mathcal{T}_\ell \Bigg)^s \Delta_{\ell'}^{k'} \le C_{\rm opt} \, \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\}, \tag{4.50}$$

where $\|u^\star\|_{\mathbb{A}_s}$ is defined in (4.47). The constant $c_{\rm opt} > 0$ depends only on $C_{\rm Céa} = L/\alpha$, $C_{\rm son}$, $C_{\rm stab}$, $C_{\rm rel}$, $\#\mathcal{T}_0$, and s, and, if $\underline{\ell} < \infty$ or $\eta_{\ell_0}(u_{\ell_0}^k) = 0$ for some $(\ell_0 + 1, 0) \in \mathcal{Q}$, additionally on $\underline{\ell}$ or ℓ_0 respectively. The constant $C_{\rm opt} > 0$ depends only on $C_{\rm stab}$, $q_{\rm red}$, $C_{\rm rel}$, $C_{\rm mesh}$, $1 - \lambda_{\rm ctr}/\lambda_{\rm opt}$, $C_{\rm mark}$, $C'_{\rm rel}$, $C_{\rm lin}$, $q_{\rm lin}$, $\#\mathcal{T}_0$, and s.

Remark 24. *The following comments underline the importance of the latter result:*

- *By definition (4.47), it holds that $\|u^\star\|_{\mathbb{A}_s} < \infty$ if and only if the quasi-error (for the exact discrete solutions) converges at least with algebraic rate $s > 0$ along a sequence of optimal meshes.*
- *If all steps of Algorithm 15 can be performed at linear costs $\mathcal{O}(\#\mathcal{T}_\ell)$, then the sum*

$$\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)\le(\ell',k')}} \#\mathcal{T}_\ell$$

  is proportional to the overall computational work (resp. the overall computational time spent) to perform the $|(\ell', k')|$-th step of the adaptive loop, since each adaptive step depends on the full adaptive history. Note that the computation of, e.g., all residual error indicators in Step (c) of Algorithm 15 as well as as the local mesh-refinement by, e.g., newest vertex bisection can be done at linear costs. The same applies to, e.g., one step of PCG with an optimal additive Schwarz preconditioner in Step (b) of Algorithm 15. For the Dörfler marking (4.17) in Step (iii) of Algorithm 15, we refer to [Ste07] for an algorithm with linear cost and $C_{\rm mark} = 2$ as well as to the recent algorithm from [PP20] with linear cost and even $C_{\rm mark} = 1$.
- *The interpretation of (4.50) thus is that the quasi-error for the computed discrete solutions u_ℓ^k decays with rate s with respect to the overall computational costs (as well as the degrees of freedom) if and only if rate s is possible with respect to the degrees of freedom (for the exact discrete solutions on optimal meshes).*
- *Since $s > 0$ is arbitrary, the proposed algorithm will asymptotically regain the best possible convergence behavior, even with respect to the computational costs.*
- *Prior works (see, e.g., [Ste07, BMS10, CG12, GHPS18]) proved linear convergence of the quasi-error only for those steps, where mesh-refinement takes place. Unlike this, we prove linear convergence (4.24) for the full sequence of discrete approximations, i.e., independently of the algorithmic decision for mesh-refinement or one step of the discrete solver.*

- *In usual applications, the* quasi-error Δ_ℓ^k *(i.e., error plus estimator) is equivalent to the so-called* total error *(i.e., error plus data oscillations) as well as to the estimator alone. Therefore, the approximability* $\|u^\star\|_{\mathbb{A}_s}$ *in* (4.47) *can equivalently be defined through the total error (see, e.g., [Ste07, CKNS08, CN12, FFP14]) or the estimator (see, e.g., [CFPP14]) instead of the quasi-error (used in* (4.47)*). The overall result will be the same.*

4.6.4 Proof of Theorem 23 (optimal convergence rates)

Recall $\|u^\star\|_{\mathbb{A}_s}$ from (4.47) and the set $\mathbb{T}(N) = \big\{\mathcal{T} \in \texttt{refine}(\mathcal{T}_0) : \#\mathcal{T} - \#\mathcal{T}_0 \le N\big\}$. Then, the following lemma proves the first inequality in (4.50).

Lemma 25. *Suppose* (R1) *as well as* (A1)–(A3)*. Let* $s > 0$*. Then, it holds that*

$$\|u^\star\|_{\mathbb{A}_s} \le c_{\mathrm{opt}} \sup_{(\ell,k)\in\mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \Delta_\ell^k, \tag{4.51}$$

where $c_{\mathrm{opt}} > 0$ *depends only on* $C_{\mathrm{Céa}} = L/\alpha$, C_{son}, C_{stab}, C_{rel}, $\#\mathcal{T}_0$, *and* s, *and, if* $\underline{\ell} < \infty$ *or* $\eta_{\ell_0}(u_{\ell_0}^k) = 0$ *for some* $(\ell_0+1, 0) \in \mathcal{Q}$*, additionally on* $\underline{\ell}$ *or* ℓ_0 *respectively.*

Proof. The proof is split into three steps. First, we recall Lemma 22 from [BHP17]: Let $\mathcal{T}_\bullet \in \mathbb{T}$ and $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$. Then, it holds that

$$\#\mathcal{T}_\circ / \#\mathcal{T}_\bullet \le \#\mathcal{T}_\circ - \#\mathcal{T}_\bullet + 1 \le \#\mathcal{T}_\circ. \tag{4.52}$$

Step 1. In this step, we consider the pathological cases that $\underline{\ell} < \infty$ or $\eta_{\ell_0}(u_{\ell_0}^k) = 0$ for some $(\ell_0+1, 0) \in \mathcal{Q}$. In the first case, Corollary 18 gives that $u^\star = u_{\underline{\ell}}^\star$ as well as $\eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) = 0$. From Proposition 16 and Lemma 11, we know that the latter implies $u_{\ell_0}^k = u^\star = u_{\ell_0}^\star$. Hence, with $\ell' := \underline{\ell}$ or $\ell' := \ell_0$ respectively, we obtain that

$$\begin{aligned}\|u^\star\|_{\mathbb{A}_s} &= \sup_{N\in\mathbb{N}_0} (N+1)^s \inf_{\mathcal{T}_{\mathrm{opt}}\in\mathbb{T}(N)} \big(|||u^\star - u_{\mathrm{opt}}^\star||| + \eta_{\mathrm{opt}}(u_{\mathrm{opt}}^\star)\big) \\ &= \max_{0\le N<\#\mathcal{T}_{\ell'}-\#\mathcal{T}_0} (N+1)^s \min_{\mathcal{T}_\bullet\in\mathbb{T}(N)} \big(|||u^\star - u_\bullet^\star||| + \eta_\bullet(u_\bullet^\star)\big).\end{aligned}$$

The term $N+1$ within the maximum can be estimated by

$$N + 1 \le \#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 \overset{\text{(R1)}}{\le} (C_{\mathrm{son}}^{\ell'} - 1)\,\#\mathcal{T}_0.$$

The Céa lemma (4.12) and (A1)–(A3) give that $|||u^\star - u_\bullet^\star||| \lesssim |||u^\star - u_0^\star|||$ and $\eta_\bullet(u_\bullet^\star) \lesssim \eta_0(u_0^\star)$ (see, e.g., [CFPP14, Lemma 3.5]). Altogether, we thus arrive at

$$\|u^\star\|_{\mathbb{A}_s} \lesssim \big(|||u^\star - u_0^\star||| + \eta_0(u_0^\star)\big). \tag{4.53}$$

Step 2. Next, we consider the generic case that $\underline{\ell} = \infty$ and $\eta_{\ell_0}(u_{\ell_0}^k) > 0$ for all $\ell_0 \in \mathbb{N}_0$. Algorithm 15 yields that $\#\mathcal{T}_\ell \to \infty$ as $\ell \to \infty$. Thus, we can argue analogously to the

proof of [CFPP14, Theorem 4.1]: Let $N \in \mathbb{N}_0$. Choose the maximal $\ell \in \mathbb{N}_0$ such that $\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 \leq N$. Then, $\mathcal{T}_\ell \in \mathbb{T}(N)$. The choice of N guarantees that

$$\begin{aligned} N+1 &\leq \#\mathcal{T}_{\ell+1} - \#\mathcal{T}_0 + 1 \\ &\overset{(4.52)}{\leq} \#\mathcal{T}_{\ell+1} \\ &\overset{(R1)}{\leq} C_{\mathrm{son}}\,\#\mathcal{T}_\ell \\ &\overset{(4.52)}{\leq} C_{\mathrm{son}}\,\#\mathcal{T}_0\,(\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1). \end{aligned} \tag{4.54}$$

This leads to

$$(N+1)^s \min_{\mathcal{T}_\bullet \in \mathbb{T}(N)} \big(|||u^\star - u_\bullet^\star||| + \eta_\bullet(u_\bullet^\star)\big) \lesssim (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \big(|||u^\star - u_\ell^\star||| + \eta_\ell(u_\ell^\star)\big).$$

Taking the supremum over all $N \in \mathbb{N}_0$, we conclude that

$$\|u^\star\|_{\mathbb{A}_s} \lesssim \sup_{\ell \in \mathbb{N}_0} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \big(|||u^\star - u_\ell^\star||| + \eta_\ell(u_\ell^\star)\big). \tag{4.55}$$

Step 3. With stability (A1) and the Céa lemma (4.12), we see for all $(\ell, 0) \in \mathcal{Q}$ that

$$\begin{aligned} |||u^\star - u_\ell^\star||| + \eta_\ell(u_\ell^\star) &\overset{(A1)}{\lesssim} |||u^\star - u_\ell^\star||| + |||u_\ell^\star - u_\ell^0||| + \eta_\ell(u_\ell^0) \\ &\leq 2\,|||u^\star - u_\ell^\star||| + |||u^\star - u_\ell^0||| + \eta_\ell(u_\ell^0) \\ &\overset{(4.12)}{\lesssim} |||u^\star - u_\ell^0||| + \eta_\ell(u_\ell^0) \\ &= \Delta_\ell^0. \end{aligned}$$

With (4.53) and (4.55), we thus obtain that

$$\begin{aligned} \|u^\star\|_{\mathbb{A}_s} &\lesssim \sup_{(\ell,0)\in\mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \big(|||u^\star - u_\ell^\star||| + \eta_\ell(u_\ell^\star)\big) \\ &\leq \sup_{(\ell,k)\in\mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s\,\Delta_\ell^k. \end{aligned}$$

This concludes the proof. □

To prove the converse estimate, we need the so-called comparison lemma for the error estimator of the exact discrete solution $u_\ell^\star \in \mathcal{X}_\ell$, i.e., Lemma 4.14 from [CFPP14].

Lemma 26. *Suppose (R1)–(R2) and (A1)–(A4). Let $0 < \theta' < \theta_{\mathrm{opt}} := (1 + C_{\mathrm{stab}}^2\,C_{\mathrm{rel}}^2)^{-1/2}$. Then, there exist constants $C_1, C_2 > 0$ such that for all $s > 0$ with $\|u^\star\|_{\mathbb{A}_s} < \infty$ and all $\mathcal{T}_\bullet \in \mathbb{T}$, there exists a subset $\mathcal{R}_\bullet \subseteq \mathcal{T}_\bullet$ which satisfies that*

$$\#\mathcal{R}_\bullet \leq C_1\,C_2^{-1/s}\,\|u^\star\|_{\mathbb{A}_s}^{1/s}\,\eta_\bullet(u_\bullet^\star)^{-1/s}, \tag{4.56}$$

and the Dörfler marking criterion

$$\theta'\eta_\bullet(u_\bullet^\star) \leq \eta_\bullet(\mathcal{R}_\bullet, u_\bullet^\star). \tag{4.57}$$

The constants C_1, C_2 depend only on the constants of (A1)–(A4). □

Proof of Theorem 23. The proof is split into six steps.

Step 1. It holds that

$$\sup_{(\ell',k')\in\mathcal{Q}} (\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1)^s \, \Delta_{\ell'}^{k'} \le \sup_{(\ell',k')\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)\le(\ell',k')}} \#\mathcal{T}_\ell \Bigg)^s \Delta_{\ell'}^{k'}.$$

Hence, in accordance with Lemma 25, it only remains to prove that

$$\sup_{(\ell',k')\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)\le(\ell',k')}} \#\mathcal{T}_\ell \Bigg)^s \Delta_{\ell'}^{k'} \lesssim \max\big\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\big\}. \tag{4.58}$$

Without loss of generality, we may assume that $\|u^\star\|_{\mathbb{A}_s} < \infty$.

Step 2. Provided that $(\ell+1,0) \in \mathcal{Q}$ (and as a consequence that $\underline{k}(\ell) < \infty$) Lemma 14(i)&(iii) and the stopping criterion (4.16) of Algorithm 15 prove that

$$\begin{aligned}
\mathrm{dl}(u_\ell^\star, u_\ell^k) &\overset{\text{Lemma 14(i)}}{\le} q_{\mathrm{ctr}} \, \mathrm{dl}(u_\ell^\star, u_\ell^{k-1}) \\
&\overset{\text{Lemma 14(iii)}}{\le} \frac{q_{\mathrm{ctr}}}{1-q_{\mathrm{ctr}}} \, \mathrm{dl}(u_\ell^k, u_\ell^{k-1}) \\
&\overset{(4.16)}{\le} \frac{q_{\mathrm{ctr}}}{1-q_{\mathrm{ctr}}} \, \lambda_{\mathrm{ctr}} \, \eta_\ell(u_\ell^k).
\end{aligned}$$

Under (C2), this leads to

$$\begin{aligned}
\|\!|\!| u_\ell^\star - u_\ell^k |\!|\!| = \mathrm{dl}(u_\ell^\star, u_\ell^k) &\le \frac{q_{\mathrm{ctr}}}{1-q_{\mathrm{ctr}}} \, \lambda_{\mathrm{ctr}} \, \eta_\ell(u_\ell^k) \\
&\overset{(4.48)}{\le} C_{\mathrm{stab}}^{-1} \, \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}} \, \eta_\ell(u_\ell^k).
\end{aligned} \tag{4.59a}$$

Under (C1), this leads to

$$\begin{aligned}
\|\!|\!| u_\ell^\star - u_\ell^k |\!|\!| &\overset{(4.9)}{\le} \sqrt{2/\alpha} \, \mathrm{dl}(u_\ell^\star, u_\ell^k) \\
&\le \sqrt{2/\alpha} \, \frac{q_{\mathrm{ctr}}}{1-q_{\mathrm{ctr}}} \, \lambda_{\mathrm{ctr}} \, \eta_\ell(u_\ell^k) \\
&\overset{(4.48)}{\le} C_{\mathrm{stab}}^{-1} \, \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}} \, \eta_\ell(u_\ell^k).
\end{aligned} \tag{4.59b}$$

Step 3. With Step 2, we see that

$$\begin{aligned}
\eta_\ell(u_\ell^k) &\overset{(A1)}{\le} \eta_\ell(u_\ell^\star) + C_{\mathrm{stab}} \, \|\!|\!| u_\ell^\star - u_\ell^k |\!|\!| \overset{(4.59)}{\le} \eta_\ell(u_\ell^\star) + \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}} \, \eta_\ell(u_\ell^k), \\
\eta_\ell(u_\ell^\star) &\overset{(A1)}{\le} \eta_\ell(u_\ell^k) + C_{\mathrm{stab}} \, \|\!|\!| u_\ell^\star - u_\ell^k |\!|\!| \overset{(4.59)}{\le} \eta_\ell(u_\ell^k) + \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}} \, \eta_\ell(u_\ell^k).
\end{aligned}$$

With $0 < \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}} < 1$, this guarantees for all $(\ell+1,0) \in \mathcal{Q}$ the equivalence

$$(1 - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}) \, \eta_\ell(u_\ell^k) \le \eta_\ell(u_\ell^\star) \le (1 + \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}) \, \eta_\ell(u_\ell^k). \tag{4.60}$$

Step 4. Let $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$ be the subset from Lemma 26 with θ' from (4.49). Note that

$$\begin{aligned}\eta_\ell(\mathcal{R}_\ell, u_\ell^\star) &\overset{\text{(A1)}}{\le} \eta_\ell(\mathcal{R}_\ell, u_\ell^k) + C_{\mathrm{stab}}\, |||u_\ell^\star - u_\ell^k||| \\ &\overset{(4.59)}{\le} \eta_\ell(\mathcal{R}_\ell, u_\ell^k) + \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}\, \eta_\ell(u_\ell^k).\end{aligned} \tag{4.61}$$

This proves that

$$\begin{aligned}(1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}})\,\theta'\,\eta_\ell(u_\ell^k) &\overset{(4.60)}{\le} \theta'\,\eta_\ell(u_\ell^\star) \\ &\overset{(4.57)}{\le} \eta_\ell(\mathcal{R}_\ell, u_\ell^\star) \\ &\overset{(4.61)}{\le} \eta_\ell(\mathcal{R}_\ell, u_\ell^k) + \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}\,\eta_\ell(u_\ell^k).\end{aligned} \tag{4.62}$$

The choice of θ' in (4.49) gives that $\theta = (1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}})\,\theta' - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}$. Thus, we obtain that

$$\theta\,\eta_\ell(u_\ell^k) \overset{(4.49)}{=} \big((1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}})\,\theta' - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}\big)\,\eta_\ell(u_\ell^k) \overset{(4.62)}{\le} \eta_\ell(\mathcal{R}_\ell, u_\ell^k).$$

Hence, $\mathcal{R}_\ell$ satisfies the Dörfler marking criterion (4.17) used in Algorithm 15(iii). By (quasi-)minimality of $\mathcal{M}_\ell$ in Algorithm 15(iii), we infer that

$$\#\mathcal{M}_\ell \lesssim \#\mathcal{R}_\ell \overset{(4.56)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\eta_\ell(u_\ell^\star)^{-1/s} \overset{(4.60)}{\simeq} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\eta_\ell(u_\ell^k)^{-1/s}.$$

Nested iteration guarantees that $u_{\ell+1}^0 = u_\ell^k$. Thus, reliability (4.20) and (A1)–(A2) lead to

$$\begin{aligned}\eta_\ell(u_\ell^k) &\overset{(4.20)}{\simeq} \Delta_\ell^k \\ &= |||u^\star - u_{\ell+1}^0||| + \eta_\ell(u_{\ell+1}^0) \\ &\ge |||u^\star - u_{\ell+1}^0||| + \eta_{\ell+1}(u_{\ell+1}^0) \\ &= \Delta_{\ell+1}^0.\end{aligned}$$

Overall, we derive that

$$\#\mathcal{M}_\ell \lesssim \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\eta_\ell(u_\ell^k)^{-1/s} \lesssim \|u^\star\|_{\mathbb{A}_s}^{1/s}\,(\Delta_{\ell+1}^0)^{-1/s} \quad \text{for all } (\ell+1,0) \in \mathcal{Q}. \tag{4.63}$$

The hidden constant depends only on C_{stab}, q_{red}, C_{rel}, $1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}$, C_{mark}, C'_{rel}, and s.

Step 5. For $(\ell,k) \in \mathcal{Q}$ with $\mathcal{T}_\ell \neq \mathcal{T}_0$, Step 4 and the closure estimate (R3) lead to

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 \simeq \#\mathcal{T}_\ell - \#\mathcal{T}_0 \overset{\text{(R3)}}{\lesssim} \sum_{n=0}^{\ell-1} \#\mathcal{M}_n \overset{(4.63)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s} \sum_{n=0}^{\ell} (\Delta_n^0)^{-1/s}.$$

Replacing $\|u^\star\|_{\mathbb{A}_s}$ with $\max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\}$, the overall estimate trivially holds for $\mathcal{T}_\ell = \mathcal{T}_0$. We thus have derived that

$$\begin{aligned}\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 &\lesssim \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\}^{1/s} \sum_{n=0}^{\ell} (\Delta_n^0)^{-1/s} \\ &\le \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\}^{1/s} \sum_{\substack{(\ell',k')\in\mathcal{Q} \\ (\ell',k')\le(\ell,k)}} (\Delta_{\ell'}^{k'})^{-1/s} \quad \text{for all } (\ell,k) \in \mathcal{Q},\end{aligned}$$

where the hidden constant depends only on C_{stab}, q_{red}, C_{rel}, C_{mesh}, $1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}$, C_{mark}, C'_{rel}, Δ_0^0, and s. Finally, we employ linear convergence (4.24) to bound the latter sum by means of the geometric series

$$\sum_{\substack{(\ell',k')\in\mathcal{Q}\\(\ell',k')\le(\ell,k)}} (\Delta_{\ell'}^{k'})^{-1/s} \overset{(4.24)}{\le} C_{\mathrm{lin}}^{1/s}\,(\Delta_\ell^k)^{-1/s} \sum_{\substack{(\ell',k')\in\mathcal{Q}\\(\ell',k')\le(\ell,k)}} (q_{\mathrm{lin}}^{1/s})^{|(\ell,k)|-|(\ell',k')|}$$
$$\le \frac{C_{\mathrm{lin}}^{1/s}}{1-q_{\mathrm{lin}}^{1/s}}\,(\Delta_\ell^k)^{-1/s}.$$

Combining the latter two estimates, we see that

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 \lesssim \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\}^{1/s}(\Delta_\ell^k)^{-1/s} \quad \text{for all } (\ell,k)\in\mathcal{Q}, \tag{4.64}$$

where the hidden constant depends only on C_{stab}, q_{red}, C_{rel}, C_{mark}, $1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}$, C_{mark}, C'_{rel}, C_{lin}, q_{lin}, Δ_0^0, and s.

Step 6. Let $(\ell',k')\in\mathcal{Q}$. Together with Step 5, the geometric series proves that

$$\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)\le(\ell',k')}} \#\mathcal{T}_\ell \overset{(4.52)}{\le} (\#\mathcal{T}_0) \sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)\le(\ell',k')}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)$$
$$\overset{(4.64)}{\lesssim} \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\}^{1/s} \sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)\le(\ell',k')}} (\Delta_\ell^k)^{-1/s}$$
$$\overset{(4.24)}{\lesssim} \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\}^{1/s}\, C_{\mathrm{lin}}^{1/s}\,(\Delta_{\ell'}^{k'})^{-1/s} \sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)\le(\ell',k')}} (q_{\mathrm{lin}}^{1/s})^{|(\ell',k')|-|(\ell,k)|}$$
$$\le \frac{C_{\mathrm{lin}}^{1/s}}{1-q_{\mathrm{lin}}^{1/s}}\,\max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\}^{1/s}\,(\Delta_{\ell'}^{k'})^{-1/s}.$$

Rearranging this estimate, we end up with

$$\sup_{(\ell',k')\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell,k)\in\mathcal{Q}\\(\ell,k)\le(\ell',k')}} \#\mathcal{T}_\ell \Bigg)^s \Delta_{\ell'}^{k'} \lesssim \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^0\},$$

where the hidden constant depends only on C_{stab}, q_{red}, C_{rel}, C_{mesh}, $1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}$, C_{mark}, C'_{rel}, C_{lin}, q_{lin}, Δ_0^0, $\#\mathcal{T}_0$, and s. This concludes the proof. □

4.7 AFEM for linear elliptic PDE with optimal PCG solver

We present our first setting which fits in the abstract framework of Section 4.2–4.6.

Model problem

We consider the elliptic boundary value problem (4.1)

$$\begin{aligned} -\operatorname{div} A(\nabla u^\star) &= f \quad \text{in } \Omega \\ u^\star &= 0 \quad \text{on } \Gamma := \partial\Omega, \end{aligned}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain with $d \in \{2,3\}$, and $f \in L^2(\Omega)$ is a given load. Recall the corresponding variational formulation (4.2): Given a load $f \in L^2(\Omega)$, find $u^\star \in \mathcal{H} := H^1_0(\Omega)$ such that

$$\langle \mathcal{A}u^\star \,,\, v\rangle_{\mathcal{H}'\times\mathcal{H}} := \int_\Omega A(\nabla u^\star)\cdot\nabla v\,\mathrm{d}x = \int_\Omega f v_\ell\,\mathrm{d}x =: \langle F \,,\, v\rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{for all } v\in\mathcal{H}.$$

We assume that $A\colon L^2(\Omega)^d \to L^2(\Omega)^d$ has the given form

$$A(\mathbf{v}) = [x \mapsto \mathbf{A}(x)\mathbf{v}(x)] \quad \text{for } \mathbf{v}\in L^2(\Omega)^d,$$

where $\mathbf{A} \in W^{1,\infty}(\Omega)^{d\times d}$ is symmetric and uniformly positive definite. The choice of $W^{1,\infty}(\Omega)$ as the domain of $\mathbf{A}$ instead of $L^\infty(\Omega)$ is only necessary to ensure that the residual error indicators (4.69) are well-defined.

We define the potential $P\colon H^1_0(\Omega)\to\mathbb{R}$ via

$$P(v) = \frac{1}{2}\int_\Omega \mathbf{A}\nabla v\cdot\nabla v\,\mathrm{d}x \quad \text{for all } v\in H^1_0(\Omega). \tag{4.65}$$

Then, it holds that

$$\begin{aligned} \lim_{\substack{t\to 0\\ t\in\mathbb{R}}} \frac{P(w+tv)-P(w)}{t} &= \lim_{\substack{t\to 0\\ t\in\mathbb{R}}} \frac{\int_\Omega \mathbf{A}\nabla(w+tv)\cdot\nabla(w+tv)\,\mathrm{d}x - \int_\Omega \mathbf{A}\nabla w\cdot\nabla w\,\mathrm{d}x}{2\,t} \\ &= \lim_{\substack{t\to 0\\ t\in\mathbb{R}}} \frac{\int_\Omega 2\,\mathbf{A}\nabla w\cdot\nabla(tv) + \mathbf{A}\nabla(tv)\cdot\nabla(tv)\,\mathrm{d}x}{2\,t} \\ &= \lim_{\substack{t\to 0\\ t\in\mathbb{R}}} \int_\Omega \mathbf{A}\nabla w\cdot\nabla v + \frac{1}{2}\mathbf{A}\nabla v\cdot\nabla(tv)\,\mathrm{d}x \\ &= \int_\Omega \mathbf{A}\nabla w\cdot\nabla v\,\mathrm{d}x \\ &= \langle \mathcal{A}u^\star \,,\, v\rangle_{\mathcal{H}'\times\mathcal{H}} \end{aligned}$$

Hence, assumption (O3) is satisfied.

We equip $H^1_0(\Omega)$ with the scalar product

$$\langle\!\langle v \,,\, w\rangle\!\rangle := \int_\Omega \mathbf{A}\nabla v\cdot\nabla w\,\mathrm{d}x \tag{4.66}$$

and the induced norm $|\!|\!| v |\!|\!|^2 := \langle\!\langle v \,,\, v\rangle\!\rangle$. Then, the assumptions (O1)–(O2) are satisfied with $\alpha = 1 = L$.

Triangulation and mesh-refinement

Let $\mathcal{T}_0$ be a conforming initial triangulation of Ω into simplices $T \in \mathcal{T}_0$. We use newest vertex bisection for the mesh-refinement $\texttt{refine}(\cdot)$ such that the axioms (R1)–(R3) are satisfied, cf. Section 3.6. In this section, we define the local mesh-width function as

$$h_\ell|_T := h_\ell(T) := \operatorname{diam}(T) \quad \text{for all } T \in \mathcal{T}_\ell,$$

which is equivalent to the definition of Section 3.1. For a node $z \in \mathcal{T}_\ell$, we additionally define the mesh-width

$$h_\ell(z) := \max_{\substack{T \in \mathcal{T}_\ell \\ T \subseteq \omega_\ell(z)}} \operatorname{diam}(T).$$

It holds that

$$h_\ell(T) \le h_\ell(z) \lesssim h_\ell(T) \quad \text{for all } z \in \mathcal{N}_\ell \text{ and } T \in \mathcal{T}_\ell \text{ with } z \in T, \tag{4.67}$$

where the hidden constant depends only on γ-shape regularity.

Discretization

For $\mathcal{T}_\ell \in \mathbb{T}$, we use the corresponding ansatz space

$$\mathcal{X}_\ell := \big\{ v \in C(\Omega) \,:\, v|_\Gamma = 0 \text{ and } v|_T \in \mathcal{P}^1 \text{ for all } T \in \mathcal{T}_\ell \big\}, \tag{4.68}$$

i.e., the space of all continuous piecewise first degree polynomials that vanish on the boundary $\Gamma = \partial\Omega$.

Error estimator

Next, we define the weighted-residual error indicators (see, e.g., [AO11, Ver13]). For all $T \in \mathcal{T}_\ell$ and $v_\ell \in \mathcal{X}_\ell$ define the error indicators $\eta_\ell(T, v_\ell)^2$ as

$$\eta_\ell(T, v_\ell)^2 := |T|^{2/d} \, \| f + \operatorname{div}(\mathbf{A}\nabla v_\ell) \|_{L^2(T)} + |T|^{1/d} \, \| [\mathbf{A}\nabla v_\ell \cdot \mathbf{n}] \|_{L^2(\partial T \cap \Omega)}, \tag{4.69}$$

where $[\cdot]$ denotes the usual jump of piecewise continuous functions across element interfaces, and $\mathbf{n}$ is the outer normal vector of the considered element. It is well-known that the resulting error estimator satisfies the axioms (A1)–(A4), see, e.g., [CFPP14, Section 6.1] and the references therein.

Galerkin system

With the usual Lagrangian basis $\{\eta_{\ell,1}, \dots, \eta_{\ell,N}\} \subseteq \mathcal{X}_\ell$ of $\mathcal{X}_\ell$, we define the Galerkin matrix $\mathbf{M}_\ell$ via

$$\mathbf{M}_\ell := \Big(\int_\Omega \mathbf{A}\nabla\eta_{\ell,j} \cdot \nabla\eta_{\ell,i} \, \mathrm{d}x \Big)_{i,j=1}^N \in \mathbb{R}^{N \times N},$$

as well as the right-hand side,

$$\mathbf{b}_\ell := \Big(\int_\Omega f \eta_{\ell,i} \, \mathrm{d}x \Big)_{i=1}^N \in \mathbb{R}^N$$

corresponding to (4.8). Hence, the coefficient vector $\mathbf{x}_\ell^\star \in \mathbb{R}^N$ of the solution $u_\ell^\star = \sum_{i=1}^N \mathbf{x}_\ell^\star[i] \, \eta_{\ell,i}$ is the unique solution of the linear system

$$\mathbf{M}_\ell \, \mathbf{x}_\ell^\star = \mathbf{b}_\ell. \tag{4.70}$$

Preconditioned conjugate gradient method (PCG) for the Galerkin system

Finally, we introduce the iteration function $\Phi_\ell : \mathcal{X}_\ell \to \mathcal{X}_\ell$ for Step (i) of Algorithm 15 as one step of the preconditioned conjugated gradient method (PCG): Given an initial guess $\mathbf{x}_\ell^0 \in \mathbb{R}^N$, PCG approximates the solution $\mathbf{x}_\ell^\star \in \mathbb{R}^N$ of (4.70).

Let $\mathbf{P}_\ell \in \mathbb{R}^{N \times N}$ be an arbitrary symmetric positive definite preconditioner and define

$$\widetilde{\mathbf{M}}_\ell := \mathbf{P}_\ell^{-1/2} \mathbf{M}_\ell \mathbf{P}_\ell^{-1/2}$$

as well as

$$\widetilde{\mathbf{b}}_\ell := \mathbf{P}_\ell^{-1/2} \mathbf{b}_\ell.$$

Now, instead of solving the linear system (4.70), the PCG iteration considers the preconditioned system

$$\widetilde{\mathbf{M}}_\ell \, \widetilde{\mathbf{x}}_\ell^\star = \widetilde{\mathbf{b}}_\ell \tag{4.71}$$

and formally applies the conjugate gradient method (CG, cf. [GVL13, Algorithm 11.3.2]) to (4.71) with the given initial guess $\mathbf{x}_\ell^0$. Note that $\mathbf{x}_\ell^\star$ and $\widetilde{\mathbf{x}}_\ell^\star$ are connected via

$$\mathbf{x}_\ell^\star = \mathbf{P}_\ell^{-1/2} \widetilde{\mathbf{x}}_\ell^\star.$$

Also, the iterates $\mathbf{x}_\ell^k \in \mathbb{R}^N$ of PCG (for $\mathbf{P}_\ell$, $\mathbf{M}_\ell$, $\mathbf{b}_\ell$, and the initial guess $\mathbf{x}_\ell^0$) and the iterates $\widetilde{\mathbf{x}}_\ell^k$ of CG (for $\widetilde{\mathbf{M}}_\ell$, $\widetilde{\mathbf{b}}_\ell$, and the initial guess $\widetilde{\mathbf{x}}_\ell^0 := \mathbf{P}_\ell^{1/2} \mathbf{x}_\ell^0$) are formally linked by

$$\mathbf{x}_\ell^k = \mathbf{P}_\ell^{-1/2} \widetilde{\mathbf{x}}_\ell^k,$$

see [GVL13, Section 11.5].

Let $v_\ell \in \mathcal{X}_\ell$ with coefficient vector $\mathbf{y}_\ell \in \mathbb{R}^N$. Then, there holds the elementary identity

$$\|v_\ell\|^2 = \mathbf{y}_\ell \cdot \mathbf{M}_\ell \, \mathbf{y}_\ell =: |\mathbf{y}_\ell|_{\mathbf{M}_\ell}^2. \tag{4.72}$$

In addition, for $\widetilde{\mathbf{y}}_\ell \in \mathbb{R}^N$ and $\mathbf{y}_\ell \in \mathbb{R}^N$ such that $\mathbf{y}_\ell = \mathbf{P}_\ell^{-1/2} \, \widetilde{\mathbf{y}}_\ell$, direct computation yields that

$$\begin{aligned} |\widetilde{\mathbf{y}}_\ell|_{\widetilde{\mathbf{M}}_\ell}^2 &:= \widetilde{\mathbf{y}}_\ell \cdot \widetilde{\mathbf{M}}_\ell \widetilde{\mathbf{y}}_\ell \\ &= (\mathbf{P}_\ell^{1/2} \, \mathbf{y}_\ell) \cdot \mathbf{P}_\ell^{-1/2} \, \mathbf{M}_\ell \, \mathbf{P}_\ell^{-1/2} \, \mathbf{P}_\ell^{1/2} \, \mathbf{y}_\ell \\ &= \mathbf{y}_\ell \cdot \mathbf{M}_\ell \, \mathbf{y}_\ell \\ &= |\mathbf{y}_\ell|_{\mathbf{M}_\ell}^2. \end{aligned} \tag{4.73}$$

Hence, [GVL13, Theorem 11.3.3] for CG (applied to $\widetilde{\mathbf{M}}_\ell$, $\widetilde{\mathbf{b}}_\ell$, $\widetilde{\mathbf{x}}_\ell^0$) yields the following lemma for PCG (which follows from the implicit steepest decent property of CG).

Lemma 27. *Let $\mathbf{M}_\ell, \mathbf{P}_\ell \in \mathbb{R}^{N\times N}$ be symmetric and positive definite, $\mathbf{b}_\ell \in \mathbb{R}^N$, $\mathbf{x}_\ell^\star := \mathbf{M}_\ell^{-1}\,\mathbf{b}_\ell$, and $\mathbf{x}_\ell^0 \in \mathbb{R}^N$. Suppose the ℓ_2-condition number estimate*

$$\operatorname{cond}_2(\mathbf{P}_\ell^{-1/2}\,\mathbf{M}_\ell\,\mathbf{P}_\ell^{-1/2}) \le C_{\mathrm{alg}}. \tag{4.74}$$

Then, the iterates $\mathbf{x}_\ell^k$ of the PCG algorithm satisfy the contraction property

$$|\mathbf{x}_\ell^\star - \mathbf{x}_\ell^{k+1}|_{\mathbf{M}_\ell} \le q_{\mathrm{pcg}}\,|\mathbf{x}_\ell^\star - \mathbf{x}_\ell^k|_{\mathbf{M}_\ell} \quad \text{for all } k \in \mathbb{N}_0, \tag{4.75}$$

where $q_{\mathrm{pcg}} := (1 - 1/C_{\mathrm{alg}})^{1/2} < 1$. □

Remark 28. *Each step of PCG has the following computational costs:*

- *$\mathcal{O}(N)$ costs for vector operations (e.g., assignment, addition, scalar product),*
- *computation of* one *matrix-vector product with $\mathbf{M}_\ell$,*
- *computation of* one *matrix-vector product with $\mathbf{P}_\ell^{-1}$.*

Optimal preconditioner

We suppose that the employed preconditioners $\mathbf{P}_\ell$ are *optimal*. This means that the constant $C_{\mathrm{alg}} > 0$ of Lemma 27 depends only on the coefficient matrix $\mathbf{A}$, the initial mesh $\mathcal{T}_0$, and the polynomial degree p. One example of such an optimal preconditioner is the multilevel additive Schwarz preconditioner from Section 4.7.1, see also, e.g., [WC06, SMPZ08, XCH10, CNX12]. We stress that the product of $\mathbf{P}_\ell$ with *one* vector can be realized in linear complexity $\mathcal{O}(N)$.

Hence, to fit the framework of the main results from Section 4.6, at least one of the contraction properties (C1)–(C2) has to be fulfilled: From the contraction property (4.75) and the identity (4.72), it follows that

$$\begin{aligned} |||u_\ell^\star - u_\ell^{k+1}||| &\overset{(4.72)}{=} |\mathbf{x}_\ell^\star - \mathbf{x}_\ell^{k+1}|_{\mathbf{M}_\ell} \\ &\overset{(4.75)}{\le} q_{\mathrm{pcg}}^2\,|\mathbf{x}_\ell^\star - \mathbf{x}_\ell^k|_{\mathbf{M}_\ell} \\ &\overset{(4.72)}{=} q_{\mathrm{pcg}}\,|||u_\ell^\star - u_\ell^k||| \end{aligned}$$

Hence, there holds the contraction property (C2) with $q_{\mathrm{ctr}} := q_{\mathrm{pcg}} = (1 - 1/C_{\mathrm{alg}})^{1/2}$.

From (4.65)–(4.66), it directly follows that

$$|\mathcal{E}(v) - \mathcal{E}(w)| = \frac{1}{2}\,|||w - v|||^2 \quad \text{for all } v, w \in H_0^1(\Omega).$$

Thus, the norm contraction property (C2) is equivalent to the energy contraction property (C1). Altogether, the main results from Section 4.6 apply to the present setting and the linear convergence (4.24) from Theorem 17 holds even for arbitrary $\lambda_{\mathrm{ctr}} > 0$ and $0 < \theta \le 1$ in Algorithm 15.

4.7.1 Optimal multilevel additive Schwarz preconditioner

In this section, we propose a multilevel additive Schwarz preconditioner for the arising Galerkin matrix and prove its optimality in the sense that the condition number of the additive Schwarz matrix is uniformly bounded.

Multilevel additive Schwarz preconditioner

In order to define the additive Schwarz preconditioner, we introduce the set of vertices $\widetilde{\mathcal{N}}_\ell$ for $\ell \in \mathbb{N}_0$ via

$$\widetilde{\mathcal{N}}_0 := \mathcal{N}_0$$

as well as

$$\widetilde{\mathcal{N}}_\ell := \mathcal{N}_\ell \setminus \mathcal{N}_{\ell-1} \cup \big\{ z \in \mathcal{N}_\ell \cap \mathcal{N}_{\ell-1} \,:\, \omega_\ell(z) \subsetneqq \omega_{\ell-1}(z) \big\} \quad \text{for } \ell \geq 1.$$

Hence, $\widetilde{\mathcal{N}}_\ell$ is the set of new vertices and their direct neighbors in the mesh $\mathcal{T}_\ell$. Additionally, we define the corresponding subspaces

$$\widetilde{\mathcal{X}}_\ell := \operatorname{span}\{\eta_{\ell,z} \,:\, z \in \widetilde{\mathcal{N}}_\ell\}$$

as well as

$$\mathcal{X}_{\ell,z} := \operatorname{span}\{\eta_{\ell,z}\}.$$

Then, for all $0 \leq L$ and with $N_\ell := \#\mathcal{N}_\ell$, the local multilevel diagonal preconditioner is given by

$$\mathbf{P}_L := \sum_{\ell=0}^{L} \mathbf{I}_\ell \widetilde{\mathbf{D}}_\ell^{-1} (\mathbf{I}_\ell)^\mathsf{T}, \tag{4.76}$$

where the appearing matrices are defined as follows:

- $\widetilde{\mathbf{D}}_\ell^{-1} \in \mathbb{R}^{N_\ell \times N_\ell}$ is a diagonal matrix with entries

$$\big(\widetilde{\mathbf{D}}_\ell^{-1}\big)(j,k) := \begin{cases} \big(\mathbf{M}_\ell(j,j)\big)^{-1} \delta_{jk} & \text{if } z_j \in \widetilde{\mathcal{N}}_\ell, \\ 0 & \text{otherwise,} \end{cases}$$

  where δ_{jk} is the usual Kronecker delta. Hence, for all degrees of freedom in $\widetilde{\mathcal{N}}_\ell$, the corresponding diagonal elements of $\widetilde{\mathbf{D}}_\ell^{-1}$ are the inverse diagonal entries of $\mathbf{M}_\ell$.

- $\mathbf{I}_\ell \in \mathbb{R}^{N_L \times N_\ell}$ is the matrix representation of the embedding operator $\mathcal{I}_\ell \colon \mathcal{X}_\ell \to \mathcal{X}_L$.

Instead of solving the linear system

$$\mathbf{M}_L \mathbf{x}_L = \mathbf{b}_L,$$

we instead consider the preconditioned linear system

$$\mathbf{P}_L \mathbf{M}_L \mathbf{x}_L = \mathbf{P}_L \mathbf{b}_L.$$

Optimal cost of matrix-vector multiplication

Let $\mathbf{I}_\ell^{\ell+1} \in \mathbb{R}^{N_{\ell+1}\times N_\ell}$ denote the matrix representation of the embedding operator from $\mathcal{X}_{\ell-1}$ to $\mathcal{X}_\ell$. Then, it holds that

$$\mathbf{I}_\ell = \mathbf{I}_{L-1}^L\, \mathbf{I}_{L-2}^{L-1} \cdots \mathbf{I}_\ell^{\ell+1}.$$

Hence, we can rewrite the preconditioner $\mathbf{P}_L$ from (4.76) as follows

$$\begin{aligned}\mathbf{P}_L &= \sum_{\ell=0}^{L} \mathbf{I}_\ell \widetilde{\mathbf{D}}_\ell^{-1} (\mathbf{I}_\ell)^\intercal \\ &= \mathbf{I}_{L-1}^L \cdots \mathbf{I}_0^1\, \widetilde{\mathbf{D}}_0^{-1}\, (\mathbf{I}_0^1)^\intercal \cdots (\mathbf{I}_{L-1}^L)^\intercal + \ldots + \mathbf{I}_{L-1}^L\, \widetilde{\mathbf{D}}_{L-1}^{-1}\, (\mathbf{I}_{L-1}^L)^\intercal + \widetilde{\mathbf{D}}_L^{-1}.\end{aligned}$$

Using this representation, we can evaluate the matrix-vector multiplication with the preconditioner $\mathbf{P}_L$ with the following algorithm.

Algorithm 29 (Evaluation of $\mathbf{y} = \mathbf{P}_L\,\mathbf{x}$). ***Input:*** $\mathbf{y} := \mathbf{x} \in \mathbb{R}^{N_L}$, *matrices* $\{\mathbf{I}_\ell^{\ell+1}\}_{\ell=0}^{L-1}$, $\{\widetilde{\mathbf{D}}_\ell^{-1}\}_{\ell=0}^{L}$, *auxiliary memory* $\mathbf{y}_0 \in \mathbb{R}^{N_0}, \ldots, \mathbf{y}_L \in \mathbb{R}^{N_L}$.

(i) **For** $\ell = L, \ldots, 1$ **do**:

$\mathbf{y}_\ell \longleftarrow \widetilde{\mathbf{D}}_\ell^{-1}\,\mathbf{y}$

$\mathbf{y} \longleftarrow (\mathbf{I}_{\ell-1}^\ell)^\intercal\,\mathbf{y}$

End for

(ii) $\mathbf{y}_0 \longleftarrow \widetilde{\mathbf{D}}_0^{-1}\,\mathbf{y}$

(iii) **For** $\ell = 0, \ldots, L-1$ **do**:

$\mathbf{y} \longleftarrow \mathbf{I}_\ell^{\ell+1}\,\mathbf{y}$

$\mathbf{y} \longleftarrow \mathbf{y} + \mathbf{y}_{\ell+1}$

End for

Output: $\mathbf{y} = \mathbf{P}_L\,\mathbf{x}$.

In order to analyze the computational costs of Algorithm 29, we first note that $\widetilde{\mathcal{N}}_\ell$ consists only of newly created nodes and some of its neighbours. This yields that

$$\widetilde{N}_\ell := \#\widetilde{\mathcal{N}}_\ell \leq C(N_\ell - N_{\ell-1}) = C\#(\mathcal{N}_\ell \setminus \mathcal{N}_{\ell-1}),$$

where the constant $C > 0$ depends only on shape regularity. Since the matrices $\widetilde{\mathbf{D}}_\ell^{-1}$ have only $\mathcal{O}(N_\ell - N_{\ell-1})$ non-zero entries, the overall storage requirements are

$$\mathcal{O}\big(N_0 + \sum_{\ell=1}^{L}(N_\ell - N_{\ell-1})\big) = \mathcal{O}(N_L).$$

The same holds for the evaluations $\mathbf{I}_{\ell-1}^{\ell}\,\mathbf{x}$ as well as $(\mathbf{I}_{\ell-1}^{\ell})^{\mathsf{T}}\,\mathbf{x}$. All values of $\mathbf{x}$ with indices corresponding to nodes in $\mathcal{N}_\ell$ remain unchanged during the evaluation and we hence only need $\mathcal{O}(N_{\ell+1}-N_\ell)$ many arithmetic operations. Summing up all operations in Algorithm 29, we then end up with linear complexity $\mathcal{O}(N_L)$ for the evaluation of the preconditioner $\mathbf{P}_L$.

Optimal condition number

The following theorem is the main result of this section.

Theorem 30. *The minimal and maximal eigenvalues of $\mathbf{P}_L\mathbf{M}_L$ satisfy*

$$c \le \lambda_{\min}(\mathbf{P}_L\mathbf{M}_L) \quad \text{and} \quad \lambda_{\max}(\mathbf{P}_L\mathbf{M}_L) \le C, \tag{4.77}$$

where the constants $c, C > 0$ depend only on Ω, d, the initial triangulation $\mathcal{T}_0$, and the diffusion coefficient $\boldsymbol{A}$. In particular, it holds that

$$\operatorname{cond}_{\mathbf{M}_L}(\mathbf{P}_L\mathbf{M}_L) \le \frac{C}{c}, \tag{4.78}$$

i.e., the condition number of the preconditioned matrix $\mathbf{P}_L\mathbf{M}_L$ is L-independently bounded and therefrom the multilevel diagonal scaling preconditioner $\mathbf{P}_L$ is optimal.

4.7.2 Auxiliary results

Level function and uniform mesh-refinement

In this section, we define the level function $\operatorname{level}_\ell(\cdot)$ as well as the sequence of uniformly refined triangulations $\widehat{\mathcal{T}}_m$ and collect some technical results.

To this end, we first define the generation $\operatorname{gen}(T) \in \mathbb{N}_0$ of an element T. Let $T \in \mathcal{T}_\ell$ be an element of the triangulation $\mathcal{T}_\ell$ and $T_0 \in \mathcal{T}_0$ the unique ancestor element of the initial triangulation $\mathcal{T}_0$ such that $T \subseteq T_0$. Then, the generation of T is defined by

$$\operatorname{gen}(T) := \frac{\log(|T|/|T_0|)}{\log(1/2)} \in \mathbb{N}_0,$$

i.e., $|T| = 2^{-\operatorname{gen}(T)}|T_0|$ and $\operatorname{gen}(T)$ returns the number of bisections to generate T from T_0. Based on the generation, we now define for each node $z \in \mathcal{N}_\ell$ the level

$$\operatorname{level}_\ell(z) := \lceil \max\{\operatorname{gen}(T)/d \,:\, T \in \mathcal{T}_\ell \text{ with } T \subseteq \omega_\ell(z)\}\rceil, \tag{4.79}$$

where $\lceil\cdot\rceil$ denotes the Gaussian ceil function, i.e., $\lceil x\rceil := \min\{n \in \mathbb{N}_0 \,:\, x \le n\}$ for $x \ge 0$.

Next, let $z \in \mathcal{N}_L$ and $k \in \mathbb{N}_0$. We define the index set

$$\widetilde{\mathcal{K}}_k(z) := \big\{\ell \in \{0,1,\dots,L\} \,:\, z \in \widetilde{\mathcal{N}}_\ell \text{ and } \operatorname{level}_\ell(z) = k\big\}, \tag{4.80}$$

which describes in how many sets $\widetilde{\mathcal{N}}_\ell$ with $\operatorname{level}_\ell(z) = k$ a given node $z \in \mathcal{N}_L$ appears. The following lemma from [WC06, Lemma 3.1] proves that the cardinality of this set can be uniformely bounded.

Lemma 31. *It holds that*

$$\#\widetilde{\mathcal{K}}_k(z) \le C \quad \text{for all } z \in \mathcal{N}_L \text{ and } k \in \mathcal{N}_0, \tag{4.81}$$

where the constant $C > 0$ depends only on $\mathcal{T}_0$. □

The sequence of uniform triangulations $\widehat{\mathcal{T}}_m$ is defined as follows: Let $\widehat{\mathcal{T}}_0 := \mathcal{T}_0$. For $m \ge 1$, the mesh $\widehat{\mathcal{T}}_m$ is obtained by uniformely refining the mesh $\widehat{\mathcal{T}}_{m-1}$, i.e., every element $T \in \widehat{\mathcal{T}}_{m-1}$ is successively bisected into 2^d many son elements $T' \in \widehat{\mathcal{T}}_m$ with measure $|T'| = 2^{-d}|T|$, cf. [Ste08, Theorem 2.1]. With $\widehat{\mathcal{N}}_m$ denoting the set of all nodes of $\widehat{\mathcal{T}}_m$, we define the local mesh-width

$$\widehat{h}_0 := \max_{T \in \mathcal{T}_0} h_0(T) \quad \text{and} \quad \widehat{h}_m := 2^{-m}\widehat{h}_0 \quad \text{for all } m \ge 1. \tag{4.82}$$

From [Ste08, Section 4], we get the equivalence

$$|T| \simeq h_\ell(T)^d = \operatorname{diam}(T)^d \simeq 2^{-\operatorname{gen}(T)} \quad \text{for all } T \in \mathcal{T}_\ell,$$

where the implicit constants depend only on $\mathcal{T}_0$ and d. Hence, it holds that

$$\widehat{h}_m = 2^{-m}\widehat{h}_0 = 2^{-\operatorname{gen}(T)/d}\,\widehat{h}_0 \simeq \operatorname{diam}(T) \quad \text{for all } T \in \widehat{\mathcal{T}}_m \text{ and } m \ge 0.$$

Lemma 32. *Let $z \in \mathcal{N}_\ell$ and $m := \operatorname{level}_\ell(z)$. Then, it holds that $z \in \widehat{\mathcal{N}}_m$ as well as $\eta_{\ell,z} \in \widehat{\mathcal{X}}_m := \mathcal{S}^1(\widehat{\mathcal{T}}_m)$. Additionally, there holds the equivalence*

$$c\,\widehat{h}_m \le h_\ell(z) \le C\,\widehat{h}_m, \tag{4.83}$$

where $h_\ell(z) := \max\{\operatorname{diam}(T) : T \in \mathcal{T}_\ell, z \in T\}$ and the constants $c, C > 0$ depend only on the initial triangulation $\mathcal{T}_0$.

Proof. For $\widehat{T} \in \widehat{\mathcal{T}}_m$ and $T \in \mathcal{T}_\ell$ with $T \subseteq \omega_\ell(z)$, it holds that

$$\operatorname{gen}(\widehat{T}) = m\,d \ge \operatorname{gen}(T). \tag{4.84}$$

Now, let $z' \in \omega_\ell(z) \cap \mathcal{N}_\ell$ and $T \in \mathcal{T}_\ell$ with $T \subseteq \omega_\ell(z)$ such that $z' \in T$. Let $T_0 \in \mathcal{T}_0$ be the unique ancestor of T. From (4.84), it follows that there exists a $\widehat{T} \in \widehat{\mathcal{T}}_m$ such that $\widehat{T} \subseteq T \subseteq T_0$ and $z' \in \widehat{\mathcal{N}}_m \cap \widehat{T}$. Hence, it holds for all nodes $z' \in \omega_\ell(z) \cap \mathcal{N}_\ell$ that $z' \in \widehat{\mathcal{N}}_m$ and consequently $\eta_{\ell,z} \in \widehat{\mathcal{X}}_m$. To see (4.83), recall the definition (4.79) of $m = \operatorname{level}_\ell(z)$, i.e., there exists $T' \in \mathcal{T}_\ell$ with $T' \subseteq \omega_\ell(z)$ such that

$$\operatorname{gen}(T') + 1 > m\,d = \operatorname{gen}(\widehat{T}) \ge \operatorname{gen}(T').$$

Therefore, it holds that

$$\operatorname{diam}(\widehat{T}) \simeq \operatorname{diam}(T') \simeq \operatorname{diam}(T) \quad \text{for all } \widehat{T} \in \widehat{\mathcal{T}}_m \text{ and } T \subseteq \omega_\ell(z).$$

This implies the equivalence (4.83). □

Let $\widehat{\Pi}_m\colon L^2(\Omega) \to \widehat{\mathcal{X}}_m$ denote the L^2-orthogonal projection onto $\widehat{\mathcal{X}}_m = \mathcal{S}^1_0(\widehat{\mathcal{T}}_m)$.

Lemma 33. *For all $v \in H^1_0(\Omega)$, it holds that*

$$\sum_{m=0}^{\infty} \widehat{h}_m^{-2} \|v - \widehat{\Pi}_m v\|^2_{L^2(\Omega)} \le C_{\mathrm{norm}} \|v\|^2_{H^1(\Omega)}, \tag{4.85}$$

where the constant $C_{\mathrm{norm}} > 0$ depends only on Ω and the initial triangulation $\mathcal{T}_0$.

Proof. Let $w \in H^1_0(\Omega)$. It follows from the orthogonality of the L^2-projection that

$$\begin{aligned} \sum_{k=1}^{N} \|(\widehat{\Pi}_k - \widehat{\Pi}_{k-1})w\|^2_{L^2(\Omega)} &= \|\sum_{k=1}^{N} (\widehat{\Pi}_k - \widehat{\Pi}_{k-1})w\|^2_{L^2(\Omega)} \\ &= \|\widehat{\Pi}_N w - \widehat{\Pi}_0 w\|^2_{L^2(\Omega)} \\ &= \|(1-\widehat{\Pi}_0)w\|^2_{L^2(\Omega)} - \|(1-\widehat{\Pi}_N)w\|^2_{L^2(\Omega)}. \end{aligned} \tag{4.86}$$

Taking the limit $N \to \infty$, we hence get that

$$\|w - \widehat{\Pi}_0 w\|^2_{L^2(\Omega)} = \sum_{k=1}^{\infty} \|(\widehat{\Pi}_k - \widehat{\Pi}_{k-1})w\|^2_{L^2(\Omega)} \quad \text{for all } w \in H^1_D(\Omega), \tag{4.87}$$

since the last term in (4.87) converges to 0 for $N \to \infty$. From [Xu96, Theorem 4.32] follows that

$$\|w - \widehat{\Pi}_0 w\|^2_{H^1(\Omega)} \simeq \sum_{k=1}^{\infty} \widehat{h}_k^{-2} \|(\widehat{\Pi}_k - \widehat{\Pi}_{k-1})w\|^2_{L^2(\Omega)} \quad \text{for all } w \in H^1_0(\Omega). \tag{4.88}$$

With $w := v - \widehat{\Pi}_m v$, and $\widehat{\Pi}_n \widehat{\Pi}_m v = \widehat{\Pi}_{\min\{m,n\}} v$, we get that

$$\begin{aligned} \|v - \widehat{\Pi}_m v\|^2_{L^2(\Omega)} = \|w\|_{L^2(\Omega)} &= \|w - \widehat{\Pi}_0 w\|^2_{L^2(\Omega)} \\ &\overset{(4.87)}{=} \sum_{k=1}^{\infty} \|(\widehat{\Pi}_k - \widehat{\Pi}_{k-1})w\|^2_{L^2(\Omega)} \\ &= \sum_{k=m+1}^{\infty} \|(\widehat{\Pi}_k - \widehat{\Pi}_{k-1})v\|^2_{L^2(\Omega)}. \end{aligned} \tag{4.89}$$

With the definition (4.82) of $\widehat{h}_m$, we infer that

$$\sum_{m=0}^{k-1} \widehat{h}_m^{-2} = \widehat{h}_0^{-2} \sum_{m=0}^{k-1} 2^{2m} < \widehat{h}_0^{-2}\, 2^{2k} = \widehat{h}_k^{-2}. \tag{4.90}$$

Combining (4.89)–(4.90), changing the order of summation, and exploiting (4.88), we derive that

$$\begin{aligned}\sum_{m=0}^{\infty} \widehat{h}_m^{-2}\|v-\widehat{\Pi}_m v\|_{L^2(\Omega)}^2 &\overset{(4.89)}{\simeq} \sum_{m=0}^{\infty}\sum_{k=m+1}^{\infty} \widehat{h}_m^{-2}\,\|(\widehat{\Pi}_k-\widehat{\Pi}_{k-1})v\|_{L^2(\Omega)}^2\\ &= \sum_{k=1}^{\infty}\sum_{m=0}^{k-1} \widehat{h}_m^{-2}\,\|(\widehat{\Pi}_k-\widehat{\Pi}_{k-1})v\|_{L^2(\Omega)}^2\\ &\overset{(4.90)}{<} \sum_{k=1}^{\infty} \widehat{h}_k^{-2}\,\|(\widehat{\Pi}_k-\widehat{\Pi}_{k-1})v\|_{L^2(\Omega)}^2\\ &\overset{(4.88)}{\simeq} \|v-\widehat{\Pi}_0 v\|_{H^1(\Omega)}^2\\ &\lesssim \|v\|_{H^1(\Omega)}^2,\end{aligned}$$

where the last inequality follows from the H^1-stability of the L^2-orthogonal projection $\widehat{\Pi}_0$, cf. [CT87, BPS02, Car02]. This concludes the proof. □

The patches $\widehat{\omega}_m^k(z)$ corresponding to the uniformly refined mesh $\widehat{\mathcal{T}}_m$ are defined analogously to the patches $\omega_m^k(z)$.

For each $z \in \mathcal{N}_L$, we define

$$r_\ell(z) := \min\big\{\mathrm{gen}(T) \,:\, T \in \mathcal{T}_{\ell-1} \text{ with } T \subseteq \omega_{\ell-1}^2(z)\big\} \tag{4.91}$$

as well as

$$R_\ell(z) := \lfloor r_\ell(z)/d \rfloor, \tag{4.92}$$

where $\lfloor\cdot\rfloor$ denotes the Gaussian floor function, i.e., $\lfloor x\rfloor := \max\big\{n \in \mathbb{N}_0 \,:\, x \geq n\big\}$.

Lemma 34. *For all $z \in \mathcal{N}_\ell$, there hold* (i)–(iii)*:*

(i) *It holds that $\mathrm{level}_\ell(z) \leq R_\ell(z) + C_1$, where the constant $C_1 > 0$ depends only on the initial triangulation $\mathcal{T}_0$.*

(ii) *For all $T \in \mathcal{T}_{\ell-1}$ with $T \subseteq \omega_{\ell-1}^2(z)$, there exists an element $\widehat{T} \in \widehat{\mathcal{T}}_{R_\ell(z)}$ such that $T \subseteq \widehat{T}$.*

(iii) *There exists an index $n \in \mathbb{N}_0$, which depends only on the initial triangulation $\mathcal{T}_0$, such that $\omega_\ell(z) \subseteq \omega_{\ell-1}^2(z) \subseteq \widehat{\omega}_{\mathrm{level}_\ell(z)}^n(z)$.*

Proof of (i). Let $T \in \mathcal{T}_\ell$ with $T \subseteq \omega_\ell(z)$ such that $\lceil \mathrm{gen}(T)/d\rceil = \mathrm{level}_\ell(z)$ and let $T' \in \mathcal{T}_{\ell-1}$ with $T' \subseteq \omega_{\ell-1}^2(z)$ such that $\lfloor \mathrm{gen}(T')/d\rfloor = R_\ell(z)$. Let $T \subseteq T_0 \in \mathcal{T}_0$ and $T' \subseteq T_0' \in \mathcal{T}_0$ be the corresponding ancestor elements in $\mathcal{T}_0$, respectively. Due to γ-shape regularity of the mesh, there exists a constant $C > 0$ which depends only on the initial triangulation $\mathcal{T}_0$ such that

$$\mathrm{gen}(T) = \frac{\log(|T|/|T_0|)}{\log(1/2)} \leq \frac{\log(C\,|T'|/|T_0'|)}{\log(1/2)} = \mathrm{gen}(T') + \frac{\log(C)}{\log(1/2)}.$$

Therefrom, we get that

$$\begin{aligned}\operatorname{level}_\ell(z) &= \lceil \operatorname{gen}(T)/d \rceil \\ &\le \lfloor \operatorname{gen}(T')/d \rfloor + 1 + \left\lceil \frac{\log(C)}{\log(1/2)} \right\rceil \\ &= R_\ell(z) + \left(1 + \left\lceil \frac{\log(C)}{\log(1/2)} \right\rceil\right).\end{aligned}$$

This concludes the proof with $C_1 := 1 + \lceil \frac{\log(C)}{\log(1/2)} \rceil$. □

Proof of (ii). Let $T \in \mathcal{T}_{\ell-1}$ with $T \subseteq \omega_{\ell-1}^2(z)$. Due to the definition (4.91) of $r_\ell(z)$, it holds that $\operatorname{gen}(T) \ge r_\ell(z) \ge \lfloor r_\ell(z)/d \rfloor = R_\ell(z)$. Since $T \in \widehat{\mathcal{T}}_{\operatorname{gen}(T)}$ and $\operatorname{gen}(T) \ge R_\ell(z)$, there exists an ancestor element $\widehat{T} \in \widehat{\mathcal{T}}_{R_\ell(z)}$ such that $T \subseteq \widehat{T}$. □

Proof of (iii). Since the mesh $\mathcal{T}_\ell$ is a refinement of $\mathcal{T}_{\ell-1}$, it holds that $\omega_\ell(z) \subseteq \omega_{\ell-1}(z) \subseteq \omega_{\ell-1}^2(z)$. Hence, it only remains to prove the second inclusion $\omega_{\ell-1}^2(z) \subseteq \widehat{\omega}^n_{\operatorname{level}_\ell(z)}(z)$. To that end, let $T \in \mathcal{T}_{\ell-1}$ with $T \subseteq \omega_{\ell-1}^2(z)$. Lemma 34(ii) provides an element $\widehat{T} \in \widehat{\mathcal{T}}_{R_\ell(z)}$ such that $T \subseteq \widehat{T}$. Furthermore, it holds that $\widehat{T} \subseteq \widehat{\omega}^2_{R_\ell(z)}(z)$ and hence $T \subseteq \widehat{T} \subseteq \widehat{\omega}^2_{R_\ell(z)}(z)$. The element $\widehat{T}$ can be rewritten with elements of $\widehat{\mathcal{T}}_{R_\ell(z)+C_1}$ the following way. Since the series $\widehat{\mathcal{T}}_m$ is generated by uniform refinement via bisection, the element $\widehat{T}$ gets bisected into $2^{d\,C_1}$ many elements $\widehat{T}'_j \in \widehat{\mathcal{T}}_{R_\ell(z)+C_1}$ such that

$$\widehat{T} = \bigcup_{j=1}^{2^{d\,C_1}} \widehat{T}'_j.$$

Since $\widehat{T} \in \widehat{\omega}^2_{R_\ell(z)}(z)$, there exists $n \in \mathbb{N}$ with $n \le 2^{d\,C_1+1}$ such that $\widehat{T} \subseteq \widehat{\omega}^n_{R_\ell(z)+C_1}$. Lemma 34(i) yields that $\operatorname{level}_\ell(z) \le R_\ell(z) + C_1$ and hence $\widehat{\omega}^n_{R_\ell(z)+C_1}(z) \subseteq \widehat{\omega}^n_{\operatorname{level}_\ell(z)}(z)$. So far, this proves that $T \subseteq \widehat{T} \subseteq \widehat{\omega}^n_{\operatorname{level}_\ell(z)}(z)$, and we conclude that $\omega_{\ell-1}^2(z) \subseteq \widehat{\omega}^n_{\operatorname{level}_\ell(z)}(z)$. □

Scott–Zhang projection

We recall a variant of the Scott–Zhang quasi-interpolation operator, cf. [SZ90] or [BS02, Section 4.8]. For $z \in \mathcal{N}_\ell$, let $T_{\ell,z} \in \mathcal{T}_\ell$ be an element with $z \in T_{\ell,z}$. Let $\psi_{\ell,z}$ denote the (unique) $L^2(T_{\ell,z})$-dual basis function with

$$\int_{T_{\ell,z}} \psi_{\ell,z}(x)\, \eta_{\ell,z'}(x)\, \mathrm{d}x = \delta_{zz'} \quad \text{for all } z' \in \mathcal{N}_\ell,$$

where $\delta_{zz'}$ denotes the Kronecker delta. Defining the Scott–Zhang operator $J_\ell \colon L^2(\Omega) \to \mathcal{S}^1(\mathcal{T}_\ell)$ by

$$J_\ell v := \sum_{z \in \mathcal{N}_\ell} \eta_{\ell,z} \int_{T_{\ell,z}} \psi_{\ell,z}(x)\, v(x)\, \mathrm{d}x \quad \text{for all } v \in L^2(\Omega),$$

we note the following properties, where the constant $C > 0$ depends only on the γ-shape regularity of $\mathcal{T}_\ell$:

- J_ℓ is a linear projection onto $\mathcal{S}^1_0(\mathcal{T}_\ell)$, i.e.,

$$J_\ell v_\ell = v_\ell \quad \text{for all } v_\ell \in \mathcal{S}^1_0(\mathcal{T}_\ell). \tag{4.93}$$

- J_ℓ is locally L^2-stable, i.e., for all $T \in \mathcal{T}_\ell$, it holds that

$$\|v - J_\ell v\|_{L^2(T)} \le C \, \|v\|_{L^2(\omega_\ell(T))} \quad \text{for all } v \in L^2(\Omega).$$

- J_ℓ is locally H^1-stable, i.e., for all $T \in \mathcal{T}_\ell$, it holds that

$$\|\nabla(v - J_\ell v)\|_{L^2(T)} \le C \, \|\nabla v\|_{L^2(\omega_\ell(T))} \quad \text{for all } v \in H^1_0(\Omega).$$

- J_ℓ has a local first-order approximation property

$$\|v - J_\ell v\|_{L^2(T)} \le C \, h_\ell(T) \, \|\nabla v\|_{L^2(\omega_\ell(T))} \quad \text{for all } v \in H^1_0(\Omega).$$

The freedom in the choice of the averaging element $T_{\ell,z}$ can be exploited to ensure additional properties. In our case, the choice of $T_{\ell,z}$ is arbitrary, but we require that $T_{\ell-1,z} = T_{\ell,z} \in \mathcal{T}_\ell \cap \mathcal{T}_{\ell-1}$ for all $z \in \mathcal{N}_\ell \setminus \widetilde{\mathcal{N}}_\ell \subseteq \mathcal{N}_{\ell-1}$. From this choice, it also follows that $\eta_{\ell,z} = \eta_{\ell-1,z}$ and $\psi_{\ell,z} = \psi_{\ell-1,z}$ for all $z \in \mathcal{N}_\ell \setminus \widetilde{\mathcal{N}}_\ell$. Hence, we get that

$$(J_\ell - J_{\ell-1})v(z) = 0 \quad \text{for all } z \in \mathcal{N}_\ell \setminus \widetilde{\mathcal{N}}_\ell,$$

as well as

$$(J_\ell - J_{\ell-1})v \in \operatorname{span}\{\eta_{\ell,z} \colon z \in \widetilde{\mathcal{N}}_\ell\} = \widetilde{\mathcal{X}}_\ell. \tag{4.94}$$

Lemma 35. *For all $v \in L^2(\Omega)$ and $z \in \widetilde{\mathcal{N}}_\ell$, it holds that*

$$\begin{aligned} |(J_\ell - J_{\ell-1})v(z)| &\le |J_\ell v(z)| + |J_{\ell-1} v(z)| \\ &\le C h_\ell(z)^{-d/2} \, \|v\|_{L^2(\omega^2_{\ell-1}(z))}, \end{aligned} \tag{4.95}$$

where $C > 0$ depends only on γ-shape regularity of $\mathcal{T}_\ell$.

Proof. The first inequality in (4.95) follows from the usual triangle inequality. Hence, it only remains to prove the second inequality. [SZ90, Lemma 3.1] states that $\|\psi_{\ell,z}\|_{L^\infty(T_{\ell,z})} \lesssim |T_{\ell,z}|^{-1}$. For $z \in \widetilde{\mathcal{N}}_\ell$, it holds that $T_{\ell,z} \subseteq \omega_\ell(z) \subseteq \omega^2_{\ell-1}(z)$. Thus, the first summand in (4.95) is bounded by

$$\begin{aligned} |J_\ell v(z)| &\le \int_{T_{\ell,z}} |\psi_{\ell,z}(x) v(x)| \, \mathrm{d}x \\ &\le \|\psi_{\ell,z}\|_{L^\infty(T_{\ell,z})} \, |T_{\ell,z}|^{1/2} \, \|v\|_{L^2(T_{\ell,z})} \\ &\lesssim |T_{\ell,z}|^{-1/2} \, \|v\|_{L^2(\omega^2_{\ell-1}(z))} \\ &\simeq h_\ell(z)^{-d/2} \, \|v\|_{L^2(\omega^2_{\ell-1}(z))}. \end{aligned} \tag{4.96}$$

To bound the second summand in (4.95), we must consider two cases: First, let $z \in \widetilde{\mathcal{N}}_\ell \cap \mathcal{N}_{\ell-1}$. It holds that $|T_{\ell,z}| \simeq h_\ell^d(z)$ as well as $T_{\ell-1,z} \subseteq \omega_{\ell-1}(z) \subseteq \omega_{\ell-1}^2(z)$. Similarly to (4.96), we get that

$$\begin{aligned} |J_{\ell-1}v(z)| &\le \int_{T_{\ell-1,z}} |\psi_{\ell-1,z}(x)v(x)| \,\mathrm{d}x \\ &\le \|\psi_{\ell-1,z}\|_{L^\infty(T_{\ell-1,z})} \, |T_{\ell-1,z}|^{1/2} \, \|v\|_{L^2(T_{\ell-1,z})} \\ &\lesssim |T_{\ell-1,z}|^{-1/2} \, \|v\|_{L^2(\omega_{\ell-1}^2(z))} \\ &\lesssim h_\ell(z)^{-d/2} \, \|v\|_{L^2(\omega_{\ell-1}^2(z))}. \end{aligned} \tag{4.97}$$

Second, let $z \in \widetilde{\mathcal{N}}_\ell \backslash \mathcal{N}_{\ell-1}$. Then, due to γ-shape regularity, there exists a uniformly bounded number of nodes $z_1, z_2, \dots, z_{n(z)} \in \mathcal{N}_{\ell-1}$ such that

$$J_{\ell-1}v(z) = \sum_{i=1}^{n(z)} \eta_{\ell-1,z_i}(z) \int_{T_{\ell-1,z_i}} \psi_{\ell-1,z_i}(x) \, v(x) \,\mathrm{d}x.$$

For $i \in \{1, 2, \dots, n(z)\}$, it again holds that $|T_{\ell-1,z_i}| \simeq h_\ell^d(z)$ as well as $T_{\ell-1,z_i} \subseteq \omega_{\ell-1}(z_i) \subseteq \omega_{\ell-1}^2(z)$. With the same arguments as for (4.97), it follows that

$$\begin{aligned} |J_{\ell-1}v(z)| &\le \sum_{i=1}^{n(z)} \int_{T_{\ell-1,z_i}} |\psi_{\ell-1,z_i}(x)v(x)| \,\mathrm{d}x \\ &\lesssim \sum_{i=1}^{n(z)} |T_{\ell-1,z_i}|^{-1/2} \, \|v\|_{L^2(T_{\ell-1,z_i})} \\ &\lesssim h_\ell(z)^{-d/2} \, \|v\|_{L^2(\omega_{\ell-1}^2(z))}. \end{aligned} \tag{4.98}$$

Combining (4.96)–(4.98), we conclude (4.95). □

4.7.3 Additive Schwarz operator

For all $z \in \widetilde{\mathcal{N}}_\ell$, we define the local orthogonal projections $\mathcal{P}_{\ell,z} \colon H_0^1(\Omega) \to \mathcal{X}_{\ell,z} = \operatorname{span}\{\eta_{\ell,z}\}$ by

$$\langle\!\langle \mathcal{P}_{\ell,z} v \,,\, w_{\ell,z} \rangle\!\rangle = \langle\!\langle v \,,\, w_{\ell,z} \rangle\!\rangle \quad \text{for all } w_{\ell,z} \in \mathcal{X}_{\ell,z}$$

with the explicit representation

$$\mathcal{P}_{\ell,z} v = \frac{\langle\!\langle v \,,\, \eta_{\ell,z} \rangle\!\rangle}{|||\eta_{\ell,z}|||^2} \eta_{\ell,z} \quad \text{for all } v \in H_0^1(\Omega). \tag{4.99}$$

Based on these projections, we define the additive Schwarz operator as

$$\mathcal{Q}_L := \sum_{\ell=0}^{L} \sum_{z \in \widetilde{\mathcal{N}}_\ell} \mathcal{P}_{\ell,z} \colon H_0^1(\Omega) \to \mathcal{X}_L. \tag{4.100}$$

Therefore, the multilevel diagonal scaling is a multilevel additive Schwarz method and we can use the abstract analysis of these methods.

The key result reads as follows.

Proposition 36. *The operator $\mathcal{Q}_L$ is linear and bounded as well as symmetric*

$$\langle\!\langle \mathcal{Q}_L v \,,\, w \rangle\!\rangle = \langle\!\langle v \,,\, \mathcal{Q}_L w \rangle\!\rangle \quad \text{for all } v, w \in H^1_0(\Omega) \tag{4.101}$$

and satisfies

$$c\, |\!|\!| v |\!|\!|^2 \leq \langle\!\langle \mathcal{Q}_L v \,,\, v \rangle\!\rangle \leq C\, |\!|\!| v |\!|\!|^2 \quad \text{for all } v \in \mathcal{X}_L. \tag{4.102}$$

The constants $c, C > 0$ depend only on Ω, the initial triangulation $\mathcal{T}_0$, and the diffusion coefficient $\boldsymbol{A}$.

While linearity, boundedness, and symmetry of additive Schwarz operators are well-known (cf. [GO94, Lemma 2]), we will provide the proof of (4.102) in Section 4.7.5 as well as Section 4.7.6.

4.7.4 Proof of Theorem 30 (optimal condition number)

Let $v := \sum_{j=0}^{N_L} \mathbf{x}_j \eta_{L,z_j} \in \mathcal{X}_L$ and $w := \sum_{j=0}^{N_L} \mathbf{y}_j \eta_{L,z_j} \in \mathcal{X}_L$. From the definition (4.76) of the local multilevel diagonal preconditioner, it follows that $\mathbf{M}_L \mathbf{P}_L \mathbf{M}_L$ is symmetric. We define the additive Schwarz matrix $\mathbf{Q}_L := \mathbf{P}_L \mathbf{M}_L$. It then holds that

$$\langle\!\langle \mathcal{Q}_L v \,,\, w \rangle\!\rangle = \langle \mathbf{Q}_L \mathbf{x} \,,\, \mathbf{y} \rangle_{\mathbf{M}_L}. \tag{4.103}$$

Combining the identity (4.103) with (4.102), we see that

$$c\, \langle \mathbf{x} \,,\, \mathbf{x} \rangle_{\mathbf{M}_L} = c\, |\!|\!| v |\!|\!|^2 \leq \langle\!\langle \mathcal{Q}_L v \,,\, v \rangle\!\rangle = \langle \mathbf{Q}_L \mathbf{x} \,,\, \mathbf{x} \rangle_{\mathbf{M}_L}$$

as well as

$$\langle \mathbf{Q}_L \mathbf{x} \,,\, \mathbf{x} \rangle_{\mathbf{M}_L} = \langle\!\langle \mathcal{Q}_L v \,,\, v \rangle\!\rangle \leq C\, |\!|\!| v |\!|\!|^2 = C\, \langle \mathbf{x} \,,\, \mathbf{x} \rangle_{\mathbf{M}_L}.$$

Due to the symmetry (4.101) and again the identity (4.103), we get that

$$\langle \mathbf{Q}_L \mathbf{x} \,,\, \mathbf{y} \rangle_{\mathbf{M}_L} = \langle\!\langle \mathcal{Q}_L v \,,\, w \rangle\!\rangle = \langle\!\langle v \,,\, \mathcal{Q}_L w \rangle\!\rangle = \langle \mathbf{x} \,,\, \mathbf{Q}_L \mathbf{y} \rangle_{\mathbf{M}_L},$$

i.e., $\mathbf{Q}_L$ is symmetric with respect to $\langle \cdot \,,\, \cdot \rangle_{\mathbf{M}_L}$. Now, [TW05, Lemma C.1] or [GVL13, Section 8.1] yield the Rayleigh quotient estimates

$$\lambda_{\min}(\mathbf{Q}_L) = \min_{\substack{\mathbf{x} \in \mathbb{R}^{N_L} \\ \mathbf{x} \neq 0}} \frac{\langle \mathbf{Q}_L \mathbf{x} \,,\, \mathbf{x} \rangle_{\mathbf{M}_L}}{\langle \mathbf{x} \,,\, \mathbf{x} \rangle_{\mathbf{M}_L}} \geq c,$$

and

$$\lambda_{\max}(\mathbf{Q}_L) = \max_{\substack{\mathbf{x} \in \mathbb{R}^{N_L} \\ \mathbf{x} \neq 0}} \frac{\langle \mathbf{Q}_L \mathbf{x} \,,\, \mathbf{x} \rangle_{\mathbf{M}_L}}{\langle \mathbf{x} \,,\, \mathbf{x} \rangle_{\mathbf{M}_L}} \leq C.$$

In particular, it follows that

$$\operatorname{cond}_{\mathbf{M}_L}(\mathbf{Q}_L) = \frac{\lambda_{\max}(\mathbf{Q}_L)}{\lambda_{\min}(\mathbf{Q}_L)} \leq \frac{C}{c}.$$

This concludes the proof. □

Lions' lemma

The last lemma we need for the proof of the lower bound in (4.102) is known as Lions's lemma, cf. [Lio88, Wid89] and [TW05, Lemma 2.5].

Lemma 37 (Lions). *Let $m \in \mathbb{N}_0$ and $v \in V$, where V is a finite-dimensional Hilbert space with scalar product $\langle\!\langle \cdot\,, \cdot \rangle\!\rangle$ and corresponding norm $|\!|\!| \cdot |\!|\!|$. Assume that there exists a decomposition of V into spaces V_ℓ with $0 \le \ell \le m$ such that $V = \sum_{\ell=0}^m V_\ell$ and orthogonal projections $\mathcal{P}_\ell \colon V \to V_\ell$ defined by*

$$\langle\!\langle \mathcal{P}_\ell\, v\,,\, w_\ell \rangle\!\rangle = \langle\!\langle v\,,\, w_\ell \rangle\!\rangle \quad \text{for all } w_\ell \in V_\ell.$$

Define $\mathcal{P}_{\mathrm{AS}} := \sum_{\ell=0}^m \mathcal{P}_\ell$. If there exists a constant $C > 0$ such that every $v \in V$ admits a decomposition $v = \sum_{\ell=0}^m v_\ell$ with $v_\ell \in V_\ell$ that satisfies

$$\sum_{\ell=0}^m |\!|\!| v_\ell |\!|\!|^2 \le C\, |\!|\!| v |\!|\!|^2,$$

then it holds that

$$|\!|\!| v |\!|\!|^2 \le C\, \langle\!\langle \mathcal{P}_{\mathrm{AS}}\, v\,,\, v \rangle\!\rangle$$

for all $v \in V$. □

4.7.5 Proof of lower bound in Proposition 36

The proof is split into 5 steps.

Step 1. With property (4.94) of the Scott–Zhang projection J_ℓ, we define the difference

$$\widetilde{v}_\ell := (J_\ell - J_{\ell-1})v \in \widetilde{\mathcal{X}}_\ell \quad \text{for } v \in \mathcal{X}_L \text{ and } 0 \le \ell \le L, \tag{4.104a}$$

where $J_{-1} := 0$. Henceforth, we can rewrite any $v \in \mathcal{X}_L$ using the projection property (4.93) of J_L as a telescoping series as follows

$$v = J_L v = (J_L - J_{-1})v = \sum_{\ell=0}^L \widetilde{v}_\ell. \tag{4.104b}$$

Using the basis representation of $\widetilde{v}_\ell$, we can decompose this further into

$$v = \sum_{\ell=0}^L \sum_{z \in \widetilde{\mathcal{N}}_\ell} \widetilde{v}_\ell(z)\eta_{\ell,z} =: \sum_{\ell=0}^L \sum_{z \in \widetilde{\mathcal{N}}_\ell} v_{\ell,z} \quad \text{with } v_{\ell,z} \in \mathcal{X}_{\ell,z}. \tag{4.104c}$$

Step 2. Let $z \in \mathcal{N}_\ell$. Then, there holds the inverse inequality

$$\|\nabla \eta_{\ell,z}\|_{L^2(\omega_\ell(z))} \lesssim h_\ell(z)^{-1}\, \|\eta_{\ell,z}\|_{L^2(\omega_\ell(z))}$$

which follows from a scaling argument with the hidden constant depending only on γ-shape regularity of $\mathcal{T}_\ell$. Combining this inequality with the equivalence (4.83), it holds that

$$|\!|\!| \eta_{\ell,z} |\!|\!|^2 \lesssim \|\nabla \eta_{\ell,z}\|^2_{L^2(\omega_\ell(z))} \lesssim h_\ell(z)^{-2}\, \|\eta_{\ell,z}\|^2_{L^2(\omega_\ell(z))} \le h_\ell(z)^{-2}\, |\omega_\ell(z)| \simeq h_\ell(z)^{d-2}.$$

Hence, we get that

$$\|\!|v_{\ell,z}|\!\|^2 = \|\!|\eta_{\ell,z}|\!\|^2 \, |\widetilde{v}_\ell(z)|^2 \lesssim h_\ell(z)^{d-2} \, |(J_\ell - J_{\ell-1})v(z)|^2. \tag{4.105}$$

Step 3. Let $\widehat{\Pi}_m := \widehat{\Pi}_0$ for $m < 0$. From Lemma 34(i), we get that

$$\widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v \in \widehat{\mathcal{X}}_{R_\ell(z)}$$

and especially that $(\widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v)|_T$ is affine on all $T \in \mathcal{T}_{\ell-1}$ with $T \subseteq \omega^2_{\ell-1}(z)$ as well as continuous on the whole patch $\omega^2_{\ell-1}(z)$. In particular, the same holds also on the patch $\omega^2_\ell(z)$. Therefore, the projection property (4.93) of the Scott–Zhang operator yields that

$$(J_\ell \widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v)(z) = (\widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v)(z) = (J_{\ell-1} \widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v)(z).$$

Together with Lemma 35, this yields that

$$\begin{aligned} |(J_\ell - J_{\ell-1})v(z)|^2 &= \big|\big(J_\ell - J_{\ell-1}\big)\big(v - \widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v\big)(z)\big|^2 \\ &\lesssim h_\ell(z)^{-d} \, \|v - \widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v\|^2_{L^2(\omega^2_{\ell-1}(z))}. \end{aligned} \tag{4.106}$$

Step 4. Combining Step 2 and Step 3, we see that

$$\|\!|v_{\ell,z}|\!\|^2 \lesssim h_\ell(z)^{-2} \, \|v - \widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v\|^2_{L^2(\omega^2_{\ell-1}(z))}. \tag{4.107}$$

Using the equivalence $h_\ell(z) \simeq \widehat{h}_{\mathrm{level}_\ell(z)}$ from (4.83), we get that

$$\begin{aligned} \sum_{\ell=0}^{L} \sum_{z \in \widetilde{\mathcal{N}}_\ell} \|\!|v_{\ell,z}|\!\|^2 &\overset{(4.107)}{\lesssim} \sum_{\ell=0}^{L} \sum_{z \in \widetilde{\mathcal{N}}_\ell} h_\ell(z)^{-2} \, \|v - \widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v\|^2_{L^2(\omega^2_{\ell-1}(z))} \\ &\overset{(4.83)}{\simeq} \sum_{\ell=0}^{L} \sum_{z \in \widetilde{\mathcal{N}}_\ell} \widehat{h}^{-2}_{\mathrm{level}_\ell(z)} \, \|v - \widehat{\Pi}_{\mathrm{level}_\ell(z)-C_1} v\|^2_{L^2(\omega^2_{\ell-1}(z))} \\ &= \sum_{m=0}^{\infty} \sum_{\ell=0}^{L} \sum_{\substack{z \in \widetilde{\mathcal{N}}_\ell \\ \mathrm{level}_\ell(z)=m}} \widehat{h}^{-2}_m \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\omega^2_{\ell-1}(z))}. \end{aligned}$$

Combining Lemma 34(iii) with the definition (4.80) of $\widetilde{\mathcal{K}}_m(z)$, we see that

$$\begin{aligned} &\sum_{m=0}^{\infty} \sum_{\ell=0}^{L} \sum_{\substack{z \in \widetilde{\mathcal{N}}_\ell \\ \mathrm{level}_\ell(z)=m}} \widehat{h}^{-2}_m \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\omega^2_{\ell-1}(z))} \\ &\qquad \le \sum_{m=0}^{\infty} \sum_{\ell=0}^{L} \sum_{\substack{z \in \widetilde{\mathcal{N}}_\ell \\ \mathrm{level}_\ell(z)=m}} \widehat{h}^{-2}_m \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\widehat{\omega}^n_m(z))} \\ &\qquad \overset{(4.80)}{=} \sum_{m=0}^{\infty} \sum_{z \in \mathcal{N}_L} \sum_{\ell \in \widetilde{\mathcal{K}}_m(z)} \widehat{h}^{-2}_m \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\widehat{\omega}^n_m(z))}. \end{aligned}$$

Lemma 32 states that $z \in \mathcal{N}_\ell$ with $\mathrm{level}_\ell(z) = m$ is also an element of $\widehat{\mathcal{N}}_m$. Together with the boundedness (4.81) of $\#\widetilde{\mathcal{K}}_m(z)$ from Lemma 31, this yields that

$$\begin{aligned}
&\sum_{m=0}^{\infty} \sum_{z\in\mathcal{N}_L} \sum_{\ell\in\widetilde{\mathcal{K}}_m(z)} \widehat{h}_m^{-2} \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\widehat{\omega}_m^n(z))} \\
&\qquad = \sum_{m=0}^{\infty} \sum_{z\in\mathcal{N}_L\cap\widehat{\mathcal{N}}_m} \sum_{\ell\in\widetilde{\mathcal{K}}_m(z)} \widehat{h}_m^{-2} \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\widehat{\omega}_m^n(z))} \\
&\qquad \overset{(4.81)}{\lesssim} \sum_{m=0}^{\infty} \sum_{z\in\mathcal{N}_L\cap\widehat{\mathcal{N}}_m} \widehat{h}_m^{-2} \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\widehat{\omega}_m^n(z))} \\
&\qquad \le \sum_{m=0}^{\infty} \sum_{z\in\widehat{\mathcal{N}}_m} \widehat{h}_m^{-2} \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\widehat{\omega}_m^n(z))}.
\end{aligned}$$

Due to uniform γ-shape regularity of $\widehat{\mathcal{T}}_m$ and the definition $\widehat{\Pi}_m = \widehat{\Pi}_0$ for $m < 0$, it follows that

$$\begin{aligned}
\sum_{m=0}^{\infty} \sum_{z\in\widehat{\mathcal{N}}_m} \widehat{h}_m^{-2} \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\widehat{\omega}_m^n(z))} &\lesssim \sum_{m=0}^{\infty} \widehat{h}_m^{-2} \, \|v - \widehat{\Pi}_{m-C_1} v\|^2_{L^2(\Omega)} \\
&\lesssim \sum_{m=0}^{\infty} \widehat{h}_m^{-2} \, \|v - \widehat{\Pi}_m v\|^2_{L^2(\Omega)}.
\end{aligned}$$

Combining the last four estimates, we end up with

$$\sum_{\ell=0}^{L} \sum_{z\in\widetilde{\mathcal{N}}_\ell} |\!|\!| v_{\ell,z} |\!|\!|^2 \lesssim \sum_{m=0}^{\infty} \widehat{h}_m^{-2} \, \|v - \widehat{\Pi}_m v\|^2_{L^2(\Omega)}. \tag{4.108}$$

Step 5: Finally, Step 4 together with Lemma 33 and norm equivalence yields that

$$\sum_{\ell=0}^{L} \sum_{z\in\widetilde{\mathcal{N}}_\ell} |\!|\!| v_{\ell,z} |\!|\!|^2 \overset{(4.108)}{\lesssim} \sum_{m=0}^{\infty} \widehat{h}_m^{-2} \|v - \widehat{\Pi}_m v\|^2_{L^2(\Omega)} \overset{(4.104)}{\lesssim} \|v\|^2_{H^1(\Omega)} \simeq |\!|\!| v |\!|\!|^2, \tag{4.109}$$

for all $v \in \mathcal{X}_L$ and the decomposition $v = \sum_{\ell=0}^{L} \sum_{z\in\widetilde{\mathcal{N}}_\ell} v_{\ell,z}$ from (4.104c). Due to Lions's lemma (cf. Lemma 37) this guarantees the ellipticity of the additive Schwarz operator $\mathcal{Q}_L$ from (4.100).

$$|\!|\!| v |\!|\!|^2 \lesssim \langle\!\langle \mathcal{Q}_L v \,,\, v \rangle\!\rangle \quad \text{for all } v \in \mathcal{X}^L,$$

which concludes the proof of the lower bound in (4.102). □

Auxiliary results

We define the maximal level $M := \max_{z\in\mathcal{N}_L s} \operatorname{level}_L(z)$ of all nodes $z \in \mathcal{N}_L$. From Lemma 32, it follows that $\mathcal{N}_L \subseteq \widehat{\mathcal{N}}_M$ and $\mathcal{X}_L \subseteq \widehat{\mathcal{X}}_M$. We rewrite the additive Schwarz operator $\mathcal{Q}_L$ as

$$\mathcal{Q}_L = \sum_{\ell=0}^{L} \sum_{z\in\widetilde{\mathcal{N}}_\ell} \mathcal{P}_{\ell,z} = \sum_{m=0}^{M} \mathcal{Q}_{L,m} \quad \text{with} \quad \mathcal{Q}_{L,m} := \sum_{\ell=0}^{L} \sum_{\substack{z\in\widetilde{\mathcal{N}}_\ell \\ \operatorname{level}_\ell(z)=m}} \mathcal{P}_{\ell,z}. \tag{4.110}$$

Then, there holds the following lemma, which is used to prove the strenghtened Cauchy–Schwarz inequality (4.118).

Lemma 38. *Let $0 \le k \le m \le M$ and $0 \le \ell \le L$. For $T \in \widehat{\mathcal{T}}_k$, $\widehat{v}_k \in \widehat{\mathcal{X}}_k$, and $z \in \widetilde{\mathcal{N}}_\ell$ with $\operatorname{level}_\ell(z) = m$, it holds that*

$$\int_T \mathbf{A}\nabla\widehat{v}_k \cdot \nabla\eta_{\ell,z}\,\mathrm{d}x \le C\left(2^{-1/2}\right)^{m-k} \widehat{h}_m^{-1} \|\nabla\widehat{v}_k\|_{L^2(T)} \|\eta_{\ell,z}\|_{L^2(T)}, \tag{4.111}$$

where the constant $C > 0$ depends only on the initial triangulation $\mathcal{T}_0$, and $\|\boldsymbol{A}\|_\infty$.

Proof. From Lemma 32, we know that $\eta_{\ell,z} \in \widehat{\mathcal{X}}_m$. Hence, we can decompose $\eta_{\ell,z}$ as follows. We define $\widehat{v}_{m,0} \in \widehat{\mathcal{X}}_m$ such that $\widehat{v}_{m,0}$ vanishes on ∂T and is equal to $\eta_{\ell,z}$ at the interior nodes in T. Let $\widehat{v}_{m,1} = \eta_{\ell,z} - \widehat{v}_{m,0}$. Then, it holds that

$$\int_T \boldsymbol{A}\nabla\widehat{v}_k \cdot \nabla\eta_{\ell,z}\,\mathrm{d}x = \int_T \boldsymbol{A}\nabla\widehat{v}_k \cdot \nabla\widehat{v}_{m,0}\,\mathrm{d}x + \int_T \boldsymbol{A}\nabla\widehat{v}_k \cdot \nabla\widehat{v}_{m,1}\,\mathrm{d}x. \tag{4.112}$$

Note that $\nabla\widehat{v}_k|_T$ is constant, since $T \in \widehat{\mathcal{T}}_k$. Moreover, note that $\widehat{v}_{m,0}|_{\partial T} = 0$. With integration by parts and $\nabla\widehat{v}_k \in \mathcal{P}^0(T)$, we get for the first summand of (4.112) that

$$\begin{aligned} \int_T \boldsymbol{A}\nabla\widehat{v}_k \cdot \nabla\widehat{v}_{m,0}\,\mathrm{d}x &= -\int_T \operatorname{div}\left(\boldsymbol{A}\nabla\widehat{v}_k\right)\widehat{v}_{m,0}\,\mathrm{d}x \\ &= -\int_T \left((\operatorname{div}\boldsymbol{A})\nabla\widehat{v}_k\right)\widehat{v}_{m,0}\,\mathrm{d}x. \end{aligned} \tag{4.113}$$

From the Cauchy–Schwarz inequality combined with $1 \lesssim (2^{-(m-k)})^{1/2}\,\widehat{h}_m^{-1}$, we estimate the latter term as follows

$$\begin{aligned} -\int_T \left((\operatorname{div}\boldsymbol{A})\nabla\widehat{v}_k\right)\widehat{v}_{m,0}\,\mathrm{d}x &\lesssim \|\nabla\widehat{v}_k\|_{L^2(T)}\|\widehat{v}_{m,0}\|_{L^2(T)} \\ &\lesssim \left(2^{-(m-k)}\right)^{1/2}\widehat{h}_m^{-1}\,\|\nabla\widehat{v}_k\|_{L^2(T)}\|\eta_{\ell,z}\|_{L^2(T)}. \end{aligned} \tag{4.114}$$

Hence, it only remains to estimate the second summand of (4.112). We define $T_m := \bigcup\{K \in \widehat{\mathcal{T}}_m : K \cap \partial T \neq \emptyset\}$, cf. Figure 4.1. It then holds that $\operatorname{supp}\widehat{v}_{m,1} \subseteq T_m$ and $|T_m| \simeq \widehat{h}_k^{d-1}\widehat{h}_m$. Again, using the Cauchy–Schwarz inequality, we see that

$$\begin{aligned} \int_T \boldsymbol{A}\nabla\widehat{v}_k \cdot \nabla\widehat{v}_{m,1}\,\mathrm{d}x &= \int_{T_m} \boldsymbol{A}\nabla\widehat{v}_k \cdot \nabla\widehat{v}_{m,1}\,\mathrm{d}x \\ &\lesssim \|\nabla\widehat{v}_k\|_{L^2(T_m)}\|\nabla\widehat{v}_{m,1}\|_{L^2(T_m)}. \end{aligned} \tag{4.115}$$

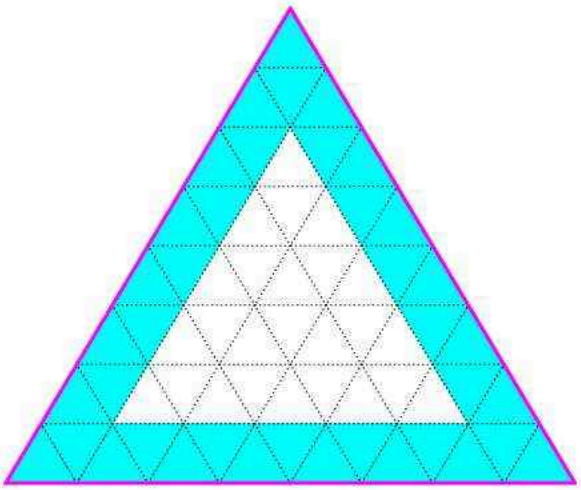


Figure 4.1: Illustration of the set $T_m := \bigcup \{K \in \widehat{\mathcal{T}}_m : K \cap \partial T \neq \emptyset\}$ from the proof of Lemma 38: The outer triangle (solid lines, pink) represents the element $T \in \widehat{\mathcal{T}}_k$, while the inner triangles (dashed lines) correspond to all elements $K \in \widehat{\mathcal{T}}_m$ such that $T \subseteq K$. Then, the set T_m is the outer cyan area.

Since $\widehat{v}_k \in \widehat{\mathcal{X}}_k$, we know that $\nabla \widehat{v}_k$ is constant on K. This yields that

$$\begin{aligned} \|\nabla \widehat{v}_k\|_{L^2(T_m)} &= \frac{|T_m|^{1/2}}{|T|^{1/2}} \|\nabla \widehat{v}_k\|_{L^2(T)} \\ &\overset{(4.83)}{\simeq} \Big(\frac{\widehat{h}_k^{d-1} \widehat{h}_m}{\widehat{h}_k^d}\Big)^{1/2} \|\nabla \widehat{v}_k\|_{L^2(T)} \\ &\overset{(4.82)}{=} \big(2^{-(m-k)}\big)^{1/2} \|\nabla \widehat{v}_k\|_{L^2(T)}. \end{aligned} \tag{4.116}$$

The remaining term $\|\nabla \widehat{v}_{m,1}\|_{L^2(T_m)}$ is estimated by an inverse estimate

$$\|\nabla \widehat{v}_{m,1}\|_{L^2(T_m)} \lesssim \widehat{h}_m^{-1} \|\widehat{v}_{m,1}\|_{L^2(T_m)} \leq h_m^{-1} \|\eta_{\ell,z}\|_{L^2(T)}. \tag{4.117}$$

Combining (4.112)–(4.117), we finally get that

$$\int_T \boldsymbol{A} \nabla \widehat{v}_k \cdot \nabla \eta_{\ell,z} \, \mathrm{d}x \lesssim \big(2^{-(m-k)}\big)^{1/2} \widehat{h}_m^{-1} \, \|\nabla \widehat{v}_k\|_{L^2(T)} \|\eta_{\ell,z}\|_{L^2(T)}.$$

This concludes the proof.

□

Now, we are able to prove the following strenghtened Cauchy–Schwarz inequality.

Lemma 39. *For all $0 \leq k \leq m \leq M$, it holds that*

$$\langle\!\langle \widehat{v}_k \,,\, \mathcal{Q}_{L,m} \widehat{w}_k \rangle\!\rangle \leq C \, \big(2^{-1/2}\big)^{m-k} \, |\!|\!| \widehat{v}_k |\!|\!| \, |\!|\!| \widehat{w}_k |\!|\!| \quad \text{for all } \widehat{v}_k, \widehat{w}_k \in \widehat{\mathcal{X}}_k, \tag{4.118}$$

where $C > 0$ depends only on Ω, the initial triangulation $\mathcal{T}_0$, $\|\boldsymbol{A}\|_\infty$, and γ-shape regularity.

Proof. The proof is split into three steps.
Step 1: Define $q := 2^{-1/2}$ and let $z \in \widetilde{\mathcal{N}}_\ell$ with $0 \le k \le m = \operatorname{level}_\ell(z)$. Then, Lemma 38, the Cauchy–Schwarz inequality, and the Friedrichs inequality yield that

$$\begin{aligned}
\langle\!\langle \widehat{v}_k \,,\, \eta_{\ell,z} \rangle\!\rangle &= \sum_{K\in\widehat{\mathcal{T}}_k} \int_K \mathbf{A}\nabla\widehat{v}_k \cdot \nabla\eta_{\ell,z}\,\mathrm{d}x \\
&\overset{(4.111)}{\lesssim} q^{m-k}\,\widehat{h}_m^{-1} \sum_{K\in\widehat{\mathcal{T}}_k} \|\nabla\widehat{v}_k\|_{L^2(K)} \|\eta_{\ell,z}\|_{L^2(K)} \\
&\le q^{m-k}\,\widehat{h}_m^{-1}\,\|\nabla\widehat{v}_k\|_{L^2(\Omega)}\,\|\eta_{\ell,z}\|_{L^2(\Omega)} \\
&\simeq q^{m-k}\,\widehat{h}_m^{-1}\,|||\widehat{v}_k|||\,\|\eta_{\ell,z}\|_{L^2(\omega_\ell(z))} \\
&\lesssim q^{m-k}\,\widehat{h}_m^{-1}\,|||\widehat{v}_k|||\,\operatorname{diam}(\omega_\ell(z))\,\|\nabla\eta_{\ell,z}\|_{L^2(\omega_\ell(z))} \\
&\simeq q^{m-k}\,\widehat{h}_m^{-1}\,|||\widehat{v}_k|||\,\widehat{h}_m\,|||\eta_{\ell,z}||| \\
&= q^{m-k}\,|||\widehat{v}_k|||\,|||\eta_{\ell,z}|||.
\end{aligned}$$

Summing up, we have that

$$\boxed{\langle\!\langle \widehat{v}_k \,,\, \eta_{\ell,z} \rangle\!\rangle \lesssim q^{m-k}\,|||\widehat{v}_k|||\,|||\eta_{\ell,z}||| \quad \text{for all } z \in \widetilde{\mathcal{N}}_\ell \text{ with } k \le m = \operatorname{level}_\ell(z)}, \tag{4.119}$$

where the hidden constant depends only on $\mathcal{T}_0$ and $\boldsymbol{A}$.
Step 2: Next, we show that

$$\boxed{\sum_{\ell=0}^{L} \sum_{\substack{z\in\widetilde{\mathcal{N}}_\ell \\ \operatorname{level}_\ell(z)=m}} |||\mathcal{P}_{\ell,z}\widehat{w}_k||| \lesssim |||\widehat{w}_k|||,} \tag{4.120}$$

where the hidden constant depends only on $\mathcal{T}_0$ and γ-shape regularity. The representation (4.99), the Cauchy–Schwarz inequality, and Lemma 34(iii) yield that

$$\begin{aligned}
|||\mathcal{P}_{\ell,z}\widehat{w}_k||| &\overset{(4.99)}{=} \frac{|\langle\!\langle \widehat{w}_k \,,\, \eta_{\ell,z} \rangle\!\rangle|}{|||\eta_{\ell,z}|||} \\
&\le |||\widehat{w}_k|||_{\omega_\ell(z)} \\
&\overset{\text{Lemma 34(iii)}}{\le} |||\widehat{w}_k|||_{\widehat{\omega}_m^n(z)}.
\end{aligned}$$

Recall the set $\widetilde{\mathcal{K}}_k(z)$ from (4.80)

$$\widetilde{\mathcal{K}}_k(z) = \big\{\ell \in \{0,1,\dots,L\} \,:\, z \in \widetilde{\mathcal{N}}_\ell \text{ and } \operatorname{level}_\ell(z) = k\big\}.$$

From Lemma 31, we know that $\sup_{k\in\mathbb{N}_0} \#\widetilde{\mathcal{K}}_k(z) \le C(\mathcal{T}_0) < \infty$ for all $z \in \widetilde{\mathcal{N}}_\ell$ with a constant $C(\mathcal{T}_0) > 0$ depending only on the initial mesh $\mathcal{T}_0$. Hence, from the last inequality and shape

regularity of the mesh $\widehat{\mathcal{T}}_m$, it follows that

$$\begin{aligned}\sum_{\ell=0}^{L} \sum_{\substack{z\in\widetilde{\mathcal{N}}_\ell\\ \mathrm{level}_\ell(z)=m}} |||\mathcal{P}_{\ell,z}\widehat{w}_k||| &= \sum_{z\in\mathcal{N}_L\cap\widehat{\mathcal{N}}_m} \sum_{\ell\in\widetilde{\mathcal{K}}_m(z)} |||\mathcal{P}_{\ell,z}\widehat{w}_k||| \\ &\le \sum_{z\in\mathcal{N}_L\cap\widehat{\mathcal{N}}_m} \sum_{\ell\in\widetilde{\mathcal{K}}_m(z)} |||\widehat{w}_k|||_{\widehat{\omega}_m^n(z)} \\ &\overset{(4.81)}{\lesssim} \sum_{z\in\widehat{\mathcal{N}}_m} |||\widehat{w}_k|||_{\widehat{\omega}_m^n(z)} \\ &\simeq |||\widehat{w}_k|||.\end{aligned}$$

Step 3: Since $\mathcal{P}_{\ell,z}\widehat{w}_k \in \mathcal{X}_{\ell,z} = \mathrm{span}\{\eta_{\ell,z}\}$, there exists $\lambda_{\ell,z}\in\mathbb{R}$ such that $\mathcal{P}_{\ell,z}\widehat{w}_k = \lambda_{\ell,z}\eta_{\ell,z}$. Based on the previous steps, the definition of $\mathcal{Q}_{L,m}$ shows that

$$\begin{aligned}\langle\!\langle \widehat{v}_k\,,\, \mathcal{Q}_{L,m}\widehat{w}_k\rangle\!\rangle &\overset{(4.110)}{=} \sum_{\ell=0}^{L} \sum_{\substack{z\in\widetilde{\mathcal{N}}_\ell\\ \mathrm{level}_\ell(z)=m}} \langle\!\langle \widehat{v}_k\,,\, \mathcal{P}_{\ell,z}\widehat{w}_k\rangle\!\rangle \\ &= \sum_{\ell=0}^{L} \sum_{\substack{z\in\widetilde{\mathcal{N}}_\ell\\ \mathrm{level}_\ell(z)=m}} |\lambda_{\ell,z}|\,\langle\!\langle \widehat{v}_k\,,\, \eta_{\ell,z}\rangle\!\rangle \\ &\overset{(4.119)}{\lesssim} q^{m-k}\,|||\widehat{v}_k|||\sum_{\ell=0}^{L} \sum_{\substack{z\in\widetilde{\mathcal{N}}_\ell\\ \mathrm{level}_\ell(z)=m}} |\lambda_{\ell,z}|\,|||\eta_{\ell,z}||| \\ &= q^{m-k}\,|||\widehat{v}_k|||\sum_{\ell=0}^{L} \sum_{\substack{z\in\widetilde{\mathcal{N}}_\ell\\ \mathrm{level}_\ell(z)=m}} |||\mathcal{P}_{\ell,z}\widehat{w}_k||| \\ &\overset{(4.120)}{\lesssim} q^{m-k}\,|||\widehat{v}_k|||\,|||\widehat{w}_k|||.\end{aligned}$$

This concludes the proof. □

Remark 40. *Due to the self-adjointness of the orthogonal projections $\mathcal{P}_{\ell,z}$, we know that $\langle\!\langle \mathcal{Q}_{L,m}\cdot\,,\,\cdot\rangle\!\rangle$ is a symmetric bilinear form on $\widehat{\mathcal{X}}_k$ for $k\le m$. By definition (4.110) of $\mathcal{Q}_{L,m}$, it holds that*

$$\langle\!\langle \mathcal{Q}_{L,m}v\,,\,v\rangle\!\rangle = \sum_{\ell=0}^{L} \sum_{\substack{z\in\widetilde{\mathcal{N}}_\ell\\ \mathrm{level}_\ell(z)=m}} \langle\!\langle \mathcal{P}_{\ell,z}v\,,\,v\rangle\!\rangle = \sum_{\ell=0}^{L} \sum_{\substack{z\in\widetilde{\mathcal{N}}_\ell\\ \mathrm{level}_\ell(z)=m}} |||\mathcal{P}_{\ell,z}v|||^2 \ge 0 \quad \text{for all } v\in\widehat{\mathcal{X}}_k.$$

Hence, $\langle\!\langle \mathcal{Q}_m^L\cdot\,,\,\cdot\rangle\!\rangle$ is even positive semi-definite. As a consequence, there holds the Cauchy–Schwarz inequality

$$\langle\!\langle \mathcal{Q}_{L,m}v\,,\,w\rangle\!\rangle \le \langle\!\langle \mathcal{Q}_{L,m}v\,,\,v\rangle\!\rangle^{1/2}\,\langle\!\langle \mathcal{Q}_{L,m}w\,,\,w\rangle\!\rangle^{1/2} \quad \text{for all } v,w\in\widehat{\mathcal{X}}_k. \tag{4.121}$$

4.7.6 Proof of upper bound in Proposition 36

First, we define the Galerkin projection $\widehat{\mathcal{G}}_m\colon H^1(\Omega) \to \widehat{\mathcal{X}}_m$ with respect to the scalar product $\langle\!\langle \cdot\,,\,\cdot \rangle\!\rangle$ via

$$\langle\!\langle \widehat{\mathcal{G}}_m v\,,\,\widehat{w}_m \rangle\!\rangle = \langle\!\langle v\,,\,\widehat{w}_m \rangle\!\rangle \quad \text{for all } \widehat{w}_m \in \widehat{\mathcal{X}}_m.$$

With $\widehat{\mathcal{G}}_{-1} := 0$, we can rewrite $\widehat{\mathcal{G}}_m v$ as a telescoping sum, i.e., $\widehat{\mathcal{G}}_m = \sum_{k=0}^m (\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1})$. Let $v \in \mathcal{X}_L \subseteq \widehat{\mathcal{X}}_M$. It holds that $\widehat{\mathcal{G}}_M v = v$.

Since $\mathcal{Q}_{L,m} v \in \widehat{\mathcal{X}}_M$, cf. Lemma 32, the representation (4.110), the symmetry of $\langle\!\langle \cdot\,,\,\cdot \rangle\!\rangle$, and the Cauchy–Schwarz inequality (4.121) yield that

$$\begin{aligned}
\langle\!\langle \mathcal{Q}_L v\,,\,v \rangle\!\rangle &= \sum_{m=0}^M \langle\!\langle \mathcal{Q}_{L,m} v\,,\,v \rangle\!\rangle \\
&= \sum_{m=0}^M \langle\!\langle \mathcal{Q}_{L,m} v\,,\,\widehat{\mathcal{G}}_m v \rangle\!\rangle \\
&= \sum_{m=0}^M \sum_{k=0}^m \langle\!\langle \mathcal{Q}_{L,m} v\,,\,(\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \rangle\!\rangle \\
&\le \sum_{m=0}^M \sum_{k=0}^m \langle\!\langle \mathcal{Q}_{L,m} v\,,\,v \rangle\!\rangle^{1/2} \langle\!\langle \mathcal{Q}_{L,m}(\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v\,,\,(\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \rangle\!\rangle^{1/2}.
\end{aligned}$$

Next, we use the strenghtened Cauchy–Schwarz inequality (4.118) with $(\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \in \widehat{\mathcal{X}}_k$ and get that

$$\begin{aligned}
&\sum_{m=0}^M \sum_{k=0}^m \langle\!\langle \mathcal{Q}_{L,m} v\,,\,v \rangle\!\rangle^{1/2} \langle\!\langle \mathcal{Q}_{L,m}(\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v\,,\,(\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \rangle\!\rangle^{1/2} \\
&\qquad \overset{(4.118)}{\le} C \sum_{m=0}^M \sum_{k=0}^m 2^{-(m-k)/4} \langle\!\langle \mathcal{Q}_{L,m} v\,,\,v \rangle\!\rangle^{1/2} \, |\!|\!| (\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v |\!|\!| \\
&\qquad = C \sum_{m=0}^M \sum_{k=0}^m 2^{-(m-k)/4} \langle\!\langle \mathcal{Q}_{L,m} v\,,\,v \rangle\!\rangle^{1/2} \langle\!\langle (\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v\,,\,v \rangle\!\rangle^{1/2},
\end{aligned}$$

where $C > 0$ is the constant from the strenghtened Cauchy–Schwarz inequality. With $\delta > 0$, which will be fixed later, we use the following variant of the Young inequality

$$ab \le \frac{\delta}{2} a^2 + \frac{\delta^{-1}}{2} b^2 \quad \text{for all } a, b \in \mathbb{R}.$$

We get that

$$C \sum_{m=0}^{M} \sum_{k=0}^{m} 2^{-(m-k)/4} \, \langle\!\langle \mathcal{Q}_{L,m} v \, , \, v \rangle\!\rangle^{1/2} \, \langle\!\langle (\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \, , \, v \rangle\!\rangle^{1/2}$$
$$\leq C \sum_{m=0}^{M} \sum_{k=0}^{m} 2^{-(m-k)/4} \, \frac{\delta}{2} \, \langle\!\langle \mathcal{Q}_{L,m} v \, , \, v \rangle\!\rangle$$
$$+ C \sum_{m=0}^{M} \sum_{k=0}^{m} 2^{-(m-k)/4} \, \frac{\delta^{-1}}{2} \, \langle\!\langle (\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \, , \, v \rangle\!\rangle .$$

The inner sum over k of the first double sum can be bounded by $\sum_{k=0}^{m} 2^{-(m-k)/4} \leq \sum_{k=0}^{\infty} 2^{-k/4} =: K < \infty$. Together with changing the summation order in the second sum, we see that

$$\langle\!\langle \mathcal{Q}_L v \, , \, v \rangle\!\rangle \leq C \sum_{m=0}^{M} \sum_{k=0}^{m} 2^{-(m-k)/4} \, \frac{\delta}{2} \, \langle\!\langle \mathcal{Q}_{L,m} v \, , \, v \rangle\!\rangle$$
$$+ C \sum_{m=0}^{M} \sum_{k=0}^{m} 2^{-(m-k)/4} \, \frac{\delta^{-1}}{2} \, \langle\!\langle (\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \, , \, v \rangle\!\rangle$$
$$\leq C \, K \, \frac{\delta}{2} \sum_{m=0}^{M} \langle\!\langle \mathcal{Q}_{L,m} v \, , \, v \rangle\!\rangle$$
$$+ C \, \frac{\delta^{-1}}{2} \sum_{k=0}^{M} \sum_{m=k}^{M} 2^{-(m-k)/4} \, \langle\!\langle (\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \, , \, v \rangle\!\rangle$$
$$\leq C \, K \, \frac{\delta}{2} \sum_{m=0}^{M} \langle\!\langle \mathcal{Q}_{L,m} v \, , \, v \rangle\!\rangle + C \, K \, \frac{\delta^{-1}}{2} \sum_{k=0}^{M} \langle\!\langle (\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \, , \, v \rangle\!\rangle$$
$$= C \, K \, \frac{\delta}{2} \, \langle\!\langle \mathcal{Q}_L v \, , \, v \rangle\!\rangle + C \, K \, \frac{\delta^{-1}}{2} \, \langle\!\langle \sum_{k=0}^{M} (\widehat{\mathcal{G}}_k - \widehat{\mathcal{G}}_{k-1}) v \, , \, v \rangle\!\rangle$$
$$= C \, K \, \frac{\delta}{2} \, \langle\!\langle \mathcal{Q}_L v \, , \, v \rangle\!\rangle + C \, K \, \frac{\delta^{-1}}{2} \, \langle\!\langle v \, , \, v \rangle\!\rangle .$$

Let $\delta < 2 \, (C \, K)^{-1}$. Then, it holds that

$$\langle\!\langle \mathcal{Q}_L v \, , \, v \rangle\!\rangle \leq \big(1 - C \, K \, \frac{\delta}{2}\big)^{-1} \, C \, K \, \frac{\delta^{-1}}{2} \, \langle\!\langle v \, , \, v \rangle\!\rangle$$
$$= \big(1 - C \, K \, \frac{\delta}{2}\big)^{-1} \, C \, K \, \frac{\delta^{-1}}{2} \, |\!|\!| v |\!|\!|^2 .$$

Hence, there holds the upper bound in (4.102). □

4.7.7 Numerical experiments

In this section, we provide numerical experiments that underpin the theoretical findings of Section 4.6, where we employ H^1-conforming lowest-order FEM in 2D. For ease of notation, we define $\lambda := \lambda_{\mathrm{ctr}}$ for this section. We present an example for AFEM with optimal PCG solver, cf. Section 4.7, and compare the performance of Algorithm 15 for

- different geometries, i.e., the domain $\Omega \subset \mathbb{R}^2$ is either the Z–shaped domain or the L–shaped domain, cf. Figure 4.2,
- different values of $\lambda \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$,
- different values of $\theta \in \{0.05, 0.1, 0.15, \dots, 1\}$,

where $\theta = 1$ corresponds to uniform mesh-refinement.
We consider the following Poisson problem with homogeneous Dirichlet boundary conditions

$$\begin{aligned} -\Delta u^\star &= 1 \quad \text{in } \Omega, \\ u^\star &= 0 \quad \text{on } \Gamma := \partial\Omega, \end{aligned} \tag{4.122}$$

for both geometries from Figure 4.2. As preconditioner for the PCG solver, we use the multilevel additive Schwarz preconditioner of Section 4.7.1 which is optimal, cf. Theorem 30.

Poisson problem (4.122) on Z-shaped domain

In Figure 4.3, we compare Algorithm 15 for different values of θ and λ, and uniform mesh-refinement on the Z-shaped domain, cf. Figure 4.2. To this end, the error estimator $\eta_\ell(u_\ell^k)$ of the last step of the PCG solver is plotted over the number of elements. Recall that $\eta_\ell(u_\ell^k) \simeq \Delta_\ell^k$ according to Proposition 16. We see that uniform mesh-refinement leads to the suboptimal rate of convergence $\mathcal{O}(N^{-2/7})$, while Algorithm 15 regains the optimal rate of convergence $\mathcal{O}(N^{-1/2})$. This empirically confirms Theorem 23. The latter rate of convergence appears to be even robust with respect to $\theta \in \{0.1, 0.3, \dots, 0.9\}$ as well as $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$.

In Figure 4.4, we aim to underpin that Algorithm 15 has the optimal rate of convergence with respect to the computational complexity. To this end, we plot the error estimator $\eta_\ell(u_\ell^k)$ of the last step of the PCG solver over the cumulative sum $\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'}$. In accordance with Theorem 23, we observe again the optimal order $\mathcal{O}\big(\big(\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'}\big)^{-1/2}\big)$.

In Figure 4.5, we take a look at the number of PCG iterations. We observe that a larger value of λ or a smaller value of θ lead to a smaller number of PCG iterations. Nonetheless, in each case, this number stays uniformly bounded.

Summing up so far, we see

- that Algorithm 15 appears to be robust with respect to the choice of θ and λ, cf. Figure 4.3,
- that a larger value of λ leads to less computational cost and a smaller value of θ leads to higher computational cost, cf. Figure 4.4, and,

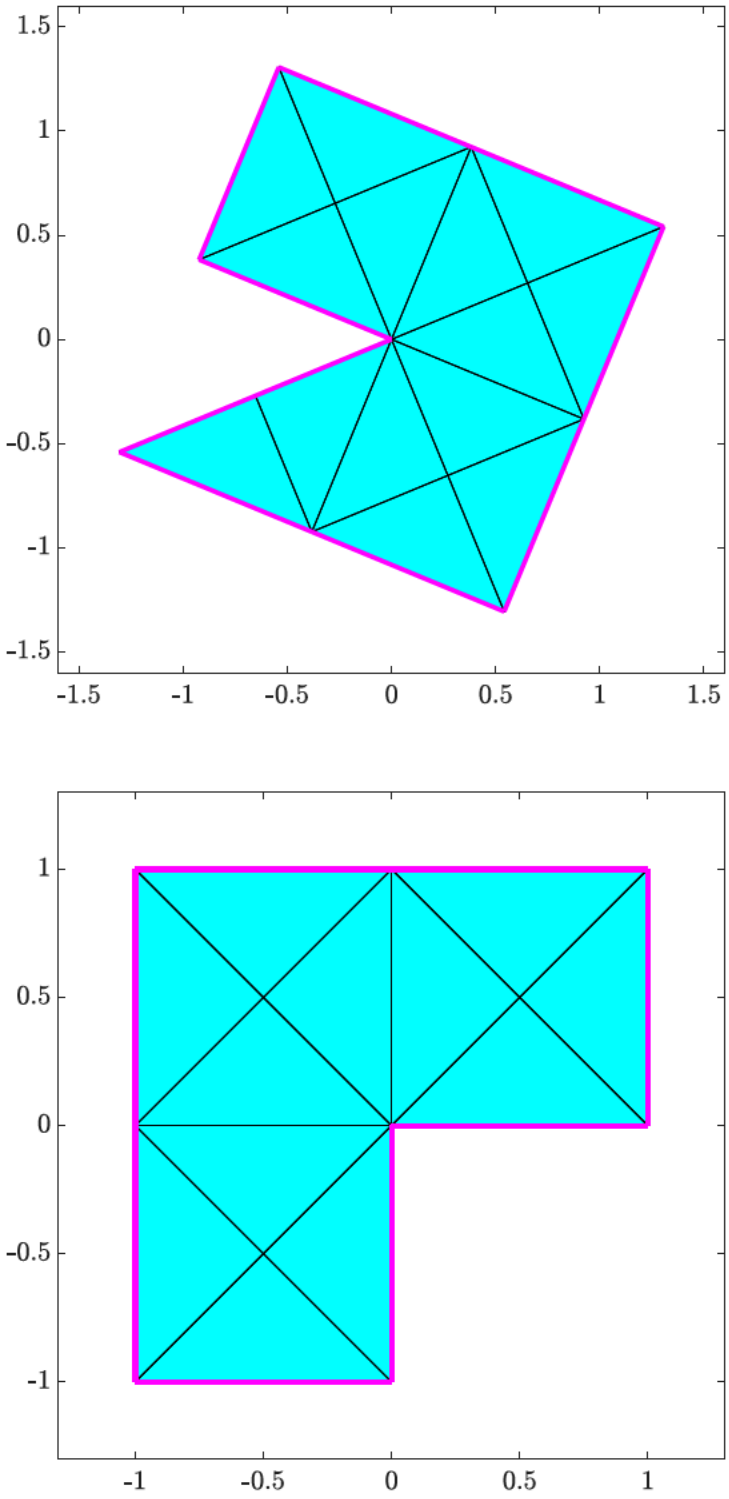


Figure 4.2: Z-shaped domain $\Omega \subset \mathbb{R}^2$ with initial mesh $\mathcal{T}_0$ (top) and L–shaped domain $\Omega \subset \mathbb{R}^2$ with initial mesh $\mathcal{T}_0$ (bottom).

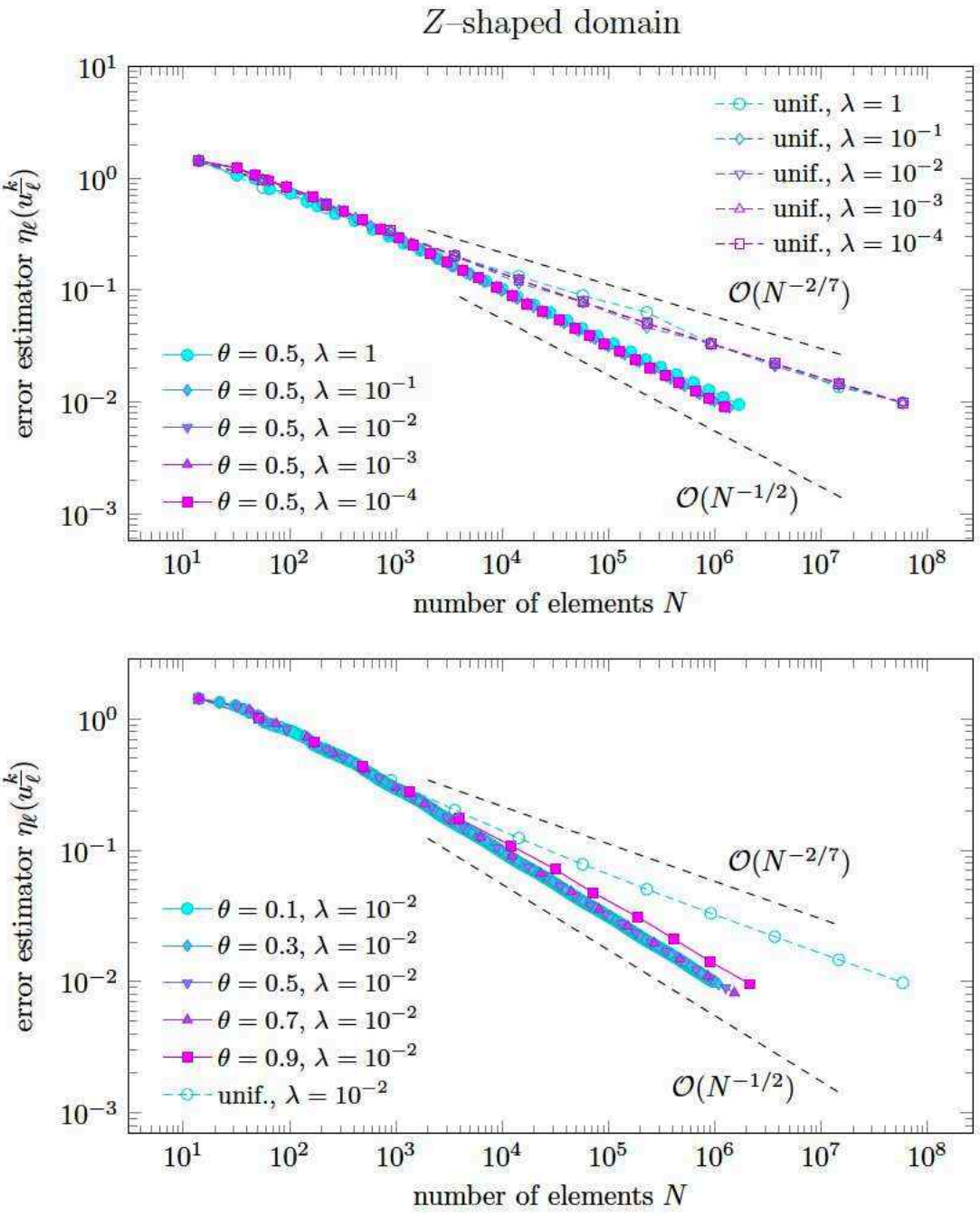


Figure 4.3: Example from Section 4.7.7 (Poisson problem on Z-shaped domain): Error estimator $\eta_\ell(u_\ell^k)$ of the last step of the PCG solver with respect to the number of elements N of the mesh $\mathcal{T}_\ell$ for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

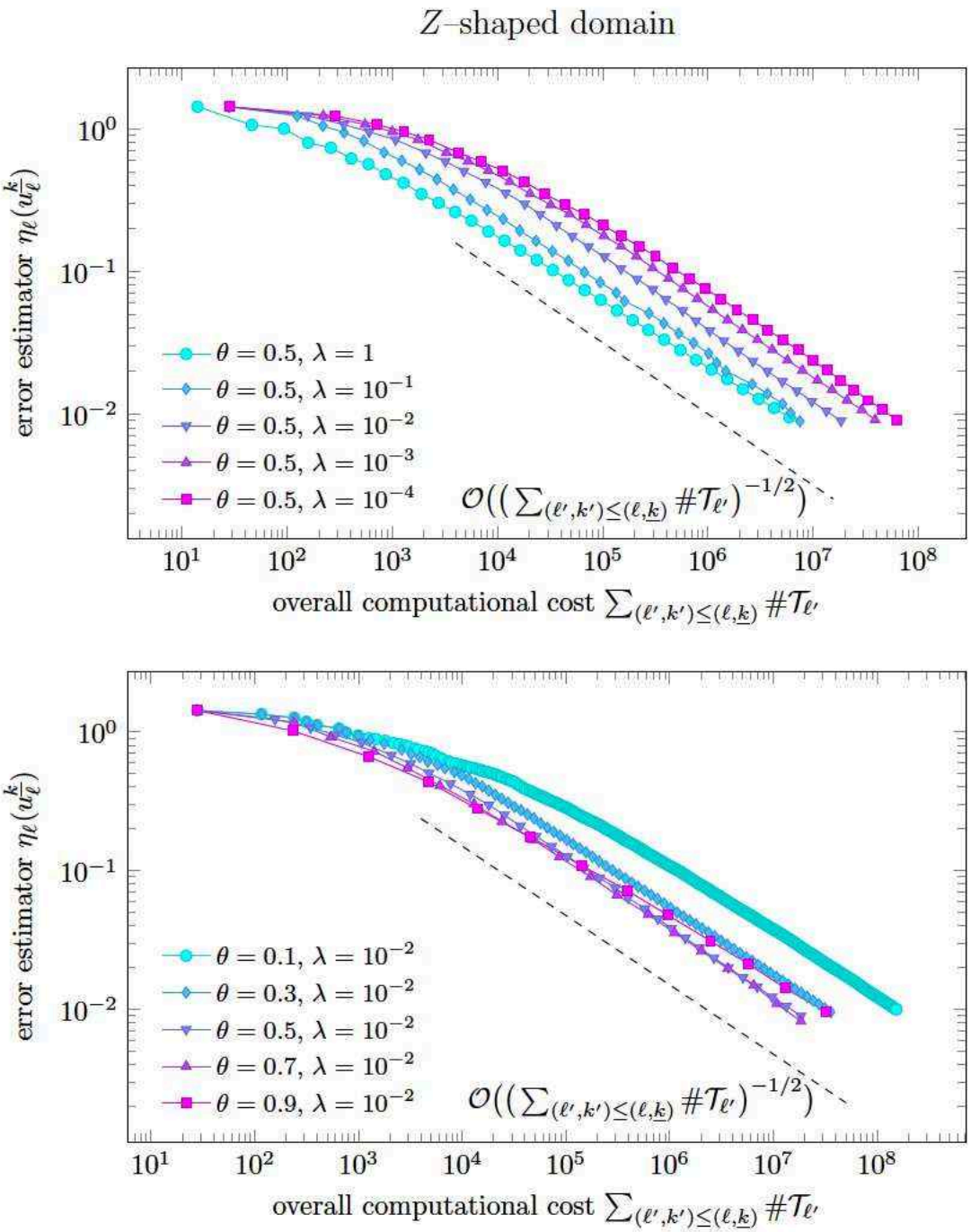


Figure 4.4: Example from Section 4.7.7 (Poisson problem on Z-shaped domain): Error estimator $\eta_\ell(u_\ell^k)$ of the last step of the PCG solver with respect to the overall computational cost expressed as the cumulative sum $\sum_{(\ell',k')\le(\ell,k)} \#\mathcal{T}_{\ell'}$ for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

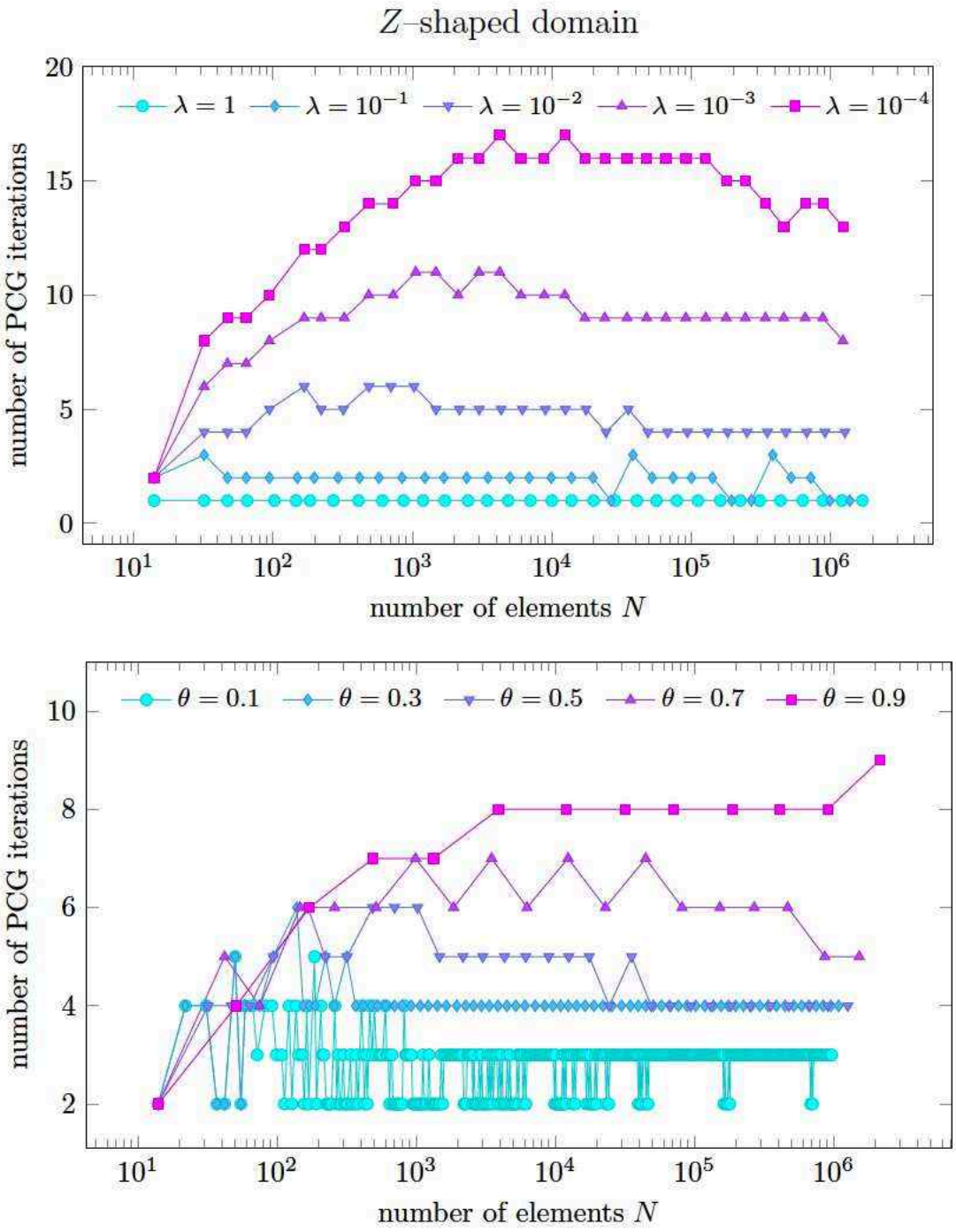


Figure 4.5: Example from Section 4.7.7 (Poisson problem on Z-shaped domain): Number of PCG iterations with respect to the number of elements N for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

θ \ λ	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
0.05	1.9e+08	1.9e+08	1.9e+08	1.9e+08	3.6e+08	5.7e+08	1.0e+09	1.5e+09	**1.9e+09**
0.1	5.3e+07	5.3e+07	5.2e+07	5.3e+07	1.5e+08	2.5e+08	3.7e+08	4.7e+08	6.0e+08
0.15	2.7e+07	2.7e+07	2.7e+07	4.3e+07	7.7e+07	1.3e+08	1.9e+08	2.5e+08	3.2e+08
0.2	1.7e+07	1.7e+07	1.7e+07	3.5e+07	5.0e+07	8.4e+07	1.3e+08	1.6e+08	2.0e+08
0.25	1.2e+07	1.2e+07	1.2e+07	2.6e+07	4.7e+07	7.1e+07	9.1e+07	1.1e+08	1.5e+08
0.3	8.5e+06	8.5e+06	9.9e+06	2.2e+07	3.6e+07	5.0e+07	7.3e+07	9.6e+07	1.2e+08
0.35	6.8e+06	6.8e+06	9.1e+06	2.1e+07	2.7e+07	4.1e+07	5.5e+07	7.1e+07	8.8e+07
0.4	6.2e+06	6.2e+06	7.8e+06	1.6e+07	2.1e+07	3.3e+07	4.6e+07	6.3e+07	7.4e+07
0.45	5.8e+06	7.1e+06	7.0e+06	1.3e+07	1.9e+07	3.0e+07	4.3e+07	5.3e+07	6.7e+07
0.5	5.9e+06	4.5e+06	7.5e+06	1.3e+07	1.8e+07	2.9e+07	3.9e+07	4.8e+07	6.2e+07
0.55	5.9e+06	4.2e+06	6.7e+06	1.0e+07	1.9e+07	2.8e+07	3.8e+07	5.0e+07	6.3e+07
0.6	1.9e+07	5.3e+06	5.7e+06	8.3e+06	1.6e+07	2.5e+07	3.2e+07	4.3e+07	5.8e+07
0.65	1.3e+07	5.0e+06	6.3e+06	1.1e+07	1.6e+07	2.4e+07	3.4e+07	4.6e+07	5.7e+07
0.7	1.5e+07	**4.0e+06**	7.5e+06	1.1e+07	1.8e+07	2.7e+07	3.8e+07	5.0e+07	6.1e+07
0.75	9.4e+06	1.7e+07	9.2e+06	1.0e+07	2.5e+07	3.7e+07	4.8e+07	6.3e+07	7.7e+07
0.8	1.3e+07	1.6e+07	2.2e+07	1.6e+07	2.0e+07	2.8e+07	3.7e+07	4.8e+07	5.7e+07
0.85	9.1e+06	1.5e+07	2.3e+07	2.2e+07	2.8e+07	3.8e+07	5.0e+07	6.1e+07	7.2e+07
0.9	2.9e+07	1.6e+07	1.8e+07	2.3e+07	3.2e+07	4.5e+07	5.7e+07	6.7e+07	7.7e+07
0.95	4.4e+07	3.3e+07	4.8e+07	6.4e+07	9.0e+07	1.2e+08	1.5e+08	1.9e+08	2.2e+08

min **max**

Figure 4.6: Example from Section 4.7.7 (Poisson problem on Z-shaped domain): Overall computational cost $\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'}$ such that $\eta_\ell(u_\ell^{\underline{k}}) < \tau$ for given precision $\tau = 10^{-2}$, $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$, and $\theta \in \{0.05, 0.1, \dots, 0.95\}$.

- that a larger value of λ as well as a smaller value of θ lead to fewer PCG iterations, cf. Figure 4.5.

Hence, the question arises, how to choose θ and λ in order to mimize the overall computational cost to reach a given bound $\tau > 0$ for the error estimator, i.e., such that $\eta_\ell(u_\ell^k) < \tau$. In Figure 4.6, we compare the computational cost to reach the precision $\tau = 10^{-2}$ for $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$ and $\theta \in \{0.05, 0.1, \dots, 0.95\}$. As a result, we get that the best choice is $\lambda = 10^{-0.5}$ and $\theta = 0.7$. For the overall computational cost it then holds that

$$\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'} = 4034040,$$

where u_ℓ^k is the first approximation such that $\eta_\ell(u_\ell^k) < 10^{-2}$.

Poisson problem (4.122) on L-shaped domain

In Figure 4.7, we compare Algorithm 15 for different values of θ and λ, and uniform mesh-refinement on the L-shaped domain, cf. Figure 4.2. To this end, the error estimator $\eta_\ell(u_\ell^k)$ of the last step of the PCG solver is plotted over the number of elements. Recall that $\eta_\ell(u_\ell^k) \simeq \Delta_\ell^k$ according to Proposition 16. We see that uniform mesh-refinement leads to the suboptimal rate of convergence $\mathcal{O}(N^{-1/3})$, while Algorithm 15 regains the optimal rate of convergence $\mathcal{O}(N^{-1/2})$. Again, this empirically confirms Theorem 23. The latter rate of convergence appears to be even robust with respect to $\theta \in \{0.1, 0.3, \dots, 0.9\}$ as well as $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$.

In Figure 4.8, the error estimator $\eta_\ell(u_\ell^k)$ of the last step of the PCG solver is plotted over the cumulative sum $\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'}$. In accordance with Theorem 23, we observe again the optimal order $\mathcal{O}\big(\big(\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'}\big)^{-1/2}\big)$.

In Figure 4.9, we take a look at the number of PCG iterations. We observe that a larger value of λ or a smaller value of θ lead to a smaller number of PCG iterations. Nonetheless, in each case, this number stays uniformly bounded.

As for the Z-shaped domain, we see

- that Algorithm 15 appears to be robust with respect to the choice of θ and λ, cf. Figure 4.3,
- that a larger value of λ leads to less computational cost and a smaller value of θ leads to higher computational cost, cf. Figure 4.4, and,
- that a larger value of λ as well as a smaller value of θ lead to fewer PCG iterations, cf. Figure 4.5.

In Figure 4.10, we compare the computational cost to reach the precision $\tau = 10^{-2}$ for $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$ and $\theta \in \{0.05, 0.1, \dots, 0.95\}$. As a result, we get that the best choice is $\lambda = 10^{-0.5}$ and $\theta = 0.8$. For the overall computational cost it then holds that

$$\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'} = 2832761,$$

where u_ℓ^k is the first approximation such that $\eta_\ell(u_\ell^k) < 10^{-2}$.

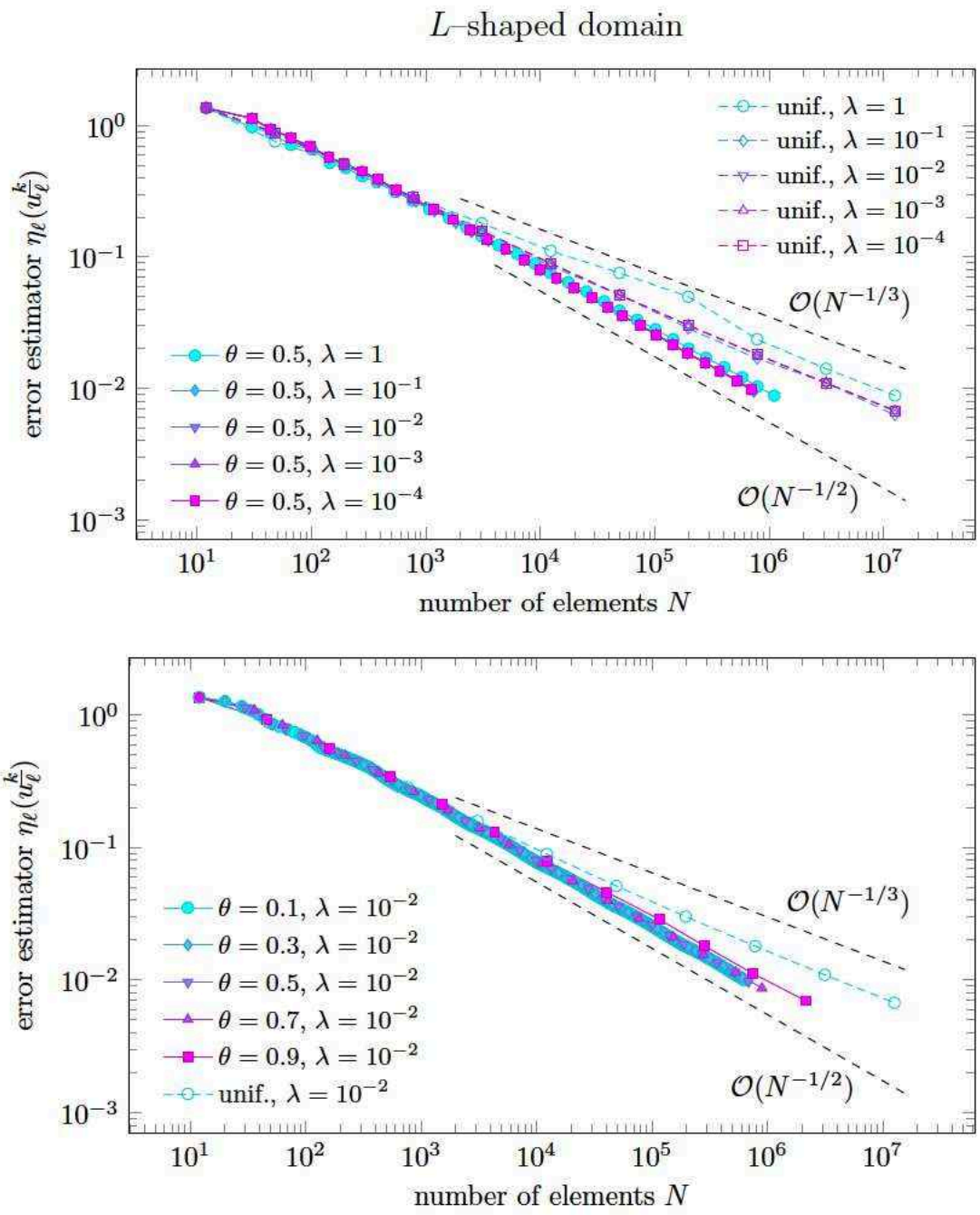


Figure 4.7: Example from Section 4.7.7 (Poisson problem on L-shaped domain): Error estimator $\eta_\ell(u_\ell^k)$ of the last step of the PCG solver with respect to the number of elements N of the mesh $\mathcal{T}_\ell$ for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

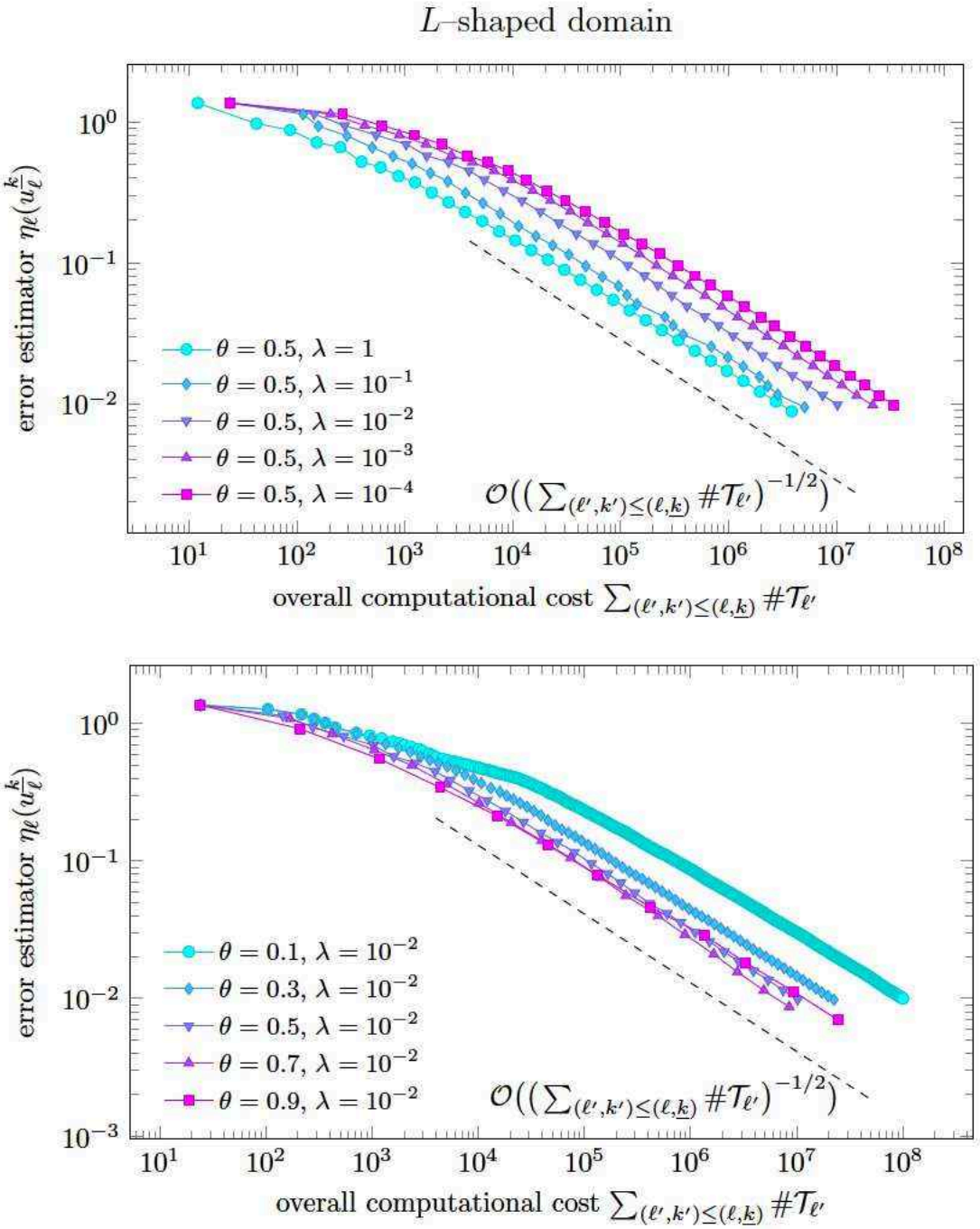


Figure 4.8: Example from Section 4.7.7 (Poisson problem on L-shaped domain): Error estimator $\eta_\ell(u_\ell^k)$ of the last step of the PCG solver with respect to the overall computational cost expressed as the cumulative sum $\sum_{(\ell',k')\leq(\ell,k)} \#\mathcal{T}_{\ell'}$ for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

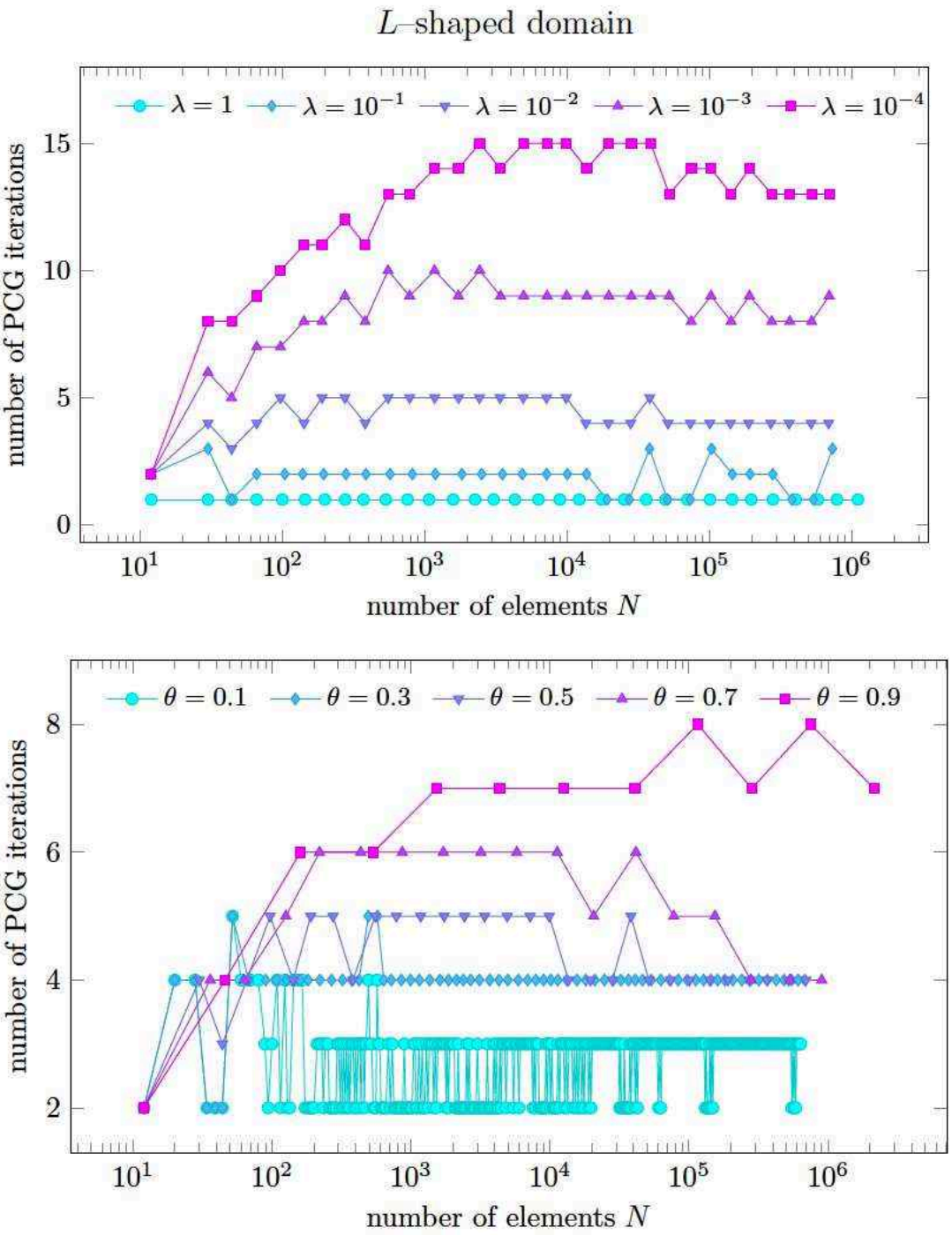


Figure 4.9: Example from Section 4.7.7 (Poisson problem on L-shaped domain): Number of PCG iterations with respect to the number of elements N for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

θ \ λ	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
0.05	1.2e+08	1.2e+08	1.2e+08	1.2e+08	2.3e+08	3.8e+08	6.6e+08	9.8e+08	**1.2e+09**
0.1	3.5e+07	3.5e+07	3.5e+07	3.4e+07	1.0e+08	1.6e+08	2.4e+08	3.1e+08	4.0e+08
0.15	1.8e+07	1.8e+07	1.8e+07	2.8e+07	5.3e+07	8.8e+07	1.3e+08	1.7e+08	2.1e+08
0.2	1.1e+07	1.1e+07	1.1e+07	2.2e+07	3.4e+07	5.6e+07	8.6e+07	1.1e+08	1.3e+08
0.25	7.5e+06	7.5e+06	7.8e+06	1.6e+07	2.9e+07	4.5e+07	6.2e+07	7.8e+07	1.0e+08
0.3	5.7e+06	5.7e+06	5.8e+06	1.4e+07	2.2e+07	3.5e+07	4.6e+07	5.9e+07	7.0e+07
0.35	4.7e+06	4.7e+06	4.7e+06	1.3e+07	1.8e+07	2.8e+07	4.0e+07	5.2e+07	6.4e+07
0.4	3.9e+06	3.9e+06	4.9e+06	1.2e+07	1.5e+07	2.3e+07	3.1e+07	4.0e+07	5.0e+07
0.45	3.4e+06	3.4e+06	4.5e+06	1.1e+07	1.2e+07	1.8e+07	2.6e+07	3.3e+07	4.1e+07
0.5	3.8e+06	5.4e+06	5.1e+06	7.5e+06	1.0e+07	1.6e+07	2.2e+07	2.8e+07	3.4e+07
0.55	3.0e+06	3.3e+06	4.4e+06	6.5e+06	8.7e+06	1.4e+07	1.8e+07	2.3e+07	2.8e+07
0.6	3.2e+06	3.6e+06	4.5e+06	5.6e+06	7.9e+06	1.2e+07	1.6e+07	2.1e+07	3.8e+07
0.65	5.1e+06	4.7e+06	4.7e+06	6.0e+06	7.9e+06	1.2e+07	1.6e+07	2.1e+07	2.7e+07
0.7	1.7e+07	2.9e+06	5.1e+06	6.6e+06	8.5e+06	1.5e+07	2.0e+07	2.6e+07	3.2e+07
0.75	9.5e+06	7.5e+06	3.8e+06	4.1e+06	1.4e+07	2.0e+07	2.7e+07	3.5e+07	4.3e+07
0.8	7.1e+06	**2.8e+06**	6.7e+06	8.0e+06	1.0e+07	1.4e+07	1.9e+07	2.4e+07	2.9e+07
0.85	4.2e+06	6.6e+06	3.8e+06	1.2e+07	1.6e+07	2.0e+07	2.6e+07	3.3e+07	3.9e+07
0.9	5.8e+06	9.2e+06	5.6e+06	1.8e+07	2.4e+07	3.4e+07	4.3e+07	5.3e+07	6.3e+07
0.95	7.6e+06	9.5e+06	1.6e+07	2.4e+07	3.1e+07	4.2e+07	5.2e+07	6.1e+07	7.2e+07

min **max**

Figure 4.10: Example from Section 4.7.7 (Poisson problem on L-shaped domain): Overall computational cost $\sum_{(\ell',k')\leq(\ell,\underline{k})} \#\mathcal{T}_{\ell'}$ such that $\eta_\ell(u_\ell^{\underline{k}}) < \tau$ for given precision $\tau = 10^{-2}$, $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$, and $\theta \in \{0.05, 0.1, \dots, 0.95\}$.

4.8 AFEM for quasi-linear elliptic PDE with strongly monotone nonlinearity

The second setting which we introduce in this chapter and which fits into the abstract framework of Section 4.2–Section 4.6 is AFEM for a boundary value problem with a strongly monotone nonlinearity.

Model problem

We consider the following boundary value problem

$$\begin{aligned} -\operatorname{div}\big(\mu(x,|\nabla u^\star(x)|^2)\nabla u^\star(x)\big) &= f(x) && \text{in } \Omega,\\ u^\star(x) &= 0 && \text{on } \Gamma_D,\\ \mu(x,|\nabla u^\star(x)|^2)\,\partial_{\mathbf{n}} u^\star(x) &= g(x) && \text{on } \Gamma_N, \end{aligned} \tag{4.123}$$

where $\Omega \subset \mathbb{R}^d$ is a bounded Lipschitz domain with $d \in \{2,3\}$ and polytopal boundary $\Gamma = \partial\Omega$, and given $f \in L^2(\Omega)$, $g \in L^2(\Gamma)$ as well as a scalar nonlinearity $\mu\colon \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}$. Let the boundary Γ be split into relatively open and disjoint Dirichlet and Neumann boundaries Γ_D, Γ_N such that $|\Gamma_D| > 0$ and $\Gamma = \overline{\Gamma}_D \cup \overline{\Gamma}_N$. The scalar nonlinearity μ satisfies the following properties (N1)–(N4) with generic constants $\gamma_1, \gamma_2, \widetilde{\gamma}_1, \widetilde{\gamma}_2, L_\mu, \widetilde{L}_\mu > 0$, which have already been considered in [GMZ12, GHPS18]:

(N1) boundedness of $\mu(x,t)$: There exist constants $\gamma_1, \gamma_2 > 0$ such that

$$\gamma_1 \leq \mu(x,t) \leq \gamma_2 \quad \text{for all } x \in \Omega \text{ and } t \geq 0.$$

(N2) boundedness of $\mu(x,t) + 2\,t\frac{\mathrm{d}}{\mathrm{d}t}\mu(x,t)$: For $x \in \Omega$, the function $\mu(x,\cdot)$ is continuously differentiable, i.e., $\mu(x,\cdot) \in C^1(\mathbb{R}_{\geq 0}, \mathbb{R})$ and there exist constants $\widetilde{\gamma}_1, \widetilde{\gamma}_2 > 0$ such that

$$\widetilde{\gamma}_1 \leq \mu(x,t) + 2\,t\frac{\mathrm{d}}{\mathrm{d}t}\mu(x,t) \leq \widetilde{\gamma}_2 \quad \text{for all } x \in \Omega \text{ and } t \geq 0.$$

(N3) Lipschitz-continuity of $\mu(x,t)$ in x: There exists a constant $L_\mu > 0$ such that

$$|\mu(x,t) - \mu(y,t)| \leq L_\mu |x-y| \quad \text{for all } x, y \in \Omega \text{ and } t \geq 0.$$

(N4) Lipschitz-continuity of $t\frac{\mathrm{d}}{\mathrm{d}t}\mu(x,t)$ in x: There exists a constant $\widetilde{L}_\mu > 0$ such that

$$|t\frac{\mathrm{d}}{\mathrm{d}t}\mu(x,t) - t\frac{\mathrm{d}}{\mathrm{d}t}\mu(y,t)| \leq \widetilde{L}_\mu |x-y| \quad \text{for all } x, y \in \Omega \text{ and } t \geq 0.$$

Weak formulation

The weak formulation of (4.123) reads as follows: Find $u \in H^1_D(\Omega) := \{w \in H^1(\Omega) : w = 0 \text{ on } \Gamma_D\}$ such that

$$\int_\Omega \mu(x, |\nabla u^\star(x)|^2)\, \nabla u^\star \cdot \nabla v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x + \int_{\Gamma_N} g v \,\mathrm{d}s \quad \text{for all } v \in H^1_D(\Omega). \tag{4.124}$$

With respect to the abstract framework of Section 4.2, we take $\mathcal{H} = H^1_D(\Omega)$, $\mathbb{K} = \mathbb{R}$, and $\langle\!\langle \cdot\,, \cdot \rangle\!\rangle = \langle\!\langle \nabla\cdot\,, \nabla\cdot \rangle\!\rangle$ with corresponding norm $|\!|\!| v |\!|\!| = \|\nabla v\|_{L^2(\Omega)}$. We obtain (4.7) with operators

$$\langle \mathcal{A} w\,,\, v\rangle_{\mathcal{H}'\times\mathcal{H}} = \int_\Omega \mu(x, |\nabla w(x)|^2)\, \nabla w(x) \cdot \nabla v(x)\,\mathrm{d}x, \tag{4.125a}$$

$$F(v) = \int_\Omega f v \,\mathrm{d}x + \int_{\Gamma_N} g v \,\mathrm{d}s \tag{4.125b}$$

for all $v, w \in \mathcal{H}$. We recall from [GHPS18, Proposition 8.2] that (N1)–(N2) implies that $\mathcal{A}$ is strongly monotone (with $\alpha := \widetilde{\gamma}_1$) and Lipschitz continuous (with $L := \widetilde{\gamma}_2$), and that there exists a potential $P\colon H^1_0(\Omega) \to \mathbb{R}$, i.e., there hold (O1)–(O3) with $\alpha = \widetilde{\gamma}_1$ and $L = \widetilde{\gamma}_2$. The assumptions (N3)–(N4) are required to prove the well-posedness and the properties (A1)–(A4) of the residual *a posteriori* error estimator.

Triangulation and mesh-refinement

Let $\mathcal{T}_0$ be a conforming initial triangulation of Ω into simplices $T \in \mathcal{T}_0$. As the refinement strategy `refine`$(\cdot)$, we employ newest vertex bisection such that the axioms (R1)–(R3) are fulfilled, cf. Section 3.6.

Discretization

For $\mathcal{T}_\ell \in \mathbb{T}$, we consider the lowest-order FEM space

$$\mathcal{X}_\ell := \big\{v \in C(\overline{\Omega}) \,:\, v|_T \in \mathcal{P}^1(T) \text{ for all } T \in \mathcal{T}_\ell\big\} \cap H^1_D(\Omega), \tag{4.126}$$

i.e., the space of all continuous piecewise affine functions that vanish on the boundary $\Gamma = \partial\Omega$.

Error estimator

For all elements $T \in \mathcal{T}_\ell$ and discrete functions $v_\ell \in \mathcal{X}_\ell$, we define the weighted-residual error indicators, cf., e.g., [GMZ12, GHPS18]) via

$$\begin{aligned}\eta_\ell(T, v_\ell)^2 :=& |T|^{2/d}\, \|f + \operatorname{div}(\mu(\cdot, |\nabla v_\ell|^2)\nabla v_\ell)\|_{L^2(T)} + |T|^{1/d}\, \|[\mu(\cdot, |\nabla v_\bullet|^2)\nabla v_\bullet \cdot \mathbf{n}]\|_{L^2(\partial T\cap\Omega)} \\ &+ |T|^{1/d} \|g - \mu(\cdot, |\nabla v_\bullet|^2)\nabla v_\bullet \cdot \mathbf{n}\|_{L^2(\partial T\cap\Gamma_N)},\end{aligned} \tag{4.127}$$

where $[\cdot]$ denotes the usual jump of piecewise continuous functions across element interfaces, and $\mathbf{n}$ is the outer normal vector of the considered element. Due to assumption (N3) on the nonlinearity $\mu(\cdot,\cdot)$, the presented error indicators are well-defined.

While reliability (A3) and discrete reliability (A4) are proved as in the linear case; cf., e.g., [CKNS08] for the linear case and [GMZ12, Theorem 3.3 and 3.4] for the present nonlinear setting, the verification of stability (A1) and reduction (A2) requires the validity of an appropriate inverse estimate. For scalar nonlinearities and under the assumptions (N1)–(N4), the latter is proved in [GMZ12, Lemma 3.7]. Using this inverse estimate, the proof of (A1)–(A2) follows as for the linear case, cf., e.g., [CKNS08] for the linear case or [GMZ12, Section 3.3] for scalar nonlinearities.

Zarantonello iteration

Since the nonlinear system (4.8) can hardly be solved exactly, we use the Zarantonello iteration, also called Banach–Picard iteration, as iteration function $\Phi_\ell\colon \mathcal{X}_\ell \to \mathcal{X}_\ell$ for Step (i) of Algorithm 15: Recall that the Riesz mapping $I_{\mathcal{H}}\colon \mathcal{H} \to \mathcal{H}'$, $I_{\mathcal{H}} w \mapsto \langle\!\langle \cdot\,, w\rangle\!\rangle$ is an isometric isomorphism, cf. [Yos80, Chapter III.6] and let $I_\ell\colon \mathcal{X}_\ell \to \mathcal{X}_\ell'$, $I_\ell v_\ell \to \langle\!\langle \cdot\,, v_\ell\rangle\!\rangle$ denote the discrete Riesz operator. Additionally, let $\mathcal{A}_\ell\colon \mathcal{X}_\ell \to \mathcal{X}_\ell'$ and $F_\ell\colon \mathcal{X}_\ell \to \mathbb{R}$ be the restrictions of $\mathcal{A}$ and F respectively to the discrete space $\mathcal{X}_\ell$. Then, define

$$\Phi_\ell\colon \mathcal{X}_\ell \to \mathcal{X}_\ell, \quad v_\ell \mapsto v_\ell - \frac{\alpha}{L^2}\, I_\ell^{-1}(\mathcal{A}_\ell v_\ell - F_\ell). \tag{4.128}$$

Given $u_\ell^k \in \mathcal{X}_\ell$, we thus compute the discrete iterate $u_\ell^{k+1} := \Phi_\ell(u_\ell^k)$ as follows:

(i) Solve the linear system $\langle\!\langle v_\ell\,, w_\ell\rangle\!\rangle = \langle \mathcal{A} u_\ell^k - F\,, v_\ell\rangle_{\mathcal{H}\times\mathcal{H}'}$ for all $v_\ell \in \mathcal{X}_\ell$.

(ii) Define $u_\ell^{k+1} := u_\ell^k - \frac{\alpha}{L^2} w_\ell$.

In explicit terms, the computation of one step of the iteration requires only the solution of one (discretized) Poisson equation with homogeneous Dirichlet data. Then, Φ_ℓ satisfies the norm contraction (C2) with $q_{\mathrm{ctr}}^2 = 1 - \alpha^2/L^2$, cf., e.g., [GHPS18, Section 3.2] and it holds that

$$\begin{aligned}
\mathcal{E}(\Phi_\ell(v_\ell)) - \mathcal{E}(u_\ell^\star) &\overset{(4.9)}{\leq} \frac{L}{2}\, |\!|\!| u_\ell^\star - \Phi_\ell(v_\ell) |\!|\!|^2 \\
&\overset{(C2)}{\leq} \frac{L}{2}\, q_{\mathrm{ctr}}^2\, |\!|\!| u_\ell^\star - v_\ell |\!|\!|^2 \\
&\overset{(4.9)}{\leq} \frac{L}{\alpha}\, q_{\mathrm{ctr}}^2 \big(\mathcal{E}(v_\ell) - \mathcal{E}(u_\ell^\star)\big).
\end{aligned}$$

In this case, the additional validity of (C1) with the modified constant $\frac{L}{\alpha}\, q_{\mathrm{ctr}}^2$ follows from an additional condition on L/α involving the *golden ratio*, namely

$$0 \leq \frac{L}{\alpha}\, q_{\mathrm{ctr}}^2 = \frac{L}{\alpha} - \frac{\alpha}{L} < 1 \quad \Longleftrightarrow \quad \frac{L}{\alpha} < \frac{1+\sqrt{5}}{2} \approx 1.618. \tag{4.129}$$

Moreover, with the same arguments, (C1) guarantees that

$$|\!|\!| u_\ell^\star - \Phi_\ell(v_\ell) |\!|\!|^2 \leq \frac{L}{\alpha}\, q_{\mathrm{ctr}}^2\, |\!|\!| u_\ell^\star - v_\ell |\!|\!|^2.$$

Hence, the condition (4.129) even yields equivalence of (C1) and (C2) (but with different contraction constants q_{ctr}).

Altogether, the present setting fits into the abstract framework of Section 4.2 and the main results from Section 4.6 apply to it.

4.8.1 Numerical experiments

In this section, we provide numerical experiments that again underpin the theoretical findings of Section 4.6. For ease of notation, we define $\lambda := \lambda_{\mathrm{ctr}}$ for this section. We present two examples for AFEM for strongly monotone nonlinearities, cf. Section 4.8, one with homogeneous Dirichlet boundary conditions on the L-shaped domain and the second with mixed boundary conditions on the Z-shaped domain, cf. Figure 4.11 where the Dirichlet boundary Γ_D is marked by a thick pink line. We compare the performance of Algorithm 15 for

- different values of $\lambda \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$,
- different values of $\theta \in \{0.05, 0.1, 0.15, \dots, 1\}$,

where $\theta = 1$ corresponds to uniform mesh-refinement.

Homogeneous problem on L-shaped domain

We consider the boundary value problem

$$\begin{aligned} -\operatorname{div}\big(\mu(\cdot, |\nabla u^\star|^2)\nabla u^\star\big) &= 1 \quad \text{in } \Omega, \\ u^\star &= 0 \quad \text{on } \Gamma, \end{aligned} \tag{4.130}$$

where the scalar nonlinearity $\mu\colon \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ is defined by

$$\mu(x, |\nabla u^\star|^2) := 1 + \frac{\ln(1 + |\nabla u^\star|^2)}{1 + |\nabla u^\star|^2}. \tag{4.131}$$

Then, (N1)–(N4) hold with $\alpha = \widetilde{\gamma}_1 \approx 0.9582898017$ and $L = \widetilde{\gamma}_2 \approx 1.542343818$.

In Figure 4.12, we compare Algorithm 15 for different values of θ and λ, and uniform mesh-refinement. To this end, the error estimator $\eta_\ell(u_\ell^k)$ of the last step of the Zarantonello iteration is plotted over the number of elements. We see that uniform mesh-refinement leads to the suboptimal rate of convergence $\mathcal{O}(N^{-1/3})$ for the L–shaped domain. Algorithm 15 regains the optimal rate of convergence $\mathcal{O}(N^{-1/2})$, independently of the actual choice of $\theta \in \{0.1, 0.3, \dots, 0.9\}$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$. Since $\eta_\ell(u_\ell^k) \simeq \Delta_\ell^k$, this again empirically confirms Theorem 23.

In Figure 4.13, we plot the estimator $\eta_\ell(u_\ell^k)$ of the last step of the Zarantonello iteration over the cumulative sum $\sum_{(\ell',k')\leq(\ell,\underline{k})} \#\mathcal{T}_{\ell'}$. As predicted in Theorem 23, we observe that Algorithm 15 regains the optimal order of convergence $\mathcal{O}\big(\big(\sum_{(\ell',k')\leq(\ell,\underline{k})} \#\mathcal{T}_{\ell'}\big)^{-1/2}\big)$ with respect to the computational complexity. The rate seems to be independent of the values of λ or θ.

In Figure 4.14, we take a look at the number of Zarantonello iterations. Similarly to the number of PCG iterations in Figure 4.5 and Figure 4.9, we observe that that a larger value of λ or a smaller value of θ lead to less iterations, while the number stays uniformly bounded in each case.

In Figure 4.15, we compare the computational cost to reach the precision $\tau = 10^{-2}$ for $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$ and $\theta \in \{0.05, 0.1, \dots, 0.95\}$. As a result, we get that the best

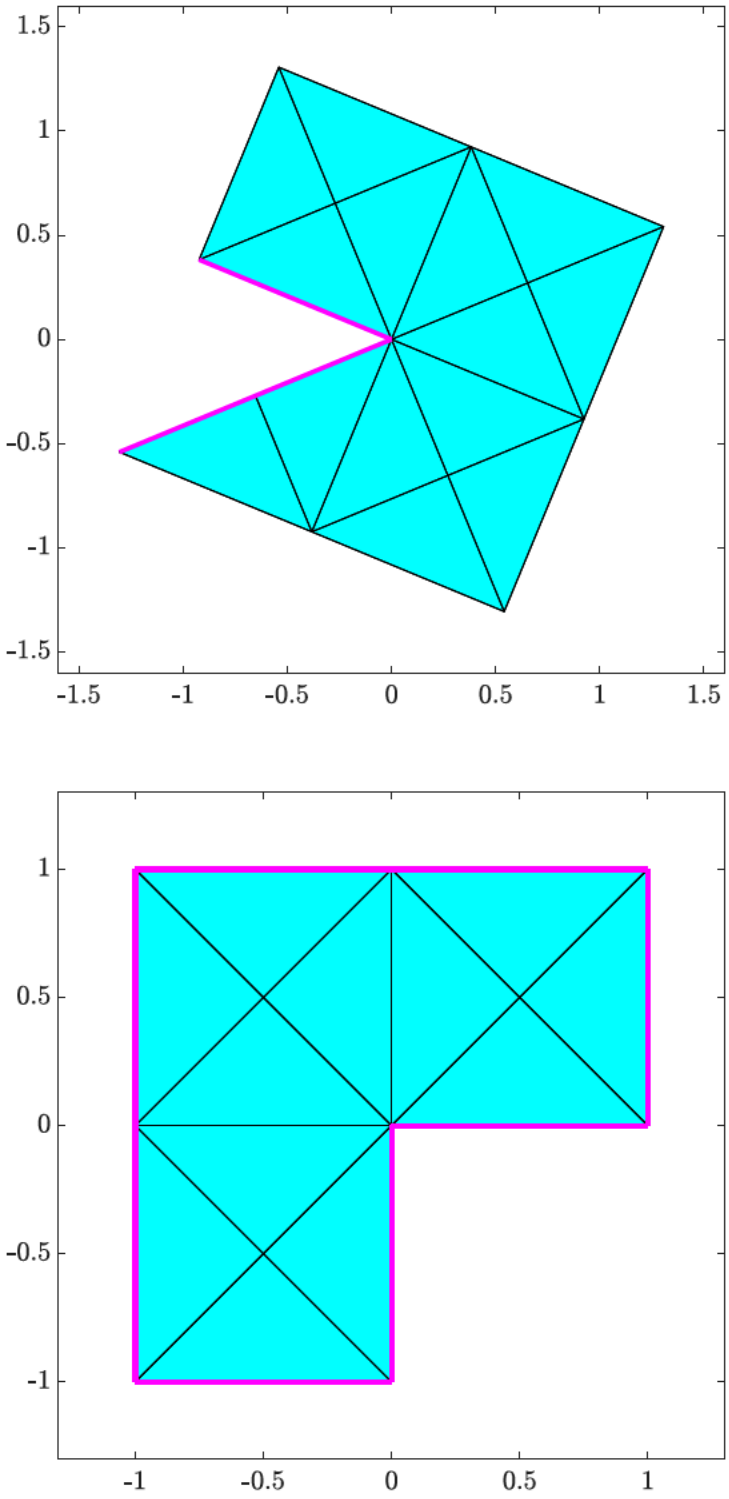


Figure 4.11: Z-shaped domain $\Omega \subset \mathbb{R}^2$ with initial mesh $\mathcal{T}_0$ (top) and L–shaped domain $\Omega \subset \mathbb{R}^2$ with initial mesh $\mathcal{T}_0$ (bottom), where Γ_D is marked by a thick pink line.

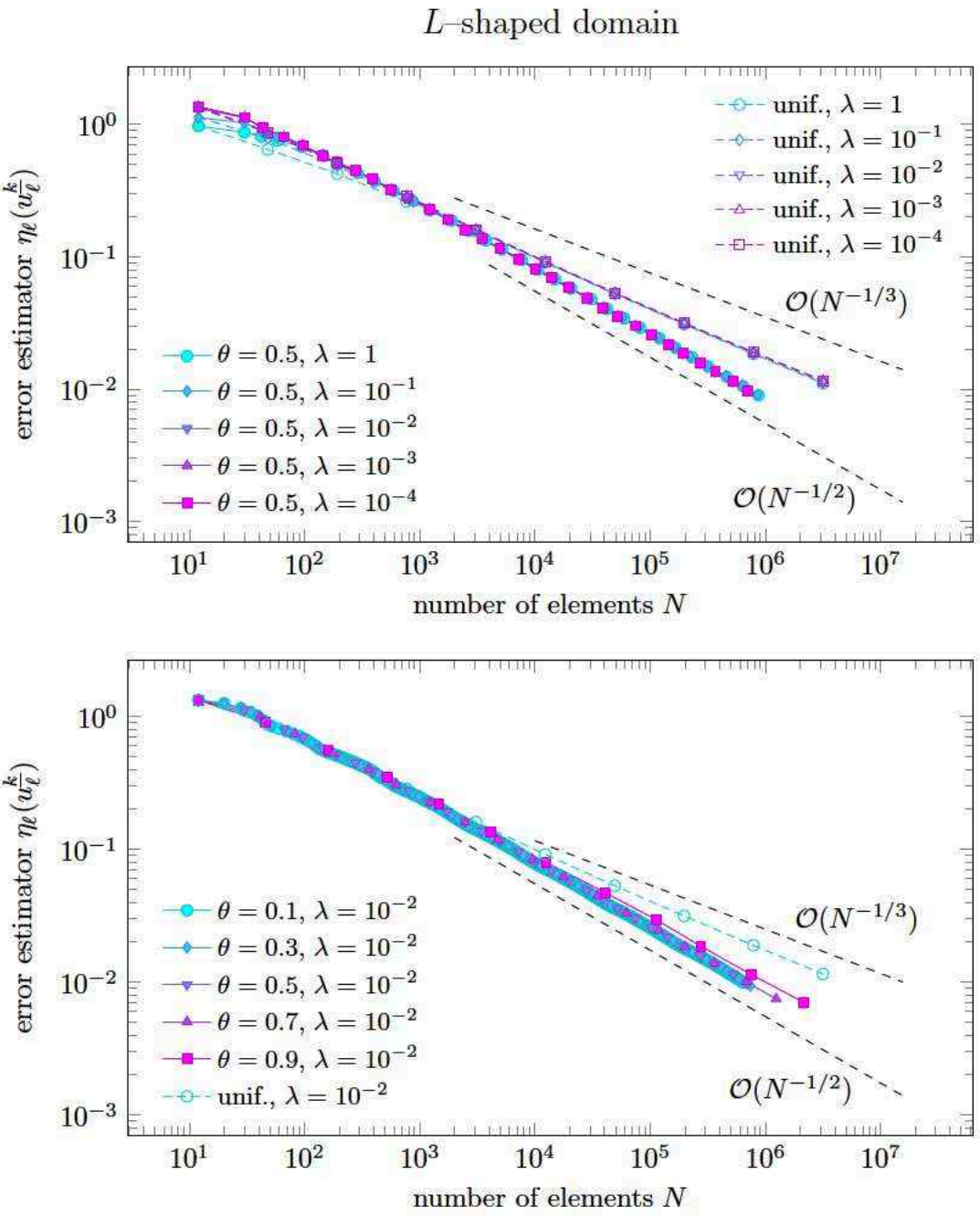


Figure 4.12: Example from Section 4.8.1 (Homogeneous problem on L-shaped domain): Error estimator $\eta_\ell(u_\ell^k)$ of the last step of the Zarantonello iteration with respect to the number of elements N of the mesh $\mathcal{T}_\ell$ for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

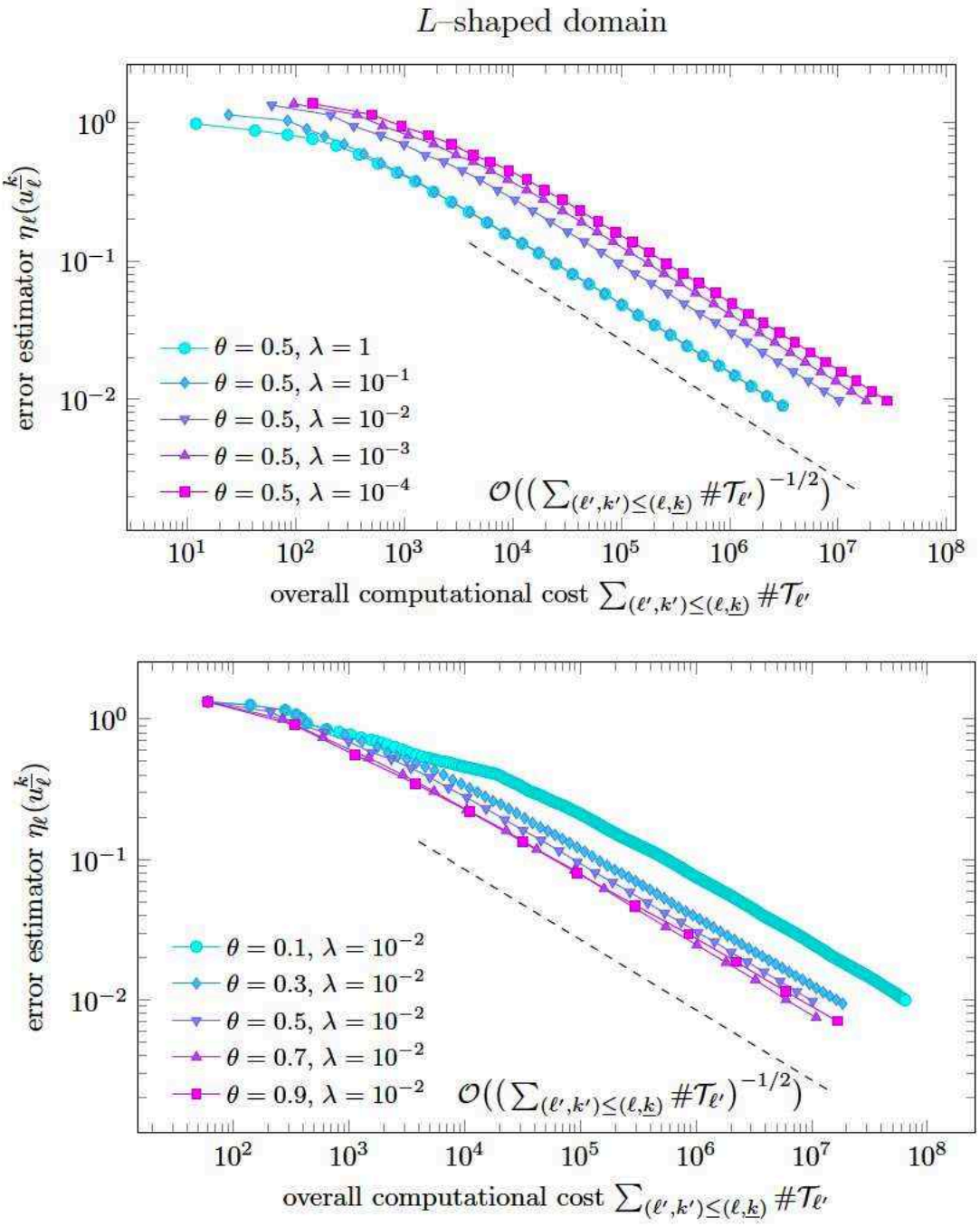


Figure 4.13: Example from Section 4.8.1 (Homogeneous problem on L-shaped domain): Error estimator $\eta_\ell(u_\ell^k)$ of the last step of the Zarantonello iteration with respect to the overall computational cost expressed as the cumulative sum $\sum_{(\ell',k')\le(\ell,k)} \#\mathcal{T}_{\ell'}$ for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

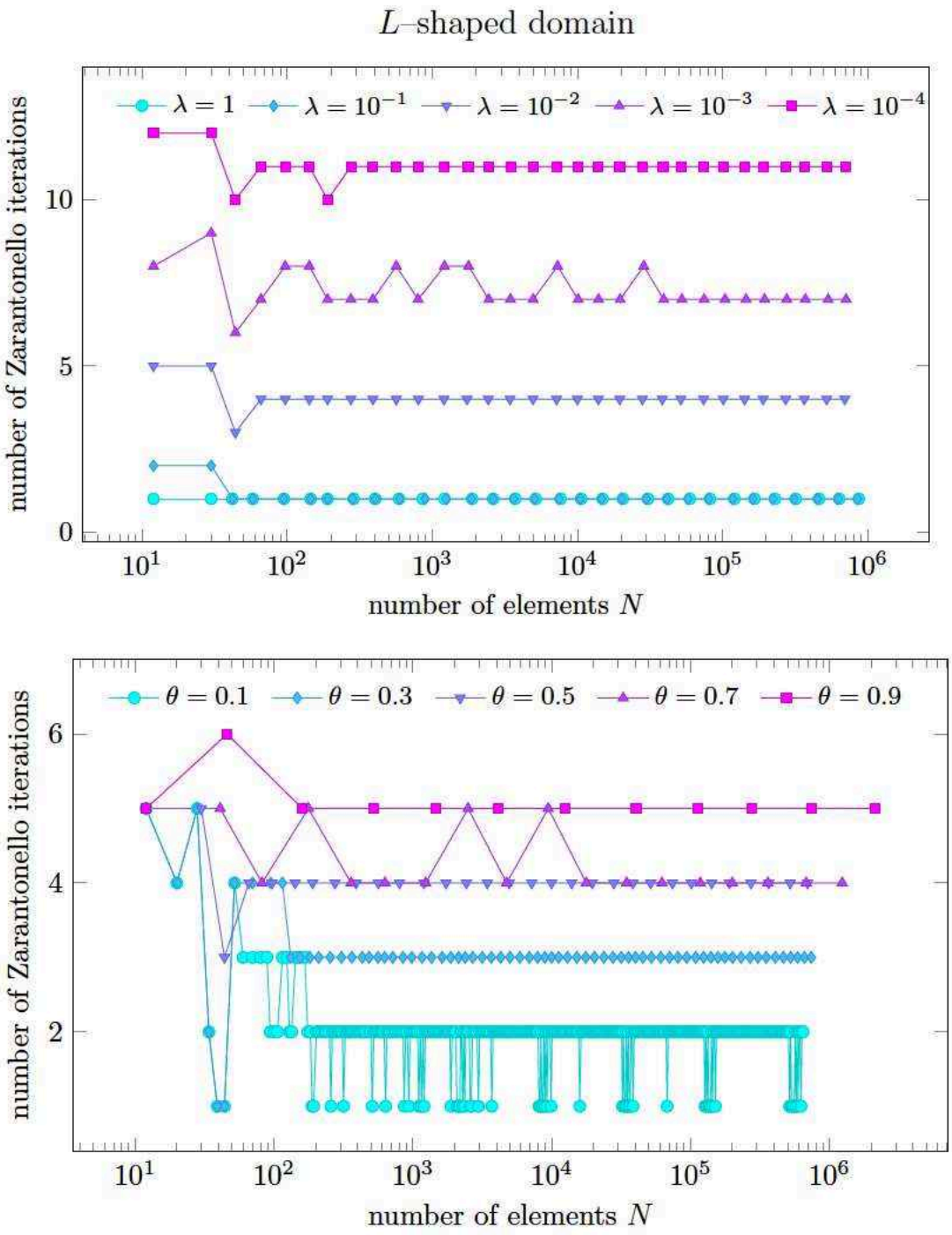


Figure 4.14: Example from Section 4.8.1 (Homogeneous problem on L-shaped domain): Number of Zarantonello iterations with respect to the number of elements N for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

θ \ λ	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
0.05	1.3e+07	1.3e+07	1.3e+07	1.3e+07	1.3e+07	2.8e+07	5.4e+07	8.0e+07	**1.0e+08**
0.1	3.8e+06	3.8e+06	3.8e+06	3.8e+06	7.2e+06	1.2e+07	1.9e+07	2.7e+07	3.4e+07
0.15	1.8e+06	1.8e+06	1.8e+06	1.9e+06	3.7e+06	7.5e+06	1.1e+07	1.3e+07	1.7e+07
0.2	1.2e+06	1.2e+06	1.2e+06	1.2e+06	2.9e+06	4.6e+06	6.8e+06	9.1e+06	1.1e+07
0.25	7.8e+05	7.8e+05	7.8e+05	8.8e+05	2.3e+06	3.9e+06	5.0e+06	6.6e+06	8.2e+06
0.3	6.5e+05	6.5e+05	6.5e+05	9.0e+05	1.9e+06	3.2e+06	4.3e+06	5.1e+06	6.5e+06
0.35	5.5e+05	5.5e+05	5.5e+05	9.2e+05	1.4e+06	2.5e+06	3.4e+06	4.3e+06	5.0e+06
0.4	3.8e+05	3.8e+05	4.0e+05	7.5e+05	1.4e+06	2.2e+06	3.1e+06	4.0e+06	4.0e+06
0.45	3.5e+05	3.5e+05	3.3e+05	6.4e+05	1.4e+06	1.9e+06	2.6e+06	3.4e+06	4.2e+06
0.5	2.9e+05	2.9e+05	2.8e+05	7.0e+05	1.4e+06	2.1e+06	2.6e+06	3.3e+06	4.0e+06
0.55	2.7e+05	2.7e+05	2.4e+05	6.5e+05	1.3e+06	2.0e+06	1.7e+06	2.1e+06	2.5e+06
0.6	3.3e+05	3.3e+05	3.0e+05	5.5e+05	8.5e+05	1.3e+06	1.7e+06	1.9e+06	2.4e+06
0.65	1.9e+05	1.9e+05	**1.9e+05**	6.7e+05	9.5e+05	1.4e+06	1.8e+06	2.3e+06	2.6e+06
0.7	2.9e+05	2.9e+05	4.3e+05	7.3e+05	1.0e+06	1.0e+06	1.4e+06	1.7e+06	2.0e+06
0.75	3.2e+05	3.2e+05	2.6e+05	5.1e+05	7.2e+05	9.9e+05	1.3e+06	1.7e+06	2.0e+06
0.8	2.0e+05	2.0e+05	4.1e+05	5.7e+05	8.2e+05	9.8e+05	1.3e+06	1.6e+06	1.9e+06
0.85	2.1e+05	2.1e+05	3.6e+05	5.5e+05	7.7e+05	1.1e+06	1.2e+06	1.5e+06	1.8e+06
0.9	2.1e+05	2.1e+05	4.2e+05	5.7e+05	8.6e+05	2.6e+06	3.4e+06	4.2e+06	5.1e+06
0.95	4.3e+05	4.3e+05	9.9e+05	1.6e+06	2.5e+06	2.9e+06	3.8e+06	4.8e+06	5.7e+06

min **max**

Figure 4.15: Example from Section 4.8.1 (Homogeneous problem on L-shaped domain): Overall computational cost $\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'}$ such that $\eta_\ell(u_\ell^{\underline{k}}) < \tau$ for given precision $\tau = 10^{-2}$, $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$, and $\theta \in \{0.05, 0.1, \dots, 0.95\}$.

choice is $\lambda = 1$ and $\theta = 0.75$. For the overall computational cost it then holds that

$$\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'} = 1531423,$$

where $u_\ell^{\underline{k}}$ is the first approximation such that $\eta_\ell(u_\ell^{\underline{k}}) < 10^{-2}$.

Experiment with known solution on Z-shaped domain

We consider the Z-shaped domain $\Omega \subset \mathbb{R}^2$ from Figure 4.11 (top) and the boundary value problem (4.123)

$$\begin{aligned} -\operatorname{div}\left(\mu(x,|\nabla u^\star(x)|^2)\nabla u^\star(x)\right) &= f(x) && \text{in } \Omega,\\ u^\star(x) &= 0 && \text{on } \Gamma_D,\\ \mu(x,|\nabla u^\star(x)|^2)\,\partial_{\mathbf{n}} u^\star(x) &= g(x) && \text{on } \Gamma_N, \end{aligned}$$

where the scalar nonlinearity $\mu\colon \Omega \times \mathbb{R}_{\ge 0} \to \mathbb{R}$ is defined by

$$\mu(x,t) := 1 + \frac{1}{\sqrt{1+t}}. \tag{4.132}$$

This leads to (N1)–(N4) with $\alpha = \widetilde{\gamma}_1 = 2$ and $L = \widetilde{\gamma}_2 = 3$.

We prescribe the solution $u^\star$ in polar coordinates $(x,y) = r(\cos\phi, \sin\phi)$ with $\phi \in (-\pi,\pi)$ by

$$u^\star(x,y) = r^\beta \cos(\beta\,\phi), \tag{4.133}$$

where $\beta = 4/7$ and compute f and g in (4.123) accordingly. We note that $u^\star$ has a generic singularity at the re-entrant corner $(x,y) = (0,0)$.

In Figure 4.16, we compare Algorithm 15 for different values of θ and λ, and uniform mesh-refinement. To this end, the error estimator $\eta_\ell(u_\ell^{\underline{k}})$ of the last step of the Zarantonello iteration is plotted over the number of elements. We see that uniform mesh-refinement leads to the suboptimal rate of convergence $\mathcal{O}(N^{-2/7})$ for the Z–shaped domain. Algorithm 15 regains the optimal rate of convergence $\mathcal{O}(N^{-1/2})$, independently of the actual choice of $\theta \in \{0.1, 0.3, \dots, 0.9\}$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$. Since $\eta_\ell(u_\ell^{\underline{k}}) \simeq \Delta_\ell^{\underline{k}}$, this once again empirically underpins Theorem 23.

In Figure 4.17, we plot the estimator $\eta_\ell(u_\ell^{\underline{k}})$ of the last step of the Zarantonello iteration over the cumulative sum $\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'}$. As predicted in Theorem 23, we observe that Algorithm 15 regains the optimal order of convergence $\mathcal{O}\big(\big(\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'}\big)^{-1/2}\big)$ with respect to the computational complexity, while the rate seems to be independent of the values of λ or θ.

In Figure 4.18, we take a look at the number of Zarantonello iterations. As in Figure 4.14, we observe that that a larger value of λ or a smaller value of θ lead to less iterations, while the number stays uniformly bounded in each case.

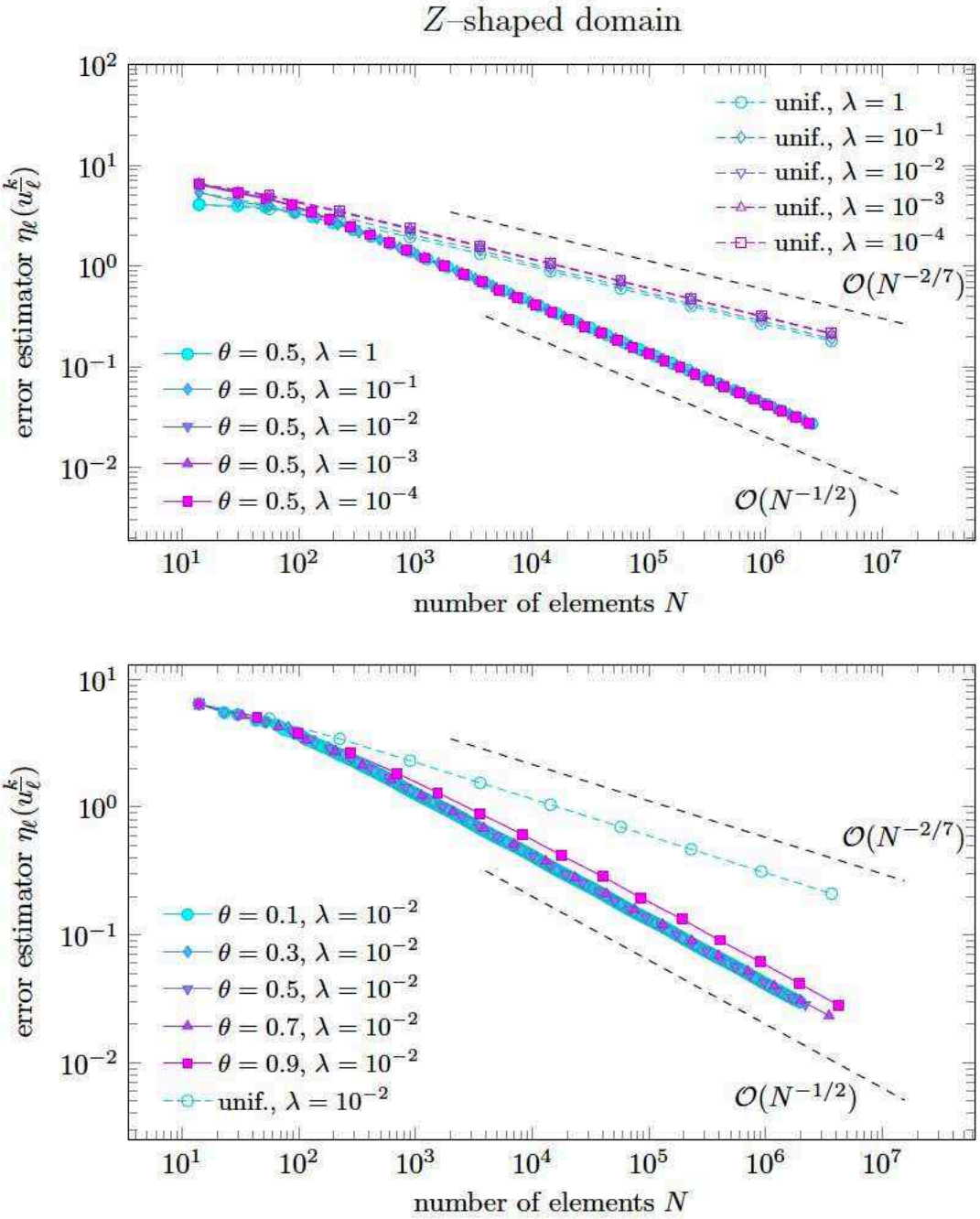


Figure 4.16: Example from Section 4.8.1 (Experiment with known solution on Z-shaped domain): Error estimator $\eta_\ell(u_\ell^k)$ of the last step of the Zarantonello iteration with respect to the number of elements N of the mesh $\mathcal{T}_\ell$ for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

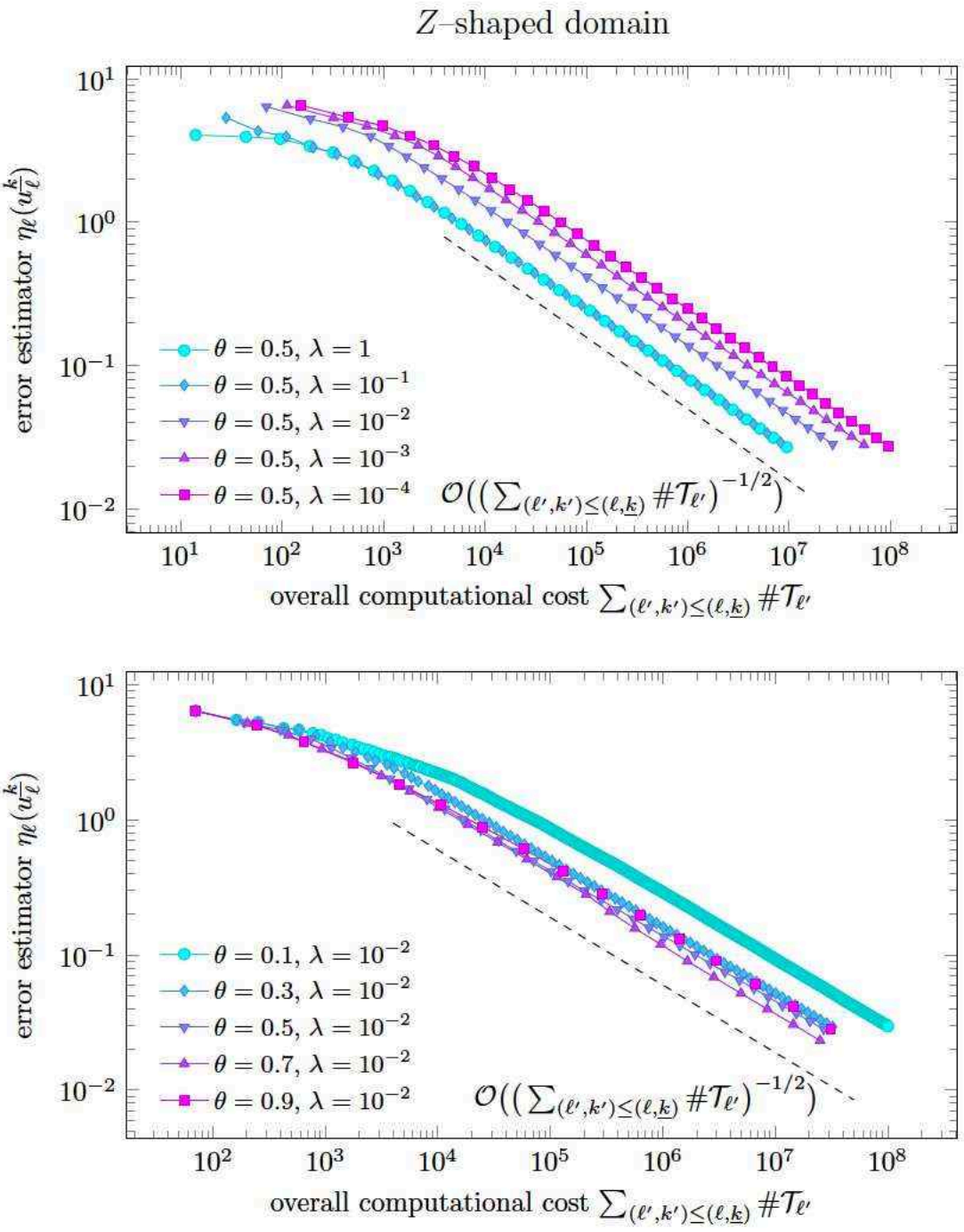


Figure 4.17: Example from Section 4.8.1 (Experiment with known solution on Z-shaped domain): Error estimator $\eta_\ell(u_\ell^k)$ of the last step of the Zarantonello iteration with respect to the overall computational cost expressed as the cumulative sum $\sum_{(\ell',k')\le(\ell,k)} \#\mathcal{T}_{\ell'}$ for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

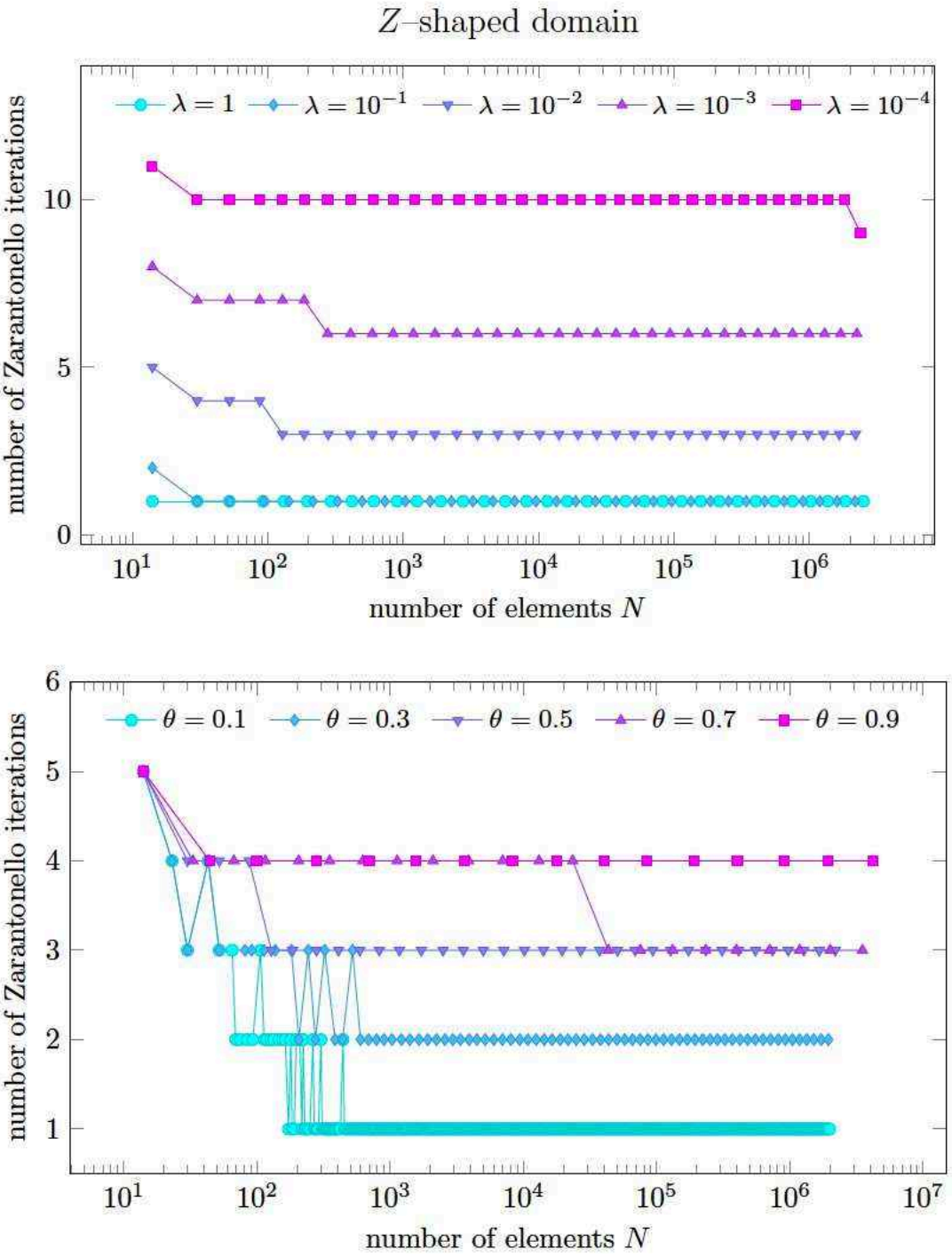


Figure 4.18: Example from Section 4.8.1 (Experiment with known solution on Z-shaped domain): Number of Zarantonello iterations with respect to the number of elements N for $\theta = 0.5$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (top) as well as for $\lambda = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

θ \ λ	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
0.05	3.7e+08	3.7e+08	3.7e+08	3.7e+08	3.7e+08	7.4e+08	1.1e+09	1.9e+09	**2.3e+09**
0.1	1.0e+08	1.0e+08	1.0e+08	1.0e+08	1.0e+08	2.0e+08	4.1e+08	6.1e+08	7.1e+08
0.15	4.9e+07	4.9e+07	4.9e+07	4.9e+07	7.3e+07	1.5e+08	2.5e+08	3.0e+08	3.9e+08
0.2	3.0e+07	3.0e+07	3.0e+07	3.0e+07	6.0e+07	8.9e+07	1.5e+08	2.2e+08	2.4e+08
0.25	2.1e+07	2.1e+07	2.1e+07	2.1e+07	4.5e+07	8.4e+07	1.1e+08	1.5e+08	2.0e+08
0.3	1.7e+07	1.7e+07	1.7e+07	1.8e+07	3.2e+07	6.9e+07	7.8e+07	1.2e+08	1.6e+08
0.35	1.4e+07	1.4e+07	1.4e+07	1.3e+07	2.6e+07	5.3e+07	8.1e+07	9.4e+07	1.2e+08
0.4	1.2e+07	1.2e+07	1.2e+07	1.3e+07	2.8e+07	4.5e+07	6.8e+07	8.5e+07	1.0e+08
0.45	1.0e+07	1.0e+07	1.0e+07	1.0e+07	3.3e+07	4.5e+07	5.5e+07	7.3e+07	8.3e+07
0.5	9.5e+06	9.5e+06	8.2e+06	1.1e+07	2.7e+07	3.7e+07	5.5e+07	7.6e+07	9.5e+07
0.55	8.2e+06	8.2e+06	8.3e+06	1.1e+07	2.8e+07	4.3e+07	5.6e+07	7.5e+07	9.4e+07
0.6	6.7e+06	6.7e+06	6.3e+06	1.6e+07	2.3e+07	3.9e+07	4.7e+07	6.3e+07	7.9e+07
0.65	8.5e+06	8.5e+06	8.5e+06	1.1e+07	2.3e+07	3.8e+07	4.6e+07	5.7e+07	7.1e+07
0.7	7.9e+06	7.9e+06	7.9e+06	1.3e+07	2.5e+07	4.0e+07	5.7e+07	6.8e+07	8.0e+07
0.75	**5.4e+06**	5.4e+06	5.5e+06	1.1e+07	2.5e+07	3.6e+07	4.9e+07	6.2e+07	7.6e+07
0.8	7.2e+06	7.2e+06	7.4e+06	1.4e+07	2.1e+07	2.5e+07	3.4e+07	4.4e+07	5.3e+07
0.85	7.4e+06	7.4e+06	7.4e+06	1.9e+07	2.6e+07	3.4e+07	4.4e+07	5.3e+07	6.0e+07
0.9	1.6e+07	1.6e+07	1.5e+07	2.6e+07	3.1e+07	4.3e+07	5.6e+07	7.0e+07	8.4e+07
0.95	1.8e+07	1.8e+07	1.7e+07	7.2e+07	1.0e+08	1.2e+08	1.6e+08	1.9e+08	2.3e+08

min **max**

Figure 4.19: Example from Section 4.8.1 (Experiment with known solution on Z-shaped domain): Overall computational cost $\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'}$ such that $\eta_\ell(u_\ell^{\underline{k}}) < \tau$ for given precision $\tau = 3 \cdot 10^{-2}$, $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$, and $\theta \in \{0.05, 0.1, \dots, 0.95\}$.

In Figure 4.19, we compare the computational cost to reach the precision $\tau = 3 \cdot 10^{-2}$ for $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$ and $\theta \in \{0.05, 0.1, \dots, 0.95\}$. As a result, we get that the best choice is $\lambda = 1$ and $\theta = 0.75$. For the overall computational cost it then holds that

$$\sum_{(\ell',k')\le(\ell,\underline{k})} \#\mathcal{T}_{\ell'} = 5439636,$$

where $u_\ell^{\underline{k}}$ is the first approximation such that $\eta_\ell(u_\ell^{\underline{k}}) < 3 \cdot 10^{-2}$.

5 Fully adaptive algorithm for AFEM for nonlinear operators

5.1 Introduction

In Chapter 4, we considered adaptive finite element methods for second-order elliptic PDEs where the arising discrete systems are not solved exactly. We showed that both AFEM for linear elliptic PDEs in combination with an optimal PCG solver for the Galerkin system, cf. Section 4.7, as well as AFEM for certain nonlinear elliptic PDEs in combination with the Zarantonello iteration, cf. Section 4.8, fit in the abstract framework of Algorithm 15. The idea of this chapter, which is based on [HPSV21], is to combine these two settings into one fully adaptive algorithm.

Let $\Omega \subset \mathbb{R}^d$ with $d \geq 1$ be a bounded Lipschitz domain with polytopal boundary. Given $f \in L^2(\Omega)$ and a nonlinear operator $A\colon \mathbb{R}^d \to \mathbb{R}^d$, we then aim to numerically approximate the weak solution $u^\star \in H^1_0(\Omega)$ of the nonlinear boundary value problem

$$\begin{aligned} -\operatorname{div} A(\nabla u^\star) &= f \quad \text{in } \Omega, \\ u^\star &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{5.1}$$

To this end, we propose an adaptive algorithm of the type

$$\boxed{\text{estimate total error and its components}} \\ \downarrow \\ \boxed{\text{advance algebra/advance linearization/mark and refine mesh elements}} \tag{5.2}$$

which monitors and adequately stops the iterative linearization and the linear algebraic solver as well as steers the local mesh-refinement. The goal of this chapter is to perform a rigorous mathematical analysis of this algorithm in terms of convergence and quasi-optimal computational cost.

5.1.1 Finite element approximation and Banach–Picard iteration

Suppose that the nonlinearity A in (5.1) is Lipschitz-continuous (with constant $L > 0$) and strongly monotone (with constant $\alpha > 0$), see Section 5.2 for details. Then, the main theorem on monotone operators yields the existence and uniqueness of the weak solution $u^\star \in H^1_0(\Omega)$, see, e.g., [Zei90, Theorem 25.B]. Given a triangulation $\mathcal{T}_\bullet$ of Ω, the lowest-order finite element method (FEM) for problem (5.1) reads as follows: Find $u^\star_\bullet \in \mathcal{X}_\bullet := \{v_\bullet \in C(\overline{\Omega}) : v_\bullet|_T \text{ is affine for all } T \in \mathcal{T}_\bullet \text{ and } v_\bullet|_{\partial\Omega} = 0\} \subset H^1_0(\Omega)$ such that

$$\langle A(\nabla u^\star_\bullet), \nabla v_\bullet\rangle_\Omega = \langle f, v_\bullet\rangle_\Omega \quad \text{for all } v_\bullet \in \mathcal{X}_\bullet. \tag{5.3}$$

The discrete solution $u_\bullet^\star \in \mathcal{X}_\bullet$ again exists and is unique, but (5.3) corresponds to a *nonlinear discrete system* which can typically only be solved *inexactly*.

The most straightforward algorithm for *iterative linearization* of (5.3) stems from the proof of the main theorem on monotone operators which is constructive and relies on the Banach fixed point theorem: Define the (nonlinear) operator $\Phi_\bullet : \mathcal{X}_\bullet \to \mathcal{X}_\bullet$ by

$$\langle \nabla \Phi_\bullet(w_\bullet) \,,\, \nabla v_\bullet \rangle_\Omega = \langle \nabla w_\bullet \,,\, \nabla v_\bullet \rangle_\Omega - \frac{\alpha}{L^2} \big[\langle A(\nabla w_\bullet) \,,\, \nabla v_\bullet \rangle_\Omega - \langle f \,,\, v_\bullet \rangle_\Omega \big] \tag{5.4}$$

for all $w_\bullet, v_\bullet \in \mathcal{X}_\bullet$. Note that (5.4) corresponds to a discrete Poisson problem and hence $\Phi_\bullet(w_\bullet) \in \mathcal{X}_H$ is well-defined. Then, it holds that

$$\| \nabla(u_\bullet^\star - \Phi_\bullet(w_\bullet)) \|_{L^2(\Omega)} \le q_{\mathrm{Pic}} \, \| \nabla(u_\bullet^\star - w_\bullet) \|_{L^2(\Omega)} \tag{5.5}$$

with

$$q_{\mathrm{Pic}} := (1 - \alpha^2/L^2)^{1/2} < 1,$$

see, e.g., [Zei90, Section 25.4]. Based on the *contraction* $\Phi_\bullet$, the Banach–Picard iteration starts from an arbitrary discrete initial guess and applies $\Phi_\bullet$ inductively to generate a sequence of discrete functions which hence converge towards $u_\bullet^\star$. Note that the computation of $\Phi_\bullet(w_h)$ by means of the discrete variational formulation (5.4) corresponds to the solution of a (generically large) *linear discrete system* with symmetric and positive definite matrix that does not change during the iterations. As mentioned before, we now suppose that also (5.4) is solved *inexactly* by means of a *contractive iterative algebraic solver* (with contraction factor $q_{\mathrm{alg}} < 1$), e.g., PCG with optimal preconditioner, see, e.g., [OT14].

5.1.2 Fully adaptive algorithm

In our approach, we compute a *sequence* of discrete approximations $u_\ell^{k,j}$ of $u^\star$ that have an index ℓ for the *mesh-refinement*, an index k for the Banach–Picard *linearization* iteration, and an index j for the *algebraic solver* iteration.

First, we design a stopping criterion for the algebraic solver such that, at linearization step $k-1 \in \mathbb{N}_0$ on the mesh $\mathcal{T}_\ell$, we stop for some index $\underline{j} \in \mathbb{N}$. At the next linearization step $k \in \mathbb{N}$, the arising linear system reads as follows:

$$\begin{aligned} &\text{Find } u_\ell^{k,\star} \in \mathcal{X}_\ell \text{ such that, for all } v_\ell \in \mathcal{X}_\ell, \\ &\langle \nabla u_\ell^{k,\star} \,,\, \nabla v_\ell \rangle_\Omega = \langle \nabla u_\ell^{k-1,\underline{j}} \,,\, \nabla v_\ell \rangle_\Omega - \frac{\alpha}{L^2} \big[\langle A(\nabla u_\ell^{k-1,\underline{j}}) \,,\, \nabla v_\ell \rangle_\Omega - \langle f \,,\, v_\ell \rangle_\Omega \big], \end{aligned} \tag{5.6}$$

with uniquely defined but not computed exact solution $u_\ell^{k,\star} = \Phi_\ell(u_\ell^{k-1,\underline{j}})$ and computed iterates $u_\ell^{k,j}$ that approximate $u_\ell^{k,\star}$. Note that (5.6) is a *perturbed* Banach–Picard iteration since it starts from the available $u_\ell^{k-1,\underline{j}}$, typically not equal to the unavailable $u_\ell^{k-1,\star}$.

Second, we design a stopping criterion for the perturbed Banach–Picard iteration at some index $\underline{k}$, producing a discrete approximation $u_\ell^{\underline{k},\underline{j}}$.

Finally, we locally refine the triangulation $\mathcal{T}_\ell$ on the basis of the Dörfler marking criterion for the local contributions of the residual error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{j}})$, and, to lower the computational effort, employ nested iteration in that the continuation on the new triangulation $\mathcal{T}_{\ell+1}$ is started with the initial guess $u_{\ell+1}^{0,0} := u_\ell^{\underline{k},\underline{j}}$.

5.1.3 State of the art

Solving the linear system (5.6) inexactly gives rise to the so-called "inexact Newton method", see, e.g., in [Deu91, EW94] and the references therein. Under appropriate conditions, these can asymptotically preserve the convergence speed of the "exact" Newton method. However, these approaches only focus on the finite-dimensional system of nonlinear algebraic equations of the form (5.3) but do not take into account the continuous problem (5.1), which is our central issue here.

Solving the nonlinear algebraic systems (5.3) "exactly" (up to machine precision), only the *discretization* error is left. Then, convergence and optimal decay rates of the error $\|\nabla(u^\star - u_\bullet^\star)\|_{L^2(\Omega)}$ with respect to the degrees of freedom of FEM adapting the approximation space (mesh) were obtained in [Vee02, DK08, BDK12, GMZ12], following the seminal contributions [Dör96, MNS00, BDD04, Ste07, CKNS08] for linear problems. We also refer to [CFPP14] for a general framework of convergence of adaptive FEM with optimal convergence rates.

Solving only the linear algebraic systems (5.6) "exactly" but (5.3) inexactly leaves the *discretization* and *linearization* errors. Such a setting has been considered in, e.g., [CS07, EAEV11], where reliable (guaranteed) and efficient *a posteriori* error estimates were derived. Adaptive algorithms aiming at a balance of the linearization and discretization errors were proposed and their optimal performance was observed numerically, see, e.g., [BDMS15, BCL15, CW17, HW18]. Later, theoretical proofs of *plain convergence* (without rates) were given in [GMZ11, HW20b], where [HW20b] builds on the unified framework of [HW20a] encompassing also the Kačanov and (damped) Newton linearizations in addition to the Banach–Picard linearization (5.6).

The works [GHPS18, GHPS21], cf. Chapter 4, considered that the linear systems (5.6) are solved exactly at *linear cost* (so that $u_\ell^{k,\underline{j}} = u_\ell^{k,\star}$ with $\underline{j}(\ell,k) = \mathcal{O}(1)$ in the present notation), as in the seminal work [Ste07] for the Poisson model problem and in [CG12] for an adaptive Laplace eigenvalue computation. Under this so-called realistic assumption on the algebraic solver, [GHPS18] proved that the overall strategy leads to optimal convergence *rates* with respect to the *number of degrees of freedom* as well as to *almost optimal* convergence rates with respect to the *overall computational cost*. The latter means that, if the total error converges with rate $s > 0$ with respect to the degrees of freedom, then, for all $\varepsilon > 0$, it also converges with rate $s - \varepsilon > 0$ with respect to the overall computational cost. The proof of [GHPS18] was based on proving first that the estimator $\eta_\ell(u_\ell^{\underline{k},\star})$ for the final Picard iterates decays with optimal rate s and second that the number of Picard iterates satisfies $\underline{k}(\ell) \lesssim 1 + \log[1 + \eta_\ell(u_{\ell+1}^{\underline{k},\star})/\eta_\ell(u_\ell^{\underline{k},\star})]$. This logarithmic bound then led to the bound $s - \varepsilon$ for the convergence rate with respect to the overall computational cost.

As shown in Chapter 4, we have improved the latter result in [GHPS21] and proved optimal computational cost (i.e., $\varepsilon = 0$), still relying on the assumption that the discrete Poisson problem (5.6) is solved exactly at linear cost. The core idea of the new proof follows ideas from adaptive Uzawa FEM for the Stokes model problem [KS08, DFFGP19]. However, besides the nonlinearity, the structural difference is that the adaptive Uzawa FEM employs an outer iteration on the continuous level (i.e., we first linearize and then discretize), while the approach of [CW17, GHPS18, HW20a, HW20b, GHPS21] is first to discretize and then to linearize.

As in the present setting, the "adaptive inexact Newton method" in [EV13] takes into account all *discretization*, *linearization*, and *algebraic* error components, see also [CPV14, DPVY15] and [Pol16] for regularizations on coarse meshes ensuring well-posedness of the discrete systems in Newton-like linearizations. The goal of this chapter is to perform a rigorous mathematical analysis of such algorithms in terms of convergence and optimal decay rate of the error with respect to computational cost.

We stress that such results have already been derived for adaptive wavelet discretizations [CDD03, Ste14] which provide inherent control of the residual error in terms of the wavelet coefficients, while the present analysis for standard finite element discretizations has to rely on the local information of appropriate *a posteriori* error estimators. Also, while the present analysis is closely related to that of [GHPS21], we stress that both works [GHPS18, GHPS21] focused only on linearization and discretization, while here, we also include the innermost algebraic loop into the adaptive algorithm. In particular, the technical challenges in the present analysis are much more involved than in [GHPS21] due to the coupling of the two nested inexact solvers.

5.1.4 Main results and outline

Similarly to Chapter 4, the sequential nature of the fully adaptive algorithm of Section 5.1.2 gives rise to an index set

$$\mathcal{Q} := \big\{(\ell,k,j) \in \mathbb{N}_0^3 : \text{discrete approximation } u_\ell^{k,j} \text{ is computed by the algorithm}\big\}$$

together with an ordering

$$|(\ell,k,j)| < |(\ell',k',j')| \quad \overset{\text{def}}{\iff} \quad u_\ell^{k,j} \text{ is computed earlier than } u_{\ell'}^{k',j'}.$$

Our first main result, formulated in Theorem 45 below, proves that the proposed adaptive strategy is *contractive* after some amount of steps and *linearly convergent* in the sense of

$$\Delta_{\ell'}^{k',j'} \le C_{\mathrm{lin}} q_{\mathrm{lin}}^{|(\ell',k',j')|-|(\ell,k,j)|}\, \Delta_\ell^{k,j} \quad \text{for all } |(\ell,k,j)| \le |(\ell',k',j')|, \tag{5.7}$$

where $C_{\mathrm{lin}} \ge 1$ and $0 < q_{\mathrm{lin}} < 1$ are generic constants and $\Delta_\ell^{k,j}$ is an appropriate quasi-error quantity involving the error $\|\nabla(u^\star - u_\ell^{k,j})\|_{L^2(\Omega)}$ as well as the error estimator $\eta_\ell(u_\ell^{k,j})$. Second, we prove the *optimal error decay rate* with respect to the number of *degrees of freedom* added with respect to the initial mesh in the sense that

$$\sup_{(\ell,k,j)\in\mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \Delta_\ell^{k,j} < \infty \tag{5.8}$$

whenever $u^\star$ is approximable at algebraic rate $s > 0$, see Theorem 49 below for the details. Finally, estimate (5.7) appears to be also the key argument to prove our most eminent result, namely the *optimal error decay rate* with respect to the *overall computational cost* of the fully adaptive algorithm which steers the mesh-refinement, the perturbed Banach–Picard linearization, and the algebraic solver. In short, this reads

$$\sup_{(\ell',k',j')\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\ (\ell,k,j)\le(\ell',k',j')}} \#\mathcal{T}_\ell \Bigg)^s \Delta_{\ell'}^{k',j'} < \infty \tag{5.9}$$

whenever $u^\star$ is approximable at algebraic rate $s > 0$; see Theorem 53 below for the details. We stress that under realistic assumptions the sum in (5.9) is indeed proportional to the overall computational cost invested into the fully adaptive numerical approximation of (5.1), if the cost of all procedures like matrix and right-hand-side assembly, one algebraic solver step, evaluation of the involved a posteriori error estimates, marking, and local adaptive mesh refinement is proportional to the number of mesh elements in $\mathcal{T}_\ell$ (i.e., the number of degrees of freedom).

The remainder of this section is organised as follows. In Section 5.2, we introduce the abstract setting for our algorithm as well as the requirements on mesh-refinement, error estimator, and algebraic solver, before we state the fully adaptive algorithm in Section 5.2.5. In Section 5.3, we then state the aforementioned main results, i.e., linear convergence of the quasi-error in each step of the adaptive algorithm (Section 5.3.4), optimal convergence rates of the quasi-error with respect to the number of degrees of freedom (Section 5.3.6), as well as optimal convergence rates of the quasi-error with respect to the overall computational cost of the fully adaptive algorithm (Section 5.3.7). Finally, numerical experiments in Section 5.4 underline the theoretical findings.

5.2 Adaptive algorithm

In this section, we introduce an abstract setting, in which all our results will be formulated, define the exact weak and finite elements solutions, introduce our requirements on mesh-refinement, error estimator, and algebraic solver, state our adaptive algorithm, and present our main results, including some discussions.

5.2.1 Abstract setting

Let $\mathcal{H}$ be a Hilbert space over $\mathbb{K} \in \{\mathbb{R}, \mathbb{C}\}$ with scalar product $\langle\!\langle \cdot\,, \cdot \rangle\!\rangle$, corresponding norm $|\!|\!| \cdot |\!|\!|$, and dual space $\mathcal{H}'$ (with canonical operator norm $|\!|\!| \cdot |\!|\!|'$). Let the operator $\mathcal{A}\colon \mathcal{H} \to \mathcal{H}'$ satisfy (O1)–(O3) from Section 4.2 with potential $P\colon \mathcal{H} \to \mathbb{K}$, i.e., we suppose that the operator $\mathcal{A}$ is *strongly monotone* and *Lipschitz-continuous*, i.e.,

$$\alpha\, |\!|\!| w - v |\!|\!|^2 \le \operatorname{Re} \langle \mathcal{A}w - \mathcal{A}v\,,\, w - v \rangle_{\mathcal{H}'\times\mathcal{H}} \quad \text{and} \quad |\!|\!| \mathcal{A}w - \mathcal{A}v |\!|\!|' \le L\, |\!|\!| w - v |\!|\!| \tag{5.10}$$

for all $v, w \in \mathcal{H}$, where $0 < \alpha \le L$ are generic real constants and P is Gâteaux-differentiable with derivative $\mathcal{A} := \mathrm{d}P\colon \mathcal{H} \to \mathcal{H}'$, i.e., there holds that

$$\langle \mathcal{A}w\,,\, v \rangle_{\mathcal{H}'\times\mathcal{H}} = \lim_{\substack{t\to 0\\ t\in\mathbb{R}}} \frac{P(w+tv) - P(w)}{t} \quad \text{for all } v, w \in \mathcal{H}.$$

Given a linear and continuous functional $F \in \mathcal{H}'$, the main theorem on monotone operators [Zei90, Section 25.4] yields existence and uniqueness of the solution $u^\star \in \mathcal{H}$ of

$$\langle \mathcal{A}u^\star\,,\, v \rangle_{\mathcal{H}'\times\mathcal{H}} = F(v) \quad \text{for all } v \in \mathcal{H}. \tag{5.11}$$

The result actually holds true for any closed subspace $\mathcal{X}_\bullet \subseteq \mathcal{H}$, which also gives rise to a unique $u^\star_\bullet \in \mathcal{X}_\bullet$ such that

$$\langle \mathcal{A}u^\star_\bullet\,,\, v_\bullet \rangle_{\mathcal{H}'\times\mathcal{H}} = F(v_\bullet) \quad \text{for all } v_\bullet \in \mathcal{X}_\bullet. \tag{5.12}$$

Finally, with the *energy functional* $\mathcal{E} := \operatorname{Re}(P - F)$, it holds that

$$\frac{\alpha}{2} \|\!| v_\bullet - u_\bullet^\star |\!\|^2 \le \mathcal{E}(v_\bullet) - \mathcal{E}(u_\bullet^\star) \le \frac{L}{2} \|\!| v_\bullet - u_\bullet^\star |\!\|^2 \quad \text{for all } v_\bullet \in \mathcal{X}_\bullet, \tag{5.13}$$

see, e.g., [GHPS18, Lemma 5.1]. In particular, $u^\star \in \mathcal{H}$ is the unique minimizer of the minimization problem

$$\mathcal{E}(u^\star) = \min_{v \in \mathcal{H}} \mathcal{E}(v) \tag{5.14}$$

as well as $u_\bullet^\star \in \mathcal{X}_\bullet^\star$ is the unique minimizer of the minimization problem

$$\mathcal{E}(u_\bullet^\star) = \min_{v_\bullet \in \mathcal{X}_\bullet} \mathcal{E}(v_\bullet). \tag{5.15}$$

As in Section 4.2, it follows from (5.10)–(5.12) that the present setting guarantees the Céa lemma

$$\|\!| u^\star - u_\bullet^\star |\!\| \le C_{\text{Céa}} \|\!| u^\star - v_\bullet |\!\| \quad \text{for all } v_\bullet \in \mathcal{X}_\bullet \quad \text{with} \quad C_{\text{Céa}} := L/\alpha. \tag{5.16}$$

5.2.2 Mesh-refinement

We briefly recall some definitions of the mesh-refinement from Section 3.4. Let $\mathcal{T}_\bullet$ be a conforming simplicial mesh of Ω, i.e., a partition of $\overline{\Omega}$ into compact simplices T such that $\bigcup_{T \in \mathcal{T}_\bullet} T = \overline{\Omega}$ and such that the intersection of two different simplices is either empty or their common vertex, edge, or face.

We assume that $\texttt{refine}(\cdot)$ is a fixed mesh-refinement strategy, e.g., newest vertex bisection, cf. Section 3.6.

We write $\mathcal{T}_\circ = \texttt{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$ for the coarsest one-level refinement of $\mathcal{T}_\bullet$, where all marked elements $\mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet$ have been refined, i.e., $\mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet \backslash \mathcal{T}_\circ$. We write $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$, if $\mathcal{T}_\circ$ can be obtained by finitely many steps of one-level refinement (with appropriate, yet arbitrary marked elements in each step). We define $\mathbb{T} := \texttt{refine}(\mathcal{T}_0)$ as the set of all meshes which can be generated from the initial simplicial mesh $\mathcal{T}_0$ of Ω by use of $\texttt{refine}(\cdot)$.

Finally, we associate to each $\mathcal{T}_\bullet \in \mathbb{T}$ a corresponding finite-dimensional subspace $\mathcal{X}_\bullet \subsetneq \mathcal{H}$, where we suppose that $\mathcal{X}_\bullet \subseteq \mathcal{X}_\circ$ whenever $\mathcal{T}_\bullet, \mathcal{T}_\circ \in \mathbb{T}$ with $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$.

For newest vertex bisection, we refer to Section 3.6 for the validity of (R1)–(R3) as well as Section 3.7 for other refinement strategies.

5.2.3 Error estimator

For each mesh $\mathcal{T}_\bullet \in \mathbb{T}$, suppose that we can compute refinement indicators

$$\eta_\bullet(T, v_\bullet) \ge 0 \quad \text{for all } T \in \mathcal{T}_\bullet \text{ and all } v_\bullet \in \mathcal{X}_\bullet. \tag{5.17}$$

We denote

$$\eta_\bullet(\mathcal{V}_\bullet, v_\bullet) := \Big(\sum_{T \in \mathcal{V}_\bullet} \eta_\bullet(T, v_\bullet)^2 \Big)^{1/2} \quad \text{for all } \mathcal{V}_\bullet \subseteq \mathcal{T}_\bullet \tag{5.18}$$

and abbreviate $\eta_\bullet(v_\bullet) := \eta_\bullet(\mathcal{T}_\bullet, v_\bullet)$. Analogously to Section 4.3, we assume the *axioms of adaptivity* (A1)–(A4) from [CFPP14] for all $\mathcal{T}_\bullet \in \mathbb{T}$ and all $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$ with generic constants $C_{\mathrm{stab}}, C_{\mathrm{rel}} > 0$, and $0 < q_{\mathrm{red}} < 1$. We stress that the exact discrete solutions $u_\bullet^\star$ (and $u_\circ^\star$ respectively) in (A3)–(A4) will never be computed but are only auxiliary quantities for the analysis.

We refer to Section 5.4 below for precise assumptions on the nonlinearity $A(\cdot)$ of problem (5.1) such that the standard residual error estimator satisfies (A1)–(A4) for lowest-order Courant finite elements, see also Section 5.4.1–5.4.2.

5.2.4 Algebraic solver

For given linear and continuous functionals $G \in \mathcal{H}'$, we consider linear systems of algebraic equations of the type

$$\langle\!\langle v_\bullet^\star, w_\bullet \rangle\!\rangle = G(w_\bullet) \quad \text{for all } w_\bullet \in \mathcal{X}_\bullet \tag{5.19}$$

with unique (but not computed) exact solution $v_\bullet^\star \in \mathcal{X}_\bullet$. We suppose here that we have at hand a contractive iterative algebraic solver for problems of the form (5.19). More precisely, let $v_\bullet^0 \in \mathcal{X}_\bullet$ be an initial guess and let the solver produce a sequence $v_\bullet^j \in \mathcal{X}_\bullet$, $j \geq 1$. Then, we suppose that there exists a generic constant $0 < q_{\mathrm{alg}} < 1$ such that

$$|\!|\!| v_\bullet^\star - v_\bullet^j |\!|\!| \leq q_{\mathrm{alg}} \, |\!|\!| v_\bullet^\star - v_\bullet^{j-1} |\!|\!| \quad \text{for all } j \geq 1. \tag{5.20}$$

Examples for such solvers are suitably preconditioned conjugate gradients or multigrid, see, e.g., Olshanskii and Tyrtyshnikov [OT14] and the references therein.

5.2.5 Adaptive algorithm

For the numerical approximation of problem (5.11), we consider an adaptive algorithm which steers mesh-refinement with index ℓ, a (perturbed) contractive Banach–Picard iteration with index k, and a contractive algebraic solver with index j. On each step (ℓ, k, j), it yields an approximation $u_\ell^{k,j} \in \mathcal{X}_\ell$ to the unique but unavailable $u_\ell^\star \in \mathcal{X}_\ell$ on the mesh $\mathcal{T}_\ell$ defined by

$$\langle \mathcal{A} u_\ell^\star, v_\ell \rangle_{\mathcal{H}' \times \mathcal{H}} = F(v_\ell) \quad \text{for all } v_\ell \in \mathcal{X}_\ell. \tag{5.21}$$

Reporting for the summary of notation to Table 5.1, the algorithm reads as follows:

Algorithm 41. ***Input:*** *Initial mesh $\mathcal{T}_0$ and initial guess $u_0^{0,0} = u_0^{0,j} \in \mathcal{X}_0$, parameters $0 < \theta \leq 1$, $0 < \lambda_{\mathrm{alg}} < 1$, $0 < \lambda_{\mathrm{Pic}}$, and $1 \leq C_{\mathrm{mark}}$, counters $\ell = k = j = 0$.*
Adaptive loop: *Iterate the following steps* (i)–(vi)*: (adaptive mesh-refinement loop)*

(i) **Repeat** *the following steps* (a)–(c)*: (linearization loop)*

(a) *Define $u_\ell^{k+1,0} := u_\ell^{k,j}$ and update counters $k := k+1$ as well as $j := 0$.*

(b) **Repeat** *the following steps* (I)–(III)*: (algebraic solver loop)*

	counter		discrete solution		
			available		unavailable
	running	stopping	running	stopping	exact
mesh	ℓ	$\underline{\ell}$	$u_\ell^{\underline{k},\underline{j}}$	$u_{\underline{\ell}}^{\underline{k},\underline{j}}$	$u_\ell^\star$ from (5.21)
linearization	k	$\underline{k}$	$u_\ell^{k,\underline{j}}$	$u_{\underline{\ell}}^{\underline{k},\underline{j}}$	$u_\ell^{k,\star}$ from (5.22)
algebraic solver	j	$\underline{j}$	$u_\ell^{k,j}$	$u_\ell^{k,\underline{j}}$	

Table 5.1: Counters and discrete solutions in Algorithm 41.

(I) *Update counter* $j := j + 1$.

(II) *Consider the problem of finding*

$$u_\ell^{k,\star} \in \mathcal{X}_\ell \text{ such that, for all } v_\ell \in \mathcal{X}_\ell,$$
$$\langle\!\langle u_\ell^{k,\star}, v_\ell \rangle\!\rangle = \langle\!\langle u_\ell^{k-1,\underline{j}}, v_\ell \rangle\!\rangle - \frac{\alpha}{L^2}\langle \mathcal{A}u_\ell^{k-1,\underline{j}} - F, v_\ell\rangle_{\mathcal{H}'\times\mathcal{H}} \tag{5.22}$$

and do one step of the algebraic solver applied to (5.22) *starting from* $u_\ell^{k,j-1}$, *which yields* $u_\ell^{k,j}$ *(an approximation to* $u_\ell^{k,\star}$*).*

(III) *Compute the local indicators* $\eta_\ell(T, u_\ell^{k,j})$ *for all* $T \in \mathcal{T}_\ell$.

Until $|||u_\ell^{k,j} - u_\ell^{k,j-1}||| \le \lambda_{\mathrm{alg}}\big[\eta_\ell(u_\ell^{k,j}) + |||u_\ell^{k,j} - u_\ell^{k-1,\underline{j}}|||\big]$. (5.23)

(c) *Define* $\underline{j} := \underline{j}(\ell, k) := j$.

Until $|||u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}}||| \le \lambda_{\mathrm{Pic}}\eta_\ell(u_\ell^{k,\underline{j}})$. (5.24)

(ii) *Define* $\underline{k} := \underline{k}(\ell) := k$.

(iii) *If* $\eta_\ell(u_\ell^{\underline{k},\underline{j}}) = 0$, *then set* $\underline{\ell} := \ell$ *and exit.*

(iv) *Determine a set* $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ *with up to the multiplicative constant* C_{mark} *minimal cardinality such that*

$$\theta\,\eta_\ell(u_\ell^{\underline{k},\underline{j}}) \le \eta_\ell(\mathcal{M}_\ell, u_\ell^{\underline{k},\underline{j}}). \tag{5.25}$$

(v) *Generate* $\mathcal{T}_{\ell+1} := \mathtt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ *and define* $u_{\ell+1}^{0,0} := u_{\ell+1}^{0,\underline{j}} := u_\ell^{\underline{k},\underline{j}}$.

(vi) *Update counters* $\ell := \ell + 1$, $k := 0$, *and* $j := 0$ *and continue with* (i).

Output: *Sequence of discrete solutions* $u_\ell^{k,j}$ *and corresponding error estimators* $\eta_\ell(u_\ell^{k,j})$.

Remark 42. *Some remarks in order to explain the nature of Algorithm 41:*

- *The innermost loop, Algorithm 41*(i)(b)*, steers the algebraic solver. Note that the exact solution* $u_\ell^{k,\star}$ *of* (5.22) *is not computed but only approximated by the computed iterates* $u_\ell^{k,j}$*. For the linear system* (5.22)*, the contraction assumption* (5.20) *reads as*

$$\|\!|u_\ell^{k,\star} - u_\ell^{k,j}|\!\| \le q_{\mathrm{alg}} \, |\!|\!|u_\ell^{k,\star} - u_\ell^{k,j-1}|\!\| \quad \text{for all } j \ge 1. \tag{5.26}$$

  Then, the triangle inequality implies that

$$\frac{1 - q_{\mathrm{alg}}}{q_{\mathrm{alg}}} \, |\!|\!|u_\ell^{k,\star} - u_\ell^{k,j}|\!\| \le |\!|\!|u_\ell^{k,j} - u_\ell^{k,j-1}|\!\| \le (1 + q_{\mathrm{alg}}) \, |\!|\!|u_\ell^{k,\star} - u_\ell^{k,j-1}|\!\|. \tag{5.27}$$

  Hence, the term $|\!|\!|u_\ell^{k,j} - u_\ell^{k,j-1}|\!\|$ *provides a means to estimate the algebraic error* $|\!|\!|u_\ell^{k,\star} - u_\ell^{k,j}|\!\|$*. In particular, the approximation* $u_\ell^{k,j}$ *is accepted and the algebraic solver is stopped if the algebraic error estimate* $|\!|\!|u_\ell^{k,j} - u_\ell^{k,j-1}|\!\|$ *is, up to the threshold* λ_{alg}*, below the estimate on the sum* $\eta_\ell(u_\ell^{k,j}) + |\!|\!|u_\ell^{k,j} - u_\ell^{k-1,\underline{j}}|\!\|$ *of the discretization and linearization error, see* (5.23)*. Since* $|\!|\!|u_\ell^{k,1} - u_\ell^{k,0}|\!\| = |\!|\!|u_\ell^{k,1} - u_\ell^{k-1,\underline{j}}|\!\|$*, the stopping criterion* (5.23) *would always terminate the algebraic solver at the first step* $j = 1$ *if* λ_{alg} *was chosen greater or equal to* 1 *which motivates the restriction* $\lambda_{\mathrm{alg}} < 1$*.*

- *The middle loop, Algorithm 41*(i)*, steers the linearization by means of the (perturbed) Banach–Picard iteration. Lemma 44 below shows that the term* $|\!|\!|u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}}|\!\|$ *estimates the linearization error* $|\!|\!|u_\ell^\star - u_\ell^{k,\underline{j}}|\!\|$*. Note that, a priori, only the non-perturbed Banach–Picard iteration corresponding to the (unavailable) exact solve of* (5.22) *yielding* $u_\ell^{k,\star}$ *would lead to the contraction*

$$|\!|\!|u_\ell^\star - u_\ell^{k,\star}|\!\| \le q_{\mathrm{Pic}} \, |\!|\!|u_\ell^\star - u_\ell^{k-1,\underline{j}}|\!\| \quad \text{for all } (\ell, k, 0) \in \mathcal{Q} \text{ with } k \ge 1, \tag{5.28}$$

  where $0 < q_{\mathrm{Pic}} := (1 - \alpha^2/L^2)^{1/2} < 1$ *and* $\mathcal{Q}$ *the index set defined in* (5.29)*. The approximation* $u_\ell^{k,\underline{j}}$ *is accepted and the linearization is stopped if the linearization error estimate* $|\!|\!|u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}}|\!\|$ *is, up to the threshold* λ_{Pic}*, below the discretization error estimate* $\eta_\ell(u_\ell^{k,\underline{j}})$*, see* (5.24) *(here* $\lambda_{\mathrm{Pic}} < 1$ *is not necessary).*

- *Finally, the outermost adaptive loop steers the local adaptive mesh-refinement. To this end, the Dörfler marking criterion* (5.25) *from [Dör96] is employed to mark elements* $T \in \mathcal{M}_\ell$ *for refinement, unless* $\eta_\ell(u_\ell^{\underline{k},\underline{j}}) = 0$*, in which case Proposition 43 below ensures that the approximation* $u_\ell^{\underline{k},\underline{j}}$ *coincides with the exact solution* $u^\star$ *of* (5.11)*.*

- *In a practical implementation, Algorithm 41 has to be complemented by appropriate stopping criteria in all of the loops so that the computation is terminated if* $u_\ell^{k,j} \in \mathcal{X}_\ell$ *is a sufficiently accurate approximation of* $u^\star$*. This can be done with the help of the reliable a posteriori error estimates summarized in Proposition 43 below.*

5.2.6 Index set $\mathcal{Q}$ for the triple loop

To analyze the asymptotic convergence behavior of Algorithm 41, we define the index set

$$\mathcal{Q} := \big\{(\ell,k,j) \in \mathbb{N}_0^3 \,:\, \text{index triple } (\ell,k,j) \text{ is used in Algorithm 41}\big\}. \tag{5.29}$$

Since Algorithm 41 is sequential, the index set $\mathcal{Q}$ is naturally ordered. For indices (ℓ,k,j), $(\ell',k',j') \in \mathcal{Q}$, we write

$$(\ell,k,j) < (\ell',k',j') \overset{\text{def}}{\iff} (\ell,k,j) \text{ appears earlier in Algorithm 41 than } (\ell',k',j'). \tag{5.30}$$

With this order, we can define

$$|(\ell,k,j)| := \#\big\{(\ell',k',j') \in \mathcal{Q} \,:\, (\ell',k',j') < (\ell,k,j)\big\},$$

which is the *total step number* of Algorithm 41. We make the following definitions, which are consistent with that of Algorithm 41, and additionally define $\underline{j}(\ell,0) := 0$:

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

Generically, it holds that $\underline{\ell} = \infty$, i.e., infinitely many steps of mesh-refinement take place. However, our analysis also covers the cases that either the k-loop (linearization) or the j-loop (algebraic solver) does not terminate, i.e.,

$$\underline{k}(\underline{\ell}) = \infty \;\text{ if }\; \underline{\ell} < \infty \quad \text{resp.} \quad \underline{j}(\underline{\ell},\underline{k}) = \infty \;\text{ if }\; \underline{\ell} < \infty \text{ and } \underline{k}(\underline{\ell}) < \infty,$$

or that the exact solution $u^\star$ is hit at Step (iii) of Algorithm 41 (note that $\eta_{\underline{\ell}}(u_{\underline{\ell}}^{\underline{k},\underline{j}}) = 0$ implies $u^\star = u_{\underline{\ell}}^{\underline{k},\underline{j}}$ by virtue of Proposition 43 below). To abbreviate notation, we make the following convention: If the mesh index $\ell \in \mathbb{N}_0$ is clear from the context, we simply write $\underline{k} := \underline{k}(\ell)$, e.g., $u_\ell^{\underline{k},j} := u_\ell^{\underline{k}(\ell),j}$. Similarly, we simply write $\underline{j} := \underline{j}(\ell,k)$, e.g., $u_\ell^{k,\underline{j}} := u_\ell^{k,\underline{j}(\ell,k)}$.

Note that there in particular holds $u_{\ell-1}^{\underline{k},\underline{j}} = u_\ell^{0,0} = u_\ell^{1,0}$ for all $(\ell,0,0) \in \mathcal{Q}$ with $\ell \geq 1$. Hence, these approximate solutions are indexed three times. This is our notational choice that will not be harmful for what follows. Alternatively, one could only index the approximate solutions that appear on Step (i)(b)(II) of Algorithm 41.

5.3 Main results

5.3.1 Reliabilty estimates of Algorithm 41

Our first proposition provides computable upper bounds for the energy error $|||u^\star - u_\ell^{k,j}|||$ of the iterates $u_\ell^{k,j}$ of Algorithm 41 at any step $(\ell,k,j) \in \mathcal{Q}$. In particular, we note that the stopping criteria (5.23)–(5.24) ensure reliability of $\eta_\ell(u_\ell^{\underline{k},\underline{j}})$ for the final perturbed Banach–Picard iterates $u_\ell^{\underline{k},\underline{j}}$.

Proposition 43 (Reliability at various stages of Algorithm 41). *Suppose (A1) and (A3). Then, for all $(\ell,k,j)\in\mathcal{Q}$, it holds that*

$$\vertiii{u^\star-u_\ell^{k,j}}\le C'_{\mathrm{rel}}\begin{cases}\eta_\ell(u_\ell^{k,j})+\vertiii{u_\ell^{k,j}-u_\ell^{k-1,\underline{j}}}+\vertiii{u_\ell^{k,j}-u_\ell^{k,j-1}}\\ \qquad\text{if } 0<k\le\underline{k}(\ell) \text{ and } 0<j\le\underline{j}(\ell,k),\\ \eta_\ell(u_\ell^{k,\underline{j}})+\vertiii{u_\ell^{k,\underline{j}}-u_\ell^{k-1,\underline{j}}}\quad\text{if } 0<k\le\underline{k}(\ell) \text{ and } j=\underline{j}(\ell,k),\\ \eta_\ell(u_\ell^{\underline{k},\underline{j}})\qquad\text{if } k=\underline{k}(\ell) \text{ and } \underline{j}=\underline{j}(\ell,\underline{k}),\\ \eta_{\ell-1}(u_{\ell-1}^{\underline{k},\underline{j}})\qquad\text{if } k=0 \text{ and } \ell>0.\end{cases}\tag{5.31}$$

The constant $C'_{\mathrm{rel}}>0$ depends only on C_{rel}, C_{stab}, q_{alg}, λ_{alg}, q_{Pic}, and λ_{Pic}.

The proof ist postponed to Section 5.3.2, because we first need some auxiliary results for Algorithm 41.

Observations on Algorithm 41

First, we collect some elementary observations on Algorithm 41 in what concerns nested iteration and stopping criteria. The given initial value of Algorithm 41 reads

$$u_0^{0,0}=u_0^{0,\underline{j}}=u_0^{0,\star}\in\mathcal{X}_0.\tag{5.32}$$

If $(\ell,0,0)\in\mathcal{Q}$ with $\ell\ge1$, then

$$u_\ell^{0,\star}:=u_\ell^{0,0}:=u_\ell^{0,\underline{j}}:=u_{\ell-1}^{\underline{k},\underline{j}}\in\mathcal{X}_{\ell-1}\subseteq\mathcal{X}_\ell.\tag{5.33}$$

If $(\ell,k,0)\in\mathcal{Q}$, then the initial guess for the algebraic solver reads

$$u_\ell^{k,0}=\begin{cases}u_0^{0,0} & \text{for } \ell=0,\\ u_{\ell-1}^{\underline{k},\underline{j}} & \text{if } k=0 \text{ and } \ell\ge1,\\ u_\ell^{k-1,\underline{j}} & \text{if } k>0,\end{cases}\tag{5.34}$$

i.e., the algebraic solver employs *nested iteration*. The stopping criterion (5.23) of Algorithm 41 guarantees that $\underline{j}(\ell,k)\ge1$ if $k>0$ and, for all $(\ell,k,j)\in\mathcal{Q}$, it holds that

$$\vertiii{u_\ell^{k,\underline{j}}-u_\ell^{k,\underline{j}-1}}\le\lambda_{\mathrm{alg}}\big[\eta_\ell(u_\ell^{k,\underline{j}})+\vertiii{u_\ell^{k,\underline{j}}-u_\ell^{k-1,\underline{j}}}\big]\quad\text{for } j=\underline{j}(\ell,k),\tag{5.35}$$

$$\vertiii{u_\ell^{k,j}-u_\ell^{k,j-1}}>\lambda_{\mathrm{alg}}\big[\eta_\ell(u_\ell^{k,j})+\vertiii{u_\ell^{k,j}-u_\ell^{k-1,\underline{j}}}\big]\quad\text{for } j<\underline{j}(\ell,k),\tag{5.36}$$

i.e., the algebraic error estimate $\vertiii{u_\ell^{k,j}-u_\ell^{k,j-1}}$ only drops below the discretization plus linearization error estimate at the stopping iteration $\underline{j}=\underline{j}(\ell,k)$.

The final iterates $u_\ell^{k,\underline{j}}$ of the algebraic solver are used to obtain the perturbed Banach–Picard iterates $u_\ell^{k+1,\underline{j}}$ for $k>0$, see (5.22). The stopping criterion (5.24) of Algorithm 41 guarantees that $\underline{k}(\ell)\ge1$ and, for all $(\ell,k,\underline{j})\in\mathcal{Q}$, it holds that

$$\vertiii{u_\ell^{\underline{k},\underline{j}}-u_\ell^{\underline{k}-1,\underline{j}}}\le\lambda_{\mathrm{Pic}}\,\eta_\ell(u_\ell^{\underline{k},\underline{j}})\quad\text{for } k=\underline{k}(\ell),\tag{5.37}$$

$$\vertiii{u_\ell^{k,\underline{j}}-u_\ell^{k-1,\underline{j}}}>\lambda_{\mathrm{Pic}}\,\eta_\ell(u_\ell^{k,\underline{j}})\quad\text{for } k<\underline{k}(\ell),\tag{5.38}$$

i.e., the linearization error estimate $\|\!|u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}}|\!\|$ only drops below the discretization error estimate at the stopping iteration $\underline{k} = \underline{k}(\ell)$.

Contraction of the perturbed Banach–Picard iteration

Assumption (5.20) immediately implies the algebraic solver contraction (5.26) and reliability (5.27) of the algebraic error estimate $\|\!|u_\ell^{k,j} - u_\ell^{k,j-1}|\!\|$. Similarly, one step of the non-perturbed Banach–Picard iteration (5.22) (i.e., with an exact algebraic solve of problem (5.22) with the datum $u_\ell^{k-1,\underline{j}}$) leads to contraction (5.28) and consequently to the reliability

$$\frac{1-q_{\mathrm{Pic}}}{q_{\mathrm{Pic}}}\,\|\!|u_\ell^\star - u_\ell^{k,\star}|\!\| \le \|\!|u_\ell^{k,\star} - u_\ell^{k-1,\underline{j}}|\!\| \le (1+q_{\mathrm{Pic}})\,\|\!|u_\ell^\star - u_\ell^{k-1,\underline{j}}|\!\| \tag{5.39}$$

of the unavailable linearization error estimate $\|\!|u_\ell^{k,\star} - u_\ell^{k-1,\underline{j}}|\!\|$. As our first result, we now show that, for sufficiently small stopping parameters $0 < \lambda_{\mathrm{alg}}$ in (5.23), we also get that the *perturbed* Banach–Picard iteration is a *contraction*.

Recall that $u_\ell^\star \in \mathcal{X}_\ell$ is the (unavailable) exact discrete solution given by (5.21), that $u_\ell^{k,\star} \in \mathcal{X}_\ell$ is the (unavailable) exact linearization solution given by (5.22), and that $u_\ell^{k,\underline{j}} \in \mathcal{X}_\ell$ is the computed solution for which the algebraic solver is stopped, see (5.23) (and (5.35)–(5.36) respectively) for the stopping criterion.

Lemma 44. *There exists $\lambda_{\mathrm{alg}}^\star > 0$ only depending on q_{alg} and q_{Pic} such that*

$$0 < q'_{\mathrm{Pic}} := \frac{q_{\mathrm{Pic}} + \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}}\,\lambda_{\mathrm{alg}}^\star}{1 - \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}}\,\lambda_{\mathrm{alg}}^\star} < 1. \tag{5.40}$$

Moreover, for all stopping parameters $0 < \lambda_{\mathrm{alg}} < 1$ and $0 < \lambda_{\mathrm{Pic}}$ from (5.23)–(5.24) such that $0 < \lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}} < \lambda_{\mathrm{alg}}^\star$, it holds that

$$\|\!|u_\ell^\star - u_\ell^{k,\underline{j}}|\!\| \le q'_{\mathrm{Pic}}\,\|\!|u_\ell^\star - u_\ell^{k-1,\underline{j}}|\!\| \quad \text{for all } 1 \le k < \underline{k}(\ell). \tag{5.41}$$

This also implies that

$$\frac{1-q'_{\mathrm{Pic}}}{q'_{\mathrm{Pic}}}\,\|\!|u_\ell^\star - u_\ell^{k,\underline{j}}|\!\| \le \|\!|u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}}|\!\| \le (1+q'_{\mathrm{Pic}})\,\|\!|u_\ell^\star - u_\ell^{k-1,\underline{j}}|\!\|. \tag{5.42}$$

Proof. Clearly, (5.42) follows from (5.41) by the triangle inequality as in (5.27) and (5.39). Moreover, (5.40) is obvious for sufficiently small $\lambda_{\mathrm{alg}}^\star$, since $q_{\mathrm{Pic}} = (1-\alpha^2/L^2)^{1/2} < 1$ from (5.28) and $0 < q_{\mathrm{alg}} < 1$ is fixed from (5.20). To see (5.41), first note that

$$\begin{aligned}\|\!|u_\ell^\star - u_\ell^{k,\underline{j}}|\!\| &\le \|\!|u_\ell^\star - u_\ell^{k,\star}|\!\| + \|\!|u_\ell^{k,\star} - u_\ell^{k,\underline{j}}|\!\| \\ &\overset{(5.28)}{\le} q_{\mathrm{Pic}}\,\|\!|u_\ell^\star - u_\ell^{k-1,\underline{j}}|\!\| + \|\!|u_\ell^{k,\star} - u_\ell^{k,\underline{j}}|\!\|,\end{aligned}$$

where the first term corresponds to the unperturbed Banach–Picard iteration (5.22) and the second to the algebraic error. Second, note that, since $1 \le k < \underline{k}(\ell)$,

$$\begin{aligned}
\|\!| u_\ell^{k,\star} - u_\ell^{k,\underline{j}} |\!\| &\overset{(5.27)}{\le} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} \|\!| u_\ell^{k,\underline{j}} - u_\ell^{k,\underline{j}-1} |\!\| \\
&\overset{(5.35)}{\le} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} \lambda_{\mathrm{alg}} \big[\eta_\ell(u_\ell^{k,\underline{j}}) + \|\!| u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}} |\!\| \big] \\
&\overset{(5.38)}{<} \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} (\lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}}) \|\!| u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}} |\!\| \\
&\le \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} (\lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}}) \big[\|\!| u_\ell^\star - u_\ell^{k,\underline{j}} |\!\| + \|\!| u_\ell^\star - u_\ell^{k-1,\underline{j}} |\!\| \big].
\end{aligned}$$

Combining the latter estimates with the assumption $\lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}} < \lambda_{\mathrm{alg}}^\star$, we see that

$$\|\!| u_\ell^\star - u_\ell^{k,\underline{j}} |\!\| \le (q_{\mathrm{Pic}} + \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} \lambda_{\mathrm{alg}}^\star) \|\!| u_\ell^\star - u_\ell^{k-1,\underline{j}} |\!\| + \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} \lambda_{\mathrm{alg}}^\star \|\!| u_\ell^\star - u_\ell^{k,\underline{j}} |\!\|.$$

If $0 < \lambda_{\mathrm{alg}}^\star$ is sufficiently small, it follows for all $1 \le k < \underline{k}(\ell)$ that

$$\begin{aligned}
\|\!| u_\ell^\star - u_\ell^{k,\underline{j}} |\!\| &\le \frac{q_{\mathrm{Pic}} + \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} \lambda_{\mathrm{alg}}^\star}{1 - \frac{q_{\mathrm{alg}}}{1-q_{\mathrm{alg}}} \lambda_{\mathrm{alg}}^\star} \|\!| u_\ell^\star - u_\ell^{k-1,\underline{j}} |\!\| \\
&= q'_{\mathrm{Pic}} \|\!| u_\ell^\star - u_\ell^{k-1,\underline{j}} |\!\|.
\end{aligned}$$

This concludes the proof. □

5.3.2 Proof of Proposition 43 (reliability estimates)

We are now ready to prove the estimates (5.31).

Proof of Proposition 43. First, let $(\ell,k,j) \in \mathcal{Q}$ with $0 < k \le \underline{k}(\ell)$ and $0 < j \le \underline{j}(\ell,k)$. Due to stability (A1), reliability (A3), and the contraction properties (5.27) resp. (5.39), it holds that

$$\begin{aligned}
\|\!| u^\star - u_\ell^{k,j} |\!\| &\le \|\!| u^\star - u_\ell^\star |\!\| + \|\!| u_\ell^\star - u_\ell^{k,j} |\!\| \\
&\overset{(A3)}{\lesssim} \eta_\ell(u_\ell^\star) + \|\!| u_\ell^\star - u_\ell^{k,j} |\!\| \\
&\overset{(A1)}{\lesssim} \eta_\ell(u_\ell^{k,j}) + \|\!| u_\ell^\star - u_\ell^{k,j} |\!\| \\
&\le \eta_\ell(u_\ell^{k,j}) + \|\!| u_\ell^\star - u_\ell^{k,\star} |\!\| + \|\!| u_\ell^{k,\star} - u_\ell^{k,j} |\!\| \\
&\overset{(5.39)}{\lesssim} \eta_\ell(u_\ell^{k,j}) + \|\!| u_\ell^{k,\star} - u_\ell^{k-1,\underline{j}} |\!\| + \|\!| u_\ell^{k,\star} - u_\ell^{k,j} |\!\| \\
&\le \eta_\ell(u_\ell^{k,j}) + \|\!| u_\ell^{k,j} - u_\ell^{k-1,\underline{j}} |\!\| + 2 \|\!| u_\ell^{k,\star} - u_\ell^{k,j} |\!\| \\
&\overset{(5.27)}{\lesssim} \eta_\ell(u_\ell^{k,j}) + \|\!| u_\ell^{k,j} - u_\ell^{k-1,\underline{j}} |\!\| + \|\!| u_\ell^{k,j} - u_\ell^{k,j-1} |\!\|.
\end{aligned} \tag{5.43}$$

This proves (5.31) for the case $0 < k \leq \underline{k}(\ell)$ and $0 < j \leq \underline{j}(\ell, k)$.

If $j = \underline{j}(\ell, k)$, we can improve this estimate using the stopping criterion (5.35) which yields that

$$|||u_\ell^{k,\underline{j}} - u_\ell^{k,\underline{j}-1}||| \overset{(5.35)}{\lesssim} \eta_\ell(u_\ell^{k,\underline{j}}) + |||u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}}|||. \tag{5.44}$$

Combined with (5.43), this proves (5.31) for $j = \underline{j}(\ell, k)$. If additionally $k = \underline{k}(\ell)$, the stopping criterion (5.37) and the previous estimate (5.44) provide that

$$|||u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k},\underline{j}-1}||| \overset{(5.44)}{\lesssim} \eta_\ell(u_\ell^{\underline{k},\underline{j}}) + |||u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k}-1,\underline{j}}||| \overset{(5.37)}{\lesssim} \eta_\ell(u_\ell^{\underline{k},\underline{j}}), \tag{5.45}$$

which proves (5.31) for this case. Finally, for $k = 0$, $\ell > 0$ and hence $j = \underline{j} = 0$, it directly follows from nested iteration (5.33) and the previous case $k = \underline{k}(\ell - 1)$ resp. $j = \underline{j}(\ell - 1, \underline{k})$ that

$$|||u^\star - u_\ell^{0,0}||| = |||u^\star - u_{\ell-1}^{\underline{k},\underline{j}}||| \lesssim \eta_{\ell-1}(u_{\ell-1}^{\underline{k},\underline{j}}). \tag{5.46}$$

This concludes the proof. □

5.3.3 Linear convergence of the quasi-error

The first main theorem states linear convergence in *each* step of the adaptive algorithm, i.e., algebraic solver *or* linearization *or* mesh-refinement.

Theorem 45 (linear convergence). *Suppose (A1)–(A3). Then, there exist* $\lambda_{\mathrm{alg}}^\star, \lambda_{\mathrm{Pic}}^\star > 0$ *such that for arbitrary* $0 < \theta, \lambda_{\mathrm{alg}}, \lambda_{\mathrm{Pic}}$ *with*

$$\begin{aligned} &0 < \theta \leq 1, \\ &0 < \lambda_{\mathrm{alg}} < 1, \\ &0 < \lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}} < \lambda_{\mathrm{alg}}^\star, \quad \textit{and,} \\ &0 < \lambda_{\mathrm{Pic}}/\theta < \lambda_{\mathrm{Pic}}^\star, \end{aligned}$$

there exist constants $C_{\mathrm{lin}} \geq 1$ *and* $0 < q_{\mathrm{lin}} < 1$ *such that the quasi-error*

$$\Delta_\ell^{k,j} := |||u^\star - u_\ell^{k,j}||| + |||u_\ell^{k,\star} - u_\ell^{k,j}||| + \eta_\ell(u_\ell^{k,j}), \tag{5.47}$$

composed of the overall error, the algebraic error, and the error estimator, is linearly convergent in the sense of

$$\Delta_{\ell'}^{k',j'} \leq C_{\mathrm{lin}}\, q_{\mathrm{lin}}^{|(\ell',k',j')|-|(\ell,k,j)|}\, \Delta_\ell^{k,j} \tag{5.48}$$

for all $(\ell, k, j), (\ell', k', j') \in \mathcal{Q}$ *with* $(\ell', k', j') \geq (\ell, k, j)$. *The constants* C_{lin} *and* q_{lin} *depend only on* C_{rel}, C_{stab}, q_{red}, θ, q_{alg}, λ_{alg}, q_{Pic}, λ_{Pic}, α, *and* L.

Note that $\Delta_{\ell'}^{k',j'} = \Delta_\ell^{k,j}$ when $(\ell', k', j') = (\ell, k, j)$, and then (5.48) holds with equality for $C_{\mathrm{lin}} = 1$. There are other cases where $u_{\ell'}^{k',j'} = u_\ell^{k,j}$ and where $u_{\ell'}^{k',j'} = u_\ell^{k,j}$ together

with $\mathcal{T}_{\ell'} = \mathcal{T}_\ell$, and consequently $\eta_{\ell'}(u_{\ell'}^{k',j'}) = \eta_\ell(u_\ell^{k,j})$, related to our notational choice for $\mathcal{Q}$ in (5.29) that also indexes nested iterates. The case with $\ell' = \ell$ arises for instance when $j = \underline{j}$, $j' = 0$, and $k' = k+1$, see Step (i)(a) of Algorithm 41. Note, however, that in such a situation, typically $u_{\ell'}^{k',\star} \neq u_\ell^{k,\star}$, and consequently $\Delta_{\ell'}^{k',j'} \neq \Delta_\ell^{k,j}$. A situation where $\Delta_{\ell'}^{k',j'} = \Delta_\ell^{k,j}$ for $(\ell',k',j') \neq (\ell,k,j)$ can nevertheless also appear, and is covered in (5.48). For instance, in the above example, when $j = \underline{j}$, $j' = 0$, $k' = k+1$, and $\ell' = \ell$, and where moreover $u_\ell^{k,j} = u_\ell^{k,\star} = u_\ell^\star$ (so that $u_\ell^{k,j} = u_\ell^{k,\star} = u_{\ell'}^{k',\star} = u_{\ell'}^{k',j'} = u_\ell^\star$), Algorithm 41 performs only one step of the algebraic solver on the linearization step k', so that $C_{\mathrm{lin}} = 1/q_{\mathrm{lin}}$ leads to equality in (5.48) where now $|(\ell',k',j')| - |(\ell,k,j)| = 1$.

In order to prove Theorem 45, we first introduce an auxiliary adaptive algorithm which we employ to prove a certain summability property of the quasi-error, before we prove linear convergence in Section 45.

An auxiliary adaptive algorithm

Due to Lemma 44, the iterates $u_\ell^{k,\underline{j}}$ are contractive in the index k. Consequently, Algorithm 41 fits into the framework of [GHPS18] upon defining u_ℓ from [GHPS18] as $u_\ell := u_\ell^{\underline{k},\underline{j}}$ for the case where $\underline{k}(\ell) < \infty$ and $\underline{j}(\ell,\underline{k}) < \infty$, i.e., both the algebraic and the linearization solvers are stopped by (5.23)–(5.24) on the mesh $\mathcal{T}_\ell$. Note that the assumption $(\ell+n+1,0,0) \in \mathcal{Q}$ below ensures this for all meshes $\mathcal{T}_{\ell'}$ with $0 \le \ell' \le \ell+n$. Then, we can rewrite [GHPS18, Lemma 4.9, equation (4.10)] and [GHPS18, Theorem 5.3, equation (5.5)] in the current setting to conclude two important properties: First, the estimators $\eta_\ell(u_\ell^{\underline{k},\underline{j}})$ available at Step (iv) of Algorithm 41 are, up to a constant, equivalent to the estimators $\eta_\ell(u_\ell^\star)$ corresponding to the unavailable exact linearization $u_\ell^\star$ of (5.21). And second, the estimators $\eta_\ell(u_\ell^{\underline{k},\underline{j}})$ are linearly convergent.

Lemma 46 ([GHPS18, Lemma 4.9, Theorem 5.3]). *Recall $\lambda_{\mathrm{alg}}^\star > 0$ and $0 < q_{\mathrm{Pic}}' < 1$ from Lemma 44. Define*

$$\lambda_{\mathrm{Pic}}^\star := \frac{1 - q_{\mathrm{Pic}}'}{q_{\mathrm{Pic}}' C_{\mathrm{stab}}} > 0$$

and note that it depends only on q_{Pic}, q_{alg}, and C_{stab}. Then, for all $0 < \theta, \lambda_{\mathrm{alg}}, \lambda_{\mathrm{Pic}}$ with

$$\begin{aligned} &0 < \theta \le 1,\\ &0 < \lambda_{\mathrm{alg}} < 1,\\ &0 < \lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}} < \lambda_{\mathrm{alg}}^\star, \quad \textit{and,}\\ &0 < \lambda_{\mathrm{Pic}}/\theta < \lambda_{\mathrm{Pic}}^\star, \end{aligned}$$

and all $(\ell,\underline{k},\underline{j}) \in \mathcal{Q}$ with $\underline{k} < \infty$ and $\underline{j} < \infty$, it holds that

$$(1 - \lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^\star)\,\eta_\ell(u_\ell^{\underline{k},\underline{j}}) \le \eta_\ell(u_\ell^\star) \le (1 + \lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^\star)\,\eta_\ell(u_\ell^{\underline{k},\underline{j}}). \tag{5.49}$$

Moreover, there exist $C_{\mathrm{GHPS}} > 0$ and $0 < q_{\mathrm{GHPS}} < 1$ such that

$$\eta_{\ell+n}(u_{\ell+n}^{\underline{k},\underline{j}}) \le C_{\mathrm{GHPS}}\, q_{\mathrm{GHPS}}^n\, \eta_\ell(u_\ell^{\underline{k},\underline{j}}) \quad \textit{for all } (\ell+n+1,0,0) \in \mathcal{Q}. \tag{5.50}$$

The constants C_{GHPS} and q_{GHPS} depend only on L, α, C_{rel}, C_{stab}, q_{red}, q_{alg}, and q_{Pic}, as well as on the adaptivity parameters θ, λ_{alg}, and λ_{Pic}. □

As a result of Lemma 46 and Proposition 43, we get the following lemma for the quasi-error of (5.47) on stopping indices $\underline{k}(\ell)$, $\underline{j}(\ell,k)$. Please note that when $\underline{\ell}<\infty$, the summation below only goes to $\underline{\ell}-1$, as the arguments rely on (5.50) which needs finite stopping indices $\underline{k}(\ell)$ and $\underline{j}(\ell,k)$ on each mesh $\mathcal{T}_\ell$.

Lemma 47. *Suppose that $0<\lambda_{\mathrm{alg}}+\lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}}<\lambda_{\mathrm{alg}}^\star$ (from Lemma 44) as well as $0<\theta\le 1$ and $0<\lambda_{\mathrm{Pic}}/\theta<\lambda_{\mathrm{Pic}}^\star$ (from Lemma 46). With the convention $\underline{\ell}-1=\infty$ if $\underline{\ell}=\infty$, there holds summability*

$$\sum_{\ell=\ell'+1}^{\underline{\ell}-1}\Delta_\ell^{\underline{k},\underline{j}}\le C\,\Delta_{\ell'}^{\underline{k},\underline{j}}\quad\text{for all }(\ell',\underline{k},\underline{j})\in\mathcal{Q},\tag{5.51}$$

where $C>0$ depends only on L, α, C_{rel}, C_{stab}, q_{red}, θ, q_{alg}, q_{Pic}, λ_{alg}, and λ_{Pic}.

Proof. Define $\widetilde{\Delta}_\ell^{\underline{k}}:=\|\!|u^\star-u_\ell^{\underline{k},\underline{j}}|\!\|+\eta_\ell(u_\ell^{\underline{k},\underline{j}})$ as the sum of overall error plus error estimator. In comparison with (5.47), $\widetilde{\Delta}_\ell^{\underline{k}}$ omits the algebraic error term but is only defined for the algebraic stopping indices $\underline{j}(\ell,k)$. With Proposition 43 and the linear convergence (5.50), we get that

$$\sum_{\ell=\ell'+1}^{\underline{\ell}-1}\widetilde{\Delta}_\ell^{\underline{k}}\overset{(5.31)}{\lesssim}\sum_{\ell=\ell'+1}^{\underline{\ell}-1}\eta_\ell(u_\ell^{\underline{k},\underline{j}})\overset{(5.50)}{\lesssim}\eta_{\ell'}(u_{\ell'}^{\underline{k},\underline{j}})\sum_{\ell=\ell'+1}^{\underline{\ell}-1}q_{\mathrm{GHPS}}^{\ell-\ell'}\lesssim\widetilde{\Delta}_{\ell'}^{\underline{k}}.$$

Let $(\ell',\underline{k},\underline{j})\in\mathcal{Q}$. By definition (5.47), it holds that

$$\Delta_{\ell'}^{\underline{k},\underline{j}}=\|\!|u^\star-u_{\ell'}^{\underline{k},\underline{j}}|\!\|+\|\!|u_{\ell'}^{\underline{k},\star}-u_{\ell'}^{\underline{k},\underline{j}}|\!\|+\eta_{\ell'}(u_{\ell'}^{\underline{k},\underline{j}})=\widetilde{\Delta}_{\ell'}^{\underline{k}}+\|\!|u_{\ell'}^{\underline{k},\star}-u_{\ell'}^{\underline{k},\underline{j}}|\!\|.$$

Moreover, note that

$$\begin{aligned}\|\!|u_{\ell'}^{\underline{k},\star}-u_{\ell'}^{\underline{k},\underline{j}}|\!\|&\overset{(5.27)}{\lesssim}\|\!|u_{\ell'}^{\underline{k},\underline{j}}-u_{\ell'}^{\underline{k},\underline{j}-1}|\!\|\\&\overset{(5.35)}{\lesssim}\eta_{\ell'}(u_{\ell'}^{\underline{k},\underline{j}})+\|\!|u_{\ell'}^{\underline{k},\underline{j}}-u_{\ell'}^{\underline{k}-1,\underline{j}}|\!\|\\&\overset{(5.37)}{\lesssim}\eta_{\ell'}(u_{\ell'}^{\underline{k},\underline{j}})\\&\le\widetilde{\Delta}_{\ell'}^{\underline{k}}.\end{aligned}$$

This proves the equivalence $\Delta_{\ell'}^{\underline{k},\underline{j}}\simeq\widetilde{\Delta}_{\ell'}^{\underline{k}}$ for all $(\ell',\underline{k},\underline{j})\in\mathcal{Q}$ and concludes the proof. □

5.3.4 Proof of Theorem 45 (linear convergence)

This section is dedicated to the proof of Theorem 45. The core is the following lemma that extends Lemma 47 to our setting with the triple indices.

Lemma 48. *Suppose that* $0 < \lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}} < \lambda_{\mathrm{alg}}^\star$ *(from Lemma 44) as well as* $0 < \theta \le 1$ *and* $0 < \lambda_{\mathrm{Pic}}/\theta < \lambda_{\mathrm{Pic}}^\star$ *(from Lemma 46). Then, there exists* $C_{\mathrm{sum}} > 0$ *such that*

$$\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \Delta_\ell^{k,j} \le C_{\mathrm{sum}}\, \Delta_{\ell'}^{k',j'} \quad \text{for all } (\ell',k',j') \in \mathcal{Q}. \tag{5.52}$$

The constant C_{sum} *depends only on* C_{rel}, C_{stab}, q_{red}, θ, q_{alg}, λ_{alg}, q_{Pic}, λ_{Pic}, α, *and* L.

Proof. **Step 1.** We prove that

$$\boxed{\mathrm{A}_\ell^{k,j} := |||u_\ell^\star - u_\ell^{k,j}||| + |||u_\ell^{k,\star} - u_\ell^{k,j}||| + \eta_\ell(u_\ell^{k,j}) \simeq \Delta_\ell^{k,j} \quad \text{for all } (\ell,k,j) \in \mathcal{Q}.} \tag{5.53}$$

Note that $\mathrm{A}_\ell^{k,j}$ and $\Delta_\ell^{k,j}$ only differ in the first term, where the overall error is replaced by the (inexact) linearization error. According to the Céa lemma (5.16), it holds that

$$|||u_\ell^\star - u_\ell^{k,j}||| \le |||u^\star - u_\ell^{k,j}||| + |||u^\star - u_\ell^\star||| \overset{(5.16)}{\lesssim} |||u^\star - u_\ell^{k,j}||| \le \Delta_\ell^{k,j}.$$

This implies that $\mathrm{A}_\ell^{k,j} \lesssim \Delta_\ell^{k,j}$. To see the converse inequality, note that

$$\begin{aligned} |||u^\star - u_\ell^{k,j}||| &\le |||u^\star - u_\ell^\star||| + |||u_\ell^\star - u_\ell^{k,j}||| \\ &\overset{(A3)}{\lesssim} \eta_\ell(u_\ell^\star) + |||u_\ell^\star - u_\ell^{k,j}||| \\ &\overset{(A1)}{\lesssim} \eta_\ell(u_\ell^{k,j}) + |||u_\ell^\star - u_\ell^{k,j}||| \\ &\le \mathrm{A}_\ell^{k,j}. \end{aligned}$$

This proves $\Delta_\ell^{k,j} \lesssim \mathrm{A}_\ell^{k,j}$ and concludes this step.

Step 2. We prove some auxiliary estimates. First, we prove that the algebraic error $|||u_\ell^{k,\star} - u_\ell^{k,j-1}|||$ dominates the modified total error $\mathrm{A}_\ell^{k,j}$, before the algebraic stopping criterion (5.23) is reached, i.e.,

$$\boxed{\mathrm{A}_\ell^{k,j} \lesssim |||u_\ell^{k,\star} - u_\ell^{k,j-1}||| \quad \text{for all } (\ell,k,j) \in \mathcal{Q} \text{ with } k \ge 1 \text{ and } 1 \le j < \underline{j}(\ell,k).} \tag{5.54}$$

To this end, note that

$$\begin{aligned} |||u_\ell^\star - u_\ell^{k,j}||| + |||u_\ell^{k,\star} - u_\ell^{k,j}||| &\le |||u_\ell^\star - u_\ell^{k,\star}||| + 2\,|||u_\ell^{k,\star} - u_\ell^{k,j}||| \\ &\overset{(5.39)}{\lesssim} |||u_\ell^{k,\star} - u_\ell^{k-1,\underline{j}}||| + |||u_\ell^{k,\star} - u_\ell^{k,j}||| \\ &\le 2\,|||u_\ell^{k,\star} - u_\ell^{k,j}||| + |||u_\ell^{k,j} - u_\ell^{k-1,\underline{j}}||| \\ &\overset{(5.27)}{\lesssim} |||u_\ell^{k,j} - u_\ell^{k,j-1}||| + |||u_\ell^{k,j} - u_\ell^{k-1,\underline{j}}|||. \end{aligned}$$

Since $1 \le j < \underline{j}(\ell,k)$, we obtain from the latter equation that

$$\begin{aligned}
\mathrm{A}_\ell^{k,j} &= \|\!|u_\ell^\star - u_\ell^{k,j}|\!\| + \|\!|u_\ell^{k,\star} - u_\ell^{k,j}|\!\| + \eta_\ell(u_\ell^{k,j}) \\
&\lesssim \|\!|u_\ell^{k,j} - u_\ell^{k,j-1}|\!\| + \|\!|u_\ell^{k,j} - u_\ell^{k-1,\underline{j}}|\!\| + \eta_\ell(u_\ell^{k,j}) \\
&\overset{(5.36)}{\lesssim} \|\!|u_\ell^{k,j} - u_\ell^{k,j-1}|\!\| \\
&\overset{(5.27)}{\lesssim} \|\!|u_\ell^{k,\star} - u_\ell^{k,j-1}|\!\|.
\end{aligned}$$

This proves (5.54).

Second, we consider the use of nested iteration when passing to the next perturbed Banach–Picard step. We prove that

$$\boxed{\|\!|u_\ell^{k,\star} - u_\ell^{k,0}|\!\| \lesssim \mathrm{A}_\ell^{k-1,\underline{j}} \quad \text{for all } (\ell,k,0) \in \mathcal{Q} \text{ with } k \ge 1,} \tag{5.55}$$

To this end, simply note that

$$\|\!|u_\ell^{k,\star} - u_\ell^{k,0}|\!\| \overset{(5.34)}{=} \|\!|u_\ell^{k,\star} - u_\ell^{k-1,\underline{j}}|\!\| \overset{(5.39)}{\lesssim} \|\!|u_\ell^\star - u_\ell^{k-1,\underline{j}}|\!\| \le \mathrm{A}_\ell^{k-1,\underline{j}}.$$

This proves (5.55).

Third, we prove that

$$\boxed{\mathrm{A}_\ell^{k,\underline{j}} \lesssim \mathrm{A}_\ell^{k,j} \quad \text{for all } (\ell,k,j) \in \mathcal{Q},} \tag{5.56}$$

related to the algebraic error contraction. Note that $k = 0$ implies $\underline{j} = 0$, so that (5.56) trivially holds for $k = 0$ with equality. Let now $k \ge 1$. We first consider the last but one algebraic iteration step $j = \underline{j}(\ell,k) - 1 \ge 0$. There holds that

$$\begin{aligned}
\mathrm{A}_\ell^{k,\underline{j}} &= \|\!|u_\ell^\star - u_\ell^{k,\underline{j}}|\!\| + \|\!|u_\ell^{k,\star} - u_\ell^{k,\underline{j}}|\!\| + \eta_\ell(u_\ell^{k,\underline{j}}) \\
&\le \|\!|u_\ell^\star - u_\ell^{k,\underline{j}-1}|\!\| + \|\!|u_\ell^{k,\star} - u_\ell^{k,\underline{j}-1}|\!\| + \eta_\ell(u_\ell^{k,\underline{j}}) + 2\,\|\!|u_\ell^{k,\underline{j}} - u_\ell^{k,\underline{j}-1}|\!\| \\
&\overset{(A1)}{\lesssim} \mathrm{A}_\ell^{k,\underline{j}-1} + \|\!|u_\ell^{k,\underline{j}} - u_\ell^{k,\underline{j}-1}|\!\| \\
&\overset{(5.27)}{\lesssim} \mathrm{A}_\ell^{k,\underline{j}-1} + \|\!|u_\ell^{k,\star} - u_\ell^{k,\underline{j}-1}|\!\| \\
&\simeq \mathrm{A}_\ell^{k,\underline{j}-1}.
\end{aligned}$$

This proves (5.56) for $j = \underline{j}(\ell,k) - 1 \ge 0$. Note that this argument also applies when $\underline{j} = 1$. If $0 \le j \le \underline{j}(\ell,k) - 2$, then we employ the last estimate and (5.54) to obtain that

$$\mathrm{A}_\ell^{k,\underline{j}} \lesssim \mathrm{A}_\ell^{k,\underline{j}-1} \overset{(5.54)}{\lesssim} \|\!|u_\ell^{k,\star} - u_\ell^{k,\underline{j}-2}|\!\| \overset{(5.26)}{\le} \|\!|u_\ell^{k,\star} - u_\ell^{k,j}|\!\| \le \mathrm{A}_\ell^{k,j},$$

also using that $q_{\mathrm{alg}} \le 1$. This concludes the proof of (5.56).

Fourth, we prove that the linearization error $|||u_\ell^\star - u_\ell^{k-1,\underline{j}}|||$ dominates the modified total error $\mathrm{A}_\ell^{k,\underline{j}}$, before the linearization stopping criterion (5.24) is reached, i.e.,

$$\boxed{\mathrm{A}_\ell^{k,\underline{j}} \lesssim |||u_\ell^\star - u_\ell^{k-1,\underline{j}}||| \quad \text{for all } (\ell,k,\underline{j}) \in \mathcal{Q} \text{ with } 1 \le k < \underline{k}(\ell).} \tag{5.57}$$

To see this, note that $1 \le k < \underline{k}(\ell)$ yields that

$$\begin{aligned}
\mathrm{A}_\ell^{k,\underline{j}} &= |||u_\ell^\star - u_\ell^{k,\underline{j}}||| + |||u_\ell^{k,\star} - u_\ell^{k,\underline{j}}||| + \eta_\ell(u_\ell^{k,\underline{j}}) \\
&\overset{(5.27)}{\lesssim} |||u_\ell^\star - u_\ell^{k,\underline{j}}||| + |||u_\ell^{k,\underline{j}} - u_\ell^{k,\underline{j}-1}||| + \eta_\ell(u_\ell^{k,\underline{j}}) \\
&\overset{(5.35)}{\lesssim} |||u_\ell^\star - u_\ell^{k,\underline{j}}||| + |||u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}}||| + \eta_\ell(u_\ell^{k,\underline{j}}) \\
&\overset{(5.42)}{\lesssim} |||u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}}||| + \eta_\ell(u_\ell^{k,\underline{j}}) \\
&\overset{(5.38)}{\lesssim} |||u_\ell^{k,\underline{j}} - u_\ell^{k-1,\underline{j}}||| \\
&\overset{(5.42)}{\lesssim} |||u_\ell^\star - u_\ell^{k-1,\underline{j}}|||,
\end{aligned}$$

where we employ Lemma 44 and hence require $0 < \lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}}$ to be sufficiently small. This proves (5.57).

Fifth, we consider the use of nested iteration when refining the mesh. We prove that

$$\boxed{\mathrm{A}_\ell^{0,\underline{j}} \lesssim \eta_{\ell-1}(u_{\ell-1}^{\underline{k},\underline{j}}) \le \mathrm{A}_{\ell-1}^{\underline{k},\underline{j}} \quad \text{for all } (\ell,\underline{k},\underline{j}) \in \mathcal{Q}.} \tag{5.58}$$

To this end, note that

$$|||u_\ell^\star - u_{\ell-1}^{\underline{k},\underline{j}}||| \le |||u^\star - u_\ell^\star||| + |||u^\star - u_{\ell-1}^{\underline{k},\underline{j}}||| \overset{(5.16)}{\lesssim} |||u^\star - u_{\ell-1}^{\underline{k},\underline{j}}||| \overset{(5.31)}{\lesssim} \eta_{\ell-1}(u_{\ell-1}^{\underline{k},\underline{j}}). \tag{5.59}$$

Next, recall from (5.33) that $u_\ell^{0,\star} = u_\ell^{0,\underline{j}} = u_{\ell-1}^{\underline{k},\underline{j}}$. From (A1) used on non-refined mesh elements and (A2) used on refined mesh elements, we hence conclude that

$$\begin{aligned}
\mathrm{A}_\ell^{0,\underline{j}} &= |||u_\ell^\star - u_\ell^{0,\underline{j}}||| + \eta_\ell(u_\ell^{0,\underline{j}}) \\
&\overset{(5.33)}{=} |||u_\ell^\star - u_{\ell-1}^{\underline{k},\underline{j}}||| + \eta_\ell(u_{\ell-1}^{\underline{k},\underline{j}}) \\
&\overset{(5.59)}{\lesssim} \eta_{\ell-1}(u_{\ell-1}^{\underline{k},\underline{j}}) + \eta_\ell(u_{\ell-1}^{\underline{k},\underline{j}}) \\
&= \eta_{\ell-1}(u_{\ell-1}^{\underline{k},\underline{j}}) + \eta_\ell(\mathcal{T}_{\ell-1} \cap \mathcal{T}_\ell, u_{\ell-1}^{\underline{k},\underline{j}}) + \eta_\ell(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell-1}, u_{\ell-1}^{\underline{k},\underline{j}}) \\
&\overset{(A1)}{\le} \eta_{\ell-1}(u_{\ell-1}^{\underline{k},\underline{j}}) + \eta_{\ell-1}(\mathcal{T}_{\ell-1} \cap \mathcal{T}_\ell, u_{\ell-1}^{\underline{k},\underline{j}}) + \eta_\ell(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell-1}, u_{\ell-1}^{\underline{k},\underline{j}}) \\
&\overset{(A2)}{\le} \eta_{\ell-1}(u_{\ell-1}^{\underline{k},\underline{j}}) + \eta_\ell(\mathcal{T}_{\ell-1} \cap \mathcal{T}_\ell, u_{\ell-1}^{\underline{k},\underline{j}}) + \eta_{\ell-1}(\mathcal{T}_{\ell-1} \setminus \mathcal{T}_\ell, u_{\ell-1}^{\underline{k},\underline{j}}) \\
&= 2\,\eta_{\ell-1}(u_{\ell-1}^{\underline{k},\underline{j}}).
\end{aligned}$$

This proves (5.58).

Sixth, we prove that

$$\boxed{\mathrm{A}_\ell^{\underline{k},\underline{j}} \lesssim \mathrm{A}_\ell^{k,\underline{j}} \quad \text{for all } (\ell,k,\underline{j}) \in \mathcal{Q},} \tag{5.60}$$

related to the linearization error contraction. We first consider $k = \underline{k}(\ell) - 1 \geq 0$. Note that

$$\|\!|u_\ell^{\underline{k},\star} - u_\ell^{\underline{k}-1,\underline{j}}|\!\| \leq \|\!|u_\ell^\star - u_\ell^{\underline{k},\star}|\!\| + \|\!|u_\ell^\star - u_\ell^{\underline{k}-1,\underline{j}}|\!\| \overset{(5.28)}{\lesssim} \|\!|u_\ell^\star - u_\ell^{\underline{k}-1,\underline{j}}|\!\| \leq \mathrm{A}_\ell^{\underline{k}-1,\underline{j}}. \tag{5.61}$$

Hence, the triangle inequality leads to

$$\begin{aligned}
\mathrm{A}_\ell^{\underline{k},\underline{j}} &= \|\!|u_\ell^\star - u_\ell^{\underline{k},\underline{j}}|\!\| + \|\!|u_\ell^{\underline{k},\star} - u_\ell^{\underline{k},\underline{j}}|\!\| + \eta_\ell(u_\ell^{\underline{k},\underline{j}}) \\
&\leq \|\!|u_\ell^\star - u_\ell^{\underline{k}-1,\underline{j}}|\!\| + \|\!|u_\ell^{\underline{k},\star} - u_\ell^{\underline{k}-1,\underline{j}}|\!\| + 2\,\|\!|u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k}-1,\underline{j}}|\!\| + \eta_\ell(u_\ell^{\underline{k},\underline{j}}) \\
&\overset{(5.61)}{\lesssim} \mathrm{A}_\ell^{\underline{k}-1,\underline{j}} + \|\!|u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k}-1,\underline{j}}|\!\| + \eta_\ell(u_\ell^{\underline{k},\underline{j}}) \\
&\overset{(A1)}{\lesssim} \mathrm{A}_\ell^{\underline{k}-1,\underline{j}} + \|\!|u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k}-1,\underline{j}}|\!\| \\
&\overset{(5.42)}{\lesssim} \mathrm{A}_\ell^{\underline{k}-1,\underline{j}} + \|\!|u_\ell^\star - u_\ell^{\underline{k}-1,\underline{j}}|\!\| \\
&\leq 2\,\mathrm{A}_\ell^{\underline{k}-1,\underline{j}}.
\end{aligned}$$

This proves (5.60) for $k = \underline{k}(\ell) - 1$. Note that the same argument also applies when $\underline{k} = 1$. If $0 \leq k \leq \underline{k}(\ell) - 2$, then

$$\mathrm{A}_\ell^{\underline{k},\underline{j}} \lesssim \mathrm{A}_\ell^{\underline{k}-1,\underline{j}} \overset{(5.57)}{\lesssim} \|\!|u_\ell^\star - u_\ell^{\underline{k}-2,\underline{j}}|\!\| \overset{(5.41)}{\leq} \|\!|u_\ell^\star - u_\ell^{\underline{k},\underline{j}}|\!\| \leq \mathrm{A}_\ell^{\underline{k},\underline{j}},$$

also using that $q'_{\mathrm{Pic}} \leq 1$. This concludes the proof of (5.60).

Seventh, we consider the use of nested iteration when passing to the next perturbed Banach–Picard step. We prove that

$$\boxed{\mathrm{A}_\ell^{k,0} \lesssim \mathrm{A}_\ell^{k-1,\underline{j}} \quad \text{for all } (\ell,k,0) \in \mathcal{Q} \text{ with } k \geq 1.} \tag{5.62}$$

Using (5.55) and recalling the definition $u_\ell^{k,0} = u_\ell^{k-1,\underline{j}}$, it holds that

$$\mathrm{A}_\ell^{k,0} = \|\!|u_\ell^\star - u_\ell^{k-1,\underline{j}}|\!\| + \|\!|u_\ell^{k,\star} - u_\ell^{k,0}|\!\| + \eta_\ell(u_\ell^{k-1,\underline{j}}) \overset{(5.55)}{\lesssim} \mathrm{A}_\ell^{k-1,\underline{j}},$$

which is the claim (5.62).

Step 3. This step collects auxiliary estimates following from the geometric series and the contraction properties of the linearization and the algebraic solver. First, with the convention $\underline{j}(\ell,k) - 1 = \infty$ when $\underline{j}(\ell,k) = \infty$, it holds that

$$\boxed{\sum_{j=i+1}^{\underline{j}(\ell,k)-1} \mathrm{A}_\ell^{k,j} \lesssim \|\!|u_\ell^{k,\star} - u_\ell^{k,i}|\!\| \leq \mathrm{A}_\ell^{k,i} \quad \text{for all } (\ell,k,i) \in \mathcal{Q} \text{ with } k \geq 1.} \tag{5.63}$$

This follows immediately from

$$\sum_{j=i+1}^{\underline{j}(\ell,k)-1} \mathrm{A}_\ell^{k,j} \overset{(5.54)}{\lesssim} \sum_{j=i+1}^{\underline{j}(\ell,k)-1} |||u_\ell^{k,\star} - u_\ell^{k,j-1}||| \overset{(5.26)}{\leq} |||u_\ell^{k,\star} - u_\ell^{k,i}||| \sum_{j=i}^{\infty} q_{\mathrm{alg}}^{j-i} \lesssim |||u_\ell^{k,\star} - u_\ell^{k,i}|||.$$

Analogously, with the convention that $\underline{k}(\ell)-1 = \infty$ when $\underline{k}(\ell) = \infty$, the contraction (5.41) of the perturbed Banach–Picard iteration leads to

$$\boxed{\sum_{k=i+1}^{\underline{k}(\ell)-1} \mathrm{A}_\ell^{k,\underline{j}} \lesssim |||u_\ell^\star - u_\ell^{i,\underline{j}}||| \leq \mathrm{A}_\ell^{i,\underline{j}} \quad \text{for all } (\ell, i, \underline{j}) \in \mathcal{Q}.} \tag{5.64}$$

This follows immediately from

$$\sum_{k=i+1}^{\underline{k}(\ell)-1} \mathrm{A}_\ell^{k,\underline{j}} \overset{(5.57)}{\lesssim} \sum_{k=i+1}^{\underline{k}(\ell)-1} |||u_\ell^\star - u_\ell^{k-1,\underline{j}}||| \overset{(5.41)}{\lesssim} |||u_\ell^\star - u_\ell^{i,\underline{j}}||| \sum_{k=i}^{\infty} (q'_{\mathrm{Pic}})^{k-i} \lesssim |||u_\ell^\star - u_\ell^{i,\underline{j}}|||.$$

With the analogous convention $\underline{\ell} - 1 = \infty$ when $\underline{\ell} = \infty$, we finally prove that

$$\boxed{\sum_{\ell=i+1}^{\underline{\ell}-1} \mathrm{A}_\ell^{\underline{k},\underline{j}} \lesssim \mathrm{A}_i^{\underline{k},\underline{j}} \quad \text{for all } (i, \underline{k}, \underline{j}) \in \mathcal{Q}.} \tag{5.65}$$

This follows from Step 1 and

$$\sum_{\ell=i+1}^{\underline{\ell}-1} \mathrm{A}_\ell^{\underline{k},\underline{j}} \overset{(5.53)}{\simeq} \sum_{\ell=i+1}^{\underline{\ell}-1} \Delta_\ell^{\underline{k},\underline{j}} \overset{(5.51)}{\lesssim} \Delta_i^{\underline{k},\underline{j}} \overset{(5.53)}{\simeq} \mathrm{A}_i^{\underline{k},\underline{j}}.$$

Step 4. From now on, let $(\ell', k', j') \in \mathcal{Q}$ be arbitrary. Suppose first that $\underline{\ell} = \infty$, i.e., both algebraic and linearization solvers terminate at some finite values $\underline{k}(\ell)$ for all $\ell \geq 0$ and $\underline{j}(\ell, k)$ for all $\ell \geq 0$ and all $k \leq \underline{k}(\ell)$, whereas infinitely many steps of mesh-refinement take place. By the definition of our index set $\mathcal{Q}$ in (5.29) (which in particular features nested

iterates), it holds that

$$\begin{aligned}\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} &= \sum_{\ell=\ell'+1}^{\infty}\Big(\mathrm{A}_\ell^{0,0}+\sum_{k=1}^{\underline{k}(\ell)}\Big(\mathrm{A}_\ell^{k,0}+\sum_{j=1}^{\underline{j}(\ell,k)}\mathrm{A}_\ell^{k,j}\Big)\Big)\\ &\quad+\sum_{k=k'+1}^{\underline{k}(\ell')}\Big(\mathrm{A}_{\ell'}^{k,0}+\sum_{j=1}^{\underline{j}(\ell',k)}\mathrm{A}_{\ell'}^{k,j}\Big)+\sum_{j=j'+1}^{\underline{j}(\ell',k')}\mathrm{A}_{\ell'}^{k',j}\\ &\lesssim \sum_{\ell=\ell'+1}^{\infty}\sum_{k=1}^{\underline{k}(\ell)}\sum_{j=1}^{\underline{j}(\ell,k)}\mathrm{A}_\ell^{k,j}+\sum_{k=k'+1}^{\underline{k}(\ell')}\sum_{j=1}^{\underline{j}(\ell',k)}\mathrm{A}_{\ell'}^{k,j}+\sum_{j=j'+1}^{\underline{j}(\ell',k')}\mathrm{A}_{\ell'}^{k',j},\end{aligned}\tag{5.66}$$

where we have employed estimates (5.58) and (5.62) in order to start all the summations from $k=1$ and $j=1$.

We consider the three summands in (5.66) separately. For the first sum, we infer that

$$\begin{aligned}\sum_{\ell=\ell'+1}^{\infty}\sum_{k=1}^{\underline{k}(\ell)}\sum_{j=1}^{\underline{j}(\ell,k)}\mathrm{A}_\ell^{k,j} &\overset{(5.63)}{\lesssim} \sum_{\ell=\ell'+1}^{\infty}\sum_{k=1}^{\underline{k}(\ell)}(\mathrm{A}_\ell^{k,\underline{j}}+|||u_\ell^{k,\star}-u_\ell^{k,0}|||)\\ &\overset{(5.55)}{\lesssim} \sum_{\ell=\ell'+1}^{\infty}\sum_{k=1}^{\underline{k}(\ell)}(\mathrm{A}_\ell^{k,\underline{j}}+\mathrm{A}_\ell^{k-1,\underline{j}})\\ &\lesssim \sum_{\ell=\ell'+1}^{\infty}\Big(\mathrm{A}_\ell^{0,\underline{j}}+\sum_{k=1}^{\underline{k}(\ell)}\mathrm{A}_\ell^{k,\underline{j}}\Big)\\ &\overset{(5.64)}{\lesssim} \sum_{\ell=\ell'+1}^{\infty}(\mathrm{A}_\ell^{0,\underline{j}}+\mathrm{A}_\ell^{\underline{k},\underline{j}})\\ &\overset{(5.58)}{\lesssim} \sum_{\ell=\ell'+1}^{\infty}(\mathrm{A}_{\ell-1}^{\underline{k},\underline{j}}+\mathrm{A}_\ell^{\underline{k},\underline{j}})\\ &\lesssim \mathrm{A}_{\ell'}^{\underline{k},\underline{j}}+\sum_{\ell=\ell'+1}^{\infty}\mathrm{A}_\ell^{\underline{k},\underline{j}}\\ &\overset{(5.65)}{\lesssim} \mathrm{A}_{\ell'}^{\underline{k},\underline{j}}\\ &\overset{(5.60)}{\lesssim} \mathrm{A}_{\ell'}^{k',\underline{j}}\\ &\overset{(5.56)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}.\end{aligned}\tag{5.67}$$

If $k' = \underline{k}(\ell')$, the second sum in the bound (5.66) disappears. If $k' < \underline{k}(\ell')$, we infer that

$$\begin{aligned}\sum_{k=k'+1}^{\underline{k}(\ell')}\sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} &\overset{(5.63)}{\lesssim} \sum_{k=k'+1}^{\underline{k}(\ell')} (\mathrm{A}_{\ell'}^{k,\underline{j}} + |||u_{\ell'}^{k,\star} - u_{\ell'}^{k,0}|||)\\ &\overset{(5.55)}{\lesssim} \sum_{k=k'+1}^{\underline{k}(\ell')} (\mathrm{A}_{\ell'}^{k,\underline{j}} + \mathrm{A}_{\ell'}^{k-1,\underline{j}})\\ &\lesssim \mathrm{A}_{\ell'}^{k',\underline{j}} + \sum_{k=k'+1}^{\underline{k}(\ell')} \mathrm{A}_{\ell'}^{k,\underline{j}}\\ &\overset{(5.64)}{\lesssim} \mathrm{A}_{\ell'}^{k',\underline{j}} + \mathrm{A}_{\ell'}^{\underline{k},\underline{j}}\\ &\overset{(5.60)}{\leq} \mathrm{A}_{\ell'}^{k',\underline{j}}\\ &\overset{(5.56)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}.\end{aligned} \tag{5.68}$$

If $j' = \underline{j}(\ell', k')$, the third sum in the bound (5.66) disappears. If $j' < \underline{j}(\ell', k')$, we infer that

$$\sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j} \overset{(5.63)}{\leq} \mathrm{A}_{\ell'}^{k',\underline{j}} + \mathrm{A}_{\ell'}^{k',j'} \overset{(5.56)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}. \tag{5.69}$$

Summing up (5.66)–(5.69), we see that, provided that $\underline{\ell} = \infty$,

$$\boxed{\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_{\ell}^{k,j} \lesssim \mathrm{A}_{\ell'}^{k',j'} \quad \text{provided that } \underline{\ell} = \infty.}$$

Step 5. Suppose that $\underline{\ell} < \infty$ and $\underline{k}(\underline{\ell}) = \infty$, i.e., for the mesh $\mathcal{T}_{\underline{\ell}}$, the linearization loop does not terminate. Moreover, let $\ell' < \underline{\ell}$. Then, it holds as in (5.66) that

$$\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_{\ell}^{k,j} \lesssim \sum_{k=1}^{\infty}\sum_{j=1}^{\underline{j}(\underline{\ell},k)} \mathrm{A}_{\underline{\ell}}^{k,j} + \sum_{\ell=\ell'+1}^{\underline{\ell}-1}\sum_{k=1}^{\underline{k}(\ell)}\sum_{j=1}^{\underline{j}(\ell,k)} \mathrm{A}_{\ell}^{k,j} + \sum_{k=k'+1}^{\underline{k}(\ell')}\sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} + \sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j}. \tag{5.70}$$

We argue as before to see that

$$\begin{aligned}&\sum_{\ell=\ell'+1}^{\underline{\ell}-1}\sum_{k=1}^{\underline{k}(\ell)}\sum_{j=1}^{\underline{j}(\ell,k)} \mathrm{A}_{\ell}^{k,j} \overset{(5.67)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'},\\ &\sum_{k=k'+1}^{\underline{k}(\ell')}\sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} \overset{(5.68)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}, \quad \text{and,}\\ &\sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j} \overset{(5.69)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}.\end{aligned} \tag{5.71}$$

It only remains to estimate

$$\begin{aligned}
\sum_{k=1}^{\infty}\sum_{j=1}^{\underline{j}(\underline{\ell},k)} \mathrm{A}_{\underline{\ell}}^{k,j} &\overset{(5.63)}{\lesssim} \sum_{k=1}^{\infty} \big(\mathrm{A}_{\underline{\ell}}^{k,\underline{j}} + |||u_{\underline{\ell}}^{k,\star} - u_{\underline{\ell}}^{k,0}|||\big) \\
&\overset{(5.55)}{\lesssim} \mathrm{A}_{\underline{\ell}}^{0,\underline{j}} + \sum_{k=1}^{\infty} \mathrm{A}_{\underline{\ell}}^{k,\underline{j}} \\
&\overset{(5.64)}{\lesssim} \mathrm{A}_{\underline{\ell}}^{0,\underline{j}} \\
&\overset{(5.58)}{\lesssim} \mathrm{A}_{\underline{\ell}-1}^{\underline{k},\underline{j}} \\
&\leq \mathrm{A}_{\ell'}^{\underline{k},\underline{j}} + \sum_{\ell=\ell'+1}^{\underline{\ell}-1} \mathrm{A}_{\ell}^{\underline{k},\underline{j}} \\
&\overset{(5.65)}{\lesssim} \mathrm{A}_{\underline{\ell}'}^{\underline{k},\underline{j}} \\
&\overset{(5.60)}{\lesssim} \mathrm{A}_{\underline{\ell}'}^{k',\underline{j}} \\
&\overset{(5.56)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}.
\end{aligned} \tag{5.72}$$

Altogether, we hence obtain that

$$\boxed{\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_{\ell}^{k,j} \lesssim \mathrm{A}_{\ell'}^{k',j'} \quad \text{provided that } \ell' < \underline{\ell} < \infty \text{ and } \underline{k}(\underline{\ell}) = \infty.}$$

Step 6. Suppose that $\underline{\ell} < \infty$ and $\underline{k}(\underline{\ell}) = \infty$, i.e., for the mesh $\mathcal{T}_{\underline{\ell}}$, the linearization loop does not terminate, and moreover, $\ell' = \underline{\ell}$. Arguing as in (5.72) and (5.69), it holds that

$$\boxed{\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_{\ell}^{k,j} \lesssim \sum_{k=k'+1}^{\infty}\sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} + \sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j} \lesssim \mathrm{A}_{\ell'}^{k',j'}.} \tag{5.73}$$

Step 7. Suppose that $\underline{\ell} < \infty$, where $\underline{k}(\underline{\ell}) < \infty$ and hence $\underline{j}(\underline{\ell},\underline{k}) = \infty$, i.e., the linear solver does not terminate for the linearization step $\underline{k}(\underline{\ell})$. Suppose moreover $\ell' < \underline{\ell}$. Then, it holds that

$$\begin{aligned}
\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_{\ell}^{k,j} \lesssim \sum_{j=1}^{\infty} \mathrm{A}_{\underline{\ell}}^{\underline{k},j} &+ \sum_{k=1}^{\underline{k}(\underline{\ell})-1}\sum_{j=1}^{\underline{j}(\underline{\ell},k)} \mathrm{A}_{\underline{\ell}}^{k,j} + \sum_{\ell=\ell'+1}^{\underline{\ell}-1}\sum_{k=1}^{\underline{k}(\ell)}\sum_{j=1}^{\underline{j}(\ell,k)} \mathrm{A}_{\ell}^{k,j} \\
&+ \sum_{k=k'+1}^{\underline{k}(\ell')}\sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} + \sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j}.
\end{aligned} \tag{5.74}$$

We argue as before to see that

$$\sum_{\ell=\ell'+1}^{\underline{\ell}-1}\sum_{k=1}^{\underline{k}(\ell)}\sum_{j=1}^{\underline{j}(\ell,k)} \mathrm{A}_\ell^{k,j} \overset{(5.67)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'},$$
$$\sum_{k=k'+1}^{\underline{k}(\ell')}\sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} \overset{(5.68)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}, \quad \text{and,}$$
$$\sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j} \overset{(5.69)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}.$$

For the first sum in (5.74), we get that

$$\sum_{j=1}^{\infty} \mathrm{A}_{\underline{\ell}}^{\underline{k},j} \overset{(5.63)}{\lesssim} |\!|\!| u_{\underline{\ell}}^{\underline{k},\star} - u_{\underline{\ell}}^{\underline{k},0} |\!|\!| \overset{(5.55)}{\lesssim} \mathrm{A}_{\underline{\ell}}^{\underline{k}-1,\underline{j}} \overset{(5.67)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}. \tag{5.75}$$

Hence, it only remains to estimate the second sum in (5.74), which can be treated analogously to (5.72) in Step 5 by $\mathrm{A}_{\ell'}^{k',j'}$. This proves that

$$\sum_{k=1}^{\underline{k}(\underline{\ell})-1}\sum_{j=1}^{\underline{j}(\underline{\ell},k)} \mathrm{A}_{\underline{\ell}}^{k,j} \overset{(5.72)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}.$$

Altogether, we obtain that

$$\boxed{\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} \lesssim \mathrm{A}_{\ell'}^{k',j'} \quad \text{provided that } \ell' < \underline{\ell} < \infty,\ \underline{k}(\underline{\ell}) < \infty, \text{ and } \underline{j}(\underline{\ell},\underline{k}) = \infty.}$$

Step 8. Suppose that $\underline{\ell} < \infty$, where $\underline{k}(\underline{\ell}) < \infty$ and hence $\underline{j}(\underline{\ell},\underline{k}) = \infty$, i.e., the linear solver does not terminate for the linearization step $\underline{k}(\underline{\ell})$. Suppose moreover $\ell' = \underline{\ell}$ but $k' < \underline{k}(\ell')$. Then, it holds that

$$\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} \lesssim \sum_{j=1}^{\infty} \mathrm{A}_{\ell'}^{\underline{k},j} + \sum_{k=k'+1}^{\underline{k}(\ell')-1}\sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} + \sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j}. \tag{5.76}$$

We argue as before to see that

$$\sum_{j=1}^{\infty} \mathrm{A}_{\ell'}^{\underline{k},j} \overset{(5.75)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'},$$
$$\sum_{k=k'+1}^{\underline{k}(\ell')-1}\sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} \overset{(5.68)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}, \quad \text{and,}$$
$$\sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j} \overset{(5.69)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}.$$

Hence, we obtain that

$$\boxed{\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} \lesssim \mathrm{A}_{\ell'}^{k',j'} \quad \text{provided that } \ell' = \underline{\ell} < \infty,\ k' < \underline{k}(\ell') < \infty, \text{ and } \underline{j}(\ell',\underline{k}) = \infty.}$$

Step 9. Suppose that $\underline{\ell} < \infty$, where $\underline{k}(\underline{\ell}) < \infty$ and hence $\underline{j}(\underline{\ell},\underline{k}) = \infty$, i.e., the linear solver does not terminate for the linearization step $\underline{k}(\underline{\ell})$. Suppose $\ell' = \underline{\ell}$ and $k' = \underline{k}(\ell')$. Then, it holds that

$$\boxed{\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} = \sum_{j=j'+1}^{\infty} \mathrm{A}_{\ell'}^{k',j} \overset{(5.63)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'}.} \tag{5.77}$$

Step 10. Suppose that $\underline{\ell}, \underline{k}(\underline{\ell}), \underline{j}(\underline{\ell},\underline{k}(\underline{\ell})) < \infty$ and that Algorithm 41 finished on Step (iii) when $\eta_{\underline{\ell}}(u_{\underline{\ell}}^{\underline{k},\underline{j}}) = 0$. From (5.31), we see that $\eta_{\underline{\ell}}(u_{\underline{\ell}}^{\underline{k},\underline{j}}) = 0$ implies $u^\star = u_{\underline{\ell}}^{\underline{k},\underline{j}}$, i.e., the exact solution was found. Moreover, through the stopping criteria (5.24) and (5.23), we see that $u_{\underline{\ell}}^{\underline{k}-1,\underline{j}} = u_{\underline{\ell}}^{\underline{k},\underline{j}-1} = u_{\underline{\ell}}^{\underline{k},\underline{j}}$, so that (5.42) gives $u_{\underline{\ell}}^\star = u_{\underline{\ell}}^{\underline{k},\underline{j}}$, and finally (5.22) gives $u_{\underline{\ell}}^{\underline{k},\star} = u_{\underline{\ell}}^{\underline{k},\underline{j}}$. Thus $\mathrm{A}_{\underline{\ell}}^{\underline{k},\underline{j}} = 0$.

Let $\ell' < \underline{\ell}$. Then, as in (5.70),

$$\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} \lesssim \sum_{k=1}^{\underline{k}(\underline{\ell})} \sum_{j=1}^{\underline{j}(\underline{\ell},k)} \mathrm{A}_{\underline{\ell}}^{k,j} + \sum_{\ell=\ell'+1}^{\underline{\ell}-1} \sum_{k=1}^{\underline{k}(\ell)} \sum_{j=1}^{\underline{j}(\ell,k)} \mathrm{A}_\ell^{k,j} + \sum_{k=k'+1}^{\underline{k}(\ell')} \sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} + \sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j}.$$

Here, the last three terms are estimated as in (5.71), whereas for the first one, we can proceed as in (5.72), crucially noting that the last summand $\mathrm{A}_{\underline{\ell}}^{\underline{k},\underline{j}}$ is zero.

If $\ell' = \underline{\ell}$, three cases are possible. The first case is $k' < \underline{k}$. Then

$$\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} \lesssim \sum_{k=k'+1}^{\underline{k}(\ell')} \sum_{j=1}^{\underline{j}(\ell',k)} \mathrm{A}_{\ell'}^{k,j} + \sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j},$$

which is controlled as in (5.71). The second case is $k' = \underline{k}$ but $j' < \underline{j}$, where directly

$$\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} \le \sum_{j=j'+1}^{\underline{j}(\ell',k')} \mathrm{A}_{\ell'}^{k',j} \overset{(5.63)}{\lesssim} \mathrm{A}_{\ell'}^{k',j'},$$

again using $\mathrm{A}_{\ell'}^{k',\underline{j}} = 0$. Finally, in the third case, $k' = \underline{k}$ and $j' = \underline{j}$, the sum is void. Altogether

$$\boxed{\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} \lesssim \mathrm{A}_{\ell'}^{k',j'}} \tag{5.78}$$

also holds in this case.

Step 11. Combining Steps 4–10 that cover all possible runs of Algorithm 41 with Step 1, we finally see that

$$\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \Delta_\ell^{k,j} \overset{(5.53)}{\simeq} \sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)>(\ell',k',j')}} \mathrm{A}_\ell^{k,j} \lesssim \mathrm{A}_{\ell'}^{k',j'} \overset{(5.53)}{\simeq} \Delta_{\ell'}^{k',j'} \quad \text{for all } (\ell',k',j')\in\mathcal{Q}.$$

This concludes the proof of (5.52). □

Proof of Theorem 45. The proof is split into two steps.

Step 1. For the convenience of the reader, we recall an argument from the proof of [CFPP14, Lemma 4.9]: For $M \in \mathbb{N}\cup\{\infty\}$, let $C>0$ and $\alpha_n \geq 0$ satisfy that

$$\sum_{n=N+1}^{M} \alpha_n \leq C\,\alpha_N \quad \text{for all } N\in\mathbb{N}_0 \text{ with } N<\min\{M,\infty\}.$$

Then,

$$(1+C^{-1})\sum_{n=N+1}^{M}\alpha_n \leq \sum_{n=N+1}^{M}\alpha_n + \alpha_N = \sum_{n=N}^{M}\alpha_n \quad \text{for all } N\in\mathbb{N}_0.$$

Inductively, it follows for all $N,m\in\mathbb{N}_0$ with $N+m<\min\{M+1,\infty\}$ that

$$(1+C^{-1})^m\sum_{n=N+m}^{M}\alpha_n \leq \sum_{n=N+1}^{M}\alpha_n + \alpha_N = \sum_{n=N}^{M}\alpha_n.$$

We thus conclude for all $N,m\in\mathbb{N}_0$ with $N+m<\min\{M+1,\infty\}$ that

$$\alpha_{N+m} \leq \sum_{n=N+m}^{M}\alpha_n \leq (1+C^{-1})^{-m}\sum_{n=N}^{M}\alpha_n \leq (1+C)\,(1+C^{-1})^{-m}\alpha_N.$$

Step 2. Since the index set $\mathcal{Q}$ is linearly ordered with respect to the total step counter $|(\cdot,\cdot,\cdot)|$, Lemma 48 and Step 1 imply that

$$\Delta_{\ell'}^{k',j'} \leq C_{\mathrm{lin}}\,q_{\mathrm{lin}}^{|(\ell',k',j')|-|(\ell,k,j)|}\,\Delta_\ell^{k,j}$$

for all $(\ell,k,j),(\ell',k',j')\in\mathcal{Q}$ with $(\ell',k',j')\geq(\ell,k,j)$, where $C_{\mathrm{lin}} = 1+C_{\mathrm{sum}}$ and $q_{\mathrm{lin}} = C_{\mathrm{sum}}/(C_{\mathrm{sum}}+1)$. This concludes the proof. □

5.3.5 Optimal convergence rates of the quasi-error

The second main result states optimal decay rate of the quasi-error $\Delta_\ell^{k,j}$ of (5.47) (and consequently of the total error $|||u^\star - u_\ell^{k,j}|||$) in terms of the number of degrees of freedom added in the space $\mathcal{X}_\ell$ with respect to $\mathcal{X}_0$. More precisely, the result states that if the

unknown weak solution u of (5.11) can be approximated at algebraic decay rate s with respect to the number of mesh elements added in the refinement of $\mathcal{T}_0$ (plus one) for a best-possible mesh, then Algorithm 41 achieves the same decay rate s with respect to the number of elements actually added in Algorithm 41, $(\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)$, up to a generic multiplicative constant. The proof of the following Theorem 49 is given in Section 5.3.6.

Theorem 49 (optimal decay rate wrt. degrees of freedom). *Suppose (A1)–(A4) and (R1)–(R3). Recall $\lambda_{\mathrm{alg}}^\star, \lambda_{\mathrm{Pic}}^\star > 0$ from Theorem 45. Let*

$$\begin{aligned} C_{\mathrm{Pic}} &:= q_{\mathrm{Pic}}/(1-q_{\mathrm{Pic}}) > 0,\\ C_{\mathrm{alg}} &:= q_{\mathrm{alg}}/(1-q_{\mathrm{alg}}) > 0, \quad \textit{and,}\\ \theta_{\mathrm{opt}} &:= (1 + C_{\mathrm{stab}}^2 C_{\mathrm{rel}}^2)^{-1}. \end{aligned}$$

Then, there exists $\theta^\star$ such that for all $0 < \theta, \lambda_{\mathrm{alg}}, \lambda_{\mathrm{Pic}}$ with

$$\begin{aligned} &0 < \theta < \min\{1, \theta^\star\},\\ &0 < \lambda_{\mathrm{alg}} < 1,\\ &0 < \lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}} < \lambda_{\mathrm{alg}}^\star, \quad \textit{and,}\\ &0 < \lambda_{\mathrm{Pic}}/\theta < \lambda_{\mathrm{Pic}}^\star, \end{aligned}$$

it holds that

$$0 < \theta' := \frac{\theta + C_{\mathrm{stab}}\Big((1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda_{\mathrm{alg}} + \big[C_{\mathrm{Pic}} + (1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda_{\mathrm{alg}}\big]\lambda_{\mathrm{Pic}}\Big)}{1 - \lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^\star} < \theta_{\mathrm{opt}}, \tag{5.79}$$

where the constant $\theta^\star > 0$ depends only on C_{stab}, q_{Pic}, and q_{alg}. Let $s > 0$ and define

$$\|u^\star\|_{\mathbb{A}_s} := \sup_{N\in\mathbb{N}_0}\Big((N+1)^s \inf_{\mathcal{T}_{\mathrm{opt}}\in\mathbb{T}(N)} \eta_{\mathrm{opt}}(u_{\mathrm{opt}}^\star)\Big) \in \mathbb{R}_{\geq 0}\cup\{\infty\}, \tag{5.80}$$

where $\eta_{\mathrm{opt}}(u_{\mathrm{opt}}^\star)$ is the error estimator corresponding to the exact solution of (5.12) with respect to the mesh $\mathcal{T}_{\mathrm{opt}}$ and

$$\mathbb{T}(N) := \big\{\mathcal{T}\in\mathbb{T} : \#\mathcal{T} - \#\mathcal{T}_0 \leq N\big\}.$$

Then, there exist $c_{\mathrm{opt}}, C_{\mathrm{opt}} > 0$ such that

$$c_{\mathrm{opt}}^{-1}\,\|u^\star\|_{\mathbb{A}_s} \leq \sup_{(\ell,k,j)\in\mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \Delta_\ell^{k,j} \leq C_{\mathrm{opt}}\,\max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}. \tag{5.81}$$

The constant $c_{\mathrm{opt}} > 0$ depends only on $C_{\mathrm{Céa}} = L/\alpha$, C_{stab}, C_{rel}, C_{son}, $\#\mathcal{T}_0$, s, and, if $\underline{\ell} < \infty$, additionally on $\underline{\ell}$. The constant $C_{\mathrm{opt}} > 0$ depends only on C_{stab}, C_{rel}, C_{mark}, $1 - \lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^\star$, $C_{\mathrm{Céa}} = L/\alpha$, C_{rel}', C_{mesh}, C_{lin}, q_{lin}, $\#\mathcal{T}_0$, and s. The maximum in the right inequality is only needed if $\ell = 0$. If $\ell \geq 1$, the maximum $\max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}$ can be replaced by $\|u^\star\|_{\mathbb{A}_s}$.

Remark 50. *Note that $\Delta_0^{0,0}$ can be arbitrarily bad due to a bad initial guess $u_0^{0,0}$. However, $\|u^\star\|_{\mathbb{A}_s}$ as well as the constant C_{opt} are independent of the initial guess, so that the upper bound in (5.81) cannot avoid $\max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}$ for the case $\ell = 0$. Such a phenomenon does not appear at later stages, since the stopping criteria (5.23) and (5.24) ensure that, though $u_\ell^{k,j}$ does not in general coincide with $u_\ell^\star$, it is sufficiently accurate. If one restricts the indices to $(\ell, k, j) \in \mathcal{Q}$ with $\ell \geq 1$, then the upper bound in (5.81) may omit $\Delta_0^{0,0}$.*

5.3.6 Proof of Theorem 49 (optimal convergence rates)

Lower bound in (5.81)

The first result of this section proves the left inequality in (5.81):

Lemma 51. *Suppose (R1) as well as (A1), (A2), and (A4). Let $s > 0$ and assume $\|u^\star\|_{\mathbb{A}_s} > 0$. Then, it holds that*

$$\|u^\star\|_{\mathbb{A}_s} \leq c_{\mathrm{opt}} \sup_{(\ell',k',j')\in\mathcal{Q}} (\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1)^s \Delta_{\ell'}^{k',j'}, \tag{5.82}$$

where the constant $c_{\mathrm{opt}} > 0$ depends only on $C_{\mathrm{Céa}} = L/\alpha$, C_{stab}, C_{rel}, C_{son}, $\#\mathcal{T}_0$, s, and, if $\underline{\ell} < \infty$, additionally on $\underline{\ell}$.

Proof. The proof is split into three steps. First, we recall from [BHP17, Lemma 22] that

$$\#\mathcal{T}_\circ/\#\mathcal{T}_\bullet \leq \#\mathcal{T}_\circ - \#\mathcal{T}_\bullet + 1 \leq \#\mathcal{T}_\circ \quad \text{for all } \mathcal{T}_\bullet \in \mathbb{T} \text{ and all } \mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet). \tag{5.83}$$

Step 1. We consider the three non-generic cases with $\underline{\ell} < \infty$. First, let $\underline{k}(\underline{\ell}) < \infty$, and $\underline{j}(\underline{\ell}, \underline{k}) < \infty$. Then, Algorithm 41 was terminated in Step (iii) with $\eta_{\underline{\ell}}(u_{\underline{\ell}}^{\underline{k},\underline{j}}) = 0$. Due to the Céa lemma (5.16) and Proposition 43, it follows that

$$|||u^\star - u_{\underline{\ell}}^\star||| \overset{(5.16)}{\lesssim} |||u^\star - u_{\underline{\ell}}^{\underline{k},\underline{j}}||| \overset{(5.31)}{\lesssim} \eta_{\underline{\ell}}(u_{\underline{\ell}}^{\underline{k},\underline{j}}) = 0$$

and hence $u^\star = u_{\underline{\ell}}^\star = u_{\underline{\ell}}^{\underline{k},\star} = u_{\underline{\ell}}^{\underline{k},\underline{j}}$ and $\eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) = 0$.

Second, let $\underline{k}(\underline{\ell}) < \infty$ but $\underline{j}(\underline{\ell}, \underline{k}) = \infty$, i.e., the algebraic solver does not stop. According to Theorem 45, it holds that

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

Hence, due to the uniqueness of the limit and the Céa lemma (5.16), we obtain that $u^\star = u_{\underline{\ell}}^\star = u_{\underline{\ell}}^{\underline{k},\star}$. From stability (A1), it follows that

$$0 \leq \eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) \overset{(A1)}{\lesssim} \eta_{\underline{\ell}}(u_{\underline{\ell}}^{\underline{k},j}) + |||u_{\underline{\ell}}^\star - u_{\underline{\ell}}^{\underline{k},j}||| \to 0 \quad \text{as} \quad j \to \infty.$$

Hence, we see that $\eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) = \eta_{\underline{\ell}}(u_{\underline{\ell}}^{\underline{k},\star}) = 0$.

Finally, let $\underline{k}(\underline{\ell}) = \infty$, i.e., the linearization solver does not stop. Analogously to the previous case, we obtain that

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

With the Céa lemma (5.16), this leads to

$$0 \le |||u_{\underline{\ell}}^\star - u_{\underline{\ell}}^{k,\underline{j}}||| \overset{(5.16)}{\le} (1 + C_{\text{Céa}})|||u^\star - u_{\underline{\ell}}^{k,\underline{j}}||| \to 0 \quad \text{as} \quad k \to \infty.$$

Hence, we get that $u^\star = u_{\underline{\ell}}^\star$. Again, stability (A1) yields that $\eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) = 0$.

In any case, $\underline{\ell} < \infty$ implies that $|||u^\star - u_{\underline{\ell}}^\star||| + \eta_{\underline{\ell}}(u_{\underline{\ell}}^\star) = 0$ and hence that

$$\|u^\star\|_{\mathbb{A}_s} = \sup_{0 \le N < \#\mathcal{T}_{\underline{\ell}} - \#\mathcal{T}_0} \Big((N+1)^s \inf_{\mathcal{T}_{\text{opt}} \in \mathbb{T}(N)} \eta_{\text{opt}}(u_{\text{opt}}^\star) \Big)$$

The term $N + 1$ within the supremum can be estimated by

$$N + 1 \le \#\mathcal{T}_{\underline{\ell}} - \#\mathcal{T}_0 \overset{(R1)}{\le} (C_{\text{son}}^{\underline{\ell}} - 1)\,\#\mathcal{T}_0.$$

Moreover, (A1), (A2), and (A4) yield quasi-monotonicity $\eta_{\text{opt}}(u_{\text{opt}}^\star) \lesssim \eta_0(u_0^\star)$ (see, e.g., [CFPP14, Lemma 3.5]). Altogether, we thus arrive at

$$\|u^\star\|_{\mathbb{A}_s} \lesssim \eta_0(u_0^\star) \le \sup_{\ell' \in \mathbb{N}_0} (\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1)^s\, \eta_{\ell'}(u_{\ell'}^\star). \tag{5.84}$$

Step 2. We consider the generic case that $\underline{\ell} = \infty$ and $\eta_\ell(u_\ell^{\underline{k},\underline{j}}) > 0$ for all $\ell \in \mathbb{N}_0$. Algorithm 41 then guarantees that $\#\mathcal{T}_\ell \to \infty$ as $\ell \to \infty$. Thus, we can argue analogously to the proof of [CFPP14, Theorem 4.1]: Let $N \in \mathbb{N}$. Choose the maximal $\ell' \in \mathbb{N}_0$ such that $\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1 \le N$. Then, $\mathcal{T}_{\ell'} \in \mathbb{T}(N)$. The choice of N guarantees that

$$\begin{aligned} N + 1 &\le \#\mathcal{T}_{\ell'+1} - \#\mathcal{T}_0 + 1 \\ &\overset{(5.83)}{\le} \#\mathcal{T}_{\ell'+1} \\ &\le C_{\text{son}} \#\mathcal{T}_{\ell'} \\ &\overset{(5.83)}{\le} C_{\text{son}} \#\mathcal{T}_0\,(\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1). \end{aligned} \tag{5.85}$$

This leads to

$$(N+1)^s \inf_{\mathcal{T}_{\text{opt}} \in \mathbb{T}(N)} \eta_{\text{opt}}(u_{\text{opt}}^\star) \lesssim (\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1)^s \eta_{\ell'}(u_{\ell'}^\star),$$

and we immediately see that this also holds for $N = 0$ with $\ell' = 0$. Taking the supremum over all $N \in \mathbb{N}_0$, we conclude that

$$\|u^\star\|_{\mathbb{A}_s} \lesssim \sup_{\ell' \in \mathbb{N}_0} (\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1)^s \eta_{\ell'}(u_{\ell'}^\star). \tag{5.86}$$

Step 3. With stability (A1) and the Céa lemma (5.16), we see for all $(\ell',0,0) \in \mathcal{Q}$ that

$$\begin{aligned}\eta_{\ell'}(u^\star_{\ell'}) &\overset{(A1)}{\lesssim} |||u^\star_{\ell'} - u^{0,0}_{\ell'}||| + \eta_{\ell'}(u^{0,0}_{\ell'}) \\ &\leq |||u^\star - u^\star_{\ell'}||| + |||u^\star - u^{0,0}_{\ell'}||| + \eta_{\ell'}(u^{0,0}_{\ell'}) \\ &\overset{(5.16)}{\lesssim} |||u^\star - u^{0,0}_{\ell'}||| + \eta_{\ell'}(u^{0,0}_{\ell'}) \\ &\leq \Delta^{0,0}_{\ell'}.\end{aligned}$$

With (5.84) and (5.86), we thus obtain that

$$\begin{aligned}\|u^\star\|_{\mathbb{A}_s} &\lesssim \sup_{(\ell',0,0)\in\mathcal{Q}} (\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1)^s \, \eta_{\ell'}(u^\star_{\ell'}) \\ &\leq \sup_{(\ell',k',j')\in\mathcal{Q}} (\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1)^s \, \Delta^{k',j'}_{\ell'}.\end{aligned}$$

This concludes the proof. □

Upper bound in (5.81)

To prove the right inequality in (5.81), we need the comparison lemma from [CFPP14, Lemma 4.14] for the error estimator of the exact discrete solution $u^\star_\ell \in \mathcal{X}_\ell$.

Lemma 52. *Suppose (R1)–(R2) as well as (A1), (A2), and (A4). Let $0 < \theta' < \theta_{\mathrm{opt}} := (1 + C^2_{\mathrm{stab}} C^2_{\mathrm{rel}})^{-1}$. Then, there exist constants $C_1, C_2 > 0$ such that for all $s > 0$ with $0 < \|u^\star\|_{\mathbb{A}_s} < \infty$ and all $\mathcal{T}_\ell \in \mathbb{T}$, there exists $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$ which satisfies*

$$\#\mathcal{R}_\ell \leq C_1 C_2^{-1/s} \|u^\star\|^{1/s}_{\mathbb{A}_s} \, \eta_\ell(u^\star_\ell)^{-1/s}, \tag{5.87}$$

as well as the Dörfler marking criterion

$$\theta' \eta_\ell(u^\star_\ell) \leq \eta_\ell(\mathcal{R}_\ell, u^\star_\ell). \tag{5.88}$$

The constants C_1, C_2 depend only on C_{stab} and C_{rel}. □

We are now ready to prove the right inequality in (5.81), which is the main result of Theorem 49:

Proof of Theorem 49. The proof is split into four steps. Without loss of generality, we may assume that $\|u^\star\|_{\mathbb{A}_s} < \infty$.

Step 1. Due to the assumptions $\lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}} \leq \lambda^\star_{\mathrm{alg}}$ (from Lemma 44) and $\lambda_{\mathrm{Pic}}/\theta < \lambda^\star_{\mathrm{Pic}}$ (from Lemma 46), we get that $\lambda_{\mathrm{alg}} \leq \lambda^\star_{\mathrm{alg}} \lambda_{\mathrm{Pic}} \leq \lambda^\star_{\mathrm{alg}} \lambda^\star_{\mathrm{Pic}} \theta$. Hence, it holds that

$$\begin{aligned}\theta' &= \frac{\theta + C_{\mathrm{stab}}\Big((1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda_{\mathrm{alg}} + \big[C_{\mathrm{Pic}} + (1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda_{\mathrm{alg}}\big]\lambda_{\mathrm{Pic}}\Big)}{1 - \lambda_{\mathrm{Pic}}/\lambda^\star_{\mathrm{Pic}}} \\ &\leq \frac{\theta + C_{\mathrm{stab}}\Big((1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda^\star_{\mathrm{alg}}\lambda^\star_{\mathrm{Pic}}\theta + \big[C_{\mathrm{Pic}} + (1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda^\star_{\mathrm{alg}}\lambda^\star_{\mathrm{Pic}}\theta\big]\lambda^\star_{\mathrm{Pic}}\theta\Big)}{1-\theta}\end{aligned}$$

which converges to 0 as $\theta \to 0$. As a consequence, (5.79) holds for sufficiently small θ.

Clearly, the parameters $\lambda_{\mathrm{alg}}, \lambda_{\mathrm{Pic}}, \theta > 0$ can be chosen such that all assumptions are fulfilled. First, choose $\theta > 0$ such that $0 < \theta < \min\{1, \theta^\star\}$. Then, choose $\lambda_{\mathrm{Pic}} > 0$ such that $0 < \lambda_{\mathrm{Pic}}/\theta < \lambda_{\mathrm{Pic}}^\star$. Finally, choose $0 < \lambda_{\mathrm{alg}} < 1$ such that $\lambda_{\mathrm{alg}} + \lambda_{\mathrm{alg}}/\lambda_{\mathrm{Pic}} < \lambda_{\mathrm{alg}}^\star$.

Step 2. Recall that $C_{\mathrm{Pic}} = q_{\mathrm{Pic}}/(1-q_{\mathrm{Pic}})$ and $C_{\mathrm{alg}} = q_{\mathrm{alg}}/(1-q_{\mathrm{alg}})$. Provided that $(\ell+1,0,0) \in \mathcal{Q}$, it follows from the contraction properties (5.27) as well as (5.39), and the stopping criteria (5.35) as well as (5.37) that

$$
\begin{aligned}
|||u_\ell^\star - u_\ell^{\underline{k},\underline{j}}||| &\le |||u_\ell^\star - u_\ell^{\underline{k},\star}||| + |||u_\ell^{\underline{k},\star} - u_\ell^{\underline{k},\underline{j}}||| \\
&\overset{(5.39)}{\le} C_{\mathrm{Pic}}\,|||u_\ell^{\underline{k},\star} - u_\ell^{\underline{k}-1,\underline{j}}||| + |||u_\ell^{\underline{k},\star} - u_\ell^{\underline{k},\underline{j}}||| \\
&\le (1+C_{\mathrm{Pic}})|||u_\ell^{\underline{k},\star} - u_\ell^{\underline{k},\underline{j}}||| + C_{\mathrm{Pic}}\,|||u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k}-1,\underline{j}}||| \\
&\overset{(5.27)}{\le} (1+C_{\mathrm{Pic}})C_{\mathrm{alg}}|||u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k},\underline{j}-1}||| + C_{\mathrm{Pic}}\,|||u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k}-1,\underline{j}}||| \\
&\overset{(5.35)}{\le} (1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda_{\mathrm{alg}}\,\eta_\ell(u_\ell^{\underline{k},\underline{j}}) + \big[C_{\mathrm{Pic}} + (1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda_{\mathrm{alg}}\big]|||u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k}-1,\underline{j}}||| \\
&\overset{(5.37)}{\le} \Big((1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda_{\mathrm{alg}} + \big[C_{\mathrm{Pic}} + (1+C_{\mathrm{Pic}})C_{\mathrm{alg}}\lambda_{\mathrm{alg}}\big]\,\lambda_{\mathrm{Pic}}\Big)\eta_\ell(u_\ell^{\underline{k},\underline{j}}) \\
&\overset{(5.79)}{=} C_{\mathrm{stab}}^{-1}\Big(\theta'\big(1-\lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^\star\big) - \theta\Big)\eta_\ell(u_\ell^{\underline{k},\underline{j}}).
\end{aligned}
$$

Step 3. Let $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$ be the subset from Lemma 52 with θ' from (5.79). From Step 2, we obtain that

$$
\begin{aligned}
\eta_\ell(\mathcal{R}_\ell, u_\ell^\star) &\overset{(A1)}{\le} \eta_\ell(\mathcal{R}_\ell, u_\ell^{\underline{k},\underline{j}}) + C_{\mathrm{stab}}|||u_\ell^\star - u_\ell^{\underline{k},\underline{j}}||| \\
&\le \eta_\ell(\mathcal{R}_\ell, u_\ell^{\underline{k},\underline{j}}) + \Big(\theta'\big(1-\lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^\star\big) - \theta\Big)\eta_\ell(u_\ell^{\underline{k},\underline{j}}).
\end{aligned}
\tag{5.89}
$$

With the equivalence (5.49), Lemma 52, and estimate (5.89), we see that

$$
\begin{aligned}
\theta'\big(1-\lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^\star\big)\eta_\ell(u_\ell^{\underline{k},\underline{j}}) &\overset{(5.49)}{\le} \theta'\eta_\ell(u_\ell^\star) \\
&\overset{(5.88)}{\le} \eta_\ell(\mathcal{R}_\ell, u_\ell^\star) \\
&\overset{(5.89)}{\le} \eta_\ell(\mathcal{R}_\ell, u_\ell^{\underline{k},\underline{j}}) + \Big(\theta'\big(1-\lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^\star\big) - \theta\Big)\eta_\ell(u_\ell^{\underline{k},\underline{j}}).
\end{aligned}
$$

Thus, we are led to

$$
\theta\,\eta_\ell(u_\ell^{\underline{k},\underline{j}}) \le \eta_\ell(\mathcal{R}_\ell, u_\ell^{\underline{k},\underline{j}}).
$$

Hence, $\mathcal{R}_\ell$ satisfies the Dörfler marking criterion (5.25) used in Algorithm 41. By the (quasi-)minimality of $\mathcal{M}_\ell$ in (5.25), we infer that

$$
\#\mathcal{M}_\ell \lesssim \#\mathcal{R}_\ell \overset{(5.87)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\eta_\ell(u_\ell^\star)^{-1/s} \overset{(5.49)}{\simeq} \|u^\star\|_{\mathbb{A}_s}^{1/s}\,\eta_\ell(u_\ell^{\underline{k},\underline{j}})^{-1/s}.
$$

Recall from (5.34) that $u_{\ell+1}^{0,\underline{j}} = u_\ell^{\underline{k},\underline{j}}$. Thus, (5.58) and the equivalence (5.53) lead to

$$\eta_\ell(u_\ell^{\underline{k},\underline{j}})^{-1/s} \overset{(5.58)}{\lesssim} (\mathrm{A}_{\ell+1}^{0,\underline{j}})^{-1/s} \overset{(5.53)}{\simeq} (\Delta_{\ell+1}^{0,\underline{j}})^{-1/s}.$$

Overall, we end up with

$$\#\mathcal{M}_\ell \lesssim \|u^\star\|_{\mathbb{A}_s}^{1/s} (\Delta_{\ell+1}^{0,\underline{j}})^{-1/s} \quad \text{for all } (\ell+1,0,0) \in \mathcal{Q}. \tag{5.90}$$

The hidden constant depends only on C_{stab}, C_{rel}, C_{mark}, $1-\lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^\star$, $C_{\mathrm{Céa}} = L/\alpha$, C_{rel}' and s.

Step 4. With linear convergence (5.48) and the geometric series, we see that

$$\begin{aligned}\sum_{\substack{(\widetilde{\ell},\widetilde{k},\widetilde{j})\in\mathcal{Q}\\(\widetilde{\ell},\widetilde{k},\widetilde{j})\le(\ell,k,j)}} (\Delta_{\widetilde{\ell}}^{\widetilde{k},\widetilde{j}})^{-1/s} &\overset{(5.48)}{\lesssim} (\Delta_\ell^{k,j})^{-1/s} \sum_{\substack{(\widetilde{\ell},\widetilde{k},\widetilde{j})\in\mathcal{Q}\\(\widetilde{\ell},\widetilde{k},\widetilde{j})\le(\ell,k,j)}} (q_{\mathrm{lin}}^{1/s})^{|(\ell,k,j)|-|(\widetilde{\ell},\widetilde{k},\widetilde{j})|}\\ &\lesssim (\Delta_\ell^{k,j})^{-1/s}\end{aligned} \tag{5.91}$$

with hidden constants depending only on C_{lin}, q_{lin}, and s. For $(\ell,k,j) \in \mathcal{Q}$ such that $(\ell+1,0,0) \in \mathcal{Q}$ and such that $\mathcal{T}_\ell \neq \mathcal{T}_0$, Step 3 and the closure estimate (R3) lead to

$$\begin{aligned}\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 &\simeq \#\mathcal{T}_\ell - \#\mathcal{T}_0\\ &\overset{(R3)}{\lesssim} \sum_{\widetilde{\ell}=0}^{\ell-1} \#\mathcal{M}_{\widetilde{\ell}}\\ &\overset{(5.90)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s} \sum_{\widetilde{\ell}=0}^{\ell} (\Delta_{\widetilde{\ell}}^{0,\underline{j}})^{-1/s}\\ &\le \|u^\star\|_{\mathbb{A}_s}^{1/s} \sum_{\substack{(\widetilde{\ell},\widetilde{k},\widetilde{j})\in\mathcal{Q}\\(\widetilde{\ell},\widetilde{k},\widetilde{j})\le(\ell,k,j)}} (\Delta_{\widetilde{\ell}}^{\widetilde{k},\widetilde{j}})^{-1/s}\\ &\overset{(5.91)}{\lesssim} \|u^\star\|_{\mathbb{A}_s}^{1/s} (\Delta_\ell^{k,j})^{-1/s}.\end{aligned}$$

Replacing $\|u^\star\|_{\mathbb{A}_s}$ with $\max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}$, the overall estimate trivially holds for $\mathcal{T}_\ell = \mathcal{T}_0$. This proves that

$$(\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \Delta_\ell^{k,j} \lesssim \begin{cases} \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}, & \text{if } (\ell+1,0,0)\in\mathcal{Q} \text{ and } \ell \ge 0,\\ \|u^\star\|_{\mathbb{A}_s}, & \text{if } (\ell+1,0,0)\in\mathcal{Q} \text{ and } \ell \ge 1.\end{cases} \tag{5.92}$$

It remains to consider the cases where $(\ell,k,j) \in \mathcal{Q}$ but $(\ell+1,0,0) \notin \mathcal{Q}$, as well as the case $\mathcal{T}_\ell = \mathcal{T}_0$. In the first case, it holds that $1 \le \ell = \underline{\ell} < \infty$, and one of the cases discussed in detail in Step 1 of Lemma 51 arises.

First, let $2 \le \ell = \underline{\ell} < \infty$. Since $\ell - 1 \ge 1$ and $(\ell,0,0) \in \mathcal{Q}$, (5.92) shows that

$$(\#\mathcal{T}_{\ell-1} - \#\mathcal{T}_0 + 1)^s \Delta_{\ell-1}^{\underline{k},\underline{j}} \lesssim \|u^\star\|_{\mathbb{A}_s}.$$

Moreover, Lemma 48 leads to $\Delta_\ell^{k,j} \lesssim \Delta_{\ell-1}^{\underline{k},\underline{j}}$. Therefore, we obtain from (5.85) that

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 \le C_{\mathrm{son}} \#\mathcal{T}_0 (\#\mathcal{T}_{\ell-1} - \#\mathcal{T}_0 + 1). \tag{5.93}$$

Altogether, (5.92) holds for this case as well.

Second, let $\ell = \underline{\ell} = 1$. Then, we can rely on the inequality

$$\begin{aligned}
(\#\mathcal{T}_1 - \#\mathcal{T}_0 + 1)^s \Delta_1^{k,j} &\overset{(5.93)}{\le} (C_{\mathrm{son}} \#\mathcal{T}_0)^s \, \Delta_1^{k,j} \\
&\overset{(5.52)}{\lesssim} \Delta_0^{\underline{k},\underline{j}} \\
&\overset{(5.47)}{=} |||u^\star - u_0^{\underline{k},\underline{j}}||| + |||u_0^{\underline{k},\star} - u_0^{\underline{k},\underline{j}}||| + \eta_0(u_0^{\underline{k},\underline{j}}) \\
&\overset{(5.27)}{\lesssim} |||u^\star - u_0^\star||| + |||u_0^\star - u_0^{\underline{k},\underline{j}}||| + |||u_0^{\underline{k},\underline{j}} - u_0^{\underline{k},\underline{j}-1}||| + \eta_0(u_0^{\underline{k},\underline{j}}) \\
&\overset{(5.23)}{\lesssim} |||u^\star - u_0^\star||| + |||u_0^\star - u_0^{\underline{k},\underline{j}}||| + |||u_0^{\underline{k},\underline{j}} - u_0^{\underline{k}-1,\underline{j}}||| + \eta_0(u_0^{\underline{k},\underline{j}}) \\
&\overset{(5.42)}{\lesssim} |||u^\star - u_0^\star||| + |||u_0^{\underline{k},\underline{j}} - u_0^{\underline{k}-1,\underline{j}}||| + \eta_0(u_0^{\underline{k},\underline{j}}) \\
&\overset{(5.24)}{\lesssim} |||u^\star - u_0^\star||| + \eta_0(u_0^{\underline{k},\underline{j}}) \\
&\overset{(5.49)}{\lesssim} |||u^\star - u_0^\star||| + \eta_0(u_0^\star) \\
&\overset{(A3)}{\lesssim} \eta_0(u_0^\star) \\
&\le \|u^\star\|_{\mathbb{A}_s}.
\end{aligned} \tag{5.94}$$

Thus, (5.92) holds for this case as well.

Finally, let $\ell = \underline{\ell} = 0$. Then, linear convergence (5.48) proves that

$$\Delta_0^{k,j} \overset{(5.48)}{\lesssim} \Delta_0^{0,0}. \tag{5.95}$$

Hence, (5.92) also holds for this case, and we conclude the proof of (5.81) □

5.3.7 Optimal computational complexity

Our last main result states that Algorithm 41 drives the quasi-error down at each possible rate s not only with respect to the number of degrees of freedom added in the space $\mathcal{X}_\ell$ in comparison with $\mathcal{X}_0$, but actually also with respect to the overall computational cost expressed as a cumulated sum of the number of degrees of freedom. This is an important improvement of Theorem 49. More precisely, under the same conditions as above, i.e., if the unknown weak solution u of (5.11) can be approximated at algebraic decay rate s with respect to the number of mesh elements added in the refinement of $\mathcal{T}_0$ (plus one), then Algorithm 41 generates a sequence of triple-(ℓ, k, j)-indexed approximations (mesh, linearization, algebraic solver) such that the quasi-error decays at rate s with respect to the overall algorithmic cost expressed as the sum of the number of simplices $\#\mathcal{T}_\ell$ over all steps $(\ell, k, j) \in \mathcal{Q}$ effectuated by Algorithm 41.

Theorem 53 (optimal decay rate wrt. overall computational cost). *Let the assumptions of Theorem 49 be verified. Then*

$$\begin{aligned} c_{\mathrm{opt}}^{-1}\,\|u^\star\|_{\mathbb{A}_s} &\le \sup_{(\ell',k',j')\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)\le(\ell',k',j')}} \#\mathcal{T}_\ell \Bigg)^s \Delta_{\ell'}^{k',j'} \\ &\le C'_{\mathrm{opt}}\,\max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}. \end{aligned} \tag{5.96}$$

The maximum in the right inequality is only needed if $\ell = 0$. If $\ell \ge 1$, the maximum $\max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}$ can be replaced by $\|u^\star\|_{\mathbb{A}_s}$. While $c_{\mathrm{opt}} > 0$ is the constant of Theorem 49, the constant $C'_{\mathrm{opt}} > 0$ reads $C'_{\mathrm{opt}} := (\#\mathcal{T}_0)^s\, C_{\mathrm{opt}}\, C_{\mathrm{lin}}\, \big(1 - q_{\mathrm{lin}}^{1/s}\big)^{-s}$.

Remark 54. *Analogously to the comments after Theorem 49, the upper estimate in (5.96) cannot avoid $\max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}$ for the case $\ell' = \ell = 0$. As above, if one restricts the indices to $(\ell',k',j'),(\ell,k,j) \in \mathcal{Q}$ with $\ell',\ell \ge 1$, then the upper bound in (5.96) may omit $\Delta_0^{0,0}$.*

Note that for any reasonable algebraic solver on mesh $\mathcal{T}_\ell$, the cost of its one step is proportional to $\#\mathcal{T}_\ell$. This also holds true for matrix and right-hand-side assembly in (5.22), evaluation of the residual estimators $\eta_\ell(u_\ell^{k,j})$, Dörfler marking, and local adaptive mesh refinement by, e.g., newest vertex bisection, while the cost of evaluation of the stopping criteria (5.23) and (5.24) is of $\mathcal{O}(1)$. Thus, the sum in (5.96) is indeed proportional to the overall computational cost invested into the numerical approximation of (5.1) by Algorithm 41.

Proof of Theorem 53. Note that $\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1 = 1 \le \#\mathcal{T}_0$ for $\ell' = 0$ and $\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 + 1 \le \#\mathcal{T}_{\ell'}$ for $\ell' > 0$, so that the left inequality in (5.96) immediately follows from the left inequality in (5.81). In order to prove the upper bound in (5.96), let $(\ell',k',j') \in \mathcal{Q}$. Employing the right inequality in (5.81) (cf. (5.92)), the geometric series proves that

$$\begin{aligned} \sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)\le(\ell',k',j')}} \#\mathcal{T}_\ell &\overset{(5.83)}{\le} \#\mathcal{T}_0 \sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)\le(\ell',k',j')}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1) \\ &\overset{(5.92)}{\le} \#\mathcal{T}_0\, C_{\mathrm{opt}}^{1/s}\, \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}^{1/s} \sum_{\substack{(\ell,k,j)\in\mathcal{Q}\\(\ell,k,j)\le(\ell',k',j')}} (\Delta_\ell^{k,j})^{-1/s} \\ &\overset{(5.48)}{\le} \#\mathcal{T}_0\, C_{\mathrm{opt}}^{1/s}\, C_{\mathrm{lin}}^{1/s}\, \frac{1}{1 - q_{\mathrm{lin}}^{1/s}}\, \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\}^{1/s} (\Delta_{\ell'}^{k',j'})^{-1/s}. \end{aligned}$$

Rearranging this estimate, we end up with

$$\sup_{(\ell',k',j')\in\mathcal{Q}} \Bigg(\sum_{\substack{(\ell,k,j)\in\mathcal{Q},\,\ell\ge1\\(\ell,k,j)\le(\ell',k',j')}} \#\mathcal{T}_\ell \Bigg)^s \Delta_{\ell'}^{k',j'} \lesssim \max\{\|u^\star\|_{\mathbb{A}_s}, \Delta_0^{0,0}\},$$

where the hidden constant depends only on C_{stab}, C_{rel}, C_{mark}, $1-\lambda_{\mathrm{Pic}}/\lambda_{\mathrm{Pic}}^{\star}$, $C_{\mathrm{Céa}} = L/\alpha$, C'_{rel}, C_{mesh}, C_{lin}, q_{lin}, $\#\mathcal{T}_0$, and s. This proves the right inequality in (5.96). □

5.4 Numerical experiments

In this section, we present numerical experiments in 2D to underpin our theoretical findings. We compare the performance of Algorithm 41 for

- different values of $\lambda_{\mathrm{alg}} \in \{10^{-0.5}, 10^{-1}, 10^{-1.5}, \dots, 10^{-4}\}$,
- different values of $\lambda_{\mathrm{Pic}} \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$,
- different values of $\theta \in \{0.05, 0.1, 0.15, \dots, 1\}$,

As model problems serve nonlinear boundary value problems which arise, e.g., from nonlinear material laws in magnetostatic computations, where the mesh-refinement is steered by newest vertex bisection.

As an algebraic solver for the linear problems arising from the Banach–Picard iteration, we use PCG with an optimal multilevel additive Schwarz preconditioner, cf. [Füh14, Section 7.4.1] and Section 4.7.1 respectively, i.e., the condition number of the preconditioned system is uniformly bounded.

Model problem

Analogously to Section 4.8, let $\Omega \subset \mathbb{R}^d$ with $d \geq 2$ be a bounded Lipschitz domain with polytopal boundary $\Gamma = \partial\Omega$. We again suppose that the boundary Γ is split into relatively open and disjoint Dirichlet and Neumann boundaries $\Gamma_D, \Gamma_N \subseteq \Gamma$ with $|\Gamma_D| > 0$, i.e., $\Gamma = \overline{\Gamma}_D \cup \overline{\Gamma}_N$. While the numerical experiments in Section 5.4.3–5.4.4 only consider $d = 2$, we stress that this model problem is covered by the abstract theory for any $d \geq 2$. For $f \in L^2(\Omega)$ and $g \in L^2(\Gamma)$, find $u^\star$ such that:

$$\begin{aligned} -\operatorname{div}\left(\mu(x, |\nabla u^\star(x)|^2)\nabla u^\star(x)\right) &= f(x) && \text{in } \Omega, \\ u^\star(x) &= 0 && \text{on } \Gamma_D, \\ \mu(x, |\nabla u^\star(x)|^2)\, \partial_{\mathbf{n}} u^\star(x) &= g(x) && \text{on } \Gamma_N, \end{aligned} \tag{5.97}$$

where the scalar nonlinearity $\mu\colon \Omega \times \mathbb{R}_{\geq 0} \to \mathbb{R}$ satisfies the properties (N1)–(N4) from Section 4.8. For the sake of completeness, we recall these properties in detail:

(N1) boundedness of $\mu(x,t)$: There exist constants $\gamma_1, \gamma_2 > 0$ such that

$$\gamma_1 \leq \mu(x,t) \leq \gamma_2 \quad \text{for all } x \in \Omega \text{ and } t \geq 0.$$

(N2) boundedness of $\mu(x,t) + 2t\frac{\mathrm{d}}{\mathrm{d}t}\mu(x,t)$: For $x \in \Omega$, the function $\mu(x,\cdot)$ is continuously differentiable, i.e., $\mu(x,\cdot) \in C^1(\mathbb{R}_{\geq 0}, \mathbb{R})$ and there exist constants $\widetilde{\gamma}_1, \widetilde{\gamma}_2 > 0$ such that

$$\widetilde{\gamma}_1 \leq \mu(x,t) + 2t\frac{\mathrm{d}}{\mathrm{d}t}\mu(x,t) \leq \widetilde{\gamma}_2 \quad \text{for all } x \in \Omega \text{ and } t \geq 0.$$

(N3) Lipschitz-continuity of $\mu(x,t)$ in x: There exists a constant $L_\mu > 0$ such that

$$|\mu(x,t) - \mu(y,t)| \le L_\mu |x-y| \quad \text{for all } x,y \in \Omega \text{ and } t \ge 0.$$

(N4) Lipschitz-continuity of $t\frac{\mathrm{d}}{\mathrm{d}t}\mu(x,t)$ in x: There exists a constant $\widetilde{L}_\mu > 0$ such that

$$|t\frac{\mathrm{d}}{\mathrm{d}t}\mu(x,t) - t\frac{\mathrm{d}}{\mathrm{d}t}\mu(y,t)| \le \widetilde{L}_\mu |x-y| \quad \text{for all } x,y \in \Omega \text{ and } t \ge 0.$$

5.4.1 Weak formulation

The weak formulation of (5.97) reads as follows: Find $u \in H^1_D(\Omega) := \{w \in H^1(\Omega) : w = 0 \text{ on } \Gamma_D\}$ such that

$$\int_\Omega \mu(x, |\nabla u^\star(x)|^2)\, \nabla u^\star \cdot \nabla v \,\mathrm{d}x = \int_\Omega f v \,\mathrm{d}x + \int_{\Gamma_N} g v \,\mathrm{d}s \quad \text{for all } v \in H^1_D(\Omega). \tag{5.98}$$

With respect to the abstract framework of Section 5.2.1, we take $\mathcal{H} = H^1_D(\Omega)$, $\mathbb{K} = \mathbb{R}$, and $\langle\!\langle \cdot\,,\cdot \rangle\!\rangle = \langle\!\langle \nabla\cdot\,, \nabla\cdot \rangle\!\rangle$ with $|\!|\!| v |\!|\!| = \|\nabla v\|_{L^2(\Omega)}$. We obtain (5.11) with operators

$$\langle \mathcal{A}w\,,\, v\rangle_{\mathcal{H}'\times\mathcal{H}} = \int_\Omega \mu(x, |\nabla w(x)|^2)\, \nabla w(x) \cdot \nabla v(x)\,\mathrm{d}x, \tag{5.99a}$$

$$F(v) = \int_\Omega f v \,\mathrm{d}x + \int_{\Gamma_N} g v \,\mathrm{d}s \tag{5.99b}$$

for all $v, w \in \mathcal{H}$. We again recall from [GHPS18, Proposition 8.2] that (N1)–(N2) implies that $\mathcal{A}$ is strongly monotone (with $\alpha := \widetilde{\gamma}_1$) and Lipschitz continuous (with $L := \widetilde{\gamma}_2$), so that (5.97) fits into the setting of Section 5.2.1. Moreover, (N3)–(N4) are required to prove the well-posedness and the properties (A1)–(A4) of the residual *a posteriori* error estimator.

5.4.2 Discretization and *a posteriori* error estimator

Let $\mathcal{T}_0$ be a conforming initial triangulation of Ω into simplices $\mathcal{T} \in \mathcal{T}_0$. For each $\mathcal{T}_\ell \in \mathbb{T}$, consider the lowest-order FEM space

$$\mathcal{H}_\ell := \big\{ v \in C(\Omega) \,:\, v|_\Gamma = 0 \text{ and } v|_T \in \mathcal{P}^1(T) \text{ for all } T \in \mathcal{T}_H \big\}. \tag{5.100}$$

As in Section 4.8, cf. [GMZ12, Section 3.2], we define for all $T \in \mathcal{T}_\ell$ and all $v_\ell \in \mathcal{H}_\ell$, the corresponding weighted residual error indicators

$$\begin{aligned} \eta_\ell(T, v_\ell)^2 := |T|^{2/d} \| f + \operatorname{div}(\mu(\cdot, |\nabla v_\ell|^2)\nabla v_\ell)\|^2_{L^2(T)} \\ + |T|^{1/d} \| [(\mu(\cdot, |\nabla v_\ell|^2)\nabla v_\ell)\cdot \mathbf{n}]\|_{L^2(\partial T \cap \Omega)^2}, \end{aligned} \tag{5.101}$$

where $[\cdot]$ denotes the usual jump of discrete functions across element interfaces, and $\mathbf{n}$ is the outer normal vector of the considered element.

Due to (N3), the error estimator is well-posed, since the nonlinearity $\mu(x,t)$ is Lipschitz continuous in x. Then, reliability (A3) and discrete reliability (A4) are proved as in the linear case, see, e.g., [CKNS08] for the linear case or [GMZ12, Theorem 3.3] and [GMZ12, Theorem 3.4], respectively, for strongly monotone nonlinearities.

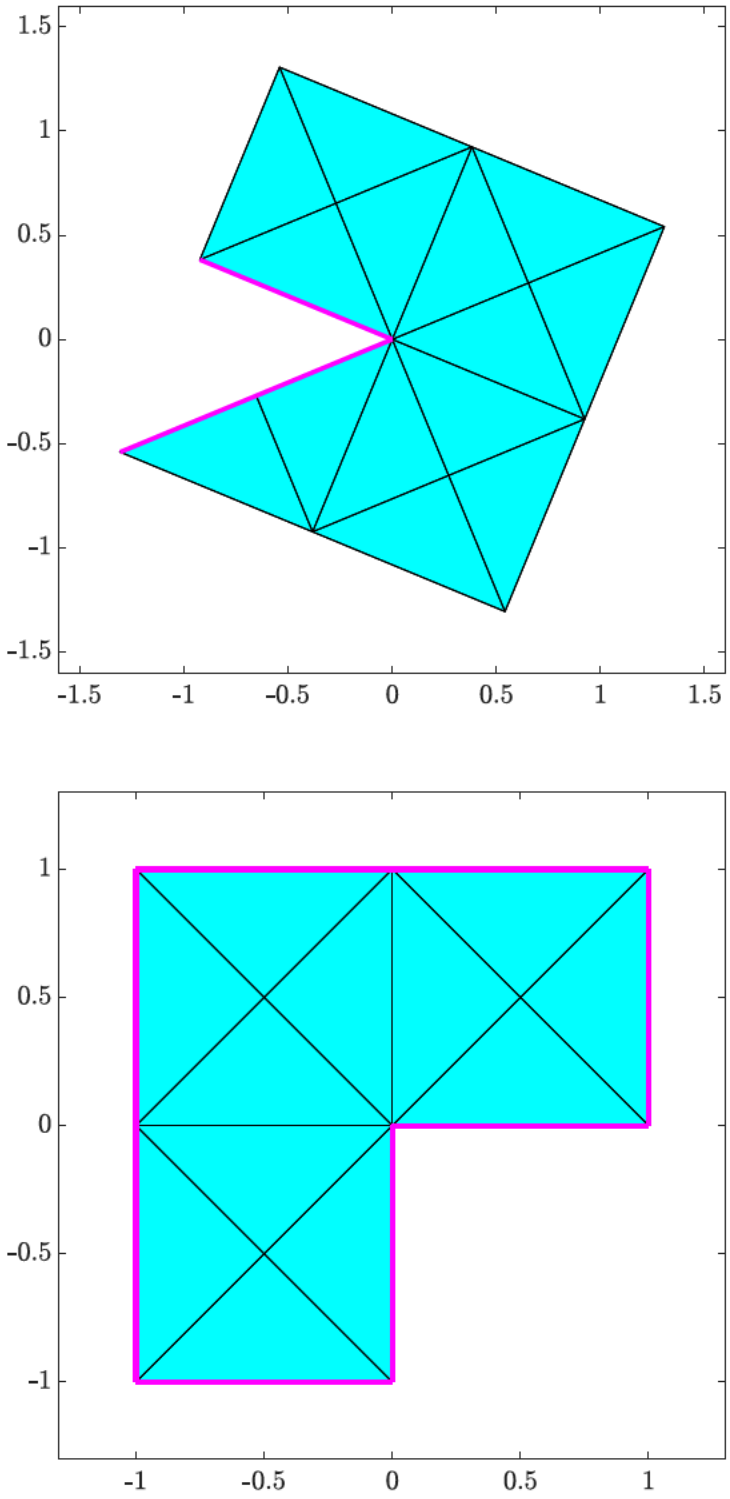


Figure 5.1: Z-shaped domain $\Omega \subset \mathbb{R}^2$ with initial mesh $\mathcal{T}_0$ (top) and L–shaped domain $\Omega \subset \mathbb{R}^2$ with initial mesh $\mathcal{T}_0$ (bottom), where Γ_D is marked by a thick pink line.

5.4.3 Experiment with known solution on Z-shaped domain

We consider the Z-shaped domain $\Omega \subset \mathbb{R}^2$ from Figure 5.1 (top) with mixed boundary conditions and the nonlinear problem (5.97) with

$$\mu(x, |\nabla u^\star(x)|^2) := 2 + \frac{1}{\sqrt{1 + |\nabla u^\star(x)|^2}}.$$

This leads to the bounds $\alpha = 2$ and $L = 3$ in (5.10). We prescribe the solution $u^\star$ in polar coordinates $(x_1, x_2) = r(\cos\xi, \sin\xi)$ with $\xi \in (-\pi, \pi)$ by

$$u^\star(x_1, x_2) = r^\beta \cos(\beta\,\xi), \tag{5.102}$$

with $\beta = 4/7$ and compute f and g in (5.97) accordingly. We note that $u^\star$ has a generic singularity at the re-entrant corner $(x, y) = (0, 0)$.

In Figure 5.2, we compare uniform mesh-refinement ($\theta = 1$) to adaptive mesh-refinement ($0 < \theta < 1$) for different values of λ_{alg} and λ_{Pic}. We plot the error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{j}})$ over the number of elements $N := \#\mathcal{T}_\ell$. First (top), we fix $\theta = 0.5$, $\lambda_{\mathrm{Pic}} = 10^{-2}$, and choose $\lambda_{\mathrm{alg}} \in \{10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$. We see that uniform mesh-refinement leads to the suboptimal rate of convergence $\mathcal{O}(N^{-2/7})$, whereas Algorithm 41 with adaptive mesh-refinement regains the optimal rate of convergence $\mathcal{O}(N^{-1/2})$, independently of the actual choice of λ_{alg}. We observe the very same if we fix $\theta = 0.5$, $\lambda_{\mathrm{alg}} = 10^{-2}$, and choose $\lambda_{\mathrm{Pic}} \in \{1, 10^{-1}, 10^{-2}, 10^{-3}, 10^{-4}\}$ (middle), or, if we fix $\lambda_{\mathrm{alg}} = \lambda_{\mathrm{Pic}} = 10^{-2}$ and vary $\theta \in \{0.1, 0.3, 0.5, 0.7, 0.9\}$ (bottom). Since we know from Proposition 43 and the estimate

$$\begin{aligned}
|||u_\ell^{\underline{k},\star} - u_\ell^{\underline{k},\underline{j}}||| &\overset{(5.27)}{\lesssim} |||u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k},\underline{j}-1}||| \\
&\overset{(5.35)}{\lesssim} \eta_\ell(u_\ell^{\underline{k},\underline{j}}) + |||u_\ell^{\underline{k},\underline{j}} - u_\ell^{\underline{k}-1,\underline{j}}||| \\
&\overset{(5.37)}{\lesssim} \eta_\ell(u_\ell^{\underline{k},\underline{j}})
\end{aligned}$$

that $\eta_\ell(u_\ell^{\underline{k},\underline{j}}) \simeq \Delta_\ell^{\underline{k},\underline{j}}$, this empirically underpins Theorem 49.

In Figure 5.3, analogously to Figure 5.2, we choose different combinations of θ, λ_{alg}, and λ_{Pic}. We plot the error estimator $\eta_{\ell'}(u_{\ell'}^{\underline{k}',\underline{j}'})$ over the cumulative sum $\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell$. Independently of the choice of θ, λ_{alg}, and λ_{Pic}, we observe the optimal order of convergence $\mathcal{O}\big(\big(\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell\big)^{-1/2}\big)$ with respect to the overall computational complexity in accordance with Theorem 53.

In Figure 5.4, we also consider the total number of PCG iterations cumulated over all Picard steps on the given mesh for different combinations of θ, λ_{alg}, and λ_{Pic}. We observe that independently of the choice of these parameters, the total number of PCG iterations stays uniformly bounded. Additionally, we see that for larger values of λ_{alg} and λ_{Pic}, as well as for smaller values of θ, the total number of PCG iterations is smaller.

In contrast to the the previous Chapters 4–6, where the corresponding algorithms steer the adaptive mesh-refinement and either incorporated an iterative linearization or an algebraic solver, our proposed Algorithm 41 combines these two concepts. Hence, to try to analyze

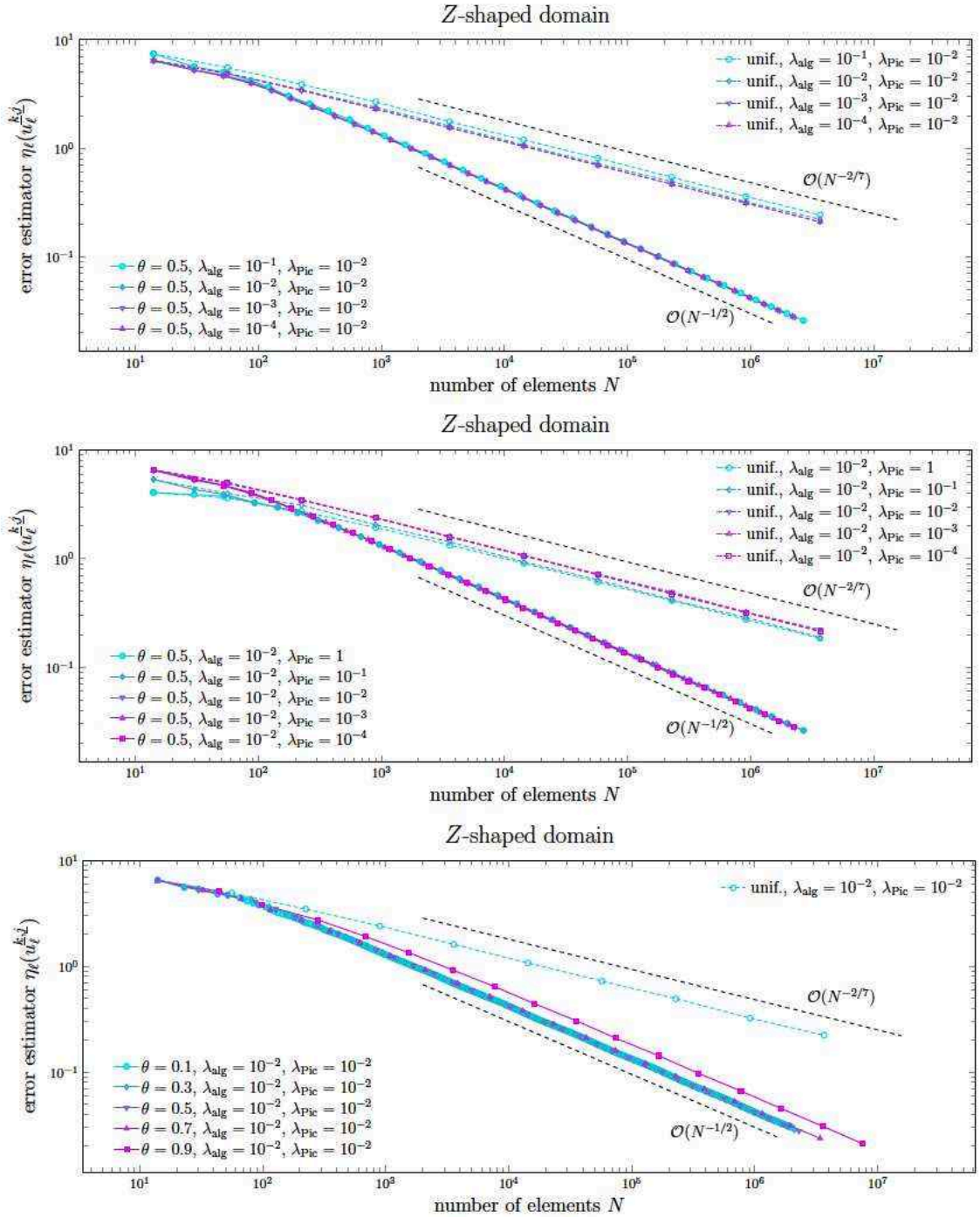


Figure 5.2: Example from Section 5.4.3 (Experiment with known solution on Z-shaped domain): Error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{j}})$ on mesh $\mathcal{T}_\ell$, perturbed Banach–Picard iteration $\underline{k}$, and PCG step $\underline{j}$ of Algorithm 41 with respect to the number of elements N of the mesh $\mathcal{T}_\ell$ for various parameters θ, λ_{Pic}, and λ_{alg}.

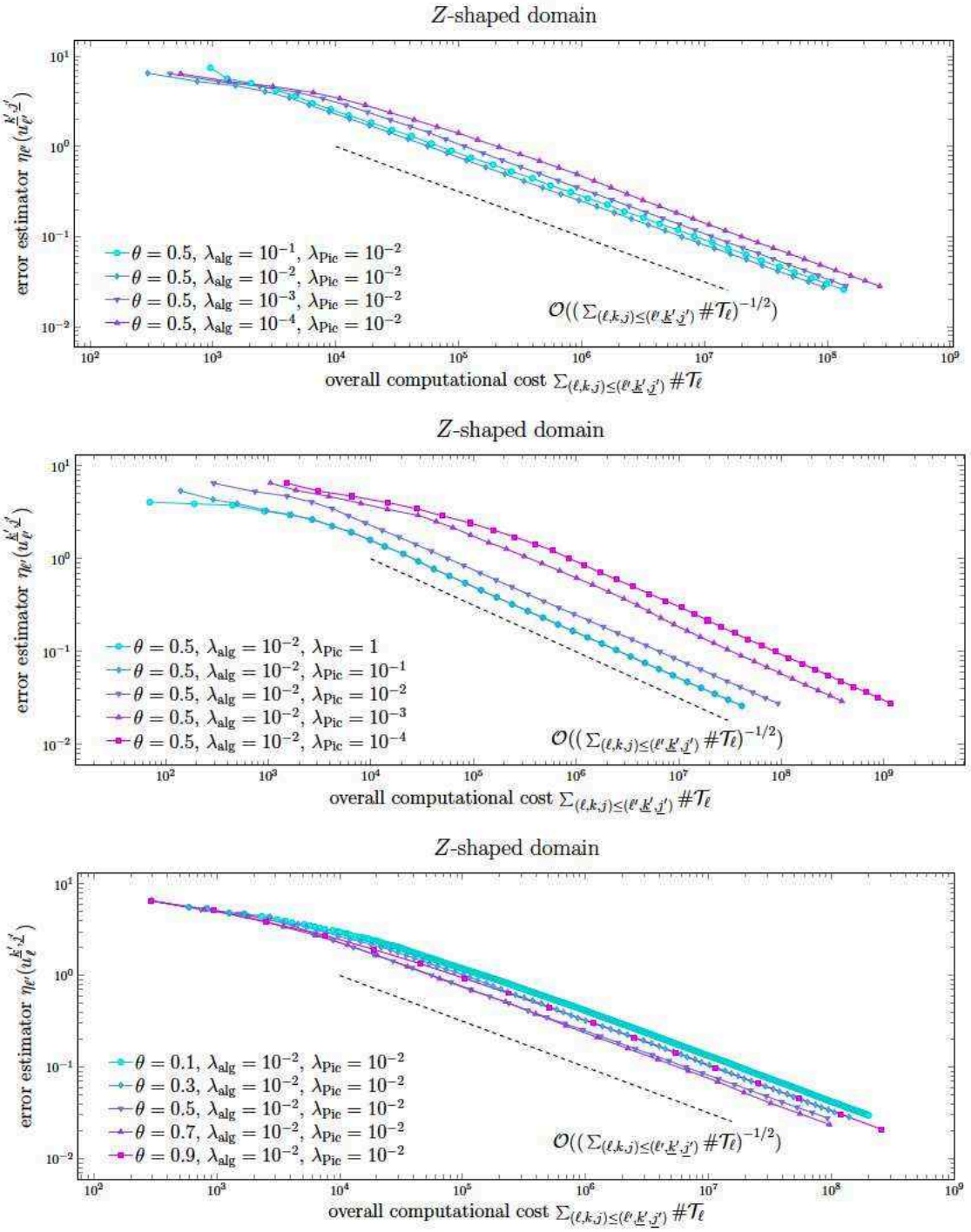


Figure 5.3: Example from Section 5.4.3 (Experiment with known solution on Z-shaped domain): Error estimator $\eta_{\ell'}(u_{\ell'}^{\underline{k}',\underline{j}'})$ on mesh $\mathcal{T}_{\ell'}$, perturbed Banach–Picard iteration $\underline{k}'$, and PCG step $\underline{j}'$ of Algorithm 41 with respect to the overall cost expressed as the cumulative sum $\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell$ for various parameters θ, λ_{Pic}, and λ_{alg}.

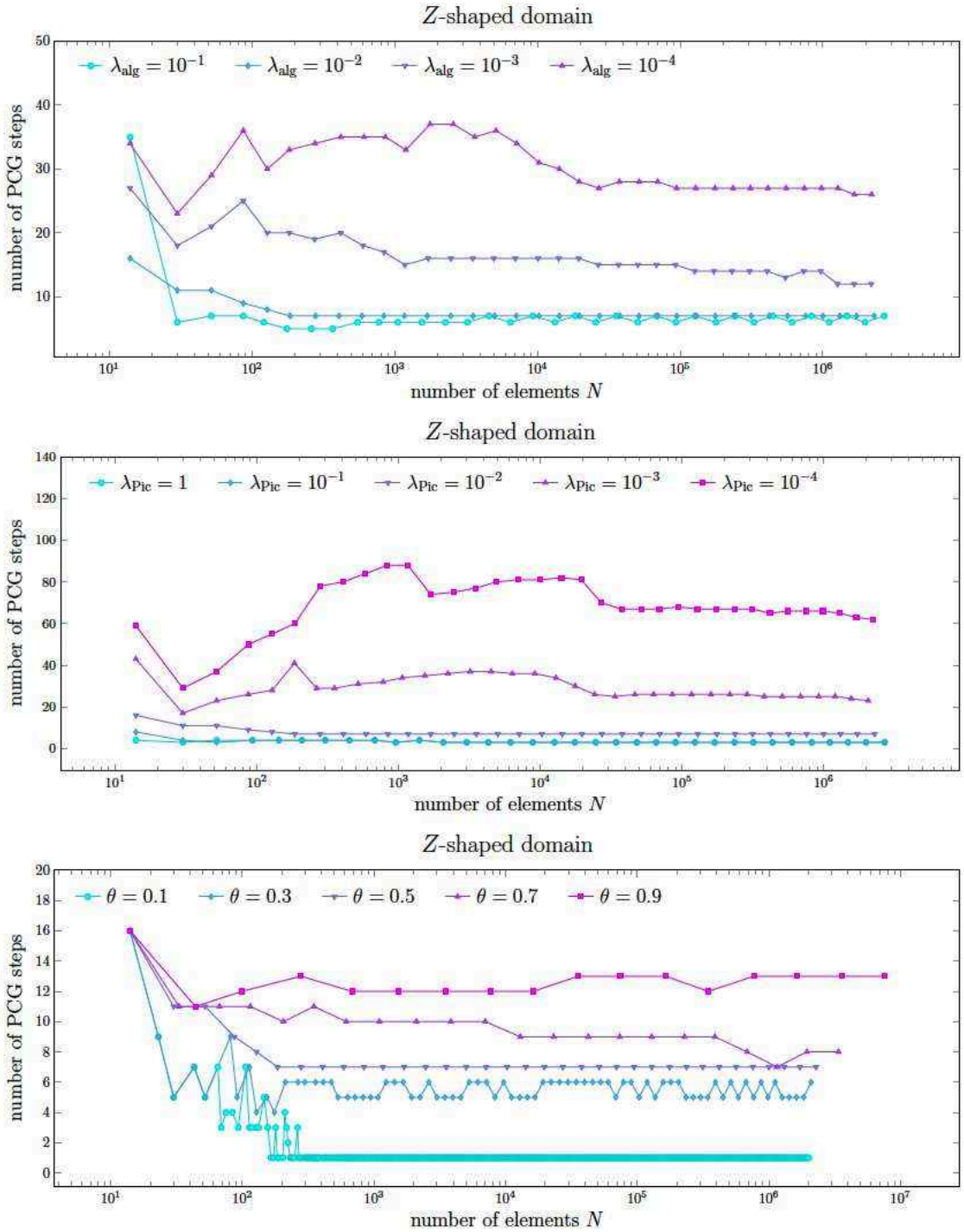


Figure 5.4: Example from Section 5.4.3 (Experiment with known solution on Z-shaped domain): Number of PCG iterations wrt. the number of elements $N := \#\mathcal{T}_\ell$ for $\theta = 0.5$, $\lambda_{\mathrm{Pic}} = 10^{-2}$, and $\lambda_{\mathrm{alg}} \in \{10^{-1}, \dots, 10^{-4}\}$ (top), for $\theta = 0.5$, $\lambda_{\mathrm{alg}} = 10^{-2}$, and $\lambda_{\mathrm{Pic}} \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (middle), and for $\lambda_{\mathrm{alg}} = \lambda_{\mathrm{Pic}} = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

what the best choice of the three parameters θ, λ_{alg}, and λ_{Pic} could be, we have to vary them all. First, we prescribe a precision $\tau = 3 \cdot 10^{-2}$ and vary $\theta \in \{0.2, 0.4, 0.6, 0.8\}$, $\lambda_{\mathrm{alg}} \in \{10^{-1}, 10^{-1.5}, \dots, 10^{-4}\}$, and $\lambda_{\mathrm{Pic}} \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$. Figure 5.5 then shows the computational cost expressed in terms of the cumulative sum $\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell$ to reach the given precision τ. It seems that a smaller value of λ_{alg} or λ_{Pic} leads to more computational cost to reach the same precision, independently of the choice of θ.

In Figure 5.6 (top), we vary $\theta \in \{0.05, 0.1, 0.15, \dots, 0.9\}$ and only print the corresponding best choices of $\lambda_{\mathrm{alg}} \in \{10^{-1}, 10^{-1.5}, \dots, 10^{-4}\}$ and $\lambda_{\mathrm{Pic}} \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$ together with the minimal overall computational cost to reach the given precision. As a result, we see that the overall best choice in terms of computational cost to reach the given precision $\tau = 3 \cdot 10^{-2}$ is $\theta = 0.7$, $\lambda_{\mathrm{alg}} = 10^{-1}$, and $\lambda_{\mathrm{Pic}} = 10^{-0.5}$ with

$$\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell = 25058328$$

where $u_\ell^{\underline{k}}$ is the first approximation such that $\eta_\ell(u_\ell^{\underline{k}}) < 3 \cdot 10^{-2}$. We also observe that the worst possible choice is $\theta = 0.05$, $\lambda_{\mathrm{alg}} = 10^{-3.5}$, and $\lambda_{\mathrm{Pic}} = 10^{-4}$. With these parameters it takes more than 1000 times the computational cost to reach the same precision in comparison to the best choice.

5.4.4 Experiment with unknown solution

We consider the L-shaped domain $\Omega \subset \mathbb{R}^2$ from Figure 5.1 (bottom) and the nonlinear problem (5.97) with $\Gamma_D = \Gamma$ and constant right-hand side $f \equiv 1$ where $\mu(\cdot,\cdot)$ is given by

$$\mu(x, |\nabla u^\star(x)|^2) := 1 + \arctan(|\nabla u^\star(x)|^2).$$

Then, according to [CW17, Example 1], there hold (N1)–(N4) with $\alpha = 1$ and $L \approx= 1 + \sqrt{3}/2 + \pi/3$, while the exact solution is unknown.

In Figure 5.7, we again test Algorithm 41 with varying θ, λ_{alg}, and λ_{Pic}. We plot the error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{j}})$ over the number of elements $N := \#\mathcal{T}_\ell$. Uniform mesh-refinement leads to the suboptimal rate of convergence $\mathcal{O}(N^{-1/3})$, whereas Algorithm 41 with adaptive mesh-refinement regains the optimal rate of convergence $\mathcal{O}(N^{-1/2})$. Again, this empirically confirms Theorem 49. The latter rate of convergence even appears to be robust with respect to θ, λ_{alg}, and λ_{Pic}.

In Figure 5.8, we plot the estimator $\eta_{\ell'}(u_{\ell'}^{\underline{k}',\underline{j}'})$ over the cumulative sum $\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell$. Independently of the choice of the parameters θ, λ_{alg}, and λ_{Pic}, we observe the optimal order of convergence $\mathcal{O}\big(\big(\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell\big)^{-1/2}\big)$ with respect to the overall computational cost, which empirically underpins Theorem 53.

In Figure 5.9, we finally consider the total number of PCG iterations cumulated over all Picard steps on the given mesh. We observe that independently of the choice of θ, λ_{alg}, and λ_{Pic}, the total number of PCG iterations stays uniformly bounded. Additionally, we see that for larger values of λ_{alg} and λ_{Pic}, as well as for smaller values of θ, the total number of PCG iterations is smaller.

$\theta = 0.2$

λ_{alg} \ λ_{Pic}	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
10^{-1}	**6.4e+07**	6.4e+07	6.5e+07	6.6e+07	1.3e+08	5.0e+08	1.6e+09	2.8e+09	**4.2e+09**
$10^{-1.5}$	6.7e+07	6.7e+07	6.7e+07	6.5e+07	1.3e+08	5.0e+08	1.6e+09	2.8e+09	4.2e+09
10^{-2}	9.4e+07	9.4e+07	9.0e+07	9.3e+07	1.2e+08	4.8e+08	1.8e+09	3.1e+09	4.2e+09
$10^{-2.5}$	1.2e+08	1.2e+08	1.2e+08	1.2e+08	2.4e+08	4.0e+08	1.4e+09	2.3e+09	3.7e+09
10^{-3}	1.8e+08	1.8e+08	1.8e+08	1.8e+08	3.5e+08	4.3e+08	7.0e+08	1.0e+09	4.0e+09
$10^{-3.5}$	2.7e+08	2.7e+08	2.7e+08	2.7e+08	5.0e+08	6.5e+08	8.7e+08	1.1e+09	1.4e+09
10^{-4}	3.4e+08	3.4e+08	3.4e+08	3.4e+08	6.1e+08	8.4e+08	1.2e+09	1.5e+09	1.7e+09

$\theta = 0.4$

λ_{alg} \ λ_{Pic}	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
10^{-1}	6.0e+07	4.6e+07	1.1e+08	9.1e+07	1.0e+08	2.8e+08	7.3e+08	1.1e+09	1.6e+09
$10^{-1.5}$	3.2e+07	3.2e+07	**3.1e+07**	5.0e+07	1.1e+08	2.5e+08	7.8e+08	1.2e+09	1.6e+09
10^{-2}	4.8e+07	4.8e+07	4.8e+07	5.1e+07	1.1e+08	2.0e+08	6.6e+08	1.1e+09	1.5e+09
$10^{-2.5}$	6.1e+07	6.1e+07	6.0e+07	6.4e+07	1.4e+08	1.8e+08	2.9e+08	7.3e+08	1.3e+09
10^{-3}	8.6e+07	8.6e+07	8.5e+07	9.0e+07	1.7e+08	2.3e+08	3.1e+08	4.1e+08	**1.8e+09**
$10^{-3.5}$	1.1e+08	1.1e+08	1.1e+08	1.2e+08	2.3e+08	3.1e+08	4.1e+08	4.8e+08	5.9e+08
10^{-4}	1.4e+08	1.4e+08	1.4e+08	1.5e+08	2.9e+08	4.1e+08	5.4e+08	6.1e+08	7.1e+08

$\theta = 0.6$

λ_{alg} \ λ_{Pic}	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
10^{-1}	2.8e+07	**2.7e+07**	1.1e+08	1.4e+08	9.3e+07	1.7e+08	5.0e+08	7.7e+08	9.7e+08
$10^{-1.5}$	2.7e+07	2.7e+07	2.8e+07	4.4e+07	1.0e+08	2.5e+08	5.1e+08	7.7e+08	**1.0e+09**
10^{-2}	3.1e+07	3.1e+07	3.1e+07	5.8e+07	7.7e+07	1.7e+08	3.8e+08	6.3e+08	1.0e+09
$10^{-2.5}$	3.7e+07	3.7e+07	3.8e+07	7.4e+07	9.5e+07	1.3e+08	2.1e+08	6.4e+08	7.2e+08
10^{-3}	6.1e+07	6.1e+07	5.8e+07	1.0e+08	1.2e+08	1.7e+08	2.2e+08	2.8e+08	6.0e+08
$10^{-3.5}$	8.6e+07	8.6e+07	8.1e+07	1.4e+08	1.7e+08	2.6e+08	2.9e+08	3.5e+08	3.9e+08
10^{-4}	1.1e+08	1.1e+08	1.1e+08	1.8e+08	2.3e+08	3.3e+08	3.8e+08	4.5e+08	5.0e+08

$\theta = 0.8$

λ_{alg} \ λ_{Pic}	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
10^{-1}	5.2e+07	5.2e+07	3.6e+07	1.4e+08	9.6e+07	2.8e+08	4.6e+08	8.4e+08	1.3e+09
$10^{-1.5}$	2.9e+07	2.9e+07	**2.6e+07**	4.1e+07	9.1e+07	3.8e+08	4.8e+08	9.2e+08	1.3e+09
10^{-2}	4.9e+07	4.9e+07	4.9e+07	5.4e+07	7.0e+07	2.1e+08	4.9e+08	9.3e+08	**1.3e+09**
$10^{-2.5}$	7.8e+07	7.8e+07	8.1e+07	8.9e+07	9.3e+07	1.1e+08	1.5e+08	5.0e+08	8.8e+08
10^{-3}	1.1e+08	1.1e+08	1.1e+08	1.3e+08	1.5e+08	1.6e+08	2.0e+08	2.5e+08	5.3e+08
$10^{-3.5}$	1.3e+08	1.3e+08	1.3e+08	1.8e+08	2.2e+08	2.5e+08	2.8e+08	3.2e+08	4.3e+08
10^{-4}	1.5e+08	1.5e+08	1.5e+08	2.3e+08	2.9e+08	3.3e+08	3.9e+08	4.3e+08	4.7e+08

min — max

Figure 5.5: Example from Section 5.4.3 (Experiment with known solution on Z-shaped domain): Overall computational cost $\sum_{(\ell,k,j)\le(\ell',\underline{k}',j')} \#\mathcal{T}_\ell$ such that $\eta_\ell(u_\ell^{\underline{k}'}) < \tau$ for given precision $\tau = 3 \cdot 10^{-2}$, $\lambda_{\mathrm{alg}} \in \{10^{-1}, 10^{-1.5}, \dots, 10^{-4}\}$, and $\lambda_{\mathrm{Pic}} \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$.

Z-shaped domain

θ	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9
min	7.4e+08	2.0e+08	9.8e+07	6.4e+07	4.7e+07	4.0e+07	4.1e+07	3.1e+07	3.4e+07	2.9e+07	2.8e+07	2.7e+07	3.0e+07	**2.5e+07**	2.5e+07	2.6e+07	3.5e+07	5.5e+07
λ_{alg}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	$10^{-1.5}$	10^{-1}	$10^{-1.5}$	$10^{-1.5}$	$10^{-1.5}$	$10^{-1.5}$	$10^{-1.5}$	10^{-1}	10^{-1}	10^{-1}	$10^{-1.5}$	$10^{-1.5}$	$10^{-1.5}$	$10^{-1.5}$
λ_{Pic}	$10^{-1.5}$	1	$10^{-1.5}$	1	10^{-1}	1	1	10^{-1}	1	10^{-1}	1	$10^{-0.5}$	1	$10^{-0.5}$	1	10^{-1}	1	1
max	**4.1e+10**	1.6e+10	6.5e+09	4.2e+09	2.9e+09	2.2e+09	1.9e+09	1.8e+09	1.3e+09	1.2e+09	1.3e+09	1.0e+09	1.1e+09	1.6e+09	1.7e+09	1.3e+09	1.7e+09	2.3e+09
λ_{alg}	$10^{-3.5}$	$10^{-3.5}$	10^{-2}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-3}	10^{-1}	$10^{-1.5}$	$10^{-1.5}$	$10^{-1.5}$	10^{-2}	10^{-2}	10^{-2}	10^{-2}	$10^{-1.5}$	$10^{-1.5}$
λ_{Pic}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}

L-shaped domain

θ	0.05	0.1	0.15	0.2	0.25	0.3	0.35	0.4	0.45	0.5	0.55	0.6	0.65	0.7	0.75	0.8	0.85	0.9
min	2.6e+07	2.4e+07	2.2e+07	2.1e+07	1.7e+07	1.2e+07	1.1e+07	1.0e+07	1.1e+07	9.4e+06	8.8e+06	1.5e+07	1.0e+07	**8.4e+06**	1.8e+07	1.7e+07	1.7e+07	2.6e+07
λ_{alg}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-1}	10^{-2}	10^{-1}	10^{-1}	$10^{-1.5}$	10^{-1}	10^{-1}	10^{-1}	$10^{-1.5}$	$10^{-1.5}$
λ_{Pic}	$10^{-1.5}$	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
max	1.7e+09	1.1e+09	9.4e+08	7.2e+08	6.1e+08	5.3e+08	5.8e+08	5.2e+08	4.9e+08	5.0e+08	4.4e+08	4.3e+08	4.5e+08	4.5e+08	9.2e+08	6.1e+08	1.1e+09	**2.1e+09**
λ_{alg}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}
λ_{Pic}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}	10^{-4}

Figure 5.6: Examples from Section 5.4: Comparison of the minimal and maximal overall computational cost $\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell$ for different values of $\theta \in \{0.05, 0.1, 0.15, \dots, 0.9\}$, $\lambda_{\text{alg}} \in \{10^{-1}, 10^{-1.5}, \dots, 10^{-4}\}$, and $\lambda_{\text{Pic}} \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$ such that $\eta_\ell(u_{\ell'}^{k'}) < \tau$ for given precision τ. Top: Example from Section 5.4.3 (Example with known solution on Z-shaped domain) with $\tau = 3 \cdot 10^{-2}$. Bottom: Example from Section 5.4.4 (Example with unknown solution on L-shaped domain) with $\tau = 10^{-2}$.

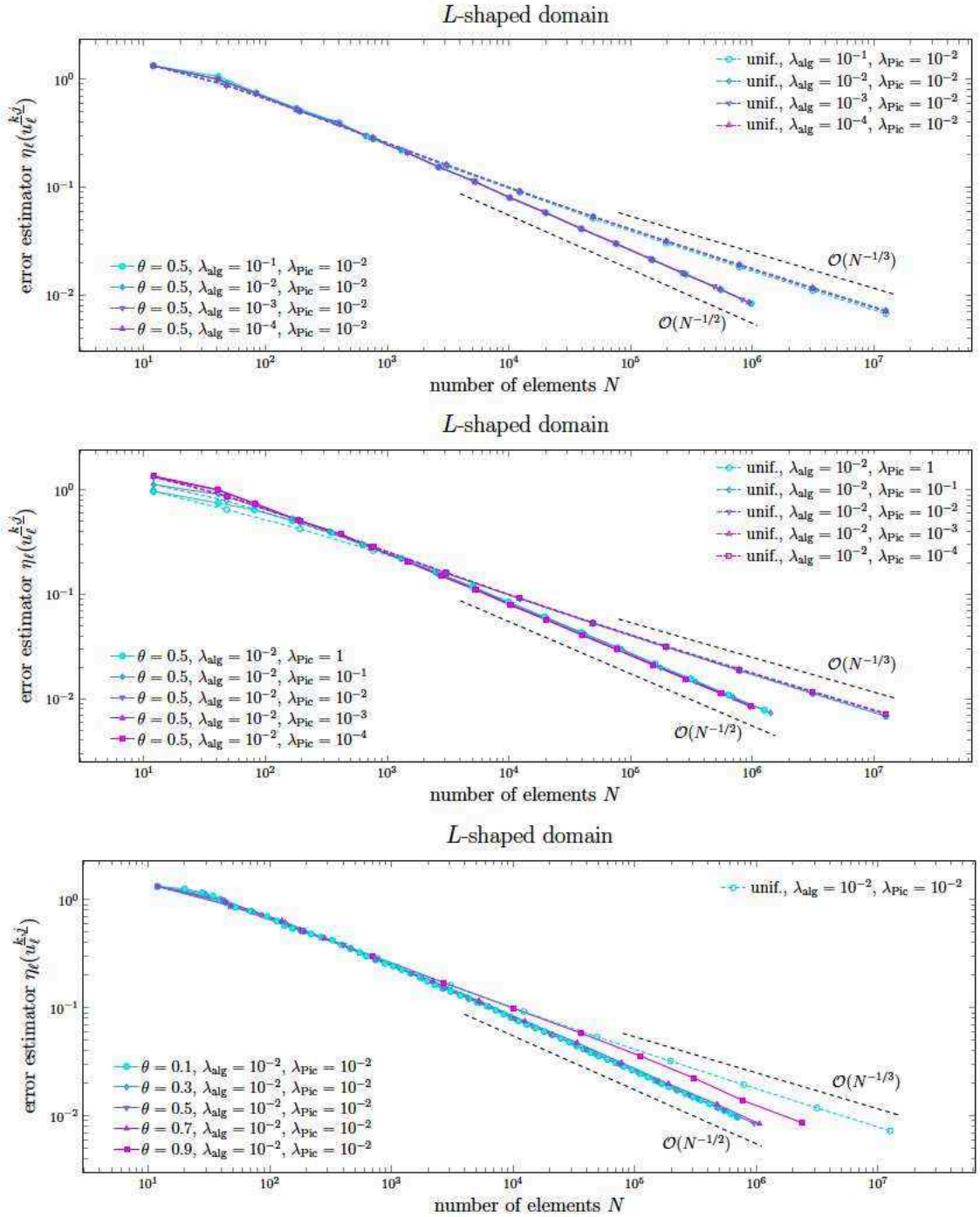


Figure 5.7: Example from Section 5.4.4 (Experiment with unknown solution on L-shaped domain): Error estimator $\eta_\ell(u_\ell^{\underline{k},\underline{j}})$ on mesh $\mathcal{T}_\ell$, perturbed Banach–Picard iteration $\underline{k}$, and PCG step $\underline{j}$ of Algorithm 41 with respect to the number of elements N of the mesh $\mathcal{T}_\ell$ for various parameters θ, λ_{Pic}, and λ_{alg}.

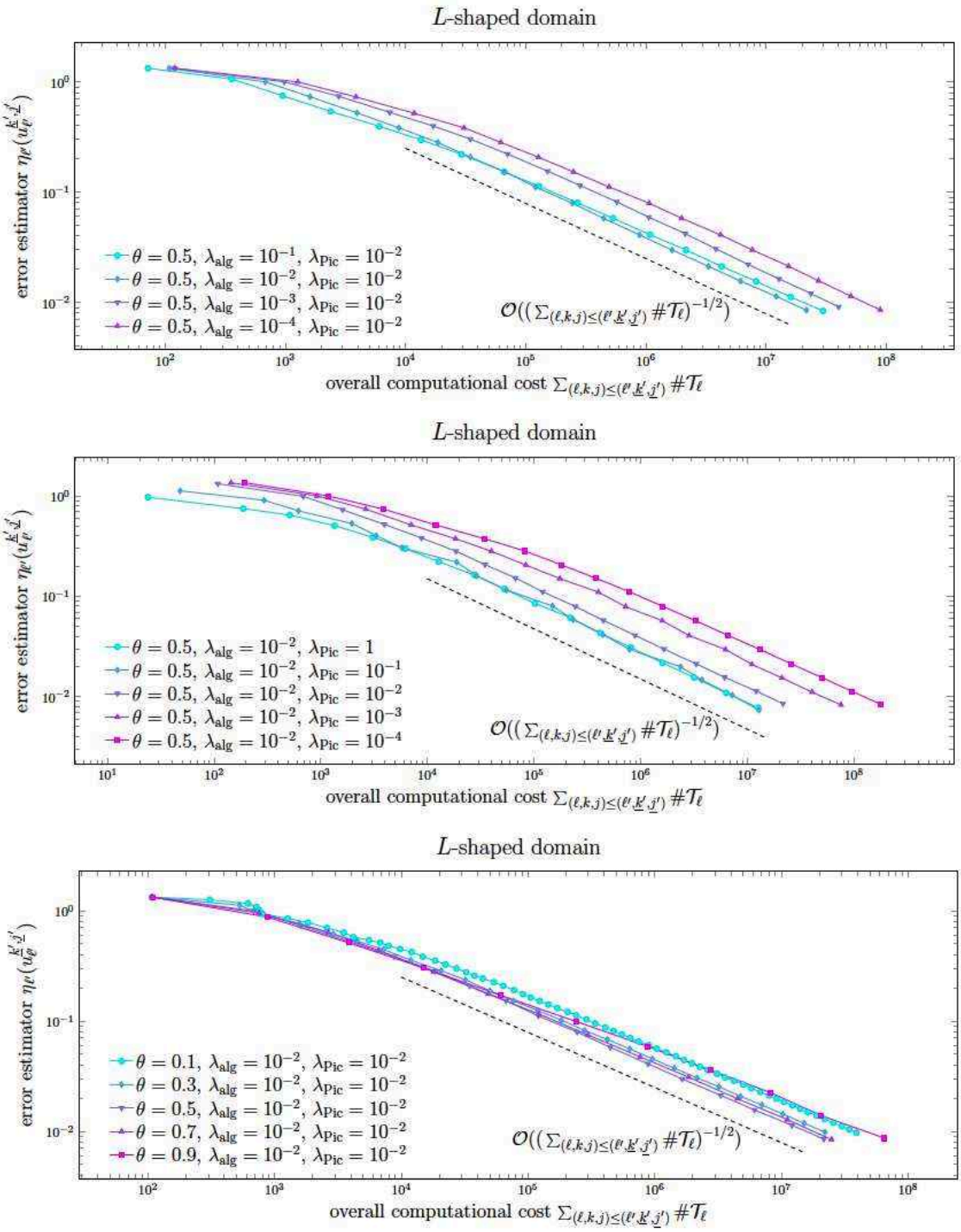


Figure 5.8: Example from Section 5.4.4 (Experiment with unknown solution on L-shaped domain): Error estimator $\eta_{\ell'}(u_{\ell'}^{\underline{k}',\underline{j}'})$ on mesh $\mathcal{T}_{\ell'}$, perturbed Banach–Picard iteration $\underline{k}'$, and PCG step $\underline{j}'$ of Algorithm 41 with respect to the overall cost expressed as the cumulative sum $\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell$ for various parameters θ, λ_{Pic}, and λ_{alg}.

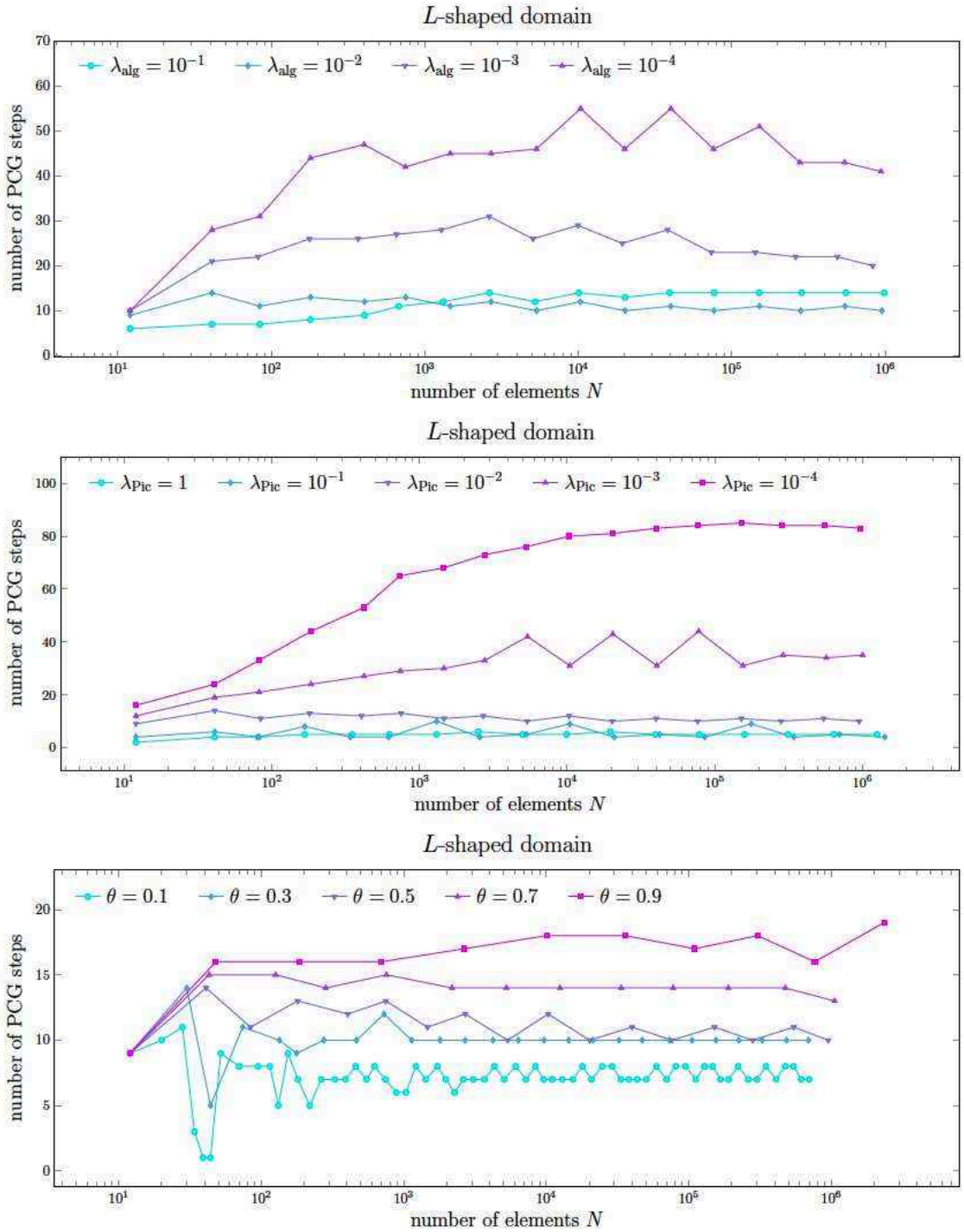


Figure 5.9: Example from Section 5.4.4 (Experiment with unknown solution on L-shaped domain): Number of PCG iterations wrt. the number of elements $N := \#\mathcal{T}_\ell$ for $\theta = 0.5$, $\lambda_{\mathrm{Pic}} = 10^{-2}$, and $\lambda_{\mathrm{alg}} \in \{10^{-1}, \dots, 10^{-4}\}$ (top), for $\theta = 0.5$, $\lambda_{\mathrm{alg}} = 10^{-2}$, and $\lambda_{\mathrm{Pic}} \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (middle), and for $\lambda_{\mathrm{alg}} = \lambda_{\mathrm{Pic}} = 10^{-2}$ and $\theta \in \{0.1, 0.3, \dots, 0.9\}$ (bottom).

In Figure 5.10, we again compare the computational cost of Algorithm 41 to reach the given precision $\tau = 10^{-2}$ for various θ, λ_{alg}, and λ_{Pic}. Also in this experiment, it seems that a smaller value of λ_{alg} or λ_{Pic} leads to more computational cost to reach the same precision, independently of the choice of θ.

In Figure 5.6 (bottom), we vary $\theta \in \{0.05, 0.1, 0.15, \dots, 0.9\}$ and print the corresponding best and worst choices of $\lambda_{\mathrm{alg}} \in \{10^{-1}, 10^{-1.5}, \dots, 10^{-4}\}$ and $\lambda_{\mathrm{Pic}} \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$ respectively, together with the overall computational cost to reach the given precision. As a result, we see that the overall best choice in terms of computational cost to reach the given precision $\tau = \cdot 10^{-2}$ is $\theta = 0.7$, $\lambda_{\mathrm{alg}} = 10^{-1}$, and $\lambda_{\mathrm{Pic}} = 1$ with

$$\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell = 25058328$$

where u_ℓ^k is the first approximation such that $\eta_\ell(u_\ell^k) < 10^{-2}$. We also observe that the worst possible choice is $\theta = 0.9$, $\lambda_{\mathrm{alg}} = 10^{-4}$, and $\lambda_{\mathrm{Pic}} = 10^{-4}$. With these parameters it takes more than 200 times the computational cost to reach the same precision in comparison to the best choice. Independently of θ, the worst choice of λ_{alg} and λ_{Pic} is always $\lambda_{\mathrm{alg}} = \lambda_{\mathrm{Pic}} = 10^{-4}$.

$\theta = 0.2$

λ_{alg} \ λ_{Pic}	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
10^{-1}	**2.1e+07**	2.1e+07	2.1e+07	4.2e+07	3.4e+07	8.3e+07	2.0e+08	2.8e+08	4.1e+08
$10^{-1.5}$	2.4e+07	2.4e+07	3.0e+07	2.5e+07	3.4e+07	8.3e+07	2.0e+08	2.8e+08	4.1e+08
10^{-2}	2.7e+07	2.7e+07	2.9e+07	3.7e+07	4.1e+07	8.3e+07	2.0e+08	2.8e+08	4.1e+08
$10^{-2.5}$	3.4e+07	3.4e+07	4.5e+07	3.6e+07	4.6e+07	9.9e+07	2.1e+08	3.0e+08	4.3e+08
10^{-3}	5.6e+07	5.6e+07	6.1e+07	5.1e+07	4.8e+07	9.9e+07	2.9e+08	4.0e+08	5.3e+08
$10^{-3.5}$	7.5e+07	7.5e+07	8.3e+07	8.6e+07	6.2e+07	1.1e+08	3.0e+08	4.7e+08	6.4e+08
10^{-4}	9.9e+07	9.9e+07	1.1e+08	1.2e+08	9.1e+07	1.2e+08	3.2e+08	4.8e+08	**7.2e+08**

$\theta = 0.4$

λ_{alg} \ λ_{Pic}	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
10^{-1}	**1.0e+07**	2.5e+07	2.1e+07	6.1e+07	4.0e+07	6.6e+07	1.4e+08	2.2e+08	3.0e+08
$10^{-1.5}$	1.4e+07	2.8e+07	4.0e+07	4.1e+07	4.0e+07	6.6e+07	1.4e+08	2.2e+08	3.0e+08
10^{-2}	1.6e+07	3.5e+07	3.9e+07	5.9e+07	5.9e+07	6.1e+07	1.4e+08	2.2e+08	3.0e+08
$10^{-2.5}$	2.4e+07	5.4e+07	5.8e+07	5.7e+07	5.9e+07	9.8e+07	1.8e+08	2.5e+08	3.3e+08
10^{-3}	3.4e+07	8.3e+07	8.2e+07	5.1e+07	6.3e+07	1.0e+08	2.3e+08	3.2e+08	4.0e+08
$10^{-3.5}$	4.6e+07	1.1e+08	1.2e+08	8.0e+07	8.1e+07	1.0e+08	2.3e+08	3.6e+08	4.7e+08
10^{-4}	5.5e+07	1.4e+08	1.6e+08	1.2e+08	1.2e+08	1.2e+08	2.4e+08	3.8e+08	**5.2e+08**

$\theta = 0.6$

λ_{alg} \ λ_{Pic}	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
10^{-1}	**1.5e+07**	5.0e+07	7.6e+07	4.1e+07	6.7e+07	6.4e+07	1.5e+08	1.8e+08	2.4e+08
$10^{-1.5}$	1.9e+07	2.8e+07	3.5e+07	4.3e+07	6.7e+07	6.4e+07	1.5e+08	1.8e+08	2.4e+08
10^{-2}	2.5e+07	4.2e+07	3.8e+07	5.6e+07	6.3e+07	6.4e+07	1.6e+08	1.8e+08	2.4e+08
$10^{-2.5}$	2.9e+07	6.0e+07	5.5e+07	5.6e+07	6.2e+07	9.5e+07	1.9e+08	2.1e+08	2.7e+08
10^{-3}	4.1e+07	8.9e+07	8.3e+07	8.3e+07	6.6e+07	9.9e+07	3.1e+08	2.6e+08	3.2e+08
$10^{-3.5}$	5.6e+07	1.2e+08	1.2e+08	1.3e+08	1.2e+08	1.0e+08	2.9e+08	3.0e+08	3.8e+08
10^{-4}	6.9e+07	1.5e+08	1.7e+08	1.9e+08	1.4e+08	1.4e+08	3.6e+08	3.3e+08	**4.3e+08**

$\theta = 0.8$

λ_{alg} \ λ_{Pic}	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
10^{-1}	**1.7e+07**	5.9e+07	4.5e+07	4.4e+07	7.2e+07	1.3e+08	1.7e+08	2.5e+08	3.3e+08
$10^{-1.5}$	2.0e+07	6.8e+07	6.5e+07	6.3e+07	7.2e+07	1.3e+08	1.7e+08	2.5e+08	3.3e+08
10^{-2}	3.2e+07	8.2e+07	6.8e+07	7.3e+07	1.1e+08	1.3e+08	1.7e+08	2.5e+08	3.2e+08
$10^{-2.5}$	4.6e+07	1.4e+08	9.8e+07	6.9e+07	1.0e+08	2.0e+08	2.3e+08	2.9e+08	3.6e+08
10^{-3}	7.1e+07	2.1e+08	1.5e+08	1.1e+08	1.1e+08	2.2e+08	2.8e+08	3.5e+08	4.2e+08
$10^{-3.5}$	9.2e+07	2.9e+08	2.2e+08	1.8e+08	1.8e+08	2.6e+08	3.1e+08	4.1e+08	5.0e+08
10^{-4}	1.1e+08	3.6e+08	3.0e+08	2.5e+08	2.9e+08	3.9e+08	3.9e+08	4.8e+08	**6.1e+08**

min max

Figure 5.10: Example from Section 5.4.3 (Experiment with known solution on Z-shaped domain): Overall computational cost $\sum_{(\ell,k,j)\le(\ell',\underline{k}',\underline{j}')} \#\mathcal{T}_\ell$ such that $\eta_\ell(u_{\ell'}^{\underline{k}'}) < \tau$ for given precision $\tau = 3 \cdot 10^{-2}$, $\lambda_{\mathrm{alg}} \in \{10^{-1}, 10^{-1.5}, \dots, 10^{-4}\}$, and $\lambda_{\mathrm{Pic}} \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$.

6 Adaptive BEM for elliptic first-kind integral equations with optimal PCG solver

6.1 Introduction

In this chapter, which is based on [FHPS19], we consider the boundary element method (BEM) subject to elliptic first-kind integral equations. We introduce our adaptive algorithm which steers both the adaptive mesh-refinement as well as the termination of the preconditioned conjugate gradient method (PCG) with optimal preconditioner, i.e., an inexact solver for the arising Galerkin system. The main results are then convergence with optimal algebraic rates as well as almost optimal computational complexity.

6.1.1 State of the art

In the last decade, the mathematical understanding of adaptive mesh-refinement has matured. We refer to [Dör96, MNS00, BDD04, Ste07, CKNS08, FFP14] for some milestones for adaptive finite element methods for second-order linear elliptic equations, [Gan13, FKMP13, FFK+14, FFK+15, AFF+17] for adaptive BEM, and [CFPP14] for a general framework of rate-optimality of adaptive mesh-refining algorithms. The interplay between adaptive mesh-refinement, optimal convergence rates, and inexact solvers has been addressed and analyzed for adaptive FEM for linear problems in [Ste07, ALMS13, AGL13], for eigenvalue problems in [CG12], and recently also for strongly monotone nonlinearities in [GHPS18]. In particular, all available results for adaptive BEM [Gan13, FKMP13, FFK+14, FFK+15, AFF+17] assume that the arising Galerkin system $\mathbf{A}_\ell \mathbf{x}_\ell^\star = \mathbf{b}_\ell$ is solved exactly. Instead, we omit the latter assumption and analyze an adaptive algorithm which steers both the local mesh-refinement and the iterations of an inexact PCG solver.

In principle, it is known [CFPP14, Section 7] that convergence and optimal convergence rates are preserved if the linear system is solved inexactly, but with sufficient accuracy. The aim now is to guarantee the latter by incorporating an appropriate stopping criterion for the PCG solver into the adaptive algorithm. Moreover, to prove that the proposed algorithm does not only lead to optimal algebraic convergence rates, but also to (almost) optimal computational cost, we provide an appropriate symmetric and positive definite preconditioner $\mathbf{P}_\ell \in \mathbb{R}^{N\times N}$ such that

- first, the matrix-vector products with $\mathbf{P}_\ell^{-1}$ can be computed at linear cost and
- second, the system matrix $\mathbf{P}_\ell^{-1/2}\mathbf{A}_\ell\mathbf{P}_\ell^{-1/2}$ of the preconditioned linear system

$$\mathbf{P}_\ell^{-1/2}\mathbf{A}_\ell\mathbf{P}_\ell^{-1/2}\widetilde{\mathbf{x}}_\ell^\star = \mathbf{P}_\ell^{-1/2}\mathbf{b}_\ell \tag{6.1}$$

has a uniformly bounded condition number which is independent of the mesh $\mathcal{T}_\ell$.

Then, $\mathbf{x}_\ell^\star = \mathbf{P}_\ell^{-1/2}\widetilde{\mathbf{x}}_\ell^\star$ solves the original system $\mathbf{A}_\ell\,\mathbf{x}_\ell^\star = \mathbf{b}_\ell$. To that end, we exploit the multilevel structure of adaptively generated meshes in the framework of additive Schwarz methods. For hyper-singular integral equations, such a multilevel additive Schwarz preconditioner has been proposed and analyzed in [FFPS17a, FMPR15] for $d = 2, 3$ and for weakly-singular integral equations in [FFPS17b] for $d = 2$. In particular, were able to close this gap by analyzing an optimal additive Schwarz preconditioner for weakly-singular integral equations for $d = 3$. Besides, we refer to [SvV20] for optimal preconditioning in Hilbert spaces of negative order. We note that the proofs of [FFPS17a, FFPS17b] do not transfer to weakly-singular integral equations for $d = 3$. Instead, we build on recent results for finite element discretizations [HWZ12, AGS16] which are then transferred to the present BEM setting by use of an abstract concept from [Osw99].

6.1.2 Outline

Section 6.2 introduces the functional analytic framework and fixes the necessary notation. In Section 6.3, we introduce the weakly-singular integral equation which serves as our model problem and give a short introduction to BEM, before we state our adaptive algorithm in Section 6.4 which steers the local mesh-refinement as well as the stopping of the PCG iteration. Section 6.5 states our main results. In Section 6.5.1, we define a local multilevel additive Schwarz preconditioner (6.36) for a sequence of locally refined meshes. Theorem 60 states that the ℓ_2-condition number of the preconditioned systems is uniformly bounded for all these meshes, i.e., the preconditioner is optimal. Theorem 68 proves

- that the overall error in the energy norm can be controlled *a posteriori*,
- that the *quasi-error* (which consists of energy norm error plus error estimator) is linearly convergent in each step of the adaptive algorithm (i.e., independent of whether the algorithm decides for local mesh-refinement or for one step of the PCG iteration),
- that the quasi-error even decays with optimal rate (i.e., with each possible algebraic rate) with respect to the degrees of freedom, i.e., Algorithm 57 is *rate optimal* in the sense of, e.g., [Ste07, CKNS08, FKMP13, CFPP14].

Finally, Section 6.5.5 considers the computational cost. Under realistic assumptions on the treatment of the arising discrete integral operators, Corollary 78 states that the quasi-error converges at almost optimal rate (i.e., with rate $s-\varepsilon$ for any $\varepsilon > 0$ if rate $s > 0$ is possible for the exact Galerkin solution) with respect to computational cost, i.e., Algorithm 57 requires *almost optimal computational time*. Section 6.6 shows that our main results also apply to the hyper-singular integral equation. The final Section 6.7 underpins the theoretical findings by some 2D and 3D experiments.

6.2 Preliminaries and notation

6.2.1 Boundary integral operators and functional analytic setting

Let $\Omega \subset \mathbb{R}^d$ with $d = 2, 3$ be a bounded Lipschitz domain with boundary $\Gamma := \partial\Omega$. We consider the usual Laplace problem

$$-\Delta u = 0 \quad \text{in } \Omega \tag{6.2}$$

with appropriate boundary conditions, i.e., Dirichlet or Neumann boundary conditions on the boundary Γ, where either u or the normal derivative $\partial_{\mathbf{n}} u$ respectively are given on Γ. Solutions to these problems can be represented via potentials which are closely related to the fundamental solution $G(\cdot)$ of the Laplace operator, i.e.,

$$G(z) = \begin{cases} -\frac{1}{2\pi} \log|z| & \text{for } d = 2, \\ \frac{1}{4\pi} \frac{1}{|z|} & \text{for } d = 3. \end{cases}$$

For smooth solutions $u \in C^2(\overline{\Gamma})$ of (6.2), there holds the following representation formula, cf. [SS11, Theorem 3.1.6],

$$u(x) = \int_\Gamma G(x-y)\, \partial_{\mathbf{n}(y)} u(y)\, \mathrm{d}s_y - \int_\Gamma \partial_{\mathbf{n}(y)} G(x-y) u(y)\, \mathrm{d}s_y \quad \text{for all } x \in \Omega, \tag{6.3}$$

where $\partial_{\mathbf{n}(y)}$ is the normal derivative with respect to $y \in \Gamma$. Hence, depending on the given boundary conditions, the unknown quantity is either $\partial_{\mathbf{n}} u$ or u.

First, we define the single-layer potential S for $\phi \in L^1(\Gamma)$ by

$$(S\phi)(x) := \int_\Gamma G(x-y)\, \phi(y)\, \mathrm{d}s_y \quad \text{for all } x \in \mathbb{R}^d \setminus \Gamma,$$

as well as the double-layer potential D for $\phi \in L^1(\Gamma)$ by

$$(D\phi)(x) := \int_\Gamma \partial_{\mathbf{n}(y)} G(x-y)\, \phi(y)\, \mathrm{d}s_y \quad \text{for all } x \in \mathbb{R}^d \setminus \Gamma.$$

Recalling the Sobolev spaces on the boundary from Section 2.3 and Section 2.4, these potentials give rise to bounded linear operators

$$S\colon \widetilde{H}^{-1/2+s}(\Gamma) \to H^1_{\mathrm{loc}}(\mathbb{R}^d) \quad \text{and} \quad D\colon H^{1/2+s}(\Gamma) \to H^1_{\mathrm{loc}}(\mathbb{R}^d) \tag{6.4}$$

with $-1/2 \le s \le 1/2$, where $H^1_{\mathrm{loc}}(\mathbb{R}^d)$ is the space of H^1-functions with compact support, cf. [SS11, Theorem 3.1.16, Remark 3.1.18].

Recalling the trace operators γ_0^{int}, γ_0^{ext} as well as the normal derivative operators γ_1^{int}, γ_1^{ext} from Section 2.5, [SS11, Theorem 3.3.1] shows that

$$\gamma_0^{\mathrm{int}} S\phi = \gamma_0^{\mathrm{ext}} S\phi \quad \text{and} \quad \gamma_1^{\mathrm{int}} D\psi = \gamma_1^{\mathrm{ext}} D\psi. \tag{6.5}$$

Thereof, we omit the superscript for ease of notation and define the following linear and continuous boundary integral operators:

- single-layer operator

$$V\colon \widetilde{H}^{-1/2+s}(\Gamma) \to H^{1/2+s}(\Gamma) \quad \text{with} \quad V\phi := \gamma_0 S\phi \tag{6.6}$$

- double-layer operator

$$K\colon H^{1/2+s}(\Gamma) \to H^{1/2+s}(\Gamma) \quad \text{with} \quad Kg := \frac{1}{2}(\gamma_0^{\mathrm{int}} D + \gamma_0^{\mathrm{ext}} D)g \tag{6.7}$$

- adjoint double-layer operator

$$K'\colon \widetilde{H}^{-1/2+s}(\Gamma) \to H^{-1/2+s}(\Gamma) \quad \text{with} \quad K'\phi := -\frac{1}{2}\mathrm{Id}\phi + \gamma_1^{\mathrm{int}} S\phi \tag{6.8}$$

- hyper-singular operator

$$W\colon \widetilde{H}^{1/2+s}(\Gamma) \to H^{-1/2+s}(\Gamma) \quad \text{with} \quad W\psi := -\gamma_1 D\psi \tag{6.9}$$

Some important properties of these operators are summerized in the following remark and we refer to [McL00, SS11] for further details and proofs.

Remark 55. *Let $-1/2 \le s \le 1/2$ and $\Gamma \subseteq \partial\Omega$ be a (relatively) open and connected subset.*

- *The single-layer operator V from (6.6) is a bounded linear operator which is even an isomorphism for $-1/2 < s < 1/2$. For $d = 2$, this requires that the domain Ω is sufficiently small, i.e., $\operatorname{diam}(\Omega) < 1$, which can always be ensured by scaling of Ω. For $s = 0$, the operator V is even symmetric and elliptic.*
- *The hyper-singular operator W from (6.9) is a bounded linear operator which is even an isomorphism for $-1/2 < s < 1/2$. For $s = 0$, the operator W is symmetric and (since Γ is connected) positive semi-definite with kernel being the constant functions. Hence, for $\Gamma \subsetneq \partial\Omega$, the operator W is an elliptic isomorphism.*

For ease of presentation, the main part of this chapter focuses on the so-called weakly-singular integral equation which corresponds to Dirichlet boundary conditions, i.e., $u = g$ on Γ for a given function $g \in H^{1/2}(\Gamma)$. Due to the representation formula (6.3), we know that the solution u is given in terms of the trace of u on Γ as well as the normal derivative $\partial_{\mathbf{n}} u$ on Γ. This normal derivative $\phi := \partial_{\mathbf{n}} u$ is given by Symm's integral equation

$$V\phi = (K + \frac{1}{2}\mathrm{Id})g \quad \text{on } \Gamma, \tag{6.10}$$

where Id is the usual identity operator.

However, we restrict ourself to an indirect formulation, where the solution u of the Dirichlet problem is given in terms of the single-layer potential

$$u = S\phi,$$

where ϕ is the solution of

$$V\phi = g \quad \text{on } \Gamma.$$

6.3 Model problem and boundary element method (BEM)

Let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with $d \in \{2, 3\}$ and polyhedral boundary $\partial\Omega$. Let $\Gamma \subseteq \partial\Omega$ be a (relatively) open and connected subset. Given $f : \Gamma \to \mathbb{R}$, we seek the density $\phi^\star \colon \Gamma \to \mathbb{R}$ of the weakly-singular integral equation

$$(V\phi^\star)(x) = \int_\Gamma G(x-y)\phi^\star(y)\,\mathrm{d}y = f(x) \quad \text{for all } x \in \Gamma. \tag{6.11}$$

From Remark 55 follows that, for $s = 0$, the operator V is even symmetric and elliptic, i.e.,

$$\langle\!\langle \phi\,,\,\psi \rangle\!\rangle := \int_\Gamma (V\phi)(x)\,\psi(x)\,\mathrm{d}x \qquad \text{for all } \phi, \psi \in \widetilde{H}^{-1/2}(\Gamma) \tag{6.12}$$

defines a scalar product and $|\!|\!| \phi |\!|\!|^2 := \langle\!\langle \phi\,,\,\phi \rangle\!\rangle$ is an equivalent norm on $\widetilde{H}^{-1/2}(\Gamma)$. For a given right-hand side $f \in H^{1/2}(\Gamma)$, the weakly-singular integral equation (6.11) can thus equivalently be reformulated as

$$\langle\!\langle \phi^\star\,,\,\psi \rangle\!\rangle = \langle f\,,\,\psi\rangle \quad \text{for all } \psi \in \widetilde{H}^{-1/2}(\Gamma). \tag{6.13}$$

In particular, the Lax–Milgram theorem proves existence and uniqueness of the solution $\phi^\star \in \widetilde{H}^{-1/2}(\Gamma)$ to (6.13).

Given a mesh $\mathcal{T}_\bullet$ of Γ, we employ a lowest-order Galerkin boundary element method (BEM) to compute a $\mathcal{T}_\bullet$-piecewise constant function $\phi_\bullet^\star \in \mathcal{P}^0(\mathcal{T}_\bullet)$, where $\mathcal{P}^0(\mathcal{T}_\bullet)$ is defined by

$$\mathcal{P}^0(\mathcal{T}_\bullet) := \big\{\psi_\bullet : \Gamma \to \mathbb{R} \,:\, \forall T \in \mathcal{T}_\bullet \quad \psi_\bullet|_T \text{ is constant}\big\}. \tag{6.14}$$

Note that $\mathcal{P}^0(\mathcal{T}_\bullet) \subset L^2(\Gamma) \subset \widetilde{H}^{-1/2}(\Gamma)$. Hence, the weakly-singular integral equation (6.11) can be reformulated for the lowest-order space $\mathcal{P}^0(\mathcal{T}_\bullet)$ as

$$\int_\Gamma (V\phi_\bullet^\star)(x)\,\psi_\bullet(x)\,\mathrm{d}x = \int_\Gamma f(x)\,\psi_\bullet(x)\,\mathrm{d}x \quad \text{for all } \psi_\bullet \in \mathcal{P}^0(\mathcal{T}_\bullet), \tag{6.15}$$

which again can be written equivalently as

$$\langle\!\langle \phi_\bullet^\star\,,\,\psi_\bullet \rangle\!\rangle = \langle f\,,\,\psi_\bullet\rangle \quad \text{for all } \psi_\bullet \in \mathcal{P}^0(\mathcal{T}_\bullet). \tag{6.16}$$

Therefore, the Lax–Milgram theorem proves existence and uniqueness of the discrete solution $\phi_\bullet^\star \in \mathcal{P}^0(\mathcal{T}_\bullet)$.

With the numbering $\mathcal{T}_\bullet = \{T_1, \dots, T_N\}$, consider the standard basis $\big\{\chi_{\bullet,j} \,:\, j = 1, \dots, N\big\}$ of $\mathcal{P}^0(\mathcal{T}_\bullet)$ consisting of characteristic functions $\chi_{\bullet,j}$ of $T_j \in \mathcal{T}_\bullet$. We make the ansatz

$$\phi_\bullet^\star = \sum_{k=1}^N \mathbf{x}_\bullet^\star[k]\,\chi_{\bullet,k} \tag{6.17}$$

with coefficient vector

$$\mathbf{x}_\bullet^\star = (\mathbf{x}_\bullet^\star[1], \dots, \mathbf{x}_\bullet^\star[N]) \in \mathbb{R}^N.$$

Figure 6.1: For newest vertex bisection (NVB) in 2D, each triangle $T \in \mathcal{T}$ has one *reference edge*, indicated by the double line (left). Bisection of T is achieved by halving the reference edge (middle). The reference edges of the sons are always opposite to the new vertex. Recursive application of this refinement rule leads to conforming triangulations.

Then, the Galerkin formulation (6.15) is equivalent to the linear system

$$\mathbf{A}_\bullet \mathbf{x}_\bullet^\star = \mathbf{b}_\bullet \tag{6.18}$$

with

$$\mathbf{A}_\bullet[j,k] := \int_{T_j} (V\chi_{\bullet,k})(x)\,\mathrm{d}x, \quad \mathbf{b}_\bullet[j] := \int_{T_j} f(x)\,\mathrm{d}x,$$

where the matrix $\mathbf{A}_\bullet \in \mathbb{R}^{N\times N}$ is positive definite and symmetric. For a given initial triangulation $\mathcal{T}_0$, we consider an adaptive mesh-refinement strategy of the type

$$\boxed{\text{solve}} \longrightarrow \boxed{\text{estimate}} \longrightarrow \boxed{\text{mark}} \longrightarrow \boxed{\text{refine}} \tag{6.19}$$

which generates a sequence $\mathcal{T}_\ell$ of successively refined triangulations $\mathcal{T}_\ell$ for all $\ell \in \mathbb{N}_0$. We note that the condition number of the Galerkin matrix $\mathbf{A}_\ell$ from (6.18) depends on the number of elements of $\mathcal{T}_\ell$, as well as the minimal and maximal diameter. Therefore, the step $\boxed{\text{solve}}$ requires an efficient preconditioner as well as an appropriate iterative solver.

6.3.1 Mesh-refinement

We briefly recall some definitions for boundary meshes and mesh-refinement from Section 3.2 and Section 3.4 respectively in the context of this chapter.

2D BEM

For $d = 2$, a mesh $\mathcal{T}_\bullet$ of Γ is a partition into non-degenerate compact line segments. It is called γ-*shape regular*, if

$$\max\big\{h_T/h_{T'} \,:\, T, T' \in \mathcal{T}_\bullet \text{ with } T \cap T' \neq \emptyset\big\} \leq \gamma. \tag{6.20}$$

Here, $h_T := \operatorname{diam}(T) > 0$ denotes the Euclidean diameter of T, i.e., the length of the line segment.

We employ the extended bisection algorithm from [AFF+13], cf. Section 3.5. For a mesh $\mathcal{T}_\bullet$ and a subset $\mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet$, let $\mathcal{T}_\circ := \texttt{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$ be the coarsest mesh such that all marked elements $T \in \mathcal{M}_\bullet$ have been refined, i.e., $\mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet \backslash \mathcal{T}_\circ$. We write $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$, if there exists $n \in \mathbb{N}_0$, conforming triangulations $\mathcal{T}_0, \dots, \mathcal{T}_n$ and corresponding sets of marked elements $\mathcal{M}_j \subseteq \mathcal{T}_j$ such that

- $\mathcal{T}_\bullet = \mathcal{T}_0$,

- $\mathcal{T}_{j+1} = \texttt{refine}(\mathcal{T}_j, \mathcal{M}_j)$ for all $j = 0, \dots, n-1$,
- $\mathcal{T}_\circ = \mathcal{T}_n$,

i.e., $\mathcal{T}_\circ$ is obtained from $\mathcal{T}_\bullet$ by finitely many steps of refinement. Note that the extended 1D bisection, i.e., Algorithm 9, guarantees, in particular, that all $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$ are uniformly γ-shape regular, where γ depends only on $\mathcal{T}_\bullet$, cf. Section 3.5.

3D BEM

For $d = 3$, a mesh $\mathcal{T}_\bullet$ of Γ is a conforming triangulation into non-degenerate compact surface triangles. In particular, we avoid hanging nodes. To ease the presentation, we suppose that the elements $T \in \mathcal{T}_\bullet$ are flat. For a γ-shape regular triangulation, it holds that

$$\max_{T \in \mathcal{T}_\bullet} \frac{\operatorname{diam}(T)}{h_T} \le \gamma, \tag{6.21}$$

cf. Lemma 8. Here, $\operatorname{diam}(T)$ denotes the Euclidean diameter of T and $h_T := |T|^{1/2}$ with $|T|$ being the two-dimensional surface measure. Note that γ-shape regularity implies that $h_T \le \operatorname{diam}(T) \le \gamma\, h_T$ and hence excludes anisotropic elements.

For 3D BEM, we employ 2D newest vertex bisection (NVB) to refine triangulations locally, cf. Section 3.6 for details and Figure 6.1 for an illustration. For a mesh $\mathcal{T}_\bullet$ and $\mathcal{M}_\bullet \subseteq \mathcal{T}_\bullet$, we employ the same notation $\mathcal{T}_\circ := \texttt{refine}(\mathcal{T}_\bullet, \mathcal{M}_\bullet)$ and $\mathcal{T}_\circ \in \texttt{refine}(\mathcal{T}_\bullet)$ respectively as for $d = 2$.

6.3.2 *A posteriori* BEM error control

For $\psi_\bullet \in \mathcal{P}^0(\mathcal{T}_\bullet)$ and $\mathcal{U}_\bullet \subseteq \mathcal{T}_\bullet$, define

$$\eta_\bullet(\mathcal{U}_\bullet, \psi_\bullet)^2 := \sum_{T \in \mathcal{U}_\bullet} \eta_\bullet(T, \psi_\bullet)^2, \tag{6.22}$$

where

$$\eta_\bullet(T, \psi_\bullet)^2 := h_T \, \|\nabla_\Gamma (f - V\psi_\bullet)\|^2_{L^2(T)} \quad \text{for all } T \in \mathcal{T}_\bullet. \tag{6.23}$$

Here $\nabla_\Gamma(\cdot)$ denotes the arclength derivative for $d = 2$ resp. the surface gradient for $d = 3$. To abbreviate notation, let $\eta_\bullet(\psi_\bullet) := \eta_\bullet(\mathcal{T}_\bullet, \psi_\bullet)$. If $\psi_\bullet = \phi^\star_\bullet$ is the discrete solution to (6.16), then there holds the reliability estimate (i.e., the global upper bound)

$$|||\phi^\star - \phi^\star_\bullet||| \le C_{\mathrm{rel}}\, \eta_\bullet(\phi^\star_\bullet), \tag{6.24}$$

where $C_{\mathrm{rel}} > 0$ depends only on Γ and γ-shape regularity of $\mathcal{T}_\bullet$, cf. [CS95, Car97] for $d = 2$ and [CMS01] for $d = 3$ respectively. Provided that $\phi^\star \in L^2(\Gamma)$, the following weak efficiency

$$|||\phi^\star - \phi^\star_\bullet||| + \eta_\bullet(\phi^\star_\bullet) \le C_{\mathrm{eff}}\, \|h_\bullet^{1/2}\, (\phi^\star - \phi^\star_\bullet)\|_{L^2(\Gamma)} \tag{6.25}$$

has recently been proved in [AFF+17], where $C_{\mathrm{eff}} > 0$ depends only on Γ and γ-shape regulartiy of $\mathcal{T}_\bullet$. We note that the weighted L^2-norm on the right-hand side of (6.25) is only slightly stronger than $|||\cdot||| \simeq \|\cdot\|_{\widetilde{H}^{-1/2}(\Gamma)}$, so that one empirically observes $\eta_\bullet(\phi^\star_\bullet) \lesssim |||\phi - \phi^\star_\bullet|||$ in practice, cf. [CS95, Car97, CMS01]. In certain situations (e.g., weakly-singular integral formulation of the interior 2D Dirichlet problem), one can rigorously prove the latter (strong) efficiency estimate up to higher-order data oscillations, cf [AFF+13].

6.3.3 Preconditioned conjugate gradient method (PCG) for the Galerkin system

Suppose that $\mathbf{P}_\bullet, \mathbf{A}_\bullet \in \mathbb{R}^{N\times N}$ are symmetric and positive definite matrices. Given $\mathbf{b}_\bullet \in \mathbb{R}^N$ and an initial guess $\mathbf{x}_\bullet^0$, PCG (see [GVL13, Algorithm 11.5.1]) aims to approximate the solution $\mathbf{x}_\bullet^\star \in \mathbb{R}^N$ to (6.18). We note that each step of PCG has the following computational costs:

- $\mathcal{O}(N)$ cost for vector operations (e.g., assignment, addition, scalar product),
- computation of *one* matrix-vector product with $\mathbf{A}_\bullet$,
- computation of *one* matrix-vector product with $\mathbf{P}_\bullet^{-1}$.

Let $\widetilde{\mathbf{x}}_\bullet^\star \in \mathbb{R}^N$ be the solution to (6.1) and recall that $\mathbf{x}_\bullet^\star = \mathbf{P}_\bullet^{-1/2}\widetilde{\mathbf{x}}_\bullet^\star$. We note that PCG formally applies the conjugate gradient method (CG, see [GVL13, Algorithm 11.3.2]) for the matrix $\widetilde{\mathbf{A}}_\bullet := \mathbf{P}_\bullet^{-1/2}\mathbf{A}_\bullet\mathbf{P}_\bullet^{-1/2}$ and the right-hand side $\widetilde{\mathbf{b}}_\bullet = \mathbf{P}_\bullet^{-1/2}\mathbf{b}_\bullet$. The iterates $\mathbf{x}_\bullet^k \in \mathbb{R}^N$ of PCG (applied to $\mathbf{P}_\bullet$, $\mathbf{A}_\bullet$, $\mathbf{b}_\bullet$, and the initial guess $\mathbf{x}_\bullet^0$) and the iterates $\widetilde{\mathbf{x}}_\bullet^k$ of CG (applied to $\widetilde{\mathbf{A}}_\bullet$, $\widetilde{\mathbf{b}}_\bullet$, and the initial guess $\widetilde{\mathbf{x}}_\bullet^0 := \mathbf{P}_\bullet^{1/2}\mathbf{x}_\bullet^0$) are formally linked by

$$\mathbf{x}_\bullet^k = \mathbf{P}_\bullet^{-1/2}\widetilde{\mathbf{x}}_\bullet^k;$$

see [GVL13, Section 11.5]. Moreover, for all $\widetilde{\mathbf{y}}_\bullet \in \mathbb{R}^N$ and $\mathbf{y}_\bullet = \mathbf{P}_\bullet^{-1/2}\widetilde{\mathbf{y}}_\bullet$, there holds that

$$\begin{aligned}\|\widetilde{\mathbf{y}}_\bullet\|_{\widetilde{\mathbf{A}}_\bullet}^2 &:= \widetilde{\mathbf{y}}_\bullet \cdot \widetilde{\mathbf{A}}_\bullet\widetilde{\mathbf{y}}_\bullet \\ &= (\mathbf{P}_\bullet^{1/2}\mathbf{y}_\bullet)\cdot\mathbf{P}_\bullet^{-1/2}\mathbf{A}_\bullet\mathbf{P}_\bullet^{-1/2}\mathbf{P}_\bullet^{1/2}\mathbf{y}_\bullet \\ &= \mathbf{y}_\bullet\cdot\mathbf{A}_\bullet\mathbf{y}_\bullet \\ &=: \|\mathbf{y}_\bullet\|_{\mathbf{A}_\bullet}^2.\end{aligned} \tag{6.26}$$

Consequently, [GVL13, Theorem 11.3.3] for CG (applied to $\widetilde{\mathbf{A}}_\bullet$, $\widetilde{\mathbf{b}}_\bullet$, $\widetilde{\mathbf{x}}_\bullet^0$) yields the following lemma for PCG (which follows from the implicit steepest decent approach of CG).

Lemma 56. *Let $\mathbf{A}_\bullet, \mathbf{P}_\bullet \in \mathbb{R}^{N\times N}$ be symmetric and positive definite, $\mathbf{b}_\bullet \in \mathbb{R}^N$, $\mathbf{x}_\bullet^\star := \mathbf{A}_\bullet^{-1}\mathbf{b}_\bullet$, and $\mathbf{x}_\bullet^0 \in \mathbb{R}^N$. Suppose the ℓ_2-condition number estimate*

$$\operatorname{cond}_2(\mathbf{P}_\bullet^{-1/2}\mathbf{A}_\bullet\mathbf{P}_\bullet^{-1/2}) \le C_{\mathrm{alg}}. \tag{6.27}$$

Then, the iterates $\mathbf{x}_\bullet^k$ of the PCG algorithm satisfy the contraction property

$$\|\mathbf{x}_\bullet^\star - \mathbf{x}_\bullet^{k+1}\|_{\mathbf{A}_\bullet} \le q_{\mathrm{pcg}}\,\|\mathbf{x}_\bullet^\star - \mathbf{x}_\bullet^k\|_{\mathbf{A}_\bullet} \quad \text{for all } k \in \mathbb{N}_0, \tag{6.28}$$

where $q_{\mathrm{pcg}} := (1 - 1/C_{\mathrm{alg}})^{1/2} < 1$. □

If the matrix $\mathbf{A}_\bullet \in \mathbb{R}^{N\times N}$ stems from the Galerkin discretization (6.18) for $\mathcal{T}_\bullet = \{T_1, \dots, T_N\}$, there is a one-to-one correspondence of vectors $\mathbf{y}_\bullet \in \mathbb{R}^N$ and discrete functions $\psi_\bullet \in \mathcal{P}^0(\mathcal{T}_\bullet)$ via

$$\psi_\bullet = \sum_{j=1}^N \mathbf{y}_\bullet[j]\,\chi_{\bullet,j},$$

where $\chi_{\bullet,j}$ is the usual characteristic function of $T_j \in \mathcal{T}_\bullet$. Let $\phi_\bullet^k \in \mathcal{P}^0(\mathcal{T}_\bullet)$ denote the discrete function corresponding to the PCG iterate $\mathbf{x}_\bullet^k \in \mathbb{R}^N$, while the Galerkin solution $\phi_\bullet^\star \in \mathcal{P}^0(\mathcal{T}_\bullet)$ of (6.16) corresponds to $\mathbf{x}_\bullet^\star = \mathbf{A}_\bullet^{-1}\mathbf{b}_\bullet$. We note the elementary identity

$$|||\phi_\bullet^\star - \phi_\bullet^k|||^2 = (\mathbf{x}_\bullet^\star - \mathbf{x}_\bullet^k) \cdot \mathbf{A}_\bullet(\mathbf{x}_\bullet^\star - \mathbf{x}_\bullet^k) = \|\mathbf{x}_\bullet^\star - \mathbf{x}_\bullet^k\|_{\mathbf{A}_\bullet}^2. \tag{6.29}$$

6.3.4 Optimal preconditioners

We say that $\mathbf{P}_\bullet$ is an *optimal* preconditioner, if $C_{\mathrm{alg}} \geq 1$ in the ℓ_2-condition number estimate (6.27) depends only on γ-shape regularity of $\mathcal{T}_\bullet$ and the initial mesh $\mathcal{T}_0$ (and is hence essentially independent of the mesh $\mathcal{T}_\bullet$).

6.4 Adaptive algorithm

Next, we introduce the following adaptive algorithm which is driven by the weighted-residual error estimator (6.22). We note that Algorithm 57 as well as the following results are independent of the precise preconditioning strategy as long as the employed preconditioners are optimal, cf. Section 6.3.4.

Algorithm 57. ***Input:*** *Initial conforming mesh $\mathcal{T}_0$ of Γ, initial guess $\phi_0^0 := 0$, adaptivity parameters $0 < \theta \leq 1$, $\lambda_{\mathrm{ctr}} > 0$, and $C_{\mathrm{mark}} > 0$, optimal preconditioning strategy $\mathbf{P}_\ell$ for all $\mathcal{T}_\ell \in \mathtt{refine}(\mathcal{T}_0)$, counters $\ell := 0 =: k$.*
Adaptive Loop: *Iterate the following Steps* (i)–(v)*:*

(i) **Repeat** *the following steps* (a)–(c)*:*

(a) *Update the counter $(\ell, k) \mapsto (\ell, k+1)$.*

(b) *Do one step of the PCG algorithm with the optimal preconditioner $\mathbf{P}_\ell$ to obtain $\phi_\ell^k \in \mathcal{P}^0(\mathcal{T}_\ell)$ from $\phi_\ell^{k-1} \in \mathcal{P}^0(\mathcal{T}_\ell)$.*

(c) *Compute the local contributions $\eta_\ell(T, \phi_\ell^k)$ of the error estimator for all $T \in \mathcal{T}_\ell$.*

Until $|||\phi_\ell^k - \phi_\ell^{k-1}||| \leq \lambda_{\mathrm{ctr}}\, \eta_\ell(\phi_\ell^k)$. (6.30)

(ii) *Define $\underline{k} := \underline{k}(\ell) := k$.*

(iii) *Determine a set $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ with up to the multiplicative constant C_{mark} minimal cardinality such that*

$$\theta\, \eta_\ell(\phi_\ell^{\underline{k}}) \leq \eta_\ell(\mathcal{M}_\ell, \phi_\ell^{\underline{k}}). \tag{6.31}$$

(iv) *Generate $\mathcal{T}_{\ell+1} := \mathtt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ and define $\phi_{\ell+1}^0 := \phi_\ell^{\underline{k}}$.*

(v) *Update the counter $(\ell, k) \mapsto (\ell+1, 0)$ and continue with* (i)*.*

Output: *Sequences of successively refined triangulations $\mathcal{T}_\ell$, discrete solutions ϕ_ℓ^k, and corresponding error estimators $\eta_\ell(\phi_\ell^k)$, for all $\ell \geq 0$ and $k \geq 0$.*

Remark 58. *The choice $\lambda_{\mathrm{ctr}} = 0$ corresponds to the case that the Galerkin system (6.16) is solved exactly, i.e., $\phi_{\ell+1}^0 = \phi_\ell^\star$. Then, optimal convergence of Algorithm 57 has already been proved in [FKMP13, Gan13, AFF+13, FFK+14] for weakly-singular integral equations and [Gan13, FFK+15] for hyper-singular integral equations. The choice $\theta = 1$ will generically lead to uniform mesh-refinement, where for each mesh all elements $\mathcal{M}_\ell = \mathcal{T}_\ell$ are refined in Step* (iv) *of Algorithm 57. Instead, small $0 < \theta \ll 1$, will lead to highly adapted meshes.*

Let $\mathcal{Q} := \big\{(\ell, k) \in \mathbb{N}_0 \times \mathbb{N}_0 : \text{index } (\ell, k) \text{ is used in Algorithm 57}\big\}$ be the set of all index pairs which appear at some point in Algorithm 57. It holds that $(0, 0) \in \mathcal{Q}$. Moreover, for $\ell, k \in \mathbb{N}_0$, it holds that

- for $\ell \geq 1$, $(\ell, 0) \in \mathcal{Q}$ implies that $(\ell - 1, 0) \in \mathcal{Q}$,
- for $k \geq 1$, $(\ell, k) \in \mathcal{Q}$ implies that $(\ell, k - 1) \in \mathcal{Q}$.

If ℓ is clear from the context, we abbreviate $\underline{k} := \underline{k}(\ell)$, e.g., $\phi_\ell^{\underline{k}} := \phi_\ell^{\underline{k}(\ell)}$. In particular, it holds that $\phi_\ell^{\underline{k}} = \phi_{\ell+1}^0$. Since PCG (like any Krylov method) provides the exact solution after at most $\#\mathcal{T}_\ell$ steps, it follows that $1 \leq \underline{k}(\ell) < \infty$. Finally, we define the ordering

$$(\ell', k') < (\ell, k) \quad \overset{\text{def}}{\iff} \quad \left\{ \begin{array}{ll} \text{either:} & \ell' < \ell \\ \text{or:} & \ell' = \ell \text{ and } k' < k \end{array} \right\} \qquad \text{for all } (\ell', k'), (\ell, k) \in \mathcal{Q}.$$

Moreover, let

$$|(\ell, k)| := \begin{cases} 0, & \text{if } \ell = 0 = k, \\ \#\big\{(\ell', k') \in \mathcal{Q} : (\ell', k') < (\ell, k) \text{ and } k' < \underline{k}(\ell')\big\}, & \text{if } \ell > 0 \text{ or } k > 0, \end{cases} \tag{6.32}$$

be the total number of PCG iterations until the computation of ϕ_ℓ^k. Note that $\ell' > \ell$ and $|(\ell', k')| = |(\ell, k)|$ imply that $\ell' = \ell + 1$, $k = \underline{k}(\ell)$, and $k' = 0$ and hence $\phi_{\ell'}^{k'} = \phi_\ell^k$.

6.5 Main results

In this section, we show the main results of this chapter, i.e., first, we introduce an additive Schwarz preconditioner and prove its optimality in the sense of Section 6.3.4, and secondly, we prove optimal convergence rates with respect to the degrees of freedom of Algorithm 57 as well as almost optimal computational complexity.

6.5.1 Optimal additive Schwarz preconditioner

We consider multilevel additive Schwarz preconditioners that build on the adaptive mesh-hierarchy.

Let $\mathcal{E}_\bullet$ denote the set of all nodes ($d = 2$) and edges ($d = 3$) respectively of the mesh $\mathcal{T}_\bullet$ which do not belong to the relative boundary $\partial\Gamma$. Only for $\Gamma = \partial\Omega$, $\mathcal{E}_\bullet$ contains all

nodes resp. edges of $\mathcal{T}_\bullet$. For $E \in \mathcal{E}_\bullet$, let $T^+, T^- \in \mathcal{T}_\bullet$ denote the two unique elements with $T^+ \cap T^- = E$. We define the Haar-type function $\varphi_{\bullet,E} \in \mathcal{P}^0(\mathcal{T}_\bullet)$ (associated to $E \in \mathcal{E}_\bullet$) by

$$\varphi_{\bullet,E}|_T := \begin{cases} \frac{|E|}{|T|} & \text{for } T = T^+, \\ -\frac{|E|}{|T|} & \text{for } T = T^-, \\ 0 & \text{else,} \end{cases} \tag{6.33}$$

where $|E| := 1$ for $d = 2$ and $|E| := \operatorname{diam}(E)$ for $d = 3$. Note that

$$\varphi_{\bullet,E} \in \mathcal{P}^0_*(\mathcal{T}_\bullet) := \big\{\psi \in \mathcal{P}^0(\mathcal{T}_\bullet) \,:\, \int_\Gamma \psi \,\mathrm{d}x = 0\big\}. \tag{6.34}$$

For $d = 3$, we additionally suppose that the orientation of each edge E is arbitrary but fixed. We choose $T^+ \in \mathcal{T}_\bullet$ such that ∂T^+ and $E \subset \partial T^+$ have the same orientation.

Given a mesh $\mathcal{T}_0$, suppose that $\mathcal{T}_\ell$ is a sequence of locally refined meshes, i.e., for all $\ell \in \mathbb{N}_0$, there exists a set $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell$ such that $\mathcal{T}_{\ell+1} = \texttt{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$. Then, define

$$\mathcal{E}_\ell^\star := \mathcal{E}_\ell \backslash \mathcal{E}_{\ell-1} \cup \big\{E \in \mathcal{E}_\ell \,:\, \operatorname{supp}(\varphi_{\ell,E}) \subsetneqq \operatorname{supp}(\varphi_{\ell-1,E})\big\} \quad \text{for all } \ell \geq 1,$$

which consist of new (interior) nodes/edges plus some of their neighbours. We note the following subspace decomposition which is, in general, *not* direct.

Lemma 59. *With $\mathcal{X}_\bullet := \mathcal{P}^0(\mathcal{T}_\bullet)$ and $\mathcal{X}_{\bullet,E} := \operatorname{span}\{\varphi_{\bullet,E}\}$, it holds that*

$$\mathcal{X}_L = \mathcal{X}_0 + \sum_{\ell=1}^{L} \sum_{E \in \mathcal{E}_\ell^\star} \mathcal{X}_{\ell,E} \quad \textit{for all } L \in \mathbb{N}_0. \tag{6.35}$$

□

Additive Schwarz preconditioners are based on (not necessarily direct) subspace decompositions. Following the standard theory (see, e.g., [TW05, Chapter 2]), (6.35) yields a (local multilevel) preconditioner. To provide its matrix formulation, let $\mathbf{I}_{k,\ell} \in \mathbb{R}^{(\#\mathcal{T}_\ell)\times(\#\mathcal{T}_k)}$ be the matrix representation of the canonical embedding $\mathcal{P}^0(\mathcal{T}_k) \hookrightarrow \mathcal{P}^0(\mathcal{T}_\ell)$ for $k < \ell$, i.e.,

$$\sum_{i=1}^{\#\mathcal{T}_k} \mathbf{x}_k[i]\, \chi_{k,i} = \sum_{i=1}^{\#\mathcal{T}_\ell} \mathbf{x}_\ell[i]\, \chi_{\ell,i} \quad \text{for all } \mathbf{x}_k \in \mathbb{R}^{\#\mathcal{T}_k} \text{ and } \mathbf{x}_\ell := \mathbf{I}_{k,\ell}\mathbf{x}_k \in \mathbb{R}^{\#\mathcal{T}_\ell}.$$

Let $\mathbf{H}_\ell \in \mathbb{R}^{(\#\mathcal{T}_\ell)\times(\#\mathcal{E}_\ell)}$ denote the matrix that represents Haar-type functions, i.e.,

$$\varphi_{\ell,E_j} = \sum_{i=1}^{\#\mathcal{T}_\ell} \mathbf{H}_\ell[i,j]\chi_{\ell,i} \quad \text{for all } E_j \in \mathcal{E}_\ell.$$

Since only two coefficients per column are non-zero, $\mathbf{H}_\ell$ is sparse, while $\mathbf{I}_{k,\ell}$ is non-sparse in general. Finally, define the (non-invertible) diagonal matrix $\mathbf{D}_\ell \in \mathbb{R}^{(\#\mathcal{E}_\ell)\times(\#\mathcal{E}_\ell)}$ by

$$(\mathbf{D}_\ell)_{jk} := \begin{cases} \|\!|\varphi_{\ell,E_j}|\!\|^{-2} & E_j \in \mathcal{E}_\ell^\star \text{ and } j = k, \\ 0 & \text{else.} \end{cases}$$

Then, the matrix representation of the preconditioner associated to (6.35) reads

$$\mathbf{P}_L^{-1} := \mathbf{I}_{0,L}\mathbf{A}_0^{-1}\mathbf{I}_{0,L}^T + \sum_{\ell=1}^{L} \mathbf{I}_{\ell,L}\mathbf{H}_\ell\mathbf{D}_\ell\mathbf{H}_\ell^T\mathbf{I}_{\ell,L}^T. \tag{6.36}$$

For $d = 2$, the subsequent Theorem 60 is already proved in [FFPS17b, Section III.B] for $\Gamma = \partial\Omega$ and in [Füh14, Section 6.3] for $\Gamma \subsetneqq \partial\Omega$. For $d = 3$, we need the following additional assumptions:

- First, suppose that $\Omega \subset \mathbb{R}^3$ is simply connected and $\Gamma = \partial\Omega$.
- Second, let $\widehat{\mathcal{T}}_0$ be a conforming triangulation of Ω into non-degenerate compact simplices such that $\mathcal{T}_0 = \widehat{\mathcal{T}}_0|_\Gamma$ is the induced boundary partition on Γ.

Then, the following theorem is our first main result.

Theorem 60. *Under the foregoing assumptions, the preconditioner $\mathbf{P}_L$ from (6.36) is optimal, i.e., there holds (6.27), where $C_{\mathrm{alg}} \geq 1$ depends only on Ω and $\widehat{\mathcal{T}}_0$, but is independent of $L \in \mathbb{N}$.*

We stress that the matrix in (6.36) will never be assembled in practice. The PCG algorithm only needs the action of $\mathbf{P}_L^{-1}$ on a vector. This can be done recursively by using the embeddings $\mathbf{I}_{\ell,\ell+1}$ which are, in fact, sparse. Up to (storing and) inverting $\mathbf{A}_0$ on the coarse mesh, the evaluation of $\mathbf{P}_L^{-1}\mathbf{x}$ can be done in $\mathcal{O}(\#\mathcal{T}_L)$ operations, see, e.g., [FFPS17a, Section 3.1] for a detailed discussion. If the mesh $\mathcal{T}_L$ is fine compared to the initial mesh $\mathcal{T}_0$ (or if $\mathbf{A}_0$ is realized with, e.g., $\mathcal{H}$-matrix techniques), then the computational costs and storage requirements associated with $\mathbf{A}_0$ can be neglected.

Remark 61. *Our proof for $d = 3$ requires additional assumptions on Ω, $\Gamma = \partial\Omega$, and $\mathcal{T}_0$. As stated above, the case $d = 2$ allows for a different proof (which, however, does not transfer to $d = 3$) and can thus avoid these assumptionss, see [FFPS17b, Füh14].*

6.5.2 Proof of Theorem 60 (optimality of additive Schwarz preconditioner)

As mentioned before, Theorem 60 is already proved for $d = 2$. Hence, we refer to [FFPS17b, Füh14] and thus focus only on $d = 3$ and $\Gamma = \partial\Omega$. Due to our additional assumption, $\mathcal{T}_0 = \widehat{\mathcal{T}}_0|_\Gamma$ is the restriction of a conforming simplicial triangulation $\widehat{\mathcal{T}}_0$ of Ω to the boundary Γ. Moreover, 2D NVB refinement of $\mathcal{T}_0$ (on the boundary Γ) is a special case of 3D NVB refinement of $\widehat{\mathcal{T}}_0$ (in the volume Ω) plus restriction to the boundary, see, e.g., [Ste08]. Hence, each mesh $\mathcal{T}_\bullet \in \mathbb{T} = \mathtt{refine}(\mathcal{T}_0)$ is the restriction of a conforming NVB refinement $\widehat{\mathcal{T}}_\bullet \in \widehat{\mathbb{T}} := \mathtt{refine}(\widehat{\mathcal{T}}_0)$, i.e., $\mathcal{T}_\bullet = \widehat{\mathcal{T}}_\bullet|_\Gamma$. Throughout, let $\widehat{\mathcal{T}}_\bullet \in \widehat{\mathbb{T}}$ be the coarsest extension of $\mathcal{T}_\bullet \in \mathbb{T}$.

Recall that NVB is a binary refinement rule. Therefore, $\mathcal{T}_\circ \in \mathtt{refine}(\mathcal{T}_\bullet)$ also implies that $\widehat{\mathcal{T}}_\circ \in \mathtt{refine}(\widehat{\mathcal{T}}_\bullet)$. Finally, we note that all triangulations $\widehat{\mathcal{T}}_\bullet \in \widehat{\mathbb{T}}$ are uniformly γ-shape regular, i.e.,

$$\max_{\widehat{T}\in\widehat{\mathcal{T}}_\bullet} \frac{\operatorname{diam}(\widehat{T})}{|\widehat{T}|^{1/3}} \leq \gamma < \infty.$$

where γ depends only on $\widehat{\mathcal{T}}_0$.

Discrete spaces and extensions

First, we recall the definition of the curl operator for a sufficiently smooth vector field $\boldsymbol{v} = (v_1, v_2, v_3)$ by

$$\operatorname{curl} \boldsymbol{v} := \nabla \times \boldsymbol{v} := \begin{pmatrix} \frac{\partial v_3}{\partial x_2} - \frac{\partial v_2}{\partial x_3} \\ \frac{\partial v_1}{\partial x_3} - \frac{\partial v_3}{\partial x_1} \\ \frac{\partial v_2}{\partial x_1} - \frac{\partial v_1}{\partial x_2} \end{pmatrix}.$$

Definition 62. *Let $\boldsymbol{v} \in L^2(\Omega)^3$. Then, we call $\operatorname{curl} \boldsymbol{v} := \boldsymbol{c} \in L^2(\Omega)^3$ the (generalized) curl of $\boldsymbol{v}$, if there holds that*

$$\int_\Omega \boldsymbol{c} \cdot \boldsymbol{w} \,\mathrm{d}x = \int_\Omega \boldsymbol{v} \cdot \operatorname{curl} \boldsymbol{w} \,\mathrm{d}x \quad \text{for all } \boldsymbol{w} \in C_0^\infty(\overline{\Omega})^3, \tag{6.37}$$

as well as $\operatorname{div} \boldsymbol{v} := d \in L^2(\Omega)$ the (generalized) divergence of $\boldsymbol{v}$, if there holds that

$$\int_\Omega d\, w \,\mathrm{d}x = -\int_\Omega \boldsymbol{v} \cdot \nabla w \,\mathrm{d}x \quad \text{for all } w \in C_0^\infty(\overline{\Omega}). \tag{6.38}$$

Moreover, we define the space of lowest-order Nédélec elements of first kind $\mathcal{ND}^1(\widehat{\mathcal{T}}_\bullet)$ by

$$\mathcal{ND}^1(\widehat{\mathcal{T}}_\bullet) := \big\{\boldsymbol{v} \in \boldsymbol{H}(\operatorname{curl}; \Omega) \,:\, \boldsymbol{v}|_K \in \mathcal{P}^0(K)^3 + \mathcal{P}^0(K)^3 \times \mathbf{x} \text{ for all } K \in \widehat{\mathcal{T}}_\bullet\big\}, \tag{6.39}$$

where

$$\boldsymbol{H}(\operatorname{curl}; \Omega) := \big\{\boldsymbol{v} \in L^2(\Omega)^3 \,:\, \operatorname{curl} \boldsymbol{v} \in L^2(\Omega)^3\big\} \tag{6.40}$$

is the space of square integrable vector fields on $\Omega \subset \mathbb{R}^3$ with square integrable curl *and corresponding norm*

$$\|\boldsymbol{v}\|^2_{\boldsymbol{H}(\operatorname{curl};\Omega)} := \|\boldsymbol{v}\|^2_{L^2(\Omega)} + \|\operatorname{curl} \boldsymbol{v}\|^2_{L^2(\Omega)}. \tag{6.41}$$

Lastly, we define the space of lowest-order Raviart–Thomas elements $\mathcal{RT}^0(\widehat{\mathcal{T}}_\bullet)$ by

$$\mathcal{RT}^0(\widehat{\mathcal{T}}_\bullet) := \big\{\boldsymbol{v} \in \boldsymbol{H}(\operatorname{div}; \Omega) \,:\, \boldsymbol{v}|_{\mathrm{K}} \in \mathcal{P}^0(\mathrm{K})^3 + \mathcal{P}^0(\mathrm{K})\,\mathbf{x} \text{ for all } \mathrm{K} \in \widehat{\mathcal{T}}_\bullet\big\}, \tag{6.42}$$

where

$$\boldsymbol{H}(\operatorname{div}; \Omega) := \big\{\boldsymbol{v} \in \mathrm{L}^2(\Omega)^3 \,:\, \operatorname{div} \boldsymbol{v} \in \mathrm{L}^2(\Omega)\big\} \tag{6.43}$$

is the space of square integrable vector fields on $\Omega \subset \mathbb{R}^3$ with square integrable div *and corresponding norm*

$$\|\boldsymbol{v}\|^2_{\boldsymbol{H}(\operatorname{div};\Omega)} := \|\boldsymbol{v}\|^2_{L^2(\Omega)} + \|\operatorname{div} \boldsymbol{v}\|^2_{L^2(\Omega)}. \tag{6.44}$$

The argument of our proof of Theorem 60 adapts ideas from [HM12], where a subspace decomposition for the lowest-order Nédélec space $\boldsymbol{\mathcal{ND}}^1(\widehat{\mathcal{T}}_\bullet)$ (see, e.g., [HZ09]) in $\boldsymbol{H}(\mathrm{curl}\,;\Omega)$ implies a decomposition of the corresponding discrete trace space. While the original idea dates back to [Osw99], a nice summary of the argument is found in [HM12, Section 2].

Remark 63. (i) *Our proof is based on the construction of an extension operator from $\mathcal{P}^0_*(\mathcal{T}_\bullet)$ to $\boldsymbol{\mathcal{ND}}^1(\widehat{\mathcal{T}}_\bullet)$, see Lemma 65 below. It is not clear if such an operator can be constructed for the case $\Gamma \subsetneqq \partial\Omega$.*

(ii) *In [HJHM15], a subspace decomposition of the lowest-order Raviart–Thomas space $\boldsymbol{\mathcal{RT}}^0(\widehat{\mathcal{T}}_\bullet)$ (see, e.g., [XCN09]) in $\boldsymbol{H}(\mathrm{div}\,;\Omega)$ implies a decomposition of the corresponding normal trace space $\mathcal{P}^0(\mathcal{T}_\bullet)$. Due to different scaling properties of the Raviart–Thomas basis functions (in the $\boldsymbol{H}(\mathrm{div}\,;\Omega)$ norm) and their normal trace (in the $H^{-1/2}(\Gamma)$ norm), this argument does not apply in our case.*

Let $\widehat{\mathcal{E}}_\bullet$ (resp. $\widehat{\mathcal{N}}_\bullet$) denote the set of all edges (resp. all nodes) of $\widehat{\mathcal{T}}_\bullet \in \widehat{\mathbb{T}}$. For each node $\mathbf{x} \in \widehat{\mathcal{N}}_\bullet$, let $\eta_{\bullet,\mathbf{x}} \in \mathcal{S}^1(\widehat{\mathcal{T}}_\bullet)$ be the corresponding hat function, i.e., $\eta_{\bullet,\mathbf{x}}$ is $\widehat{\mathcal{T}}_\bullet$-piecewise affine and globally continuous with $\eta_{\bullet,\mathbf{x}}(\mathbf{y}) = \delta_{\mathbf{xy}}$ for all $\mathbf{x},\mathbf{y} \in \widehat{\mathcal{N}}_\bullet$. For $E \in \widehat{\mathcal{E}}_\bullet$, let $\boldsymbol{u}_{\bullet,E} \in \boldsymbol{\mathcal{ND}}^1(\widehat{\mathcal{T}}_\bullet)$ denote the corresponding Nédélec basis function, i.e., for $K \in \widehat{\mathcal{T}}_\bullet$ with $E = \mathrm{conv}\{\mathbf{x},\mathbf{y}\} \subset \partial K$, it holds that

$$\boldsymbol{u}_{\bullet,E}|_K = C(\eta_{\bullet,\mathbf{x}}\nabla\eta_{\bullet,\mathbf{y}} - \eta_{\bullet,\mathbf{y}}\nabla\eta_{\bullet,\mathbf{x}}), \tag{6.45}$$

where $C > 0$ is chosen such that for the path integrals holds that

$$\int_{E'} \boldsymbol{u}_{\bullet,E}\,ds = |E|\,\delta_{EE'} \quad \text{for all } E, E' \in \widehat{\mathcal{E}}_\bullet. \tag{6.46}$$

Scaling arguments yield the next lemma. The proof follows the lines of [HM12, Lemma 5.7].

Lemma 64. *For $E \in \mathcal{E}_\bullet$, recall the Haar function $\varphi_{\bullet,E} \in \mathcal{P}^0(\mathcal{T}_\bullet)$ from (6.33). Let $\boldsymbol{u}_{\bullet,E} \in \boldsymbol{\mathcal{ND}}^1(\widehat{\mathcal{T}}_\bullet)$ denote the corresponding Nédélec basis function, see (6.45). Then,*

$$\varphi_{\bullet,E} = \mathrm{curl}\,\boldsymbol{u}_{\bullet,E} \cdot \mathbf{n}|_\Gamma \tag{6.47}$$

and

$$C^{-1}\|\varphi_{\bullet,E}\|_{H^{-1/2}(\Gamma)} \le \|\boldsymbol{u}_{\bullet,E}\|_{\boldsymbol{H}(\mathrm{curl}\,;\Omega)} \le C\,\|\varphi_{\bullet,E}\|_{H^{-1/2}(\Gamma)}, \tag{6.48}$$

where $C > 0$ depends only on Ω and the γ-shape regularity of $\widehat{\mathcal{T}}_\bullet$. □

Proof. By using (6.45)–(6.46) we get that $\varphi_{\bullet,E} = \mathrm{curl}\,\boldsymbol{u}_{\bullet,E} \cdot \mathbf{n}|_\Gamma$. Then, continuity of the normal trace operator and the fact that the divergence of the curl is zero yield that

$$\begin{aligned}
\|\varphi_{\bullet,E}\|_{H^{-1/2}(\Gamma)} &\lesssim \|\mathrm{curl}\,\boldsymbol{u}_{\bullet,E}\|_{\boldsymbol{H}(\mathrm{div}\,;\Omega)} \\
&= \big(\|\mathrm{curl}\,\boldsymbol{u}_{\bullet,E}\|^2_{L^2(\Omega)} + \|\mathrm{div}\,\mathrm{curl}\,\boldsymbol{u}_{\bullet,E}\|^2_{L^2(\Omega)}\big)^{1/2} \\
&= \|\mathrm{curl}\,\boldsymbol{u}_{\bullet,E}\|_{L^2(\Omega)} \\
&\le \|\boldsymbol{u}_{\bullet,E}\|_{\boldsymbol{H}(\mathrm{curl}\,;\Omega)}.
\end{aligned}$$

Furthermore, scaling arguments prove that

$$\begin{aligned}\|\boldsymbol{u}_{\bullet,E}\|_{\boldsymbol{H}(\operatorname{curl};\Omega)} &\simeq \|\operatorname{curl}\boldsymbol{u}_{\bullet,E}\|_{L^2(\Omega)}\\ &\simeq |E|^{1/2}\\ &\simeq |E|^{1/2}\|\varphi_{\bullet,E}\|_{L^2(\Gamma)}\\ &\lesssim \|\varphi_{\bullet,E}\|_{H^{-1/2}(\Gamma)},\end{aligned}$$

where we have finally applied an inverse estimate, cf. [HM12, Lemma 5.4]. This concludes the proof. □

The following lemma holds for (simply) connected Lipschitz domains Ω and follows essentially from [AGS16]. Recall $\mathcal{P}^0_*(\mathcal{T}_\bullet)$ from (6.34).

Lemma 65. *There exists a linear operator $\mathbf{E}_\bullet : \mathcal{P}^0_*(\mathcal{T}_\bullet) \to \boldsymbol{\mathcal{ND}}^1(\widehat{\mathcal{T}}_\bullet)$ such that*

$$\operatorname{curl}(\mathbf{E}_\bullet\psi_\bullet)\cdot\mathbf{n}|_\Gamma = \psi_\bullet \tag{6.49}$$

as well as

$$\|\mathbf{E}_\bullet\psi_\bullet\|_{\boldsymbol{H}(\operatorname{curl};\Omega)} \le C\,\|\psi_\bullet\|_{H^{-1/2}(\Gamma)} \quad \text{for all } \psi_\bullet \in \mathcal{P}^0_*(\mathcal{T}_\bullet). \tag{6.50}$$

The constant $C > 0$ depends only on Ω and γ-shape regularity of $\widehat{\mathcal{T}}_\bullet$.

Proof. Let $\psi_\bullet \in \mathcal{P}^0_*(\mathcal{T}_\bullet)$. First, [AGS16, Theorem 2.1] provides $\boldsymbol{\sigma}_\bullet \in \boldsymbol{\mathcal{RT}}^0(\widehat{\mathcal{T}}_\bullet)$ with

$$\boldsymbol{\sigma}_\bullet\cdot\mathbf{n}|_\Gamma = \psi_\bullet, \quad \operatorname{div}\boldsymbol{\sigma}_\bullet = 0, \quad \text{and} \quad \|\boldsymbol{\sigma}_\bullet\|_{\boldsymbol{H}(\operatorname{div};\Omega)} \lesssim \|\psi_\bullet\|_{\mathrm{H}^{-1/2}(\partial\Omega)}.$$

Then, [AGS16, Lemma 4.3] provides $\mathbf{E}_\bullet\psi_\bullet := \boldsymbol{v}_\bullet \in \boldsymbol{\mathcal{ND}}^1(\widehat{\mathcal{T}}_\bullet)$ such that

$$\operatorname{curl}\boldsymbol{v}_\bullet = \boldsymbol{\sigma}_\bullet \quad \text{and} \quad \|\boldsymbol{v}_\bullet\|_{\boldsymbol{H}(\operatorname{curl};\Omega)} \lesssim \|\boldsymbol{\sigma}_\bullet\|_{\boldsymbol{H}(\operatorname{div};\Omega)}.$$

Combining these results, we get that

$$\begin{aligned}\operatorname{curl}(\mathbf{E}_\bullet\psi_\bullet)\cdot\mathbf{n}|_\Gamma &= \operatorname{curl}\boldsymbol{v}_\bullet\cdot\mathbf{n}|_\Gamma\\ &= \boldsymbol{\sigma}_\bullet\cdot\mathbf{n}|_\Gamma\\ &= \psi_\bullet,\end{aligned}$$

as well as

$$\begin{aligned}\|\mathbf{E}_\bullet\psi_\bullet\|_{\boldsymbol{H}(\operatorname{curl};\Omega)} &= \|\boldsymbol{v}_\bullet\|_{\boldsymbol{H}(\operatorname{curl};\Omega)}\\ &\lesssim \|\boldsymbol{\sigma}_\bullet\|_{\boldsymbol{H}(\operatorname{div};\Omega)}\\ &\lesssim \|\psi_\bullet\|_{H^{-1/2}(\Gamma)},\end{aligned}$$

which concludes the proof. □

Abstract additive Schwarz preconditioners

Let $\mathcal{X}$ denote some finite dimensional Hilbert space with norm $\|\cdot\|_{\mathcal{X}}$ and subspace decomposition

$$\mathcal{X} = \sum_{i\in\mathcal{I}} \mathcal{X}_i,$$

where $\mathcal{I}$ is a finite index set. The associated additive Schwarz operator is given by

$$\mathcal{S} = \sum_{i\in\mathcal{I}} \mathcal{S}_i,$$

where $\mathcal{S}_i$ is the $\mathcal{X}$-orthogonal projection onto $\mathcal{X}_i$, i.e.,

$$\langle \mathcal{S}_i x\,,\, x_i\rangle_{\mathcal{X}} = \langle x\,,\, x_i\rangle_{\mathcal{X}} \quad \text{for all } x_i \in \mathcal{X}_i \text{ and all } x \in \mathcal{X},$$

where $\langle\cdot\,,\,\cdot\rangle_{\mathcal{X}}$ denotes the scalar product on $\mathcal{X}$. Then, the operator $\mathcal{S}$ is positive definite and symmetric (with respect to $\langle\cdot\,,\,\cdot\rangle_{\mathcal{X}}$). For $x \in \mathcal{X}$, define the multilevel norm

$$|\!|\!| x |\!|\!|_{\mathcal{X}}^2 := \inf\Big\{ \sum_{i\in\mathcal{I}} \|x_i\|_{\mathcal{X}}^2 \,:\, x = \sum_{i\in\mathcal{I}} x_i \quad \text{with } x_i \in \mathcal{X}_i \text{ for all } i \in \mathcal{I}\Big\}. \tag{6.51}$$

It is proved, e.g., in [Osw94, Theorem 16] that $\langle \mathcal{S}^{-1} x\,,\, x\rangle_{\mathcal{X}} = |\!|\!| x |\!|\!|_{\mathcal{X}}^2$. If there exists a constant $C > 0$ such that

$$C^{-1}\|x\|_{\mathcal{X}} \le |\!|\!| x |\!|\!|_{\mathcal{X}} \le C\,\|x\|_{\mathcal{X}} \quad \text{for all } x \in \mathcal{X},$$

then the extreme eigenvalues of $\mathcal{S}^{-1}$ (and hence those of $\mathcal{S}$) are bounded (from above and below). In particular, the additive Schwarz operator $\mathcal{S}$ is optimal in the sense that its condition number (ratio of largest and smallest eigenvalues) depends only on $C > 0$.

Let $\mathbf{S}$ denote the matrix representation of $\mathcal{S}$. Then, the norm equivalence from above and the latter observations imply that the condition number of $\mathbf{S}$ is bounded. The abstract theory on additive Schwarz operators given in [TW05, Chapter 2] shows that $\mathbf{S}$ has the form $\mathbf{S} = \mathbf{P}^{-1}\mathbf{A}$, where $\mathbf{A}$ is the Galerkin matrix of $\langle\cdot\,,\,\cdot\rangle_{\mathcal{X}}$. Therefore, boundedness of the condition number of $\mathbf{S}$ implies *optimality* of the preconditioner $\mathbf{P}^{-1}$.

We shortly discuss the matrix representation (6.36) of the additive Schwarz preconditioner

$$\mathbf{P}_L^{-1} := \mathbf{I}_{0,L}\mathbf{A}_0^{-1}\mathbf{I}_{0,L}^T + \sum_{\ell=1}^{L} \mathbf{I}_{\ell,L}\mathbf{H}_\ell\mathbf{D}_\ell\mathbf{H}_\ell^T\mathbf{I}_{\ell,L}^T.$$

Following [TW05, Chapter 2], let $\mathbf{A}_i$ denote the Galerkin matrix of $\langle\cdot\,,\,\cdot\rangle_{\mathcal{X}}$ restricted to $\mathcal{X}_i$, and let $\mathbf{I}_i$ denote the matrix that realizes the embedding from $\mathcal{X}_i \to \mathcal{X}$. We consider the matrix representation $\mathbf{S}_i$ of $\mathcal{S}_i\colon \mathcal{X} \to \mathcal{X}_i \subset \mathcal{X}$. Let $x \in \mathcal{X}$ with coordinate vector $\mathbf{x}$, and let $x_i \in \mathcal{X}_i$ be arbitrary with coordinate vector $\mathbf{x}_i$. The defining relation of $\mathcal{S}_i$, i.e.,

$$\langle \mathcal{S}_i x\,,\, x_i\rangle_{\mathcal{X}} = \langle x\,,\, x_i\rangle_{\mathcal{X}} \quad \text{for all } x_i \in \mathcal{X}_i,$$

then reads in matrix-vector form as

$$\mathbf{x}_i \cdot (\mathbf{A}_i \mathbf{S}_i \mathbf{x}) = (\mathbf{I}_i \mathbf{x}_i) \cdot (\mathbf{A}\mathbf{x}) \quad \text{for all coefficient vectors } \mathbf{x}_i,$$

or equivalently

$$\mathbf{A}_i \mathbf{S}_i \mathbf{x} = \mathbf{I}_i^T \mathbf{A} \mathbf{x}.$$

Since $\mathbf{A}_i$ is invertible, we have that

$$\mathbf{S}_i = \mathbf{A}_i^{-1} \mathbf{I}_i^T \mathbf{A}.$$

Note that the range of the operator $\mathcal{S}_i$ is $\mathcal{X}_i$ and correspondingly for the matrix representation $\mathbf{S}_i$. We therefore apply the embedding $\mathbf{I}_i$ and obtain the representation

$$\mathbf{S} = \mathbf{P}^{-1}\mathbf{A}, \quad \text{where} \quad \mathbf{P}^{-1} = \sum_{i \in \mathcal{I}} \mathbf{I}_i \mathbf{A}_i^{-1} \mathbf{I}_i^T.$$

To finally prove (6.36), note that for one-dimensional subspaces $\mathcal{X}_i$, $\mathbf{A}_i$ reduces to the diagonal entry of the matrix $\mathbf{A}$. Overall, we thus derive the matrix representation (6.36).

Subspace decomposition of $\mathcal{ND}^1(\widehat{\mathcal{T}}_\bullet)$ in $H(\mathbf{curl}\,;\Omega)$

The following result is taken from [HWZ12, Theorem 4.1], see also the references therein. In particular, we note that their proof requires the assumption that Ω is simply connected.

Proposition 66. *Let* $\mathcal{Y}_\bullet := \mathcal{ND}^1(\widehat{\mathcal{T}}_\bullet)$, $\mathcal{Y}_{\bullet,E} := \operatorname{span}\{\boldsymbol{u}_{\bullet,E}\}$, $\mathcal{Y}_{\bullet,\mathbf{x}} := \operatorname{span}\{\nabla \eta_{\bullet,\mathbf{x}}\}$, *and*

$$\widehat{\mathcal{E}}_\ell^\star := (\widehat{\mathcal{E}}_\ell \setminus \widehat{\mathcal{E}}_{\ell-1}) \cup \big\{E \in \widehat{\mathcal{E}}_\ell \,:\, \operatorname{supp} \boldsymbol{u}_{\ell,E} \subsetneqq \operatorname{supp} \boldsymbol{u}_{\ell-1,E}\big\},$$
$$\widehat{\mathcal{N}}_\ell^\star := (\widehat{\mathcal{N}}_\ell \setminus \widehat{\mathcal{N}}_{\ell-1}) \cup \big\{\mathbf{x} \in \widehat{\mathcal{N}}_\ell \,:\, \operatorname{supp} \eta_{\ell,\mathbf{x}} \subsetneqq \operatorname{supp} \eta_{\ell-1,\mathbf{x}}\big\}.$$

Then, it holds that

$$\mathcal{Y}_L = \mathcal{Y}_0 + \sum_{\ell=1}^{L} \Big(\sum_{E \in \widehat{\mathcal{E}}_\ell^\star} \mathcal{Y}_{\ell,E} + \sum_{\mathbf{x} \in \widehat{\mathcal{N}}_\ell^\star} \mathcal{Y}_{\ell,\mathbf{x}} \Big). \tag{6.52}$$

Moreover, it holds that

$$C^{-1} \|\boldsymbol{v}\|_{\boldsymbol{H}(\mathbf{curl}\,;\Omega)} \le |\!|\!| \boldsymbol{v} |\!|\!|_{\mathcal{Y}_L} \le C \, \|\boldsymbol{v}\|_{\boldsymbol{H}(\mathbf{curl}\,;\Omega)} \quad \textit{for all } \boldsymbol{v} \in \mathcal{Y}_L, \tag{6.53}$$

where $C > 0$ *depends only on* Ω *and* $\widehat{\mathcal{T}}_0$. □

Subspace decomposition of $\mathcal{P}^0(\mathcal{T}_\bullet)$ in $H^{-1/2}(\Gamma)$

It remains to prove the following proposition to conclude the proof of Theorem 60 since then we get from the abstract theory that the proposed additive Schwarz operator is optimal.

Proposition 67. *The multilevel norm $|||\cdot|||_{\mathcal{X}_L}$ associated with the decomposition (6.35) satisfies the equivalence*

$$C^{-1}\|\psi\|_{H^{-1/2}(\Gamma)} \le |||\psi|||_{\mathcal{X}_L} \le C\,\|\psi\|_{H^{-1/2}(\Gamma)} \quad \text{for all } \psi \in \mathcal{P}^0(\mathcal{T}_L), \tag{6.54}$$

where $C > 0$ depends only on Ω and $\widehat{\mathcal{T}}_0$.

Proof of lower estimate in (6.54). Let $\psi \in \mathcal{P}^0(\mathcal{T}_L)$, $\mathcal{X}_\ell := \mathcal{P}^0(\mathcal{T}_\ell)$, and $\mathcal{X}_{\ell,E} := \operatorname{span}\{\varphi_{\ell,E}\}$. Lemma 59 shows that we can decompose ψ (not necessarily uniquely) into

$$\psi = \psi_0 + \psi_* \tag{6.55}$$

where

$$\psi_* = \sum_{\ell=1}^{L} \sum_{E\in\mathcal{E}_\ell^\star} \psi_{\ell,E} \quad \text{with } \psi_0 \in \mathcal{X}_0 \text{ and } \psi_{\ell,E} \in \mathcal{X}_{\ell,E}.$$

Note that $\mathcal{X}_{\ell,E} \subset \mathcal{P}^0_*(\mathcal{T}_\ell)$. Recall the extension operator $\mathbf{E}_\ell$ from Lemma 65. Define

$$\boldsymbol{v}_* := \sum_{\ell=1}^{L} \sum_{E\in\mathcal{E}_\ell^\star} \mathbf{E}_\ell\psi_{\ell,E} \in \mathcal{Y}_L. \tag{6.56}$$

Then, due to the linearity of the curl operator and Lemma 65, it follows that

$$\begin{aligned}
\operatorname{curl}\boldsymbol{v}_*\cdot\mathbf{n}|_\Gamma &= \operatorname{curl}\Big(\sum_{\ell=1}^{L}\sum_{E\in\mathcal{E}_\ell^\star}\mathbf{E}_\ell\psi_{\ell,E}\Big)\cdot\mathbf{n}|_\Gamma \\
&= \sum_{\ell=1}^{L}\sum_{E\in\mathcal{E}_\ell^\star}\operatorname{curl}\left(\mathbf{E}_\ell\psi_{\ell,E}\right)\cdot\mathbf{n}|_\Gamma \\
&\overset{(6.49)}{=} \sum_{\ell=1}^{L}\sum_{E\in\mathcal{E}_\ell^\star}\psi_{\ell,E} \\
&= \psi_*
\end{aligned}$$

and hence the continuity of the trace operator in $\boldsymbol{H}(\operatorname{div};\Omega)$ yields that

$$\begin{aligned}
\|\psi_*\|_{H^{-1/2}(\Gamma)} &\lesssim \|\operatorname{curl}\boldsymbol{v}_*\|_{\boldsymbol{H}(\operatorname{div};\Omega)} \\
&= \|\operatorname{curl}\boldsymbol{v}_*\|_{L^2(\Omega)} \\
&\le \|\boldsymbol{v}_*\|_{\boldsymbol{H}(\operatorname{curl};\Omega)} \\
&\overset{(6.53)}{\lesssim} |||\boldsymbol{v}_*|||_{\mathcal{Y}_L}.
\end{aligned}$$

Moreover, the triangle inequality, the definition of the multilevel norm $|||\cdot|||_{\mathcal{Y}_L}$, and Lemma 65 show that

$$\begin{aligned}
\|\psi\|^2_{H^{-1/2}(\Gamma)} &\lesssim \|\psi_0\|^2_{H^{-1/2}(\Gamma)} + \|\psi_*\|^2_{H^{-1/2}(\Gamma)} \\
&\lesssim \|\psi_0\|^2_{H^{-1/2}(\Gamma)} + |||\boldsymbol{v}_*|||^2_{\mathcal{Y}_L} \\
&\overset{(6.51)}{\leq} \|\psi_0\|^2_{H^{-1/2}(\Gamma)} + \sum_{\ell=1}^{L} \sum_{E\in\mathcal{E}_\ell^\star} \|\mathbf{E}_\ell \psi_{\ell,E}\|^2_{\boldsymbol{H}(\operatorname{curl};\Omega)} \\
&\overset{(6.50)}{\lesssim} \|\psi_0\|^2_{H^{-1/2}(\Gamma)} + \sum_{\ell=1}^{L} \sum_{E\in\mathcal{E}_\ell^\star} \|\psi_{\ell,E}\|^2_{H^{-1/2}(\Gamma)}.
\end{aligned}$$

Taking the infimum over all possible decompositions (6.55), we derive the lower estimate in (6.54) by definition (6.51) of the multilevel norm. □

Proof of upper estimate in (6.54). Let $\psi \in \mathcal{P}^0(\mathcal{T}_L)$. Define $\psi_{00} := \langle \psi\,,\, 1\rangle_\Gamma/|\Gamma|$ as the integral mean of ψ over Γ. Moreover, let $\psi_* := \psi - \psi_{00} \in \mathcal{P}^0_*(\mathcal{T}_L)$. Note that

$$\begin{aligned}
\|\psi_*\|_{H^{-1/2}(\Gamma)} &\leq \|\psi\|_{H^{-1/2}(\Gamma)} + \|\psi_{00}\|_{H^{-1/2}(\Gamma)} \\
&\leq \big(1 + \|1/|\Gamma|\|_{H^{1/2}(\Gamma)}\big)\, \|\psi\|_{H^{-1/2}(\Gamma)} \\
&\lesssim \|\psi\|_{H^{-1/2}(\Gamma)}.
\end{aligned} \tag{6.57}$$

With Lemma 65, choose $\boldsymbol{v} = \mathbf{E}_L\psi_* \in \mathcal{Y}_L = \boldsymbol{\mathcal{ND}}^1(\widehat{\mathcal{T}}_L)$. Hence, we get that

$$\psi_* = \operatorname{curl} \boldsymbol{v} \cdot \mathbf{n}|_\Gamma$$

as well as

$$\|\boldsymbol{v}\|_{\boldsymbol{H}(\operatorname{curl};\Omega)} \lesssim \|\psi_*\|_{H^{-1/2}(\Gamma)}. \tag{6.58}$$

The upper bound in Proposition 66 further provides $\boldsymbol{v}_0 \in \mathcal{Y}_0$, $\boldsymbol{v}_{\ell,E} \in \mathcal{Y}_{\ell,E}$, and $\boldsymbol{v}_{\ell,\mathbf{x}} \in \mathcal{Y}_{\ell,\mathbf{x}}$ such that

$$\boldsymbol{v} = \boldsymbol{v}_0 + \sum_{\ell=1}^{L} \Bigg(\sum_{E\in\widehat{\mathcal{E}}_\ell^\star} \boldsymbol{v}_{\ell,E} + \sum_{\mathbf{x}\in\widehat{\mathcal{N}}_\ell^\star} \boldsymbol{v}_{\ell,\mathbf{x}} \Bigg)$$

as well as

$$\begin{aligned}
&\|\boldsymbol{v}_0\|^2_{\boldsymbol{H}(\operatorname{curl};\Omega)} + \sum_{\ell=1}^{L} \Bigg(\sum_{E\in\widehat{\mathcal{E}}_\ell^\star} \|\boldsymbol{v}_{\ell,E}\|^2_{\boldsymbol{H}(\operatorname{curl};\Omega)} + \sum_{\mathbf{x}\in\widehat{\mathcal{N}}_\ell^\star} \|\boldsymbol{v}_{\ell,\mathbf{x}}\|^2_{\boldsymbol{H}(\operatorname{curl};\Omega)} \Bigg) \\
&\qquad \overset{(6.53)}{\lesssim} \|\boldsymbol{v}\|^2_{\boldsymbol{H}(\operatorname{curl};\Omega)}.
\end{aligned} \tag{6.59}$$

Since $\boldsymbol{v}_{\ell,\mathbf{x}} \in \mathcal{Y}_{\ell,\mathbf{x}} = \operatorname{span}\{\nabla\eta_{\ell,\mathbf{x}}\}$ and the curl of the gradient vanishes, we observe that

$$\operatorname{curl} \boldsymbol{v}_{\ell,\mathbf{x}} = 0.$$

Thus, we see that

$$\begin{aligned}\psi &= \psi_{00} + \psi_* \\ &= \psi_{00} + \operatorname{curl} \boldsymbol{v} \cdot \mathbf{n}|_\Gamma \\ &= \psi_{00} + \operatorname{curl} \boldsymbol{v}_0 \cdot \mathbf{n}|_\Gamma + \sum_{\ell=1}^{L} \sum_{E \in \widehat{\mathcal{E}}_\ell^\star} \operatorname{curl} \boldsymbol{v}_{\ell,E} \cdot \mathbf{n}|_\Gamma \\ &= \psi_{00} + \operatorname{curl} \boldsymbol{v}_0 \cdot \mathbf{n}|_\Gamma + \sum_{\ell=1}^{L} \sum_{E \in \mathcal{E}_\ell^\star} \operatorname{curl} \boldsymbol{v}_{\ell,E} \cdot \mathbf{n}|_\Gamma,\end{aligned}$$

where the latter sum reduces to a sum over all $E \in \mathcal{E}_\ell^\star$ (instead of all $E \in \widehat{\mathcal{E}}_\ell^\star$) due to the restriction $(\cdot)|_\Gamma$ to the boundary. Note that $\psi_{*0} := \operatorname{curl} \boldsymbol{v}_0 \cdot \mathbf{n}|_\Gamma \in \mathcal{X}_0 = \mathcal{P}^0(\mathcal{T}_0)$ and hence $\psi_{00} + \psi_{*0} \in \mathcal{X}_0$. Moreover, it holds that

$$\begin{aligned}\|\psi_{00} + \psi_{*0}\|_{H^{-1/2}(\Gamma)} &\le \|\psi_{00}\|_{H^{-1/2}(\Gamma)} + \|\operatorname{curl} \boldsymbol{v}_0 \cdot \mathbf{n}\|_{H^{-1/2}(\Gamma)} \\ &\lesssim \|\psi\|_{H^{-1/2}(\Gamma)} + \|\operatorname{curl} \boldsymbol{v}_0\|_{\boldsymbol{H}(\operatorname{div};\Omega)} \\ &= \|\psi\|_{H^{-1/2}(\Gamma)} + \|\operatorname{curl} \boldsymbol{v}_0\|_{L^2(\Omega)}.\end{aligned} \tag{6.60}$$

Due to Lemma 64 and $\boldsymbol{v}_{\ell,E} \in \mathcal{Y}_{\ell,E} = \operatorname{span}\{\boldsymbol{u}_{\ell,E}\}$, it holds that

$$\psi_{\ell,E} := \operatorname{curl} \boldsymbol{v}_{\ell,E} \cdot \mathbf{n}|_\Gamma \in \mathcal{X}_{\ell,E} = \operatorname{span}\{\varphi_{\ell,E}\}$$

with

$$\|\psi_{\ell,E}\|_{H^{-1/2}(\Gamma)} \simeq \|\boldsymbol{v}_{\ell,E}\|_{\boldsymbol{H}(\operatorname{curl};\Omega)}.$$

We hence see that

$$\psi = (\psi_{00} + \psi_{*0}) + \sum_{\ell=1}^{L} \sum_{E \in \mathcal{E}_\ell^\star} \psi_{\ell,E}$$

with

$$\begin{aligned}|\!|\!|\psi|\!|\!|^2_{\mathcal{P}^0(\mathcal{T}_L)} &\overset{(6.51)}{\le} \|\psi_{00} + \psi_{*0}\|^2_{H^{-1/2}(\Gamma)} + \sum_{\ell=1}^{L} \sum_{E \in \mathcal{E}_\ell^\star} \|\psi_{\ell,E}\|^2_{H^{-1/2}(\Gamma)} \\ &\overset{(6.60)}{\lesssim} \|\psi\|^2_{H^{-1/2}(\Gamma)} + \|\boldsymbol{v}_0\|^2_{\boldsymbol{H}(\operatorname{curl};\Omega)} + \sum_{\ell=1}^{L} \sum_{E \in \mathcal{E}_\ell^\star} \|\boldsymbol{v}_{\ell,E}\|^2_{\boldsymbol{H}(\operatorname{curl};\Omega)} \\ &\overset{(6.59)}{\lesssim} \|\psi\|^2_{H^{-1/2}(\Gamma)} + \|\boldsymbol{v}\|^2_{\boldsymbol{H}(\operatorname{curl};\Omega)} \\ &\overset{(6.58)}{\lesssim} \|\psi\|^2_{H^{-1/2}(\Gamma)} + \|\psi_*\|^2_{H^{-1/2}(\Gamma)} \\ &\overset{(6.57)}{\lesssim} \|\psi\|^2_{H^{-1/2}(\Gamma)}.\end{aligned}$$

This concludes the proof. □

6.5.3 Optimal convergence

In this section we present the first main result for the adaptive Algorithm 57. We note, that Algorithm 57 as well as the following theorem are independent of the precise preconditioning strategy as long as the employed preconditioners are optimal in the sense of Section 6.3.4.

First, we recall the index set $\mathcal{Q}$ of Section 6.4 which is defined by

$$\mathcal{Q} := \big\{(\ell,k) \in \mathbb{N}_0 \times \mathbb{N}_0 \,:\, \text{index } (\ell,k) \text{ is used in Algorithm 57}\big\}$$

Then, we get the following theorem.

Theorem 68. *The output of Algorithm 57 satisfies the following assertions* (a)–(c).

(a) *There exists a constant* $C^\star_{\mathrm{rel}} > 0$ *such that*

$$|||\phi^\star - \phi^k_\ell||| \le C^\star_{\mathrm{rel}} \big(\eta_\ell(\phi^k_\ell) + |||\phi^k_\ell - \phi^{k-1}_\ell|||\big). \tag{6.61}$$

for all $(\ell,k) \in \mathcal{Q}$ *with* $k \ge 1$.

There exists a constant $C^\star_{\mathrm{eff}} > 0$ *such that, provided that* $\phi^\star \in L^2(\Gamma)$, *it holds that*

$$\eta_\ell(\phi^k_\ell) \le C^\star_{\mathrm{eff}} \big(\|h_\ell^{1/2}(\phi^\star - \phi^k_\ell)\|_{L^2(\Gamma)} + |||\phi^k_\ell - \phi^{k-1}_\ell|||\big). \tag{6.62}$$

for all $(\ell,k) \in \mathcal{Q}$ *with* $k \ge 1$.

(b) *For arbitrary* $0 < \theta \le 1$ *and arbitrary* $\lambda_{\mathrm{ctr}} > 0$, *there exist constants* $C_{\mathrm{lin}} \ge 1$ *and* $0 < q_{\mathrm{lin}} < 1$ *such that the quasi-error*

$$\Lambda^k_\ell := \big(|||\phi^\star - \phi^k_\ell|||^2 + \eta_\ell(\phi^k_\ell)^2\big)^{1/2} \tag{6.63}$$

is linearly convergent in the sense of

$$\Lambda^{k'}_{\ell'} \le C_{\mathrm{lin}}\, q_{\mathrm{lin}}^{|(\ell',k')|-|(\ell,k)|}\, \Lambda^k_\ell \tag{6.64}$$

for all $(\ell,k),(\ell',k') \in \mathcal{Q}$ *with* $(\ell',k') \ge (\ell,k)$.

(c) *For* $s > 0$, *define the approximation class*

$$\|\phi^\star\|_{\mathbb{A}_s} := \sup_{N\in\mathbb{N}_0} \Big((N+1)^s \min_{\substack{\mathcal{T}_\bullet \in \mathtt{refine}(\mathcal{T}_0)\\ \#\mathcal{T}_\bullet - \#\mathcal{T}_0 \le N}} \eta_\bullet(\phi^\star_\bullet)\Big). \tag{6.65}$$

Then, for sufficiently small $0 < \theta \ll 1$ *and* $0 < \lambda_{\mathrm{ctr}} \ll 1$, *cf. Assumption* (6.86) *below, and all* $s > 0$, *it holds that*

$$\begin{aligned} &\|\phi^\star\|_{\mathbb{A}_s} < \infty \\ &\iff \\ &\exists\, C_{\mathrm{opt}} > 0 : \sup_{(\ell,k)\in\mathcal{Q}} \big(\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1\big)^s\, \Lambda^k_\ell \le C_{\mathrm{opt}}\, \|\phi^\star\|_{\mathbb{A}_s} < \infty. \end{aligned} \tag{6.66}$$

The constants

- $C^\star_{\mathrm{rel}}, C^\star_{\mathrm{eff}} > 0$ *depend only on* q_{pcg}, Γ, *and the uniform* γ*-shape regularity of* $\mathcal{T}_j \in \texttt{refine}(\mathcal{T}_0)$,
- $C_{\mathrm{lin}} > 1$ *and* $0 < q_{\mathrm{lin}} < 1$ *depend additionally only on* θ *and* λ_{ctr}, *and*
- $C_{\mathrm{opt}} > 0$ *depends additionally only on* s, $\mathcal{T}_0$, *and* Λ_0^k.

Remark 69. *By definition, it holds that*

$$\eta_\ell(\phi_\ell^k) \le \Lambda_\ell^k \quad \text{for all } (\ell, k) \in \mathcal{Q}.$$

If $\phi_\ell^k \in \{\phi_\ell^\star, \phi_\ell^{\underline{k}}\}$, *then there also holds the converse inequality and hence*

$$\eta_\ell(\phi_\ell^k) \simeq \Lambda_\ell^k.$$

To see this, note that $\phi_\ell^k = \phi_\ell^\star$ *and* (6.24) *prove that*

$$\Lambda_\ell^k \le (1 + C_{\mathrm{rel}})\, \eta_\ell(\phi_\ell^k).$$

If $\phi_\ell^k = \phi_\ell^{\underline{k}}$, *then Theorem 68*(a) *and the stopping criterion* (6.30) *of Algorithm 57 prove that*

$$\begin{aligned}\Lambda_\ell^{\underline{k}} &\le (1 + C^\star_{\mathrm{rel}})\, \eta_\ell(\phi_\ell^{\underline{k}}) + |||\phi_\ell^{\underline{k}} - \phi_\ell^{\underline{k}-1}||| \\ &\le (1 + C^\star_{\mathrm{rel}} + \lambda_{\mathrm{ctr}})\, \eta_\ell(\phi_\ell^{\underline{k}}).\end{aligned}$$

6.5.4 Proof of Theorem 68 (optimal convergence rates)

First, we give an abstract analysis in the spirit of [CFPP14], where the precise problem and discretization (i.e., Galerkin BEM with piecewise constants for the weakly-singular integral equation for the 2D and 3D Laplacian) enter only through certain properties of the error estimator. These properties are explicitly stated in the next subsection, before we provide general PCG estimates afterwards. The remaining sections, i.e., the proofs of Theorem 68(a)–(c) then only exploit these abstract frameworks.

Axioms of adaptivity

In this section, similarly to Section 4.3, we recall some structural properties of the residual error estimator (6.22) which have been identified in [CFPP14] to be important and sufficient for the numerical analysis of Algorithm 57.

For ease of notation, let $\mathcal{T}_0$ be the fixed initial mesh of Algorithm 57. Let $\mathbb{T} := \texttt{refine}(\mathcal{T}_0)$ be the set of all possible meshes that can be obtained by successively refining $\mathcal{T}_0$.

Proposition 70. *There exist constants* $C_{\mathrm{stb}}, C_{\mathrm{red}}, C_{\mathrm{rel}} > 0$ *and* $0 < q_{\mathrm{red}} < 1$ *which depend only on* Γ *and the* γ*-shape regularity, such that the following properties* (A1)–(A4) *hold:*

(A1) stability on non-refined element domains: For each mesh $\mathcal{T}_\bullet \in \mathbb{T}$, all refinements $\mathcal{T}_\circ \in \operatorname{refine}(\mathcal{T}_\bullet)$, arbitrary discrete functions $v_\circ \in \mathcal{P}^0(\mathcal{T}_\circ)$ and $w_\bullet \in \mathcal{P}^0(\mathcal{T}_\bullet)$, and an arbitrary set $\mathcal{U}_\bullet \subseteq \mathcal{T}_\bullet \cap \mathcal{T}_\circ$ of non-refined elements, it holds that

$$|\eta_\circ(\mathcal{U}_\bullet, v_\circ) - \eta_\bullet(\mathcal{U}_\bullet, w_\bullet)| \leq C_{\mathrm{stb}} \, |\!|\!| v_\circ - w_\bullet |\!|\!|.$$

(A2) reduction on refined element domains: For each mesh $\mathcal{T}_\bullet \in \mathbb{T}$, all refinements $\mathcal{T}_\circ \in \operatorname{refine}(\mathcal{T}_\bullet)$, arbitrary discrete functions $v_\circ \in \mathcal{P}^0(\mathcal{T}_\circ)$ and $w_\bullet \in \mathcal{P}^0(\mathcal{T}_\bullet)$, it holds that

$$\eta_\circ(\mathcal{T}_\circ \backslash \mathcal{T}_\bullet, v_\circ)^2 \leq q_{\mathrm{red}} \, \eta_\bullet(\mathcal{T}_\bullet \backslash \mathcal{T}_\circ, w_\bullet)^2 + C_{\mathrm{red}} \, |\!|\!| v_\circ - w_\bullet |\!|\!|^2.$$

(A3) reliability: For each mesh $\mathcal{T}_\bullet \in \mathbb{T}$, the error of the exact discrete solution $\phi^\star_\bullet \in \mathcal{P}^0(\mathcal{T}_\bullet)$ of (6.16) is controlled by

$$|\!|\!| \phi^\star - \phi^\star_\bullet |\!|\!| \leq C_{\mathrm{rel}} \, \eta_\bullet(\phi^\star_\bullet).$$

(A4) discrete reliability: For each mesh $\mathcal{T}_\bullet \in \mathbb{T}$ and all refinements $\mathcal{T}_\circ \in \operatorname{refine}(\mathcal{T}_\bullet)$, there exists a set $\mathcal{R}_{\bullet,\circ} \subseteq \mathcal{T}_\bullet$ with $\mathcal{T}_\bullet \backslash \mathcal{T}_\circ \subseteq \mathcal{R}_{\bullet,\circ}$ as well as $\#\mathcal{R}_{\bullet,\circ} \leq C_{\mathrm{drl}} \, \#(\mathcal{T}_\bullet \backslash \mathcal{T}_\circ)$ such that the difference of $\phi^\star_\bullet \in \mathcal{P}^0(\mathcal{T}_\bullet)$ and $\phi^\star_\circ \in \mathcal{P}^0(\mathcal{T}_\circ)$ is controlled by

$$|\!|\!| \phi^\star_\circ - \phi^\star_\bullet |\!|\!| \leq C_{\mathrm{drl}} \, \eta_\bullet(\mathcal{R}_{\bullet,\circ}, \phi^\star_\bullet).$$

□

Remark 71. *For the proof of Proposition 70, we refer to [FKMP13, FFK+14]. We only note that (A4) already implies (A3) with $C_{\mathrm{rel}} \leq C_{\mathrm{drl}}$ in general, cf. [CFPP14, Section 3.3].*

Energy estimates for the PCG solver

This section collects some auxiliary results which rely on the use of PCG and, in particular, PCG with an optimal preconditioner. We first note the following Pythagoras identity.

Lemma 72. *Let $\mathbf{A}_\bullet, \mathbf{P}_\bullet \in \mathbb{R}^{N \times N}$ be symmetric and positive definite, $\mathbf{b}_\bullet \in \mathbb{R}^N$, $\mathbf{x}^\star_\bullet := \mathbf{A}_\bullet^{-1}\mathbf{b}_\bullet$, $\mathbf{x}^0_\bullet \in \mathbb{R}^N$, and $\mathbf{x}^k_\bullet \in \mathbb{R}^N$ the iterates of the PCG algorithm.*
There holds the Pythagoras identity

$$|\!|\!| \phi^\star_\bullet - \phi^k_\bullet |\!|\!|^2 = |\!|\!| \phi^\star_\bullet - \phi^{k+1}_\bullet |\!|\!|^2 + |\!|\!| \phi^{k+1}_\bullet - \phi^k_\bullet |\!|\!|^2 \quad \text{for all } k \in \mathbb{N}_0. \tag{6.67}$$

Proof. Recall that $\widetilde{\mathbf{x}}^\star_\bullet$ is the solution to (6.1) and $\widetilde{\mathbf{x}}^k_\bullet = \mathbf{P}_\bullet^{1/2}\mathbf{x}^k_\bullet$. According to the definition of PCG (and CG), it then holds that

$$\|\widetilde{\mathbf{x}}^\star_\bullet - \widetilde{\mathbf{x}}^k_\bullet\|_{\widetilde{\mathbf{A}}_\bullet} = \min_{\widetilde{\mathbf{y}}_\bullet \in \mathcal{K}_k(\widetilde{\mathbf{A}}_\bullet, \widetilde{\mathbf{b}}_\bullet, \widetilde{\mathbf{x}}^0_\bullet)} \|\widetilde{\mathbf{x}}^\star_\bullet - \widetilde{\mathbf{y}}_\bullet\|_{\widetilde{\mathbf{A}}_\bullet},$$

where

$$\mathcal{K}_k(\widetilde{\mathbf{A}}_\bullet, \widetilde{\mathbf{b}}_\bullet, \widetilde{x}_\bullet^0) := \operatorname{span}\{\widetilde{r}_\bullet^0, \widetilde{\mathbf{A}}_\bullet \widetilde{r}_\bullet^0, \dots, \widetilde{\mathbf{A}}_\bullet^{k-1} \widetilde{r}_\bullet^0\} \quad \text{with} \quad \widetilde{r}_\bullet^0 := \widetilde{\mathbf{b}}_\bullet - \widetilde{\mathbf{A}}_\bullet \widetilde{\mathbf{x}}_\bullet^0.$$

According to Linear Algebra, $\widetilde{\mathbf{x}}_\bullet^k$ is the orthogonal projection of $\widetilde{\mathbf{x}}_\bullet^\star$ in $\mathcal{K}_k(\widetilde{\mathbf{A}}_\bullet, \widetilde{\mathbf{b}}_\bullet, \widetilde{\mathbf{x}}_\bullet^0)$ with respect to the matrix norm $\|\cdot\|_{\widetilde{\mathbf{A}}_\bullet}$. From nestedness $\mathcal{K}_k(\widetilde{\mathbf{A}}_\bullet, \widetilde{\mathbf{b}}_\bullet, \widetilde{\mathbf{x}}_\bullet^0) \subseteq \mathcal{K}_{k+1}(\widetilde{\mathbf{A}}_\bullet, \widetilde{\mathbf{b}}_\bullet, \widetilde{\mathbf{x}}_\bullet^0)$, it thus follows that

$$\|\widetilde{\mathbf{x}}_\bullet^\star - \widetilde{\mathbf{x}}_\bullet^k\|_{\widetilde{\mathbf{A}}_\bullet}^2 = \|\widetilde{\mathbf{x}}_\bullet^\star - \widetilde{\mathbf{x}}_\bullet^{k+1}\|_{\widetilde{\mathbf{A}}_\bullet}^2 + \|\widetilde{\mathbf{x}}_\bullet^{k+1} - \widetilde{\mathbf{x}}_\bullet^k\|_{\widetilde{\mathbf{A}}_\bullet}^2.$$

Hence, together with (6.26) and (6.29), we get that

$$\begin{aligned}
|||\phi_\bullet^\star - \phi_\bullet^k|||^2 &\overset{(6.29)}{=} \|\mathbf{x}_\bullet^\star - \mathbf{x}_\bullet^k\|_{\mathbf{A}_\bullet}^2 \\
&\overset{(6.26)}{=} \|\widetilde{\mathbf{x}}_\bullet^\star - \widetilde{\mathbf{x}}_\bullet^k\|_{\widetilde{\mathbf{A}}_\bullet}^2 \\
&= \|\widetilde{\mathbf{x}}_\bullet^\star - \widetilde{\mathbf{x}}_\bullet^{k+1}\|_{\widetilde{\mathbf{A}}_\bullet}^2 + \|\widetilde{\mathbf{x}}_\bullet^{k+1} - \widetilde{\mathbf{x}}_\bullet^k\|_{\widetilde{\mathbf{A}}_\bullet}^2 \\
&\overset{(6.26)}{=} \|\mathbf{x}_\bullet^\star - \mathbf{x}_\bullet^{k+1}\|_{\mathbf{A}_\bullet}^2 + \|\mathbf{x}_\bullet^{k+1} - \mathbf{x}_\bullet^k\|_{\mathbf{A}_\bullet}^2 \\
&\overset{(6.29)}{=} |||\phi_\bullet^\star - \phi_\bullet^{k+1}|||^2 + |||\phi_\bullet^{k+1} - \phi_\bullet^k|||^2
\end{aligned}$$

which proves (6.67). □

The following lemma collects some estimates which follow from the contraction property (6.28) of PCG.

Lemma 73. *Algorithm 57 guarantees the following estimates for all* $(\ell, k) \in \mathcal{Q}$ *with* $k \geq 1$*:*

(i) $|||\phi_\ell^\star - \phi_\ell^k||| \leq q_{\mathrm{pcg}} |||\phi_\ell^\star - \phi_\ell^{k-1}|||$

(ii) $|||\phi_\ell^k - \phi_\ell^{k-1}||| \leq (1 + q_{\mathrm{pcg}}) |||\phi_\ell^\star - \phi_\ell^{k-1}|||$

(iii) $|||\phi_\ell^\star - \phi_\ell^{k-1}||| \leq (1 - q_{\mathrm{pcg}})^{-1} |||\phi_\ell^k - \phi_\ell^{k-1}|||$

(iv) $|||\phi_\ell^\star - \phi_\ell^k||| \leq q_{\mathrm{pcg}}(1 - q_{\mathrm{pcg}})^{-1} |||\phi_\ell^k - \phi_\ell^{k-1}|||$

Proof. Combining (6.29) and the contraction property (6.28) of PCG, we get that

$$\begin{aligned}
|||\phi_\ell^\star - \phi_\ell^k||| &\overset{(6.29)}{=} \|\mathbf{x}_\ell^\star - \mathbf{x}_\ell^k\|_{\mathbf{A}_\ell} \\
&\overset{(6.28)}{\leq} q_{\mathrm{pcg}} \|\mathbf{x}_\ell^\star - \mathbf{x}_\ell^{k-1}\|_{\mathbf{A}_\ell} \\
&\overset{(6.29)}{=} q_{\mathrm{pcg}} |||\phi_\ell^\star - \phi_\ell^{k-1}|||
\end{aligned}$$

which proves (i). Estimate (ii) follows from (i) and the triangle inequality by

$$\begin{aligned}
|||\phi_\ell^k - \phi_\ell^{k-1}||| &\leq |||\phi_\ell^\star - \phi_\ell^k||| + |||\phi_\ell^\star - \phi_\ell^{k-1}||| \\
&\overset{\text{(i)}}{\leq} (1 + q_{\mathrm{pcg}}) |||\phi_\ell^\star - \phi_\ell^{k-1}|||.
\end{aligned}$$

Estimates (iii) follows again from (i) and the triangle inequality by

$$\begin{aligned}\|\!|\phi_\ell^\star - \phi_\ell^{k-1}|\!\| &\le \|\!|\phi_\ell^\star - \phi_\ell^k|\!\| + \|\!|\phi_\ell^k - \phi_\ell^{k-1}|\!\| \\ &\overset{\text{(i)}}{\le} q_{\mathrm{pcg}} \|\!|\phi_\ell^\star - \phi_\ell^{k-1}|\!\| + \|\!|\phi_\ell^k - \phi_\ell^{k-1}|\!\|,\end{aligned}$$

which is equivalent to estimate (iii). The last estimate (iv) follows from

$$\begin{aligned}\|\!|\phi_\ell^\star - \phi_\ell^k|\!\| &\overset{\text{(i)}}{\le} q_{\mathrm{pcg}} \|\!|\phi_\ell^\star - \phi_\ell^{k-1}|\!\| \\ &\overset{\text{(iii)}}{\le} q_{\mathrm{pcg}}(1-q_{\mathrm{pcg}})^{-1} \|\!|\phi_\ell^k - \phi_\ell^{k-1}|\!\|.\end{aligned}$$

This concludes the proof. □

Proof of Theorem 68(a)

With reliability (A3) and stability (A1), we see that for all $(\ell, k) \in \mathcal{Q}$ it holds that

$$\begin{aligned}\|\!|\phi^\star - \phi_\ell^k|\!\| &\le \|\!|\phi^\star - \phi_\ell^\star|\!\| + \|\!|\phi_\ell^\star - \phi_\ell^k|\!\| \\ &\overset{\text{(A3)}}{\lesssim} \eta_\ell(\phi_\ell^\star) + \|\!|\phi_\ell^\star - \phi_\ell^k|\!\| \\ &\overset{\text{(A1)}}{\lesssim} \eta_\ell(\phi_\ell^k) + \|\!|\phi_\ell^\star - \phi_\ell^k|\!\|.\end{aligned}$$

With Lemma 73(iv), we hence prove the reliability estimate (6.61), i.e.,

$$\begin{aligned}\|\!|\phi^\star - \phi_\ell^k|\!\| &\lesssim \eta_\ell(\phi_\ell^k) + \|\!|\phi_\ell^\star - \phi_\ell^k|\!\| \\ &\overset{73\text{(iv)}}{\lesssim} \eta_\ell(\phi_\ell^k) + \|\!|\phi_\ell^k - \phi_\ell^{k-1}|\!\|.\end{aligned}$$

According to [AFF+17], it holds that

$$\begin{aligned}\eta_\ell(\phi_\ell^k) &\lesssim \|h_\ell^{1/2}(\phi^\star - \phi_\ell^k)\|_{L^2(\Gamma)} + \|\!|\phi^\star - \phi_\ell^k|\!\| \\ &\le \|h_\ell^{1/2}(\phi^\star - \phi_\ell^k)\|_{L^2(\Gamma)} + \|\!|\phi^\star - \phi_\ell^\star|\!\| + \|\!|\phi_\ell^\star - \phi_\ell^k|\!\|.\end{aligned}$$

Let $\mathbb{G}_\ell : \widetilde{H}^{-1/2}(\Gamma) \to \mathcal{P}^0(\mathcal{T}_\ell)$ be the Galerkin projection. Let $\Pi_\ell : L^2(\Gamma) \to \mathcal{P}^0(\mathcal{T}_\ell)$ be the L^2-orthogonal projection. With the Céa lemma and a duality argument (see, e.g., [CP06, Theorem 4.1]), we see for all $\psi \in L^2(\Gamma)$ that

$$\|\!|(1-\mathbb{G}_\ell)\psi|\!\| \le \|\!|(1-\Pi_\ell)\psi|\!\| \lesssim \|h_\ell^{1/2}\,\psi\|_{L^2(\Gamma)}.$$

Hence, for $\psi = \phi^\star - \phi_\ell^k$, it follows that

$$\begin{aligned}\|\!|\phi^\star - \phi_\ell^\star|\!\| &= \|\!|(1-\mathbb{G}_\ell)\phi^\star|\!\| \\ &= \|\!|(1-\mathbb{G}_\ell)(\phi^\star - \phi_\ell^k)|\!\| \\ &\lesssim \|h_\ell^{1/2}(\phi^\star - \phi_\ell^k)\|_{L^2(\Gamma)}.\end{aligned}$$

Combining the latter estimates, we see that

$$\eta_\ell(\phi_\ell^k) \lesssim \|h_\ell^{1/2}(\phi^\star - \phi_\ell^k)\|_{L^2(\Gamma)} + |||\phi_\ell^\star - \phi_\ell^k|||.$$

Lemma 73(iv) yields that

$$\begin{aligned}\eta_\ell(\phi_\ell^k) &\lesssim \|h_\ell^{1/2}(\phi^\star - \phi_\ell^k)\|_{L^2(\Gamma)} + |||\phi_\ell^\star - \phi_\ell^k||| \\ &\overset{73(\mathrm{iv})}{\lesssim} \|h_\ell^{1/2}(\phi^\star - \phi_\ell^k)\|_{L^2(\Gamma)} + |||\phi_\ell^k - \phi_\ell^{k-1}|||\end{aligned}$$

and hence concludes the proof of the efficiency estimate (6.62). □

Proof of Theorem 68(b)

The following lemma is the core part of the proof of Theorem 68(b).

Lemma 74. *Consider Algorithm 57 for arbitrary parameters $0 < \theta \le 1$ and $\lambda_{\mathrm{ctr}} > 0$. There exist constants $0 < \mu, q_{\mathrm{ctr}} < 1$ such that*

$$\Delta_\ell^k := \mu\,\eta_\ell(\phi_\ell^k)^2 + |||\phi^\star - \phi_\ell^k|||^2 \quad \text{for } (\ell,k) \in \mathcal{Q}$$

satisfies, for all $\ell \in \mathbb{N}_0$, that

$$\Delta_\ell^{k+1} \le q_{\mathrm{ctr}}\,\Delta_\ell^k \quad \text{for all } 0 \le k < k+1 < \underline{k} \tag{6.68}$$

as well as

$$\Delta_{\ell+1}^0 \le q_{\mathrm{ctr}}\,\Delta_\ell^{\underline{k}-1} \quad \text{for } k = 0. \tag{6.69}$$

Moreover, for all $(\ell',k'),(\ell,k) \in \mathcal{Q}$, it holds that

$$\Delta_{\ell'}^{k'} \le q_{\mathrm{ctr}}^{|(\ell',k')|-|(\ell,k)|}\,\Delta_\ell^k \tag{6.70}$$

provided that $(\ell',k') > (\ell,k)$, $k' < \underline{k}(\ell')$, and $k < \underline{k}(\ell)$.
The constants $0 < \mu, q_{\mathrm{ctr}} < 1$ depend only on λ_{ctr}, θ, q_{pcg}, and the constants in (A1)–(A3).

Proof. The proof is split into five steps.

Step 1. We fix some constants, which are needed below. We note that all these constants depend on $0 < \theta \le 1$ and $\lambda_{\mathrm{ctr}} > 0$, but do not require any additional constraint. First, define

$$0 < q_{\mathrm{est}} := 1 - (1 - q_{\mathrm{red}})\,\theta^2 < 1. \tag{6.71}$$

Second, choose $\gamma > 0$ such that

$$(1+\gamma)\,q_{\mathrm{est}} \overset{(6.71)}{<} 1. \tag{6.72}$$

Third, choose $\mu > 0$ such that

$$\mu\,(1+\gamma^{-1})\,q_{\mathrm{est}}\,C_{\mathrm{stb}}^2\,(1+q_{\mathrm{pcg}})^2 < \frac{1-q_{\mathrm{pcg}}^2}{2} \quad \text{and} \quad \mu\,\lambda_{\mathrm{ctr}}^{-2} \le \frac{1}{2}. \tag{6.73}$$

Fourth, choose $\varepsilon > 0$ such that

$$\varepsilon\,(1-q_{\mathrm{pcg}})^{-2} + 2\,\varepsilon\,C_{\mathrm{rel}}^2\,C_{\mathrm{stb}}^2\,(1-q_{\mathrm{pcg}})^{-2} \;\le\; \frac{1}{2} \quad\text{and}\quad 2\,\varepsilon\,C_{\mathrm{rel}}^2 \;\le\; (1-\varepsilon)\,\mu. \tag{6.74}$$

Fifth, choose $\kappa > 0$ such that

$$2\,\kappa\,C_{\mathrm{rel}}^2 \overset{(6.72)}{<} \big(1-(1+\gamma)\,q_{\mathrm{est}}\big)\,\mu \quad\text{and}\quad 2\,\kappa\,C_{\mathrm{rel}}^2\,C_{\mathrm{stb}}^2 \;<\; \frac{1-q_{\mathrm{pcg}}^2}{2}. \tag{6.75}$$

With (6.73)–(6.75), we finally define

$$\begin{aligned} 0 < q_{\mathrm{ctr}} := \max\Big\{1-\varepsilon,\, \big(\mu\,(1+\gamma)\,q_{\mathrm{est}} + 2\,\kappa\,C_{\mathrm{rel}}^2\big)\,\mu^{-1},\, 1-\kappa,\\ \big(\mu\,(1+\gamma^{-1})\,q_{\mathrm{est}}\,C_{\mathrm{stb}}^2\,(1+q_{\mathrm{pcg}})^2 + q_{\mathrm{pcg}}^2 + 2\,\kappa\,C_{\mathrm{rel}}^2 C_{\mathrm{stb}}^2\big)\Big\} < 1. \end{aligned} \tag{6.76}$$

Step 2. Due to reliability (A3), stability (A1), and Lemma 73(iii), it follows that

$$\begin{aligned} |||\phi^\star - \phi_\ell^\star|||^2 &= (1-\varepsilon)|||\phi^\star - \phi_\ell^\star|||^2 + \varepsilon\,|||\phi^\star - \phi_\ell^\star|||^2 \\ &\overset{(A3)}{\le} (1-\varepsilon)|||\phi^\star - \phi_\ell^\star|||^2 + \varepsilon\,C_{\mathrm{rel}}^2\,\eta_\ell(\phi_\ell^\star)^2 \\ &\overset{(A1)}{\le} (1-\varepsilon)|||\phi^\star - \phi_\ell^\star|||^2 + 2\,\varepsilon\,C_{\mathrm{rel}}^2\,\big(\eta_\ell(\phi_\ell^k)^2 + C_{\mathrm{stb}}^2\,|||\phi_\ell^\star - \phi_\ell^k|||^2\big) \\ &\overset{73(\mathrm{iii})}{\le} (1-\varepsilon)|||\phi^\star - \phi_\ell^\star|||^2 + 2\,\varepsilon\,C_{\mathrm{rel}}^2\,\eta_\ell(\phi_\ell^k)^2 \\ &\qquad + 2\,\varepsilon\,C_{\mathrm{rel}}^2\,C_{\mathrm{stb}}^2\,(1-q_{\mathrm{pcg}})^{-2}\,|||\phi_\ell^{k+1} - \phi_\ell^k|||^2. \end{aligned}$$

Step 3. We consider the case $k+1 < \underline{k}(\ell)$. The stopping criterion (6.30) of Algorithm 57 yields that

$$\eta_\ell(\phi_\ell^{k+1})^2 < \lambda_{\mathrm{ctr}}^{-2}|||\phi_\ell^{k+1} - \phi_\ell^k|||^2. \tag{6.77}$$

Moreover, the Pythagoras identity (6.67) implies that

$$\begin{aligned} |||\phi_\ell^\star - \phi_\ell^{k+1}|||^2 &= |||\phi_\ell^\star - \phi_\ell^k|||^2 - |||\phi_\ell^{k+1} - \phi_\ell^k|||^2 \\ &= (1-\varepsilon)\,|||\phi_\ell^\star - \phi_\ell^k|||^2 + \varepsilon\,|||\phi_\ell^\star - \phi_\ell^k|||^2 - |||\phi_\ell^{k+1} - \phi_\ell^k|||^2. \end{aligned} \tag{6.78}$$

Further, we note the Pythagoras identity

$$|||\phi^\star - \phi_\ell^\star|||^2 + |||\phi_\ell^\star - \psi_\ell|||^2 = |||\phi^\star - \psi_\ell|||^2 \quad \text{for all } \psi_\ell \in \mathcal{P}^0(\mathcal{T}_\ell). \tag{6.79}$$

Combining (6.77)–(6.79) and applying Lemma 73(iii), we see that

$$\begin{aligned} \Delta_\ell^{k+1} &= \mu\,\eta_\ell(\phi_\ell^{k+1})^2 + |||\phi_\ell^\star - \phi_\ell^{k+1}|||^2 + |||\phi^\star - \phi_\ell^\star|||^2 \\ &< (1-\varepsilon)\,|||\phi_\ell^\star - \phi_\ell^k|||^2 + \varepsilon\,|||\phi_\ell^\star - \phi_\ell^k|||^2 \\ &\qquad + (\mu\,\lambda_{\mathrm{ctr}}^{-2} - 1)|||\phi_\ell^{k+1} - \phi_\ell^k|||^2 + |||\phi^\star - \phi_\ell^\star|||^2 \\ &\overset{73(\mathrm{iii})}{\le} (1-\varepsilon)\,|||\phi_\ell^\star - \phi_\ell^k|||^2 \\ &\qquad + \big(\varepsilon\,(1-q_{\mathrm{pcg}})^{-2} + \mu\,\lambda_{\mathrm{ctr}}^{-2} - 1\big)|||\phi_\ell^{k+1} - \phi_\ell^k|||^2 + |||\phi^\star - \phi_\ell^\star|||^2. \end{aligned}$$

Step 2 further yields that

$$\begin{aligned}\Delta_\ell^{k+1} &\le (1-\varepsilon)\big(|||\phi_\ell^\star - \phi_\ell^k|||^2 + |||\phi^\star - \phi_\ell^\star|||^2\big) + 2\,\varepsilon\,C_{\mathrm{rel}}^2\,\eta_\ell(\phi_\ell^k)^2 \\ &\quad + \big(\varepsilon\,(1-q_{\mathrm{pcg}})^{-2} + \mu\,\lambda_{\mathrm{ctr}}^{-2} - 1 + 2\,\varepsilon\,C_{\mathrm{rel}}^2\,C_{\mathrm{stb}}^2\,(1-q_{\mathrm{pcg}})^{-2}\big)\,|||\phi_\ell^{k+1} - \phi_\ell^k|||^2.\end{aligned}$$

Using (6.73)–(6.74), (6.79), and (6.76), we thus see that

$$\begin{aligned}\Delta_\ell^{k+1} &\overset{(6.73)-(6.74)}{\le} (1-\varepsilon)\big(|||\phi_\ell^\star - \phi_\ell^k|||^2 + |||\phi^\star - \phi_\ell^\star|||^2\big) + (1-\varepsilon)\,\mu\,\eta_\ell(\phi_\ell^k)^2 \\ &\overset{(6.79)}{\le} (1-\varepsilon)\big(\mu\,\eta_\ell(\phi_\ell^k)^2 + |||\phi^\star - \phi_\ell^k|||^2\big) \\ &\overset{(6.76)}{\le} q_{\mathrm{ctr}}\,\Delta_\ell^k,\end{aligned}$$

if $k+1 < \underline{k}(\ell)$. This concludes the proof of (6.68).

Step 4. We use the definition $\phi_{\ell+1}^0 := \phi_\ell^k$ from Step (iv) of Algorithm 57 to see that

$$\begin{aligned}\Delta_{\ell+1}^0 &= \mu\,\eta_{\ell+1}(\phi_{\ell+1}^0)^2 + |||\phi^\star - \phi_{\ell+1}^0|||^2 \\ &= \mu\,\eta_{\ell+1}(\phi_\ell^k)^2 + |||\phi^\star - \phi_\ell^k|||^2.\end{aligned} \tag{6.80}$$

For the first summand of (6.80), we use stability (A1) and reduction (A2). Together with the Dörfler marking strategy in Step (iii) of Algorithm 57 and $\mathcal{M}_\ell \subseteq \mathcal{T}_\ell \backslash \mathcal{T}_{\ell+1}$, we see that

$$\begin{aligned}\eta_{\ell+1}(\phi_\ell^k)^2 &= \eta_{\ell+1}(\mathcal{T}_{\ell+1}\backslash\mathcal{T}_\ell, \phi_\ell^k)^2 + \eta_{\ell+1}(\mathcal{T}_{\ell+1}\cap\mathcal{T}_\ell, \phi_\ell^k)^2 \\ &\overset{(A1)-(A2)}{\le} q_{\mathrm{red}}\,\eta_\ell(\mathcal{T}_\ell\backslash\mathcal{T}_{\ell+1}, \phi_\ell^k)^2 + \eta_\ell(\mathcal{T}_{\ell+1}\cap\mathcal{T}_\ell, \phi_\ell^k)^2 \\ &= \eta_\ell(\phi_\ell^k)^2 - (1-q_{\mathrm{red}})\,\eta_\ell(\mathcal{T}_\ell\backslash\mathcal{T}_{\ell+1}, \phi_\ell^k)^2 \\ &\overset{(6.31)}{\le} \eta_\ell(\phi_\ell^k)^2 - (1-q_{\mathrm{red}})\theta^2\,\eta_\ell(\phi_\ell^k)^2 \\ &\overset{(6.71)}{=} q_{\mathrm{est}}\,\eta_\ell(\phi_\ell^k)^2.\end{aligned} \tag{6.81}$$

With this and stability (A1), the Young inequality and Lemma 73(ii) yield that

$$\begin{aligned}\eta_{\ell+1}(\phi_\ell^k)^2 &\overset{(6.81)}{\le} q_{\mathrm{est}}\,\eta_\ell(\phi_\ell^k)^2 \\ &\overset{(A1)}{\le} (1+\gamma)\,q_{\mathrm{est}}\,\eta_\ell(\phi_\ell^{k-1})^2 + (1+\gamma^{-1})\,q_{\mathrm{est}}\,C_{\mathrm{stb}}^2\,|||\phi_\ell^k - \phi_\ell^{k-1}|||^2 \\ &\overset{73(\mathrm{ii})}{\le} (1+\gamma)\,q_{\mathrm{est}}\,\eta_\ell(\phi_\ell^{k-1})^2 \\ &\qquad + (1+\gamma^{-1})\,q_{\mathrm{est}}\,C_{\mathrm{stb}}^2\,(1+q_{\mathrm{pcg}})^2\,|||\phi_\ell^\star - \phi_\ell^{k-1}|||^2.\end{aligned} \tag{6.82}$$

For the second summand of (6.80), we apply the Pythagoras identity (6.79) together with Lemma 73(i) and obtain that

$$\begin{aligned}|||\phi^\star - \phi_\ell^k|||^2 &\overset{(6.79)}{=} |||\phi^\star - \phi_\ell^\star|||^2 + |||\phi_\ell^\star - \phi_\ell^k|||^2 \\ &\overset{73(\mathrm{i})}{\le} |||\phi^\star - \phi_\ell^\star|||^2 + q_{\mathrm{pcg}}^2\,|||\phi_\ell^\star - \phi_\ell^{k-1}|||^2.\end{aligned} \tag{6.83}$$

Combining (6.80)–(6.83), we end up with

$$\begin{aligned}\Delta_{\ell+1}^0 &= \mu\,\eta_{\ell+1}(\phi_\ell^k)^2 + |||\phi^\star - \phi_\ell^k|||^2\\ &\le \mu\,(1+\gamma)\,q_{\mathrm{est}}\,\eta_\ell(\phi_\ell^{k-1})^2\\ &\quad + \big(\mu\,(1+\gamma^{-1})\,q_{\mathrm{est}}\,C_{\mathrm{stb}}^2\,(1+q_{\mathrm{pcg}})^2 + q_{\mathrm{pcg}}^2\big)|||\phi_\ell^\star - \phi_\ell^{k-1}|||^2 + |||\phi^\star - \phi_\ell^\star|||^2.\end{aligned}$$

Using the same arguments as in Step 2, we get that

$$\begin{aligned}\Delta_{\ell+1}^0 &\le \mu\,(1+\gamma)\,q_{\mathrm{est}}\,\eta_\ell(\phi_\ell^{k-1})^2\\ &\quad + \big(\mu\,(1+\gamma^{-1})\,q_{\mathrm{est}}\,C_{\mathrm{stb}}^2\,(1+q_{\mathrm{pcg}})^2 + q_{\mathrm{pcg}}^2\big)|||\phi_\ell^\star - \phi_\ell^{k-1}|||^2\\ &\quad + (1-\kappa)|||\phi^\star - \phi_\ell^\star|||^2 + 2\,\kappa\,C_{\mathrm{rel}}^2\,\eta_\ell(\phi_\ell^{k-1})^2 + 2\,\kappa\,C_{\mathrm{rel}}^2\,C_{\mathrm{stb}}^2\,|||\phi_\ell^\star - \phi_\ell^{k-1}|||^2\\ &= \big(\mu\,(1+\gamma)\,q_{\mathrm{est}} + 2\,\kappa\,C_{\mathrm{rel}}^2\big)\,\eta_\ell(\phi_\ell^{k-1})^2 + (1-\kappa)|||\phi^\star - \phi_\ell^\star|||^2\\ &\quad + \big(\mu\,(1+\gamma^{-1})\,q_{\mathrm{est}}\,C_{\mathrm{stb}}^2\,(1+q_{\mathrm{pcg}})^2 + q_{\mathrm{pcg}}^2 + 2\,\kappa\,C_{\mathrm{rel}}^2 C_{\mathrm{stb}}^2\big)|||\phi_\ell^\star - \phi_\ell^{k-1}|||^2\\ &\overset{(6.76)}{\le} q_{\mathrm{ctr}}\,\mu\,\eta_j(\phi_\ell^{k-1})^2 + q_{\mathrm{ctr}}\,|||\phi^\star - \phi_\ell^\star|||^2 + q_{\mathrm{ctr}}\,|||\phi_\ell^\star - \phi_\ell^{k-1}|||^2\\ &\overset{(6.79)}{=} q_{\mathrm{ctr}}\,\Delta_\ell^{k-1}.\end{aligned}$$

This concludes the proof of (6.69).

Step 5. Inequality (6.70) follows by induction. This concludes the proof. □

***Proof of Theorem 68*(b).** The proof is split into three steps.

Step 1. Let $\ell \in \mathbb{N}$. Recall the Pythagoras identity (6.79). We use stability (A1) and the stopping criterion (6.30) of Algorithm 57 to see that

$$\begin{aligned}\Delta_\ell^{k-1} &\overset{(6.79)}{=} \mu\,\eta_\ell(\phi_\ell^{k-1})^2 + |||\phi_\ell^\star - \phi_\ell^{k-1}|||^2 + |||\phi^\star - \phi_\ell^\star|||^2\\ &\overset{(A1)}{\lesssim} \eta_\ell(\phi_\ell^k)^2 + |||\phi_\ell^k - \phi_\ell^{k-1}|||^2 + |||\phi_\ell^\star - \phi_\ell^k|||^2 + |||\phi^\star - \phi_\ell^\star|||^2\\ &\overset{(6.30)}{\lesssim} \eta_\ell(\phi_\ell^k)^2 + |||\phi_\ell^\star - \phi_\ell^k|||^2 + |||\phi^\star - \phi_\ell^\star|||^2\\ &\overset{(6.79)}{\simeq} \Delta_\ell^k.\end{aligned}$$

With the Pythagoras identity (6.67), we argue similarly to obtain that

$$\begin{aligned}\Delta_\ell^k &\overset{(6.79)}{=} \mu\,\eta_\ell(\phi_\ell^k)^2 + |||\phi_\ell^\star - \phi_\ell^k|||^2 + |||\phi^\star - \phi_\ell^\star|||^2\\ &\overset{(A1)}{\lesssim} \eta_\ell(\phi_\ell^{k-1})^2 + |||\phi_\ell^k - \phi_\ell^{k-1}|||^2 + |||\phi_\ell^\star - \phi_\ell^k|||^2 + |||\phi^\star - \phi_\ell^\star|||^2\\ &\overset{(6.67)}{=} \eta_\ell(\phi_\ell^{k-1})^2 + |||\phi_\ell^\star - \phi_\ell^{k-1}|||^2 + |||\phi^\star - \phi_\ell^\star|||^2\\ &\overset{(6.79)}{\simeq} \Delta_\ell^{k-1}.\end{aligned}$$

Hence, it follows that $\Delta_\ell^k \simeq \Delta_\ell^{k-1}$.

Step 2. For $0 \le \ell \le \ell'$, define $\widehat{k}(\ell) := \widehat{k} \in \mathbb{N}_0$ by

$$\widehat{k} := \begin{cases} \underline{k}(\ell) & \text{if } \ell < \ell', \\ k' & \text{if } \ell = \ell'. \end{cases}$$

From Step 1, Lemma 74, and the geometric series (for the sum over k), it follows that

$$\begin{aligned} \sum_{\ell=0}^{\ell'} \sum_{k=0}^{\widehat{k}(\ell)} (\Delta_\ell^k)^{-1} &\lesssim (\Delta_{\ell'}^{k'})^{-1} + \sum_{\ell=0}^{\ell'} \sum_{k=0}^{\widehat{k}(\ell)-1} (\Delta_\ell^k)^{-1} \\ &\overset{(6.70)}{\le} (\Delta_{\ell'}^{k'})^{-1} + \sum_{\ell=0}^{\ell'} \sum_{k=0}^{\widehat{k}(\ell)-1} q_{\mathrm{ctr}}^{|(\ell,\widehat{k}-1)|-|(\ell,k)|} (\Delta_\ell^{\widehat{k}-1})^{-1} \\ &\lesssim (\Delta_{\ell'}^{k'})^{-1} + \sum_{\ell=0}^{\ell'} (\Delta_\ell^{\widehat{k}-1})^{-1}. \end{aligned}$$

For $k' < \underline{k}(\ell')$, inequality (6.70) and the geometric series (for the sum over ℓ) yield that

$$\sum_{\ell=0}^{\ell'} (\Delta_\ell^{\widehat{k}-1})^{-1} \overset{(6.70)}{\lesssim} \sum_{\ell=0}^{\ell'} q_{\mathrm{ctr}}^{|(\ell',k')|-|(\ell,\widehat{k}-1)|} (\Delta_{\ell'}^{k'})^{-1} \lesssim (\Delta_{\ell'}^{k'})^{-1}.$$

For $k' = \underline{k}(\ell')$, inequality (6.70), the geometric series, and Step 1 yield that

$$\begin{aligned} \sum_{\ell=0}^{\ell'} (\Delta_\ell^{\widehat{k}-1})^{-1} &= (\Delta_{\ell'}^{\underline{k}-1})^{-1} + \sum_{\ell=0}^{\ell'-1} (\Delta_\ell^{\underline{k}-1})^{-1} \\ &\overset{(6.70)}{\lesssim} \Big(1 + \sum_{\ell=0}^{\ell'-1} q_{\mathrm{ctr}}^{|(\ell',\underline{k}-1)|-|(\ell,\underline{k}-1)|}\Big) (\Delta_{\ell'}^{\underline{k}-1})^{-1} \\ &\lesssim (\Delta_{\ell'}^{\underline{k}-1})^{-1} \\ &\simeq (\Delta_{\ell'}^{\underline{k}})^{-1} \\ &= (\Delta_{\ell'}^{k'})^{-1}. \end{aligned}$$

Overall, it follows that

$$\sum_{\ell=0}^{\ell'} \sum_{k=0}^{\widehat{k}(\ell)} (\Delta_\ell^k)^{-1} \lesssim (\Delta_{\ell'}^{k'})^{-1} \quad \text{for all } (\ell', k') \in \mathcal{Q}. \tag{6.84}$$

Step 3. For the convenience of the reader, we recall an argument from the proof of [CFPP14, Lemma 4.9]: Let $s > 0$. Let $C > 0$ and $\alpha_n \ge 0$ satisfy that

$$\sum_{n=0}^{N-1} \alpha_n^{-1/s} \le C\, \alpha_N^{-1/s} \qquad \text{for all } N \in \mathbb{N}.$$

Then, it holds that

$$(1+C^{-1})\sum_{n=0}^{N-1}\alpha_n^{-1/s}\le\sum_{n=0}^{N-1}\alpha_n^{-1/s}+\alpha_N=\sum_{n=0}^{N}\alpha_n^{-1/s}\quad\text{for all } N\in\mathbb{N}.$$

Inductively, it follows that

$$(1+C^{-1})^m\sum_{n=0}^{N}\alpha_n^{-1/s}\le\sum_{n=0}^{N+m}\alpha_n^{-1/s}\quad\text{for all } N,m\in\mathbb{N}_0.$$

This implies that

$$\begin{aligned}\alpha_N^{-1/s}&\le\sum_{n=0}^{N}\alpha_n^{-1/s}\\&\le(1+C^{-1})^{-m}\sum_{n=0}^{N+m}\alpha_n^{-1/s}\\&\le(1+C)\,(1+C^{-1})^{-m}\,\alpha_{N+m}^{-1/s}\end{aligned}$$

for all $N,m\in\mathbb{N}_0$. This is equivalent to

$$\alpha_{N+m}^{1/s}\le(1+C)\,(1+C^{-1})^{-m}\,\alpha_N^{1/s}.$$

Step 4. Since the index set $\mathcal{Q}$ is linearly ordered with respect to the total step counter $|(\cdot,\cdot)|$, Step 2 and Step 3 with $s=2$ imply the existence of $0<q_{\mathrm{lin}}<1$ such that

$$(\Delta_{\ell'}^{k'})^{1/2}\lesssim q_{\mathrm{lin}}^{|(\ell',k')|-|(\ell,k)|}\,(\Delta_\ell^k)^{1/2}\tag{6.85}$$

for all $(\ell,k),(\ell',k')\in\mathcal{Q}$ with $(\ell',k')>(\ell,k)$. Clearly, it holds that $\Lambda_\ell^k\simeq(\Delta_\ell^k)^{1/2}$ for all $(\ell,k)\in\mathcal{Q}$. This and (6.85) conclude the proof. □

Proof of Theorem 68(c)

As in Chapter 4, the proof of optimal convergence rates requires the assumptions (R1)–(R3) on the mesh-refinement strategy. For 3D BEM (with 2D NVB from Section 3.6) and 2D BEM (with extended 1D bisection from Section 3.5) these properties are fulfilled, cf. Section 3.5 and Section 3.6 respectively.

Recall the constants $C_{\mathrm{stab}}>0$ from (A1) and $C_{\mathrm{drl}}>0$ from (A4). Suppose that $0<\theta\le1$ and $\lambda_{\mathrm{ctr}}>0$ are sufficiently small such that

$$0<\theta'':=\frac{\theta+\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}}{1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}}<\theta_{\mathrm{opt}}:=\big(1+C_{\mathrm{stb}}^2\,C_{\mathrm{drl}}^2\big)^{-1/2},\tag{6.86}$$

where

$$\lambda_{\mathrm{opt}}:=\Big(C_{\mathrm{stab}}\,\frac{q_{\mathrm{pcg}}}{1-q_{\mathrm{pcg}}}\Big)^{-1}.$$

In particular, it holds that $0 < \theta < \theta_{\mathrm{opt}}$ and $0 < \lambda_{\mathrm{ctr}} < \lambda_{\mathrm{opt}}$. We need the following comparison lemma which is found in [CFPP14, Lemma 4.14].

Lemma 75. *Suppose (R2), (A1), (A2), and (A4). Recall the assumption (6.86). There exist constants $C_1, C_2 > 0$ such that for all $s > 0$ with $\|\phi^\star\|_{\mathbb{A}_s} < \infty$ and all $\ell \in \mathbb{N}_0$, there exists $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$ which satisfies*

$$\#\mathcal{R}_\ell \le C_1 C_2^{-1/s} \, \|\phi^\star\|_{\mathbb{A}_s}^{1/s} \eta_\ell(\phi_\ell^\star)^{-1/s}, \tag{6.87}$$

as well as the Dörfler marking criterion

$$\theta'' \eta_\ell(\phi_\ell^\star) \le \eta_\ell(\mathcal{R}_\ell, \phi_\ell^\star). \tag{6.88}$$

The constants C_1, C_2 depend only on the constants of (A1), (A2), and (A4). □

Another lemma, which we need for the proof of Theorem 68(c), shows that the iterates ϕ_ℓ^k of Algorithm 57 are close to the exact Galerkin approximation $\phi_\ell^\star \in \mathcal{P}^0(\mathcal{T}_\ell)$.

Lemma 76. *Let $0 < \lambda_{\mathrm{ctr}} < \lambda_{\mathrm{opt}}$. For all $\ell \in \mathbb{N}_0$, it holds that*

$$|||\phi_\ell^\star - \phi_\ell^k||| \le \lambda_{\mathrm{ctr}} \frac{q_{\mathrm{pcg}}}{1 - q_{\mathrm{pcg}}} \min\Big\{\eta_\ell(\phi_\ell^k), \frac{1}{1 - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}} \eta_\ell(\phi_\ell^\star)\Big\}. \tag{6.89}$$

Moreover, there holds equivalence

$$(1 - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}) \, \eta_\ell(\phi_\ell^k) \le \eta_\ell(\phi_\ell^\star) \le (1 + \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}) \, \eta_\ell(\phi_\ell^k). \tag{6.90}$$

Proof. Stability (A1) yields that

$$|\eta_\ell(\phi_\ell^\star) - \eta_\ell(\phi_\ell^k)| \le C_{\mathrm{stab}} \, |||\phi_\ell^\star - \phi_\ell^k|||.$$

Therefore, Lemma 73(iv) and the stopping criterion (6.30) of Algorithm 57 imply that

$$\begin{aligned} |||\phi_\ell^\star - \phi_\ell^k||| &\overset{73(\mathrm{iv})}{\le} \frac{q_{\mathrm{pcg}}}{1 - q_{\mathrm{pcg}}} |||\phi_\ell^k - \phi_\ell^{k-1}||| \\ &\overset{(6.30)}{\le} \lambda_{\mathrm{ctr}} \frac{q_{\mathrm{pcg}}}{1 - q_{\mathrm{pcg}}} \eta_\ell(\phi_\ell^k) \\ &\overset{(A1)}{\le} \lambda_{\mathrm{ctr}} \frac{q_{\mathrm{pcg}}}{1 - q_{\mathrm{pcg}}} \big(\eta_\ell(\phi_\ell^\star) + C_{\mathrm{stab}} \, |||\phi_\ell^\star - \phi_\ell^k|||\big). \end{aligned}$$

Since $0 < \lambda_{\mathrm{ctr}} < \lambda_{\mathrm{opt}}$ and hence

$$\lambda_{\mathrm{ctr}} \, C_{\mathrm{stab}} \frac{q_{\mathrm{pcg}}}{1 - q_{\mathrm{pcg}}} = \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}} < 1,$$

this yields that

$$\begin{aligned} |||\phi_\ell^\star - \phi_\ell^k||| &\le \frac{\lambda_{\mathrm{ctr}} \frac{q_{\mathrm{pcg}}}{1 - q_{\mathrm{pcg}}}}{1 - \lambda_{\mathrm{ctr}} \, C_{\mathrm{stab}} \frac{q_{\mathrm{pcg}}}{1 - q_{\mathrm{pcg}}}} \eta_\ell(\phi_\ell^\star) \\ &= \lambda_{\mathrm{ctr}} \frac{q_{\mathrm{pcg}}}{1 - q_{\mathrm{pcg}}} \frac{1}{1 - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}} \eta_\ell(\phi_\ell^\star). \end{aligned}$$

Altogether, this proves (6.89). Moreover, with stability (A1), we see that

$$\begin{aligned}\eta_\ell(\phi_\ell^\star) &\overset{\text{(A1)}}{\leq} \eta_\ell(\phi_\ell^k) + C_{\mathrm{stab}}\,|||\phi_\ell^\star - \phi_\ell^k|||\\ &\overset{(6.89)}{\leq} (1+\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}})\,\eta_\ell(\phi_\ell^k)\end{aligned}$$

as well as

$$\begin{aligned}\eta_\ell(\phi_\ell^k) &\overset{\text{(A1)}}{\leq} \eta_\ell(\phi_\ell^\star) + C_{\mathrm{stab}}\,|||\phi_\ell^\star - \phi_\ell^k|||\\ &\overset{(6.89)}{\leq} \Big(1 + \frac{\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}}{1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}}\Big)\,\eta_\ell(\phi_\ell^\star)\\ &= \frac{1}{1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}}\,\eta_\ell(\phi_\ell^\star).\end{aligned}$$

This concludes the proof. □

The following lemma immediately shows "⟸" in (68).

Lemma 77. *Suppose (R1). For $\ell \in \mathbb{N}_0$, let $\widehat{\mathcal{T}}_{\ell+1} = \texttt{refine}(\widehat{\mathcal{T}}_\ell, \widehat{\mathcal{M}}_\ell)$ with arbitrary, but non-empty $\widehat{\mathcal{M}}_\ell \subseteq \widehat{\mathcal{T}}_\ell$ and $\widehat{\mathcal{T}}_0 = \mathcal{T}_0$. Let $\widehat{\mathcal{Q}} \subseteq \mathbb{N}_0 \times \mathbb{N}_0$ be an index set and $\widehat{\phi}_\ell^k \in \mathcal{P}^0(\widehat{\mathcal{T}}_\ell)$ for all $(\ell,k) \in \widehat{\mathcal{Q}}$. Let $s > 0$ and suppose that the corresponding quasi-errors $\widehat{\Lambda}_\ell^k := \big(|||\phi^\star - \widehat{\phi}_\ell^k|||^2 + \widehat{\eta}_\ell(\widehat{\phi}_\ell^\star)^2\big)^{1/2}$ satisfy that*

$$\sup_{(\ell,k)\in\widehat{\mathcal{Q}}} \big(\#\widehat{\mathcal{T}}_\ell - \#\mathcal{T}_0 + 1\big)^s\,\widehat{\Lambda}_\ell^k < \infty. \tag{6.91}$$

Then, it follows that $\|\phi^\star\|_{\mathbb{A}_s} < \infty$.

Proof. Due to the Pythagoras identity (6.79) and stability (A1), it holds that

$$\begin{aligned}(\widehat{\Lambda}_\ell^k)^2 &= |||\phi^\star - \widehat{\phi}_\ell^k|||^2 + \widehat{\eta}_\ell(\widehat{\phi}_\ell^k)^2\\ &\overset{(6.79)}{=} |||\phi^\star - \widehat{\phi}_\ell^\star|||^2 + |||\widehat{\phi}_\ell^\star - \widehat{\phi}_\ell^k|||^2 + \widehat{\eta}_\ell(\widehat{\phi}_\ell^k)^2\\ &\overset{\text{(A1)}}{\gtrsim} \widehat{\eta}_\ell(\widehat{\phi}_\ell^\star)^2.\end{aligned} \tag{6.92}$$

Additionally, [BHP17, Lemma 22] shows that

$$\#\mathcal{T}_\bullet - \#\mathcal{T}_0 + 1 \leq \#\mathcal{T}_\bullet \leq \#\mathcal{T}_0\,\big(\#\mathcal{T}_\bullet - \#\mathcal{T}_0 + 1\big) \quad \text{for all } \mathcal{T}_\bullet \in \mathbb{T}. \tag{6.93}$$

Given $N \in \mathbb{N}_0$, there exists an index $\ell \in \mathbb{N}_0$ such that

$$\begin{aligned}\#\widehat{\mathcal{T}}_\ell - \#\mathcal{T}_0 \leq N < N+1 &\leq \#\widehat{\mathcal{T}}_{\ell+1} - \#\mathcal{T}_0 + 1\\ &\overset{(6.93)}{\leq} \#\widehat{\mathcal{T}}_{\ell+1} \overset{\text{(R1)}}{\lesssim} \#\widehat{\mathcal{T}}_\ell \overset{(6.93)}{\lesssim} \#\widehat{\mathcal{T}}_\ell - \#\mathcal{T}_0 + 1.\end{aligned} \tag{6.94}$$

With (6.92)–(6.94), it follows that

$$\begin{aligned}(N+1)^s \min_{\substack{\mathcal{T}_\bullet \in \mathtt{refine}(\mathcal{T}_0)\\ \#\mathcal{T}_\bullet - \#\mathcal{T}_0 \le N}} \eta_\bullet(\phi_\bullet^\star) &\overset{(6.94)}{\lesssim} \big(\#\widehat{\mathcal{T}}_\ell - \#\mathcal{T}_0 + 1\big)^s\, \widehat{\eta}_\ell(\widehat{\phi}_\ell^\star)\\ &\overset{(6.92)}{\lesssim} \sup_{(\ell,k)\in\widehat{\mathcal{Q}}} \big(\#\widehat{\mathcal{T}}_\ell - \#\mathcal{T}_0 + 1\big)^s\, \widehat{\Lambda}_\ell^k\\ &\overset{(6.91)}{<} \infty.\end{aligned}$$

Since the upper bound is finite and independent of N, this implies that $\|\phi^\star\|_{\mathbb{A}_s} < \infty$. □

***Proof of Theorem 68*(c).** With Lemma 77, it only remains to prove the implication "$\Longrightarrow$" in (68). The proof is split into three steps, where we suppose that $\|\phi^\star\|_{\mathbb{A}_s} < \infty$.

Step 1. By Assumption (6.86), Lemma 75 provides a set $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$ with (6.87)–(6.88). Due to stability (A1) and $\lambda_{\mathrm{opt}}^{-1} = C_{\mathrm{stb}}\, \frac{q_{\mathrm{pcg}}}{1-q_{\mathrm{pcg}}}$, it holds that

$$\begin{aligned}\eta_\ell(\mathcal{R}_\ell, \phi_\ell^\star) &\overset{(A1)}{\le} \eta_\ell(\mathcal{R}_\ell, \phi_\ell^k) + C_{\mathrm{stb}}\, |||\phi_\ell^\star - \phi_\ell^k|||\\ &\overset{(6.89)}{\le} \eta_\ell(\mathcal{R}_\ell, \phi_\ell^k) + \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}\, \eta_\ell(\phi_\ell^k).\end{aligned}$$

Together with $\theta''\eta_\ell(\phi_\ell^\star) \le \eta_\ell(\mathcal{R}_\ell, \phi_j^\star)$, this proves that

$$\begin{aligned}(1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}})\, \theta''\, \eta_\ell(\phi_\ell^k) &\overset{(6.90)}{\le} \theta''\, \eta_\ell(\phi_\ell^\star)\\ &\le \eta_\ell(\mathcal{R}_\ell, \phi_\ell^\star)\\ &\le \eta_\ell(\mathcal{R}_\ell, \phi_\ell^k) + \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}\, \eta_\ell(\phi_\ell^k)\end{aligned}$$

and results in

$$\theta\, \eta_\ell(\phi_\ell^k) \overset{(6.86)}{=} \Big((1-\lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}})\theta'' - \lambda_{\mathrm{ctr}}/\lambda_{\mathrm{opt}}\Big)\, \eta_\ell(\phi_\ell^k) \le \eta_\ell(\mathcal{R}_\ell, \phi_\ell^k). \tag{6.95}$$

Hence, $\mathcal{R}_\ell$ satisfies the Dörfler marking for ϕ_ℓ^k with parameter θ. By choice of $\mathcal{M}_\ell$ in Step (iii) of Algorithm 57, we thus infer that

$$\#\mathcal{M}_\ell \overset{(6.95)}{\lesssim} \#\mathcal{R}_\ell \overset{(6.87)}{\lesssim} \eta_\ell(\phi_\ell^\star)^{-1/s} \overset{(6.90)}{\simeq} \eta_\ell(\phi_\ell^k)^{-1/s} \quad \text{for all } \ell \in \mathbb{N}_0.$$

The mesh-closure estimate (R3) guarantees that

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 \overset{(R3)}{\lesssim} \sum_{j=0}^{\ell-1} \#\mathcal{M}_j \lesssim \sum_{j=0}^{\ell-1} \eta_j(\phi_j^k)^{-1/s} \quad \text{for all } \ell > 0. \tag{6.96}$$

Step 2. For $\ell = 0$, it holds that $1 \lesssim (\Lambda_0^k)^{-1/s}$. For $\ell > 0$, we proceed as follows: Remark 69 yields that $\eta_j(\phi_j^k) \simeq \Lambda_j^k$. Theorem 68(b) and the geometric series prove that

$$\begin{aligned}\sum_{j=0}^{\ell-1} \eta_j(\phi_j^k)^{-1/s} &\simeq \sum_{j=0}^{\ell-1} (\Lambda_j^k)^{-1/s} \\ &\overset{(6.64)}{\lesssim} \sum_{j=0}^{\ell-1} (q_{\mathrm{lin}}^{1/s})^{|(\ell,\underline{k})-(j,\underline{k})|}\, (\Lambda_\ell^k)^{-1/s} \\ &\lesssim (\Lambda_\ell^k)^{-1/s}.\end{aligned}$$

Combining this with (6.96) and including the estimate for $\ell = 0$, we derive that

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 \lesssim (\Lambda_\ell^k)^{-1/s} \quad \text{for all } \ell \in \mathbb{N}_0. \tag{6.97}$$

Step 3. Arguing as in (6.94) and employing Theorem 68(b), we see that

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 \overset{(6.94)}{\simeq} \#\mathcal{T}_{\ell-1} - \#\mathcal{T}_0 + 1 \overset{(6.97)}{\lesssim} (\Lambda_{\ell-1}^k)^{-1/s} \overset{(6.64)}{\lesssim} (\Lambda_\ell^k)^{-1/s}$$

for all $(\ell, k) \in \mathcal{Q}$ with $\ell > 0$. Since $\underline{k}(0) \le \#\mathcal{T}_0 < \infty$, we hence conclude that

$$\sup_{(\ell,k)\in\mathcal{Q}} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s\, \Lambda_\ell^k < \infty.$$

This concludes the proof of Theorem 68. □

6.5.5 Almost optimal computational complexity

In order to get an efficient implementation, we suppose that we use $\mathcal{H}^2$-matrices for the efficient treatment of the discrete single-layer integral operator. Then, the storage requirements (and the cost for one matrix-vector multiplication respectively) of an $\mathcal{H}^2$-matrix are of order $\mathcal{O}(Np^2)$, where N is the matrix size and $p \in \mathbb{N}$ is the local block rank. For $\mathcal{H}^2$-matrices (unlike $\mathcal{H}$-matrices), these costs are, in particular, independent of a possibly unbalanced binary tree which underlies the hierarchical data structure [Hac15].

For a mesh $\mathcal{T}_\bullet \in \mathbb{T}$, we employ the local block rank $p = \mathcal{O}(\log(1 + \#\mathcal{T}_\bullet))$ to ensure that the matrix compression is asymptotically exact as $N = \#\mathcal{T}_\bullet \to \infty$, i.e., the error between the exact matrix and the $\mathcal{H}$-matrix decays exponentially fast, cf. [Hac15]. We stress that we neglect this error in the following and assume that the matrix-vector multiplication (based on the $\mathcal{H}^2$-matrix) yields the exact matrix-vector product.

The computational cost for storing $\mathbf{A}_\bullet$ (as well as for one matrix-vector multiplication) is $\mathcal{O}((\#\mathcal{T}_\bullet)\log^2(1+\#\mathcal{T}_\bullet))$. In an idealized optimal case, the computation of $\phi_\bullet^\star$ is hence (at least) of cost $\mathcal{O}((\#\mathcal{T}_\bullet)\log^2(1+\#\mathcal{T}_\bullet))$.

We consider the computational cost for one step of Algorithm 57:

- We assume that one step of the PCG algorithm with the employed optimal preconditioner is of cost $\mathcal{O}\big((\#\mathcal{T}_\ell)\log^2(1+\#\mathcal{T}_\ell)\big)$, since the evaluation of one matrix-vector multiplication with the preconditioner $\mathbf{P}_L$ can be done in $\mathcal{O}\big(\#\mathcal{T}_L\big)$, cf. Section 6.5.1.

- We assume that we can compute $\eta_\ell(\psi_\ell)$ for any $\psi_\ell \in \mathcal{P}^0(\mathcal{T}_\ell)$ (by means of numerical quadrature) with $\mathcal{O}\big((\#\mathcal{T}_j)\log^2(1+\#\mathcal{T}_j)\big)$ operations.
- Clearly, the Dörfler marking in Step (iii) can be done in $\mathcal{O}\big((\#\mathcal{T}_j)\log(1+\#\mathcal{T}_j)\big)$ operations by sorting. Moreover, for $C_{\mathrm{mark}} = 2$, Stevenson [Ste07] proposed a realization of the Dörfler marking based on binning, which can be performed at linear cost $\mathcal{O}(\#\mathcal{T}_j)$.
- Finally, the mesh-refinement in Step (iv) can be done in linear complexity $\mathcal{O}(\#\mathcal{T}_j)$ if the data structure is appropriate.

Overall, one step of Algorithm 57 is thus done in $\mathcal{O}((\#\mathcal{T}_\ell)\log^2(1+\#\mathcal{T}_j))$ operations. However, an adaptive step $(\ell',k') \in \mathcal{Q}$ depends on the full history of previous steps.

- Hence, the cumulative computational complexity for the adaptive step $(\ell',k') \in \mathcal{Q}$ is of order

$$\mathcal{O}\Big(\sum_{(\ell,k)\le(\ell',k')} (\#\mathcal{T}_\ell)\log^2(1+\#\mathcal{T}_\ell)\Big).$$

The following corollary proves that Algorithm 57 does not only lead to convergence of the quasi-error Λ_ℓ^k with optimal rate with respect to the degrees of freedom (see Theorem 68), but also with *almost* optimal rate with respect to the computational costs.

Corollary 78. *For $\ell \in \mathbb{N}_0$, let $\widehat{\mathcal{T}}_{\ell+1} = \mathtt{refine}(\widehat{\mathcal{T}}_\ell, \widehat{\mathcal{M}}_\ell)$ with arbitrary $\widehat{\mathcal{M}}_\ell \subseteq \widehat{\mathcal{T}}_\ell$ and $\widehat{\mathcal{T}}_0 = \mathcal{T}_0$. Let $s > 0$ and suppose that the corresponding error estimator $\widehat{\eta}_\ell(\widehat{\phi}_\ell^\star)$ converges at rate s with respect to the single-step computational cost, i.e.,*

$$\sup_{\ell\in\mathbb{N}_0} \big[(\#\widehat{\mathcal{T}}_\ell)\,\log^2(1+\#\widehat{\mathcal{T}}_\ell)\big]^s\,\widehat{\eta}_\ell(\widehat{\phi}_\ell^\star) < \infty. \tag{6.98}$$

Suppose that λ_{ctr} and θ satisfy the assumptions of Theorem 68(c). Then, the quasi-errors Λ_ℓ^k generated by Algorithm 57 converge almost at rate s with respect to the cumulative computational cost, i.e.,

$$\sup_{(\ell',k')\in\mathcal{Q}} \Big[\sum_{(\ell,k)\le(\ell',k')} (\#\mathcal{T}_\ell)\log^2(1+\#\mathcal{T}_\ell)\Big]^{s-\varepsilon}\,\Lambda_{\ell'}^{k'} < \infty \quad \text{for all } \varepsilon > 0. \tag{6.99}$$

Proof. For all $\delta > 0$, it holds that

$$\#\mathcal{T}_\bullet - \#\mathcal{T}_0 + 1 \overset{(6.93)}{\simeq} \#\mathcal{T}_\bullet \le (\#\mathcal{T}_\bullet)\log^2(1+\#\mathcal{T}_\bullet) \lesssim (\#\mathcal{T}_\bullet)^{1+\delta} \quad \text{for all } \mathcal{T}_\bullet \in \mathbb{T},$$

where the hidden constant depends only on δ. From (6.98), it thus follows that

$$\sup_{\ell\in\mathbb{N}_0} \big[\#\widehat{\mathcal{T}}_\ell - \#\mathcal{T}_0 + 1\big]^s\,\widehat{\eta}_\ell(\widehat{\phi}_\ell^\star) \lesssim \sup_{\ell\in\mathbb{N}_0} \big[(\#\widehat{\mathcal{T}}_\ell)\,\log^2(1+\#\widehat{\mathcal{T}}_\ell)\big]^s\,\widehat{\eta}_\ell(\widehat{\phi}_\ell^\star) < \infty.$$

From Lemma 77, we derive that $\|\phi^\star\|_{\mathbb{A}_s} < \infty$. Hence, Theorem 68(c) yields that

$$\sup_{(\ell,k)\in\mathcal{Q}} \big[\#\mathcal{T}_\ell\big]^s\,\Lambda_\ell^k \simeq \sup_{(\ell,k)\in\mathcal{Q}} \big[\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1\big]^s\,\Lambda_\ell^k < \infty. \tag{6.100}$$

Let $0 < \varepsilon < s$ and choose $\delta > 0$ such that

$$0 < s - \varepsilon = \frac{s}{1+\delta} =: t.$$

This leads to

$$(\#\mathcal{T}_\ell)\log^2(1+\#\mathcal{T}_\ell) \lesssim (\#\mathcal{T}_\ell)^{1+\delta} \overset{(6.100)}{\lesssim} (\Lambda_\ell^k)^{-(1+\delta)/s} = (\Lambda_\ell^k)^{-1/t} \quad \text{for all } (\ell,k) \in \mathcal{Q}.$$

From Theorem 68(b) and the geometric series, it follows that

$$\sum_{(\ell,k)\le(\ell',k')} (\Lambda_\ell^k)^{-1/t} \overset{(6.64)}{\lesssim} \sum_{(\ell,k)\le(\ell',k')} (q_{\mathrm{lin}}^{1/t})^{|(\ell',k')|-|(\ell,k)|}(\Lambda_{\ell'}^{k'})^{-1/t} \lesssim (\Lambda_{\ell'}^{k'})^{-1/t} \quad \text{for all } (\ell',k') \in \mathcal{Q}.$$

Combining the last two estimates, we see that

$$\Big[\sum_{(\ell,k)\le(\ell',k')} (\#\mathcal{T}_\ell)\log^2(1+\#\mathcal{T}_\ell)\Big]^{s-\varepsilon} \lesssim (\Lambda_{\ell'}^{k'})^{-(s-\varepsilon)/t} = (\Lambda_{\ell'}^{k'})^{-1} \text{ for all } (\ell',k') \in \mathcal{Q}.$$

This concludes the proof. □

6.6 Hyper-singular integral equation

In this section, we briefly introduce the setting of the hyper-singular integral equation and show that it fits into our abstract framework and that the main results still hold true.

Given $g\colon \Gamma \to \mathbb{R}$, the hyper-singular integral equation seeks $u^\star : \Gamma \to \mathbb{R}$ such that

$$(Wu^\star)(x) = -\partial_{\mathbf{n}(x)} \int_\Gamma \partial_{\mathbf{n}(y)} G(x-y) u^\star(y)\,\mathrm{d}y = g(x) \quad \text{for all } x \in \Gamma, \tag{6.101}$$

where $\partial_{\mathbf{n}}$ denotes the normal derivative with the outer unit normal vector $\mathbf{n}(\cdot)$ on $\Gamma \subseteq \partial\Omega$.

Recall from Remark 55 that the hyper-singular integral operator

$$W\colon \widetilde{H}^{1/2+s}(\Gamma) \to H^{-1/2+s}(\Gamma)$$

is a bounded linear operator for all $-1/2 \le s \le 1/2$ which is even an isomorphism for $-1/2 < s < 1/2$. For $s = 0$, the operator W is symmetric and positive semi-definite with kernel being the constant functions. Hence, for $\Gamma \subsetneq \partial\Omega$, the operator $W\colon \widetilde{H}^{1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ is an elliptic isomorphism. Moreover, for $\Gamma = \partial\Omega$ and

$$H^{1/2}_*(\Gamma) := \big\{\psi \in H^{\pm 1/2}(\Gamma) : \langle \psi\,,\,1\rangle = 0\big\},$$

the operator $W\colon H^{1/2}_*(\Gamma) \to H^{-1/2}_*(\Gamma)$ is an elliptic isomorphism. Therefore,

$$\langle\!\langle u\,,\,v\rangle\!\rangle := \begin{cases} \langle Wu\,,\,v\rangle, & \text{if } \Gamma \subsetneq \partial\Omega, \\ \langle Wu\,,\,v\rangle + \langle u\,,\,v\rangle, & \text{if } \Gamma = \partial\Omega \end{cases}$$

defines a scalar product on $\widetilde{H}^{1/2}(\Gamma)$, and the induced norm

$$|||u||| := \langle\!\langle u\,,\, u\rangle\!\rangle^{1/2}$$

is an equivalent norm on $\widetilde{H}^{1/2}(\Gamma)$. Let $g \in H^{-1/2}(\Gamma)$. If $\Gamma \subsetneq \partial\Omega$, suppose additionally that $g \in H_*^{-1/2}(\partial\Omega)$. Then, (6.101) admits unique solutions $u^\star \in \widetilde{H}^{1/2}(\Gamma)$ and $u^\star \in H_*^{1/2}(\partial\Omega)$ respectively, such that $u^\star \in \widetilde{H}^{1/2}(\Gamma)$ is also the unique solution of the variational formulation

$$\langle\!\langle u^\star\,,\, v\rangle\!\rangle = \langle g\,,\, v\rangle \quad \text{for all } v \in \widetilde{H}^{1/2}(\Gamma).$$

Given a mesh $\mathcal{T}_\bullet$ of Γ, let

$$\widetilde{\mathcal{S}}^1(\mathcal{T}_\bullet) := \big\{v \in \widetilde{H}^{1/2}(\Gamma) \,:\, \forall T \in \mathcal{T}_\bullet \quad v|_T \text{ is affine}\big\}.$$

The Lax–Milgram theorem yields existence and uniqueness of $u_\bullet^\star \in \widetilde{\mathcal{S}}^1(\mathcal{T}_\bullet)$ such that

$$\langle\!\langle u_\bullet^\star\,,\, v_\bullet\rangle\!\rangle = \langle g\,,\, v_\bullet\rangle \quad \text{for all } v_\bullet \in \widetilde{\mathcal{S}}^1(\mathcal{T}_\bullet).$$

With the corresponding weighted-residual error estimator, it holds that

$$|||u^\star - u_\bullet^\star||| \le C_{\mathrm{rel}}\,\eta_\bullet(u_\bullet^\star) := \Big(\sum_{T\in\mathcal{T}_\bullet} \eta_\bullet(T,u_\bullet^\star)^2\Big)^{1/2},$$

where

$$\eta_\bullet(T,u_\bullet^\star)^2 := h_T\,\|g - Wu_\bullet^\star\|_{L^2(T)}^2,$$

cf. [CS95, Car97] for $d = 2$ and [CMPS04] for $d = 3$ respectively.

In [Füh14, FFPS17a], optimal additive Schwarz preconditioners are derived for this setting. Hence, Algorithm 57 can also be used in the present setting. We refer to [FFK+15, Section 3.3] for the fact that the *axioms of adaptivity*, i.e., (A1)–(A4) from Proposition 70 remain valid for the hyper-singular integral equation. All other arguments in Section 6.5.4 rely only on general properties of the PCG algorithm (Section 6.5.4), the properties (A1)–(A4), and the Hilbert space setting of $|||\cdot|||$. Overall, this proves that our main results (Theorem 68 and Corollary 78) also cover the hyper-singular integral equation.

6.7 Numerical experiments

In this section, we present numerical experiments that underpin our theoretical findings. We use lowest-order BEM for direct and indirect formulations in 2D as well as 3D. For ease of notation, we define $\lambda := \lambda_{\mathrm{ctr}}$ for this section. We compare the performance of Algorithm 57 for

- different values of $\lambda \in \{1, 10^{-0.5}, 10^{-1}, \dots, 10^{-4}\}$,
- different values of $\theta \in \{0.05, 0.1, 0.15, \dots, 1\}$,

where $\theta = 1$ corresponds to uniform mesh-refinement. In particular, we monitor the condition numbers of the arising BEM systems for diagonal preconditioning [AMT99], the proposed additive Schwarz preconditioning from Section 6.5.1, and no preconditioning. The 2D implementation is based on the MATLAB implementation HILBERT [AEF+14], while the 3D implementation relies on an extension of the BEM++ library [SBA+13].

6.7.1 Slit problem in 2D

Let $\Gamma := (-1, 1) \times \{0\}$, cf. Figure 6.2. We consider the weakly-singular integral equation

$$V\phi = 1 \quad \text{on } \Gamma. \tag{6.102}$$

The unique exact solution of (6.102) reads

$$\phi^\star(x, 0) := -\frac{2x}{\sqrt{1 - x^2}}.$$

For uniform mesh-refinement, we thus expect a convergence order of $\mathcal{O}(N^{-1/2})$, while the optimal rate is $\mathcal{O}(N^{-3/2})$ with respect to the number of elements.

Figure 6.2 shows the condition numbers for an artificial refinement towards the left end point $(-1, 0)$ and for Algorithm 57 with $\lambda = 10^{-3}$ and $\theta = 0.5$. For the proposed additive Schwarz preconditioner, we see that the condition number of the preconditioned Galerkin matrix stays uniformly bounded in both cases, which underpins Theorem 60.

In Figure 6.3–6.4, we compare Algorithm 57 for different values for θ and λ as well as uniform mesh-refinement. Uniform mesh-refinement leads only to the rate $\mathcal{O}(N^{-1/2})$, while adaptivity, independently of the value of θ and λ, regains the optimal rate $\mathcal{O}(N^{-3/2})$.

In Figure 6.5, we compare the computational cost to reach the precision $\tau = 10^{-4}$ for $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$ and $\theta \in \{0.05, 0.1, \dots, 0.95\}$. As a result, we get that the best choice is $\lambda = 1$ and $\theta = 0.65$. For the overall computational cost it then holds that

$$\sum_{(\ell', k') \le (\ell, k)} (\#\mathcal{T}_{\ell'}) \log^2(\#\mathcal{T}_{\ell'}) \approx 353116.2086,$$

where ϕ_ℓ^k is the first approximation such that $\eta_\ell(\phi_\ell^k) < 10^{-4}$.

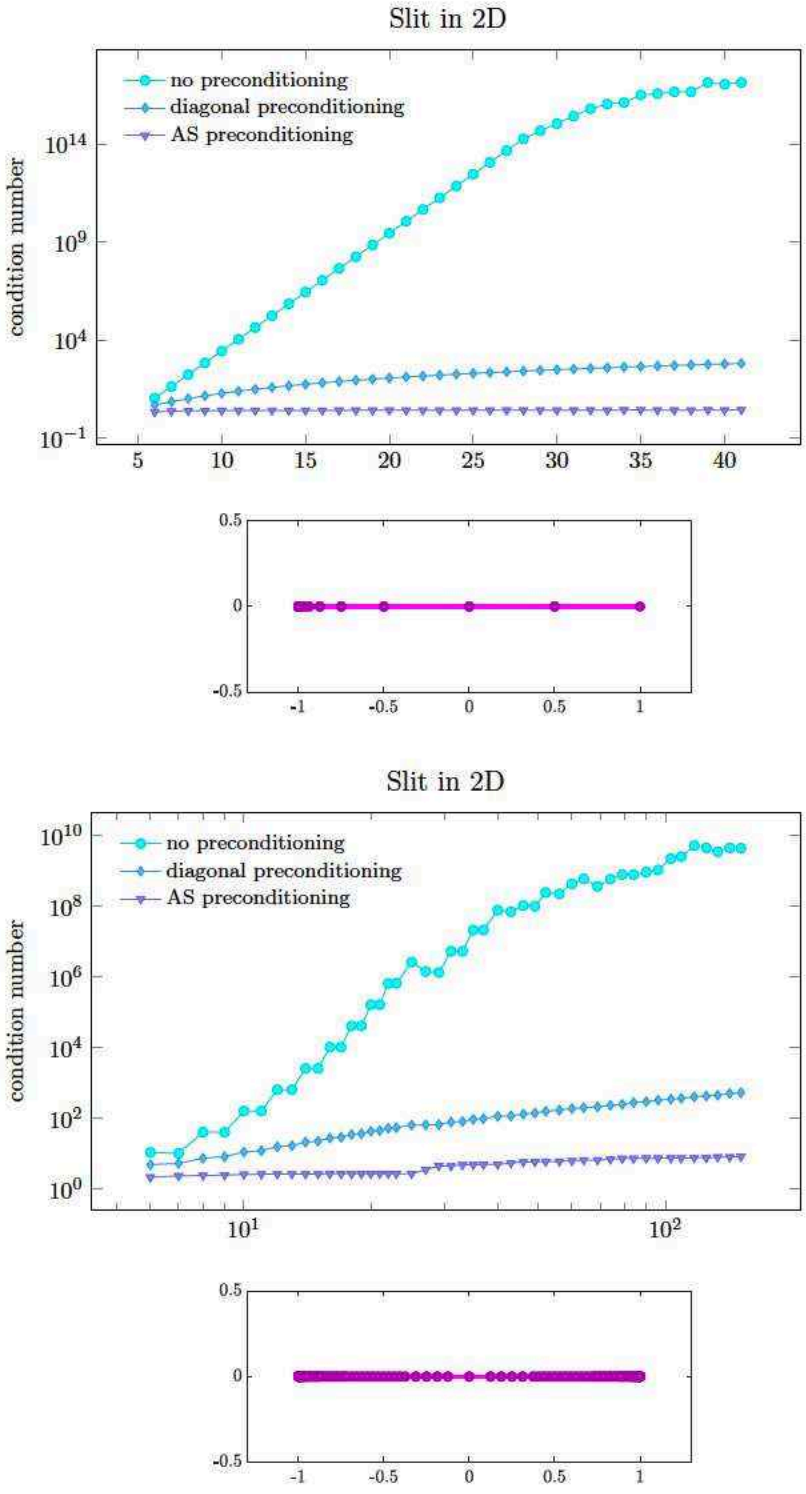


Figure 6.2: Example from Section 6.7.1 (Slit problem in 2D): Condition numbers of the preconditioned and non-preconditioned Galerkin matrix for an artificial refinement towards the left end point (top) and for the matrices arising from Algorithm 57 (bottom).

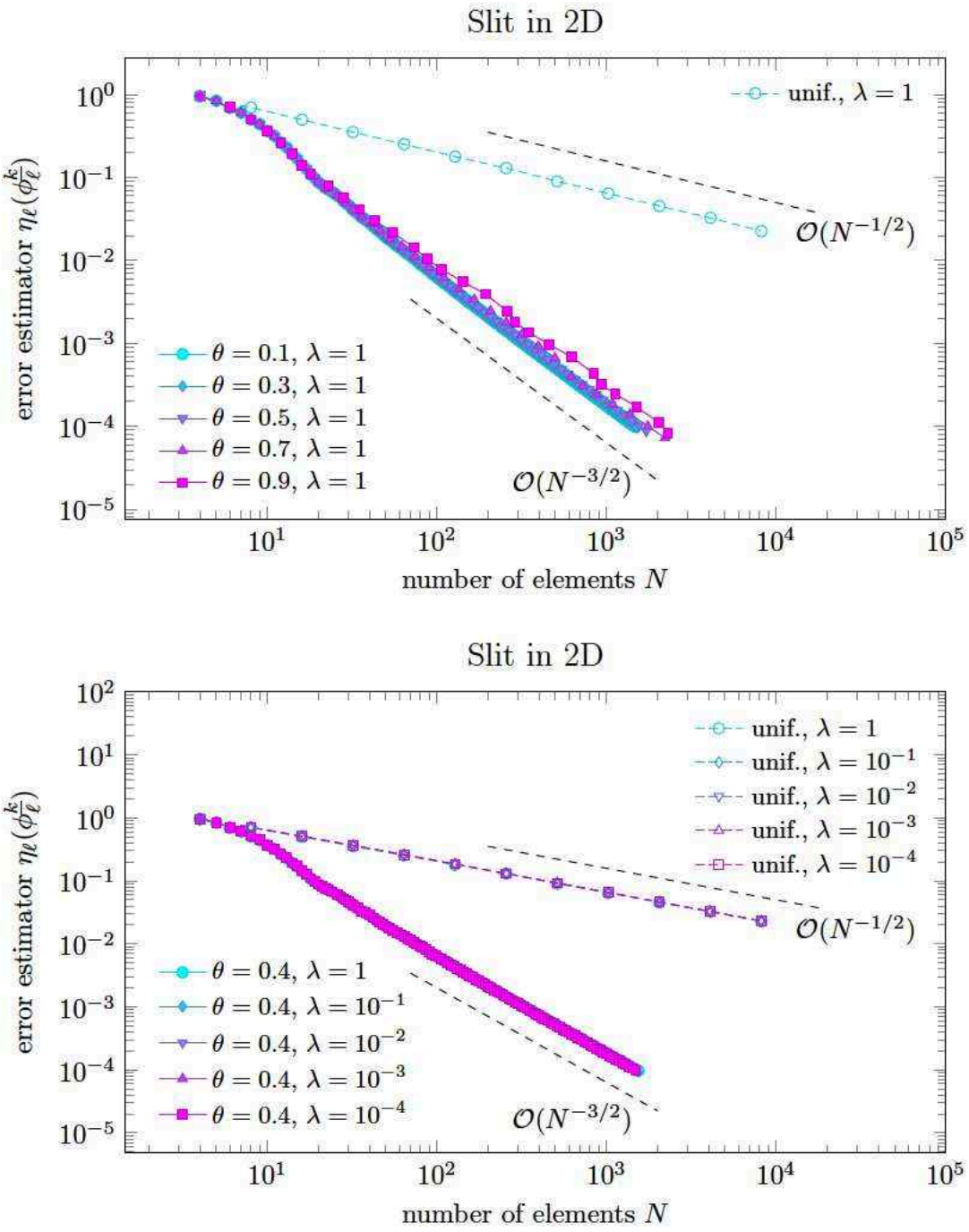


Figure 6.3: Example from Section 6.7.1 (Slit problem in 2D): Estimator convergence for fixed values of λ (left: $\lambda = 1$, right: $\lambda = 10^{-3}$) and $\theta \in \{0.2, 0.4, 0.6, 0.8\}$ (top) and for fixed values of θ (left: $\theta = 0.4$, right: $\theta = 0.6$) and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (bottom).

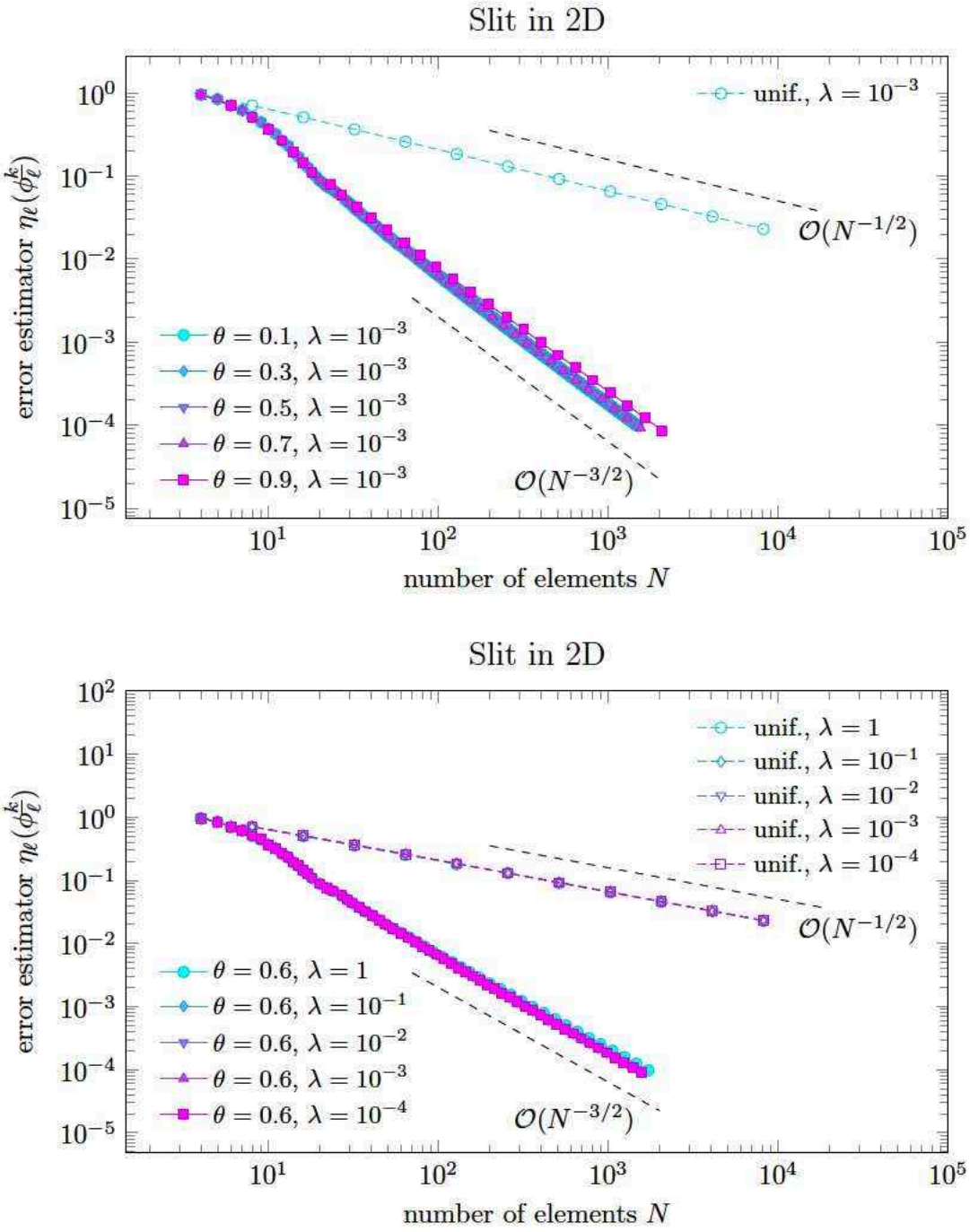


Figure 6.4: Example from Section 6.7.1 (Slit problem in 2D): Estimator convergence for fixed values of λ (left: $\lambda = 1$, right: $\lambda = 10^{-3}$) and $\theta \in \{0.2, 0.4, 0.6, 0.8\}$ (top) and for fixed values of θ (left: $\theta = 0.4$, right: $\theta = 0.6$) and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (bottom).

θ \ λ	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
0.05	5.0e+07	5.0e+07	5.0e+07	1.0e+08	1.0e+08	1.4e+08	1.8e+08	2.3e+08	**3.2e+08**
0.1	1.9e+07	1.9e+07	2.0e+07	3.9e+07	4.4e+07	6.0e+07	7.8e+07	1.1e+08	1.4e+08
0.15	9.5e+06	9.5e+06	1.9e+07	2.0e+07	2.5e+07	3.4e+07	4.7e+07	6.1e+07	7.6e+07
0.2	5.3e+06	5.4e+06	1.2e+07	1.2e+07	1.7e+07	2.3e+07	2.9e+07	4.0e+07	4.7e+07
0.25	3.4e+06	3.4e+06	7.8e+06	8.0e+06	1.2e+07	1.6e+07	2.0e+07	2.7e+07	3.1e+07
0.3	2.3e+06	2.3e+06	5.3e+06	5.6e+06	8.1e+06	1.1e+07	1.5e+07	1.9e+07	2.2e+07
0.35	1.7e+06	1.7e+06	3.8e+06	4.2e+06	5.7e+06	7.9e+06	1.1e+07	1.4e+07	1.6e+07
0.4	1.2e+06	1.7e+06	2.9e+06	3.1e+06	4.4e+06	6.1e+06	8.6e+06	1.0e+07	1.2e+07
0.45	9.3e+05	2.1e+06	2.2e+06	2.5e+06	3.4e+06	4.9e+06	6.6e+06	8.1e+06	9.5e+06
0.5	7.3e+05	1.7e+06	1.8e+06	2.0e+06	2.7e+06	4.1e+06	5.3e+06	6.4e+06	7.5e+06
0.55	5.4e+05	1.4e+06	1.4e+06	1.6e+06	2.2e+06	3.2e+06	4.1e+06	5.0e+06	5.9e+06
0.6	4.2e+05	1.0e+06	1.0e+06	1.4e+06	1.8e+06	2.7e+06	3.7e+06	4.3e+06	5.0e+06
0.65	**3.5e+05**	8.2e+05	8.5e+05	1.1e+06	1.6e+06	2.2e+06	3.0e+06	3.5e+06	4.2e+06
0.7	4.2e+05	6.4e+05	7.0e+05	9.2e+05	1.3e+06	1.8e+06	2.3e+06	2.7e+06	3.3e+06
0.75	4.4e+05	6.2e+05	6.5e+05	9.2e+05	1.3e+06	1.8e+06	2.2e+06	2.7e+06	3.0e+06
0.8	4.9e+05	5.5e+05	5.8e+05	1.0e+06	1.3e+06	1.6e+06	2.0e+06	2.3e+06	2.5e+06
0.85	4.9e+05	7.0e+05	9.3e+05	1.2e+06	1.6e+06	2.1e+06	2.5e+06	2.8e+06	3.2e+06
0.9	7.5e+05	7.9e+05	1.0e+06	1.4e+06	1.8e+06	2.4e+06	2.7e+06	3.1e+06	3.7e+06
0.95	8.1e+05	1.2e+06	1.6e+06	2.1e+06	2.6e+06	3.2e+06	4.2e+06	4.7e+06	5.3e+06

min **max**

Figure 6.5: Example from Section 6.7.1 (Slit problem in 2D): Overall computational cost $\sum_{(\ell',k')\le(\ell,k)}(\#\mathcal{T}_{\ell'})\log^2(\#\mathcal{T}_{\ell'})$ such that $\eta_\ell(\phi_\ell^k) < \tau$ for given precision $\tau = 10^{-4}$, $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$, and $\theta \in \{0.05, 0.1, \dots, 0.95\}$.

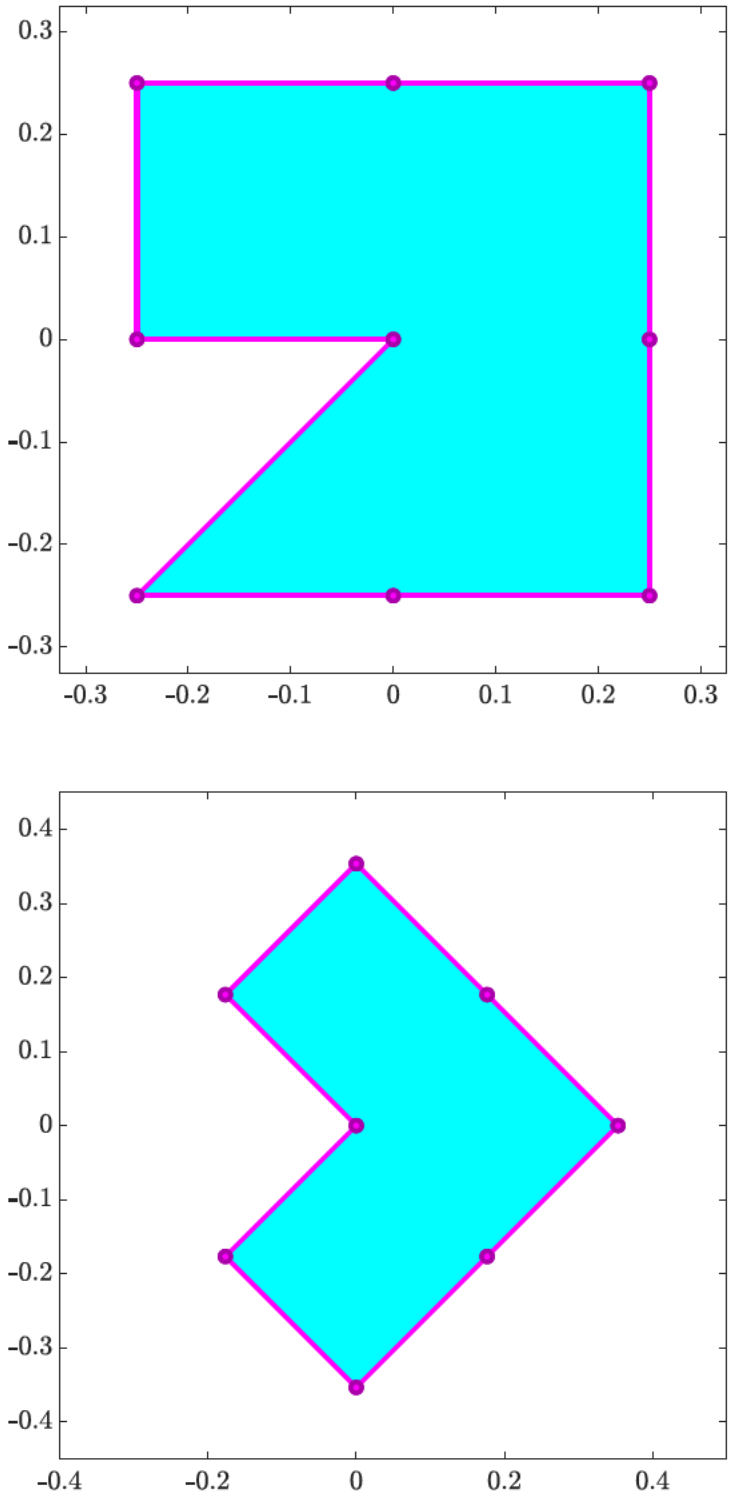


Figure 6.6: Z–shaped domain $\Omega \subset \mathbb{R}^2$ with initial mesh $\mathcal{T}_0$ (top) and L–shaped domain $\Omega \subset \mathbb{R}^2$ with initial mesh $\mathcal{T}_0$ (bottom).

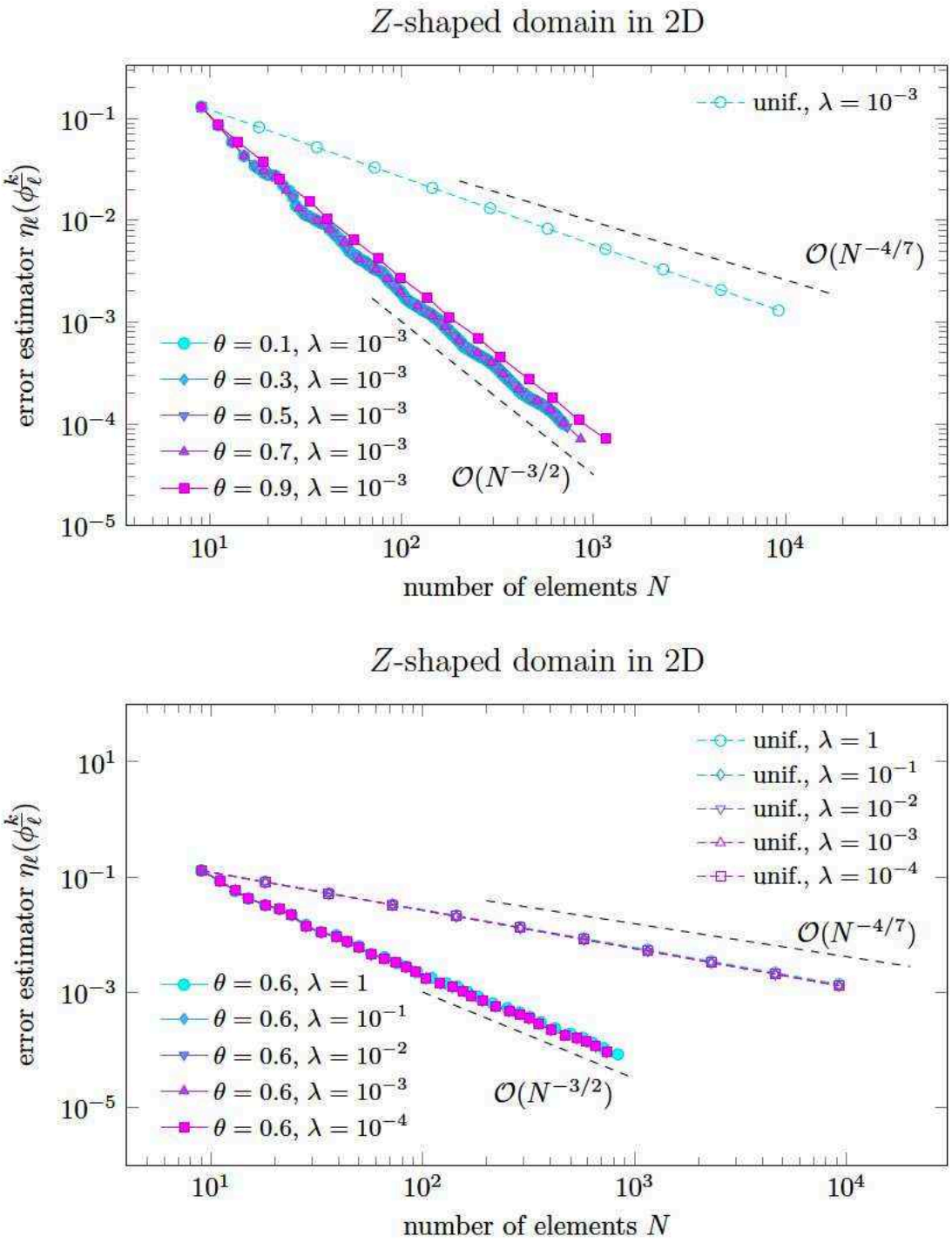


Figure 6.7: Example from Section 6.7.2 (Z-shaped domain in 2D): Estimator convergence for fixed value of $\lambda = 10^{-3}$ and $\theta \in \{0.2, 0.4, 0.6, 0.8\}$ (top) and for fixed value of $\theta = 0.6$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (bottom).

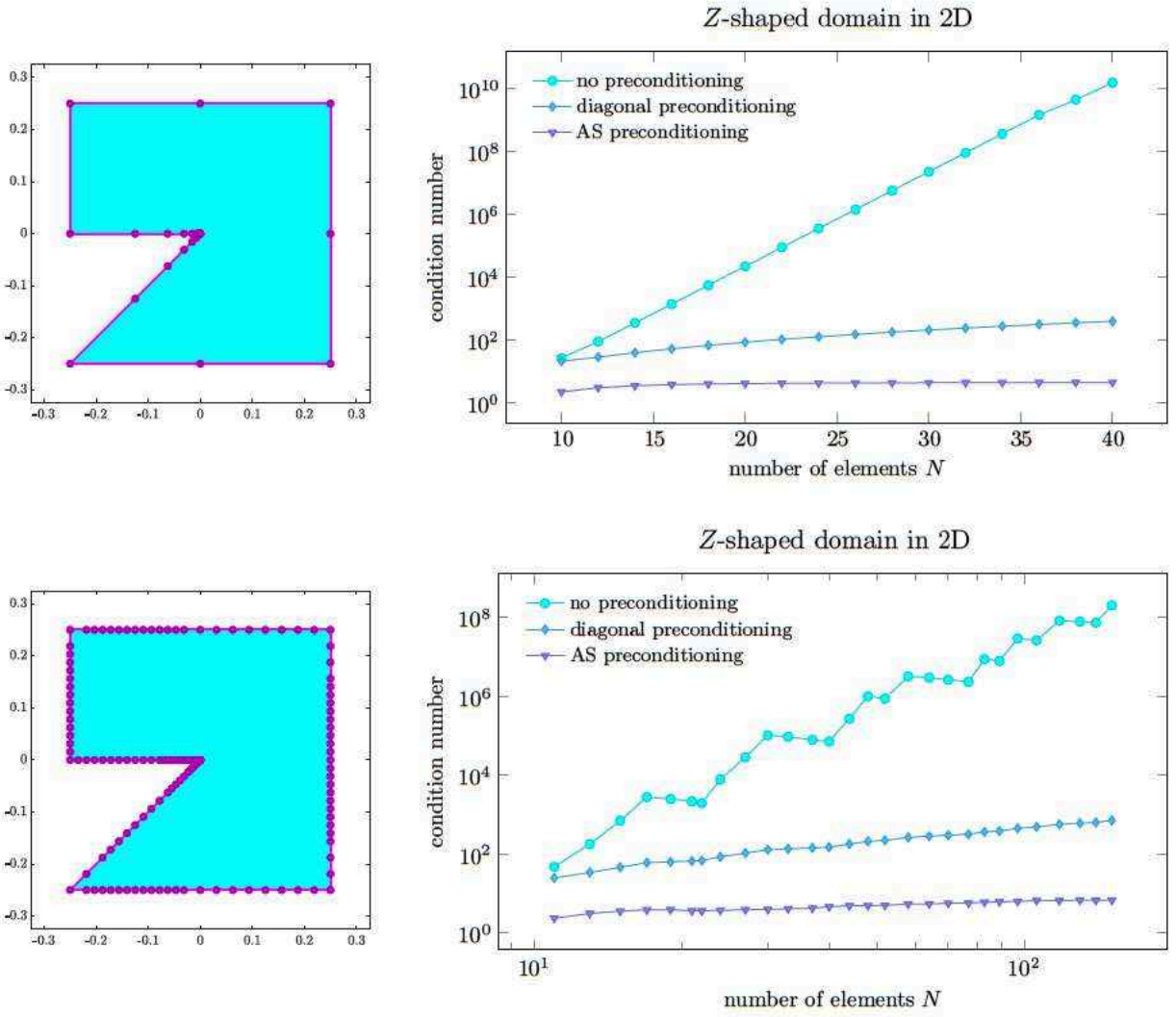


Figure 6.8: Example from Section 6.7.2 (Z-shaped domain in 2D): Condition numbers of the preconditioned and non-preconditioned Galerkin matrix for an artificial refinement towards the reentrant corner (top) and for Algorithm 57 (bottom), where $\lambda = 10^{-3}$ and $\theta = 0.5$.

θ \ λ	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
0.05	8.7e+06	8.7e+06	9.2e+06	1.6e+07	2.6e+07	3.6e+07	4.4e+07	5.3e+07	**6.5e+07**
0.1	5.3e+06	5.3e+06	5.9e+06	1.3e+07	1.9e+07	2.4e+07	2.9e+07	3.5e+07	4.3e+07
0.15	2.9e+06	2.9e+06	4.0e+06	8.1e+06	1.1e+07	1.4e+07	1.7e+07	2.1e+07	2.5e+07
0.2	1.8e+06	1.8e+06	3.0e+06	5.2e+06	7.2e+06	9.1e+06	1.1e+07	1.3e+07	1.6e+07
0.25	1.1e+06	1.1e+06	2.3e+06	3.8e+06	5.1e+06	6.3e+06	7.7e+06	9.2e+06	1.1e+07
0.3	8.2e+05	8.2e+05	2.0e+06	3.0e+06	3.9e+06	4.7e+06	5.6e+06	6.8e+06	7.9e+06
0.35	6.0e+05	6.1e+05	1.6e+06	2.4e+06	3.0e+06	3.6e+06	4.3e+06	5.2e+06	6.0e+06
0.4	4.4e+05	5.0e+05	1.3e+06	1.9e+06	2.3e+06	2.8e+06	3.4e+06	4.1e+06	4.7e+06
0.45	3.4e+05	4.8e+05	1.1e+06	1.5e+06	1.8e+06	2.2e+06	2.7e+06	3.2e+06	3.7e+06
0.5	2.8e+05	4.7e+05	8.5e+05	1.1e+06	1.4e+06	1.7e+06	2.1e+06	2.5e+06	2.9e+06
0.55	2.2e+05	4.1e+05	7.1e+05	9.4e+05	1.2e+06	1.5e+06	1.8e+06	2.1e+06	2.4e+06
0.6	1.9e+05	3.4e+05	5.5e+05	7.5e+05	9.2e+05	1.1e+06	1.4e+06	1.7e+06	1.9e+06
0.65	1.5e+05	3.3e+05	4.8e+05	6.3e+05	7.7e+05	9.8e+05	1.2e+06	1.4e+06	1.6e+06
0.7	1.1e+05	2.6e+05	3.6e+05	5.7e+05	7.1e+05	8.8e+05	1.1e+06	1.3e+06	1.4e+06
0.75	1.1e+05	2.2e+05	3.0e+05	3.8e+05	4.7e+05	6.3e+05	7.5e+05	8.4e+05	9.7e+05
0.8	**1.1e+05**	2.1e+05	2.7e+05	3.3e+05	4.3e+05	5.7e+05	6.6e+05	7.4e+05	8.5e+05
0.85	1.4e+05	2.6e+05	3.3e+05	4.7e+05	6.1e+05	7.1e+05	8.4e+05	9.4e+05	1.1e+06
0.9	1.5e+05	2.9e+05	3.5e+05	5.6e+05	6.8e+05	8.3e+05	9.6e+05	1.1e+06	1.2e+06
0.95	2.3e+05	2.8e+05	3.8e+05	5.2e+05	7.6e+05	8.9e+05	1.0e+06	1.1e+06	1.4e+06

min **max**

Figure 6.9: Example from Section 6.7.2 (Weakly-singular integral equation on Z-shaped domain in 2D): Overall computational cost $\sum_{(\ell',k')\leq(\ell,k)}(\#\mathcal{T}_{\ell'})\log^2(\#\mathcal{T}_{\ell'})$ such that $\eta_\ell(\phi_\ell^k) < \tau$ for given precision $\tau = 10^{-4}$, $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$, and $\theta \in \{0.05, 0.1, \dots, 0.95\}$.

6.7.2 Weakly-singular integral equation on Z-shaped domain in 2D

Let $\Gamma := \partial\Omega$ be the boundary of the Z-shaped domain with reentrant corner at the origin $(0,0)$, cf. Figure 6.6 (top). We consider the weakly-singular integral equation (6.11) with the right-hand side $f = (K + 1/2)g$ where $K\colon H^{1/2}(\Gamma) \to H^{1/2}(\Gamma)$ is the double-layer operator from Section 6.2.1. We note that the weakly-singular integral equation (6.11) is then equivalent to the Dirichlet problem

$$\begin{aligned} -\Delta u &= 0 \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma. \end{aligned} \tag{6.103}$$

We prescribe the exact solution $u(x_1, x_2)$ in 2D polar coordinates

$$(x_1, x_2) = r(\cos\xi, \sin\xi) \quad \text{with} \quad \xi \in (-\pi, \pi)$$

as follows

$$u(x_1, x_2) := r^{4/7}\cos(4\,\xi/7). \tag{6.104}$$

Then, u admits a generic singularity at the reentrant corner. The exact solution $\phi^\star$ of (6.11) is just the normal derivative of the solution u.

We expect a convergence order of $\mathcal{O}(N^{-4/7})$ for uniform mesh-refinement, and the optimal rate $\mathcal{O}(N^{-3/2})$ for the adaptive strategy, which is seen in Figure 6.7 for different values of θ and λ.

Figure 6.8 shows the condition numbers for an artificial refinement towards the reentrant corner as well as the condition numbers for Algorithm 57 with $\lambda = 10^{-3}$ and $\theta = 0.5$.

In Figure 6.9, we compare the computational cost to reach the precision $\tau = 10^{-4}$ for $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$ and $\theta \in \{0.05, 0.1, \dots, 0.95\}$. As a result, we get that the best choice is $\lambda = 1$ and $\theta = 0.8$. For the overall computational cost it then holds that

$$\sum_{(\ell',k')\le(\ell,k)} (\#\mathcal{T}_{\ell'})\log^2(\#\mathcal{T}_{\ell'}) \approx 105563.4255,$$

where ϕ_ℓ^k is the first approximation such that $\eta_\ell(\phi_\ell^k) < 10^{-4}$.

6.7.3 Hyper-singular integral equation on L-shaped domain in 2D

Let $\Gamma := \partial\Omega$ be the boundary of the L-shaped domain with reentrant corner at the origin $(0,0)$, cf. Figure 6.6 (bottom). We consider the hyper-singular integral equation (6.101) with the right-hand side $g = (1/2 - K')\phi$ where $K'\colon \widetilde{H}^{-1/2}(\Gamma) \to H^{-1/2}(\Gamma)$ is the adjoint double-layer operator from Section 6.2.1. We note that the hyper-singular integral equation (6.101) is then equivalent to the Neumann problem

$$\begin{aligned} -\Delta P &= 0 \quad \text{in } \Omega \\ \partial_{\mathbf{n}} P &= \phi \quad \text{on } \Gamma. \end{aligned} \tag{6.105}$$

We prescribe the exact solution $P(x_1, x_2)$ of the Laplace problem in 2D polar coordinates

$$(x_1, x_2) = r(\cos\xi, \sin\xi) \quad \text{with} \quad \xi \in (-\pi, \pi)$$

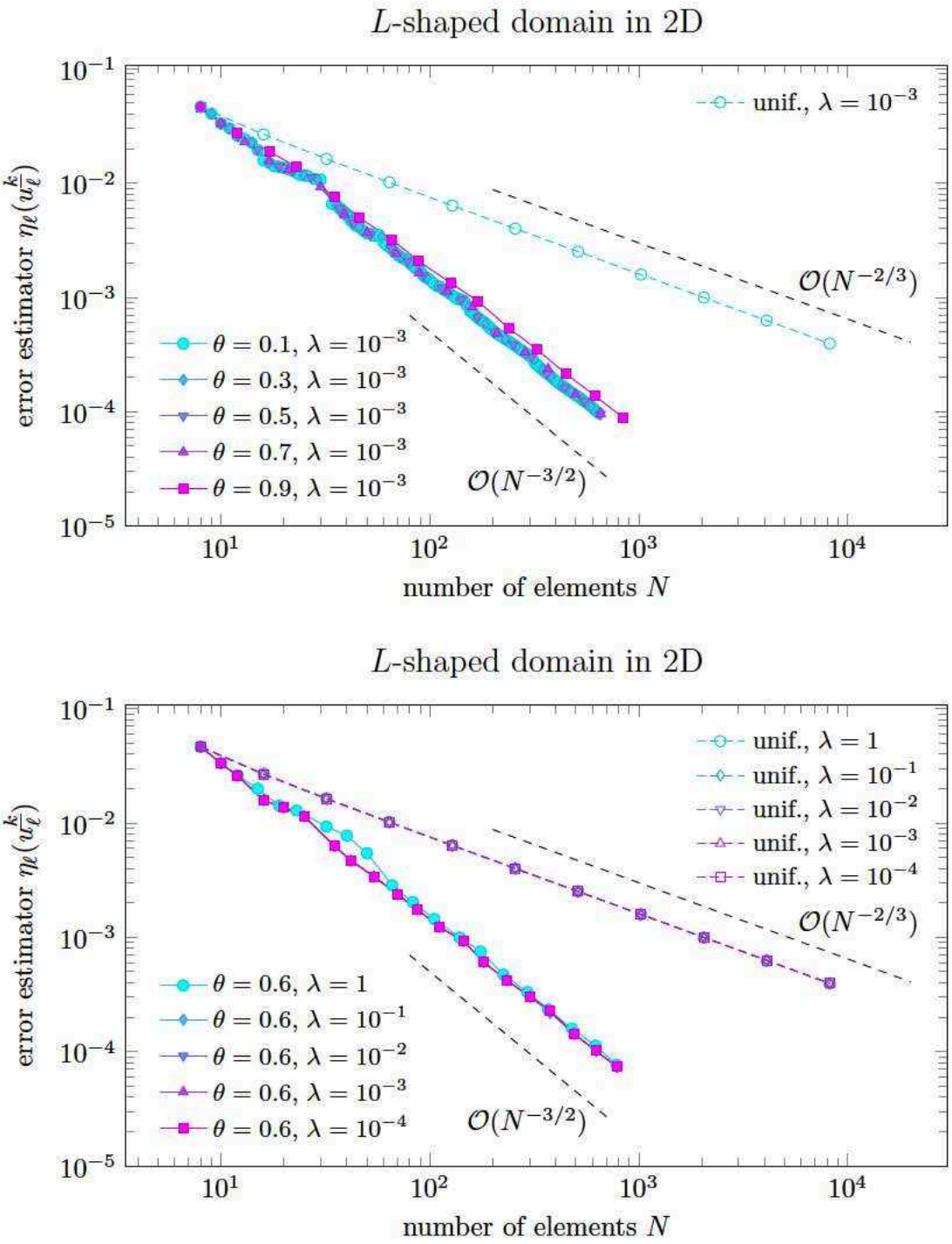


Figure 6.10: Example from Section 6.7.3 (L-shaped domain in 2D): Estimator convergence for fixed value of $\lambda = 10^{-3}$ and $\theta \in \{0.2, 0.4, 0.6, 0.8\}$ (top) and for fixed value of $\theta = 0.6$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (bottom).

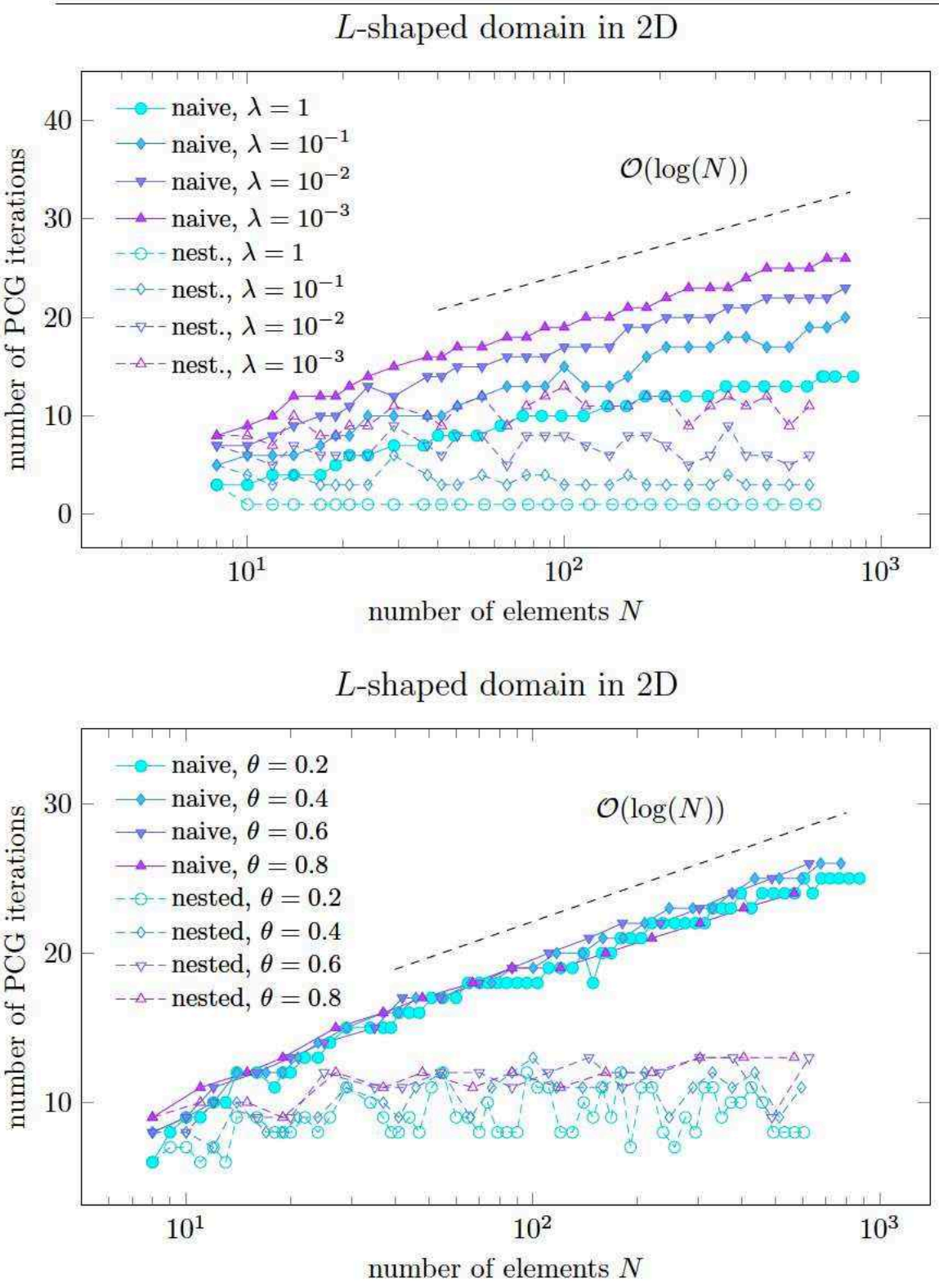


Figure 6.11: Example from Section 6.7.3 (L-shaped domain in 2D): Number of PCG iterations in Algorithm 57 for nested iteration (dashed lines), i.e., $u_{\ell+1}^0 := u_\ell^k$ in Step (iv) of Algorithm 57, and naive initial guess (solid lines), i.e., $u_{\ell+1}^0 := 0$. We compare a fixed value of $\theta = 0.4$ and $\lambda \in \{1, 10^{-1}, 10^{-2}, 10^{-3}\}$ (top) as well as a fixed value of $\lambda = 10^{-3}$ and $\theta \in \{0.2, 0.4, 0.6, 0.8\}$ (bottom).

θ \ λ	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
0.05	1.1e+06	1.1e+06	2.2e+06	3.3e+06	4.7e+06	6.2e+06	7.9e+06	9.7e+06	**1.2e+07**
0.1	6.0e+05	6.2e+05	1.4e+06	2.1e+06	2.8e+06	3.7e+06	4.7e+06	5.7e+06	6.8e+06
0.15	3.8e+05	4.5e+05	9.9e+05	1.5e+06	1.9e+06	2.6e+06	3.3e+06	4.0e+06	4.8e+06
0.2	2.8e+05	3.4e+05	7.6e+05	1.1e+06	1.5e+06	2.1e+06	2.6e+06	3.1e+06	3.7e+06
0.25	2.3e+05	3.7e+05	7.1e+05	9.9e+05	1.4e+06	1.8e+06	2.3e+06	2.7e+06	3.2e+06
0.3	2.0e+05	3.4e+05	5.4e+05	7.6e+05	1.1e+06	1.4e+06	1.7e+06	2.1e+06	2.5e+06
0.35	1.5e+05	3.2e+05	5.1e+05	7.3e+05	9.9e+05	1.3e+06	1.7e+06	1.9e+06	2.3e+06
0.4	1.4e+05	2.6e+05	4.2e+05	6.2e+05	8.5e+05	1.1e+06	1.4e+06	1.6e+06	2.0e+06
0.45	1.4e+05	2.4e+05	3.4e+05	5.0e+05	6.8e+05	9.1e+05	1.1e+06	1.4e+06	1.5e+06
0.5	9.8e+04	1.9e+05	3.0e+05	4.4e+05	6.4e+05	8.6e+05	1.0e+06	1.3e+06	1.4e+06
0.55	9.4e+04	1.9e+05	2.9e+05	4.0e+05	5.7e+05	7.9e+05	1.0e+06	1.2e+06	1.3e+06
0.6	9.7e+04	1.7e+05	3.3e+05	4.6e+05	6.6e+05	9.0e+05	1.0e+06	1.2e+06	1.4e+06
0.65	9.5e+04	1.6e+05	2.6e+05	3.5e+05	4.9e+05	6.3e+05	7.8e+05	9.0e+05	9.8e+05
0.7	**9.1e+04**	1.6e+05	2.4e+05	3.5e+05	4.6e+05	6.0e+05	7.0e+05	8.1e+05	9.5e+05
0.75	1.3e+05	2.1e+05	3.1e+05	4.3e+05	5.9e+05	7.2e+05	9.0e+05	1.0e+06	1.2e+06
0.8	9.9e+04	2.1e+05	2.9e+05	3.9e+05	5.1e+05	6.5e+05	8.1e+05	9.0e+05	9.9e+05
0.85	1.7e+05	2.5e+05	3.7e+05	5.2e+05	7.1e+05	8.5e+05	1.0e+06	1.2e+06	1.3e+06
0.9	1.5e+05	2.2e+05	3.0e+05	4.4e+05	5.2e+05	7.1e+05	8.5e+05	9.7e+05	1.1e+06
0.95	2.1e+05	3.2e+05	5.1e+05	6.3e+05	7.5e+05	1.0e+06	1.1e+06	1.4e+06	1.6e+06

min **max**

Figure 6.12: Example from Section 6.7.3 (Hyper-singular integral equation on L-shaped domain in 2D): Overall computational cost $\sum_{(\ell',k')\leq(\ell,k)}(\#\mathcal{T}_{\ell'})\log^2(\#\mathcal{T}_{\ell'})$ such that $\eta_\ell(u_\ell^k) < \tau$ for given precision $\tau = 10^{-4}$, $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$, and $\theta \in \{0.05, 0.1, \dots, 0.95\}$.

as follows

$$P(x_1, x_2) := r^{2/3} \cos(2\,\xi/3). \tag{6.106}$$

Then, ϕ is just the normal derivative of P which has a generic singularity at the reentrant corner. The exact solution $u^\star$ of the hyper-singular integral equation (6.101) is simply the restriction of the function P to the boundary Γ minus the integral mean of P on Γ.

We expect a convergence order of $\mathcal{O}(N^{-2/3})$ for uniform mesh-refinement, and the optimal rate $\mathcal{O}(N^{-3/2})$ for the adaptive strategy, which is seen in Figure 6.10 for different values of θ and λ.

A naive initial guess in Step (iv) of Algorithm 57 (i.e., if $u_{\ell+1}^0 := 0$) leads to a logarithmical growth of the number of PCG iterations, whereas for nested iteration $u_{\ell+1}^0 := u_\ell^k$ the number of PCG iterations stays uniformly bounded, cf. Figure 6.11.

In Figure 6.12, we compare the computational cost to reach the precision $\tau = 10^{-4}$ for $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$ and $\theta \in \{0.05, 0.1, \dots, 0.95\}$. As a result, we get that the best choice is $\lambda = 1$ and $\theta = 0.7$. For the overall computational cost it then holds that

$$\sum_{(\ell',k')\le(\ell,k)} (\#\mathcal{T}_{\ell'}) \log^2(\#\mathcal{T}_{\ell'}) \approx 90975.1021,$$

where u_ℓ^k is the first approximation such that $\eta_\ell(u_\ell^k) < 10^{-4}$.

6.7.4 Weakly-singular integral equation on L-shaped domain in 3D

Let $\Gamma := \partial\Omega$ be the boundary of the L-shaped domain

$$\Omega = (-1,1)^3 \backslash ([-1,0] \times [0,1] \times [-1,1]),$$

cf. Figure 6.13. We consider the weakly-singular integral equation (6.11) with the right-hand side $f = (K + 1/2)g$ where $K\colon H^{1/2}(\Gamma) \to H^{1/2}(\Gamma)$ is the double-layer operator from Section 6.2.1. Again, the weakly-singular integral equation (6.11) is then equivalent to the Dirichlet problem

$$\begin{aligned} -\Delta u &= 0 \quad \text{in } \Omega \\ u &= g \quad \text{on } \Gamma. \end{aligned} \tag{6.107}$$

We prescribe the exact solution $u(x_1, x_2, x_3)$ in 3D cylindrical coordinates

$$(x_1, x_2, x_3) = (r\cos\xi, r\sin\xi, x_3) \quad \text{with} \quad \xi \in (-\pi, \pi)$$

as follows

$$u(x_1, x_2, x_3) = x_3\, r^{2/3} \cos(2/3\,(\xi - \pi/4)). \tag{6.108}$$

Note that u admits a singularity along the reentrant edge. The exact solution $\phi^\star$ of (6.11) is just the normal derivative of the exact solution u.

Figure 6.13 shows the condition numbers for (diagonal or additive Schwarz) preconditioning and no preconditioning for artificial refinements towards one reentrant corner or the

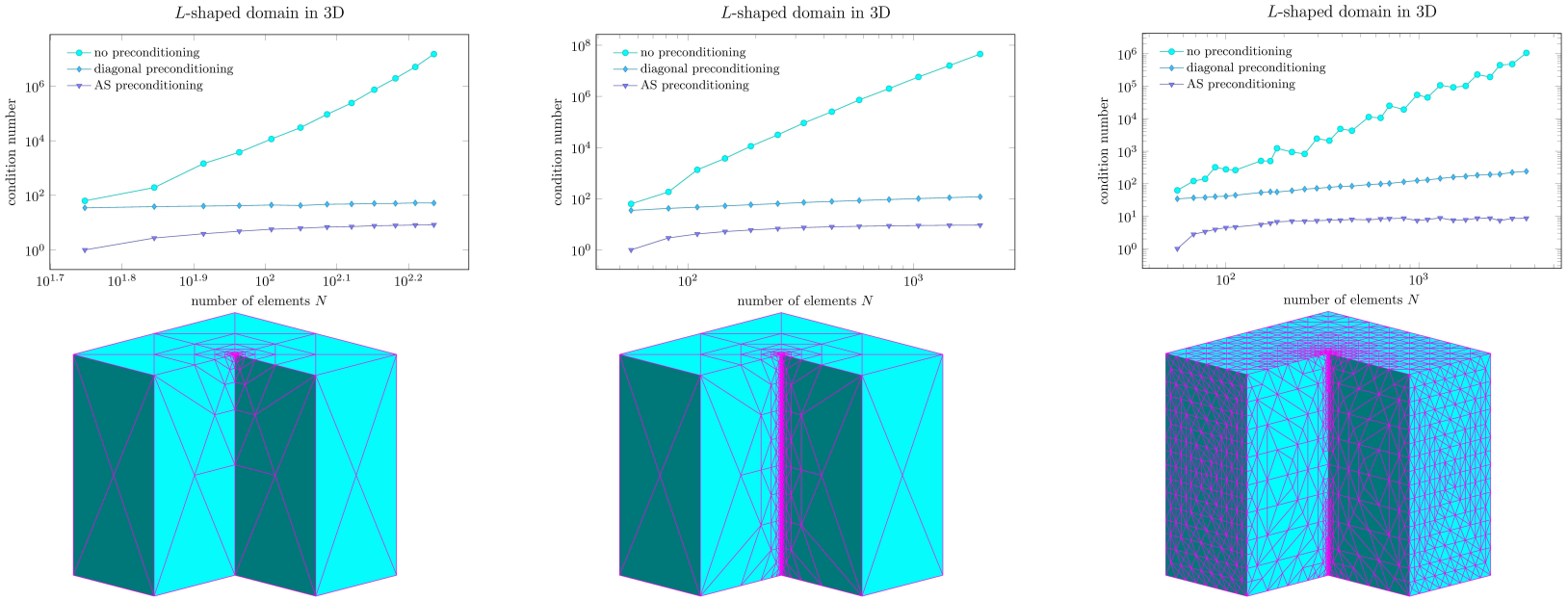


Figure 6.13: Example from Section 6.7.4 (L-shaped domain in 3D): Condition numbers of the preconditioned and non-preconditioned Galerkin matrix for an artificial refinement towards one reentrant corner (left) or edge (middle), and for Algorithm 57 (right).

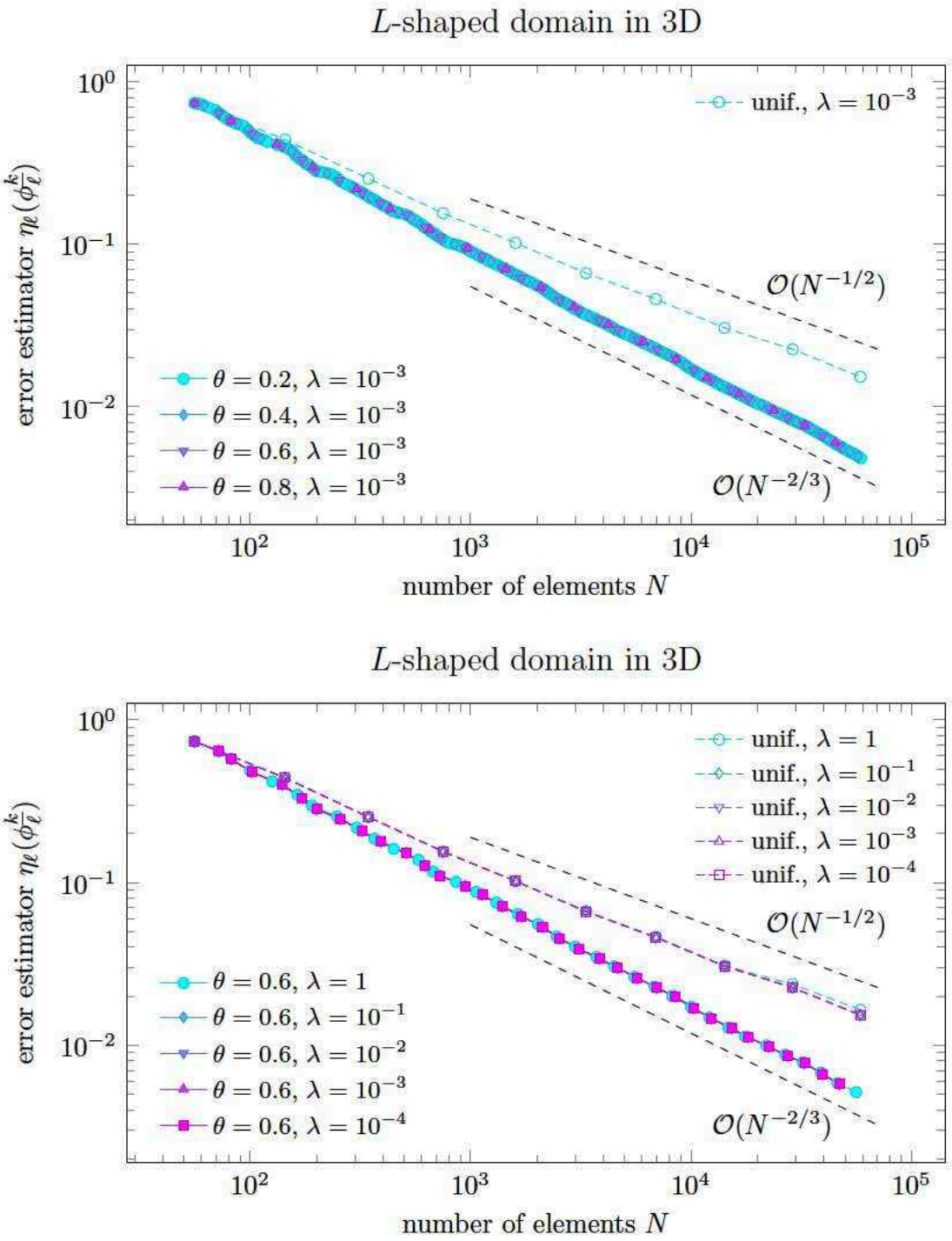


Figure 6.14: Example from Section 6.7.4 (L-shaped domain in 3D): Estimator convergence for fixed values of $\lambda = 10^{-3}$ and $\theta \in \{0.2, 0.4, 0.6, 0.8\}$ (top) and for fixed value of $\theta = 0.6$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$ (bottom).

reentrant edge as well as the condition numbers of the matrices arising from Algorithm 57 with $\lambda = 10^{-3}$ and $\theta = 0.5$.

In Figure 6.14, we compare Algorithm 57 with different values for θ and λ to uniform mesh-refinement. Uniform mesh-refinement leads only to a reduced rate of $\mathcal{O}(N^{-1/2})$, while adaptivity, independently of θ and λ, leads to the improved rate of approximately $\mathcal{O}(N^{-2/3})$. While one would expect $\mathcal{O}(N^{-3/4})$ for smooth exact solutions $\phi^\star$, this would require anisotropic elements along the reentrant edge for the present solution $\phi^\star = \partial_n u$. Since NVB guarantees uniform γ-shape regularity of the meshes, the latter is not possible and hence leads to a reduced optimal rate.

In Figure 6.15, we compare the computational cost to reach the precision $\tau = 10^{-2}$ for $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$ and $\theta \in \{0.05, 0.1, \dots, 0.95\}$. As a result, we get that the best choice is $\lambda = 1$ and $\theta = 0.8$. For the overall computational cost it then holds that

$$\sum_{(\ell',k')\le(\ell,k)} (\#\mathcal{T}_{\ell'}) \log^2(\#\mathcal{T}_{\ell'}) \approx 1067163.4947,$$

where ϕ_ℓ^k is the first approximation such that $\eta_\ell(\phi_\ell^k) < 10^{-2}$.

6.7.5 Computational complexity

With Figure 6.16–6.17, we aim to underpin the almost optimal computational complexity of Algorithm 57 (see Corollary 78). To this end, we plot the error estimator $\eta_\ell(\phi_\ell^k)$ over the cumulative sums

$$\sum_{(\ell',k')\le(\ell,k)} \#\mathcal{T}_{\ell'}$$

as well as

$$\sum_{(\ell',k')\le(\ell,k)} (\#\mathcal{T}_{\ell'}) \log^2(\#\mathcal{T}_{\ell'})$$

for $\theta = 0.4$ and $\lambda \in \{1, 10^{-3}\}$. The negative impact of the logarithmic term on the (preasymptotic) convergence rate is clearly visible.

θ \ λ_{ctr}	1	$10^{-0.5}$	10^{-1}	$10^{-1.5}$	10^{-2}	$10^{-2.5}$	10^{-3}	$10^{-3.5}$	10^{-4}
0.05	7.2e+07	7.5e+07	1.6e+08	5.6e+08	1.2e+09	2.1e+09	3.4e+09	5.0e+09	**6.9e+09**
0.1	2.5e+07	3.0e+07	1.2e+08	3.1e+08	5.8e+08	9.8e+08	1.5e+09	2.2e+09	2.9e+09
0.15	1.2e+07	1.8e+07	8.9e+07	1.9e+08	3.5e+08	5.8e+08	8.7e+08	1.2e+09	1.6e+09
0.2	7.1e+06	1.8e+07	6.9e+07	1.4e+08	2.3e+08	3.7e+08	5.5e+08	7.6e+08	1.0e+09
0.25	5.0e+06	1.8e+07	5.2e+07	1.1e+08	1.8e+08	2.8e+08	4.1e+08	5.6e+08	7.3e+08
0.3	3.5e+06	1.6e+07	4.4e+07	8.4e+07	1.4e+08	2.2e+08	3.2e+08	4.3e+08	5.8e+08
0.35	2.7e+06	1.3e+07	3.8e+07	6.7e+07	1.1e+08	1.8e+08	2.5e+08	3.3e+08	4.3e+08
0.4	2.1e+06	1.3e+07	3.1e+07	5.5e+07	9.0e+07	1.4e+08	2.1e+08	2.7e+08	3.5e+08
0.45	1.7e+06	1.0e+07	2.6e+07	4.3e+07	7.6e+07	1.2e+08	1.6e+08	2.2e+08	2.8e+08
0.5	1.9e+06	1.0e+07	2.2e+07	3.8e+07	6.4e+07	9.8e+07	1.3e+08	1.8e+08	2.3e+08
0.55	1.4e+06	1.0e+07	1.6e+07	3.0e+07	6.3e+07	9.2e+07	1.3e+08	1.7e+08	2.2e+08
0.6	1.6e+06	8.6e+06	1.6e+07	2.8e+07	4.6e+07	7.1e+07	9.8e+07	1.3e+08	1.7e+08
0.65	2.2e+06	8.2e+06	1.7e+07	3.0e+07	4.7e+07	6.7e+07	9.3e+07	1.2e+08	1.6e+08
0.7	1.6e+06	6.5e+06	1.2e+07	2.3e+07	3.5e+07	5.1e+07	7.0e+07	9.5e+07	1.1e+08
0.75	2.4e+06	4.7e+06	8.4e+06	1.8e+07	2.5e+07	3.9e+07	5.6e+07	6.8e+07	9.1e+07
0.8	**1.1e+06**	4.0e+06	8.8e+06	1.5e+07	2.1e+07	3.4e+07	4.2e+07	6.0e+07	7.1e+07
0.85	2.6e+06	6.2e+06	1.5e+07	2.3e+07	4.1e+07	5.2e+07	7.6e+07	1.0e+08	1.2e+08
0.9	2.7e+06	6.3e+06	1.5e+07	2.3e+07	3.8e+07	5.5e+07	7.5e+07	1.0e+08	1.2e+08
0.95	3.2e+06	7.1e+06	1.4e+07	2.5e+07	3.6e+07	5.6e+07	7.3e+07	1.0e+08	1.4e+08

min **max**

Figure 6.15: Example from Section 6.7 (L-shaped domain in 3D): Overall computational cost $\sum_{(\ell',k')\le(\ell,k)}(\#\mathcal{T}_{\ell'})\log^2(\#\mathcal{T}_{\ell'})$ such that $\eta_\ell(\phi_\ell^k) < \tau$ for given precision $\tau = 10^{-2}$, $\lambda \in \{1, 10^{-0.5}, \dots, 10^{-4}\}$, and $\theta \in \{0.05, 0.1, \dots, 0.95\}$.

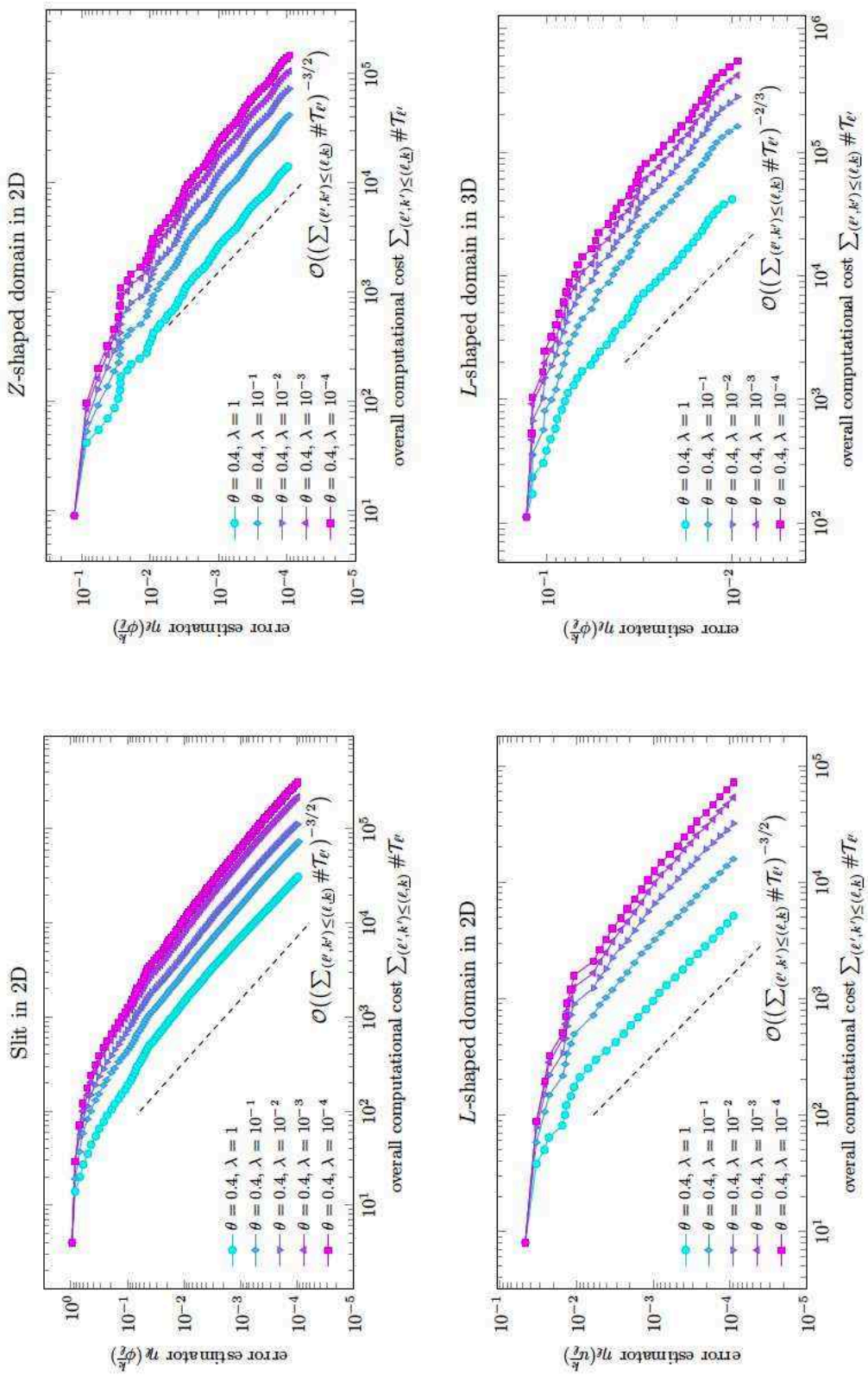


Figure 6.16: Examples from Section 6.7: Error estimator η_ℓ of the last step of the PCG iteration with respect to the cumulative sum $\sum_{(\ell',k')\le(\ell,k)} \#\mathcal{T}_{\ell'}$ for the different experiments of Section 6.7 with $\theta = 0.4$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$.

Figure 6.17: Examples from Section 6.7: Error estimator η_ℓ of the last step of the PCG iteration with respect to the cumulative sum $\sum_{(\ell',k')\le(\ell,k)}(\#\mathcal{T}_{\ell'}\log^2(\#\mathcal{T}_{\ell'})$ for the different experiments of Section 6.7 with $\theta = 0.4$ and $\lambda \in \{1, 10^{-1}, \dots, 10^{-4}\}$.

Bibliography

[AEF+14] Markus Aurada, Michael Ebner, Michael Feischl, Samuel Ferraz-Leite, Thomas Führer, Petra Goldenits, Michael Karkulik, Markus Mayr, and Dirk Praetorius. HILBERT — a MATLAB implementation of adaptive 2D-BEM. *Numer. Algorithms*, 67(1):1–32, 2014.

[AFF+13] Markus Aurada, Michael Feischl, Thomas Führer, Michael Karkulik, and Dirk 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.

[AFF+17] Markus Aurada, Michael Feischl, Thomas Führer, Michael Karkulik, J. Markus Melenk, and Dirk Praetorius. Local inverse estimates for non-local boundary integral operators. *Math. Comp.*, 86(308):2651–2686, 2017.

[AGL13] Mario Arioli, Emmanuil H. Georgoulis, and Daniel Loghin. Stopping criteria for adaptive finite element solvers. *SIAM J. Sci. Comput.*, 35(3):A1537–A1559, 2013.

[AGS16] Mark Ainsworth, Johnny Guzmán, and Francisco-Javier Sayas. Discrete extension operators for mixed finite element spaces on locally refined meshes. *Math. Comp.*, 85(302):2639–2650, 2016.

[ALMS13] Mario Arioli, Jörg Liesen, Agnieszka Międlar, and Zdeněk Strakoš. Interplay between discretization and algebraic computation in adaptive numerical solution of elliptic PDE problems. *GAMM-Mitt.*, 36(1):102–129, 2013.

[AMT99] Mark Ainsworth, William McLean, and Thanh Tran. The conditioning of boundary element equations on locally refined meshes and preconditioning by diagonal scaling. *SIAM J. Numer. Anal.*, 36(6):1901–1932, 1999.

[AO11] Mark Ainsworth and J. Tinsley Oden. *A posteriori error estimation in finite element analysis*. John Wiley & Sons, New York, 2011.

[AW15] Mario Amrein and Thomas P. Wihler. Fully adaptive Newton-Galerkin methods for semilinear elliptic partial differential equations. *SIAM J. Sci. Comput.*, 37(4):A1637–A1657, 2015.

[BCL15] Roland Becker, Daniela Capatina, and Robert Luce. Stopping criteria based on locally reconstructed fluxes. In *Numerical mathematics and advanced applications—ENUMATH 2013*, volume 103 of *Lect. Notes Comput. Sci. Eng.*, pages 243–251. Springer, Cham, 2015.

[BDD04] Peter Binev, Wolfgang Dahmen, and Ron DeVore. Adaptive finite element methods with convergence rates. *Numer. Math.*, 97(2):219–268, 2004.

[BDK12] Liudmila Belenki, Lars Diening, and Christian Kreuzer. Optimality of an adaptive finite element method for the p-Laplacian equation. *IMA J. Numer. Anal.*, 32(2):484–510, 2012.

[BDMS15] Christine Bernardi, Jad Dakroub, Gihane Mansour, and Toni Sayah. A posteriori analysis of iterative algorithms for a nonlinear problem. *J. Sci. Comput.*, 65(2):672–697, 2015.

[BGMP16] Annalisa Buffa, Carlotta Giannelli, Philipp Morgenstern, and Daniel Peterseim. Complexity of hierarchical refinement for a class of admissible mesh configurations. *Comput. Aided Geom. Design*, 47:83–92, 2016.

[BHP17] Alex Bespalov, Alexander Haberl, and Dirk 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.

[BMS10] Roland Becker, Shipeng Mao, and Zhongci Shi. A convergent nonconforming adaptive finite element method with quasi-optimal complexity. *SIAM J. Numer. Anal.*, 47(6):4639–4659, 2010.

[BN10] Andrea Bonito and Ricardo H. Nochetto. Quasi-optimal convergence rate of an adaptive discontinuous Galerkin method. *SIAM J. Numer. Anal.*, 48(2):734–771, 2010.

[BPS02] James Bramble, Joseph Pasciak, and Olaf Steinbach. On the stability of the L^2 projection in $H^1(\Omega)$. *Math. Comp.*, 71:147–156, 01 2002.

[BS02] Susanne C. Brenner and L. Ridgway Scott. *The mathematical theory of finite element methods*. Springer, New York, second edition, 2002.

[Car97] Carsten Carstensen. An a posteriori error estimate for a first-kind integral equation. *Math. Comp.*, 66(217):139–155, 1997.

[Car02] Carsten Carstensen. Merging the Bramble-Pasciak-Steinbach and the Crouzeix-Thomée criterion for H^1 -stability of the L^2 -projection onto finite element spaces. *Math. Comp.*, 71:157–163, 01 2002.

[CDD03] Albert Cohen, Wolfgang Dahmen, and Ronald Devore. Adaptive wavelet schemes for nonlinear variational problems. *SIAM J. Numer. Anal.*, 41(5):1785–1823, 2003.

[CFPP14] Carsten Carstensen, Michael Feischl, Marcus Page, and Dirk Praetorius. Axioms of adaptivity. *Comput. Math. Appl.*, 67(6):1195–1253, 2014.

[CG12] Carsten Carstensen and Joscha Gedicke. An adaptive finite element eigenvalue solver of asymptotic quasi-optimal computational complexity. *SIAM J. Numer. Anal.*, 50(3):1029–1057, 2012.

[CKNS08] J. Manuel Cascón, Christian Kreuzer, Ricardo H. Nochetto, and Kunibert G. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46(5):2524–2550, 2008.

[CMPS04] Carsten Carstensen, M. Maischak, D. Praetorius, and Ernst P. Stephan. Residual-based a posteriori error estimate for hypersingular equation on surfaces. *Numer. Math.*, 97(3):397–425, 2004.

[CMS01] Carsten Carstensen, Matthias Maischak, and Ernst P. Stephan. A posteriori error estimate and h-adaptive algorithm on surfaces for Symm's integral equation. *Numer. Math.*, 90(2):197–213, 2001.

[CN12] J. Manuel Cascón and Ricardo H. Nochetto. Quasioptimal cardinality of AFEM driven by nonresidual estimators. *IMA J. Numer. Anal.*, 32(1):1–29, 2012.

[CNX12] Long Chen, Ricardo H. Nochetto, and Jinchao Xu. Optimal multilevel methods for graded bisection grids. *Numer. Math.*, 120(1):1–34, 2012.

[CP06] Carsten Carstensen and Dirk Praetorius. Averaging techniques for the effective numerical solution of Symm's integral equation of the first kind. *SIAM J. Sci. Comput.*, 27(4):1226–1260, 2006.

[CPV14] Clément Cancès, Iuliu S. Pop, and Martin Vohralík. An a posteriori error estimate for vertex-centered finite volume discretizations of immiscible incompressible two-phase flow. *Math. Comp.*, 83(285):153–188, 2014.

[CS95] Carsten Carstensen and Ernst P. Stephan. A posteriori error estimates for boundary element methods. *Math. Comp.*, 64(210):483–500, 1995.

[CS07] Alexandra L. Chaillou and Manil Suri. A posteriori estimation of the linearization error for strongly monotone nonlinear operators. *J. Comput. Appl. Math.*, 205(1):72–87, 2007.

[CT87] Michel Crouzeix and Vidar Thomée. The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces. *Math. Comp.*, 48(178):521–532, 1987.

[CW17] Scott Congreve and Thomas P. Wihler. Iterative Galerkin discretizations for strongly monotone problems. *J. Comput. Appl. Math.*, 311:457–472, 2017.

[Deu91] Peter Deuflhard. Global inexact Newton methods for very large scale nonlinear problems. *Impact Comput. Sci. Engrg.*, 3(4):366–393, 1991.

[DFFGP19] Giovanni Di Fratta, Thomas Führer, Gregor Gantner, and Dirk Praetorius. Adaptive Uzawa algorithm for the Stokes equation. *ESAIM Math. Model. Numer. Anal.*, 53(6):1841–1870, 2019.

[DK08] Lars Diening and Christian Kreuzer. Linear convergence of an adaptive finite element method for the p-Laplacian equation. *SIAM J. Numer. Anal.*, 46(2):614–638, 2008.

[Dör96] Willy Dörfler. A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996.

[DPVY15] Daniele A. Di Pietro, Martin Vohralík, and Soleiman Yousef. Adaptive regularization, linearization, and discretization and a posteriori error control for the two-phase Stefan problem. *Math. Comp.*, 84(291):153–186, 2015.

[EAEV11] Linda El Alaoui, Alexandre Ern, and Martin 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.

[EV13] Alexandre Ern and Martin Vohralík. Adaptive inexact Newton methods with a posteriori stopping criteria for nonlinear diffusion PDEs. *SIAM J. Sci. Comput.*, 35(4):A1761–A1791, 2013.

[EW94] Stanley C. Eisenstat and Homer F. Walker. Globally convergent inexact Newton methods. *SIAM J. Optim.*, 4(2):393–422, 1994.

[Fei15] Michael Feischl. *Rate optimality of adaptive algorithms*. PhD thesis, TU Wien, Institute of Analysis and Scientific Computing, Wien, 2015.

[FFK+14] Michael Feischl, Thomas Führer, Michael Karkulik, J. Markus Melenk, and Dirk Praetorius. Quasi-optimal convergence rates for adaptive boundary element methods with data approximation. Part I: Weakly-singular integral equation. *Calcolo*, 51:531–562, 2014.

[FFK+15] Michael Feischl, Thomas Führer, Michael Karkulik, J. Markus Melenk, and Dirk Praetorius. Quasi-optimal convergence rates for adaptive boundary element methods with data approximation. Part II: Hyper-singular integral equation. *Electron. Trans. Numer. Anal.*, 44:153–176, 2015.

[FFP14] Michael Feischl, Thomas Führer, and Dirk 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.

[FFPS17a] Michael Feischl, Thomas Führer, Dirk Praetorius, and Ernst P. Stephan. Optimal additive Schwarz preconditioning for hypersingular integral equations on locally refined triangulations. *Calcolo*, 54(1):367–399, 2017.

[FFPS17b] Michael Feischl, Thomas Führer, Dirk Praetorius, and Ernst P. Stephan. Optimal preconditioning for the symmetric and nonsymmetric coupling of adaptive finite elements and boundary elements. *Numer. Methods Partial Differential Equations*, 33(3):603–632, 2017.

[FHPS19] Thomas Führer, Alexander Haberl, Dirk Praetorius, and Stefan Schimanko. Adaptive BEM with inexact PCG solver yields almost optimal computational costs. *Numer. Math.*, 141:967–1008, 2019.

[FKMP13] Michael Feischl, Michael Karkulik, J. Markus Melenk, and Dirk Praetorius. Quasi-optimal convergence rate for an adaptive boundary element method. *SIAM J. Numer. Anal.*, 51:1327–1348, 2013.

[FMPR15] Thomas Führer, J. Markus Melenk, Dirk Praetorius, and Alexander Rieder. Optimal additive Schwarz methods for the hp-BEM: the hypersingular integral operator in 3D on locally refined meshes. *Comput. Math. Appl.*, 70:1583–1605, 2015.

[Füh14] Thomas Führer. *Zur Kopplung von finiten Elementen und Randelementen.* PhD thesis, TU Wien, Institute of Analysis and Scientific Computing, Wien, 2014.

[Gan13] Tsogtgerel Gantumur. Adaptive boundary element methods with convergence rates. *Numer. Math.*, 124(3):471–516, 2013.

[GHP17] Gregor Gantner, Daniel Haberlik, and Dirk Praetorius. Adaptive IGAFEM with optimal convergence rates: Hierarchical B-splines. *Math. Models Methods Appl. Sci.*, 27(14):2631–2674, 2017.

[GHPS18] Gregor Gantner, Alexander Haberl, Dirk Praetorius, and Bernhard Stiftner. Rate optimal adaptive FEM with inexact solver for nonlinear operators. *IMA J. Numer. Anal.*, 38:1797–1831, 2018.

[GHPS21] Gregor Gantner, Alexander Haberl, Dirk Praetorius, and Stefan Schimanko. Rate optimality of adaptive finite element methods with respect to the overall computational costs. *Math. Comp., accepted for publication*, 2021.

[GMZ11] Eduardo M. Garau, Pedro Morin, and Carlos Zuppa. Convergence of an adaptive Kačanov FEM for quasi-linear problems. *Appl. Numer. Math.*, 61(4):512–529, 2011.

[GMZ12] Eduardo M. Garau, Pedro Morin, and Carlos 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.

[GO94] Michael Griebel and Peter Oswald. On additive Schwarz preconditioners for sparse grid discretizations. *Numer. Math.*, 66(4):449–463, 1994.

[GSS14] Dietmar Gallistl, Mira Schedensack, and Rob P. Stevenson. A remark on newest vertex bisection in any space dimension. *Comput. Methods Appl. Math.*, 14(3):317–320, 2014.

[GVL13] Gene H. Golub and Charles F. Van Loan. *Matrix computations.* Johns Hopkins University Press, Baltimore, fourth edition, 2013.

[Hac15] Wolfgang Hackbusch. *Hierarchical matrices: Algorithms and analysis.* Springer, Heidelberg, 2015.

[HJHM15] Ralf Hiptmair, Carlos Jerez-Hanckes, and Shipeng Mao. Extension by zero in discrete trace spaces: inverse estimates. *Math. Comp.*, 84(296):2589–2615, 2015.

[HM12] Ralf Hiptmair and Shipeng Mao. Stable multilevel splittings of boundary edge element spaces. *BIT*, 52(3):661–685, 2012.

[HPSV21] Alexander Haberl, Dirk Praetorius, Stefan Schimanko, and Martin Vohralík. Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver. *Numer. Math.*, pages 1–47, 2021.

[HW08] George C. Hsiao and Wolfgang L. Wendland. *Boundary integral equations.* Springer, Berlin, 2008.

[HW18] Paul Houston and Thomas P. Wihler. An hp-adaptive Newton-discontinuous-Galerkin finite element approach for semilinear elliptic boundary value problems. *Math. Comp.*, 87(314):2641–2674, 2018.

[HW20a] Pascal Heid and Thomas P. Wihler. Adaptive iterative linearization Galerkin methods for nonlinear problems. *Math. Comp.*, 89:2707–2734, 2020.

[HW20b] Pascal Heid and Thomas P. Wihler. On the convergence of adaptive iterative linearized Galerkin methods. *Calcolo*, 57:24, 2020.

[HWZ12] Ralf Hiptmair, Haijun Wu, and Weiying Zheng. Uniform convergence of adaptive multigrid methods for elliptic problems and Maxwell's equations. *Numer. Math. Theory Methods Appl.*, 5(3):297–332, 2012.

[HZ09] Ralf Hiptmair and Weiying Zheng. Local multigrid in H(curl). *J. Comput. Math.*, 27(5):573–603, 2009.

[KPP13] Michael Karkulik, David Pavlicek, and Dirk 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.

[KS08] Yaroslav Kondratyuk and Rob Stevenson. An optimal adaptive finite element method for the Stokes problem. *SIAM J. Numer. Anal.*, 46(2):747–775, 2008.

[Lio88] Pierre-Louis Lions. On the Schwarz alternating method. I. In *First International Symposium on Domain Decomposition Methods for Partial Differential Equations (Paris, 1987)*, pages 1–42. SIAM, Philadelphia, PA, 1988.

[McL00] William McLean. *Strongly elliptic systems and boundary integral equations.* Cambridge University Press, Cambridge, 2000.

[MNS00] Pedro Morin, Ricardo H. Nochetto, and Kunibert G. Siebert. Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.*, 38(2):466–488, 2000.

[MP15] Philipp Morgenstern and Daniel Peterseim. Analysis-suitable adaptive t-mesh refinement with linear complexity. *Comput. Aided Geom. Design*, 34:50–66, 2015.

[Osw94] Peter Oswald. *Multilevel finite element approximation.* B. G. Teubner, Stuttgart, 1994.

[Osw99] Peter Oswald. Interface preconditioners and multilevel extension operators. In *Eleventh International Conference on Domain Decomposition Methods (London, 1998)*, pages 97–104. DDM.org, Augsburg, 1999.

[OT14] Maxim A. Olshanskii and Eugene E. Tyrtyshnikov. *Iterative methods for linear systems.* Society for Industrial and Applied Mathematics, Philadelphia, PA, 2014.

[Pol16] Sara Pollock. Stabilized and inexact adaptive methods for capturing internal layers in quasilinear PDE. *J. Comput. Appl. Math.*, 308:243–262, 2016.

[PP20] Carl-Martin Pfeiler and Dirk Praetorius. Dörfler marking with minimal cardinality is a linear complexity problem. *Math. Comp.*, 89(326):2735–2752, 2020.

[SBA+13] Wojciech Śmigaj, Timo Betcke, Simon Arridge, Joel Phillips, and Martin Schweiger. Solving boundary integral problems with BEM++. *ACM Trans. Math. Softw.*, 2013.

[SMPZ08] Joachim Schöberl, J. Markus Melenk, Clemens Pechstein, and Sabine Zaglmayr. Additive Schwarz preconditioning for p-version triangular and tetrahedral finite elements. *IMA J. Numer. Anal.*, 28(1):1–24, 2008.

[SS11] Stefan A. Sauter and Christoph Schwab. *Boundary element methods.* Springer, Berlin, 2011.

[Ste07] Rob Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007.

[Ste08] Rob Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comp.*, 77(261):227–241, 2008.

[Ste14] Rob Stevenson. Adaptive wavelet methods for linear and nonlinear least-squares problems. *Found. Comput. Math.*, 14(2):237–283, 2014.

[SvV20] Rob Stevenson and Raymond van Venetië. Uniform preconditioners for problems of negative order. *Math. Comp.*, 89(322):645–674, 2020.

[SZ90] L. Ridgway Scott and Shangyou Zhang. Finite element interpolation of non-smooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[Tri95] Hans Triebel. *Interpolation theory, function spaces, differential operators.* Johann Ambrosius Barth, Heidelberg, second edition, 1995.

[TW05] Andrea Toselli and Olof Widlund. *Domain decomposition methods—algorithms and theory.* Springer, Berlin, 2005.

[Vee02] Andreas Veeser. Convergent adaptive finite elements for the nonlinear Laplacian. *Numer. Math.*, 92(4):743–770, 2002.

[Ver13] Rüdiger Verfürth. *A posteriori error estimation techniques for finite element methods.* Oxford University Press, Oxford, 2013.

[WC06] Haijun Wu and Zhiming Chen. Uniform convergence of multigrid V-cycle on adaptively refined finite element meshes for second order elliptic problems. *Sci. China Math.*, 49(10):1405–1429, 2006.

[Wid89] Olof B. Widlund. Optimal iterative refinement methods. In *Domain decomposition methods (Los Angeles, CA, 1988)*, pages 114–125. SIAM, Philadelphia, PA, 1989.

[XCH10] Xuejun Xu, Huangxin Chen, and Ronald H. W. Hoppe. Optimality of local multilevel methods on adaptively refined meshes for elliptic boundary value problems. *J. Numer. Math.*, 18(1):59–90, 2010.

[XCN09] Jinchao Xu, Long Chen, and Ricardo H. Nochetto. Optimal multilevel methods for h(grad), h(curl), and h(div) systems son graded and unstructured grids. In *Multiscale, nonlinear and adaptive approximation*, pages 599–659. Springer, Berlin, 2009.

[Xu96] Jinchao Xu. *An Introduction to Multigrid Convergence Theory.* CAM report. Department of Mathematics, University of California, Los Angeles, 1996.

[Yos80] Kosaku Yosida. *Functional analysis.* Springer, Berlin, 1980.

[Zei90] Eberhard Zeidler. *Nonlinear functional analysis and its applications. II/B.* Springer, New York, 1990.

Curriculum Vitae

Personal Data

Name	**Stefan Schimanko**
Date of birth	[redacted]
Place of birth	Amstetten / Lower Austria
Citizenship	Austria
Email	stefan.schimanko@asc.tuwien.ac.at
Homepage	http://www.asc.tuwien.ac.at/~sschiman/

Education

since 01/2017	PhD student, supervised by Dirk Praetorius, Institute of Analysis and Scientific Computing, TU Wien
10/2013–11/2016	Master studies in Technical Mathematics, TU Wien
10/2009–10/2013	Bachelor studies in Technical Mathematics, TU Wien
09/2000–06/2008	BG Amstetten

Scientific Publications

G. Gantner, A. Haberl, D. Praetorius, S. Schimanko: *Rate optimality of adaptive finite element methods with respect to the overall computational costs*, accepted for publication in Mathematics of Computation (2021).

A. Haberl, D. Praetorius, S. Schimanko, M. Vohralík: *Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver*, Numerische Mathematik, 147 (2021), 679–725.

G. Gantner, D. Praetorius, S. Schimanko: *Adaptive isogeometric boundary element methods with local smoothness control*, Mathematical Models and Methods in Applied Sciences (M3AS), 30 (2020), 261–307.

T. Führer, G. Gantner, D. Praetorius, S. Schimanko: *Optimal additive Schwarz preconditioning for adaptive 2D IGA boundary element methods*, Computer Methods in Applied Mechanics and Engineering (CMAME), 351 (2019), 571–598.

T. Führer, A. Haberl, D. Praetorius, S. Schimanko: *Adaptive BEM with inexact PCG solver yields almost optimal computational costs*, Numerische Mathematik, 141 (2019), 967–1008.

S. Schimanko (Supervisor: G. Gantner, D. Praetorius): *Adaptive isogeometric boundary element method for the hyper-singular integral equation*, Master's thesis, Institute of Analysis and Scientific Computing, TU Wien, 2016.

S. Schimanko (Supervisor: M. Ludwig): *Voronoi Diagrams*, Bachelor's thesis, Institute of Discrete Mathematics and Geometry, TU Wien, 2013.

Scientific Talks

Rate optimal adaptive FEM with inexact solver for nonlinear operators, ENUMATH 2019 – European Numerical Mathematics and Advanced Applications Conference 2019, Egmond aan Zee, Netherlands, 30.09.2019–04.10.2019.

Adaptive BEM with inexact PCG solver yields almost optimal computational costs, SIAM Conference on Computational Science and Engineering (CSE 2019), Spokane, Washington, USA, 25.02.2019–01.03.2019.

Adaptive BEM with inexact PCG solver yields almost optimal computational costs, IABEM 2018 – Symposium of the International Association for Boundary Element Methods, Paris, France, 26.06.2018–29.06.2018.

Adaptive BEM with inexact PCG solver yields almost optimal computational costs, 14th Austrian Numerical Analysis Day, Klagenfurt, Austria, 03.05.2018–04.05.2018.

Adaptive isogeometric boundary element method, Masters in Research, Wien, Austria, 29.03.2017.

Wien, June 23, 2021

Stefan Schimanko