



HAL
open science

Adaptive finite element methods with optimally preconditioned GMRES guarantee optimal complexity

Thomas Führer, Paula Hilbert, Ani Miraçi, Dirk Praetorius

► To cite this version:

Thomas Führer, Paula Hilbert, Ani Miraçi, Dirk Praetorius. Adaptive finite element methods with optimally preconditioned GMRES guarantee optimal complexity. 2026. ⟨hal-05596791⟩

HAL Id: hal-05596791

<https://hal.science/hal-05596791v1>

Preprint submitted on 20 Apr 2026

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire HAL, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons CC BY 4.0 - Attribution - International License

ADAPTIVE FINITE ELEMENT METHODS WITH OPTIMALLY PRECONDITIONED GMRES GUARANTEE OPTIMAL COMPLEXITY

THOMAS FÜHRER[Ⓛ], PAULA HILBERT[Ⓛ], ANI MIRAÇI[Ⓛ], AND DIRK PRAETORIUS[Ⓛ]

ABSTRACT. We analyze optimal complexity of adaptive finite element methods (AFEMs) for general second-order linear elliptic partial differential equations (PDEs) in the Lax–Milgram setting. To this end, we formulate an adaptive algorithm which steers the local mesh-refinement as well as the termination of a generalized minimal residual solver (GMRES) with optimal preconditioner to solve the arising non-symmetric finite element systems. Algorithmic interplay of mesh-refinement and iterative solver is shown to be optimal: A natural and fully computable quasi-error monitoring discretization error and algebraic solver error guarantees unconditional convergence for any choice of adaptivity parameters, i.e., the algorithm cannot fail to converge. This is ensured algorithmically via a novel adaptive feedback-control for the solver-termination parameter that monitors and ensures full R-linear convergence. Finally, the quasi-error even decays with optimal rates with respect to the overall computational complexity if the adaptivity parameters are chosen sufficiently small.

1. INTRODUCTION

The mathematical understanding of adaptive finite element methods (AFEMs) has matured over the last decades. The seminal works [Dör96; MNS00; BDD04; Ste07; CKNS08] proved optimal convergence rates of AFEM for symmetric second-order linear elliptic partial differential equations (PDEs). Therein, developments on optimal rates were mainly initiated by corresponding results for adaptive wavelet FEM [CDD01; CDD03]. This was later extended to general non-symmetric second-order linear elliptic PDEs [MN05; CN12; FFP14] and put into the abstract framework of the *axioms of adaptivity* [CFPP14], which also covered available AFEM results for nonlinear PDEs like, e.g., [GMZ12; BDK12].

While the first works [CDD01; CDD03] and [Ste07] also discussed optimal complexity of AFEM, i.e., the optimal decay of the error (resp. error estimator) with respect to the overall computational cost, main developments started by [CKNS08; CFPP14] were focusing on the optimal decay of the error (resp. error estimator) with respect to the number of degrees of freedom. Beyond 1D, where usual finite element (FE) matrices have a block-tridiagonal structure and can hence be solved at linear cost with respect to the number of degrees of freedom, optimal complexity of AFEM can only be achieved by including an iterative algebraic solver. Moreover, optimal complexity indeed requires a thorough interplay of the local mesh refinement with an appropriate iterative solver. While this is most important for nonlinear PDEs, it should also not be neglected for

2020 *Mathematics Subject Classification.* 41A25, 65N15, 65N30, 65F10, 65Y20, 68W40.

Key words and phrases. adaptive finite element method, non-symmetric problem, optimal convergence rates, cost-optimality, inexact solver, full linear convergence.

Acknowledgment. This research was funded by the Austrian Science Fund (FWF) projects [10.55776/F65](#) (SFB F65 “Taming complexity in PDE systems”), [10.55776/I6802](#) (international project “Functional error estimates for PDEs on unbounded domains”), [10.55776/PAT3446525](#) (standalone project “Adaptive Uzawa-type FEM for nonlinear PDEs”), and [10.55776/PAT3699424](#) (standalone project “Optimal robust solvers for reliable and efficient AFEMs”) and by ANID FONDECYT Regular 1250070 project.

linear PDEs. Recent works [GHPS21; BMP24; BFM⁺25] laid the paths of a general theory for such an interplay, where, however, the main focus was on symmetric PDEs (resp. nonlinear PDEs, where the linearization leads to symmetric PDE formulations).

A key ingredient in the analysis of [GHPS21] is that the iterative solver guarantees uniform contraction with respect to the variational structure of the PDE, i.e., the PDE-related energy norm. In the case of symmetric linear elliptic PDEs, admissible solvers include preconditioned conjugate gradients (PCG) with optimal multilevel additive Schwarz preconditioners [CNX12], geometric multigrid methods tailored to the multilevel structure of the adaptive mesh hierarchy [WZ17; IMPS24], or generalized PCG with multigrid preconditioners [HMP26]. As far as non-symmetric second-order linear elliptic PDEs are concerned, the work [BIM⁺24] formulates an AFEM strategy with optimal complexity for PDEs in the Lax–Milgram setting. However, adapting ideas from optimal-complexity AFEM for nonlinear PDEs [HPW21; HPSV21], the iterative solver in [BIM⁺24] relies on a nested combination of the Zarantonello iteration (to symmetrize the nonsymmetric linear FE systems) with an optimal multigrid method (to solve the symmetrized linear FE system). Then, [BIM⁺24] proves optimal complexity provided that the user-chosen parameters are sufficiently small: The damping $\delta > 0$ of the Zarantonello iteration, the marking parameter $0 < \theta \leq 1$ as well as the parameters $\lambda_{\text{sym}}, \lambda_{\text{alg}} > 0$ for terminating the Zarantonello iteration and the multigrid solver. However, convergence fails to be unconditional, since only θ can be chosen arbitrarily. In explicit terms, the algorithm from [BIM⁺24] might fail to converge if $\lambda_{\text{sym}}, \lambda_{\text{alg}}$, or δ are chosen too large. Moreover, the standard strategy for solving the nonsymmetric linear FE systems arising for PDEs in the Lax–Milgram setting appears to be GMRES [SS86; EES83]. However, without preconditioning, GMRES convergence can be slow or even fail [GPS96]. To improve convergence, a dedicated preconditioner for the symmetric part of the PDE can be used [Spi24], as well as using problem-adapted inner products or weights [PW13; GP14; SS24].

This work focuses on AFEM with GMRES and its unconditional convergence for general second-order linear elliptic PDEs. At the solver level, we adopt the left-preconditioning framework of [SS07] and use an optimal symmetric additive Schwarz preconditioner which additionally serves as weight. From [HMP26], the resulting weighted preconditioned GMRES method contracts the preconditioned algebraic residual robustly (in the local mesh-size h and in the polynomial degree p) in the preconditioner-weighted discrete norm. While this discrete contraction is crucial, as is, it is incompatible with the current AFEM analysis [BFM⁺25] as the latter requires contraction of the algebraic error in the PDE-induced energy norm. Moreover, to comply with the linear complexity of AFEM, GMRES is restarted after a maximum number of iterations and an a-posteriori error criterion determines when to stop the overall solver and when to proceed with mesh-refinement. Overall, the manuscript presents the following novelties: (i) an adaptive algorithm for non-symmetric problems without nested solvers, (ii) a new a-posteriori parameter control criterion included into the adaptive loop, (iii) a computable quasi-error monitoring the accuracy of the (different sources of) error, (iv) the proof of only finitely many parameter adaptation steps. These contributions allow us to prove unconditional full R-linear convergence of the quasi-error independently of the adaptivity parameters provided by the user and optimal convergence rates with respect to total computational cost for sufficiently small adaptivity parameters.

Outline. The manuscript is organized as follows: Section 2 presents the model problem and its FE discretization. Section 3 recaps the preconditioned GMRES solver in Algorithm A and optimal preconditioners for the principal part from [HMP26], and collects

potentially well-known properties like equivalence of the discrete GMRES residuals to the solver errors in the functional setting (Lemma 6) as well as uniform contraction of the discrete residuals (Proposition 7). Section 4 formulates the new AFEM algorithm with GMRES solver (Algorithm B) with the corresponding quasi-error and a-posteriori error control in Lemma 14. Unconditional convergence of Algorithm B is stated and proved in Section 5; see Theorem 18. Optimal complexity is formulated in Section 6; see Theorem 21. Finally, numerical experiments in Section 7 underline the developed theory.

2. SETTING

2.1. Model problem. On a polyhedral Lipschitz domain $\Omega \subset \mathbb{R}^d$ with $d \geq 1$ and given $f \in L^2(\Omega)$ and $\mathbf{f} \in [L^2(\Omega)]^d$, we consider the diffusion-convection-reaction equation

$$-\operatorname{div}(\mathbf{K}\nabla u^*) + \mathbf{b} \cdot \nabla u^* + cu^* = f - \operatorname{div} \mathbf{f} \quad \text{in } \Omega \quad \text{with } u^* = 0 \quad \text{on } \partial\Omega. \quad (1)$$

With the Hilbert space $\mathcal{X} := H_0^1(\Omega)$ and $a(v, w) := \langle \mathbf{K} \nabla v, \nabla w \rangle_{L^2(\Omega)}$ denoting the principal part of the bilinear form $b: \mathcal{X} \times \mathcal{X} \rightarrow \mathbb{R}$ defined by

$$b(v, w) := a(v, w) + \langle \mathbf{b} \cdot \nabla v, w \rangle_{L^2(\Omega)} + \langle cv, w \rangle_{L^2(\Omega)} \quad \text{for all } v, w \in \mathcal{X},$$

the weak formulation of (1) reads as follows: Find $u^* \in \mathcal{X}$ such that

$$b(u^*, v) = \langle f, v \rangle_{L^2(\Omega)} + \langle \mathbf{f}, \nabla v \rangle_{L^2(\Omega)} =: F(v) \quad \text{for all } v \in \mathcal{X}. \quad (2)$$

To ensure well-posedness of this problem, we assume that $\mathbf{b} \in [L^\infty(\Omega)]^d$ and $c \in L^\infty(\Omega)$, that the matrix-valued function $\mathbf{K} \in [L^\infty(\Omega)]^{d \times d}$ is symmetric with smallest eigenvalue uniformly bounded from below, and that $\operatorname{div} \mathbf{b} \in L^\infty(\Omega)$ with

$$-\operatorname{div}(\mathbf{b}) + 2c \geq 0 \quad \text{a.e. in } \Omega.$$

Under these assumptions, the bilinear form $b(\cdot, \cdot)$ is coercive and bounded on \mathcal{X} with respect to the energy norm $\|v\| := a(v, v)^{1/2}$, i.e., there exist $0 < C_{\text{ell}} \leq C_{\text{bnd}}$ such that

$$|b(v, w)| \leq C_{\text{bnd}} \|v\| \|w\| \quad \text{and} \quad b(v, v) \geq C_{\text{ell}} \|v\|^2 \quad \text{for all } v, w \in \mathcal{X}. \quad (3)$$

The Lax–Milgram lemma yields existence and uniqueness of the solution u^* to (2). For the well-posedness of the standard residual-based error estimator (see (7) below), we will require additional regularity $\mathbf{f}|_T \in [H^1(T)]^d$ and $\mathbf{K}|_T \in [W^{1,\infty}(T)]^{d \times d}$ for all $T \in \mathcal{T}_0$ with \mathcal{T}_0 being the initial triangulation of Ω introduced in the following section.

2.2. Finite element method. Let \mathcal{T}_0 be a conforming initial triangulation of Ω consisting of compact simplices. The adaptive mesh refinement employs newest vertex bisection (NVB), where we refer to [Ste08] for NVB with admissible initial triangulation and dimension $d \geq 2$, to [KPP13] for NVB with general \mathcal{T}_0 in the case $d = 2$, to the recent work [DGS25] for NVB with general \mathcal{T}_0 in any dimension $d \geq 2$, and to [AFF⁺15] for NVB in the case $d = 1$. For any triangulation \mathcal{T}_H and any subset $\mathcal{M}_H \subseteq \mathcal{T}_H$, we define $\mathcal{T}_h := \operatorname{refine}(\mathcal{T}_H, \mathcal{M}_H)$ as the coarsest refinement of \mathcal{T}_H by newest vertex bisection such that at least all elements in \mathcal{M}_H are refined, i.e., $\mathcal{M}_H \subseteq \mathcal{T}_H \setminus \mathcal{T}_h$. The notation $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ abbreviates that \mathcal{T}_h can be obtained from \mathcal{T}_H by a finite number of newest vertex bisection steps. To each triangulation \mathcal{T}_H we associate the finite element space

$$\mathcal{X}_H := \mathbb{S}_0^p(\mathcal{T}_H) := \mathbb{S}^p(\mathcal{T}_H) \cap \mathcal{X} \quad (4)$$

consisting of globally continuous \mathcal{T}_H -piecewise polynomials of degree at most $p \in \mathbb{N}$. Note that $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$ implies nestedness of the associated finite element spaces $\mathcal{X}_H \subseteq \mathcal{X}_h$.

The finite element discretization for (2) reads as follows: Find $u_H^* \in \mathcal{X}_H$ such that

$$b(u_H^*, v_H) = F(v_H) \quad \text{for all } v_H \in \mathcal{X}_H. \quad (5)$$

The Lax–Milgram lemma ensures the existence and uniqueness of the solution u_H^* to (5) and the solution is the quasi-best approximation in \mathcal{X}_H to the solution u^* of (2) in the sense that there exists a constant $1 \leq C_{\text{Céa}} \leq C_{\text{bnd}}/C_{\text{ell}}$ such that

$$\| \|u^* - u_H^* \| \| \leq C_{\text{Céa}} \min_{v_H \in \mathcal{X}_H} \| \|u^* - v_H \| \|, \quad (6)$$

where we note that $C_{\text{Céa}} \rightarrow 1$ as $\| \|u^* - u_H^* \| \| \rightarrow 0$; see [BHP17].

2.3. A-posteriori error estimation. Based on the additional regularity of the data, we consider the residual error estimator $\eta_H(\cdot)$ defined, for $T \in \mathcal{T}_H$ and $v_H \in \mathcal{X}_H$, by

$$\begin{aligned} \eta_H(T, v_H)^2 &:= |T|^{2/d} \| \operatorname{div}(\mathbf{K} \nabla v_H - \mathbf{f}) - \mathbf{b} \cdot \nabla v_H - c v_H + f \|_{L^2(T)}^2 \\ &\quad + |T|^{1/d} \| [(\mathbf{K} \nabla v_H - \mathbf{f}) \cdot \mathbf{n}] \|_{L^2(\partial T \cap \Omega)}^2, \end{aligned} \quad (7a)$$

where $[\cdot]$ denotes the jump over the $(d-1)$ -dimensional facets of T . Clearly, the well-posedness of (7a) requires the additional regularity of the data \mathbf{K} and \mathbf{f} stated above. For any subset $\mathcal{U}_H \subseteq \mathcal{T}_H$ and all $v_H \in \mathcal{X}_H$, we define the global error estimator by

$$\eta_H(v_H) := \eta_H(\mathcal{T}_H, v_H) \quad \text{with} \quad \eta_H(\mathcal{U}_H, v_H) := \left(\sum_{T \in \mathcal{U}_H} \eta_H(T, v_H)^2 \right)^{1/2}. \quad (7b)$$

It is well-known that the error estimator satisfies the axioms of adaptivity; see [CFPP14].

Proposition 1 (axioms of adaptivity). *There exist constants $C_{\text{stab}}, C_{\text{rel}}, C_{\text{drel}}, C_{\text{mon}} > 0$ and $0 < q_{\text{red}} < 1$ such that the following properties are satisfied for any triangulation $\mathcal{T}_H \in \mathbb{T}(\mathcal{T}_0)$, any refinement $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$, and arbitrary $v_H \in \mathcal{X}_H$, $v_h \in \mathcal{X}_h$:*

- (A1) **stability:** $|\eta_h(\mathcal{T}_h \cap \mathcal{T}_H, v_h) - \eta_H(\mathcal{T}_h \cap \mathcal{T}_H, v_H)| \leq C_{\text{stab}} \| \|v_h - v_H \| \|;$
- (A2) **reduction:** $\eta_h(\mathcal{T}_h \setminus \mathcal{T}_H, v_H) \leq q_{\text{red}} \eta_H(\mathcal{T}_H \setminus \mathcal{T}_h, v_H);$
- (A3) **reliability:** $\| \|u^* - u_H^* \| \| \leq C_{\text{rel}} \eta_H(u_H^*);$
- (A3⁺) **discrete reliability:** $\| \|u_h^* - u_H^* \| \| \leq C_{\text{drel}} \eta_H(\mathcal{T}_H \setminus \mathcal{T}_h, u_H^*);$
- (QM) **quasi-monotonicity:** $\eta_h(u_h^*) \leq C_{\text{mon}} \eta_H(u_H^*).$

The constant C_{rel} depends only on uniform shape regularity, the dimension d , the ellipticity constant C_{ell} , and $\| \mathbf{K} \|_\infty + \| \mathbf{b} \|_\infty + \| c \|_\infty$, whereas C_{stab} and C_{drel} additionally depend on the polynomial degree p . Moreover, there hold the bounds $q_{\text{red}} \leq 2^{-1/(2d)}$ and $C_{\text{mon}} \leq \min\{1 + C_{\text{stab}}(1 + C_{\text{Céa}})C_{\text{rel}}, 1 + C_{\text{stab}}C_{\text{drel}}\}$. \square

In addition to the estimator properties in Proposition 1, we recall the following quasi-orthogonality result from [Fei22] as one cornerstone of the analysis in this paper.

Proposition 2 (validity of quasi-orthogonality). *There exist $C_{\text{orth}} > 0$ and $0 < \delta \leq 1$ such that the following holds: For any sequence $\mathcal{X}_\ell \subseteq \mathcal{X}_{\ell+1} \subset H_0^1(\Omega)$ of nested finite-dimensional subspaces, the corresponding Galerkin solutions $u_\ell^* \in \mathcal{X}_\ell$ to (5) satisfy*

(A4) **quasi-orthogonality:**

$$\sum_{\ell'=\ell}^{\ell+N} \| \|u_{\ell'+1}^* - u_{\ell'}^* \| \|^2 \leq C_{\text{orth}}(N+1)^{1-\delta} \| \|u^* - u_\ell^* \| \|^2 \quad \text{for all } \ell, N \in \mathbb{N}_0.$$

Here, C_{orth} and δ depend only on the dimension d , the elliptic bilinear form $b(\cdot, \cdot)$, and the chosen norm $\| \| \cdot \| \|$, but are independent of the spaces \mathcal{X}_ℓ . \square

It remains to treat the discrete problem (5). Note that, in the case of a symmetric setting $b(\cdot, \cdot) = a(\cdot, \cdot)$, optimal solvers are well-established; see, e.g., [CNX12] for an optimally preconditioned conjugate gradient method (CG), [WZ17; IMPS24] for optimal local

geometric multigrid methods, and the recent work [HMP26] for multigrid preconditioners for (generalized) CG. In the next section, we revisit the literature on preconditioned GMRES methods and propose an optimally preconditioned GMRES method for the numerical solution of the non-symmetric discrete problem (5).

3. PRECONDITIONED GMRES METHOD (PGMRES)

3.1. Abstract setting. Fix $\ell \in \mathbb{N}$ and let us consider the discrete problem (5): let $\varphi_{\ell,1}, \dots, \varphi_{\ell,N_\ell}$ be the basis functions of \mathcal{X}_ℓ with $N_\ell := \dim \mathcal{X}_\ell$. We define the Galerkin matrix $\mathbf{B}_\ell \in \mathbb{R}^{N_\ell \times N_\ell}$ associated with the bilinear form $b(\cdot, \cdot)$ by $(\mathbf{B}_\ell)_{jk} := b(\varphi_{\ell,k}, \varphi_{\ell,j})$ for all $j, k = 1, \dots, N_\ell$. Similarly, discretizing the right-hand side leads to $\mathbf{d}_\ell \in \mathbb{R}^{N_\ell}$. The arising linear problem from the finite element discretization aims to find $\mathbf{x}_\ell^* \in \mathbb{R}^{N_\ell}$ solving

$$\mathbf{B}_\ell \mathbf{x}_\ell^* = \mathbf{d}_\ell. \quad (8)$$

Let $\mathbf{P}_\ell \in \mathbb{R}^{N_\ell \times N_\ell}$ be an SPD preconditioner of \mathbf{B}_ℓ , i.e., $\mathbf{P}_\ell \approx \mathbf{B}_\ell^{-1}$ is symmetric and positive definite. Note that $(\cdot, \cdot)_{\mathbf{P}_\ell^{-1}} := (\mathbf{P}_\ell^{-1} \cdot, \cdot)_2$ then defines a scalar product on \mathbb{R}^{N_ℓ} . We suppose that \mathbf{P}_ℓ is optimal in the following sense: first, the matrix-vector multiplication with \mathbf{P}_ℓ can be realised in linear complexity; second, the discrete norm induced by the inverse is equivalent to the PDE-related energy norm $\|\cdot\|$ defined in Section 2.1, i.e., there exists a constant $C_{\text{pre}} \geq 1$ independent of the discretization parameters, i.e., the local mesh size h and the polynomial degree p , such that

$$C_{\text{pre}}^{-1} \|\|v_\ell\|\| \leq \|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}} := (\mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{P}_\ell^{-1}}^{1/2} \leq C_{\text{pre}} \|\|v_\ell\|\| \quad \text{for all } v_\ell = \sum_{j=1}^{N_\ell} (\mathbf{x}_\ell)_j \varphi_{\ell,j} \in \mathcal{X}_\ell. \quad (9)$$

We now formulate the optimally preconditioned GMRES algorithm for (8). It is derived from [SS07, Algorithm 3.1], where all the scalar products are replaced by the preconditioner inner product $(\cdot, \cdot)_{\mathbf{P}_\ell^{-1}}$. Moreover, a restart strategy is included to ensure that the computational cost of the algorithm is linear in the number of degrees of freedom. To make the computations explicit at step k of the algorithm, we compute and store the Arnoldi basis vectors denoted by \mathbf{v}^j as well as the unpreconditioned Arnoldi basis vectors denoted by $\tilde{\mathbf{v}}^j$ for the current run, i.e., these vectors are deleted after every restart.

Algorithm A (optimally preconditioned GMRES method with restart). *Input:* Matrix \mathbf{B}_ℓ , preconditioner matrix \mathbf{P}_ℓ , right-hand side vector \mathbf{d}_ℓ , initial guess \mathbf{x}_ℓ^0 , absolute tolerance function $\Lambda : \mathbb{R}^{N_\ell} \rightarrow \mathbb{R}_{\geq 0}$, PGMRES restart length $k_{\text{max}} \in \mathbb{N}$.

- (a) Compute $\mathbf{r}_\ell^0 := \mathbf{d}_\ell - \mathbf{B}_\ell \mathbf{x}_\ell^0$, $\mathbf{s}_\ell^0 := \mathbf{P}_\ell \mathbf{r}_\ell^0$, $\tilde{\mathbf{v}}^1 := \mathbf{r}_\ell^0 / \|\mathbf{s}_\ell^0\|_{\mathbf{P}_\ell^{-1}} = \mathbf{r}_\ell^0 / (\mathbf{s}_\ell^0, \mathbf{r}_\ell^0)_2^{1/2}$, $\mathbf{v}^1 := \mathbf{P}_\ell \tilde{\mathbf{v}}^1$.
- (b) For all $k = 1, 2, 3, \dots$ repeat the following steps (i)–(vi) **until**

$$\|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} \leq \Lambda(\mathbf{x}_\ell^k) \quad \text{and then set } \underline{k}[\ell] := k. \quad (10)$$

- (i) Compute $\tilde{\mathbf{w}} := \mathbf{B}_\ell \mathbf{v}^k$, $\mathbf{w} := \mathbf{P}_\ell \tilde{\mathbf{w}}$ and set $K := \text{mod}(k-1, k_{\text{max}}) + 1$ as well as the restart index $R := \lfloor k/k_{\text{max}} \rfloor$.
- (ii) For all $j = 1, \dots, K$ compute $\mathbf{H}_{jK} := \langle \mathbf{w}, \mathbf{v}^{k+j-K} \rangle_{\mathbf{P}_\ell^{-1}} = \langle \tilde{\mathbf{w}}, \mathbf{v}^{k+j-K} \rangle_2$ and

$$\tilde{\mathbf{w}} := \tilde{\mathbf{w}} - \mathbf{H}_{jK} \tilde{\mathbf{v}}^{k+j-K} \quad \text{and} \quad \mathbf{w} := \mathbf{w} - \mathbf{H}_{jK} \mathbf{v}^{k+j-K}.$$

- (iii) Compute $\mathbf{H}_{K+1,K} := \|\mathbf{w}\|_{\mathbf{P}_\ell^{-1}} = \langle \tilde{\mathbf{w}}, \mathbf{w} \rangle_2^{1/2}$.
- (iv) Compute $\mathbf{y} := \text{argmin}_{\mathbf{z} \in \mathbb{R}^K} \|\|\mathbf{s}_\ell^{Rk_{\text{max}}}\|\|_{\mathbf{P}_\ell^{-1}} \mathbf{e}^1 - \mathbf{H}_{[1:K+1, 1:K]} \mathbf{z}\|_2$.

- (v) Compute $\mathbf{x}_\ell^k := \mathbf{x}_\ell^{Rk_{\max}} + \sum_{j=1}^K \mathbf{y}_j \mathbf{v}^{k+j-K}$, $\mathbf{r}_\ell^k := \mathbf{d}_\ell - \mathbf{B}_\ell \mathbf{x}_\ell^k$, $\mathbf{s}_\ell^k := \mathbf{P}_\ell \mathbf{r}_\ell^k$, and $\|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} = (\mathbf{s}_\ell^k, \mathbf{r}_\ell^k)_2^{1/2}$.
- (vi) If $\text{mod}(k, k_{\max}) = 0$, then define $\tilde{\mathbf{v}}^{k+1} := \mathbf{r}_\ell^k / \|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}}$, $\mathbf{v}^{k+1} := \mathbf{s}_\ell^k / \|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}}$, and delete \mathbf{H} , \mathbf{v}^j , and $\tilde{\mathbf{v}}^j$ for $j \leq k$; Else, compute $\tilde{\mathbf{v}}^{k+1} := \tilde{\mathbf{w}} / \mathbf{H}_{k+1,k}$, $\mathbf{v}^{k+1} := \mathbf{w} / \mathbf{H}_{k+1,k}$.

Output: Approximations \mathbf{x}_ℓ^k , norms of preconditioned residuals $\|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}}$, and index $\underline{k}[\ell]$.

Remark 3 (notation of Algorithm A). Define $K := \text{mod}(k-1, k_{\max}) + 1$ as

$$k \mapsto \begin{cases} k_{\max} & \text{for } \text{mod}(k, k_{\max}) = 0, \\ \text{mod}(k, k_{\max}) & \text{else,} \end{cases}$$

and thus yields the number of GMRES steps since the last restart. Moreover, the quantity $R := \lfloor k/k_{\max} \rfloor$ gives the number of restarts performed so far. In step (ii), the orthogonalization loop is carried out over the last K Arnoldi basis vectors \mathbf{v}^k . This is realized through the index $k+j-K$, where $j = 1, \dots, K$. Finally, $\mathbf{x}_\ell^{Rk_{\max}}$ denotes the initial guess of the current GMRES run, and $\mathbf{s}_\ell^{Rk_{\max}}$ the corresponding initial preconditioned residual.

Remark 4 (computational complexity of PGMRES with restart). For every k , Algorithm A performs two applications of \mathbf{B}_ℓ as well as two applications of \mathbf{P}_ℓ , which due to sparsity of \mathbf{B}_ℓ and optimality assumptions on the preconditioner \mathbf{P}_ℓ can be performed with linear cost $\mathcal{O}(N_\ell)$. Since $K \leq k_{\max}$ the j -loop is bounded and the calculation of \mathbf{y} in step (iv) is of constant cost with respect to k and the degrees of freedom. Therefore, one step of optimal preconditioned GMRES with restart is of cost $\mathcal{O}(N_\ell)$.

The analysis of Algorithm A relies on an equivalence of the discrete and level-induced norm $\|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}}$ of the k -th residual vector and the functional energy norm $\|u_\ell^* - u_\ell^k\|$ of the algebraic error, where $u_\ell^k := \sum_{j=1}^{N_\ell} (\mathbf{x}_\ell^k)_j \varphi_{\ell,j}$ and \mathbf{x}_ℓ^k is the k -th iterate of Algorithm A. First, let us show a preliminary equivalence result on the preconditioned norm $\|\cdot\|_{\mathbf{P}_\ell^{-1}}$.

Lemma 5 (optimality of preconditioner in discrete norm). For any $\mathbf{x}_\ell \in \mathbb{R}^{N_\ell}$ and $v_\ell := \sum_{j=1}^{N_\ell} (\mathbf{x}_\ell)_j \varphi_{\ell,j} \in \mathcal{X}_\ell$, there hold the estimates

$$C_{\text{pre}}^{-1} \|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}} \stackrel{(9)}{\leq} \|v_\ell\| \leq C_{\text{ell}}^{-1} C_{\text{pre}} \|\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}} \quad (11a)$$

and

$$C_{\text{bnd}}^{-1} C_{\text{pre}}^{-1} \|\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}} \leq \|v_\ell\| \stackrel{(9)}{\leq} C_{\text{pre}} \|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}. \quad (11b)$$

In particular, all estimates are h - and p -robust.

Proof. The estimates are trivial for $\mathbf{x}_\ell = 0$. For the case $\mathbf{x}_\ell \neq 0$, note that

$$\|\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}} = \sup_{\mathbf{y}_\ell \in \mathbb{R}^{N_\ell} \setminus \{0\}} \frac{(\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{y}_\ell)_{\mathbf{P}_\ell^{-1}}}{\|\mathbf{y}_\ell\|_{\mathbf{P}_\ell^{-1}}} \geq \frac{(\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{P}_\ell^{-1}}}{\|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}} = \frac{(\mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_2}{\|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}}.$$

Moreover, (3) and (9) result in

$$\frac{(\mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_2}{\|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}} = \frac{b(v_\ell, v_\ell)}{\|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}} \stackrel{(3)}{\geq} \frac{C_{\text{ell}} \|v_\ell\|^2}{\|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}} \stackrel{(9)}{\geq} \frac{C_{\text{ell}} \|v_\ell\|}{C_{\text{pre}}} \stackrel{(9)}{\geq} \frac{C_{\text{ell}}}{C_{\text{pre}}^2} \|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}.$$

The combination of the previous two formulas establishes (11a). For (11b), set $\mathbf{y}_\ell := \mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell \in \mathbb{R}^{N_\ell}$ and consider $w_\ell := \sum_{j=1}^{N_\ell} (\mathbf{y}_\ell)_j \varphi_{\ell,j} \in \mathcal{X}_\ell$. Then, (3) and (9) result in

$$\begin{aligned} \|\mathbf{y}_\ell\|_{\mathbf{P}_\ell^{-1}}^2 &= (\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{y}_\ell)_{\mathbf{P}_\ell^{-1}} = (\mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{y}_\ell)_2 \\ &= b(v_\ell, w_\ell) \stackrel{(3)}{\leq} C_{\text{bnd}} \|v_\ell\| \|w_\ell\| \stackrel{(9)}{\leq} C_{\text{bnd}} C_{\text{pre}} \|v_\ell\| \|\mathbf{y}_\ell\|_{\mathbf{P}_\ell^{-1}}. \end{aligned}$$

Rearranging the estimate and employing (9), we prove (11b) and conclude the proof. \square

Lemma 5 can be used to establish the desired equivalence of the energy norm of the algebraic error and the algebraic residual norm.

Lemma 6 (equivalence of discrete norm and functional algebraic error). *Let \mathbf{x}_ℓ^k denote the k -th iterate computed in Algorithm A, \mathbf{s}_ℓ^k the preconditioned residual and $u_\ell^k := \sum_{j=1}^{N_\ell} (\mathbf{x}_\ell^k)_j \varphi_{\ell,j} \in \mathcal{X}_\ell$ the corresponding function. Then, there holds the equivalence*

$$C_{\text{ell}} C_{\text{pre}}^{-1} \|u_\ell^* - u_\ell^k\| \leq \|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} \leq C_{\text{bnd}} C_{\text{pre}} \|u_\ell^* - u_\ell^k\|. \quad (12)$$

In particular, the equivalence is h - and p -robust.

Proof. The definition of \mathbf{s}_ℓ^k in Algorithm A and (8) yield

$$\|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} = \|\mathbf{P}_\ell(\mathbf{d}_\ell - \mathbf{B}_\ell \mathbf{x}_\ell^k)\|_{\mathbf{P}_\ell^{-1}} \stackrel{(8)}{=} \|\mathbf{P}_\ell(\mathbf{B}_\ell \mathbf{x}_\ell^* - \mathbf{B}_\ell \mathbf{x}_\ell^k)\|_{\mathbf{P}_\ell^{-1}}.$$

Combining this with (11) and $u_\ell^* - u_\ell^k = \sum_{j=1}^{N_\ell} (\mathbf{x}_\ell^* - \mathbf{x}_\ell^k)_j \varphi_{\ell,j}$ finishes the proof. \square

Proposition 7 (discrete contraction of optimally preconditioned GMRES). *Let $k_{\text{max}} \in \mathbb{N}$ be arbitrary. Suppose formally that $\Lambda < 0$ so that Algorithm A does not terminate. Then, the optimally preconditioned GMRES method from Algorithm A guarantees*

$$\|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} \leq q_{\text{alg}} \|\mathbf{s}_\ell^{k-1}\|_{\mathbf{P}_\ell^{-1}} \quad \text{for all } k \in \mathbb{N}, \quad (13)$$

with $0 < q_{\text{alg}} < 1$ depending only on C_{bnd} , C_{ell} , and C_{pre} .

Proof. By construction, the preconditioned GMRES method is a Krylov subspace method which minimizes residuals with respect to the preconditioned inner product over the affine Krylov space generated by the preconditioned matrix $\mathbf{P}_\ell \mathbf{B}_\ell$. Note that the Krylov spaces are reset at every restart of Algorithm A. Recalling $K = \text{mod}(k-1, k_{\text{max}}) + 1 \geq 1$ and the parameter R that counts the number of restarts and noting that $\mathbf{s}_\ell^{Rk_{\text{max}}}$ is the initial residual of the current GMRES run, it hence follows

$$\|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} = \min_{p \in \mathbb{P}_K, p(0)=1} \|p(\mathbf{P}_\ell \mathbf{B}_\ell) \mathbf{s}_\ell^{Rk_{\text{max}}}\|_{\mathbf{P}_\ell^{-1}}, \quad (14)$$

where $\mathbf{s}_\ell^k = \mathbf{P}_\ell(\mathbf{d}_\ell - \mathbf{B}_\ell \mathbf{x}_\ell^k) = \mathbf{P}_\ell \mathbf{B}_\ell(\mathbf{x}_\ell - \mathbf{x}_\ell^k)$ and \mathbb{P}_K denotes the space of polynomials (in one variable) of degree $\leq K$; see [EES83; SS07] and references therein. This yields

$$\|\mathbf{s}_\ell^{k-1}\|_{\mathbf{P}_\ell^{-1}} = \min_{p \in \mathbb{P}_{K-1}, p(0)=1} \|p(\mathbf{P}_\ell \mathbf{B}_\ell) \mathbf{s}_\ell^{Rk_{\text{max}}}\|_{\mathbf{P}_\ell^{-1}} = \|p_{K-1}(\mathbf{P}_\ell \mathbf{B}_\ell) \mathbf{s}_\ell^{Rk_{\text{max}}}\|_{\mathbf{P}_\ell^{-1}}, \quad (15)$$

Let $\alpha \in \mathbb{R}$ (to be fixed below) and define the polynomial \tilde{p} by $\tilde{p}(t) = (1 - \alpha t)p_{K-1}(t)$, $t \in \mathbb{R}$. Then, $\tilde{p} \in \mathbb{P}_K$. Furthermore, $\tilde{p}(0) = (1 - \alpha 0)p_{K-1}(0) = 1$. Therefore,

$$\begin{aligned} \|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} &= \min_{p \in \mathbb{P}_K, p(0)=1} \|p(\mathbf{P}_\ell \mathbf{B}_\ell) \mathbf{s}_\ell^{Rk_{\max}}\|_{\mathbf{P}_\ell^{-1}} \leq \|\tilde{p}(\mathbf{P}_\ell \mathbf{B}_\ell) \mathbf{s}_\ell^{Rk_{\max}}\|_{\mathbf{P}_\ell^{-1}} \\ &= \|(1 - \alpha \mathbf{P}_\ell \mathbf{B}_\ell) p_{K-1}(\mathbf{P}_\ell \mathbf{B}_\ell) \mathbf{s}_\ell^{Rk_{\max}}\|_{\mathbf{P}_\ell^{-1}} \\ &\leq \|(1 - \alpha \mathbf{P}_\ell \mathbf{B}_\ell)\|_{\mathbf{P}_\ell^{-1}} \|p_{K-1}(\mathbf{P}_\ell \mathbf{B}_\ell) \mathbf{s}_\ell^{Rk_{\max}}\|_{\mathbf{P}_\ell^{-1}} \stackrel{(15)}{=} \|(1 - \alpha \mathbf{P}_\ell \mathbf{B}_\ell)\|_{\mathbf{P}_\ell^{-1}} \|\mathbf{s}_\ell^{k-1}\|_{\mathbf{P}_\ell^{-1}}. \end{aligned}$$

To prove $\|(1 - \alpha \mathbf{P}_\ell \mathbf{B}_\ell)\|_{\mathbf{P}_\ell^{-1}} \leq q_{\text{alg}}$, let $\mathbf{x}_\ell \in \mathbb{R}^{N_\ell}$ with $\|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}} = 1$. Then, there holds

$$\|(1 - \alpha \mathbf{P}_\ell \mathbf{B}_\ell) \mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}^2 = \|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}^2 - 2\alpha (\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{P}_\ell^{-1}} + \alpha^2 \|\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}^2.$$

Viewing the right-hand side as a quadratic polynomial in α , it attains its minimum for $\alpha = (\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{P}_\ell^{-1}} / \|\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}^2$. This choice of α leads to

$$\|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}^2 - 2\alpha (\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{P}_\ell^{-1}} + \alpha^2 \|\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}^2 = 1 - \frac{(\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{P}_\ell^{-1}}^2}{\|\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}^2}.$$

With $v_\ell := \sum_{j=1}^{N_\ell} (\mathbf{x}_\ell)_j \varphi_{\ell,j} \in \mathcal{X}_\ell$, (3) and (9) yield

$$(\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{P}_\ell^{-1}} = (\mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_2 = b(v_\ell, v_\ell) \stackrel{(3)}{\geq} C_{\text{ell}} \|v_\ell\|^2.$$

In combination with Lemma 5, we further get that

$$1 - \frac{(\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{P}_\ell^{-1}}^2}{\|\mathbf{P}_\ell \mathbf{B}_\ell \mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}^2} \stackrel{(11)}{\leq} 1 - \frac{C_{\text{ell}}^2 \|v_\ell\|^4}{C_{\text{bnd}}^2 C_{\text{pre}}^2 \|v_\ell\|^2} \stackrel{(9)}{\leq} 1 - \frac{C_{\text{ell}}^2}{C_{\text{bnd}}^2 C_{\text{pre}}^4} \|\mathbf{x}_\ell\|_{\mathbf{P}_\ell^{-1}}^2$$

Altogether, we conclude that

$$\|1 - \alpha \mathbf{P}_\ell \mathbf{B}_\ell\|_{\mathbf{P}_\ell^{-1}} \leq \left(1 - \frac{C_{\text{ell}}^2}{C_{\text{bnd}}^2 C_{\text{pre}}^4} \frac{1}{C_{\text{pre}}^4}\right)^{1/2} =: q_{\text{alg}} < 1.$$

This shows (13) and hence concludes the proof. \square

Remark 8. Note that Proposition 7 ensures, in particular, discrete contraction for $k_{\max} = 1$. In this case only a one-dimensional Arnoldi basis is built in every step of PGMRES.

3.2. Multilevel additive Schwarz preconditioner. To define the multilevel additive Schwarz preconditioner $\mathbf{P}_\ell^{\text{AS}} \in \mathbb{R}^{N_\ell \times N_\ell}$, we first introduce the necessary notation. Let \mathcal{V}_ℓ denote the set of all vertices of the triangulation \mathcal{T}_ℓ . For each vertex $z \in \mathcal{V}_\ell$, we define the associated vertex patch by

$$\mathcal{T}_\ell(z) := \{T \in \mathcal{T}_\ell : z \in T\} \quad \text{and} \quad \omega_\ell(z) := \text{interior}\left(\bigcup_{T \in \mathcal{T}_\ell(z)} T\right). \quad (16)$$

Denote by \mathcal{V}_ℓ^+ the set of newly created vertices in \mathcal{T}_ℓ and their direct neighbors, i.e.,

$$\mathcal{V}_\ell^+ := \mathcal{V}_\ell \setminus \mathcal{V}_{\ell-1} \cup \{z' \in \mathcal{V}_\ell \cap \mathcal{V}_{\ell-1} : \omega_\ell(z') \neq \omega_{\ell-1}(z')\}. \quad (17)$$

To define the correction subspaces underlying the preconditioner, we set

$$\mathcal{X}_\ell^q := \mathcal{S}_0^q(\mathcal{T}_\ell), \quad \mathcal{X}_{\ell,z}^q := \mathcal{S}_0^q(\mathcal{T}_\ell(z)) \quad \text{for all } z \in \mathcal{V}_\ell, \text{ and } q \in \{1, p\}.$$

To build the preconditioner on level ℓ we employ $q = 1$ for $0 \leq \ell' < \ell$ respectively $q = p$ for $\ell' = \ell$. Let $\mathbf{A}_\ell \in \mathbb{R}^{N_\ell \times N_\ell}$ denote the Galerkin matrix for principal part $a(\cdot, \cdot)$, i.e.,

$$(\mathbf{A}_\ell)_{jk} := a(\varphi_{\ell,k}, \varphi_{\ell,j}), \quad \text{for all } j, k = 1, \dots, N_\ell.$$

Note that the induced norm satisfies $\|\mathbf{x}_\ell\|_{\mathbf{A}_\ell} = \|u_\ell\|$ for $u_\ell = \sum_{j=1}^{N_\ell} (\mathbf{x})_j \varphi_{\ell,j}$ and all $\mathbf{x}_\ell \in \mathbb{R}^{N_\ell}$. For $0 \leq \ell' < \ell$, let $\mathcal{I}_{\ell'}^1 : \mathcal{X}_{\ell'}^1 \rightarrow \mathcal{X}_\ell^p = \mathcal{X}_\ell$ denote the canonical embedding, with matrix representation $\mathbf{I}_{\ell'}^1 \in \mathbb{R}^{N_\ell \times N_{\ell'}^1}$, where $N_{\ell'}^1 := \dim(\mathcal{X}_{\ell'}^1)$. Similarly, the embeddings $\mathcal{I}_{\ell,z}^p : \mathcal{X}_{\ell,z}^p \rightarrow \mathcal{X}_\ell^p = \mathcal{X}_\ell$ are represented by matrices $\mathbf{I}_{\ell,z}^p \in \mathbb{R}^{N_\ell \times N_{\ell,z}^p}$, with $N_{\ell,z}^p := \dim(\mathcal{X}_{\ell,z}^p)$.

Denote by $\mathbf{A}_{\ell'}^1$ and $\mathbf{A}_{\ell,z}^p$ the Galerkin matrices corresponding to the spaces $\mathcal{X}_{\ell'}^1$ and $\mathcal{X}_{\ell,z}^p$, respectively. For each $0 \leq \ell' < \ell$, we define the diagonal matrix $(\mathbf{D}_{\ell'}^+)^{-1} \in \mathbb{R}^{N_{\ell'}^1 \times N_{\ell'}^1}$ by

$$[(\mathbf{D}_{\ell'}^+)^{-1}]_{jk} := \begin{cases} \delta_{jk} ([\mathbf{A}_{\ell'}^1]_{jj})^{-1}, & \text{if } z_j \in \mathcal{V}_{\ell'}^+, \\ 0, & \text{otherwise.} \end{cases}$$

With this notation, the additive Schwarz preconditioner is defined as

$$\mathbf{P}_\ell^{\text{AS}} := \mathbf{I}_0^1 (\mathbf{A}_0^1)^{-1} (\mathbf{I}_0^1)^T + \sum_{\ell'=1}^{\ell-1} \mathbf{I}_{\ell'}^1 (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^1)^T + \sum_{z \in \mathcal{V}_\ell} \mathbf{I}_{\ell,z}^p (\mathbf{A}_{\ell,z}^p)^{-1} (\mathbf{I}_{\ell,z}^p)^T.$$

Remark 9. Note that the matrices $\mathbf{I}_{\ell'}^1$ and $\mathbf{I}_{\ell,z}^p$ are introduced solely for the purpose of the analysis and are not used in the implementation. In practice, only the prolongation and restriction matrices between consecutive levels are required. As is standard practice in multigrid methods, these can be assembled in $\mathcal{O}(\#\mathcal{V}_{\ell'}^+)$ operations on the intermediate levels and in $\mathcal{O}(\#\mathcal{T}_\ell)$ operations on the finest level, thus preserving the overall linear complexity of the method; see, e.g., [XQ94].

Lemma 10 (robust additive Schwarz preconditioner). The additive Schwarz preconditioner $\mathbf{P}_\ell^{\text{AS}}$ is SPD and satisfies the norm equivalence (9), where the constant C_{pre} depends only on Ω , \mathcal{T}_0 , d , γ -shape regularity, and \mathbf{K} .

Proof. From [HMP26, Corollary 19] and [TW05, Theorem C.1], it follows that $\mathbf{P}_\ell^{\text{AS}}$ is SPD and satisfies the spectral equivalence

$$c \leq \lambda_{\min}(\mathbf{P}_\ell^{\text{AS}} \mathbf{A}_\ell) = \|(\mathbf{P}_\ell^{\text{AS}} \mathbf{A}_\ell)^{-1}\|_{\mathbf{A}_\ell}^{-1} \quad \text{and} \quad \|\mathbf{P}_\ell^{\text{AS}} \mathbf{A}_\ell\|_{\mathbf{A}_\ell} = \lambda_{\max}(\mathbf{P}_\ell^{\text{AS}} \mathbf{A}_\ell) \leq C, \quad (18)$$

where the constants c and C depend only on Ω , \mathcal{T}_0 , d , γ -shape regularity, and \mathbf{K} . Let $v_\ell = \sum_{j=1}^{N_\ell} (\mathbf{x}_\ell)_j \varphi_{\ell,j} \in \mathcal{X}_\ell$. Using (18), we obtain

$$\|\mathbf{x}_\ell\|_{(\mathbf{P}_\ell^{\text{AS}})^{-1}}^2 = ((\mathbf{P}_\ell^{\text{AS}} \mathbf{A}_\ell)^{-1} \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell} \leq \|(\mathbf{P}_\ell^{\text{AS}} \mathbf{A}_\ell)^{-1}\|_{\mathbf{A}_\ell} \|\mathbf{x}_\ell\|_{\mathbf{A}_\ell}^2 \stackrel{(18)}{\leq} c^{-1} \|v_\ell\|^2.$$

Moreover,

$$(\mathbf{P}_\ell^{\text{AS}} \mathbf{A}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell} \leq \|\mathbf{P}_\ell^{\text{AS}} \mathbf{A}_\ell\|_{\mathbf{A}_\ell} \|\mathbf{x}_\ell\|_{\mathbf{A}_\ell}^2 \stackrel{(18)}{\leq} C \|v_\ell\|^2.$$

Setting $\mathbf{y}_\ell := \mathbf{P}_\ell^{\text{AS}} \mathbf{A}_\ell \mathbf{x}_\ell$, we obtain that

$$\|\mathbf{x}_\ell\|_{(\mathbf{P}_\ell^{\text{AS}})^{-1}} = \sup_{\mathbf{y}_\ell \neq 0} \frac{(\mathbf{x}_\ell, \mathbf{y}_\ell)_{(\mathbf{P}_\ell^{\text{AS}})^{-1}}}{\|\mathbf{y}_\ell\|_{(\mathbf{P}_\ell^{\text{AS}})^{-1}}} \geq \frac{(\mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell}}{((\mathbf{P}_\ell^{\text{AS}})^{-1} \mathbf{A}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell}^{1/2}} \geq C^{-1/2} \|v_\ell\|.$$

This concludes the proof with $C_{\text{pre}} = \max\{\sqrt{C}, 1/\sqrt{c}\}$. \square

3.3. Symmetric multigrid preconditioner. As an alternative, we consider the symmetric multigrid preconditioner from [HMP26, Section 5] as symmetrized and linearized version of [IMPS24]. In the following, we abbreviate $\prod_{i=0}^{\ell} \mathbf{C}_i := \mathbf{C}_0 \mathbf{C}_1 \cdots \mathbf{C}_\ell$ for any matrices $\mathbf{C}_0, \dots, \mathbf{C}_\ell$. Let $\mu := 1/(d+1)$ and define, for $\ell' \in \{1, \dots, \ell\}$,

$$\mathbf{M}_{\ell'}^\downarrow := \mathbf{I}_{\ell'}^1 (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^1)^T \prod_{i=\ell'+1}^{\ell} (\mathbf{I} - \mu \mathbf{A}_\ell \mathbf{I}_i^1 (\mathbf{D}_i^+)^{-1} (\mathbf{I}_i^1)^T),$$

for $\ell' = 0$,

$$\mathbf{M}_0^\uparrow := \mathbf{I}_0^1 (\mathbf{A}_0^1)^{-1} (\mathbf{I}_0^1)^T \prod_{i=1}^{\ell} (\mathbf{I} - \mu \mathbf{A}_\ell \mathbf{I}_i^1 (\mathbf{D}_i^+)^{-1} (\mathbf{I}_i^1)^T),$$

and, for $\ell' \in \{1, \dots, \ell\}$,

$$\mathbf{M}_{\ell'}^\uparrow := \mathbf{I}_{\ell'}^1 (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^1)^T \prod_{i=1}^{\ell'-1} (\mathbf{I} - \mu \mathbf{A}_\ell \mathbf{I}_{\ell'-i}^1 (\mathbf{D}_{\ell'-i}^+)^{-1} (\mathbf{I}_{\ell'-i}^1)^T) \prod_{i=0}^{\ell} (\mathbf{I} - \mu \mathbf{A}_\ell \mathbf{I}_i^1 (\mathbf{D}_i^+)^{-1} (\mathbf{I}_i^1)^T).$$

For notational convenience, we set here $\mathbf{D}_0^+ := \mu \mathbf{A}_0^1$, $(\mathbf{D}_\ell^+)^{-1} := \sum_{z \in \mathcal{V}_\ell} \mathbf{I}_{\ell,z}^p (\mathbf{A}_{\ell,z}^p)^{-1} (\mathbf{I}_{\ell,z}^p)^T$, and $\mathbf{I}_\ell^1 := \mathbf{I}$. The symmetric multigrid preconditioner is then defined as

$$\mathbf{P}_\ell^{\text{sMG}} := \mu \sum_{\ell'=1}^{\ell} \mathbf{M}_{\ell'}^\downarrow + \mathbf{M}_0^\uparrow + \mu \sum_{\ell'=1}^{\ell} \mathbf{M}_{\ell'}^\uparrow. \quad (19)$$

Lemma 11 (symmetric multigrid preconditioner). *The symmetric multigrid preconditioner $\mathbf{P}_\ell^{\text{sMG}}$ is SPD and satisfies (9) with constant $C_{\text{pre}} = \max\{\sqrt{1+q}, 1/\sqrt{1-q}\}$, where $0 < q < 1$ depends only on Ω , \mathcal{T}_0 , d , γ -shape regularity, and \mathbf{K} .*

Proof. From the proof of [HMP26, Corollary 16], we obtain estimates analogous to (18). In particular, there exists a constant $0 < q < 1$, independent of h and p , such that

$$\|(\mathbf{P}_\ell^{\text{sMG}} \mathbf{A}_\ell)^{-1}\|_{\mathbf{A}_\ell} \leq \frac{1}{1-q}, \quad \|\mathbf{P}_\ell^{\text{sMG}} \mathbf{A}_\ell\|_{\mathbf{A}_\ell} \leq 1+q.$$

Moreover, [HMP26, Lemma 15] states that $\mathbf{P}_\ell^{\text{sMG}}$ is SPD. With the arguments from the proof of Lemma 10, this concludes the proof. \square

4. AFEM WITH A-POSTERIORI-STEERED PARAMETER CONTROL

In this section, we present a new adaptive algorithm that employs PGMRES from Algorithm B as its algebraic solver. First, we recall a result from [CFPP14] establishing the equivalence between tail summability, inverse summability, and R-linear convergence for any nonnegative sequence of scalars. This result is central both to the design of the algorithm and to the proof of unconditional full R-linear convergence in Section 5.

Lemma 12 (summability and R-linear convergence). *Let $(a_\ell)_{\ell \in \mathbb{N}_0}$ be a sequence in $\mathbb{R}_{>0}$ and $\beta, \gamma > 0$. Then, the following statements (i)–(iii) are pairwise equivalent*

(i) **tail summability:** *There exists a constant $C_\beta > 0$, such that*

$$\sum_{\ell'=\ell+1}^{\infty} a_{\ell'}^\beta \leq C_\beta a_\ell^\beta \quad \text{for all } \ell \in \mathbb{N}_0. \quad (20)$$

(ii) **inverse summability:** There exists a constant $C_\gamma > 0$, such that

$$\sum_{\ell'=0}^{\ell-1} a_{\ell'}^{-\gamma} \leq C_\gamma a_\ell^{-\gamma} \quad \text{for all } \ell \in \mathbb{N}_0. \quad (21)$$

(iii) **R-linear convergence:** There exist constants $\tilde{C}_{\text{lin}} > 1$ and $0 < \tilde{q}_{\text{lin}} < 1$ with

$$a_\ell \leq \tilde{C}_{\text{lin}} \tilde{q}_{\text{lin}}^{\ell-\ell'} a_{\ell'} \quad \text{for all } 0 \leq \ell' \leq \ell. \quad (22)$$

Let $(a_\ell)_{\ell \in \mathbb{N}_0}$ have the additional property

$$\forall \ell \in \mathbb{N}_0 : \left(a_\ell = 0 \implies a_{\ell'} = 0 \quad \text{for all } \ell' > \ell \right). \quad (23)$$

Let $\ell_0 \in \mathbb{N}$. If any of the statements (i)–(iii) hold only for $\ell \geq \ell_0$, then they already hold for all $\ell \in \mathbb{N}_0$ with potentially larger constants depending also on $(a_{\ell'})_{\ell'=0}^{\ell_0}$.

Proof. The equivalence of statements (i)–(iii) was proved in [CFPP14, Lemma 4.9], albeit without the exponent in the tail summability condition; see also [BFM⁺25, Lemma 2]. Let $(a_\ell)_{\ell \in \mathbb{N}_0}$ satisfy (23). We show that tail summability (20) for all $\ell \geq \ell_0$ implies tail summability (20) for all $\ell \in \mathbb{N}_0$. The corresponding implications for (ii) and (iii) follow similarly. To this end, let $\ell < \ell_0$ and define $N := \{0 \leq \ell' < \ell_0 : a_{\ell'} \neq 0\}$. Moreover, set

$$\tilde{C}_m := \frac{C_m a_{\ell_0}^m + \sum_{\ell'=1}^{\ell_0} a_{\ell'}^m}{\min_{\ell' \in N} a_{\ell'}^m}.$$

Then, using (20) for ℓ_0 and the assumption (23) in the case $a_\ell = 0$, we obtain

$$\sum_{\ell'=\ell+1}^{\infty} a_{\ell'}^m = \sum_{\ell'=\ell_0+1}^{\infty} a_{\ell'}^m + \sum_{\ell'=\ell+1}^{\ell_0} a_{\ell'}^m \stackrel{(20)}{\leq} C_m a_{\ell_0}^m + \sum_{\ell'=\ell+1}^{\ell_0} a_{\ell'}^m \stackrel{(23)}{\leq} \tilde{C}_m a_\ell^m.$$

This concludes the proof. \square

Given an approximation $u_\ell \in \mathcal{X}_\ell$ of the solution $u_\ell^* \in \mathcal{X}_\ell$ to (5) and its vector representation $u_\ell = \sum_{j=1}^{N_\ell} (\mathbf{x}_\ell)_j \varphi_{\ell,j}$, Algorithm A returns an improved approximation $\mathbf{y}_\ell \in \mathbb{R}^{N_\ell}$. This gives rise to the map $\Psi_\ell: \mathcal{X}_\ell \rightarrow \mathcal{X}_\ell$, $\Psi_\ell(u_\ell) := \sum_{j=1}^{N_\ell} (\mathbf{y}_\ell)_j \varphi_{\ell,j}$. In the following algorithm, the iterates satisfy $u_\ell^k = \sum_{j=1}^{N_\ell} (\mathbf{x}_\ell^k)_j \varphi_{\ell,j}$ and $\mathbf{s}_\ell^k = \mathbf{P}_\ell(\mathbf{d}_\ell - \mathbf{B}_\ell \mathbf{x}_\ell^k)$, where \mathbf{x}_ℓ^k are the iterates generated by Algorithm A with initial guess \mathbf{x}_ℓ^0 .

Algorithm B (unconditional AFEM with optimal PGMRES solver). *Input:* Initial mesh \mathcal{T}_0 , adaptivity parameters $0 < \theta \leq 1$ and $C_{\text{mark}} \geq 1$, initial guess $u_0^0 \in \mathcal{X}_0$, as well as $C_{\text{alg}}, \lambda_{\text{alg}} > 0$, $\mathbf{S} := 0$, k_{max} , and tolerance $\tau \geq 0$.

Perform the following steps (i)–(iv) for all $\ell = 0, 1, 2, \dots$:

- (i) **Solve & Estimate:** Compute $u_\ell^{k[\ell]} := \Psi_\ell(u_\ell^0)$ via the optimally preconditioned GMRES method from Algorithm A using $\Lambda(\mathbf{x}_\ell^k) := \lambda_{\text{alg}} \eta_\ell(u_\ell^k)$ and k_{max} for $k \leq k[\ell]$, where $k[\ell] \in \mathbb{N}$ denotes the number of PGMRES iterations on level ℓ .
- (ii) **Adaptive Parameter Control:** If $\eta_\ell(u_\ell^{k[\ell]}) + \|\mathbf{s}_\ell^{k[\ell]}\|_{\mathbf{P}_\ell^{-1}} \leq \tau$, then stop Algorithm B and terminate; else do:
 - (a) If $\ell \geq 1$ and $\mathbf{S} > C_{\text{alg}}(\eta_\ell(u_\ell^{k[\ell]}) + \|\mathbf{s}_\ell^{k[\ell]}\|_{\mathbf{P}_\ell^{-1}})^{-1}$, then $C_{\text{alg}} \leftarrow 2C_{\text{alg}}$, $\lambda_{\text{alg}} \leftarrow \lambda_{\text{alg}}/2$.
 - (b) Always update $\mathbf{S} \leftarrow \mathbf{S} + (\eta_\ell(u_\ell^{k[\ell]}) + \|\mathbf{s}_\ell^{k[\ell]}\|_{\mathbf{P}_\ell^{-1}})^{-1}$.

- (iii) **Mark:** Determine a set $\mathcal{M}_\ell \in \mathbb{M}_\ell[\theta, u_\ell^k] := \{\mathcal{U}_\ell \subseteq \mathcal{T}_\ell : \theta \eta_\ell(u_\ell^k)^2 \leq \eta_\ell(\mathcal{U}_\ell, u_\ell^k)^2\}$ satisfying the Dörfler marking criterion

$$\#\mathcal{M}_\ell \leq C_{\text{mark}} \min_{\mathcal{U}_\ell \in \mathbb{M}_\ell[\theta, u_\ell^k]} \#\mathcal{U}_\ell. \quad (24)$$

- (iv) **Refine:** Generate $\mathcal{T}_{\ell+1} := \text{refine}(\mathcal{T}_\ell, \mathcal{M}_\ell)$ and employ nested iteration $u_{\ell+1}^0 := u_\ell^k$.

Output: Sequence of meshes $\{\mathcal{T}_\ell\}_{\ell \in \mathbb{N}_0}$ and discrete solutions $\{u_\ell^k\}_{\ell \in \mathbb{N}_0}$.

Note that Algorithm B consists of two nested loops, where we denote the discretization with the index ℓ and the iterations of the algebraic solver, which in this case is PGMRES, with k . More precisely, the function u_ℓ^k for $k \leq \underline{k}[\ell]$ denotes the k -th approximation of u_ℓ^* calculated by Algorithm A on the discretization level ℓ with initial guess u_ℓ^0 . The sequential nature of Algorithm B results in the countably infinite index set

$$\mathcal{Q} := \{(\ell, k) \in \mathbb{N}_0^2 : u_\ell^k \in \mathcal{X}_\ell \text{ appears in Algorithm B}\}$$

equipped with the lexicographic ordering

$$(\ell', k') \leq (\ell, k) \quad :\iff \quad u_{\ell'}^{k'} \text{ is defined not later than } u_\ell^k \text{ in Algorithm B}$$

and the total step counter

$$|\ell, k| := \#\{(\ell', k') \in \mathcal{Q} : (\ell', k') \leq (\ell, k)\} \in \mathbb{N}_0 \quad \text{for all } (\ell, k) \in \mathcal{Q}.$$

Moreover, we define the stopping indices

$$\begin{aligned} \underline{\ell} &:= \sup\{\ell \in \mathbb{N}_0 : (\ell, 0) \in \mathcal{Q}\} \in \mathbb{N}_0 \cup \{\infty\}, \\ \underline{k}[\ell] &:= \sup\{k \in \mathbb{N}_0 : (\ell, k) \in \mathcal{Q}\} \in \mathbb{N} \cup \{\infty\}, \quad \text{whenever } (\ell, 0) \in \mathcal{Q}. \end{aligned} \quad (25)$$

We stress that the definition of $\underline{k}[\ell]$ is consistent with that of Algorithm B(ii).

Remark 13 (adaptive parameter control). Let us define the computable quasi-error

$$\mathbf{H}_\ell^k := \eta_\ell(u_\ell^k) + \|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} \quad \text{for all } (\ell, k) \in \mathcal{Q} \quad (26)$$

and note that $\mathbf{S} = \sum_{j=0}^{\ell} (\mathbf{H}_j^k)^{-1}$ in Algorithm B(ii.b). Comparing Algorithm B to standard adaptive algorithms (see, e.g., [BFM⁺25, Algorithm B]), we observe that the difference lies in the adaptive parameter control performed in step (ii). This new a-posteriori-steered criterion aims to algorithmically guarantee R -linear convergence (22) of \mathbf{H}_ℓ^k across the levels ℓ by testing the equivalent inverse summability condition (21) for $\gamma = 1$.

The following lemma provides a-posteriori error control of the total error.

Lemma 14 (a-posteriori error control). Suppose that the estimator satisfies (A1)–(A3⁺) and that the employed preconditioner satisfies (9). Recall the computable quasi-error \mathbf{H}_ℓ^k from (26). With the constants $C_{\text{stab}}, C_{\text{ell}}, C_{\text{bnd}}$, and C_{pre} from (A1), (3), and (9), define $C_{\text{eq}} := \max\{1, (1 + C_{\text{stab}}) C_{\text{ell}}^{-1} C_{\text{pre}}, (C_{\text{bnd}} C_{\text{pre}} + C_{\text{stab}})\} > 0$. Then, there holds

$$C_{\text{eq}}^{-1} (\eta_\ell(u_\ell^*) + \|\|u_\ell^* - u_\ell^k\|\|) \leq \mathbf{H}_\ell^k \leq C_{\text{eq}} (\eta_\ell(u_\ell^*) + \|\|u_\ell^* - u_\ell^k\|\|) \quad \text{for all } (\ell, k) \in \mathcal{Q}. \quad (27)$$

With $C'_{\text{rel}} := C_{\text{eq}} \max\{C_{\text{rel}}, 1\} > 0$, this yields

$$\|\|u^* - u_\ell^k\|\| \leq C'_{\text{rel}} \mathbf{H}_\ell^k, \quad \text{for all } (\ell, k) \in \mathcal{Q}. \quad (28)$$

Proof. We start with the lower bound in (27). Stability (A1) and equivalence (12) yield

$$\begin{aligned} \eta_\ell(u_\ell^\star) + \lll u_\ell^\star - u_\ell^k \lll &\stackrel{\text{(A1)}}{\leq} \eta_\ell(u_\ell^k) + (1 + C_{\text{stab}}) \lll u_\ell^\star - u_\ell^k \lll \\ &\stackrel{\text{(12)}}{\leq} \eta_\ell(u_\ell^k) + (1 + C_{\text{stab}}) C_{\text{ell}}^{-1} C_{\text{pre}} \|r_\ell^k\|_{\mathbf{P}_\ell^{-1}} \leq C_{\text{eq}} \mathbf{H}_\ell^k. \end{aligned}$$

Similarly, for the upper bound in (27), equivalence (12) and stability (A1) imply

$$\begin{aligned} \mathbf{H}_\ell^k &= \eta_\ell(u_\ell^k) + \|r_\ell^k\|_{\mathbf{P}_\ell^{-1}} \stackrel{\text{(12)}}{\leq} \eta_\ell(u_\ell^k) + C_{\text{bnd}} C_{\text{pre}} \lll u_\ell^\star - u_\ell^k \lll \\ &\stackrel{\text{(A1)}}{\leq} \eta_\ell(u_\ell^\star) + (C_{\text{bnd}} C_{\text{pre}} + C_{\text{stab}}) \lll u_\ell^\star - u_\ell^k \lll \leq C_{\text{eq}} (\eta_\ell(u_\ell^\star) + \lll u_\ell^\star - u_\ell^k \lll). \end{aligned}$$

To show (28), we use reliability (A3) and the established equivalence (27) to see that

$$\lll u^\star - u_\ell^k \lll \leq \lll u^\star - u_\ell^\star \lll + \lll u_\ell^\star - u_\ell^k \lll \stackrel{\text{(A3)}}{\leq} C_{\text{rel}} \eta_\ell(u_\ell^\star) + \lll u_\ell^\star - u_\ell^k \lll \stackrel{\text{(27)}}{\leq} C'_{\text{rel}} \mathbf{H}_\ell^k. \quad (29)$$

This concludes the proof. \square

5. UNCONDITIONAL FULL R-LINEAR CONVERGENCE OF ALGORITHM B

To prove unconditional full R-linear convergence of Algorithm B, we first show conditional tail summability of \mathbf{H}_ℓ^k in Lemma 16. This combined with the new adaptive parameter criterion of Algorithm B(ii.a) implies that C_{alg} and λ_{alg} are only updated finitely often; see Lemma 17. Then, Theorem 18 uses this result to derive unconditional full R-linear convergence of the quasi-error \mathbf{H}_ℓ^k , i.e., with no additional assumptions on the input parameters $\theta, C_{\text{alg}}, \lambda_{\text{alg}}$, and k_{max} . The following lemma establishes estimator equivalence in the spirit of [GHPS18, Lemma 4.9] that enables to transfer the Dörfler criterion from the exact discrete solution to the inexact discrete solution, and vice-versa.

Lemma 15 (estimator equivalence). *Suppose that the estimator satisfies (A1) and that the employed preconditioner satisfies (9). Let $\lambda_{\text{alg}}^\star := \min\{1, C_{\text{stab}}^{-1} C_{\text{solve}}^{-1}\}$ with $C_{\text{solve}} := C_{\text{ell}}^{-1} C_{\text{pre}}$. Then, for all $0 < \theta \leq 1$ and all $0 < \lambda_{\text{alg}} < \lambda_{\text{alg}}^\star$, there holds the equivalence*

$$[1 - \lambda_{\text{alg}}/\lambda_{\text{alg}}^\star] \eta_\ell(u_\ell^k) \leq \eta_\ell(u_\ell^\star) \leq [1 + \lambda_{\text{alg}}/\lambda_{\text{alg}}^\star] \eta_\ell(u_\ell^k). \quad (30)$$

Let

$$\theta' := \frac{(\theta^{1/2} - \lambda_{\text{alg}}/\lambda_{\text{alg}}^\star)^2}{(1 + \lambda_{\text{alg}}/\lambda_{\text{alg}}^\star)^2} \quad \text{and} \quad \theta_{\text{mark}} := \frac{(\theta^{1/2} + \lambda_{\text{alg}}/\lambda_{\text{alg}}^\star)^2}{(1 - \lambda_{\text{alg}}/\lambda_{\text{alg}}^\star)^2}. \quad (31)$$

Then, Dörfler marking for (u_ℓ^k, θ) implies Dörfler marking for (u_ℓ^\star, θ') , i.e., for any $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$, there holds the following implication:

$$\theta \eta_\ell(u_\ell^k)^2 \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^k)^2 \implies \theta' \eta_\ell(u_\ell^\star)^2 \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^\star)^2. \quad (32)$$

Conversely, for any $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$, there also holds the following implication:

$$\theta_{\text{mark}} \eta_\ell(u_\ell^\star)^2 \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^\star)^2 \implies \theta \eta_\ell(u_\ell^k)^2 \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^k)^2. \quad (33)$$

Proof. Norm equivalence (12) and the stopping criterion (10) allow to estimate

$$\lll u_\ell^\star - u_\ell^k \lll \stackrel{\text{(12)}}{\leq} C_{\text{ell}}^{-1} C_{\text{pre}} \|s_\ell^k\|_{\mathbf{P}_\ell^{-1}} \stackrel{\text{(10)}}{\leq} C_{\text{ell}}^{-1} C_{\text{pre}} \lambda_{\text{alg}} \eta_\ell(u_\ell^k) = C_{\text{solve}} \lambda_{\text{alg}} \eta_\ell(u_\ell^k). \quad (34)$$

For any $\mathcal{U}_\ell \subseteq \mathcal{T}_\ell$, stability (A1) then allows to obtain

$$\begin{aligned} \eta_\ell(\mathcal{U}_\ell; u_\ell^k) &\stackrel{(A1)}{\leq} \eta_\ell(\mathcal{U}_\ell; u_\ell^*) + C_{\text{stab}} \|u_\ell^* - u_\ell^k\| \stackrel{(34)}{\leq} \eta_\ell(\mathcal{U}_\ell; u_\ell^*) + C_{\text{stab}} C_{\text{solve}} \lambda_{\text{alg}} \eta_\ell(u_\ell^k) \\ &\leq \eta_\ell(\mathcal{U}_\ell; u_\ell^*) + (\lambda_{\text{alg}}/\lambda_{\text{alg}}^*) \eta_\ell(u_\ell^k). \end{aligned} \quad (35)$$

Then, the lower bound of (30) follows by taking $\mathcal{U}_\ell = \mathcal{T}_\ell$, $0 < \lambda_{\text{alg}} < \lambda_{\text{alg}}^*$, and rearranging the latter estimate. Analogously, it follows

$$\eta_\ell(\mathcal{U}_\ell; u_\ell^*) \stackrel{(A1)}{\leq} \eta_\ell(\mathcal{U}_\ell; u_\ell^k) + C_{\text{stab}} \|u_\ell^* - u_\ell^k\| \leq \eta_\ell(\mathcal{U}_\ell; u_\ell^k) + (\lambda_{\text{alg}}/\lambda_{\text{alg}}^*) \eta_\ell(u_\ell^k).$$

For $\mathcal{U}_\ell = \mathcal{T}_\ell$, this concludes the proof of (30). Suppose that $\mathcal{R}_\ell \subseteq \mathcal{T}_\ell$ satisfies $\theta^{1/2} \eta_\ell(u_\ell^k) \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^k)$. Using (35) for \mathcal{R}_ℓ , we obtain

$$\theta^{1/2} \eta_\ell(u_\ell^k) \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^k) \stackrel{(35)}{\leq} \eta_\ell(\mathcal{R}_\ell; u_\ell^*) + (\lambda_{\text{alg}}/\lambda_{\text{alg}}^*) \eta_\ell(u_\ell^k).$$

Together with (30), this yields

$$(\theta')^{1/2} \eta_\ell(u_\ell^*) = \frac{\theta^{1/2} - \lambda_{\text{alg}}/\lambda_{\text{alg}}^*}{1 + \lambda_{\text{alg}}/\lambda_{\text{alg}}^*} \eta_\ell(u_\ell^*) \stackrel{(30)}{\leq} [\theta^{1/2} - \lambda_{\text{alg}}/\lambda_{\text{alg}}^*] \eta_\ell(u_\ell^k) \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^*).$$

For the reverse implication, we assume $\theta_{\text{mark}}^{1/2} \eta_\ell(u_\ell^*) \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^*)$ and similarly obtain

$$\begin{aligned} [1 - \lambda_{\text{alg}}/\lambda_{\text{alg}}^*] \theta_{\text{mark}}^{1/2} \eta_\ell(u_\ell^k) &\stackrel{(30)}{\leq} \theta_{\text{mark}}^{1/2} \eta_\ell(u_\ell^*) \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^*) \\ &\leq \eta_\ell(\mathcal{R}_\ell; u_\ell^k) + (\lambda_{\text{alg}}/\lambda_{\text{alg}}^*) \eta_\ell(u_\ell^k) \stackrel{(31)}{=} \eta_\ell(\mathcal{R}_\ell; u_\ell^k) + \left(\theta_{\text{mark}}^{1/2} [1 - \lambda_{\text{alg}}/\lambda_{\text{alg}}^*] - \theta^{1/2} \right) \eta_\ell(u_\ell^k). \end{aligned}$$

This yields $\theta \eta_\ell(u_\ell^k)^2 \leq \eta_\ell(\mathcal{R}_\ell; u_\ell^k)^2$ and concludes the proof. \square

Adapting the proof techniques from [BFM⁺25, Proof of Theorem 1], we establish the following *conditional* tail summability result.

Lemma 16 (conditional tail summability of final iterates). *Suppose formally that $\tau < 0$ so that Algorithm B does not terminate. Suppose that the estimator satisfies (A1)–(A3) and that the employed preconditioner satisfies (9). Let $0 < \theta \leq 1$, $C_{\text{mark}} \geq 1$, $u_0^0 \in \mathcal{X}_0$, and $C_{\text{alg}} > 0$ be arbitrary. Recall λ_{alg}^* from Lemma 15. If there exists $\ell_0 \in \mathbb{N}$ such that the solver parameter $\lambda_{\text{alg}} > 0$ satisfies $0 < \lambda_{\text{alg}} < \lambda_{\text{alg}}^* \theta^{1/2}$ for all $\ell \geq \ell_0$, then Algorithm B guarantees tail summability of the quasi-error \mathbf{H}_ℓ^k of the final iterate, i.e., there exist a constant $C_{\text{tail}} \geq 1$ such that*

$$\sum_{\ell'=\ell+1}^{\underline{\ell}-1} \mathbf{H}_{\ell'}^k \leq C_{\text{tail}} \mathbf{H}_\ell^k \quad \text{for all } 0 \leq \ell < \underline{\ell}. \quad (36)$$

Proof. We first show that $(\mathbf{H}_\ell^k)_{\ell=0}^{\underline{\ell}}$ satisfies assumption (23) in Lemma 12. Suppose that $\mathbf{H}_\ell^k = 0$, i.e., $\eta_\ell(u_\ell^k) = 0$ and $\|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} = 0$. Then, it follows immediately that $u_\ell^k = u_\ell^* = u^*$. By C ea's lemma (6) and nested iteration, we obtain $u_\ell^k = u_\ell^* = u^* = u_{\ell'}^* = u_{\ell'}^k$ for all $\ell' > \ell$, and hence also $\mathbf{H}_{\ell'}^k = 0$ for all $\ell' > \ell$. Therefore,

$$\forall \ell < \ell' < \underline{\ell}: \left(\mathbf{H}_\ell^k = 0 \implies \mathbf{H}_{\ell'}^k = 0 \right). \quad (37)$$

The case $\underline{\ell} < \infty$ directly follows from the proof of Lemma 12 for a constant depending on $\{\mathbf{H}_\ell^k\}_{\ell=0}^{\underline{\ell}}$. Let $\ell \in \mathbb{N}$ with $\ell_0 \leq \ell$ and $(\ell + 1, k) \in \mathcal{Q}$. The stopping criterion (10) yields

$$\eta_\ell(u_\ell^k) \leq \eta_\ell(u_\ell^k) + \|\mathbf{s}_\ell^k\|_{\mathcal{P}_\ell^{-1}} \leq (1 + \lambda_{\text{alg}}) \eta_\ell(u_\ell^k).$$

Hence, estimator equivalence (30) yields the equivalence

$$\mathbf{H}_\ell^k \stackrel{(26)}{=} \eta_\ell(u_\ell^k) + \|\mathbf{s}_\ell^k\|_{\mathcal{P}_\ell^{-1}} \simeq \eta_\ell(u_\ell^k) \stackrel{(30)}{\simeq} \eta_\ell(u_\ell^*). \quad (38)$$

Lemma 15 shows that Dörfler marking (24) implies Dörfler marking for the exact discrete solution u_ℓ^* with parameter $\theta' = \theta_{\text{mark}}$ defined in (51), where $\theta' > 0$ since $\lambda_{\text{alg}} < \lambda_{\text{alg}}^* \theta^{1/2} \leq \lambda_{\text{alg}}^*$. Together with stability (A1), reduction (A2), and $q_\theta := 1 - (1 - q_{\text{red}}^2)\theta'$, we obtain

$$\begin{aligned} \eta_{\ell+1}(u_\ell^*)^2 &\stackrel{(A1),(A2)}{\leq} \eta_\ell(\mathcal{T}_{\ell+1} \cap \mathcal{T}_\ell; u_\ell^*)^2 + q_{\text{red}}^2 \eta_\ell(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+1}; u_\ell^*)^2 \\ &= \eta_\ell(u_\ell^*)^2 - (1 - q_{\text{red}}^2) \eta_\ell(\mathcal{T}_\ell \setminus \mathcal{T}_{\ell+1}; u_\ell^*)^2 \\ &\leq \eta_\ell(u_\ell^*)^2 - (1 - q_{\text{red}}^2) \eta_\ell(\mathcal{M}_\ell; u_\ell^*)^2 \stackrel{(24)}{\leq} q_\theta^2 \eta_\ell(u_\ell^*)^2. \end{aligned} \quad (39)$$

Applying stability (A1) again yields

$$\eta_{\ell+1}(u_{\ell+1}^*) \stackrel{(A1)}{\leq} \eta_{\ell+1}(u_\ell^*) + C_{\text{stab}} \|\|u_{\ell+1}^* - u_\ell^*\|\| \stackrel{(39)}{\leq} q_\theta \eta_\ell(u_\ell^*) + C_{\text{stab}} \|\|u_{\ell+1}^* - u_\ell^*\|\| \quad (40)$$

for all $\ell_0 \leq \ell < \underline{\ell}$. Using quasi-orthogonality (A4) and reliability (A3), we can estimate

$$\sum_{\ell'=\ell}^{\ell+N} \|\|u_{\ell'+1}^* - u_{\ell'}^*\|\|^2 \stackrel{(A4)}{\lesssim} (N+1)^{1-\delta} \|\|u_\ell^* - u_{\ell'}^*\|\|^2 \stackrel{(A3)}{\lesssim} (N+1)^{1-\delta} \eta_\ell(u_\ell^*)^2. \quad (41)$$

for all $0 \leq \ell \leq \ell + N < \underline{\ell}$. Hence, we can apply [BLP26, Lemma 9] for $\ell_0 \leq \ell$ and conclude tail summability of $\{\mathbf{H}_\ell^k\}_{\ell \geq \ell_0}$, i.e.,

$$\boxed{\sum_{\ell'=\ell+1}^{\underline{\ell}-1} \mathbf{H}_{\ell'}^k \lesssim \mathbf{H}_\ell^k \quad \text{for all } \ell_0 \leq \ell < \underline{\ell}.} \quad (42)$$

For a potentially larger constant depending also on $(\mathbf{H}_\ell^k)_{\ell=0}^{\ell_0}$, (37) and Lemma 12 yield tail summability of \mathbf{H}_ℓ^k for all levels $0 \leq \ell < \underline{\ell}$, which concludes the proof. \square

The following result guarantees that C_{alg} and λ_{alg} are updated only finitely often.

Lemma 17 (bounded number of updates of C_{alg} and λ_{alg}). *Suppose formally that $\tau < 0$ so that Algorithm B does not terminate. Suppose that the estimator satisfies (A1)–(A3) and that the employed preconditioner satisfies (9). Let $0 < \theta \leq 1$, $C_{\text{mark}} \geq 1$, $u_0^0 \in \mathcal{X}_0$, $\lambda_{\text{alg}} > 0$, and $C_{\text{alg}} > 0$ be arbitrary. Then, the input parameters C_{alg} and λ_{alg} in Algorithm B are updated only finitely many times.*

Proof. We argue by contradiction and assume that the parameters C_{alg} and λ_{alg} are updated infinitely many times. Then, there exists a subsequence $\ell(n)$ such that

$$\sum_{j=0}^{\ell(n)-1} (\mathbf{H}_j^k)^{-1} > C_{\ell(n)} (\mathbf{H}_{\ell(n)}^k)^{-1} \quad \text{for all } n \in \mathbb{N}, \quad (43)$$

where $C_{\ell(n)} \rightarrow \infty$ for $n \rightarrow \infty$. Since $\lambda_{\text{alg}} \rightarrow 0$ as $\ell \rightarrow \infty$ even for the whole sequence, there exists an $\ell_0 \in \mathbb{N}$ such that $\lambda_{\text{alg}} < \lambda_{\text{alg}}^* \theta^{1/2}$ for all $\ell \geq \ell_0$. Consequently, the assumptions of

Proposition 16 are satisfied and yield tail summability of \mathbf{H}_ℓ^k in the levels ℓ . Lemma 12 shows that (36) is equivalent to inverse summability (21), i.e.,

$$\sum_{j=0}^{\ell-1} (\mathbf{H}_j^k)^{-1} \leq C_{\text{inv}} (\mathbf{H}_\ell^k)^{-1} \quad \text{for all } \ell \in \mathbb{N}_0$$

for some constant $C_{\text{inv}} > 0$. This contradicts (43) and concludes the proof. \square

As a consequence of Lemmas 16 and 17, adapting the proof techniques from [BFM⁺25, Proof of Theorem 2] yields unconditional full R-linear convergence of Algorithm B.

Theorem 18 (unconditional full R-linear convergence of Algorithm B). *Suppose formally that $\tau < 0$ so that Algorithm B does not terminate. Suppose that the estimator satisfies (A1)–(A3) and that the employed preconditioner satisfies (9). Let $0 < \theta \leq 1$, $C_{\text{mark}} \geq 1$, $u_0^0 \in \mathcal{X}_0$, $\lambda_{\text{alg}} > 0$, and $C_{\text{alg}} > 0$ be arbitrary. Then, Algorithm B guarantees full R-linear convergence of the quasi-error \mathbf{H}_ℓ^k from (26), i.e., there exist constants $0 < q_{\text{lin}} < 1$ and $C_{\text{lin}} > 0$ such that*

$$\mathbf{H}_\ell^k \leq C_{\text{lin}} q_{\text{lin}}^{|\ell, k| - |\ell', k'|} \mathbf{H}_{\ell'}^{k'} \quad \text{for all } (\ell', k'), (\ell, k) \in \mathcal{Q} \text{ with } |\ell', k'| \leq |\ell, k|. \quad (44)$$

Proof. The proof is split into two steps.

Step 1 (tail summability with respect to ℓ). Lemma 17 guarantees that C_{alg} and λ_{alg} are updated only finitely often, i.e., there exists $\ell_0 \in \mathbb{N}_0$ and $C > 0$ such that

$$\sum_{j=0}^{\ell-1} (\mathbf{H}_j^k)^{-1} \leq C (\mathbf{H}_\ell^k)^{-1} \quad \text{for all } \ell_0 \leq \ell \in \mathbb{N}_0. \quad (45)$$

For a potentially bigger constant $C > 0$ depending also on $(\mathbf{H}_\ell^k)_{\ell=0}^{\ell_0}$, (37) and Lemma 12 yield tail summability of \mathbf{H}_ℓ^k in ℓ .

Step 2 (tail summability with respect to ℓ and k). First, for $0 \leq k < k' < \underline{k}[\ell]$, using that the solver stopping criterion (10) has not been met, we have

$$\mathbf{H}_\ell^{k'} \stackrel{(26)}{=} \eta_\ell(u_\ell^{k'}) + \|\mathbf{s}_\ell^{k'}\|_{\mathbf{P}_\ell^{-1}} \stackrel{(10)}{\lesssim} \|\mathbf{s}_\ell^{k'}\|_{\mathbf{P}_\ell^{-1}} \stackrel{(13)}{\leq} q_{\text{alg}}^{k'-k} \|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} \leq q_{\text{alg}}^{k'-k} \mathbf{H}_\ell^k.$$

Second, stability (A1), equivalence (12), and contraction (13) prove that

$$\begin{aligned} \mathbf{H}_\ell^k &\stackrel{(26)}{=} \eta_\ell(u_\ell^k) + \|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} \stackrel{(A1)}{\lesssim} \eta_\ell(u_\ell^{k-1}) + \|\|u_\ell^k - u_\ell^{k-1}\|\| + \|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} \\ &\stackrel{(12)}{\lesssim} \eta_\ell(u_\ell^{k-1}) + \|\mathbf{s}_\ell^k\|_{\mathbf{P}_\ell^{-1}} + \|\mathbf{s}_\ell^{k-1}\|_{\mathbf{P}_\ell^{-1}} \stackrel{(13)}{\lesssim} \eta_\ell(u_\ell^{k-1}) + \|\mathbf{s}_\ell^{k-1}\|_{\mathbf{P}_\ell^{-1}} = \mathbf{H}_\ell^{k-1} \end{aligned}$$

for all $(\ell, \underline{k}) \in \mathcal{Q}$. Hence, we may conclude

$$\boxed{\mathbf{H}_\ell^{k'} \lesssim q_{\text{alg}}^{k'-k} \mathbf{H}_\ell^k \quad \text{for all } 0 \leq k \leq k' \leq \underline{k}[\ell].} \quad (46)$$

Nested iteration $u_{\ell+1}^0 = u_\ell^k$, reduction (A2), and (41) with $N = 0$, proves

$$\begin{aligned} \mathbf{H}_{\ell+1}^0 &= \eta_{\ell+1}(u_{\ell+1}^0) + \|\mathbf{s}_{\ell+1}^0\|_{\mathbf{P}_{\ell+1}^{-1}} \stackrel{(12)}{\lesssim} \eta_{\ell+1}(u_\ell^k) + \|\|u_{\ell+1}^* - u_\ell^k\|\| \\ &\stackrel{(39)}{\lesssim} \eta_\ell(u_\ell^k) + \|\|u_{\ell+1}^* - u_\ell^k\|\| \leq \mathbf{H}_\ell^k + \|\|u_{\ell+1}^* - u_\ell^k\|\| \stackrel{(41)}{\lesssim} \mathbf{H}_\ell^k + \eta_\ell(u_\ell^*) \stackrel{(27)}{\simeq} \mathbf{H}_\ell^k \end{aligned}$$

for all $(\ell, \underline{k}) \in \mathcal{Q}$. This leads to

$$\boxed{\mathbf{H}_{\ell+1}^0 \lesssim \mathbf{H}_\ell^k \quad \text{for all } (\ell, \underline{k}) \in \mathcal{Q}.} \quad (47)$$

Overall, the geometric series proves tail summability

$$\begin{aligned} \sum_{\substack{(\ell', k') \in \mathcal{Q} \\ |\ell', k'| > |\ell, k|}} \mathbf{H}_{\ell'}^{k'} &= \sum_{k'=k+1}^{\underline{k}[\ell]} \mathbf{H}_\ell^{k'} + \sum_{\ell'=\ell+1}^{\underline{\ell}} \sum_{k'=0}^{\underline{k}[\ell']} \mathbf{H}_{\ell'}^{k'} \\ &\stackrel{(46)}{\lesssim} \mathbf{H}_\ell^k + \sum_{\ell'=\ell+1}^{\underline{\ell}} \mathbf{H}_{\ell'}^0 \stackrel{(47)}{\lesssim} \mathbf{H}_\ell^k + \sum_{\ell'=\ell}^{\underline{\ell}-1} \mathbf{H}_{\ell'}^k \stackrel{(42)}{\lesssim} \mathbf{H}_\ell^k + \mathbf{H}_\ell^k \stackrel{(46)}{\lesssim} \mathbf{H}_\ell^k \quad \text{for all } (\ell, k) \in \mathcal{Q}. \end{aligned}$$

Finally, Lemma 12 concludes the proof of full R-linear convergence (44). \square

Corollary 19 (unconditional plain convergence of Algorithm B). *Under the assumptions of Theorem 18, there holds convergence*

$$\| \| u^* - u_\ell^k \| \| + \| \| u^* - u_\ell^* \| \| + \| \| u_\ell^* - u_\ell^k \| \| + \eta_\ell(u_\ell^*) + \eta_\ell(u_\ell^k) \rightarrow 0 \quad \text{as } |\ell, k| \rightarrow \infty. \quad (48)$$

Proof. From reliability (A3) and (27), we deduce that

$$\begin{aligned} \| \| u^* - u_\ell^k \| \| + \| \| u^* - u_\ell^* \| \| + \| \| u_\ell^* - u_\ell^k \| \| + \eta_\ell(u_\ell^*) &\leq 2(\| \| u^* - u_\ell^* \| \| + \| \| u_\ell^* - u_\ell^k \| \|) + \eta_\ell(u_\ell^*) \\ &\stackrel{(A3)}{\leq} (1 + 2C_{\text{rel}}) \eta_\ell(u_\ell^*) + 2 \| \| u_\ell^* - u_\ell^k \| \| \stackrel{(27)}{\leq} \max\{1 + 2C_{\text{rel}}, 2\} C_{\text{eq}} \mathbf{H}_\ell^k. \end{aligned}$$

Full R-linear convergence (44) for $(\ell', k') = (0, 0) \in \mathcal{Q}$ concludes the proof of (48) via

$$\begin{aligned} \| \| u^* - u_\ell^k \| \| + \| \| u^* - u_\ell^* \| \| + \| \| u_\ell^* - u_\ell^k \| \| + \eta_\ell(u_\ell^*) + \eta_\ell(u_\ell^k) &\leq (1 + \max\{1 + 2C_{\text{rel}}, 2\} C_{\text{eq}}) \mathbf{H}_\ell^k \\ &\stackrel{(44)}{\leq} (1 + \max\{1 + 2C_{\text{rel}}, 2\} C_{\text{eq}}) C_{\text{lin}} q_{\text{lin}}^{|\ell, k|} \mathbf{H}_0^0 \xrightarrow{|\ell, k| \rightarrow \infty} 0. \quad \square \end{aligned}$$

6. OPTIMAL COMPLEXITY OF ALGORITHM B

First, we comment on the computational cost of Algorithm B: The cost of PGMRES for any step k on a level ℓ is $O(\#\mathcal{T}_\ell)$ as discussed in Remark 4. Calculating the error estimator $\eta_\ell(u_\ell^k)$ also has (up to quadrature) cost $O(\#\mathcal{T}_\ell)$. The marking step can be implemented with cost $O(\#\mathcal{T}_\ell)$; see [Ste07] for $C_{\text{mark}} = 2$ and [PP20] for $C_{\text{mark}} = 1$. Finally, the cost of mesh-refinement by NVB is also $O(\#\mathcal{T}_\ell)$; see, e.g., [Ste08; DGS25]. Due to the cumulative structure of Algorithm B, the total computational cost to compute u_ℓ^k is therefore, up to a multiplicative constant, given by

$$\text{cost}(\ell, k) := \sum_{\substack{(\ell', k') \in \mathcal{Q} \\ |\ell', k'| \leq |\ell, k|}} \#\mathcal{T}_{\ell'}. \quad (49)$$

A crucial consequence of full R-linear convergence is the following, which is also implicitly found in [GHPS21]. If the rate of convergence $s > 0$ is achievable with respect to the degrees of freedom, it is also achievable with respect to the cumulative computational costs. Hence, full R-linear convergence is the key to optimal complexity.

Corollary 20 (rates = complexity [BFM⁺25, Corollary 1]). *Suppose full R -linear convergence (44). Then, for any $s > 0$, it holds that*

$$M(s) := \sup_{(\ell,k) \in \mathcal{Q}} (\#\mathcal{T}_\ell)^s \mathbf{H}_\ell^k \leq \sup_{(\ell,k) \in \mathcal{Q}} \left(\sum_{\substack{(\ell',k') \in \mathcal{Q} \\ |\ell',k'| \leq |\ell,k|}} \#\mathcal{T}_{\ell'} \right)^s \mathbf{H}_\ell^k \leq \frac{C_{\text{lin}}}{(1 - q_{\text{lin}}^{1/s})^s} M(s). \quad (50)$$

Moreover, there exists $s_0 > 0$ such that $M(s) < \infty$ for all $0 < s \leq s_0$. \square

To describe whether the solution u^* can be approximated at rate $s > 0$, we use the notion of nonlinear approximation classes [BDDP02; BDD04; Ste07; CKNS08; CFPP14]

$$\|u^*\|_{\mathbb{A}_s} := \sup_{N \in \mathbb{N}_0} \left((N+1)^s \min_{\mathcal{T}_{\text{opt}} \in \mathbb{T}_N} \eta_{\text{opt}}(u_{\text{opt}}^*) \right),$$

where $\eta_{\text{opt}}(u_{\text{opt}}^*)$ is the estimator for the (unavailable) exact Galerkin solution u_{opt}^* on an optimal mesh $\mathcal{T}_{\text{opt}} \in \mathbb{T}_N$. The second main result of this work states that for sufficiently small parameters θ and λ_{alg} the adaptive algorithm leads to optimal (in the sense of the nonlinear approximation classes) decrease of the quasi-error with respect to overall computational cost. The proof can be derived following steps analogous to [GHPS21, Proof of Theorem 8]. For details, we refer to Appendix A.

Theorem 21 (optimal complexity). *Consider arbitrary $0 < \theta \leq 1$, $0 < \lambda_{\text{alg}}$, and arbitrary $u_0^0 \in \mathcal{X}_0$. Suppose that the estimator satisfies (A1)–(A3⁺) and that the preconditioner satisfies (9). Recall λ_{alg}^* and θ_{mark} from Lemma 15. Let $\theta^* := (1 + C_{\text{stab}}^2 C_{\text{drel}}^2)^{-1}$. Suppose that θ , λ_{alg} (as the input for Algorithm B) are sufficiently small in the sense of*

$$0 < \lambda_{\text{alg}} < \lambda_{\text{alg}}^* \quad \text{and} \quad 0 < \theta_{\text{mark}} = \frac{(\theta^{1/2} + \lambda_{\text{alg}}/\lambda_{\text{alg}}^*)^2}{(1 - \lambda_{\text{alg}}/\lambda_{\text{alg}}^*)^2} < \theta^* < 1, \quad (51)$$

Then, Algorithm B guarantees, for all $s > 0$, that

$$c_{\text{opt}} \|u^*\|_{\mathbb{A}_s} \leq \sup_{(\ell,k) \in \mathcal{Q}} \left(\sum_{\substack{(\ell',k') \in \mathcal{Q} \\ |\ell',k'| \leq |\ell,k|}} \#\mathcal{T}_{\ell'} \right)^s \mathbf{H}_\ell^k \leq C_{\text{opt}} \max\{\|u^*\|_{\mathbb{A}_s}, \mathbf{H}_0^0\}. \quad (52)$$

The constant $c_{\text{opt}} > 0$ depends only on C_{stab} , the use of NVB, and s ; while the constant C_{opt} depends only on C_{stab} , C_{drel} , C_{mark} , C_{mesh} , q_{alg} , λ_{alg} , θ , $\#\mathcal{T}_0$, and s .

7. NUMERICAL EXPERIMENTS

The performance of Algorithm B is tested on the L-shaped domain $\Omega := (-1, 1)^2 \setminus [0, 1] \times [-1, 0]$ with the diffusion coefficient $\mathbf{K} = I_{2 \times 2}$, convection coefficient $\mathbf{b} := (1, 25)^\top$, and reaction coefficient $c = 0$. The unknown analytical solution $u^* \in H_0^1(\Omega)$ satisfies

$$-\Delta u^* + \begin{pmatrix} 1 \\ 25 \end{pmatrix} \cdot \nabla u^* = 1 \quad \text{in } \Omega.$$

As seen in Figure 1, this problem induces refinement in the reentrant corner of the domain as well as additionally on the boundary layer due to convection.

7.1. Optimality of the adaptive algorithm. In the following, k_{max} will denote the number of maximal iteration steps of PGMRES. Then, $k_\ell[\ell]$ refers to the overall iterations, with the understanding that for a given maximal iteration steps k_{max} of PGMRES, $R = \lfloor k/k_{\text{max}} \rfloor$ denotes the number of PGMRES restart steps.

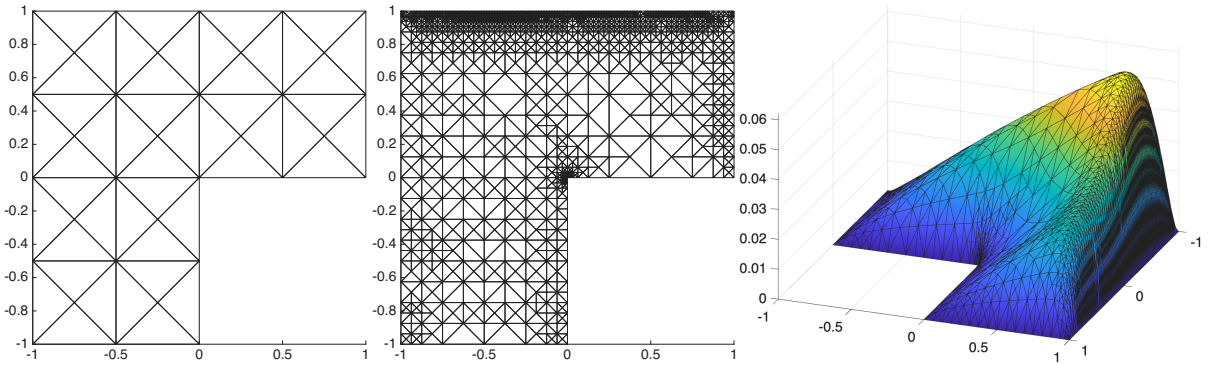


FIGURE 1. Initial mesh (left), with $\#\mathcal{T}_0 = 48$ elements, used to initialize Algorithm B together with polynomial degree $p = 2$ (leading to $N_0 = 81$ initial degrees of freedom), bulk parameter $\theta = 0.5$, and maximum PGMRES steps $k_{\max} = 5$. The resulting mesh (center) after 10 refinements with $\#\mathcal{T}_{10} = 3,396$ elements (leading to $N_{10} = 7,773$ degrees of freedom) and corresponding approximation u_{10}^k (right).

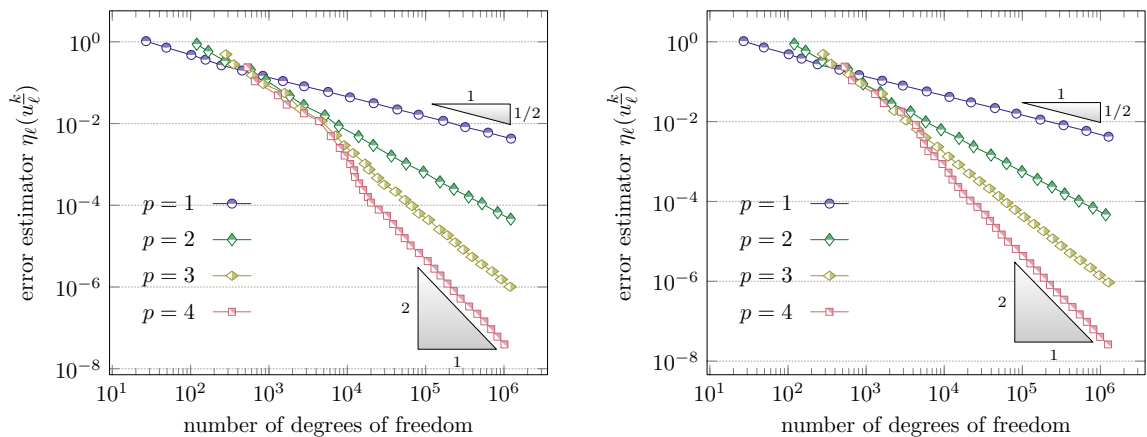


FIGURE 2. Rate-optimality of the Algorithm B using the additive Schwarz preconditioner from Section 3.2 for GMRES on the linear system with $\theta = 0.5$ and $k_{\max} = 5$ (left), $k_{\max} = 1$ (right).

First, we see in Figure 2 that Algorithm B indeed exhibits optimal convergence rates with respect to number of degrees of freedom. Next, we are interested in optimal complexity of our adaptive algorithm, as stated in Theorem 21. Indeed, we observe that optimal convergence rates are achieved with respect to overall computational cost in the sense of (49), see Figure 3, as well as with respect to time, see Figure 4. In particular, optimal complexity holds also for the minimal approach of $k_{\max} = 1$ PGMRES steps.

7.2. Adaptive stopping criteria. First, we consider the performance of the new adaptive control of the parameters C_{alg} and λ_{alg} in step (ii.a) in Algorithm B. This plays a crucial role in proving unconditional convergence in Theorem 18. Indeed, we know from Lemma 17 that the updates can only occur finitely often. In practice, Figure 5 illustrates that for several polynomial degrees, these parameters only need to be updated twice when initialized with 1. Thus, the resulting parameter choices appear to be in a standard range

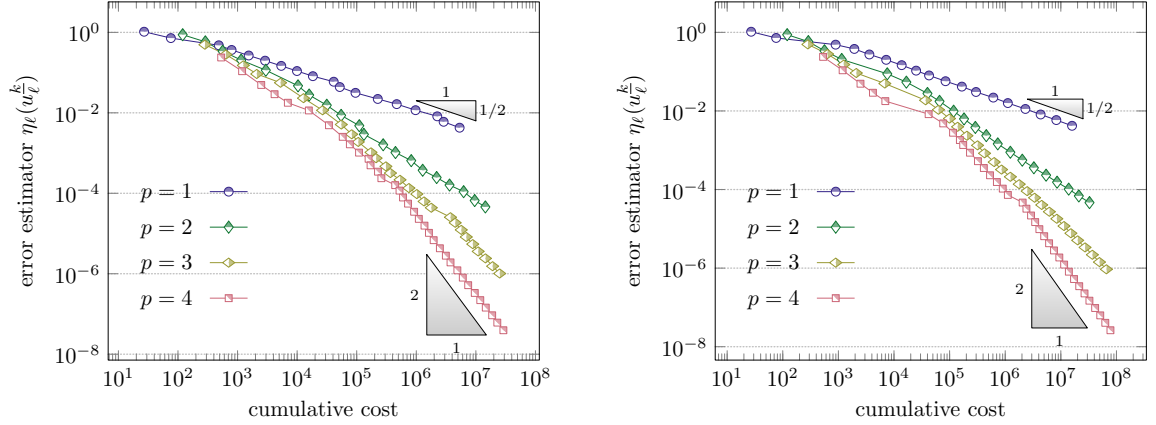


FIGURE 3. Cost-optimality of the Algorithm B using the additive Schwarz preconditioner from Section 3.2 for GMRES on the linear system, $\theta = 0.5$ and $k_{\max} = 5$ (left), $k_{\max} = 1$ (right).

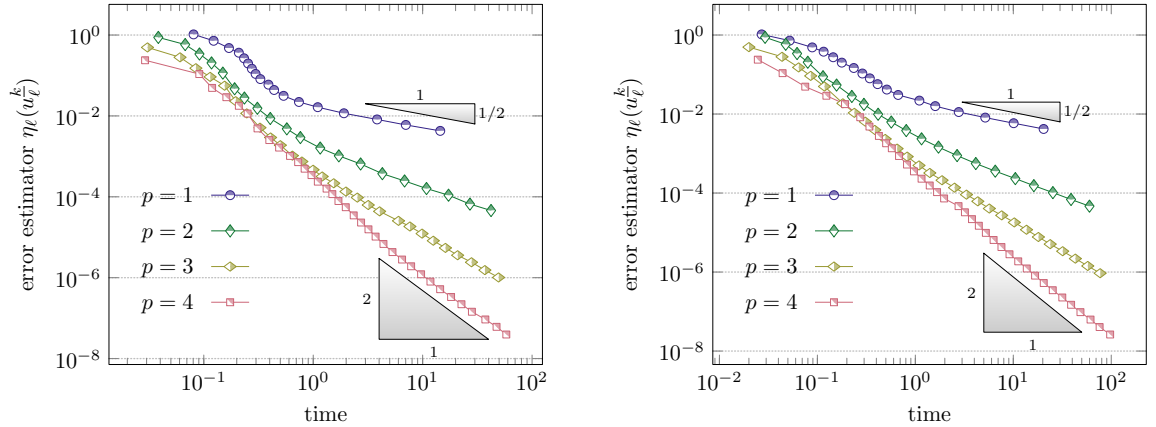


FIGURE 4. Optimal complexity of the Algorithm B using the additive Schwarz preconditioner from Section 3.2 for GMRES on the linear system, $\theta = 0.5$ and $k_{\max} = 5$ (left), $k_{\max} = 1$ (right).

(neither too large, nor too small) for a well-performing adaptive algorithm. In particular, the update of C_{alg} and λ_{alg} seems to be independent of the choice of k_{\max} .

Next, let us discuss the stopping of the algebraic solver. While the equibalancing of discretization and algebra errors is now ensured by the adaptive regulation of the parameter λ_{alg} , we test two different numbers k_{\max} of maximal iteration steps for PGMRES. As seen in Figure 6, there is a trade-off between a low k_{\max} and higher number of potential restarts and vice-versa. Nonetheless, it is noticeable that the total number of performed steps \underline{k} remains generally smaller for $k_{\max} = 5$ when compared to $k_{\max} = 1$.

7.3. Solver robustness. To conduct some experiments tailored on the robustness of the iterative solver, we start by pre-computing a mesh hierarchy of 20 levels. We do so via adaptive refinement employing $p = 2$ and a direct solver. Next, we run optimally preconditioned GMRES via additive Schwarz from Section 3.2, from a zero initial guess and stop the iterations when the algebraic relative residual drops below 10^{-5} .

Figure 7 shows overall p -robust contraction factors, as expected theoretically on the contraction bound of the preconditioned residuals $\|\mathbf{r}_\ell^k\|_{\mathbf{P}_\ell^{-1}}$ from Proposition 7. Note

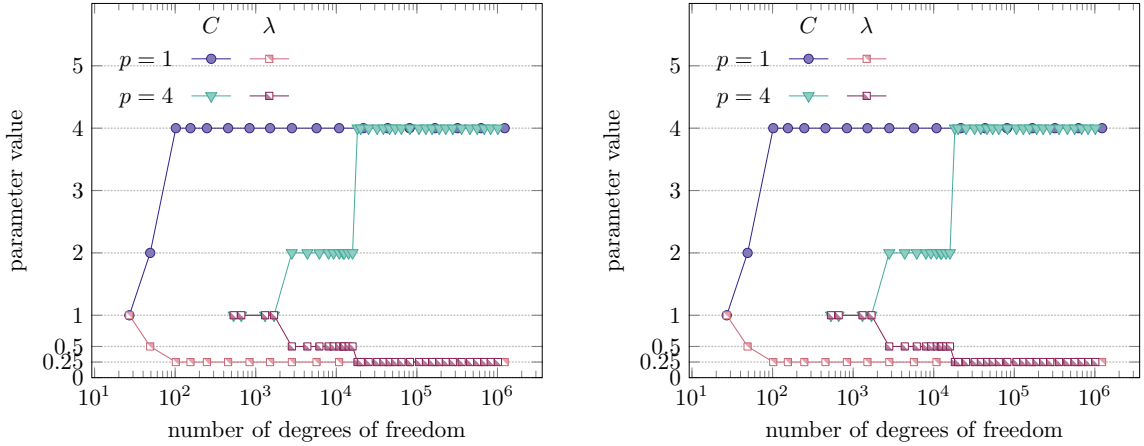


FIGURE 5. Parameter adaptation in Algorithm B(ii.a) with respect to the degrees of freedom for $\theta = 0.5$ and $k_{\max} = 5$ (left), $k_{\max} = 1$ (right).

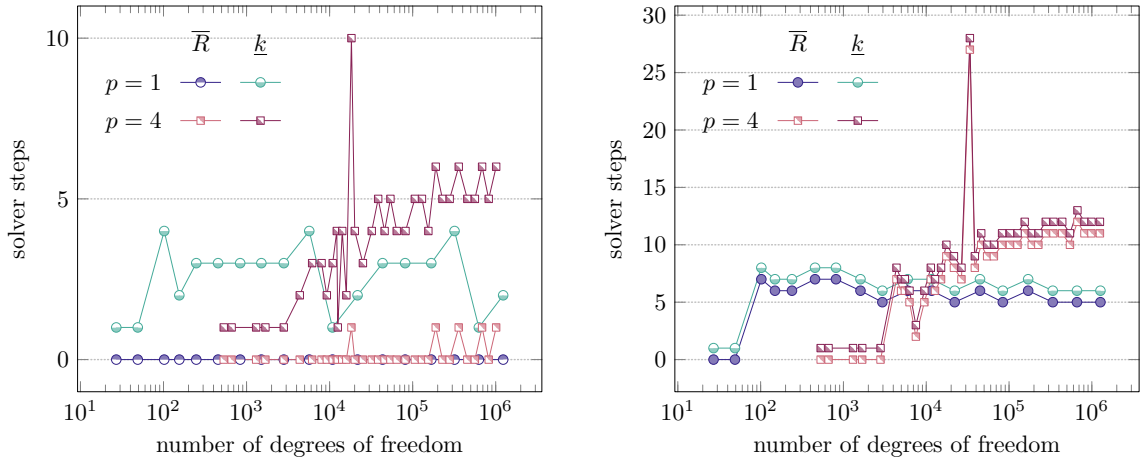


FIGURE 6. Iteration steps and number of restarts $\bar{R} = \lfloor k[\ell]/k_{\max} \rfloor$ for PGMRES with additive Schwarz preconditioner from Section 3.2 when employing Algorithm B for $\theta = 0.5$ and $k_{\max} = 5$ (left) and $k_{\max} = 1$ (right).

that, since Proposition 7 is ensured at each solver step, we indeed observe it numerically for different choices of k_{\max} of PGMRES, in particular also for $k_{\max} = 1$, i.e., we can restart at each step. Figure 7 (left) illustrates how the number of restarts can make the contraction factors very oscillatory. For $\text{mod}(k, k_{\max}) = 0$, the Krylov space attains its maximal dimension k_{\max} , so the contraction factor is minimal within the corresponding GMRES run. After each restart, the contraction factor increases; see Figure 7 (left).

Experimentally we observe in Figure 8 (left) even better algebraic error contraction in the energy norm $\|u_\ell^* - u_\ell^k\|$, though this is outside the scope of the theoretic framework. Finally, we observe in Figure 8 the behavior for a non-restarted PGMRES. This is done here using $k_{\max} = 300$ which leads to a full Arnoldi basis and thus to better contraction as seen in Figure 8 (left) and faster algebraic error decay as seen in Figure 8 (right).

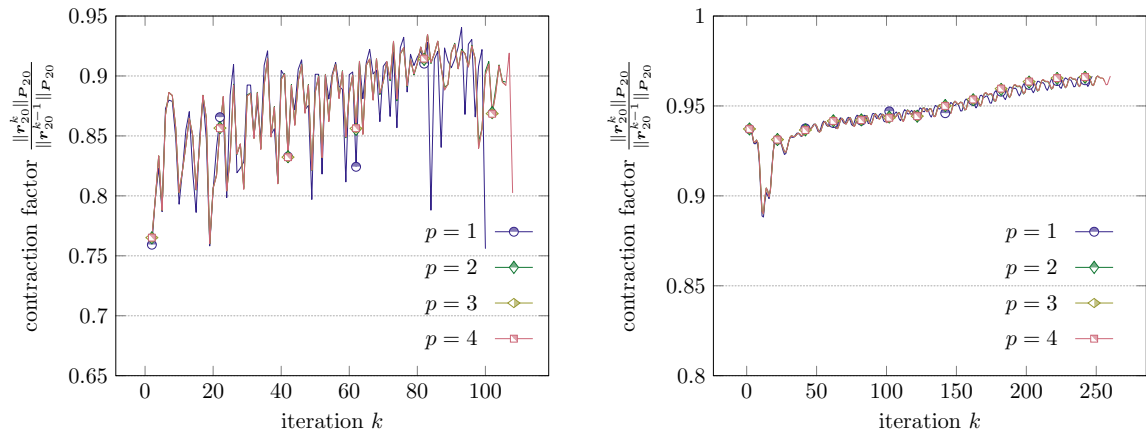


FIGURE 7. Contraction factors of the preconditioned residual for $k_{\max} = 5$ (left) and $k_{\max} = 1$ (right) for different polynomial degrees on a fixed mesh \mathcal{T}_{20} (obtained by Algorithm B for $p = 2$) with $\#\mathcal{T}_{20} = 99,085$ elements.

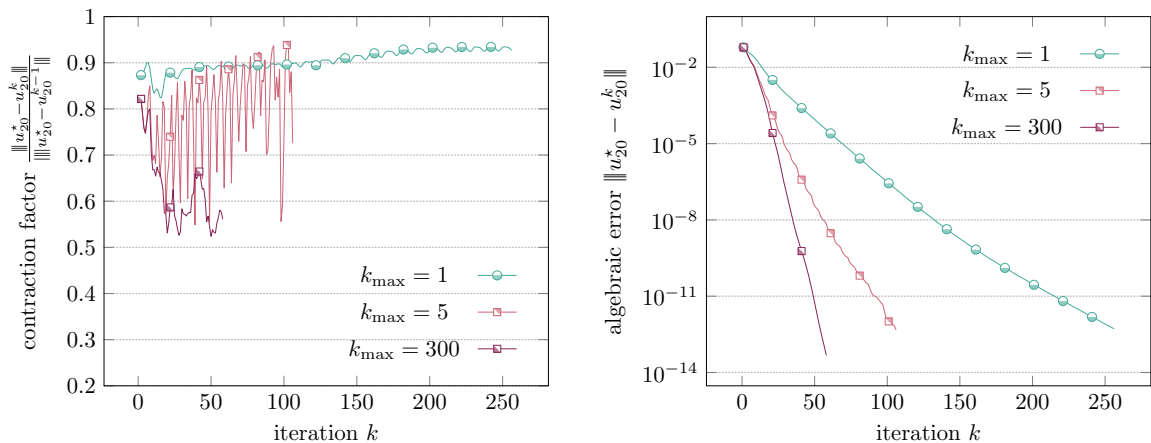


FIGURE 8. Contraction factors (left) and decrease (right) of the algebraic error in energy norm for $p = 3$ and different choices of k_{\max} on a fixed mesh \mathcal{T}_{20} (obtained by Algorithm B for $p = 2$) with $\#\mathcal{T}_{20} = 99,085$ elements.

REFERENCES

- [AFF⁺15] M. Aurada, M. Feischl, T. Führer, M. Karkulik, and D. Praetorius. Energy norm based error estimators for adaptive BEM for hypersingular integral equations. *Appl. Numer. Math.*, 95:15–35, 2015. DOI: [10.1016/j.apnum.2013.12.004](https://doi.org/10.1016/j.apnum.2013.12.004).
- [BDD04] P. Binev, W. Dahmen, and R. DeVore. Adaptive finite element methods with convergence rates. *Numer. Math.*, 97(2):219–268, 2004. DOI: [10.1007/s00211-003-0492-7](https://doi.org/10.1007/s00211-003-0492-7).
- [BDDP02] P. Binev, W. Dahmen, R. DeVore, and P. Petrushev. Approximation classes for adaptive methods. *Serdica Math. J.*, 28(4):391–416, 2002.
- [BDK12] L. Belenki, L. Diening, and C. Kreuzer. Optimality of an adaptive finite element method for the p -Laplacian equation. *IMA J. Numer. Anal.*, 32(2):484–510, 2012. DOI: [10.1093/imanum/drr016](https://doi.org/10.1093/imanum/drr016).
- [BFM⁺25] P. Bringmann, M. Feischl, A. Miraçi, D. Praetorius, and J. Streitberger. On full linear convergence and optimal complexity of adaptive FEM with inexact solver. *Comput. Math. Appl.*, 180:102–129, 2025. DOI: [10.1016/j.camwa.2024.12.013](https://doi.org/10.1016/j.camwa.2024.12.013).

- [BHP17] A. Bespalov, A. Haberl, and D. Praetorius. Adaptive FEM with coarse initial mesh guarantees optimal convergence rates for compactly perturbed elliptic problems. *Comput. Methods Appl. Mech. Engrg.*, 317:318–340, 2017. DOI: [10.1016/j.cma.2016.12.014](https://doi.org/10.1016/j.cma.2016.12.014).
- [BIM⁺24] M. Brunner, M. Innerberger, A. Miraçi, D. Praetorius, J. Streitberger, and P. Heid. Adaptive FEM with quasi-optimal overall cost for nonsymmetric linear elliptic PDEs. *IMA J. Numer. Anal.*, 44(3):1560–1596, 2024. DOI: [10.1093/imanum/drad039](https://doi.org/10.1093/imanum/drad039).
- [BLP26] P. Bringmann, C. Lietz, and D. Praetorius. Unconditional full linear convergence and quasi-optimal complexity of smoothed adaptive finite element methods, 2026. arXiv: [2601.20677](https://arxiv.org/abs/2601.20677).
- [BMP24] P. Bringmann, A. Miraçi, and D. Praetorius. Iterative solvers in adaptive FEM. *Advances in Applied Mechanics*, 59:147–212, 2024. DOI: [10.1016/bs.aams.2024.08.002](https://doi.org/10.1016/bs.aams.2024.08.002).
- [CDD01] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet methods for elliptic operator equations: convergence rates. *Math. Comp.*, 70(233):27–75, 2001. DOI: [10.1090/S0025-5718-00-01252-7](https://doi.org/10.1090/S0025-5718-00-01252-7).
- [CDD03] A. Cohen, W. Dahmen, and R. DeVore. Adaptive wavelet schemes for nonlinear variational problems. *SIAM J. Numer. Anal.*, 41(5):1785–1823, 2003. DOI: [10.1137/S0036142902412269](https://doi.org/10.1137/S0036142902412269).
- [CFPP14] C. Carstensen, M. Feischl, M. Page, and D. Praetorius. Axioms of adaptivity. *Comput. Math. Appl.*, 67(6):1195–1253, 2014. DOI: [10.1016/j.camwa.2013.12.003](https://doi.org/10.1016/j.camwa.2013.12.003).
- [CKNS08] J. Cascón, C. Kreuzer, R. Nochetto, and K. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46(5):2524–2550, 2008. DOI: [10.1137/07069047x](https://doi.org/10.1137/07069047x).
- [CN12] J. Cascón and R. Nochetto. Quasioptimal cardinality of AFEM driven by nonresidual estimators. *IMA J. Numer. Anal.*, 32(1):1–29, 2012. DOI: [10.1093/imanum/drr014](https://doi.org/10.1093/imanum/drr014).
- [CNX12] L. Chen, R. Nochetto, and J. Xu. Optimal multilevel methods for graded bisection grids. *Numer. Math.*, 120(1):1–34, 2012. DOI: [10.1007/s00211-011-0401-4](https://doi.org/10.1007/s00211-011-0401-4).
- [DGS25] L. Diening, L. Gehring, and J. Storn. Adaptive mesh refinement for arbitrary initial triangulations. *Found. Comput. Math.*, published online first, 2025. DOI: [10.1007/s10208-025-09698-7](https://doi.org/10.1007/s10208-025-09698-7).
- [Dör96] W. Dörfler. A convergent adaptive algorithm for Poisson’s equation. *SIAM J. Numer. Anal.*, 33(3):1106–1124, 1996. DOI: [10.1137/0733054](https://doi.org/10.1137/0733054).
- [EES83] S. C. Eisenstat, H. C. Elman, and M. H. Schultz. Variational iterative methods for nonsymmetric systems of linear equations. *SIAM J. Numer. Anal.*, 20(2):345–357, 1983. DOI: [10.1137/0720023](https://doi.org/10.1137/0720023).
- [Fei22] M. Feischl. Inf-sup stability implies quasi-orthogonality. *Math. Comp.*, 91(337):2059–2094, 2022. DOI: [10.1090/mcom/3748](https://doi.org/10.1090/mcom/3748).
- [FFP14] M. Feischl, T. Führer, and D. Praetorius. Adaptive FEM with optimal convergence rates for a certain class of nonsymmetric and possibly nonlinear problems. *SIAM J. Numer. Anal.*, 52(2):601–625, 2014. DOI: [10.1137/120897225](https://doi.org/10.1137/120897225).
- [GHPS18] G. Gantner, A. Haberl, D. Praetorius, and B. Stiftner. Rate optimal adaptive FEM with inexact solver for nonlinear operators. *IMA J. Numer. Anal.*, 38(4):1797–1831, 2018. DOI: [10.1093/imanum/drx050](https://doi.org/10.1093/imanum/drx050).
- [GHPS21] G. Gantner, A. Haberl, D. Praetorius, and S. Schimanko. Rate optimality of adaptive finite element methods with respect to overall computational costs. *Math. Comp.*, 90(331):2011–2040, 2021. DOI: [10.1090/mcom/3654](https://doi.org/10.1090/mcom/3654).
- [GMZ12] E. Garau, P. Morin, and C. Zuppa. Quasi-Optimal Convergence Rate of an AFEM for quasi-linear Problems of Monotone Type. *Numer. Math: Theory, Meth. Appl.*, 5(2):131–156, 2012. DOI: [10.4208/nmtma.2012.m1023](https://doi.org/10.4208/nmtma.2012.m1023).
- [GP14] S. Güttel and J. Pestana. Some observations on weighted GMRES. *Numer. Algorithms*, 67(4):733–752, 2014. DOI: [10.1007/s11075-013-9820-x](https://doi.org/10.1007/s11075-013-9820-x).

- [GPS96] A. Greenbaum, V. Pták, and Z. Strakoš. Any nonincreasing convergence curve is possible for GMRES. *SIAM J. Matrix Anal. Appl.*, 17(3):465–469, 1996. DOI: [10.1137/S0895479894275030](https://doi.org/10.1137/S0895479894275030).
- [HMP26] P. Hilbert, A. Miraçi, and D. Praetorius. Generalized preconditioned conjugate gradients for adaptive FEM with optimal complexity, 2026. arXiv: [2601.14911](https://arxiv.org/abs/2601.14911).
- [HPSV21] A. Haberl, D. Praetorius, S. Schimanko, and M. Vohralík. Convergence and quasi-optimal cost of adaptive algorithms for nonlinear operators including iterative linearization and algebraic solver. *Numer. Math.*, 147(3):679–725, 2021. DOI: [10.1007/s00211-021-01176-w](https://doi.org/10.1007/s00211-021-01176-w).
- [HPW21] P. Heid, D. Praetorius, and T. Wihler. Energy contraction and optimal convergence of adaptive iterative linearized finite element methods. *Comput. Methods Appl. Math.*, 21(2):407–422, 2021. DOI: [10.1515/cmam-2021-0025](https://doi.org/10.1515/cmam-2021-0025).
- [IMPS24] M. Innerberger, A. Miraçi, D. Praetorius, and J. Streitberger. *hp*-robust multigrid solver on locally refined meshes for FEM discretizations of symmetric elliptic PDEs. *ESAIM Math. Model. Numer. Anal.*, 58(1):247–272, 2024. DOI: [10.1051/m2an/2023104](https://doi.org/10.1051/m2an/2023104).
- [KPP13] M. Karkulik, D. Pavlicek, and D. Praetorius. On 2D newest vertex bisection: optimality of mesh-closure and H^1 -stability of L_2 -projection. *Constr. Approx.*, 38(2):213–234, 2013. DOI: [10.1007/s00365-013-9192-4](https://doi.org/10.1007/s00365-013-9192-4).
- [MN05] K. Mekchay and R. H. Nochetto. Convergence of adaptive finite element methods for general second order linear elliptic PDEs. *SIAM J. Numer. Anal.*, 43(5):1803–1827, 2005. DOI: [10.1137/04060929X](https://doi.org/10.1137/04060929X).
- [MNS00] P. Morin, R. Nochetto, and K. G. Siebert. Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.*, 38(2):466–488, 2000. DOI: [10.1137/s0036142999360044](https://doi.org/10.1137/s0036142999360044).
- [PP20] C. Pfeiler and D. Praetorius. Dörfler marking with minimal cardinality is a linear complexity problem. *Math. Comp.*, 89(326):2735–2752, 2020. DOI: [10.1090/mcom/3553](https://doi.org/10.1090/mcom/3553).
- [PW13] J. Pestana and A. J. Wathen. On the choice of preconditioner for minimum residual methods for non-Hermitian matrices. *J. Comput. Appl. Math.*, 249:57–68, 2013. DOI: [10.1016/j.cam.2013.02.020](https://doi.org/10.1016/j.cam.2013.02.020).
- [Spi24] N. Spillane. Hermitian preconditioning for a class of non-Hermitian linear systems. *SIAM J. Sci. Comput.*, 46(3):A1903–A1922, 2024. DOI: [10.1137/23M1559026](https://doi.org/10.1137/23M1559026).
- [SS07] M. Sarkis and D. B. Szyld. Optimal left and right additive Schwarz preconditioning for minimal residual methods with Euclidean and energy norms. *Comput. Methods Appl. Mech. Engrg.*, 196(8):1612–1621, 2007. DOI: [10.1016/j.cma.2006.03.027](https://doi.org/10.1016/j.cma.2006.03.027).
- [SS24] N. Spillane and D. B. Szyld. New convergence analysis of GMRES with weighted norms, preconditioning, and deflation, leading to a new deflation space. *SIAM J. Matrix Anal. Appl.*, 45(4):1721–1745, 2024. DOI: [10.1137/23M1622398](https://doi.org/10.1137/23M1622398).
- [SS86] Y. Saad and M. Schultz. GMRES: a generalized minimal residual algorithm for solving nonsymmetric linear systems. *SIAM J. Sci. Statist. Comput.*, 7(3):856–869, 1986. DOI: [10.1137/0907058](https://doi.org/10.1137/0907058).
- [Ste07] R. Stevenson. Optimality of a standard adaptive finite element method. *Found. Comput. Math.*, 7(2):245–269, 2007. DOI: [10.1007/s10208-005-0183-0](https://doi.org/10.1007/s10208-005-0183-0).
- [Ste08] R. Stevenson. The completion of locally refined simplicial partitions created by bisection. *Math. Comp.*, 77(261):227–241, 2008. DOI: [10.1090/s0025-5718-07-01959-x](https://doi.org/10.1090/s0025-5718-07-01959-x).
- [TW05] A. Toselli and O. Widlund. *Domain decomposition methods—algorithms and theory*. Springer, Berlin, 2005. DOI: [10.1007/b137868](https://doi.org/10.1007/b137868).
- [WZ17] J. Wu and H. Zheng. Uniform convergence of multigrid methods for adaptive meshes. *Appl. Numer. Math.*, 113:109–123, 2017. DOI: [10.1016/j.apnum.2016.11.005](https://doi.org/10.1016/j.apnum.2016.11.005).
- [XQ94] J. Xu and J. Qin. Some remarks on a multigrid preconditioner. *SIAM J. Sci. Comput.*, 15(1):172–184, 1994. DOI: [10.1137/0915012](https://doi.org/10.1137/0915012).

APPENDIX A. PROOF OF THEOREM 21

Proof of Theorem 21. Thanks to Corollary 20, it suffices to show

$$\|u\|_{\mathbb{A}_s} \lesssim \sup_{(\ell, k) \in \mathcal{Q}} (\#\mathcal{T}_\ell)^s \mathbf{H}_\ell^k \lesssim \max\{\|u\|_{\mathbb{A}_s}, \mathbf{H}_0^0\}. \quad (53)$$

Step 1. We first show the lower bound in (53) for $\underline{\ell} = \infty$. In this case, Algorithm B leads to $\#\mathcal{T}_\ell \rightarrow \infty$ as $\ell \rightarrow \infty$. For any $N \in \mathbb{N}$, we follow the proof of [CFPP14, Proposition 4.15] and choose the maximal index $\ell' \in \mathbb{N}_0$ with $\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 \leq N$. Since NVB refinement guarantees $\#\mathcal{T}_{\ell'+1} \lesssim \#\mathcal{T}_{\ell'}$, we are led to

$$N + 1 < \#\mathcal{T}_{\ell'+1} - \#\mathcal{T}_0 + 1 \leq \#\mathcal{T}_{\ell'+1} \lesssim \#\mathcal{T}_{\ell'}.$$

Then, using $\mathcal{T}_{\ell'} \in \mathbb{T}$ and (27) we show for all $(\ell', k') \in \mathcal{Q}$ that

$$\min_{\mathcal{T}_{\text{opt}} \in \mathbb{T}_N} \eta_{\text{opt}}(u_{\text{opt}}^*) \leq \eta_{\ell'}(u_{\ell'}^*) \stackrel{(27)}{\lesssim} \mathbf{H}_{\ell'}^{k'}.$$

A combination of the two previous estimates verifies

$$(N + 1)^s \min_{\mathcal{T}_{\text{opt}} \in \mathbb{T}_N} \eta_{\text{opt}}(u_{\text{opt}}^*) \lesssim (\#\mathcal{T}_{\ell'})^s \mathbf{H}_{\ell'}^{k'} \lesssim \sup_{(\ell, k) \in \mathcal{Q}} (\#\mathcal{T}_\ell)^s \mathbf{H}_\ell^k.$$

Taking the supremum over all $N \in \mathbb{N}$, we prove the lower bound in (53) for $\underline{\ell} = \infty$.

Step 2. We prove the lower bound in (53) for $\underline{\ell} < \infty$. Corollary 19 yields that $\eta_{\underline{\ell}}(u_{\underline{\ell}}^*) = 0$ and $u_{\underline{\ell}}^* = u^*$. Since $\|u^*\|_{\mathbb{A}_s} = 0$ for $\underline{\ell} = 0$, we may assume $\underline{\ell} > 0$. Following Step 1, we let $0 \leq N < \#\mathcal{T}_{\underline{\ell}} - \#\mathcal{T}_0$ and choose the maximal index $0 \leq \ell' < \underline{\ell}$ with $\#\mathcal{T}_{\ell'} - \#\mathcal{T}_0 \leq N$. With the estimates from Step 1, we arrive at

$$\|u^*\|_{\mathbb{A}_s} = \sup_{0 \leq N < \#\mathcal{T}_{\underline{\ell}} - \#\mathcal{T}_0} \left((N + 1)^s \min_{\mathcal{T}_{\text{opt}} \in \mathbb{T}_N} \eta_{\text{opt}}(u_{\text{opt}}^*) \right) \lesssim \sup_{(\ell, k) \in \mathcal{Q}} (\#\mathcal{T}_\ell)^s \mathbf{H}_\ell^k.$$

This proves the lower bound in (53) also for $\underline{\ell} < \infty$.

Step 3. We proceed to prove the upper bound in (53). To this end, we may suppose $\|u^*\|_{\mathbb{A}_s} < \infty$ since the result is trivial otherwise. Let $(\ell' + 1, 0) \in \mathcal{Q}$. Then, [CFPP14, Lemma 4.14] and $0 < \theta_{\text{mark}} < \theta^*$ guarantee the existence of a set $\mathcal{R}_{\ell'} \subseteq \mathcal{T}_{\ell'}$ such that

$$\#\mathcal{R}_{\ell'} \lesssim \|u^*\|_{\mathbb{A}_s}^{1/s} \eta_{\ell'}(u_{\ell'}^*)^{-1/s} \quad \text{and} \quad \theta_{\text{mark}} \eta_{\ell'}(u_{\ell'}^*)^2 \leq \eta_{\ell'}(\mathcal{R}_{\ell'}; u_{\ell'}^*)^2. \quad (54)$$

Note that (33) verifies that $\mathcal{R}_{\ell'}$ also satisfies the Dörfler marking criterion (24) for $u_{\ell'}^k$ with parameter θ , i.e., $\theta \eta_{\ell'}(u_{\ell'}^k)^2 \leq \eta_{\ell'}(\mathcal{R}_{\ell'}; u_{\ell'}^k)^2$. Therefore, it follows that

$$\#\mathcal{M}_{\ell'} \leq C_{\text{mark}} \#\mathcal{R}_{\ell'}. \quad (55)$$

Full linear convergence (44) and estimating as in (38), leads to

$$\mathbf{H}_{\ell'+1}^0 \stackrel{(44)}{\lesssim} \mathbf{H}_{\ell'}^k \stackrel{(38)}{\leq} \eta_{\ell'}(u_{\ell'}^k). \quad (56)$$

Combining all estimates and using the estimator equivalence (30), we obtain

$$\#\mathcal{M}_{\ell'} \stackrel{(55)}{\lesssim} \#\mathcal{R}_{\ell'} \stackrel{(54)}{\lesssim} \|u^*\|_{\mathbb{A}_s}^{1/s} \eta_{\ell'}(u_{\ell'}^*)^{-1/s} \stackrel{(30)}{\lesssim} \|u^*\|_{\mathbb{A}_s}^{1/s} \eta_{\ell'}(u_{\ell'}^k)^{-1/s} \stackrel{(56)}{\lesssim} \|u^*\|_{\mathbb{A}_s}^{1/s} (\mathbf{H}_{\ell'+1}^0)^{-1/s}. \quad (57)$$

Step 4. Consider $(\ell, k) \in \mathcal{Q}$, then full linear convergence (44) yields that

$$\sum_{\substack{(\ell', k') \in \mathcal{Q} \\ |\ell', k'| \leq |\ell, k|}} (\mathbf{H}_{\ell'}^{k'})^{-1/s} \stackrel{(44)}{\lesssim} (\mathbf{H}_\ell^k)^{-1/s} \sum_{\substack{(\ell', k') \in \mathcal{Q} \\ |\ell', k'| \leq |\ell, k|}} (q_{\text{lin}}^{1/s})^{|\ell, k| - |\ell', k'|} \lesssim (\mathbf{H}_\ell^k)^{-1/s}. \quad (58)$$

Recall that NVB refinement satisfies the mesh-closure estimate, i.e., there holds that

$$\#\mathcal{T}_\ell - \#\mathcal{T}_0 \leq C_{\text{mesh}} \sum_{\ell'=0}^{\ell-1} \#\mathcal{M}_{\ell'} \quad \text{for all } \ell \geq 0, \quad (59)$$

where $C_{\text{mesh}} > 1$ depends only on \mathcal{T}_0 . Thus, for $(\ell, k) \in \mathcal{Q}$ with $\ell > 0$, we have by the above formulas (57)–(59) that

$$\begin{aligned} \#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1 &\leq 2(\#\mathcal{T}_\ell - \#\mathcal{T}_0) \stackrel{(59)}{\leq} 2C_{\text{mesh}} \sum_{\ell'=0}^{\ell-1} \#\mathcal{M}_{\ell'} \stackrel{(57)}{\lesssim} \|u^*\|_{\mathbb{A}_s}^{1/s} \sum_{\ell'=0}^{\ell-1} (\mathbf{H}_{\ell'+1}^0)^{-1/s} \\ &\lesssim \|u^*\|_{\mathbb{A}_s}^{1/s} \sum_{\substack{(\ell', k') \in \mathcal{Q} \\ |\ell', k'| \leq |\ell, k|}} (\mathbf{H}_{\ell'}^{k'})^{-1/s} \stackrel{(58)}{\lesssim} \|u^*\|_{\mathbb{A}_s}^{1/s} (\mathbf{H}_\ell^k)^{-1/s}. \end{aligned}$$

Rearranging the terms, we obtain that

$$(\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \mathbf{H}_\ell^k \lesssim \|u^*\|_{\mathbb{A}_s} \quad \text{for all } \ell > 0. \quad (60a)$$

Trivially, full linear convergence proves that

$$(\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \mathbf{H}_0^k = \mathbf{H}_0^k \lesssim \mathbf{H}_0^0 \quad \text{for } \ell = 0. \quad (60b)$$

Moreover, recall from [BHP17, Lemma 22] that, for all $\mathcal{T}_H \in \mathbb{T}$ and all $\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_H)$,

$$\#\mathcal{T}_h - \#\mathcal{T}_H + 1 \leq \#\mathcal{T}_h \leq \#\mathcal{T}_H (\#\mathcal{T}_h - \#\mathcal{T}_H + 1). \quad (61)$$

Overall, we thus have shown that

$$(\#\mathcal{T}_\ell)^s \mathbf{H}_\ell^k \stackrel{(61)}{\lesssim} (\#\mathcal{T}_\ell - \#\mathcal{T}_0 + 1)^s \mathbf{H}_\ell^k \stackrel{(60)}{\lesssim} \max\{\|u^*\|_{\mathbb{A}_s}, \mathbf{H}_0^0\} \quad \text{for all } (\ell, k) \in \mathcal{Q}.$$

This concludes the proof of the upper bound in (53) and hence that of (52). \square

FACULTAD DE MATEMÁTICAS, PONTIFICIA UNIVERSIDAD CATÓLICA DE CHILE, SANTIAGO, CHILE
Email address: thfuhrer@uc.cl

SORBONNE UNIVERSITÉ, UNIVERSITÉ PARIS CITÉ, CNRS, LABORATOIRE JACQUES-LOUIS LIONS,
LJLL, F-75005 PARIS, FRANCE

Email address: ani.miraci@sorbonne-universite.fr

TU WIEN, INSTITUTE OF ANALYSIS AND SCIENTIFIC COMPUTING, WIEDNER HAUPTSTR. 8–
10/E101/4, 1040 VIENNA, AUSTRIA

Email address: paula.hilbert@asc.tuwien.ac.at (corresponding author)

Email address: dirk.praetorius@asc.tuwien.ac.at