



**HAL**  
open science

# Generalized preconditioned conjugate gradients for adaptive FEM with optimal complexity

Paula Hilbert, Ani Miraçi, Dirk Praetorius

## ► To cite this version:

Paula Hilbert, Ani Miraçi, Dirk Praetorius. Generalized preconditioned conjugate gradients for adaptive FEM with optimal complexity. 2026. <hal-05475870v2>

**HAL Id: hal-05475870**

**<https://hal.science/hal-05475870v2>**

Preprint submitted on 9 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

# GENERALIZED PRECONDITIONED CONJUGATE GRADIENTS FOR ADAPTIVE FEM WITH OPTIMAL COMPLEXITY

PAULA HILBERT<sup>1</sup>, ANI MIRAÇI<sup>1</sup>, AND DIRK PRAETORIUS<sup>1</sup>

**ABSTRACT.** We consider adaptive finite element methods (AFEMs) with inexact algebraic solvers for second-order symmetric linear elliptic diffusion problems. Optimal complexity of AFEM, i.e., optimal convergence rates with respect to the overall computational cost, hinges on two requirements on the solver. First, each solver step is of linear cost with respect to the number of degrees of freedom. Second, each solver step guarantees uniform contraction of the solver error with respect to the PDE-related energy norm. Both properties must be ensured robustly with respect to the local mesh size  $h$  (i.e.,  $h$ -robustness). While existing literature on geometric multigrid methods (MG) or symmetric additive Schwarz preconditioners for the preconditioned conjugate gradient method (PCG) that are appropriately adapted to adaptive mesh-refinement satisfy these requirements, this paper aims to consider more general solvers. Our main focus is on preconditioners stemming from contractive solvers which need not be symmetrized to be used with Krylov methods and which are not only  $h$ -robust but also  $p$ -robust, i.e., the contraction constant is independent of the polynomial degree  $p$ . In particular, we show that generalized PCG (GPCG) with an  $h$ - and  $p$ -robust contractive MG as a preconditioner satisfies the requirements for optimal-complexity AFEM and that it numerically outperforms AFEM using MG as a solver. While this is certainly known for (quasi-)uniform meshes, the main contribution of the present work is the rigorous analysis of the interplay of the solver with adaptive mesh-refinement. Numerical experiments underline the theoretical findings.

## 1. INTRODUCTION

Given a bounded Lipschitz domain  $\Omega \subset \mathbb{R}^d$  for  $d \in \mathbb{N}$ , a right-hand side  $f \in L^2(\Omega)$ , and a symmetric and uniformly positive definite diffusion coefficient  $\mathbf{K} \in [L^\infty(\Omega)]_{\text{sym}}^{d \times d}$ , we consider the second-order symmetric linear elliptic diffusion problem

$$\begin{aligned} -\operatorname{div}(\mathbf{K}\nabla u^*) &= f && \text{in } \Omega, \\ u^* &= 0 && \text{on } \partial\Omega. \end{aligned} \tag{1}$$

Adaptive finite element methods (AFEMs) allow to achieve optimal convergence rates for such problems even if the solution exhibits singularities. These methods usually require the solution of a sequence of linear systems arising from the discretization on successively refined meshes. Solving these systems by direct methods is computationally too expensive to guarantee optimal error decay with respect to the overall computing time, making the use of iterative solvers into the adaptive algorithm essential. Indeed, it has been shown that uniformly contractive iterative solvers are integral to achieve optimal complexity of AFEMs, i.e., optimal convergence rates with respect to the *overall computational cost* and hence time; see, e.g., [Ste07; GHPS21; BFM<sup>+</sup>25].

For the model problem (1), discretized by  $H^1$ -conforming finite elements, the resulting Galerkin matrix  $\mathbf{A}_\ell$  is symmetric and positive definite (SPD), which naturally suggests the use of the conjugate gradient method (CG) as an algebraic solver. However, the contraction factor of CG

---

2020 *Mathematics Subject Classification.* 65N12, 65N30, 65F10, 65N55, 65Y20.

*Key words and phrases.* adaptive finite element method, inexact solver, geometric multigrid, optimal preconditioner, additive Schwarz, generalized precondition conjugate gradient method.

**Acknowledgment.** This research was funded in whole or in part by the Austrian Science Fund (FWF) projects [10.55776/PAT3699424](https://doi.org/10.55776/PAT3699424) (standalone project PAT3699424 “Optimal robust solvers for reliable and efficient AFEMs”), and [10.55776/PAT3446525](https://doi.org/10.55776/PAT3446525) (standalone project PAT3446525 “Adaptive Uzawa-type FEM for nonlinear PDEs”). Additionally, Paula Hilbert is supported by the Vienna School of Mathematics.

degrades with respect to the discretization parameters, i.e., the local mesh size  $h$  and the polynomial degree  $p$  (see, e.g., [CG22]). Therefore, appropriate  $h$ - and  $p$ -robust preconditioning is required to obtain optimal convergence rates. Designing such preconditioners, which we will denote by  $\mathbf{B}_\ell$ , is nontrivial, as they must, first, approximate the inverse of the Galerkin matrix sufficiently well, i.e.,  $\mathbf{B}_\ell \approx \mathbf{A}_\ell^{-1}$ , and, second, be cheap to apply, ideally with linear computational complexity. Robust iterative solvers themselves, such as, e.g., the geometric multigrid method introduced in [IMPS24], are natural candidates for this purpose, as they are designed to efficiently solve the Galerkin system.

The use of additive or multiplicative multilevel methods as preconditioners for CG in finite element discretizations has a long history. Early results include [BP93] for multilevel preconditioners in the setting of quasi-uniform and locally refined grids under a *nested refinement* assumption. For multiplicative methods, this assumption is not merely technical and often necessitates additional preprocessing, such as the reconstruction of a virtual refinement hierarchy; see, e.g., [BEK93; BY93; HZ09]. For graded bisection grids, where only a single bisection is performed per refinement step, uniform convergence with respect to  $h$  was shown in [XCN09; CNX12]. Even in this setting, additional coarsening algorithms are typically required to meet the theoretical assumptions; see, e.g., [CZ10]. In [WC06], the idea of restricting levelwise smoothing to newly created vertices and their direct neighbors was introduced, yielding uniform convergence of multigrid methods for adaptive meshes generated by newest vertex bisection (NVB) in two dimensions. This  $h$ -robust approach was later extended to three dimensions in [WZ17]. An  $h$ - and  $p$ -robust geometric multigrid method has been proposed and analyzed in [IMPS24]. Its  $h$ -robustness relies on the strengthened Cauchy–Schwarz inequality from [HWZ12] and the stable decomposition established in [WZ17]. Its  $p$ -robustness follows from the stable decomposition in [SMPZ08], which was previously also used in [MPV20; MPV21] to construct a  $p$ -robust contractive multigrid method (however lacking  $h$ -robustness in case of non-uniform refinement).

Most of the aforementioned works on preconditioners focus on linear and symmetric multilevel methods in order to comply with the standard assumptions of the preconditioned conjugate gradient method (PCG). From a computational perspective, however, it can be advantageous to consider multigrid methods with post-smoothing only, which seem to preserve comparable contraction properties at reduced cost; see, e.g., [DHM<sup>+</sup>21; MPV21; IMPS24]. Moreover, the analysis of multiplicative methods often requires strong assumptions on the levelwise operators, such as contraction [CNX12], or norm bounds strictly smaller than two [BPWX91]. In contrast, [MPV21; IMPS24] employ optimal step-sizes on each level, simplifying the contraction proofs and yielding better numerical behavior, but resulting in non-linear and non-symmetric multigrid methods. In numerical linear algebra, however, it is well known that PCG may fail if the preconditioner is not symmetric and positive definite. That said, PCG has been adapted in the literature to also accommodate more general preconditioners; see, e.g., [AV91; GY99; Not00; Bla02]. In particular, the generalized preconditioned conjugate gradient method (GPCG) introduced in [Bla02] also allows for non-symmetric and non-linear preconditioners.

This work establishes a direct connection between uniformly contractive iterative solvers arising in AFEM and their use as preconditioners within conjugate gradient methods. In particular, we show that *any* uniformly contractive solver with linear computational complexity induces an *optimal preconditioner* for GPCG from [Bla02], i.e., the resulting algebraic solver has a contraction factor independent of  $h$  and  $p$  and is of linear complexity. While this is certainly well-known in the numerical linear algebra community for uniform triangulations, we are not aware of a general presentation with a focus on adaptively generated triangulations. Indeed, a central feature of the analysis in this work is that no additional preprocessing of the NVB-generated mesh hierarchy is required, allowing for seamless integration of the considered solvers into the adaptive finite-element algorithm. Moreover, the presented framework unifies techniques from numerical linear algebra and AFEM analysis, thereby allowing contractive solvers to be used as preconditioners for GPCG irrespective of symmetry or linearity. To the best of our knowledge, this connection has not been addressed in the literature on optimal-complexity AFEM yet.

Based on the  $h$ - and  $p$ -robust stable decomposition from [IMPS24], we derive various optimal preconditioners for GPCG and PCG. We first show that the geometric multigrid method from [IMPS24] directly induces a preconditioner that indeed fits in the general framework we develop in this work. In addition, we show that this multigrid method can be linearized and symmetrized, resulting in an SPD preconditioner suitable for standard PCG without loss of optimality. As additive and multiplicative Schwarz methods both rely on stable decompositions; see, e.g., [BP93; XCN09; CNX12], we use the same decomposition to also construct an optimal additive Schwarz preconditioner; see also [FFPS17] for a similar construction in the context of the  $hp$ -adaptive boundary element method. Our numerical experiments show that the non-linear and non-symmetric multigrid preconditioner applied in GPCG outperforms the standalone multigrid solver as well as the additive Schwarz preconditioner used within PCG, both in terms of the experimental contraction factor and its application within the AFEM loop. The linear and symmetric multigrid preconditioner used in PCG generally exhibits better contraction factors as the iteration (and thus the Krylov space dimension) increase. However, its worse contraction in the initial iterations impacts the overall adaptive algorithm relying typically on few solver steps. Moreover, the linear and symmetric version is computationally more expensive since it entails that pre-smoothing steps have to be added.

Finally, we note that all results in this work rely on the refined analysis from [Hil25], which shows that the constants in the strengthened Cauchy–Schwarz inequality as well as in the  $h$ - and  $p$ -stable subspace decompositions depend only on the local behaviour of the diffusion coefficient. In contrast, the constants in [IMPS24] and other works depend on the ratio of the maximal and minimal eigenvalue of the diffusion matrix, i.e., global diffusion contrast.

**Outline.** In Section 2, we introduce the problem setting. Section 3 is devoted to GPCG (Algorithm B) and its analysis (Proposition 2). In Section 4, we derive an optimal preconditioner for GPCG (Theorem 8) based on the geometric multigrid method (Algorithm C) from [IMPS24]. Section 5 presents a linearized and symmetrized variant (Algorithm D) and shows that this approach likewise yields an optimal preconditioner (Theorem 16), where GPCG reduces to PCG. Section 6 introduces an additive Schwarz preconditioner based on the same space decomposition as for the proposed multigrid preconditioners and establishes its optimality (Theorem 19). We conclude the theoretical analysis in Section 7 with a brief discussion of the application of the proposed solvers within the AFEM framework (Algorithm E) to solve the symmetric model problem (1). However, we already note here that the developments of the present work play a central role in the analysis of optimal-complexity AFEM for general second-order linear elliptic PDEs in the framework of the Lax–Milgram lemma, where the solution of the linear finite element system relies on optimal preconditioners for the principal part of the PDE; see [FHMP26]. Finally, Section 8 presents numerical experiments comparing the developed algebraic solvers.

## 2. SETTING

**2.1. Mesh and space hierarchy.** Let  $\mathcal{T}_0$  be an initial conforming mesh of  $\Omega$  into compact simplices  $T \in \mathcal{T}_0$ , which is admissible in the sense of [Ste08]. From now on, we consider a sequence  $(\mathcal{T}_\ell)_{\ell \in \mathbb{N}_0}$  of successively refined triangulations obtained by newest vertex bisection; see, e.g., [Tra97; Ste08] for  $d \geq 2$  and [AFF<sup>+</sup>13] for  $d = 1$ . Thus, for all  $\ell \geq 1$ , it holds that  $\mathcal{T}_\ell = \text{refine}(\mathcal{T}_{\ell-1}, \mathcal{M}_{\ell-1})$ , where  $\mathcal{T}_\ell$  is the coarsest conforming triangulation obtained by NVB ensuring that all marked elements  $\mathcal{M}_{\ell-1} \subseteq \mathcal{T}_{\ell-1}$  have been bisected. The generated sequence is uniformly  $\gamma$ -shape regular, i.e.,

$$\sup_{\ell \in \mathbb{N}_0} \max_{T \in \mathcal{T}_\ell} \frac{\text{diam}(T)}{|T|^{1/d}} \leq \gamma < \infty \quad \text{and} \quad \sup_{\ell \in \mathbb{N}_0} \max_{T \in \mathcal{T}_\ell} \max_{\substack{T' \in \mathcal{T}_\ell \\ T \cap T' \neq \emptyset}} \frac{\text{diam}(T)}{\text{diam}(T')} \leq \gamma < \infty, \quad (2)$$

where  $\gamma$  depends only on  $\mathcal{T}_0$ ; see, e.g., [Ste08, Theorem 2.1] for  $d \geq 2$  and [AFF<sup>+</sup>13] for  $d = 1$ . Furthermore, for every mesh  $\mathcal{T}_\ell$ , we denote by  $\mathcal{V}_\ell$  the set of vertices. For  $z \in \mathcal{V}_\ell$ , we define the

$n$ -patch  $\mathcal{T}_\ell^n(z)$  inductively via

$$\mathcal{T}_\ell(z) := \mathcal{T}_\ell^1(z) := \{T \in \mathcal{T}_\ell : z \in T\}, \quad \mathcal{T}_\ell^{n+1}(z) := \{T \in \mathcal{T}_\ell : T \cap \overline{\omega_\ell^n(z)} \neq \emptyset\},$$

where  $\omega_\ell^n(z) := \text{int}(\bigcup_{T \in \mathcal{T}_\ell^n(z)} T)$  denotes the corresponding  $n$ -patch subdomain. With the element size  $h_T := |T|^{1/d}$  for all  $T \in \mathcal{T}_\ell$ , the size of a patch subdomain is given by  $h_{\ell,z} := \max_{T \in \mathcal{T}_\ell(z)} h_T$ . With  $\omega_\ell(z) := \omega_\ell^1(z)$ , we finally define

$$\mathcal{V}_0^+ := \mathcal{V}_0 \quad \text{and} \quad \mathcal{V}_\ell^+ := \mathcal{V}_\ell \setminus \mathcal{V}_{\ell-1} \cup \{\mathcal{V}_\ell \cap \mathcal{V}_{\ell-1} : \omega_\ell(z) \neq \omega_{\ell-1}(z)\} \quad \text{for } \ell \in \mathbb{N},$$

where  $\mathcal{V}_\ell^+$  consists of the new vertices of  $\mathcal{T}_\ell$  along with their neighboring vertices in  $\mathcal{V}_\ell \cap \mathcal{V}_{\ell-1}$ .

Let  $q \geq 1$  and  $T \in \mathcal{T}_\ell$ . Then,  $\mathbb{P}^q(T)$  denotes the space of all polynomials on  $T$  of degree at most  $q$ . For  $\ell \in \mathbb{N}_0$ , we define the finite-dimensional subspaces

$$\mathcal{X}_\ell^q := \mathbb{S}_0^q(\mathcal{T}_\ell) := \{v_\ell \in H_0^1(\Omega) : v_\ell|_T \in \mathbb{P}^q(T) \text{ for all } T \in \mathcal{T}_\ell\} \subset \mathcal{X} := H_0^1(\Omega)$$

and observe that

$$\mathcal{X}_0^1 \subseteq \mathcal{X}_1^1 \subseteq \dots \subseteq \mathcal{X}_{\ell-1}^1 \subseteq \mathcal{X}_\ell^1 \subseteq \mathcal{X}_\ell^p,$$

where  $p \geq 1$  is a fixed polynomial degree. Furthermore, the local spaces  $\mathcal{X}_{\ell,z}^q$  are given by

$$\mathcal{X}_{\ell,z}^q := \mathbb{S}_0^q(\mathcal{T}_\ell(z)) := \{v_\ell \in \mathcal{X}_\ell^q : v_\ell|_T = 0 \text{ for all } T \in \mathcal{T}_\ell \setminus \mathcal{T}_\ell(z)\}.$$

Denote by  $\varphi_{\ell,z}^1$  the usual hat-function associated with the vertex  $z \in \mathcal{V}_\ell$ , i.e.,  $\varphi_{\ell,z}^1|_T \in \mathbb{P}^1(T)$  for all  $T \in \mathcal{T}_\ell$  and  $\varphi_{\ell,z}^1(z') = \delta_{zz'}$  for all  $z' \in \mathcal{V}_\ell$ . The set  $\{\varphi_{\ell,z}^1 : z \in \mathcal{V}_\ell \setminus \partial\Omega\}$  is a basis of  $\mathcal{X}_\ell^1$ . Therefore, we define the spaces  $\mathcal{X}_\ell^+$  induced by the set of vertices  $\mathcal{V}_\ell^+$  as

$$\mathcal{X}_\ell^+ := \text{span}\{\varphi_{\ell,z}^1 : z \in \mathcal{V}_\ell^+ \setminus \partial\Omega\} \subseteq \mathcal{X}_\ell^1.$$

**2.2. Weak formulation.** For the analysis, we need some more assumptions on the model problem (1). Firstly, we will only consider  $d \in \{1, 2, 3\}$ . For the diffusion coefficient  $\mathbf{K}$ , we require the stronger regularity  $\mathbf{K}|_T \in [W^{1,\infty}(T)]^{d \times d}$  for all  $T \in \mathcal{T}_0$ , where  $\mathcal{T}_0$  is the initial triangulation. More precisely, this is needed to show the strengthened Cauchy–Schwarz inequality of Lemma 24 and to define the residual error estimator (72). For  $x \in \Omega$ , the expressions  $\lambda_{\max}(\mathbf{K}(x))$  and  $\lambda_{\min}(\mathbf{K}(x))$  denote the maximal and minimal eigenvalue of  $\mathbf{K}(x) \in \mathbb{R}^{d \times d}$ , respectively. For any measurable set  $\omega \subseteq \Omega$ , we denote the  $L^2(\omega)$ -scalar product with  $\langle \cdot, \cdot \rangle_\omega$ . The weak formulation of (1) reads: Find  $u^* \in \mathcal{X}$  that solves

$$\langle\langle u^*, v \rangle\rangle_\Omega := \langle \mathbf{K} \nabla u^*, \nabla v \rangle_\Omega = \langle f, v \rangle_\Omega =: F(v) \quad \text{for all } v \in \mathcal{X}. \quad (3)$$

In particular, the Riesz theorem yields existence and uniqueness of the weak solution  $u^* \in \mathcal{X}$  to (3). From here on, we omit the index  $\omega$  whenever  $\omega = \Omega$ . Furthermore, we define the induced norm  $\|v\|^2 := \langle\langle v, v \rangle\rangle$  and observe that

$$\inf_{y \in \omega} \lambda_{\min}(\mathbf{K}(y)) \|\nabla v\|_\omega^2 \leq \|v\|_\omega^2 \leq \sup_{y \in \omega} \lambda_{\max}(\mathbf{K}(y)) \|\nabla v\|_\omega^2 \quad \text{for all } v \in \mathcal{X} \text{ and all } \omega \subseteq \Omega.$$

For a given triangulation  $\mathcal{T}_\ell$  and polynomial degree  $p \geq 1$ , the Galerkin discretization of (1) reads: Find  $u_\ell^* \in \mathcal{X}_\ell^p$  such that

$$\langle\langle u_\ell^*, v_\ell \rangle\rangle = F(v_\ell) \quad \text{for all } v_\ell \in \mathcal{X}_\ell^p. \quad (4)$$

Let  $N_\ell^p := \dim(\mathcal{X}_\ell^p)$  and  $\{\varphi_{\ell,j}^p\}_{j=1}^{N_\ell^p}$  be a basis of  $\mathcal{X}_\ell^p$ . Then, the discrete problem (4) can equivalently be rewritten as  $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$ , with the symmetric and positive definite Galerkin matrix

$$(\mathbf{A}_\ell)_{jk} := \langle\langle \varphi_{\ell,j}^p, \varphi_{\ell,k}^p \rangle\rangle \quad \text{for } j, k = 1, \dots, N_\ell^p \quad (5)$$

and the right-hand side vector

$$(\mathbf{b}_\ell)_j := F(\varphi_{\ell,j}^p) \quad \text{for } j = 1, \dots, N_\ell^p. \quad (6)$$

The solution vector  $\mathbf{x}_\ell^* \in \mathbb{R}^{N_\ell^p}$  is the coefficient vector of the discrete solution  $u_\ell^* = \sum_{j=1}^{N_\ell^p} (\mathbf{x}_\ell^*)_j \varphi_{\ell,j}^p$  with respect to the fixed basis. To describe this connection, we note that  $\mathbf{x}_\ell^* = \chi_\ell^p(u_\ell^*)$ , where

$$\chi_\ell^p: \mathcal{X}_\ell^p \rightarrow \mathbb{R}^{N_\ell^p}, \quad \text{is defined via} \quad v_\ell = \sum_{j=1}^{N_\ell^p} \chi_\ell^p(v_\ell)_j \varphi_{\ell,j}^p. \quad (7)$$

### 3. GENERALIZED PRECONDITIONED CONJUGATE GRADIENT METHOD

In this section, we recall the generalized preconditioned conjugate gradient method from [Bla02]. The contraction of the method relies on a discrete assumption of the preconditioner matrix; see [Bla02, Theorem 3.4]. We use this to show that any uniformly contractive algebraic solver with linear cost induces an optimal preconditioner for GPCG. Let  $(\mathbf{x}, \mathbf{y})_2 := \mathbf{x}^T \mathbf{y}$  denote the Euclidean scalar product with corresponding norm  $|\mathbf{x}|_2 := (\mathbf{x}, \mathbf{x})_2^{1/2}$ . For an SPD matrix  $\mathbf{A} \in \mathbb{R}^{N \times N}$ , we denote by  $(\mathbf{x}, \mathbf{y})_{\mathbf{A}} := \mathbf{x}^T \mathbf{A} \mathbf{y}$  the induced scalar product with corresponding norm  $|\mathbf{x}|_{\mathbf{A}} := (\mathbf{x}, \mathbf{x})_{\mathbf{A}}^{1/2}$ . Lastly, we denote by  $\text{cond}_2(\mathbf{M})$  and  $\text{cond}_{\mathbf{A}}(\mathbf{M})$  the condition number of a regular matrix  $\mathbf{M} \in \mathbb{R}^{N \times N}$  with respect to the norms  $|\cdot|_2$  and  $|\cdot|_{\mathbf{A}}$ , respectively.

**3.1. Symmetric preconditioners.** In this section, we briefly recall the classical and preconditioned conjugate gradient method for solving linear systems  $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$  with a symmetric and positive definite matrix  $\mathbf{A}_\ell \in \mathbb{R}^{N_\ell^p \times N_\ell^p}$ . Let us begin with the CG algorithm; see, e.g., [HS52].

---

**Algorithm A (CG).** *Input:* SPD matrix  $\mathbf{A}_\ell \in \mathbb{R}^{N_\ell^p \times N_\ell^p}$ , right-hand side  $\mathbf{b}_\ell \in \mathbb{R}^{N_\ell^p}$ , initial guess  $\mathbf{x}_\ell^0 \in \mathbb{R}^{N_\ell^p}$ , and tolerance  $\tau \geq 0$ .  
Set  $\mathbf{p}_\ell^0 := \mathbf{r}_\ell^0 = \mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell^0$  and repeat for all  $k = 0, 1, 2, \dots$  until  $|\mathbf{r}_\ell^k|_2^2 < \tau$ :

- (i) **Update approximation:**  $\mathbf{x}_\ell^{k+1} := \mathbf{x}_\ell^k + \alpha_k \mathbf{p}_\ell^k$  with  $\alpha_k = |\mathbf{r}_\ell^k|_2^2 / |\mathbf{p}_\ell^k|_{\mathbf{A}_\ell}^2$
- (ii) **Update residual:**  $\mathbf{r}_\ell^{k+1} := \mathbf{r}_\ell^k - \alpha_k \mathbf{A}_\ell \mathbf{p}_\ell^k$
- (iii) **Compute new search direction:**  $\mathbf{p}_\ell^{k+1} := \mathbf{r}_\ell^{k+1} + \beta_k \mathbf{p}_\ell^k$  with  $\beta_k = |\mathbf{r}_\ell^{k+1}|_2^2 / |\mathbf{r}_\ell^k|_2^2$

**Output:** Approximation  $\mathbf{x}_\ell^k$  to the solution  $\mathbf{x}_\ell^*$  of  $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$ .

---

The following result gives a convergence estimate for the CG method. For more details and a proof, we refer to [GV13, Theorem 11.3.3].

---

**Proposition 1 (Contraction of CG).** *Let  $\mathbf{A}_\ell \in \mathbb{R}^{N_\ell^p \times N_\ell^p}$  be a symmetric and positive definite matrix and let  $\mathbf{x}_\ell^* \in \mathbb{R}^{N_\ell^p}$  be the unique solution of the linear system  $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$ . Then, for any initial guess  $\mathbf{x}_\ell^0 \in \mathbb{R}^{N_\ell^p}$ , the sequence  $\mathbf{x}_\ell^k \in \mathbb{R}^{N_\ell^p}$  produced by Algorithm A guarantees contraction*

$$|\mathbf{x}_\ell^* - \mathbf{x}_\ell^{k+1}|_{\mathbf{A}_\ell} \leq \left(1 - \frac{1}{\text{cond}_2(\mathbf{A}_\ell)}\right)^{1/2} |\mathbf{x}_\ell^* - \mathbf{x}_\ell^k|_{\mathbf{A}_\ell}. \quad \square$$


---

For the Galerkin matrix  $\mathbf{A}_\ell$ , the condition number  $\text{cond}_2(\mathbf{A}_\ell)$  depends on and grows with the local mesh size  $h$  and the polynomial degree  $p$ . One way to achieve better contraction is to introduce an SPD preconditioner  $\mathbf{B}_\ell$  such that the condition number of  $\mathbf{B}_\ell^{1/2} \mathbf{A}_\ell \mathbf{B}_\ell^{1/2}$  is bounded independently of  $h$  and  $p$  and to apply CG to the modified system

$$\tilde{\mathbf{A}}_\ell \tilde{\mathbf{x}}_\ell^* := \mathbf{B}_\ell^{1/2} \mathbf{A}_\ell \mathbf{B}_\ell^{1/2} \tilde{\mathbf{x}}_\ell^* = \mathbf{B}_\ell^{1/2} \mathbf{b}_\ell =: \tilde{\mathbf{b}}_\ell. \quad (8)$$

This is known as the preconditioned CG method (PCG). Note that  $\tilde{\mathbf{A}}_\ell$  is SPD, so indeed CG can be applied. For the initial guess  $\mathbf{x}_\ell^0 \in \mathbb{R}^{N_\ell^p}$ , let  $\tilde{\mathbf{x}}_\ell^0 := \mathbf{B}_\ell^{-1/2} \mathbf{x}_\ell^0$  and let  $\tilde{\mathbf{x}}_\ell^k$ ,  $\tilde{\mathbf{r}}_\ell^k$ , and  $\tilde{\mathbf{p}}_\ell^k$  denote the sequences generated by Algorithm A for the preconditioned system (8). Furthermore, let

$\mathbf{x}_\ell^k := \mathbf{B}_\ell^{1/2} \tilde{\mathbf{x}}_\ell^k$ ,  $\mathbf{r}_\ell^k := \mathbf{B}_\ell^{-1/2} \tilde{\mathbf{r}}_\ell^k$ , and  $\mathbf{p}_\ell^k := \mathbf{B}_\ell^{1/2} \tilde{\mathbf{p}}_\ell^k$ . Then, there holds

$$\begin{aligned} \mathbf{p}_\ell^0 &= \mathbf{B}_\ell \mathbf{r}_\ell^0, \\ \mathbf{x}_\ell^{k+1} &= \mathbf{x}_\ell^k + \alpha_k \mathbf{p}_\ell^k, & \text{where } \alpha_k &= |\mathbf{r}_\ell^k|_{\mathbf{B}_\ell}^2 / |\mathbf{p}_\ell^k|_{\mathbf{A}_\ell}^2, \\ \mathbf{r}_\ell^{k+1} &= \mathbf{r}_\ell^k - \alpha_k \mathbf{A}_\ell \mathbf{p}_\ell^k, \\ \mathbf{p}_\ell^{k+1} &= \mathbf{B}_\ell \mathbf{r}_\ell^{k+1} + \beta_k \mathbf{p}_\ell^k, & \text{where } \beta_k &= |\mathbf{r}_\ell^{k+1}|_{\mathbf{B}_\ell}^2 / |\mathbf{r}_\ell^k|_{\mathbf{B}_\ell}^2 \end{aligned}$$

Moreover, for  $\mathbf{y}_\ell \in \mathbb{R}^{N_\ell^p}$  and  $\tilde{\mathbf{y}}_\ell := \mathbf{B}_\ell^{-1/2} \mathbf{y}_\ell$ , we have the identity

$$\begin{aligned} |\mathbf{y}_\ell|_{\mathbf{A}_\ell}^2 &= (\mathbf{A}_\ell \mathbf{y}_\ell, \mathbf{y}_\ell)_2 = (\mathbf{A}_\ell \mathbf{B}_\ell^{1/2} \tilde{\mathbf{y}}_\ell, \mathbf{B}_\ell^{1/2} \tilde{\mathbf{y}}_\ell)_2 \\ &= (\mathbf{B}_\ell^{1/2} \mathbf{A}_\ell \mathbf{B}_\ell^{1/2} \tilde{\mathbf{y}}_\ell, \tilde{\mathbf{y}}_\ell)_2 = |\tilde{\mathbf{y}}_\ell|_{\mathbf{B}_\ell^{1/2} \mathbf{A}_\ell \mathbf{B}_\ell^{1/2}}^2. \end{aligned} \quad (9)$$

With  $q := (1 - 1/\text{cond}_2(\mathbf{B}_\ell^{1/2} \mathbf{A}_\ell \mathbf{B}_\ell^{1/2}))^{1/2}$ , it hence follows from Proposition 1 that

$$|\mathbf{x}_\ell^* - \mathbf{x}_\ell^k|_{\mathbf{A}_\ell} \stackrel{(9)}{=} |\tilde{\mathbf{x}}_\ell^* - \tilde{\mathbf{x}}_\ell^k|_{\mathbf{B}_\ell^{1/2} \mathbf{A}_\ell \mathbf{B}_\ell^{1/2}} \leq q |\tilde{\mathbf{x}}_\ell^* - \tilde{\mathbf{x}}_\ell^0|_{\mathbf{B}_\ell^{1/2} \mathbf{A}_\ell \mathbf{B}_\ell^{1/2}} \stackrel{(9)}{=} q |\mathbf{x}_\ell^* - \mathbf{x}_\ell^0|_{\mathbf{A}_\ell}.$$

Finally, [TW05, Theorem C.1] shows that

$$\text{cond}_2(\mathbf{B}_\ell^{1/2} \mathbf{A}_\ell \mathbf{B}_\ell^{1/2}) = \text{cond}_{\mathbf{A}_\ell}(\mathbf{B}_\ell \mathbf{A}_\ell). \quad (10)$$

Therefore, it suffices to show  $h$ - and  $p$ -independent boundedness of  $\text{cond}_{\mathbf{A}_\ell}(\mathbf{B}_\ell \mathbf{A}_\ell)$ .

**3.2. Non-linear and non-symmetric preconditioners.** Let  $\mathbf{B}_\ell : \mathbb{R}^{N_\ell^p} \rightarrow \mathbb{R}^{N_\ell^p}$  denote a generally non-linear and non-symmetric preconditioner. Then, the generalized preconditioned conjugate gradient method introduced in [Bla02] is given by the following algorithm.

---

**Algorithm B (GPCG).** *Input:* SPD matrix  $\mathbf{A}_\ell \in \mathbb{R}^{N_\ell^p \times N_\ell^p}$ , preconditioner  $\mathbf{B}_\ell : \mathbb{R}^{N_\ell^p} \rightarrow \mathbb{R}^{N_\ell^p}$ , right-hand side  $\mathbf{b}_\ell \in \mathbb{R}^{N_\ell^p}$ , initial guess  $\mathbf{x}_\ell^0 \in \mathbb{R}^{N_\ell^p}$ , and tolerance  $\tau > 0$ .

Set  $\mathbf{r}_\ell^0 := \mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell^0$  and  $\mathbf{p}_\ell^0 := \mathbf{B}_\ell[\mathbf{r}_\ell^0]$  and repeat for all  $k = 0, 1, 2, \dots$  until  $|\mathbf{r}_\ell^k|_2^2 < \tau$ :

- (i) **Update approximation:**  $\mathbf{x}_\ell^{k+1} := \mathbf{x}_\ell^k + \alpha_k \mathbf{p}_\ell^k$  with  $\alpha_k := (\mathbf{B}_\ell[\mathbf{r}_\ell^k], \mathbf{r}_\ell^k)_2 / |\mathbf{p}_\ell^k|_{\mathbf{A}_\ell}^2$
- (ii) **Update residual:**  $\mathbf{r}_\ell^{k+1} := \mathbf{r}_\ell^k - \alpha_k \mathbf{A}_\ell \mathbf{p}_\ell^k$
- (iii) **Compute new search direction:**  $\mathbf{p}_\ell^{k+1} := \mathbf{B}_\ell[\mathbf{r}_\ell^{k+1}] + \beta_k \mathbf{p}_\ell^k$  with

$$\beta_k := \frac{(\mathbf{B}_\ell[\mathbf{r}_\ell^{k+1}], \mathbf{r}_\ell^{k+1})_2 - (\mathbf{B}_\ell[\mathbf{r}_\ell^{k+1}], \mathbf{r}_\ell^k)_2}{(\mathbf{B}_\ell[\mathbf{r}_\ell^k], \mathbf{r}_\ell^k)_2}$$

**Output:** Approximation  $\mathbf{x}_\ell^k$  to the solution  $\mathbf{x}_\ell^*$  of  $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$ .

---

In [Bla02], different assumptions on the preconditioner  $\mathbf{B}_\ell$  are investigated. We will only consider the case where  $\mathbf{B}_\ell$  is a good approximation of the SPD matrix  $\mathbf{A}_\ell^{-1} \in \mathbb{R}^{N_\ell^p \times N_\ell^p}$ , in the sense that there exists an  $h$ - and  $p$ -robust factor  $q \in (0, 1)$  such that

$$|\mathbf{B}_\ell[\mathbf{x}_\ell] - \mathbf{A}_\ell^{-1} \mathbf{x}_\ell|_{\mathbf{A}_\ell} \leq q |\mathbf{A}_\ell^{-1} \mathbf{x}_\ell|_{\mathbf{A}_\ell} \quad \text{for all } \mathbf{x}_\ell \in \mathbb{R}^{N_\ell^p}. \quad (11)$$

In [Bla02, Theorem 3.4], contraction of GPCG under the assumption (11) is shown. While (11) is a discrete assumption, we use it to show the following central result.

---

**Proposition 2 (Contractive solvers are general preconditioners).** *Let*  $\mathcal{S}_\ell : \mathcal{X}_\ell^p \rightarrow \mathcal{X}_\ell^p$  *denote the error propagation operator of a given algebraic solver, i.e.,*  $u_\ell^{k+1} - u_\ell^k = \mathcal{S}_\ell(u_\ell^* - u_\ell^k)$ . *Assume the solver is contractive, i.e., there exists a constant*  $q \in (0, 1)$  *such that*

$$\| (I - \mathcal{S}_\ell) u_\ell \| \leq q \| u_\ell \| \quad \text{for all } u_\ell \in \mathcal{X}_\ell^p. \quad (12)$$

Let  $\mathbf{B}_\ell : \mathbb{R}^{N_\ell^p} \rightarrow \mathbb{R}^{N_\ell^p}$  denote the representation of  $\mathcal{S}_\ell$  acting on coefficient vectors, meaning that  $\chi_\ell^p(\mathcal{S}_\ell u_\ell) = \mathbf{B}_\ell[\mathbf{A}_\ell \chi_\ell^p(u_\ell)]$  for all  $u_\ell \in \mathcal{X}_\ell^p$ , where  $\mathbf{A}_\ell$  is the Galerkin matrix satisfying

$\|\cdot\| = |\chi_\ell^p(\cdot)|_{\mathbf{A}_\ell}$ . Then, the GPCG method with preconditioner  $\mathbf{B}_\ell$  applied to the Galerkin system  $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$  yields an iterative solver with contraction factor  $q$ , i.e.,

$$|\mathbf{x}_\ell^* - \tilde{\mathbf{x}}_\ell^{k+1}|_{\mathbf{A}_\ell} \leq q |\mathbf{x}_\ell^* - \tilde{\mathbf{x}}_\ell^k|_{\mathbf{A}_\ell} \quad \text{for all } k \in \mathbb{N}_0. \quad (13)$$

For  $u_\ell^* = \sum_{j=1}^{N_\ell^p} (\mathbf{x}_\ell^*)_j \varphi_{\ell,j}^p$  and  $\tilde{u}_\ell^k = \sum_{j=1}^{N_\ell^p} (\tilde{\mathbf{x}}_\ell^k)_j \varphi_{\ell,j}^p$ , this directly translates to

$$\|u_\ell^* - \tilde{u}_\ell^{k+1}\| \leq q \|u_\ell^* - \tilde{u}_\ell^k\| \quad \text{for all } k \in \mathbb{N}_0.$$

**Proof.** Let  $u_\ell \in \mathcal{X}_\ell^p$  and  $\mathbf{y}_\ell = \chi_\ell^p(u_\ell)$ . By the assumption  $\chi_\ell^p(\mathcal{S}_\ell u_\ell) = \mathbf{B}_\ell[\mathbf{A}_\ell \mathbf{y}_\ell]$ , the definition of the energy norm, and (12), we have

$$|\mathbf{y}_\ell - \mathbf{B}_\ell[\mathbf{A}_\ell \mathbf{y}_\ell]|_{\mathbf{A}_\ell} = \|u_\ell - \mathcal{S}_\ell u_\ell\| \leq q \|u_\ell\| = q |\mathbf{y}_\ell|_{\mathbf{A}_\ell} \quad \text{for all } \mathbf{y}_\ell \in \mathbb{R}^{N_\ell^p}.$$

Choosing  $\mathbf{y}_\ell = \mathbf{A}_\ell^{-1} \mathbf{x}_\ell$ , we obtain assumption (11). Applying [Bla02, Theorem 3.4] gives the desired contraction (13).  $\square$

**Remark 3.** The result of Proposition 2 is entirely to be expected, since all quantities have been defined and denoted so as to fit into the framework of [Bla02]. We emphasize, however, that in practice one must still verify that a given contractive solver can indeed be represented as a preconditioner, i.e., the property  $\chi_\ell^p(\mathcal{S}_\ell u_\ell) = \mathbf{B}_\ell[\mathbf{A}_\ell \chi_\ell^p(u_\ell)]$  for all  $u_\ell \in \mathcal{X}_\ell^p$  has to be validated. For the geometric multigrid method from [IMPS24], this is done in the next section.

#### 4. OPTIMAL NON-LINEAR AND NON-SYMMETRIC MULTIGRID PRECONDITIONER

In this section, we consider the  $h$ - and  $p$ -robustly contractive geometric multigrid method from [IMPS24] and rewrite the algorithm in matrix formulation with the purpose of verifying (11) and using this solver as a (non-symmetric and non-linear) preconditioner for GPCG.

**4.1.  $h$ - and  $p$ -robust geometric multigrid method.** For fixed  $\ell$ , we define the residual functional  $R_\ell : \mathcal{X}_\ell^p \rightarrow \mathbb{R}$  associated with the current approximation  $u_\ell$  of  $u_\ell^*$  by  $R_\ell(v_\ell) := \langle\langle u_\ell^* - u_\ell, v_\ell \rangle\rangle$ . Then, one V-cycle of the geometric multigrid method is given by the following algorithm.

**Algorithm C (V-cycle of local multigrid method [IMPS24]).** *Input:* Current approximation  $u_\ell \in \mathcal{X}_\ell^p$ , triangulations  $\{\mathcal{T}_{\ell'}\}_{\ell'=0}^\ell$  and polynomial degree  $p \geq 1$ .

Perform the following steps (i)–(iii):

- (i) **Lowest-order coarse solve:** Find  $\rho_0 \in \mathcal{X}_0^1$  such that

$$\langle\langle \rho_0, v_0 \rangle\rangle = R_\ell(v_0) \quad \text{for all } v_0 \in \mathcal{X}_0^1. \quad (14)$$

Define  $\sigma_0 := \rho_0$  and  $\lambda_0 := 1$ .

- (ii) **Local lowest-order correction:** For all intermediate levels  $\ell' = 1, \dots, \ell - 1$  and all  $z \in \mathcal{V}_{\ell'}^+$ , compute  $\rho_{\ell',z} \in \mathcal{X}_{\ell',z}^1$  such that

$$\langle\langle \rho_{\ell',z}, v_{\ell',z} \rangle\rangle = R_\ell(v_{\ell',z}) - \langle\langle \sigma_{\ell'-1}, v_{\ell',z} \rangle\rangle \quad \text{for all } v_{\ell',z} \in \mathcal{X}_{\ell',z}^1. \quad (15)$$

Define  $\rho_{\ell'} := \sum_{z \in \mathcal{V}_{\ell'}^+} \rho_{\ell',z}$ ,  $\nu_{\ell'} := \frac{R_\ell(\rho_{\ell'}) - \langle\langle \sigma_{\ell'-1}, \rho_{\ell'} \rangle\rangle}{\|\rho_{\ell'}\|^2}$ , and

$$\sigma_{\ell'} := \sigma_{\ell'-1} + \lambda_{\ell'} \rho_{\ell'}, \quad \text{where } \lambda_{\ell'} := \begin{cases} \nu_{\ell'} & \text{if } \nu_{\ell'} \leq d + 1, \\ (d + 1)^{-1} & \text{otherwise.} \end{cases}$$

- (iii) **Local high-order correction:** For all  $z \in \mathcal{V}_\ell$ , compute  $\rho_{\ell,z} \in \mathcal{X}_{\ell,z}^p$  such that

$$\langle\langle \rho_{\ell,z}, v_{\ell,z} \rangle\rangle = R_\ell(v_{\ell,z}) - \langle\langle \sigma_{\ell-1}, v_{\ell,z} \rangle\rangle \quad \text{for all } v_{\ell,z} \in \mathcal{X}_{\ell,z}^p. \quad (16)$$

Define  $\rho_\ell := \sum_{z \in \mathcal{V}_\ell} \rho_{\ell,z}$  and

$$\sigma_\ell := \sigma_{\ell-1} + \lambda_\ell \rho_\ell, \quad \text{where } \lambda_\ell := \frac{R_\ell(\rho_\ell) - \langle\langle \sigma_{\ell-1}, \rho_\ell \rangle\rangle}{\|\rho_\ell\|^2}.$$

**Output:** Improved approximation  $\Phi_\ell(u_\ell) := u_\ell + \sigma_\ell$ .

---

**Remark 4.** In the case  $p = 1$ , it suffices to restrict step (iii) of Algorithm C to  $\mathcal{V}_\ell^+$ , i.e., the local problems (16) need only be solved for  $z \in \mathcal{V}_\ell^+$ . All comments and results then hold accordingly.

---

Using the Galerkin matrix  $\mathbf{A}_\ell$  from (5) and the vector  $\mathbf{b}_\ell$  from (6), we want to rewrite Algorithm C in terms of linear algebra. To this end, we introduce the following notation: For  $\ell' \in \{0, \dots, \ell\}$ , consider, only for an analytical perspective, the embedding  $\mathcal{I}_{\ell'}^+ : \mathcal{X}_{\ell'}^+ \rightarrow \mathcal{X}_\ell^p$ , i.e., the formal identity, with matrix representation  $\mathbf{I}_{\ell'}^+ \in \mathbb{R}^{N_\ell^p \times N_{\ell'}^+}$ , where  $N_{\ell'}^+ := \dim(\mathcal{X}_{\ell'}^+)$ . Similarly, the embeddings  $\mathcal{I}_{\ell',z}^p : \mathcal{X}_{\ell',z}^p \rightarrow \mathcal{X}_\ell^p$  also have matrix representations  $\mathbf{I}_{\ell',z}^p \in \mathbb{R}^{N_\ell^p \times N_{\ell',z}^p}$  with  $N_{\ell',z}^p := \dim(\mathcal{X}_{\ell',z}^p)$ . Denote by  $\mathbf{A}_{\ell'}$ ,  $\mathbf{A}_{\ell'}^+$ , and  $\mathbf{A}_{\ell',z}^p$  the Galerkin matrices with respect to  $\mathcal{X}_{\ell'}^1$ ,  $\mathcal{X}_{\ell'}^+$ , and  $\mathcal{X}_{\ell',z}^p$ , respectively. We define the diagonal matrices  $\mathbf{D}_{\ell'}^+ \in \mathbb{R}^{N_{\ell'}^+ \times N_{\ell'}^+}$  via  $(\mathbf{D}_{\ell'}^+)_{jk} := \delta_{jk}(\mathbf{A}_{\ell'}^+)_{jj}$ . Furthermore, consider the levelwise smoothers

$$\begin{aligned} \mathbf{S}_0 &:= \mathbf{I}_0^+ \mathbf{A}_0^{-1} (\mathbf{I}_0^+)^T \mathbf{A}_\ell, & \mathbf{S}_{\ell'} &:= \mathbf{I}_{\ell'}^+ (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^+)^T \mathbf{A}_\ell \quad \text{for } \ell' = 1, \dots, \ell - 1, \quad \text{and} \\ \mathbf{S}_\ell &:= \sum_{z \in \mathcal{V}_\ell} \mathbf{I}_{\ell,z}^p (\mathbf{A}_{\ell,z}^p)^{-1} (\mathbf{I}_{\ell,z}^p)^T \mathbf{A}_\ell. \end{aligned} \tag{17}$$

Lastly, we note that matrix products, and analogously products of operators, are generally non-commutative. In particular, for matrices or operators  $C_0, \dots, C_\ell$ , we define

$$\prod_{\ell'=0}^{\ell} C_{\ell'} := C_0 C_1 \cdots C_\ell \quad \text{as well as} \quad \prod_{\ell'=\ell}^0 C_{\ell'} := C_\ell C_{\ell-1} \cdots C_0.$$

In [IMPS24], it is already remarked that the solver from Algorithm C is of linear complexity per step. However, let us provide some more details.

**Remark 5 (Computational complexity of Algorithm C).** The computational cost on the initial mesh depends only on  $\#\mathcal{T}_0$ . The matrices for the local high-order problems (16) have dimension  $\mathcal{O}(p^d)$ , where the notationally hidden constant depends only on  $\gamma$ -shape regularity. Since we solve such a system on every patch, the computational complexity on the finest level is of order  $\mathcal{O}(p^{3d} \#\mathcal{T}_\ell)$ . As  $\dim(\mathcal{X}_{\ell',z}^1) = 1$ , each of local lowest-order problems (15) can be solved in  $\mathcal{O}(1)$  operations. Hence, the computation of  $\rho_{\ell'}$  is of order  $\mathcal{O}(\#\mathcal{V}_{\ell'}^+)$  for  $\ell' = 1, \dots, \ell - 1$ . Moreover, let us discuss the calculation of the step-size  $\lambda_{\ell'}$  when it is bounded by  $(d + 1)$ . By definition of the local problems (15), we have

$$R_\ell(\rho_{\ell'}) - \langle \sigma_{\ell'-1}, v_{\ell',z} \rangle = \sum_{z \in \mathcal{V}_{\ell'}^+} \|\rho_{\ell',z}\|^2$$

and the resulting sum can be computed in  $\mathcal{O}(\#\mathcal{V}_{\ell'}^+)$  operations. Let  $\mathbf{s}_{\ell'} = \chi_{\ell'}^1[\rho_{\ell'}]$  denote the vector corresponding to  $\rho_{\ell'}$ . Since the Galerkin matrix  $\mathbf{A}_{\ell'}$  is sparse and we are solely interested in the value  $\|\rho_{\ell'}\|^2 = \mathbf{s}_{\ell'}^T \mathbf{A}_{\ell'} \mathbf{s}_{\ell'}$  and  $\rho_{\ell'} \in \mathcal{X}_{\ell'}^+$ , we only need to compute the entries of  $\mathbf{A}_{\ell'} \mathbf{s}_{\ell'}$  corresponding to the vertices in  $\mathcal{V}_{\ell'}^+$ . This can be done in  $\mathcal{O}(\#\mathcal{V}_{\ell'}^+)$  operations. Thus, the overall computational complexity on an intermediate level  $\ell'$  is of order  $\mathcal{O}(\#\mathcal{V}_{\ell'}^+)$ . By the definition of  $\mathcal{V}_{\ell'}^+$ , we have

$$\sum_{\ell'=1}^{\ell-1} \#\mathcal{V}_{\ell'}^+ \lesssim \sum_{\ell'=1}^{\ell-1} \#(\mathcal{V}_{\ell'} \setminus \mathcal{V}_{\ell'-1}) = \sum_{\ell'=1}^{\ell-1} (\#\mathcal{V}_{\ell'} - \#\mathcal{V}_{\ell'-1}) = \#\mathcal{V}_{\ell-1} - \#\mathcal{V}_0 \leq \#\mathcal{V}_\ell \simeq \#\mathcal{T}_\ell.$$

Therefore, the overall computational complexity of Algorithm C is of order  $\mathcal{O}(\#\mathcal{T}_\ell)$  and the notationally hidden constant depends only on the initial mesh  $\mathcal{T}_0$ , the polynomial degree  $p$ , the dimension  $d$ , and  $\gamma$ -shape regularity. In particular, the overall complexity does not depend on the number  $\ell + 1$  of triangulations  $\mathcal{T}_0, \dots, \mathcal{T}_\ell$ .

Note that the matrices  $\mathbf{I}_{\ell'}^+$  and  $\mathbf{I}_{\ell,z}^p$  above are introduced solely for the purpose of the analysis and are not computed in the actual implementation. In practice, Algorithm C only employs the prolongation and restriction matrices between consecutive levels. 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. This recursive prolongation and restriction between consecutive levels is standard practice in multigrid methods; see, e.g., [XQ94].

Let us denote the algebraic residual of the current iterate  $u_\ell \in \mathcal{X}_\ell$  by  $\mathbf{r}_\ell := \mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell \in \mathbb{R}^{N_\ell^p}$ , where  $\mathbf{x}_\ell = \chi_\ell^p(u_\ell) \in \mathbb{R}^{N_\ell^p}$ . With the matrices from (17), we obtain

- (i) **Lowest-order coarse solve:** Let  $\tilde{\mathbf{s}}_0 = \chi_0^1(\rho_0) \in \mathbb{R}^{N_0^1}$  denote the coefficient vector of the solution  $\rho_0 \in \mathcal{X}_0^1$  to the coarse problem (14). Since  $\mathcal{V}_0^+ = \mathcal{V}_0$  and hence  $\mathcal{X}_0^+ = \mathcal{X}_0$ , problem (14) can equivalently be reformulated by the linear system  $\mathbf{A}_0 \tilde{\mathbf{s}}_0 = (\mathbf{I}_0^+)^T \mathbf{r}_\ell$ . Due to the nestedness of the finite element spaces, it holds  $\rho_0 \in \mathcal{X}_\ell^p$  and thus  $\mathbf{s}_0 = \mathbf{I}_0^+ \tilde{\mathbf{s}}_0$  for  $\mathbf{s}_0 = \chi_\ell^p(\rho_0) \in \mathbb{R}^{N_\ell^p}$ . Hence, we have

$$\mathbf{s}_0 = \mathbf{I}_0^+ \tilde{\mathbf{s}}_0 = \mathbf{I}_0^+ (\mathbf{A}_0)^{-1} (\mathbf{I}_0^+)^T \mathbf{r}_\ell =: \mathbf{B}_0[\mathbf{r}_\ell].$$

- (ii) **Local lowest-order correction:** Combining the local contributions  $\rho_{1,z}$  on the first level, we consider  $\mathbf{s}_1 = \chi_\ell^p(\rho_1)$ . From (15), it follows that

$$\begin{aligned} \mathbf{s}_1 &= \mathbf{I}_1^+ (\mathbf{D}_1^+)^{-1} (\mathbf{I}_1^+)^T (\mathbf{r}_\ell - \lambda_0 \mathbf{A}_\ell \mathbf{s}_0) \\ &= \mathbf{I}_1^+ (\mathbf{D}_1^+)^{-1} (\mathbf{I}_1^+)^T (\mathbf{I} - \lambda_0 \mathbf{A}_\ell \mathbf{I}_0^+ (\mathbf{A}_0)^{-1} (\mathbf{I}_0^+)^T) \mathbf{r}_\ell =: \mathbf{B}_1[\mathbf{r}_\ell]. \end{aligned}$$

For  $\mathbf{s}_2 = \chi_\ell^p(\rho_2) \in \mathbb{R}^{N_\ell^p}$ , we can therefore show

$$\begin{aligned} \mathbf{s}_2 &= \mathbf{I}_2^+ (\mathbf{D}_2^+)^{-1} (\mathbf{I}_2^+)^T (\mathbf{r}_\ell - \mathbf{A}_\ell (\lambda_1 \mathbf{s}_1 + \lambda_0 \mathbf{s}_0)) \\ &= \mathbf{I}_2^+ (\mathbf{D}_2^+)^{-1} (\mathbf{I}_2^+)^T \left( \mathbf{r}_\ell - \lambda_1 \mathbf{A}_\ell \mathbf{I}_1^+ (\mathbf{D}_1^+)^{-1} (\mathbf{I}_1^+)^T (\mathbf{r}_\ell - \lambda_0 \mathbf{A}_\ell \mathbf{B}_0[\mathbf{r}_\ell]) - \lambda_0 \mathbf{A}_\ell \mathbf{B}_0[\mathbf{r}_\ell] \right) \\ &= \mathbf{I}_2^+ (\mathbf{D}_2^+)^{-1} (\mathbf{I}_2^+)^T (\mathbf{I} - \lambda_1 \mathbf{A}_\ell \mathbf{I}_1^+ (\mathbf{D}_1^+)^{-1} (\mathbf{I}_1^+)^T) (\mathbf{I} - \lambda_0 \mathbf{A}_\ell \mathbf{I}_0^+ (\mathbf{A}_0)^{-1} (\mathbf{I}_0^+)^T) \mathbf{r}_\ell =: \mathbf{B}_2[\mathbf{r}_\ell]. \end{aligned}$$

For easier notation, we set  $\mathbf{D}_0^+ := \mathbf{A}_0$ . By induction on  $\ell'$ , we obtain the update

$$\mathbf{s}_{\ell'} = \mathbf{I}_{\ell'}^+ (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^+)^T \prod_{i=\ell'-1}^0 (\mathbf{I} - \lambda_i \mathbf{A}_\ell \mathbf{I}_i^+ (\mathbf{D}_i^+)^{-1} (\mathbf{I}_i^+)^T) \mathbf{r}_\ell =: \mathbf{B}_{\ell'}[\mathbf{r}_\ell], \quad (18)$$

where  $\mathbf{s}_{\ell'} = \chi_\ell^p(\rho_{\ell'}) \in \mathbb{R}^{N_\ell^p}$  and  $\ell' = 1, \dots, \ell - 1$ .

- (iii) **Local high-order correction:** On the finest level, we have

$$\mathbf{s}_\ell = \sum_{z \in \mathcal{V}_\ell} \mathbf{I}_{\ell,z}^p (\mathbf{A}_{\ell,z}^p)^{-1} (\mathbf{I}_{\ell,z}^p)^T \prod_{\ell'=0}^{\ell-1} (\mathbf{I} - \lambda_{\ell'} \mathbf{A}_\ell \mathbf{I}_{\ell'}^+ (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^+)^T) \mathbf{r}_\ell =: \mathbf{B}_\ell[\mathbf{r}_\ell].$$

Thus, we obtain the update

$$\chi_\ell^p(\Phi_\ell(u_\ell)) = \mathbf{x}_\ell + \sum_{\ell'=0}^{\ell-1} \lambda_{\ell'} \mathbf{s}_{\ell'} = \mathbf{x}_\ell + \sum_{\ell'=0}^{\ell-1} \lambda_{\ell'} \mathbf{B}_{\ell'}[\mathbf{r}_\ell]. \quad (19)$$

**Remark 6 (Nonlinearity of the preconditioner).** We note that the step-sizes  $\lambda_{\ell'}$  for  $\ell' \in \{1, \dots, \ell\}$  are also functions of the current residual  $\mathbf{r}_\ell$ , i.e.,  $\lambda_{\ell'} = \lambda_{\ell'}[\mathbf{r}_\ell]$ . More precisely, for  $\ell' < \ell$  the step-sizes are defined as

$$\lambda_{\ell'}[\mathbf{r}_\ell] := \begin{cases} \nu_{\ell'}[\mathbf{r}_\ell] & \text{if } \nu_{\ell'}[\mathbf{r}_\ell] \leq d + 1 \\ (d + 1)^{-1} & \text{otherwise,} \end{cases}$$

where

$$\nu_{\ell'}[\mathbf{r}_\ell] := \frac{(\mathbf{r}_\ell, \mathbf{B}_{\ell'}[\mathbf{r}_\ell])_2 - \left(\sum_{i=0}^{\ell'-1} \lambda_i[\mathbf{r}_\ell] \mathbf{B}_i[\mathbf{r}_\ell], \mathbf{B}_{\ell'}[\mathbf{r}_\ell]\right)_{\mathbf{A}_\ell}}{|\mathbf{B}_{\ell'}[\mathbf{r}_\ell]|_{\mathbf{A}_\ell}^2}.$$

In the case  $\ell' = \ell$ , we have  $\lambda_\ell[\mathbf{r}_\ell] = \nu_\ell[\mathbf{r}_\ell]$ . From this, it is clear that the operators  $\mathbf{B}_{\ell'} : \mathbb{R}^{N_\ell^p} \rightarrow \mathbb{R}^{N_\ell^p}$  are indeed non-linear. For easier notation, we will omit the dependence on  $\mathbf{r}_\ell$  in the following when it is clear from the context. Moreover, we set  $\lambda_0[\mathbf{r}_\ell] := 1$ .

Consider the orthogonal projections  $\mathcal{P}_{\ell'}^q : \mathcal{X}_\ell^p \rightarrow \mathcal{X}_{\ell'}^q$  and  $\mathcal{P}_{\ell',z}^q : \mathcal{X}_\ell^p \rightarrow \mathcal{X}_{\ell',z}^q$  which, for  $q \in \{1, p\}$  and each  $v \in \mathcal{X}_\ell^p$ , are given by

$$\langle\langle \mathcal{P}_{\ell'}^q v, w_{\ell'} \rangle\rangle = \langle\langle v, w_{\ell'} \rangle\rangle \quad \text{for all } w_{\ell'} \in \mathcal{X}_{\ell'}^q, \quad (20)$$

$$\langle\langle \mathcal{P}_{\ell',z}^q v, w_{\ell',z} \rangle\rangle = \langle\langle v, w_{\ell',z} \rangle\rangle \quad \text{for all } w_{\ell',z} \in \mathcal{X}_{\ell',z}^q. \quad (21)$$

With these, we can define levelwise operators  $\mathcal{S}_0 : \mathcal{X}_\ell^p \rightarrow \mathcal{X}_0^1$ ,  $\mathcal{S}_{\ell'} : \mathcal{X}_\ell^p \rightarrow \mathcal{X}_{\ell'}^+$  for  $\ell' = 1, \dots, \ell-1$ , and  $\mathcal{S}_\ell : \mathcal{X}_\ell^p \rightarrow \mathcal{X}_\ell^p$  via

$$\mathcal{S}_0 := \mathcal{P}_0^1, \quad \mathcal{S}_{\ell'} := \sum_{z \in \mathcal{V}_{\ell'}^+} \mathcal{P}_{\ell',z}^1 \quad \text{for } \ell' = 1, \dots, \ell-1, \quad \text{and} \quad \mathcal{S}_\ell := \sum_{z \in \mathcal{V}_\ell} \mathcal{P}_{\ell,z}^p. \quad (22)$$

The following lemma shows the connection between the levelwise operators  $\mathcal{S}_{\ell'}$  from (22) and the matrices  $\mathbf{S}_{\ell'}$  from (17).

**Lemma 7 (Levelwise smoothers functional/matrix representation).** *Let  $\ell' \in \{0, \dots, \ell\}$ . Then, it holds that*

$$\langle\langle \mathcal{S}_{\ell'} v_\ell, w_\ell \rangle\rangle = (\mathbf{S}_{\ell'} \mathbf{x}_\ell, \mathbf{y}_\ell)_{\mathbf{A}_\ell} \quad \text{for all } v_\ell, w_\ell \in \mathcal{X}_\ell^p \text{ and } \mathbf{x}_\ell = \chi_\ell^p(v_\ell), \mathbf{y}_\ell = \chi_\ell^p(w_\ell). \quad (23)$$

**Proof.** Let  $v_\ell, w_\ell \in \mathcal{X}_\ell^p$ ,  $\mathbf{y}_\ell = \chi_\ell^p(w_\ell)$ ,  $\mathbf{x}_\ell = \chi_\ell^p(v_\ell)$ ,  $\mathbf{x}_0 = \chi_\ell^p(\mathcal{P}_0^1 v_\ell)$ , and  $\tilde{\mathbf{x}}_0 = \chi_0^1(\mathcal{P}_0^1 v_\ell) \in \mathbb{R}^{N_0^1}$ . Then, the definition (20) of  $\mathcal{P}_0^1$  implies that  $\mathbf{A}_0 \tilde{\mathbf{x}}_0 = (\mathbf{I}_0^+)^T \mathbf{A}_\ell \mathbf{x}_\ell$ . Using  $\mathcal{S}_0 = \mathcal{P}_0^1$ , we thus get

$$\mathbf{x}_0 = \mathbf{I}_0^+ \tilde{\mathbf{x}}_0 = \mathbf{I}_0^+ \mathbf{A}_0^{-1} (\mathbf{I}_0^+)^T \mathbf{A}_\ell \mathbf{x}_\ell \stackrel{(17)}{=} \mathbf{S}_0 \mathbf{x}_\ell. \quad (24)$$

Therefore, we have

$$\langle\langle \mathcal{S}_0 v_\ell, w_\ell \rangle\rangle = (\mathbf{x}_0, \mathbf{y}_\ell)_{\mathbf{A}_\ell} \stackrel{(24)}{=} (\mathbf{S}_0 \mathbf{x}_\ell, \mathbf{y}_\ell)_{\mathbf{A}_\ell}.$$

For  $\ell' = 1, \dots, \ell-1$ , we similarly find

$$\mathbf{x}_{\ell'} = \mathbf{I}_{\ell'}^+ (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^+)^T \mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{S}_{\ell'} \mathbf{x}_\ell$$

and

$$\langle\langle \mathcal{S}_{\ell'} v_\ell, w_\ell \rangle\rangle = (\mathbf{x}_{\ell'}, \mathbf{y}_\ell)_{\mathbf{A}_{\ell'}} = (\mathbf{S}_{\ell'} \mathbf{x}_\ell, \mathbf{y}_\ell)_{\mathbf{A}_\ell}.$$

Finally, for the finest level  $\ell$ , we obtain (23) directly from the definition (21) of  $\mathcal{P}_{\ell,z}^p$  and summation over  $z \in \mathcal{V}_\ell$ . This concludes the proof.  $\square$

**4.2. Induced multigrid preconditioner and corresponding main result.** With the notation from the last section, we define the multigrid preconditioner

$$\mathbf{B}_\ell^{\text{MG}} := \sum_{\ell'=0}^{\ell} \lambda_{\ell'} \mathbf{B}_{\ell'} : \mathbb{R}^{N_\ell^p} \rightarrow \mathbb{R}^{N_\ell^p}. \quad (25)$$

The subsequent theorem shows that GPCG with preconditioner  $\mathbf{B}_\ell^{\text{MG}}$  contracts by an  $h$ - and  $p$ -independent factor, where the proof is postponed to Section 4.3. Together with Remark 5, it follows that the preconditioner  $\mathbf{B}_\ell^{\text{MG}}$  is indeed optimal.

**Theorem 8 (GPCG with non-linear and non-symmetric MG preconditioner).** Let  $\mathbf{x}_\ell^k$  be the GPCG iterates generated by Algorithm B applied to the Galerkin system  $\mathbf{A}_\ell \mathbf{x}_\ell^* = \mathbf{b}_\ell$  employing the non-linear and non-symmetric multigrid preconditioner  $\mathbf{B}_\ell^{\text{MG}}$  defined in (25). Then, there holds

$$|\mathbf{x}_\ell^* - \mathbf{x}_\ell^{k+1}|_{\mathbf{A}_\ell} \leq q |\mathbf{x}_\ell^* - \mathbf{x}_\ell^k|_{\mathbf{A}_\ell} \quad \text{for all } k \in \mathbb{N}_0, \quad (26)$$

where  $q$  is independent of  $h$ ,  $p$ , and only depends locally on the diffusion contrast  $\mathbf{K}$ . For  $u_\ell^* = \sum_{j=1}^{N_\ell^p} (\mathbf{x}_\ell^*)_j \varphi_{\ell,j}^p$  and  $u_\ell^k = \sum_{j=1}^{N_\ell^p} (\mathbf{x}_\ell^k)_j \varphi_{\ell,j}^p$ , this directly translates to

$$\|u_\ell^* - u_\ell^{k+1}\| \leq q \|u_\ell^* - u_\ell^k\| \quad \text{for all } k \in \mathbb{N}_0.$$

The following auxiliary lemma simplifies the application of the multigrid preconditioner (25).

**Lemma 9.** Let  $\{\mathbf{M}_{\ell'}\}_{\ell'=0}^\ell$  be a family of matrices. Then, it holds that

$$\sum_{\ell'=0}^{\ell} \mathbf{M}_{\ell'} \prod_{i=\ell'-1}^0 (\mathbf{I} - \mathbf{M}_i) = \mathbf{I} - \prod_{\ell'=\ell}^0 (\mathbf{I} - \mathbf{M}_{\ell'}). \quad (27)$$

*Proof.* The proof is done by induction on  $\ell$ . For  $\ell = 0$ , we have

$$\mathbf{M}_0 = \mathbf{I} - (\mathbf{I} - \mathbf{M}_0).$$

Using the induction hypothesis for  $\ell - 1$ , we obtain

$$\begin{aligned} \mathbf{I} - \prod_{\ell'=\ell}^0 (\mathbf{I} - \mathbf{M}_{\ell'}) &= \mathbf{I} - (\mathbf{I} - \mathbf{M}_\ell) \prod_{\ell'=\ell-1}^0 (\mathbf{I} - \mathbf{M}_{\ell'}) = \mathbf{I} - \prod_{\ell'=\ell-1}^0 (\mathbf{I} - \mathbf{M}_{\ell'}) + \mathbf{M}_\ell \prod_{i=\ell-1}^0 (\mathbf{I} - \mathbf{M}_i) \\ &= \sum_{\ell'=0}^{\ell-1} \mathbf{M}_{\ell'} \prod_{i=\ell'-1}^0 (\mathbf{I} - \mathbf{M}_i) + \mathbf{M}_\ell \prod_{i=\ell-1}^0 (\mathbf{I} - \mathbf{M}_i) = \sum_{\ell'=0}^{\ell} \mathbf{M}_{\ell'} \prod_{i=\ell'-1}^0 (\mathbf{I} - \mathbf{M}_i). \end{aligned}$$

This concludes the proof.  $\square$

The application of the multigrid preconditioner  $\mathbf{B}_\ell^{\text{MG}}$  to  $\mathbf{A}_\ell \mathbf{x}_\ell$  for  $\mathbf{x}_\ell \in \mathbb{R}^{N_\ell^p}$  can hence be simplified using the levelwise matrices  $\mathbf{S}_{\ell'}$  defined in (17). First, we note that

$$\begin{aligned} \mathbf{B}_{\ell'}[\mathbf{A}_\ell \mathbf{x}_\ell] &\stackrel{(18)}{=} \mathbf{I}_{\ell'}^+ (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^+)^T \prod_{i=\ell'-1}^0 (\mathbf{I} - \lambda_i [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{A}_\ell \mathbf{I}_i^+ (\mathbf{D}_i^+)^{-1} (\mathbf{I}_i^+)^T) \mathbf{A}_\ell \mathbf{x}_\ell \\ &= \mathbf{I}_{\ell'}^+ (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^+)^T \mathbf{A}_\ell \prod_{i=\ell'-1}^0 (\mathbf{I} - \lambda_i [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{I}_i^+ (\mathbf{D}_i^+)^{-1} (\mathbf{I}_i^+)^T \mathbf{A}_\ell) \mathbf{x}_\ell \\ &= \mathbf{S}_{\ell'} \prod_{i=\ell'-1}^0 (\mathbf{I} - \lambda_i [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_i) \mathbf{x}_\ell. \end{aligned} \quad (28)$$

Applying Lemma 9 to  $\mathbf{M}_{\ell'} = \lambda_{\ell'} [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_{\ell'}$ , we thus observe that

$$\begin{aligned} \mathbf{B}_\ell^{\text{MG}}[\mathbf{A}_\ell \mathbf{x}_\ell] &\stackrel{(25)}{=} \sum_{\ell'=0}^{\ell} \lambda_{\ell'} [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{B}_{\ell'}[\mathbf{A}_\ell \mathbf{x}_\ell] \stackrel{(28)}{=} \mathbf{S}_0 \mathbf{x}_\ell + \sum_{\ell'=1}^{\ell} \lambda_{\ell'} [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_{\ell'} \prod_{i=\ell'-1}^0 (\mathbf{I} - \lambda_i [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_i) \mathbf{x}_\ell \\ &\stackrel{(27)}{=} \left( \mathbf{I} - \prod_{\ell'=\ell}^0 (\mathbf{I} - \lambda_{\ell'} [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_{\ell'}) \right) \mathbf{x}_\ell. \end{aligned}$$

With the identity  $I$ , let us define the operator  $\mathcal{S}_\ell^{\text{MG}} : \mathcal{X}_\ell^p \rightarrow \mathcal{X}_\ell^p$  as

$$\mathcal{S}_\ell^{\text{MG}} v_\ell := \left( I - \prod_{\ell'=\ell}^0 (I - \lambda_{\ell'} [\mathbf{A}_\ell \mathcal{X}_\ell^p(v_\ell)] \mathbf{S}_{\ell'}) \right) v_\ell. \quad (29)$$

---

**Remark 10 (Link multigrid and Schwarz methods).** *The presented multigrid preconditioner (25) is related to multiplicative Schwarz methods. To be precise, the operator  $\mathcal{S}_\ell^{\text{MG}}$  has a multiplicative structure with levelwise operators  $\lambda_{\ell'}\mathcal{S}_{\ell'}$ . However, these operators are non-linear since the step-sizes  $\lambda_{\ell'}$  depend on the input  $v_\ell \in \mathcal{X}_\ell^p$  via  $\lambda_{\ell'} = \lambda_{\ell'}[\mathbf{A}_\ell\chi_\ell^p(v_\ell)]$ . If one were to omit the step-sizes  $\lambda_{\ell'}$ , then the operator  $\mathcal{S}_\ell^{\text{MG}}$  would be a multiplicative Schwarz operator in the classical sense. For more details on multiplicative Schwarz methods, we refer to [BPWX91; CW93].*

---

The subsequent proposition shows the connection between the operator  $\mathcal{S}_\ell^{\text{MG}}$  from (29) and the preconditioner  $\mathbf{B}_\ell^{\text{MG}}$  from (25).

**Proposition 11 (Multigrid functional/matrix representation).** *Let  $\mathcal{S}_\ell^{\text{MG}}$  be the multiplicative Schwarz operator defined in (29) and  $\mathbf{B}_\ell^{\text{MG}}$  the multigrid preconditioner defined in (25). Then, it holds that*

$$\langle\langle \mathcal{S}_\ell^{\text{MG}}v_\ell, w_\ell \rangle\rangle = (\mathbf{B}_\ell^{\text{MG}}[\mathbf{A}_\ell\mathbf{x}_\ell], \mathbf{y}_\ell)_{\mathbf{A}_\ell} \quad \text{for all } v_\ell, w_\ell \in \mathcal{X}_\ell^p \text{ and } \mathbf{x}_\ell = \chi_\ell^p(v_\ell), \mathbf{y}_\ell = \chi_\ell^p(w_\ell). \quad (30)$$

Moreover, this implies

$$\chi_\ell^p(\mathcal{S}_\ell^{\text{MG}}v_\ell) = \mathbf{B}_\ell^{\text{MG}}[\mathbf{A}_\ell\chi_\ell^p(v_\ell)] \quad \text{for all } v_\ell \in \mathcal{X}_\ell^p. \quad (31)$$

Lastly, for the update  $\sigma_\ell = \Phi_\ell(u_\ell) - u_\ell$  defined in Algorithm C(iii), we have

$$\mathcal{S}_\ell^{\text{MG}}(u_\ell^* - u_\ell) = \sigma_\ell \quad \text{for all } u_\ell \in \mathcal{X}_\ell^p. \quad (32)$$


---

**Proof.** Let  $\ell \in \mathbb{N}_0$  be fixed. Let  $v_\ell, w_\ell \in \mathcal{X}_\ell^p$  and  $\mathbf{x}_\ell = \chi_\ell^p(v_\ell)$ ,  $\mathbf{y}_\ell = \chi_\ell^p(w_\ell)$ . We show that

$$\left\langle\left\langle \prod_{\ell'=\ell}^0 (I - \lambda_{\ell'}[\mathbf{A}_{\ell'}\chi_{\ell'}^p(v_{\ell'})]\mathcal{S}_{\ell'}v_\ell, w_\ell \right\rangle\right\rangle = \left( \prod_{\ell'=\ell}^0 (\mathbf{I} - \lambda_{\ell'}[\mathbf{A}_{\ell'}\mathbf{x}_{\ell'}]\mathbf{S}_{\ell'})\mathbf{x}_\ell, \mathbf{y}_\ell \right)_{\mathbf{A}_\ell} \quad (33)$$

by induction on  $\ell'$ , i.e., the induction hypothesis reads as

$$\left\langle\left\langle \prod_{i=\ell'}^0 (I - \lambda_i[\mathbf{A}_\ell\chi_\ell^p(v_\ell)]\mathcal{S}_i)v_\ell, w_\ell \right\rangle\right\rangle = \left( \prod_{i=\ell'}^0 (\mathbf{I} - \lambda_i[\mathbf{A}_\ell\mathbf{x}_\ell]\mathbf{S}_i)\mathbf{x}_\ell, \mathbf{y}_\ell \right)_{\mathbf{A}_\ell} \quad (34)$$

for  $\ell' \in \{0, \dots, \ell - 1\}$ . From Lemma 7, we immediately obtain the base case  $\ell' = 0$ . Since the operators  $\mathcal{S}_{\ell'+1}$  from (22) and  $\mathbf{S}_{\ell'+1}$  from (17) are symmetric with respect to the scalar products  $\langle\langle \cdot, \cdot \rangle\rangle$  and  $(\cdot, \cdot)_{\mathbf{A}_\ell}$ , respectively, it follows from the induction hypothesis (34) that

$$\begin{aligned} & \left\langle\left\langle \prod_{i=\ell'+1}^0 (I - \lambda_i[\mathbf{A}_\ell\chi_\ell^p(v_\ell)]\mathcal{S}_i)v_\ell, w_\ell \right\rangle\right\rangle \\ &= \left\langle\left\langle \prod_{i=\ell'}^0 (I - \lambda_i[\mathbf{A}_\ell\chi_\ell^p(v_\ell)]\mathcal{S}_i)v_\ell, (I - \lambda_{\ell'+1}[\mathbf{A}_\ell\chi_\ell^p(v_\ell)]\mathcal{S}_{\ell'+1})w_\ell \right\rangle\right\rangle \\ &\stackrel{(34)}{=} \left( \prod_{i=\ell'}^0 (\mathbf{I} - \lambda_i[\mathbf{A}_\ell\mathbf{x}_\ell]\mathbf{S}_i)\mathbf{x}_\ell, \tilde{\mathbf{y}}_\ell \right)_{\mathbf{A}_\ell}, \end{aligned}$$

where  $\tilde{\mathbf{y}}_\ell = \chi_\ell^p((I - \lambda_{\ell'+1}[\mathbf{A}_\ell\chi_\ell^p(v_\ell)]\mathcal{S}_{\ell'+1})w_\ell)$ . Let  $\tilde{v}_\ell \in \mathcal{X}_\ell^p$  denote the discrete function such that  $\prod_{i=\ell'}^0 (\mathbf{I} - \lambda_i[\mathbf{A}_\ell\mathbf{x}_\ell]\mathbf{S}_i)\mathbf{x}_\ell = \chi_\ell^p(\tilde{v}_\ell)$ . Utilizing the connection (23) and the symmetry of  $\mathbf{S}_{\ell'+1}$ ,

we get

$$\begin{aligned}
\left( \prod_{i=\ell'}^0 (\mathbf{I} - \lambda_i [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_i) \mathbf{x}_\ell, \tilde{\mathbf{y}}_\ell \right)_{\mathbf{A}_\ell} &= \langle \tilde{v}_\ell, (I - \lambda_{\ell'+1} [\mathbf{A}_\ell \chi_\ell^p(v_\ell)] \mathcal{S}_{\ell'+1}) w_\ell \rangle \\
&\stackrel{(23)}{=} \left( \prod_{i=\ell'}^0 (\mathbf{I} - \lambda_i [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_i) \mathbf{x}_\ell, (I - \lambda_{\ell'+1} [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_{\ell'+1}) \mathbf{y}_\ell \right)_{\mathbf{A}_\ell} \\
&= \left( \prod_{i=\ell'+1}^0 (\mathbf{I} - \lambda_i [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_i) \mathbf{x}_\ell, \mathbf{y}_\ell \right)_{\mathbf{A}_\ell}.
\end{aligned}$$

This concludes the induction step and thus proves (33). Since  $\langle v_\ell, w_\ell \rangle = (\mathbf{x}_\ell, \mathbf{y}_\ell)_{\mathbf{A}_\ell}$ , we obtain

$$\begin{aligned}
\langle \mathcal{S}_\ell^{\text{MG}} v_\ell, w_\ell \rangle &\stackrel{(29)}{=} \langle v_\ell, w_\ell \rangle - \left\langle \prod_{\ell'=\ell}^0 (I - \lambda_{\ell'} [\mathbf{A}_\ell \chi_\ell^p(v_\ell)] \mathcal{S}_{\ell'}) v_\ell, w_\ell \right\rangle \\
&\stackrel{(33)}{=} (\mathbf{x}_\ell, \mathbf{y}_\ell)_{\mathbf{A}_\ell} - \left( \prod_{\ell'=\ell}^0 (\mathbf{I} - \lambda_{\ell'} [\mathbf{A}_\ell \mathbf{x}_\ell] \mathbf{S}_{\ell'}) \mathbf{x}_\ell, \mathbf{y}_\ell \right)_{\mathbf{A}_\ell} \stackrel{(25)}{=} (\mathbf{B}_\ell^{\text{MG}} [\mathbf{A}_\ell \mathbf{x}_\ell], \mathbf{y}_\ell)_{\mathbf{A}_\ell}.
\end{aligned}$$

From (30), we immediately have

$$(\chi_\ell^p(\mathcal{S}_\ell^{\text{MG}} v_\ell), \mathbf{y}_\ell)_{\mathbf{A}_\ell} = \langle \mathcal{S}_\ell^{\text{MG}} v_\ell, w_\ell \rangle \stackrel{(30)}{=} (\mathbf{B}_\ell^{\text{MG}} [\mathbf{A}_\ell \mathbf{x}_\ell], \mathbf{y}_\ell)_{\mathbf{A}_\ell}.$$

Since this holds for every  $\mathbf{y}_\ell \in \mathbb{R}^{N_\ell^p}$ , we obtain (31). Let  $u_\ell \in \mathcal{X}_\ell^p$ ,  $\sigma_\ell = \Phi_\ell(u_\ell) - u_\ell$  be the update constructed by Algorithm C, and  $\mathbf{s}_\ell = \chi_\ell^p(\sigma_\ell)$ . In (19), we derived that  $\mathbf{s}_\ell = \mathbf{B}_\ell^{\text{MG}}[\mathbf{r}_\ell]$ , where  $\mathbf{r}_\ell = \mathbf{A}_\ell(\mathbf{x}_\ell^* - \mathbf{x}_\ell)$ , with  $\mathbf{x}_\ell^* = \chi_\ell^p(u_\ell^*)$  and  $\mathbf{x}_\ell = \chi_\ell^p(u_\ell)$ . Finally, (31) implies  $\mathbf{B}_\ell^{\text{MG}}[\mathbf{A}_\ell(\mathbf{x}_\ell^* - \mathbf{x}_\ell)] = \chi_\ell^p(\mathcal{S}_\ell^{\text{MG}}(u_\ell^* - u_\ell))$  and hence  $\mathcal{S}_\ell^{\text{MG}}(u_\ell^* - u_\ell) = \sigma_\ell$ . This concludes the proof.  $\square$

Recall the following result on Algorithm C from [IMPS24, Theorem 2.5], where the local dependence on the diffusion coefficient is proved in [Hil25].

---

**Proposition 12 (Robust contraction of multigrid [IMPS24; Hil25]).** *Let  $u_\ell^* \in \mathcal{X}_\ell^p$  be the solution of (4) and  $u_\ell \in \mathcal{X}_\ell^p$  an approximation. Then, there exists a constant  $q \in (0, 1)$  such that*

$$\| \|u_\ell^* - \Phi_\ell(u_\ell)\| \| \leq q \| \|u_\ell^* - u_\ell\| \| \quad \text{for all } u_\ell \in \mathcal{X}_\ell^p. \quad (35)$$

The factor  $q$  depends only on the space dimension  $d$ , the initial mesh  $\mathcal{T}_0$ , the  $\gamma$ -shape regularity (2), and the local constants  $C_{\text{loc}}^{(1)}$  and  $C_{\text{loc}}^{(2)}$ , which are defined as

$$C_{\text{loc}}^{(1)} := \max \left\{ \sup_{z_0 \in \mathcal{V}_0} \frac{\max_{T \subseteq \omega_0^2(z_0)} \| \text{div}(\mathbf{K}) \|_{L^\infty(T)}}{\inf_{y \in \omega_0^2(z_0)} \lambda_{\min}(\mathbf{K}(y))}, \sup_{z_0 \in \mathcal{V}_0} \frac{\sup_{y \in \omega_0^2(z_0)} \lambda_{\max}(\mathbf{K}(y))}{\inf_{y \in \omega_0^2(z_0)} \lambda_{\min}(\mathbf{K}(y))} \right\}. \quad (36)$$

and

$$C_{\text{loc}}^{(2)} := \sup_{z \in \mathcal{V}_0} \frac{\sup_{y \in \omega_0^3(z)} \lambda_{\max}(\mathbf{K}(y))}{\inf_{y \in \omega_0^3(z)} \lambda_{\min}(\mathbf{K}(y))}. \quad \square \quad (37)$$

---

**Remark 13.** *Note that the analytical contraction factors  $q$  in Proposition 12 for the standalone geometric MG and in Theorem 8 for GPCG with non-linear and non-symmetric MG preconditioner are the same, i.e., GPCG can only improve contraction in practice. Indeed this is what we observe in our numerical experiments; see Section 8.*

---

**Remark 14.** *Instead of the optimal step-sizes  $\lambda_{\ell'}$  defined in Algorithm C(ii), one can alternatively use the fixed step-size  $\lambda := (d+1)^{-1}$  for  $\ell' \geq 1$  in Algorithm C. In this case the multigrid*

preconditioner  $\mathbf{B}_\ell^{\text{MG}}$  becomes linear and we denote it with  $\mathbf{B}_\ell^{\text{nsMG}}$ . All results of this section remain valid. In particular, Theorem 2 implies that  $\mathbf{B}_\ell^{\text{nsMG}}$  fulfills assumption (11), i.e.,

$$|(\mathbf{I} - \mathbf{B}_\ell^{\text{nsMG}} \mathbf{A}_\ell) \mathbf{x}_\ell|_{\mathbf{A}_\ell} \leq q |\mathbf{x}_\ell|_{\mathbf{A}_\ell} \quad \text{for all } \mathbf{x}_\ell \in \mathbb{R}^{N_\ell^p}, \quad (38)$$

where  $\mathbf{B}_\ell^{\text{nsMG}} \mathbf{A}_\ell = \mathbf{I} - \left( \prod_{\ell'=\ell}^1 (\mathbf{I} - \lambda \mathbf{S}_{\ell'}) \right) (\mathbf{I} - \mathbf{S}_0)$  and  $q \in (0, 1)$  is the constant from (35).

**4.3. Proof of Theorem 8.** We show that the multigrid preconditioner  $\mathbf{B}_\ell^{\text{MG}}$  fulfills the assumption of Theorem 2, i.e., contraction (12) of the operator  $\mathcal{S}_\ell^{\text{MG}}$ . Let  $u_\ell \in \mathcal{X}_\ell^p$ . With (32), we can rewrite the estimate (35) as

$$\| (I - \mathcal{S}_\ell^{\text{MG}})(u_\ell^* - u_\ell) \| = \| u_\ell^* - (u_\ell + \sigma_\ell) \| = \| u_\ell^* - \Phi_\ell(u_\ell) \| \stackrel{(35)}{\leq} q \| u_\ell^* - u_\ell \|.$$

Since this holds for all  $u_\ell \in \mathcal{X}_\ell^p$ , we also have

$$\| (I - \mathcal{S}_\ell^{\text{MG}})v_\ell \| \leq q \| v_\ell \| \quad \text{for all } v_\ell \in \mathcal{X}_\ell^p.$$

Applying Theorem 2 concludes the proof.  $\square$

## 5. OPTIMAL LINEAR AND SYMMETRIC MULTIGRID PRECONDITIONER

In this section, we formulate and analyze a linear and symmetric multigrid preconditioner  $\mathbf{B}_\ell^{\text{sMG}}$ . First, we linearize the multigrid of [IMPS24] by replacing the optimal step-sizes  $\lambda_{\ell'}$  with the constant  $\lambda = (d+1)^{-1}$  for all  $\ell' \in \{1, \dots, \ell\}$ , as already described in Remark 14. Second, we symmetrize the approach by adding suitable pre-smoothing steps to Algorithm C.

**5.1. Preconditioner and corresponding main result.** We build on the notation already introduced in Section 3. The additional superscript is introduced to distinguish algorithmically the pre-smoothing ( $\downarrow$ ) and post-smoothing ( $\uparrow$ ) in the V-cycle.

**Algorithm D (V-cycle of symmetric and linear multigrid method).** *Input:* Current approximation  $u_\ell \in \mathcal{X}_\ell^p$ , triangulations  $\{\mathcal{T}_{\ell'}\}_{\ell'=0}^\ell$ , and polynomial degree  $p \geq 1$ . Follow the steps (i)–(v):

- (i) **High-order correction:** For all  $z \in \mathcal{V}_\ell$ , compute  $\rho_{\ell,z}^\downarrow \in \mathcal{X}_{\ell,z}^p$  such that

$$\langle \rho_{\ell,z}^\downarrow, v_{\ell,z} \rangle = R_\ell(v_{\ell,z}) \quad \text{for all } v_{\ell,z} \in \mathcal{X}_{\ell,z}^p.$$

Define  $\rho_\ell^\downarrow := \sum_{z \in \mathcal{V}_\ell} \rho_{\ell,z}^\downarrow$  and  $\sigma_\ell^\downarrow := \lambda \rho_\ell^\downarrow = (d+1)^{-1} \rho_\ell^\downarrow$ .

- (ii) **Lowest-order correction:** For all intermediate levels  $\ell' = \ell - 1, \dots, 1$  and all  $z \in \mathcal{V}_{\ell'}^+$ , compute  $\rho_{\ell',z}^\downarrow \in \mathcal{X}_{\ell',z}^1$  such that

$$\langle \rho_{\ell',z}^\downarrow, v_{\ell',z} \rangle = R_{\ell'}(v_{\ell',z}) - \langle \sigma_{\ell'+1}^\downarrow, v_{\ell',z} \rangle \quad \text{for all } v_{\ell',z} \in \mathcal{X}_{\ell',z}^1.$$

Define  $\rho_{\ell'}^\downarrow := \sum_{z \in \mathcal{V}_{\ell'}^+} \rho_{\ell',z}^\downarrow$  and  $\sigma_{\ell'}^\downarrow := \sigma_{\ell'+1}^\downarrow + \lambda \rho_{\ell'}^\downarrow = \sigma_{\ell'+1}^\downarrow + (d+1)^{-1} \rho_{\ell'}^\downarrow$ .

- (iii) **Coarse level solve:** Compute  $\rho_0^\uparrow \in \mathcal{X}_0^1$  such that

$$\langle \rho_0^\uparrow, v_0 \rangle = R_\ell(v_0) - \langle \sigma_1^\downarrow, v_0 \rangle \quad \text{for all } v_0 \in \mathcal{X}_0^1. \quad (39)$$

Define  $\sigma_0^\uparrow := \sigma_1^\downarrow + \rho_0^\uparrow$ .

- (iv) **Lowest-order correction:** For all intermediate levels  $\ell' = 1, \dots, \ell - 1$  and all  $z \in \mathcal{V}_{\ell'}^+$ , compute  $\rho_{\ell',z}^\uparrow \in \mathcal{X}_{\ell',z}^1$  such that

$$\langle \rho_{\ell',z}^\uparrow, v_{\ell',z} \rangle = R_{\ell'}(v_{\ell',z}) - \langle \sigma_{\ell'-1}^\uparrow, v_{\ell',z} \rangle \quad \text{for all } v_{\ell',z} \in \mathcal{X}_{\ell',z}^1. \quad (40)$$

Define  $\rho_{\ell'}^\uparrow := \sum_{z \in \mathcal{V}_{\ell'}^+} \rho_{\ell',z}^\uparrow$  and  $\sigma_{\ell'}^\uparrow := \sigma_{\ell'-1}^\uparrow + \lambda \rho_{\ell'}^\uparrow = \sigma_{\ell'-1}^\uparrow + (d+1)^{-1} \rho_{\ell'}^\uparrow$ .

(v) **High-order correction:** For all  $z \in \mathcal{V}_\ell$ , compute  $\rho_{\ell,z}^\uparrow \in \mathcal{X}_{\ell,z}^p$  such that

$$\langle\langle \rho_{\ell,z}^\uparrow, v_{\ell,z} \rangle\rangle = R_\ell(v_{\ell,z}) - \langle\langle \sigma_{\ell-1}^\uparrow, v_{\ell,z} \rangle\rangle \quad \text{for all } v_{\ell,z} \in \mathcal{X}_{\ell,z}^p.$$

Define  $\rho_\ell^\uparrow := \sum_{z \in \mathcal{V}_\ell} \rho_{\ell,z}^\uparrow$  and  $\sigma_\ell := \sigma_\ell^\uparrow := \sigma_{\ell-1}^\uparrow + \lambda \rho_\ell^\uparrow = \sigma_{\ell-1}^\uparrow + (d+1)^{-1} \rho_\ell^\uparrow$ .

**Output:** Improved approximation  $\tilde{\Phi}_\ell(u_\ell) := u_\ell + \sigma_\ell$ .

**Remark 15 (Computational complexity of Algorithm D).** With the same arguments as in Remark 5, we obtain that the overall computational cost of Algorithm D is of order  $\mathcal{O}(\#\mathcal{T}_\ell)$ . Note that the calculation of the optimal step-sizes  $\lambda_{\ell'}$  is not applicable in Algorithm D.

Similarly to (18) in Section 4, we define the matrices  $\mathbf{B}_{\ell'}^\downarrow$  for  $\ell' \in \{1, \dots, \ell\}$  and  $\mathbf{B}_{\ell'}^\uparrow$  for  $\ell' \in \{0, \dots, \ell\}$  by

$$\begin{aligned} \mathbf{B}_{\ell'}^\downarrow &:= \mathbf{I}_{\ell'}^+(\mathbf{D}_{\ell'}^+)^{-1}(\mathbf{I}_{\ell'}^+)^T \prod_{i=\ell'+1}^{\ell} (\mathbf{I} - \lambda \mathbf{A}_\ell \mathbf{I}_i^+(\mathbf{D}_i^+)^{-1}(\mathbf{I}_i^+)^T), \\ \mathbf{B}_0^\uparrow &:= \mathbf{I}_0^+(\mathbf{A}_0)^{-1}(\mathbf{I}_0^+)^T \prod_{i=1}^{\ell} (\mathbf{I} - \lambda \mathbf{A}_\ell \mathbf{I}_i^+(\mathbf{D}_i^+)^{-1}(\mathbf{I}_i^+)^T) \quad \text{and} \\ \mathbf{B}_{\ell'}^\uparrow &:= \mathbf{I}_{\ell'}^+(\mathbf{D}_{\ell'}^+)^{-1}(\mathbf{I}_{\ell'}^+)^T \prod_{i=\ell'-1}^1 (\mathbf{I} - \lambda \mathbf{A}_\ell \mathbf{I}_i^+(\mathbf{D}_i^+)^{-1}(\mathbf{I}_i^+)^T) \prod_{i=0}^{\ell} (\mathbf{I} - \lambda \mathbf{A}_\ell \mathbf{I}_i^+(\mathbf{D}_i^+)^{-1}(\mathbf{I}_i^+)^T), \end{aligned}$$

where we set  $\mathbf{D}_0^+ := \lambda \mathbf{A}_0$  and  $(\mathbf{D}_\ell^+)^{-1} := \sum_{z \in \mathcal{V}_\ell} \mathbf{I}_{\ell,z}^p (\mathbf{A}_{\ell,z}^p)^{-1} (\mathbf{I}_{\ell,z}^p)^T$  for easier notation.

Let  $u_\ell$  be the current approximation of  $u_\ell^*$  and  $\mathbf{x}_\ell = \chi_\ell^p[u_\ell]$ ,  $\mathbf{x}_\ell^* = \chi_\ell^p[u_\ell^*] \in \mathbb{R}^{N_\ell^p}$ . Moreover, we set  $\mathbf{r}_\ell = \mathbf{b}_\ell - \mathbf{A}_\ell \mathbf{x}_\ell = \mathbf{A}_\ell (\mathbf{x}_\ell^* - \mathbf{x}_\ell)$ . Along the lines of (17)–(19) in Section 4.1, one also shows that the coefficient vectors  $\mathbf{s}_{\ell'}^\downarrow = \chi_\ell^p(\rho_{\ell'}^\downarrow)$ ,  $\mathbf{s}_{\ell'}^\uparrow = \chi_\ell^p(\rho_{\ell'}^\uparrow) \in \mathbb{R}^{N_\ell^p}$  are given by

$$\mathbf{s}_{\ell'}^\downarrow = \mathbf{B}_{\ell'}^\downarrow \mathbf{r}_\ell, \quad \mathbf{s}_0^\uparrow = \mathbf{B}_0^\uparrow \mathbf{r}_\ell \quad \text{and} \quad \mathbf{s}_{\ell'}^\uparrow = \mathbf{B}_{\ell'}^\uparrow \mathbf{r}_\ell.$$

Proceeding analogously as in Section 4, we obtain that the total error update  $\sigma_\ell$  satisfies

$$\chi_\ell^p(\sigma_\ell) = \mathbf{s} = \lambda \sum_{\ell'=1}^{\ell} \mathbf{s}_{\ell'}^\downarrow + \mathbf{s}_0^\uparrow + \lambda \sum_{\ell'=1}^{\ell} \mathbf{s}_{\ell'}^\uparrow = \left( \lambda \sum_{\ell'=1}^{\ell} \mathbf{B}_{\ell'}^\downarrow + \mathbf{B}_0^\uparrow + \lambda \sum_{\ell'=1}^{\ell} \mathbf{B}_{\ell'}^\uparrow \right) \mathbf{r}_\ell.$$

With these preparations, we define the linear and symmetric multigrid preconditioner  $\mathbf{B}_\ell^{\text{sMG}}$

$$\mathbf{B}_\ell^{\text{sMG}} := \lambda \sum_{\ell'=1}^{\ell} \mathbf{B}_{\ell'}^\downarrow + \mathbf{B}_0^\uparrow + \lambda \sum_{\ell'=1}^{\ell} \mathbf{B}_{\ell'}^\uparrow \in \mathbb{R}^{N_\ell^p \times N_\ell^p}. \quad (41)$$

Lemma 9 yields the representation

$$\mathbf{B}_\ell^{\text{sMG}} \mathbf{A}_\ell = \mathbf{I} - \prod_{\ell'=\ell}^1 (\mathbf{I} - \lambda \mathbf{S}_{\ell'}) (\mathbf{I} - \mathbf{S}_0) \prod_{\ell'=1}^{\ell} (\mathbf{I} - \lambda \mathbf{S}_{\ell'}). \quad (42)$$

The subsequent theorem shows that  $\text{cond}_{\mathbf{A}_\ell}(\mathbf{B}_\ell^{\text{sMG}} \mathbf{A}_\ell)$  is bounded independently of  $h$  and  $p$ . Thus, Proposition 1 yields that PCG with preconditioner  $\mathbf{B}_\ell^{\text{sMG}}$  contracts  $h$ - and  $p$ -robustly. The proof is postponed to Section 5.2. With Remark 15, it follows that  $\mathbf{B}_\ell^{\text{sMG}}$  is indeed optimal.

**Theorem 16 (PCG with linear and symmetric MG preconditioner).** The symmetric multigrid preconditioner  $\mathbf{B}_\ell^{\text{sMG}}$  defined in (41) is SPD and optimal, i.e., there holds

$$\text{cond}_2((\mathbf{B}_\ell^{\text{sMG}})^{1/2} \mathbf{A}_\ell (\mathbf{B}_\ell^{\text{sMG}})^{1/2}) = \text{cond}_{\mathbf{A}_\ell}(\mathbf{B}_\ell^{\text{sMG}} \mathbf{A}_\ell) \leq \frac{1+q^2}{1-q^2}, \quad (43)$$

with the contraction factor  $0 < q < 1$  from Proposition 12. For the iterates  $\mathbf{x}_\ell^k$  generated by PCG employing the preconditioner (41) with initial guess  $\mathbf{x}_\ell^0$ , there holds

$$|\mathbf{x}_\ell^* - \mathbf{x}_\ell^{k+1}|_{\mathbf{A}_\ell} \leq \left(1 - \frac{1 - q^2}{1 + q^2}\right)^{1/2} |\mathbf{x}_\ell^* - \mathbf{x}_\ell^k|_{\mathbf{A}_\ell} \quad \text{for all } k \in \mathbb{N}_0. \quad (44)$$

For  $u_\ell^* = \sum_{j=1}^{N_\ell^p} (\mathbf{x}_\ell^*)_j \varphi_{\ell,j}^p$  and  $u_\ell^k = \sum_{j=1}^{N_\ell^p} (\mathbf{x}_\ell^k)_j \varphi_{\ell,j}^p$ , this directly translates to

$$\|u_\ell^* - u_\ell^{k+1}\| \leq \left(1 - \frac{1 - q^2}{1 + q^2}\right)^{1/2} \|u_\ell^* - u_\ell^k\| \quad \text{for all } k \in \mathbb{N}_0.$$

---

The following lemma provides a connection between the notion of symmetrizing the multigrid via pre-smoothing and symmetrizing matrices via their transpose.

**Lemma 17 (Symmetrized MG is a symmetric preconditioner).** *It holds that  $\mathbf{B}_\ell^{\text{smMG}} \mathbf{A}_\ell$  is symmetric with respect to  $(\cdot, \cdot)_{\mathbf{A}_\ell}$  and hence  $\mathbf{B}_\ell^{\text{smMG}}$  is a symmetric matrix. Let us define the error propagation matrix*

$$\mathbf{E}_\ell := \prod_{\ell'=\ell}^1 (\mathbf{I} - \lambda \mathbf{S}_{\ell'}) (\mathbf{I} - \mathbf{S}_0) = \mathbf{I} - \mathbf{B}_\ell^{\text{nsMG}} \mathbf{A}_\ell, \quad (45)$$

where  $\mathbf{B}_\ell^{\text{nsMG}}$  is defined in Remark 14. Then, it follows that

$$\mathbf{B}_\ell^{\text{smMG}} \mathbf{A}_\ell = \mathbf{I} - \mathbf{E}_\ell \mathbf{E}_\ell^{T_{\mathbf{A}_\ell}}. \quad (46)$$

where  $T_{\mathbf{A}_\ell}$  denotes the transpose with respect to the inner product  $(\cdot, \cdot)_{\mathbf{A}_\ell}$ .

---

**Proof.** Let  $\mathbf{A}$  be an SPD matrix and  $\mathbf{N}$  and  $\mathbf{M}$  be any two matrices of the same size as  $\mathbf{A}$ . Then, it holds that  $(\mathbf{N}\mathbf{M})^{T_{\mathbf{A}}} = \mathbf{M}^{T_{\mathbf{A}}}\mathbf{N}^{T_{\mathbf{A}}}$ . Lemma 7 yields that the levelwise matrices  $\mathbf{I} - \lambda \mathbf{S}_{\ell'}$  and  $\mathbf{I} - \mathbf{S}_0$  are symmetric with respect to  $(\cdot, \cdot)_{\mathbf{A}_\ell}$ . With (42), we conclude that  $\mathbf{B}_\ell^{\text{smMG}} \mathbf{A}_\ell$  is symmetric with respect to  $(\cdot, \cdot)_{\mathbf{A}_\ell}$ . Due to the symmetry of the levelwise matrices, we can write

$$\mathbf{E}_\ell \mathbf{E}_\ell^{T_{\mathbf{A}_\ell}} = \prod_{\ell'=\ell}^1 (\mathbf{I} - \lambda \mathbf{S}_{\ell'}) (\mathbf{I} - \mathbf{S}_0) (\mathbf{I} - \mathbf{S}_0) \prod_{\ell'=1}^{\ell} (\mathbf{I} - \lambda \mathbf{S}_{\ell'}).$$

Since  $\mathbf{S}_0$  is the matrix representation of the lowest-order Galerkin projection  $\mathcal{P}_0^1$ , we get

$$(\mathbf{I} - \mathbf{S}_0)(\mathbf{I} - \mathbf{S}_0) = \mathbf{I} - \mathbf{S}_0$$

and (46) follows from (42). This concludes the proof.  $\square$

**5.2. Proof of Theorem 16.** For any SPD matrix  $\mathbf{A}$  and any square matrix  $\mathbf{M}$  satisfy

$$(\mathbf{x}, \mathbf{M}^T \mathbf{y})_2 = (\mathbf{M}\mathbf{x}, \mathbf{y})_2 = (\mathbf{M}\mathbf{x}, \mathbf{A}^{-1}\mathbf{y})_{\mathbf{A}} = (\mathbf{x}, \mathbf{M}^{T_{\mathbf{A}}}\mathbf{A}^{-1}\mathbf{y})_{\mathbf{A}} = (\mathbf{x}, \mathbf{A}\mathbf{M}^{T_{\mathbf{A}}}\mathbf{A}^{-1}\mathbf{y})_2 \quad (47)$$

for all  $\mathbf{x}, \mathbf{y} \in \mathbb{R}^{N_\ell}$  and hence  $\mathbf{M}^T = \mathbf{A}\mathbf{M}^{T_{\mathbf{A}}}\mathbf{A}^{-1}$ . Furthermore, we have

$$|\mathbf{M}|_{\mathbf{A}} = \max_{|\mathbf{x}|_{\mathbf{A}}=1} |\mathbf{M}\mathbf{x}|_{\mathbf{A}} = \max_{|\mathbf{y}|_2=1} |\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{-1/2}\mathbf{y}|_2 = |\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{-1/2}|_2. \quad (48)$$

Recall the familiar identity

$$|\mathbf{M}|_2^2 = |\mathbf{M}\mathbf{M}^T|_2 = |\mathbf{M}^T|_2^2. \quad (49)$$

Altogether, it follows that

$$|\mathbf{M}|_{\mathbf{A}}^2 \stackrel{(48)}{=} |\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{-1/2}|_2^2 \stackrel{(49)}{=} |\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{-1}\mathbf{M}^T\mathbf{A}^{1/2}|_2^2 \stackrel{(47)}{=} |\mathbf{A}^{1/2}\mathbf{M}\mathbf{M}^{T_{\mathbf{A}}}\mathbf{A}^{-1/2}|_2^2 \stackrel{(48)}{=} |\mathbf{M}\mathbf{M}^{T_{\mathbf{A}}}|_{\mathbf{A}}$$

as well as

$$|\mathbf{M}|_{\mathbf{A}} \stackrel{(48)}{=} |\mathbf{A}^{1/2}\mathbf{M}\mathbf{A}^{-1/2}|_2 \stackrel{(49)}{=} |\mathbf{A}^{1/2}\mathbf{M}^T\mathbf{A}^{-1/2}|_2 = |\mathbf{M}^{T_{\mathbf{A}}}|_{\mathbf{A}}.$$

This proves

$$|\mathbf{M}|_{\mathbf{A}}^2 = |\mathbf{M}\mathbf{M}^{T_{\mathbf{A}}}|_{\mathbf{A}} = |\mathbf{M}^{T_{\mathbf{A}}}|_{\mathbf{A}}^2. \quad (50)$$

Symmetry of  $\mathbf{B}_\ell^{\text{sMG}}$  follows from Lemma 17. To show that  $\mathbf{B}_\ell^{\text{sMG}}$  is positive definite, note that (46), (50), and (38) imply that

$$\begin{aligned} (\mathbf{B}_\ell^{\text{sMG}} \mathbf{A}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell} &\stackrel{(46)}{=} ((\mathbf{I} - \mathbf{E}_\ell \mathbf{E}_\ell^T \mathbf{A}_\ell) \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell} \stackrel{(50)}{=} (\mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell} - |\mathbf{E}_\ell \mathbf{x}_\ell|_{\mathbf{A}_\ell}^2 \\ &\stackrel{(38)}{\geq} (1 - q) |\mathbf{x}_\ell|_{\mathbf{A}_\ell}^2 > 0 \quad \text{for all } \mathbf{x}_\ell \in \mathbb{R}^{N_\ell^p} \setminus \{0\}. \end{aligned}$$

Substituting  $\mathbf{y}_\ell = \mathbf{A}_\ell \mathbf{x}_\ell$  yields  $(\mathbf{B}_\ell^{\text{sMG}} \mathbf{y}_\ell, \mathbf{y}_\ell)_2 > 0$  for all  $\mathbf{y}_\ell \neq 0$ . Hence,  $\mathbf{B}_\ell^{\text{sMG}}$  is SPD. Using identity (46), the Neumann series, identity (50), and (38), we get

$$\begin{aligned} \text{cond}_{\mathbf{A}_\ell}(\mathbf{B}_\ell^{\text{sMG}} \mathbf{A}_\ell) &= |\mathbf{B}_\ell^{\text{sMG}} \mathbf{A}_\ell|_{\mathbf{A}_\ell} |(\mathbf{B}_\ell^{\text{sMG}} \mathbf{A}_\ell)^{-1}|_{\mathbf{A}_\ell} \stackrel{(46)}{=} |\mathbf{I} - \mathbf{E}_\ell \mathbf{E}_\ell^T \mathbf{A}_\ell|_{\mathbf{A}_\ell} |(\mathbf{I} - \mathbf{E}_\ell \mathbf{E}_\ell^T \mathbf{A}_\ell)^{-1}|_{\mathbf{A}_\ell} \\ &\leq \frac{|1 - \mathbf{E}_\ell \mathbf{E}_\ell^T \mathbf{A}_\ell|_{\mathbf{A}_\ell}}{1 - |\mathbf{E}_\ell \mathbf{E}_\ell^T \mathbf{A}_\ell|_{\mathbf{A}_\ell}} \leq \frac{1 + |\mathbf{E}_\ell \mathbf{E}_\ell^T \mathbf{A}_\ell|_{\mathbf{A}_\ell}}{1 - |\mathbf{E}_\ell \mathbf{E}_\ell^T \mathbf{A}_\ell|_{\mathbf{A}_\ell}} \stackrel{(50)}{=} \frac{1 + |\mathbf{E}_\ell|_{\mathbf{A}_\ell}^2}{1 - |\mathbf{E}_\ell|_{\mathbf{A}_\ell}^2} \stackrel{(38)}{\leq} \frac{1 + q^2}{1 - q^2}. \end{aligned}$$

This proves (43). Contraction (44) follows from Proposition 1 and thus concludes the proof.  $\square$

## 6. OPTIMAL ADDITIVE SCHWARZ PRECONDITIONER

In this section, we consider an additive Schwarz preconditioner  $\mathbf{B}_\ell^{\text{AS}}$  induced by the levelwise matrices  $\mathbf{S}_{\ell'}$  from (17).

**6.1. Preconditioner and corresponding main result.** Define the additive Schwarz preconditioner  $\mathbf{B}_\ell^{\text{AS}}$  as

$$\mathbf{B}_\ell^{\text{AS}} := \mathbf{I}_0^+ \mathbf{A}_0^{-1} (\mathbf{I}_0^+)^T + \sum_{\ell'=1}^{\ell-1} \mathbf{I}_{\ell'}^+ (\mathbf{D}_{\ell'}^+)^{-1} (\mathbf{I}_{\ell'}^+)^T + \sum_{z \in \mathcal{V}_\ell} \mathbf{I}_{\ell,z}^p (\mathbf{A}_{\ell,z}^p)^{-1} (\mathbf{I}_{\ell,z}^p)^T. \quad (51)$$

From its definition (51) and the definition (17) of the matrices  $\mathbf{S}_{\ell'}$ , it follows that  $\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell = \sum_{\ell'=0}^{\ell} \mathbf{S}_{\ell'}$ . We define the additive Schwarz operator  $\mathcal{S}_\ell^{\text{AS}}$  via

$$\mathcal{S}_\ell^{\text{AS}} := \sum_{\ell'=0}^{\ell} \mathcal{S}_{\ell'}. \quad (52)$$

---

**Remark 18 (Computational complexity of additive Schwarz preconditioner).** *Arguing as in Remark 5, we conclude that the overall computational cost of a single application of the additive Schwarz preconditioner  $\mathbf{B}_\ell^{\text{AS}}$  is  $\mathcal{O}(\#\mathcal{T}_\ell)$ . Moreover, since this is an additive method, in contrast to the multiplicative structure of  $\mathbf{B}_\ell^{\text{MG}}$  and  $\mathbf{B}_\ell^{\text{sMG}}$ , its application can be efficiently parallelized, thereby reducing the overall computational time.*

---

The subsequent theorem shows that  $\text{cond}_{\mathbf{A}_\ell}(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell)$  is bounded independently of  $h$  and  $p$ . Thus, Proposition 1 yields that PCG with preconditioner  $\mathbf{B}_\ell^{\text{AS}}$  contracts  $h$ - and  $p$ -robustly. The proof is postponed to Section 6.4. Together with Remark 18, it follows that  $\mathbf{B}_\ell^{\text{AS}}$  is optimal.

---

**Theorem 19 (PCG with additive Schwarz preconditioner).** *The additive Schwarz preconditioner  $\mathbf{B}_\ell^{\text{AS}}$  defined in (51) is an SPD matrix. For the minimal and maximal eigenvalues of  $\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell$  it holds that*

$$c \leq \lambda_{\min}(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell) \quad \text{and} \quad \lambda_{\max}(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell) \leq C,$$

where  $c, C > 0$  are independent of  $h, p$ , and only depend locally on the diffusion contrast  $\mathbf{K}$ . In particular, this yields

$$\text{cond}_2((\mathbf{B}_\ell^{\text{AS}})^{1/2} \mathbf{A}_\ell (\mathbf{B}_\ell^{\text{AS}})^{1/2}) = \text{cond}_{\mathbf{A}_\ell}(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell) \leq C/c,$$

For the iterates  $\mathbf{x}_\ell^k$  generated by PCG with preconditioner  $\mathbf{B}_\ell^{\text{AS}}$  and initial guess  $\mathbf{x}_\ell^0$ , there holds

$$|\mathbf{x}_\ell^* - \mathbf{x}_\ell^{k+1}|_{\mathbf{A}_\ell} \leq \left(1 - \frac{c}{C}\right)^{1/2} |\mathbf{x}_\ell^* - \mathbf{x}_\ell^k|_{\mathbf{A}_\ell} \quad \text{for all } k \in \mathbb{N}_0. \quad (53)$$

For  $u_\ell^* = \sum_{j=1}^{N_\ell^p} (\mathbf{x}_\ell^*)_j \varphi_{\ell,j}^p$  and  $u_\ell^k = \sum_{j=1}^{N_\ell^p} (\mathbf{x}_\ell^k)_j \varphi_{\ell,j}^p$ , this directly translates to

$$\|u_\ell^* - u_\ell^{k+1}\| \leq \left(1 - \frac{c}{C}\right)^{1/2} \|u_\ell^* - u_\ell^k\| \quad \text{for all } k \in \mathbb{N}_0.$$

The following lemma provides a translation between functional and matrix notation for the additive Schwarz method, analogous to Lemma 11 for the geometric multigrid method. The proof follows directly from Lemma 7 and is thus omitted.

**Lemma 20 (Additive Schwarz functional/matrix representation).** *Let  $\mathbf{B}_\ell^{\text{AS}}$  and  $\mathcal{S}_\ell^{\text{AS}}$  denote the preconditioner and operator from (51)–(52). Then, there holds*

$$\langle\langle \mathcal{S}_\ell^{\text{AS}} v_\ell, w_\ell \rangle\rangle = (\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell \mathbf{x}_\ell, \mathbf{y}_\ell)_{\mathbf{A}_\ell} \quad \text{for all } v_\ell, w_\ell \in \mathcal{X}_\ell^p \text{ and } \mathbf{x}_\ell = \chi_\ell^p(v_\ell), \mathbf{y}_\ell = \chi_\ell^p(w_\ell). \quad \square \quad (54)$$

Similar to the multiplicative case, we show properties of the operator  $\mathcal{S}_\ell^{\text{AS}}$  and translate them via (54) to prove Theorem 19. For this purpose, we need some auxiliary results.

**6.2. Auxiliary results.** First, recall Lions' lemma [Lio88].

**Lemma 21 (Lions' lemma).** *Let  $\mathcal{V}$  be a finite-dimensional Hilbert space with scalar product  $\langle\langle \cdot, \cdot \rangle\rangle_{\mathcal{V}}$  and corresponding norm  $\|\cdot\|_{\mathcal{V}}$ . Suppose the decomposition  $\mathcal{V} = \sum_{j=0}^m \mathcal{V}_j$  with corresponding orthogonal projections  $\tilde{\mathcal{S}}_j : \mathcal{V} \rightarrow \mathcal{V}_j$  defined by*

$$\langle\langle \tilde{\mathcal{S}}_j v, w_j \rangle\rangle_{\mathcal{V}} = \langle\langle v, w_j \rangle\rangle_{\mathcal{V}} \quad \text{for all } v \in \mathcal{V} \text{ and all } w_j \in \mathcal{V}_j. \quad (55)$$

If there exists a stable decomposition  $v = \sum_{j=0}^m v_j$  with  $v_j \in \mathcal{V}_j$  for every  $v \in \mathcal{V}$  such that

$$\sum_{j=0}^m \|v_j\|_{\mathcal{V}}^2 \leq C \|v\|_{\mathcal{V}}^2, \quad (56)$$

then the operator  $\tilde{\mathcal{S}} := \sum_{j=0}^m \tilde{\mathcal{S}}_j$  satisfies, even with the same constant  $C$ ,

$$\|v\|_{\mathcal{V}}^2 \leq C \langle\langle \tilde{\mathcal{S}} v, v \rangle\rangle_{\mathcal{V}} \quad \text{for all } v \in \mathcal{V}. \quad \square \quad (57)$$

We also need a strengthened Cauchy–Schwarz inequality. This requires the use of generation constraints. For every  $\ell' \in \{0, \dots, \ell\}$  and  $z \in \mathcal{V}_{\ell'}$ , define the generation  $g_{\ell',z}$  of  $\mathcal{T}_{\ell'}(z)$  by

$$g_{\ell',z} := \max_{T \in \mathcal{T}_{\ell'}(z)} \text{level}(T) := \max_{T \in \mathcal{T}_{\ell'}(z)} \log_2(|T_0|/|T|) \in \mathbb{N}_0, \quad (58)$$

where  $T_0 \in \mathcal{T}_0$  is the unique ancestor of  $T$ , i.e.,  $T \subseteq T_0$ .

Let  $M := \max_{T \in \mathcal{T}_\ell} \text{level}(T)$ . We denote by  $\{\hat{\mathcal{T}}_j\}_{j=0}^M$  the sequence of uniformly refined triangulations that satisfy  $\hat{\mathcal{T}}_{j+1} := \text{REFINE}(\hat{\mathcal{T}}_j, \hat{\mathcal{T}}_j)$  and  $\hat{\mathcal{T}}_0 := \mathcal{T}_0$ . Since  $\mathcal{T}_0$  is admissible, each element  $T \in \hat{\mathcal{T}}_j$  satisfies  $\text{level}(T) = j$  and hence is only bisected once during uniform refinement; see [Ste08, Theorem 4.3]. Moreover, denote by  $\hat{h}_j := \max_{T \in \hat{\mathcal{T}}_j} |T|^{1/d}$  the mesh-size of the uniform triangulation  $\hat{\mathcal{T}}_j$ . Importantly, there holds  $\hat{h}_j \simeq h_T$  for all  $T \in \hat{\mathcal{T}}_j$  and all  $j \in \mathbb{N}_0$  and the hidden constants depend only on  $\mathcal{T}_0$ . Every object associated with uniform meshes will be indicated with a hat, e.g.,  $\hat{\mathcal{X}}_j^1$  is the lowest-order FEM space induced by  $\hat{\mathcal{T}}_j$ .

**Lemma 22.** *Let  $\ell' \in \{0, \dots, \ell\}$  and  $z \in \mathcal{V}_{\ell'}$ . Define  $r_{\ell',z} := \min_{T \in \mathcal{T}_{\ell'}(z)} \text{level}(T)$ . Then, there exist integers  $C_{\text{span}}, n \in \mathbb{N}$  depending only on  $\gamma$ -shape regularity such that  $m := g_{\ell',z} \leq r_{\ell',z} + C_{\text{span}}$  and  $\omega_{\ell'}(z) \subseteq \hat{\omega}_m^n(z)$ .*

**Proof.** The proof is split into two steps.

**Step 1:** Let  $z \in \mathcal{V}_{\ell'}$ . Then, there exists elements  $T, T' \in \mathcal{T}_{\ell'}(z)$  such that  $g_{\ell',z} = \text{level}(T)$  and  $r_{\ell',z} = \text{level}(T')$ . Moreover, we have  $|T| \simeq |T'|$  due to  $\gamma$ -shape regularity. Denote by  $T_0, T'_0 \in \mathcal{T}_0$  the unique ancestors of  $T$  and  $T'$ , respectively. With the quasi-uniformity of  $\mathcal{T}_0$ , it follows that

$$\text{level}(T) = \log_2(|T_0|/|T|) \leq \log_2(C |T'_0|/|T'|) = \log_2(C) + \text{level}(T'),$$

where  $C > 0$  depends only  $\gamma$ -shape regularity. For  $C_{\text{span}} := \lceil \log_2(C) \rceil$ , we get  $m = g_{\ell',z} \leq r_{\ell',z} + C_{\text{span}}$ .

**Step 2:** By definition of  $r_{\ell',z}$ , we have  $\omega_{\ell'}(z) \subseteq \widehat{\omega}_{r_{\ell',z}}(z)$ . Every element  $T \in \widehat{\mathcal{T}}_{r_{\ell',z}}$  can be decomposed into elements  $T_j \in \widehat{\mathcal{T}}_{r_{\ell',z} + C_{\text{span}}}$  with  $j = 1, \dots, 2^{C_{\text{span}}}$ , i.e.,  $T = \bigcup_{j=1}^{2^{C_{\text{span}}}} T_j$ . Thus, for every  $T \subseteq \widehat{\omega}_{r_{\ell',z}}(z)$  we also have  $T \subseteq \widehat{\omega}_{r_{\ell',z} + C_{\text{span}}}^{2^{C_{\text{span}}}}(z)$ . Hence, there exists an integer  $n \in \mathbb{N}$  with  $n \leq 2^{C_{\text{span}}}$  such that  $\widehat{\omega}_{r_{\ell',z}}(z) \subseteq \widehat{\omega}_{r_{\ell',z} + C_{\text{span}}}^n(z)$ . Finally, Step 1 yields

$$\omega_{\ell'}(z) \subseteq \widehat{\omega}_{r_{\ell',z}}(z) \subseteq \widehat{\omega}_{r_{\ell',z} + C_{\text{span}}}^n(z) \subseteq \widehat{\omega}_m^n(z).$$

This concludes the proof.  $\square$

Let  $\mathcal{T}$  be a refinement of the initial mesh  $\mathcal{T}_0$  and  $\mathcal{M} \subseteq \mathcal{T}$ . For  $\omega := \text{int}(\bigcup_{T \in \mathcal{M}} T)$ , we define

$$C[\omega] := \max\left\{\max_{T \in \mathcal{M}} \|\text{div}(\mathbf{K})\|_{L^\infty(T)}, \sup_{y \in \omega} \lambda_{\max}(\mathbf{K}(y))\right\}. \quad (59)$$

From [IMPS24, Lemma 5.6] and [Hil25], the following strengthened Cauchy–Schwarz inequality on uniform meshes is already known.

---

**Lemma 23 (Strengthened Cauchy–Schwarz inequality).** *Let  $0 \leq i \leq j$ ,  $\widehat{\mathcal{M}}_i \subseteq \widehat{\mathcal{T}}_i$ , and  $\omega_i := \text{int}(\bigcup_{T \in \widehat{\mathcal{M}}_i} T)$ . Then, it holds that*

$$\langle\langle \widehat{u}_i, \widehat{v}_j \rangle\rangle_{\widehat{\omega}_i} \leq C_{\text{SCS}} C[\widehat{\omega}_i] \delta^{j-i} \widehat{h}_j^{-1} \|\nabla \widehat{u}_i\|_{\widehat{\omega}_i} \|\widehat{v}_j\|_{\widehat{\omega}_i} \quad \text{for all } \widehat{u}_i \in \widehat{\mathcal{X}}_i^1 \text{ and } \widehat{v}_j \in \widehat{\mathcal{X}}_j^1, \quad (60)$$

where  $\delta = 2^{-1/(2d)}$ . The constant  $C_{\text{SCS}}$  depends only on  $\Omega$ ,  $d$ ,  $\mathcal{T}_0$ , and  $\gamma$ -shape regularity.  $\square$

---

For  $0 \leq m \leq M$ , we define the operator  $\mathcal{G}_{\ell,m}^1 : \mathcal{X}_\ell^p \rightarrow \mathcal{X}_\ell^1$  by

$$\mathcal{G}_{\ell,m}^1 := \sum_{\ell'=1}^{\ell-1} \sum_{\substack{z \in \mathcal{V}_{\ell'}^+ \\ g_{\ell',z} = m}} \mathcal{P}_{\ell',z}^1. \quad (61)$$

It immediately follows that

$$\mathcal{S}_\ell^{\text{AS}} = \mathcal{P}_0^1 + \sum_{m=0}^M \mathcal{G}_{\ell,m}^1 + \sum_{z \in \mathcal{V}_\ell} \mathcal{P}_{\ell,z}^p. \quad (62)$$

We show a strengthened Cauchy–Schwarz inequality for the operators  $\mathcal{G}_{\ell,m}^1$ .

---

**Lemma 24 (Strengthened Cauchy–Schwarz inequality for  $\mathcal{G}_{\ell,m}^1$ ).** *For  $0 \leq k \leq m \leq M$ , it holds that*

$$\langle\langle \widehat{v}_k, \mathcal{G}_{\ell,m}^1 \widehat{w}_k \rangle\rangle \leq C_{\text{loc}}^{(1)} C_{\text{SCS}} \delta^{m-k} \|\widehat{v}_k\| \|\widehat{w}_k\| \quad \text{for all } \widehat{v}_k, \widehat{w}_k \in \widehat{\mathcal{X}}_k^1, \quad (63)$$

where the constant  $C_{\text{loc}}^{(1)}$  is defined in (36) and the constant  $C_{\text{SCS}} > 0$  depends only on  $\Omega$ ,  $d$ ,  $\mathcal{T}_0$ , and  $\gamma$ -shape regularity.

---

**Proof.** Let  $0 \leq k \leq m \leq M$  and  $\widehat{v}_k, \widehat{w}_k \in \widehat{\mathcal{X}}_k^1$ . The proof is split into two steps.

**Step 1:** Let  $\ell' \in \{1, \dots, \ell-1\}$  and  $z \in \mathcal{V}_{\ell'}^+$  with  $g_{\ell',z} = m$ . Recall  $C[\cdot]$  from (59). Since  $\mathcal{P}_{\ell',z}^1 \widehat{w}_k \in \widehat{\mathcal{X}}_{m,z}^1$ , the strengthened Cauchy–Schwarz inequality (60) implies

$$\langle\langle \widehat{v}_k, \mathcal{P}_{\ell',z}^1 \widehat{w}_k \rangle\rangle_T \stackrel{(60)}{\lesssim} C[T] \delta^{m-k} \widehat{h}_m^{-1} \|\nabla \widehat{v}_k\|_T \|\mathcal{P}_{\ell',z}^1 \widehat{w}_k\|_T \quad \text{for all } T \in \widehat{\mathcal{T}}_k.$$

As  $\text{supp } \mathcal{P}_{\ell',z}^1 \widehat{w}_k \subseteq \overline{\omega_{\ell'}(z)}$ , we only need to consider  $T \in \widehat{\mathcal{T}}_k$  with  $|T \cap \overline{\omega_{\ell'}(z)}| > 0$ . For these elements, we can find a vertex  $z_0 \in \mathcal{V}_0$  depending only on  $z$  such that  $T \cup \omega_{\ell'}(z) \subseteq \overline{\omega_0(z_0)}$ . Using

the local norm equivalence, we obtain

$$\|\nabla \widehat{v}_k\|_T \leq \left( \inf_{y \in T} \lambda_{\min}(\mathbf{K}(y)) \right)^{1/2} \|\widehat{v}_k\|_T.$$

Summation over  $T \in \widehat{\mathcal{T}}_k$ , the existence of  $z_0$ , and the discrete Cauchy-Schwarz inequality yield

$$\begin{aligned} \langle \widehat{v}_k, \mathcal{P}_{\ell', z}^1 \widehat{w}_k \rangle &\lesssim C[\omega_0(z_0)] \left( \inf_{y \in \omega_0(z_0)} \lambda_{\min}(\mathbf{K}(y)) \right)^{1/2} \delta^{m-k} \widehat{h}_m^{-1} \sum_{T \in \widehat{\mathcal{T}}_k} \|\widehat{v}_k\|_T \|\mathcal{P}_{\ell', z}^1 \widehat{w}_k\|_T \\ &\leq C[\omega_0(z_0)] \left( \inf_{y \in \omega_0(z_0)} \lambda_{\min}(\mathbf{K}(y)) \right)^{1/2} \delta^{m-k} \widehat{h}_m^{-1} \|\widehat{v}_k\| \|\mathcal{P}_{\ell', z}^1 \widehat{w}_k\|. \end{aligned}$$

The generation constraint  $g_{\ell', z} = m$  and quasi-uniformity lead to  $\widehat{h}_m \simeq h_{\ell', z}$ . Due to the local support of  $\mathcal{P}_{\ell', z}^1$ , the Poincaré inequality, and the local norm equivalence, we obtain

$$\widehat{h}_m^{-1} \|\mathcal{P}_{\ell', z}^1 \widehat{w}_k\| \simeq \widehat{h}_{\ell', z}^{-1} \|\mathcal{P}_{\ell', z}^1 \widehat{w}_k\|_{\omega_{\ell'}(z)} \lesssim \|\nabla \mathcal{P}_{\ell', z}^1 \widehat{w}_k\| \leq \left( \inf_{y \in \omega_0(z_0)} \lambda_{\min}(\mathbf{K}(y)) \right)^{1/2} \|\mathcal{P}_{\ell', z}^1 \widehat{w}_k\|$$

and thus, with  $C_{\text{loc}}^{(1)}$  from (36),

$$\langle \widehat{v}_k, \mathcal{P}_{\ell', z}^1 \widehat{w}_k \rangle \lesssim C_{\text{loc}}^{(1)} \delta^{m-k} \|\widehat{v}_k\| \|\mathcal{P}_{\ell', z}^1 \widehat{w}_k\|. \quad (64)$$

**Step 2:** Based on (64) and the definition (61) of  $\mathcal{G}_{\ell, m}^1$ , it only remains to prove that

$$\sum_{\ell'=1}^{\ell-1} \sum_{\substack{z \in \mathcal{V}_{\ell'}^+ \\ g_{\ell', z} = m}} \|\mathcal{P}_{\ell', z}^1 \widehat{w}_k\| \lesssim \|\widehat{w}_k\|.$$

The definition (21) of  $\mathcal{P}_{\ell', z}^1$  implies  $\mathcal{P}_{\ell', z}^1 v = \frac{\langle v, \varphi_{\ell', z}^1 \rangle}{\|\varphi_{\ell', z}^1\|^2} \varphi_{\ell', z}^1$ . Since  $g_{\ell', z} = m$ , Lemma 22 yields

$$\|\mathcal{P}_{\ell', z}^1 \widehat{w}_k\| = \frac{|\langle \widehat{w}_k, \varphi_{\ell', z}^1 \rangle|}{\|\varphi_{\ell', z}^1\|} \leq \|\widehat{w}_k\|_{\omega_{\ell'}(z)} \leq \|\widehat{w}_k\|_{\widehat{\omega}_m^n(z)}. \quad (65)$$

Let  $z \in \mathcal{V}_\ell$  and  $0 \leq j \leq M$ . To keep track of the levels, where the patch associated to the vertex  $z$  has been modified in the refinement and remains of generation  $j$ , we define

$$\mathcal{L}_{\ell, \bar{\ell}}(z, j) := \{\ell' \in \{\ell, \dots, \bar{\ell}\} : z \in \mathcal{V}_{\ell'}^+ \text{ and } g_{\ell', z} = j\} \quad \text{for all } 0 \leq \ell \leq \bar{\ell} \leq \ell.$$

According to [WC06, Lemma 3.1], there exists a constant  $C_{\text{lev}} > 0$  depending only on  $\gamma$ -shape regularity such that

$$\max_{\substack{z \in \mathcal{V}_\ell \\ 0 \leq j \leq M}} \#(\mathcal{L}_{0, \ell}(z, j)) \leq C_{\text{lev}} < \infty. \quad (66)$$

Moreover, it holds that

$$\begin{aligned} \{(\ell', z) \in \mathbb{N}_0 \times \mathcal{V}_\ell : \ell' \in \{\ell, \dots, \bar{\ell}\}, z \in \mathcal{V}_{\ell'}^+ \text{ with } g_{\ell', z} = j\} \\ = \{(\ell', z) \in \mathbb{N}_0 \times \mathcal{V}_\ell : z \in \widehat{\mathcal{V}}_j, \ell' \in \mathcal{L}_{\ell, \bar{\ell}}(z, j)\}. \end{aligned} \quad (67)$$

With finite patch overlap, this leads to

$$\begin{aligned} \sum_{\ell'=1}^{\ell-1} \sum_{\substack{z \in \mathcal{V}_{\ell'}^+ \\ g_{\ell', z} = m}} \|\mathcal{P}_{\ell', z}^1 \widehat{w}_k\| &\stackrel{(65)}{\leq} \sum_{\ell'=1}^{\ell-1} \sum_{\substack{z \in \mathcal{V}_{\ell'}^+ \\ g_{\ell', z} = m}} \|\widehat{w}_k\|_{\widehat{\omega}_m^n(z)} \stackrel{(67)}{=} \sum_{z \in \widehat{\mathcal{V}}_m} \sum_{\ell' \in \mathcal{L}_{1, \ell-1}(z, m)} \|\widehat{w}_k\|_{\widehat{\omega}_m^n(z)} \\ &\stackrel{(66)}{\leq} C_{\text{lev}} \sum_{z \in \widehat{\mathcal{V}}_m} \|\widehat{w}_k\|_{\widehat{\omega}_m^n(z)} \lesssim \|\widehat{w}_k\|. \end{aligned}$$

Overall, we hence obtain

$$\langle \widehat{v}_k, \mathcal{G}_{\ell, m}^1 \widehat{w}_k \rangle \lesssim \delta^{m-k} \|\widehat{v}_k\| \|\widehat{w}_k\|.$$

This concludes the proof.  $\square$

**6.3. Additive Schwarz operator.** We require the following proposition on the additive Schwarz operator (52) to finally show Theorem 19 in the next section. For the proof of Proposition 25, we adapt [FFPS17] for the boundary element setting with energy space  $\tilde{H}^{1/2}(\Gamma)$  to our setting with energy space  $H_0^1(\Omega)$ .

---

**Proposition 25 (Properties of the additive Schwarz operator).** *The operator  $\mathcal{S}_\ell^{\text{AS}}$  defined in (52) is linear, bounded, and symmetric, i.e., there holds*

$$\langle\langle \mathcal{S}_\ell^{\text{AS}} v, w \rangle\rangle = \langle\langle v, \mathcal{S}_\ell^{\text{AS}} w \rangle\rangle \quad \text{for all } v, w \in \mathcal{X}. \quad (68)$$

Moreover, it holds that

$$c \lll v_\ell \lll^2 \leq \langle\langle \mathcal{S}_\ell^{\text{AS}} v_\ell, v_\ell \rangle\rangle \leq C \lll v_\ell \lll^2 \quad \text{for all } v_\ell \in \mathcal{X}_\ell^p, \quad (69)$$

where the constants  $c$  and  $C$  depend only on  $\Omega$ ,  $\mathcal{T}_0$ ,  $d$ ,  $\gamma$ -shape regularity,  $C_{\text{loc}}^{(1)}$ , and  $C_{\text{loc}}^{(2)}$ .

---

*Proof.* The proof is divided into three steps.

**Step 1 (basic properties):** The linearity, boundedness, and symmetry follow directly from the additive structure of  $\mathcal{S}_\ell^{\text{AS}}$  and the respective property of the levelwise operators  $\mathcal{S}_{\ell'}$  from (22).

**Step 2 (lower bound in (69)):** We show the lower bound using Lemma 21. For our application, we consider the space decomposition

$$\mathcal{X}_\ell^p = \mathcal{X}_0^1 + \sum_{\ell'=1}^{\ell-1} \sum_{z \in \mathcal{V}_{\ell'}^+} \mathcal{X}_{\ell',z}^1 + \sum_{z \in \mathcal{V}_\ell} \mathcal{X}_{\ell,z}^p.$$

In [IMPS24, Proposition 5.5] and [Hil25], it is shown that there exists an  $hp$ -robust stable decomposition, i.e., for every  $v_\ell \in \mathcal{X}_\ell^p$ , there exist functions  $v_0 \in \mathcal{X}_0^1$ ,  $v_{\ell',z} \in \mathcal{X}_{\ell',z}^1$ , and  $v_{\ell,z} \in \mathcal{X}_{\ell,z}^p$  such that  $v_\ell = v_0 + \sum_{\ell'=1}^{\ell-1} \sum_{z \in \mathcal{V}_{\ell'}^+} v_{\ell',z} + \sum_{z \in \mathcal{V}_\ell} v_{\ell,z}$  and

$$\lll v_0 \lll^2 + \sum_{\ell'=1}^{\ell-1} \sum_{z \in \mathcal{V}_{\ell'}^+} \lll v_{\ell',z} \lll^2 + \sum_{z \in \mathcal{V}_\ell} \lll v_{\ell,z} \lll^2 \leq \tilde{C} \lll v_\ell \lll^2,$$

where the constant  $\tilde{C} > 0$  depends only on the initial triangulation  $\mathcal{T}_0$ ,  $\gamma$ -shape regularity, and  $C_{\text{loc}}^{(2)}$ . Therefore, Lemma 21 yields the lower bound in (69) with  $c := 1/\tilde{C}$ .

**Step 2 (upper bound in (69)):** Let  $v_\ell \in \mathcal{X}_\ell^p$ . Recall  $M = \max_{T \in \mathcal{T}_\ell} \text{level}(T)$  and observe  $\mathcal{G}_{\ell,m}^1 v_\ell \in \hat{\mathcal{X}}_m^1 \subseteq \hat{\mathcal{X}}_M^1$ . Utilizing (62), we have

$$\langle\langle \mathcal{S}_\ell^{\text{AS}} v_\ell, v_\ell \rangle\rangle = \langle\langle \mathcal{P}_0^1 v_\ell, v_\ell \rangle\rangle + \sum_{m=0}^M \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \rangle\rangle + \sum_{z \in \mathcal{V}_\ell} \langle\langle \mathcal{P}_{\ell,z}^p v_\ell, v_\ell \rangle\rangle.$$

To complete the proof, we must bound each of the summands by  $\lesssim \lll v_\ell \lll^2$ .

On the initial level, we have  $\langle\langle \mathcal{P}_0^1 v_\ell, v_\ell \rangle\rangle \leq \lll v_\ell \lll^2$ .

On the finest level  $\ell$ , we apply the Cauchy-Schwarz inequality, Young inequality for  $\mu > 0$ , and finite patch overlap to see

$$\begin{aligned} \sum_{z \in \mathcal{V}_\ell} \langle\langle \mathcal{P}_{\ell,z}^p v_\ell, v_\ell \rangle\rangle &\leq \lll v_\ell \lll \left\| \sum_{z \in \mathcal{V}_\ell} \mathcal{P}_{\ell,z}^p v_\ell \right\| \leq \frac{\mu}{2} \left\| \sum_{z \in \mathcal{V}_\ell} \mathcal{P}_{\ell,z}^p v_\ell \right\|^2 + \frac{1}{2\mu} \lll v_\ell \lll^2 \\ &\leq \frac{\mu}{2} (d+1) \sum_{z \in \mathcal{V}_\ell} \lll \mathcal{P}_{\ell,z}^p v_\ell \lll^2 + \frac{1}{2\mu} \lll v_\ell \lll^2 = \frac{\mu}{2} (d+1) \sum_{z \in \mathcal{V}_\ell} \langle\langle \mathcal{P}_{\ell,z}^p v_\ell, v_\ell \rangle\rangle + \frac{1}{2\mu} \lll v_\ell \lll^2. \end{aligned}$$

With  $\mu = (d+1)^{-1}$ , this proves

$$\sum_{z \in \mathcal{V}_\ell} \langle\langle \mathcal{P}_{\ell,z}^p v_\ell, v_\ell \rangle\rangle \leq (d+1) \lll v_\ell \lll^2.$$

It remains to consider the intermediate levels, where we will exploit the strengthened Cauchy–Schwarz inequality from Lemma 24. Let us denote by  $\widehat{\mathcal{Q}}_m : H_0^1(\Omega) \rightarrow \widehat{\mathcal{X}}_m^1$  the Galerkin projections for the uniform meshes, i.e.,  $\langle\langle \widehat{\mathcal{Q}}_m v, \widehat{w}_m \rangle\rangle = \langle\langle v, \widehat{w}_m \rangle\rangle$  for all  $v \in H_0^1(\Omega)$  and all  $\widehat{w}_m \in \widehat{\mathcal{X}}_m^1$ . With  $\widehat{\mathcal{Q}}_{-1} := 0$ , it holds that

$$\widehat{\mathcal{Q}}_m = \sum_{k=0}^m (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}).$$

Since the operators  $\mathcal{P}_{\ell',z}^1$  are symmetric, the bilinear form  $\langle\langle \mathcal{G}_{\ell,m}^1 \cdot, \cdot \rangle\rangle$  is symmetric on the space  $\mathcal{X}_\ell^p$ . The positivity

$$\langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \rangle\rangle = \sum_{\ell'=1}^{\ell-1} \sum_{\substack{z \in \mathcal{V}_{\ell'}^+ \\ g_{\ell',z}=m}} \langle\langle \mathcal{P}_{\ell',z}^1 v_\ell, v_\ell \rangle\rangle = \sum_{\ell'=1}^{\ell-1} \sum_{\substack{z \in \mathcal{V}_{\ell'}^+ \\ g_{\ell',z}=m}} \|\mathcal{P}_{\ell',z}^1 v_\ell\|^2 \geq 0 \quad \text{for all } v_\ell \in \mathcal{X}_\ell^p$$

implies the Cauchy–Schwarz inequality

$$\langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, w_\ell \rangle\rangle \leq \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \rangle\rangle^{1/2} \langle\langle \mathcal{G}_{\ell,m}^1 w_\ell, w_\ell \rangle\rangle^{1/2} \quad \text{for all } v_\ell, w_\ell \in \mathcal{X}_\ell^p. \quad (70)$$

Consequently, we can estimate the sum over the intermediate levels by

$$\begin{aligned} \sum_{m=0}^M \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \rangle\rangle &= \sum_{m=0}^M \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, \widehat{\mathcal{Q}}_m v_\ell \rangle\rangle = \sum_{m=0}^M \sum_{k=0}^m \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell \rangle\rangle \\ &\stackrel{(70)}{\leq} \sum_{m=0}^M \sum_{k=0}^m \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \rangle\rangle^{1/2} \langle\langle \mathcal{G}_{\ell,m}^1 (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell, (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell \rangle\rangle^{1/2}. \end{aligned}$$

As  $(\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell \in \widehat{\mathcal{X}}_k^1$ , the strengthened Cauchy–Schwarz inequality from Lemma 24 yields

$$\langle\langle \mathcal{G}_{\ell,m}^1 (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell, (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell \rangle\rangle \lesssim \delta^{m-k} \|\mathcal{G}_{\ell,m}^1 (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell\|^2.$$

Moreover, we observe that

$$\|\mathcal{G}_{\ell,m}^1 (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell\|^2 = \langle\langle \widehat{\mathcal{Q}}_k v_\ell, v_\ell \rangle\rangle - 2\langle\langle v_\ell, \widehat{\mathcal{Q}}_{k-1} v_\ell \rangle\rangle + \langle\langle \widehat{\mathcal{Q}}_{k-1} v_\ell, v_\ell \rangle\rangle = \langle\langle (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell, v_\ell \rangle\rangle.$$

Applying the Young inequality and the summability of the geometric series, we get

$$\begin{aligned} \sum_{m=0}^M \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \rangle\rangle &\lesssim \sum_{m=0}^M \sum_{k=0}^m \delta^{(m-k)/2} \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \rangle\rangle^{1/2} \langle\langle (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell, v_\ell \rangle\rangle^{1/2} \\ &\leq \frac{\mu}{2} \sum_{m=0}^M \sum_{k=0}^m \delta^{m-k} \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \rangle\rangle + \frac{1}{2\mu} \sum_{m=0}^M \sum_{k=0}^m \delta^{m-k} \langle\langle (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell, v_\ell \rangle\rangle \\ &\lesssim \frac{\mu}{2} \left\langle\left\langle \sum_{m=0}^M \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \right\rangle\right\rangle + \frac{1}{2\mu} \sum_{k=0}^M \sum_{m=k}^M \delta^{m-k} \langle\langle (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell, v_\ell \rangle\rangle \\ &\lesssim \frac{\mu}{2} \left\langle\left\langle \sum_{m=0}^M \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \right\rangle\right\rangle + \frac{1}{2\mu} \left\langle\left\langle \sum_{k=0}^M (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell, v_\ell \right\rangle\right\rangle. \end{aligned}$$

Since  $\widehat{\mathcal{Q}}_M$  is an orthogonal projection, it holds that

$$\left\langle\left\langle \sum_{k=0}^M (\widehat{\mathcal{Q}}_k - \widehat{\mathcal{Q}}_{k-1}) v_\ell, v_\ell \right\rangle\right\rangle = \langle\langle \widehat{\mathcal{Q}}_M v_\ell, v_\ell \rangle\rangle \leq \|v_\ell\|^2.$$

Choosing  $\mu$  sufficiently small, we obtain

$$\sum_{m=0}^M \langle\langle \mathcal{G}_{\ell,m}^1 v_\ell, v_\ell \rangle\rangle \lesssim \|v_\ell\|^2.$$

This concludes the proof.  $\square$

**6.4. Proof of Theorem 19.** By definition (51), it is clear that  $\mathbf{B}_\ell^{\text{AS}}$  is an SPD matrix. Let  $\mathbf{x}_\ell, \mathbf{y}_\ell \in \mathbb{R}^{N_\ell^p}$ ,  $v_\ell := \sum_{j=1}^{N_\ell^p} (\mathbf{x}_\ell)_j \varphi_{\ell,j}^p \in \mathcal{X}_\ell^p$  and  $w_\ell := \sum_{j=1}^{N_\ell^p} (\mathbf{y}_\ell)_j \varphi_{\ell,j}^p \in \mathcal{X}_\ell^p$ . Due to the identity (54) and the symmetry of  $\mathcal{S}_\ell^{\text{AS}}$ , it follows that

$$(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell \mathbf{x}_\ell, \mathbf{y}_\ell)_{\mathbf{A}_\ell} \stackrel{(54)}{=} \langle\langle \mathcal{S}_\ell^{\text{AS}} v_\ell, w_\ell \rangle\rangle \stackrel{(68)}{=} \langle\langle v_\ell, \mathcal{S}_\ell^{\text{AS}} w_\ell \rangle\rangle \stackrel{(54)}{=} (\mathbf{x}_\ell, \mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell \mathbf{y}_\ell)_{\mathbf{A}_\ell}.$$

Thus, the matrix  $\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell$  is symmetric with respect to the scalar product  $(\cdot, \cdot)_{\mathbf{A}_\ell}$ . We use [TW05, Theorem C.1] to write down the expressions for the maximal and minimal eigenvalue of  $\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell$ . Together with the identity (54) and the bounds in (69), we get

$$\lambda_{\min}(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell) = \min_{\substack{\mathbf{x}_\ell \in \mathbb{R}^{N_\ell^p} \\ \mathbf{x}_\ell \neq 0}} \frac{(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell}}{(\mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell}} \stackrel{(54)}{=} \min_{\substack{v_\ell \in \mathcal{X}_\ell^p \\ v_\ell \neq 0}} \frac{\langle\langle \mathcal{S}_\ell^{\text{AS}} v_\ell, v_\ell \rangle\rangle}{\|v_\ell\|^2} \stackrel{(69)}{\geq} c$$

and

$$\lambda_{\max}(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell) = \max_{\substack{\mathbf{x}_\ell \in \mathbb{R}^{N_\ell^p} \\ \mathbf{x}_\ell \neq 0}} \frac{(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell \mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell}}{(\mathbf{x}_\ell, \mathbf{x}_\ell)_{\mathbf{A}_\ell}} \stackrel{(54)}{=} \max_{\substack{v_\ell \in \mathcal{X}_\ell^p \\ v_\ell \neq 0}} \frac{\langle\langle \mathcal{S}_\ell^{\text{AS}} v_\ell, v_\ell \rangle\rangle}{\|v_\ell\|^2} \stackrel{(69)}{\leq} C.$$

Hence, we also obtain that

$$\text{cond}_{\mathbf{A}_\ell}(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell) = \frac{\lambda_{\max}(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell)}{\lambda_{\min}(\mathbf{B}_\ell^{\text{AS}} \mathbf{A}_\ell)} \leq \frac{C}{c}.$$

Uniform contraction (53) follows directly from Proposition 1 and thus concludes the proof.  $\square$

## 7. ADAPTIVE FEM WITH CONTRACTIVE SOLVER

In this section, we present the application of uniformly contractive solvers in AFEM and the optimal complexity of the resulting adaptive algorithm. Let us denote by  $\Psi_\ell : \mathcal{X}_\ell^p \rightarrow \mathcal{X}_\ell^p$  the iteration map of an iterative solver of linear complexity which is uniformly contractive, i.e., there exists a constant  $q \in (0, 1)$  such that

$$\|u_\ell^* - \Psi_\ell(u_\ell)\| \leq q \|u_\ell^* - u_\ell\| \quad \text{for all } u_\ell \in \mathcal{X}_\ell^p. \quad (71)$$

We consider the refinement indicators of the standard residual error estimator

$$\eta_\ell(T, u_\ell)^2 := h_T^2 \|f + \text{div}(\mathbf{K} \nabla u_\ell)\|_T^2 + h_T \|[\![ \mathbf{K} \nabla u_\ell ]\!] \cdot \mathbf{n}\|_{\partial T \cap \Omega}^2 \quad \text{for } T \in \mathcal{T}_\ell, \quad (72a)$$

where  $\mathbf{n}$  denotes the outer normal vector of the element  $T$  and  $[\![ \cdot ]\!]$  denotes the jump across the element boundary. Define

$$\eta_\ell(\mathcal{U}_\ell, u_\ell)^2 := \sum_{T \in \mathcal{U}_\ell} \eta_\ell(T, u_\ell)^2 \quad \text{for } \mathcal{U}_\ell \subseteq \mathcal{T}_\ell \text{ and all } u_\ell \in \mathcal{X}_\ell^p. \quad (72b)$$

For  $\mathcal{U}_\ell = \mathcal{T}_\ell$ , we abbreviate  $\eta_\ell(u_\ell)^2 := \eta_\ell(\mathcal{T}_\ell, u_\ell)^2$ . Consider the adaptive algorithm with iterative solver from, e.g., [GHPS21].

---

**Algorithm E (AFEM with optimal iterative solver).** *Input:* Initial triangulation  $\mathcal{T}_0$ , polynomial degree  $p \geq 1$ , initial guess  $u_0^0 := 0$ , adaptivity parameters  $0 < \theta \leq 1$ ,  $C_{\text{mark}} \geq 1$ , and  $\mu > 0$ . Then, for all  $\ell = 0, 1, 2, \dots$ , perform the following steps (i)–(iii):

(i) **Solve & Estimate:** For all  $k = 1, 2, 3, \dots$ , repeat (a)–(b) until

$$\|u_\ell^k - u_\ell^{k-1}\| \leq \mu \eta_\ell(u_\ell^k) :$$

(a) Compute  $u_\ell^k := \Psi_\ell(u_\ell^{k-1})$  with one step of the algebraic solver.

(b) Compute the refinement indicators  $\eta_\ell(T, u_\ell^k)$  for all  $T \in \mathcal{T}_\ell$ .

Upon termination of the  $k$ -loop, define the index  $\underline{k}[\ell] := k \in \mathbb{N}$  and  $u_\ell^{\underline{k}} := u_\ell^k$ .

- (ii) **Mark:** Employ Dörfler marking to determine a set  $\mathcal{M}_\ell \in \mathbb{M}_\ell[\theta, u_\ell^k] := \{\mathcal{U}_\ell \subset \mathcal{T}_\ell : \theta \eta_\ell(u_\ell^k)^2 \leq \eta_\ell(\mathcal{U}_\ell, u_\ell^k)^2\}$  that fulfills

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

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

**Output:** Sequence of triangulations  $\{\mathcal{T}_\ell\}_{\ell \geq 0}$  and discrete approximations  $\{u_\ell^k\}_{\ell \geq 0}$ .

In order to formulate optimal complexity, we first define the countably infinite set

$$\mathcal{Q} := \{(\ell, k) \in \mathbb{N}_0^2 : u_\ell^k \text{ is defined in Algorithm E}\}.$$

The set  $\mathcal{Q}$  can be equipped with the natural order

$$(\ell', k') \leq (\ell, k) := \iff u_{\ell'}^{k'} \text{ is computed earlier than or equal to } u_\ell^k \text{ in Algorithm E}.$$

Furthermore, we define the total step counter by

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

Finally, we introduce the notion of approximation classes following [BDDP02; BDD04; Ste07; CKNS08; CFPP14]. For any rate  $s > 0$ , define

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

where  $\mathbb{T}_N(\mathcal{T}_0) := \{\mathcal{T}_H \in \mathbb{T}(\mathcal{T}_0) : \#\mathcal{T}_H - \#\mathcal{T}_0 \leq N\}$ . The notation  $\mathcal{T}_H \in \mathbb{T}(\mathcal{T}_0)$  abbreviates that  $\mathcal{T}_H$  can be obtained from  $\mathcal{T}_0$  by a finite number of newest vertex bisection steps. If  $\|u^*\|_{\mathbb{A}_s} < \infty$ , then the error estimator  $\eta_{\text{opt}}(u_{\text{opt}}^*)$  decays with rate  $-s$  with respect to the number of elements of a sequence of (practically unknown) optimal triangulations  $\mathcal{T}_{\text{opt}}$ . However, due to the iterative solver, rates of the adaptive algorithm should rather be considered with respect to the computational time or, equivalently, the total computational cost. Hence, we comment on the computational cost of implementing Algorithm E: Calculating the error estimator  $\eta_\ell(u_\ell^k)$  has (up to quadrature) cost  $O(\#\mathcal{T}_\ell)$  as this consist only of element-wise operations. 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]. If one step of the algebraic solver  $\Psi_\ell$  is of linear complexity, then total computational cost to compute  $u_\ell^k$  via Algorithm E is given by

$$\text{cost}(\ell, k) := \sum_{\substack{(\ell', k') \in \mathcal{Q} \\ |\ell', k'| \leq |\ell, k|}} \#\mathcal{T}_{\ell'}. \quad (73)$$

We can now state optimal complexity of Algorithm E. The proof fits within the setting of [BFM<sup>+</sup>25, Theorem 2.3] relying on [GHPS21, Theorem 8] and is thus omitted here.

**Theorem 26 (Optimal complexity of Algorithm E).** *Let  $s > 0$ . Let the algebraic solver  $\Psi_\ell$  be given by either GPCG with  $\mathbf{B}_\ell^{\text{MG}}$  preconditioner (25) or PCG with  $\mathbf{B}_\ell^{\text{sMG}}$  preconditioner (41) or  $\mathbf{B}_\ell^{\text{AS}}$  preconditioner (51) satisfying linear complexity and uniform contraction (71). Define the quasi-error by*

$$\mathbf{H}_\ell^k := \|\|u_\ell^* - u_\ell^k\|\| + \eta_\ell(u_\ell^*) \quad \text{for all } (\ell, k) \in \mathcal{Q}. \quad (74)$$

*For arbitrary  $0 < \theta \leq 1$ ,  $C_{\text{mark}} \geq 1$  and  $\mu > 0$ , there holds full  $R$ -linear convergence, i.e., there exist constants  $C_{\text{lin}} \geq 1$  and  $0 < q_{\text{lin}} < 1$  such that*

$$\mathbf{H}_{\ell'}^{k'} \leq C_{\text{lin}} q_{\text{lin}}^{|\ell', k'| - |\ell, k|} \mathbf{H}_\ell^k \quad \text{for all } (\ell, k), (\ell', k') \in \mathcal{Q} \text{ with } (\ell', k') \geq (\ell, k). \quad (75)$$

*With  $C_{\text{cost}} := C_{\text{lin}}(1 - q_{\text{lin}}^{1/s})^{-s}$ , this yields*

$$\sup_{(\ell, k) \in \mathcal{Q}} (\#\mathcal{T}_\ell)^s \mathbf{H}_\ell^k \leq \sup_{(\ell, k) \in \mathcal{Q}} \text{cost}(\ell, k)^s \mathbf{H}_\ell^k \leq C_{\text{cost}} \sup_{(\ell, k) \in \mathcal{Q}} (\#\mathcal{T}_\ell)^s \mathbf{H}_\ell^k \quad \text{for all } s > 0. \quad (76)$$

Moreover, there exists  $0 < \theta^* \leq 1$  and  $\mu^* > 0$  such that, for sufficiently small parameters

$$0 < \mu < \mu^* \quad \text{and} \quad 0 < \frac{(\theta^{1/2} + \mu/\mu^*)^2}{(1 - \mu/\mu^*)^2} < \theta^*, \quad (77)$$

Algorithm E guarantees that

$$c_{\text{opt}} \|u^*\|_{\mathbb{A}_s} \leq \sup_{(\ell, k) \in \mathcal{Q}} \text{cost}(\ell, k)^s \mathbf{H}_\ell^k \leq C_{\text{opt}} \max\{\|u^*\|_{\mathbb{A}_s}, \mathbf{H}_0^0\}. \quad (78)$$

The constants  $C_{\text{opt}}, c_{\text{opt}} > 0$  depend only on the polynomial degree  $p$ , the initial triangulation  $\mathcal{T}_0$ , the rate  $s$ , the adaptivity parameters  $\theta$  and  $\mu$ , the solver contraction constant  $q$ , constants stemming from the axioms of adaptivity [CFPP14] for the residual error estimator (72), and the properties of NVB. In particular, every possible convergence rate is achieved with respect to the overall computational cost.  $\square$

The interpretation of Theorem 26 reads as follows: Unconditionally of the adaptivity parameters, Algorithm E leads essentially to contraction (75) of the quasi-error (74), independently of the algorithmic decision for mesh-refinement or yet another solver step. This yields (76), which states that the convergence rate  $-s$  with respect to the number of degrees of freedom  $\dim(\mathcal{X}_\ell^p) \simeq \#\mathcal{T}_\ell$  and with respect to the total computational cost (73) (and hence total computing time) coincide. Moreover, for sufficiently small parameters (77), Algorithm E will achieve optimal complexity (78): If the error estimator evaluated for the exact FE solutions decays with rate  $-s$  along a sequence of optimal meshes, then Algorithm E guarantees that the quasi-error (74) decays with rate  $-s$  with respect to the computational cost, i.e., inexact solution does indeed not spoil the overall convergence behavior.

---

**Remark 27.** Algorithm E and Theorem 26 can be generalized to non-symmetric second-order linear elliptic PDEs that fit into the framework of the Lax–Milgram lemma, see [FHMP26]. Then, the algebraic solver employs a preconditioned GMRES method with an optimal symmetric and positive definite preconditioner for the principal part. Based on the present work, canonical preconditioning includes the linear and symmetric multigrid method from Section 5 as well as the multilevel additive Schwarz preconditioner from Section 6.

---

## 8. NUMERICAL EXPERIMENTS

In this section, we compare the behavior of the proposed algebraic solvers, where the multigrid solver from [IMPS24] is used as a baseline. For this, we investigate both the performance of the solver itself as well as its application in the adaptive Algorithm E. The experiments are conducted in the open-source Matlab package MooAFEM [IP23].

**8.1. Considered model problems.** We consider the following two test cases of model problem (1):

- *Poisson*: Let  $\Omega = (-1, 1)^2 \setminus ([0, 1] \times [-1, 0])$  be the L-shaped domain with diffusion coefficient  $\mathbf{K} = \mathbf{I}$  and right-hand side  $f = 1$ . An example of a mesh obtained via AFEM for this problem is displayed in Figure 1 (left).
- *Checkerboard*: Let  $\Omega = (0, 1)^2$  be the unit square and  $\mathbf{K}$  the  $2 \times 2$  checkerboard diffusion with values 1 (white) and 100 (gray); see Figure 1 (right). We refer to [Kel75] for an exact solution of this problem.

**8.2. Considered iterative solvers.** We give an overview of all considered iterative solvers for the numerical experiments.

- **MG**: The geometric multigrid solver from [IMPS24] with  $h$ - and  $p$ -robust contraction factor; see Proposition 12.
- **GPCG+MG**: GPCG of Algorithm B with the non-linear and non-symmetric multigrid preconditioner defined in (25), for which Theorem 8 establishes its  $h$ - and  $p$ -robust contraction.

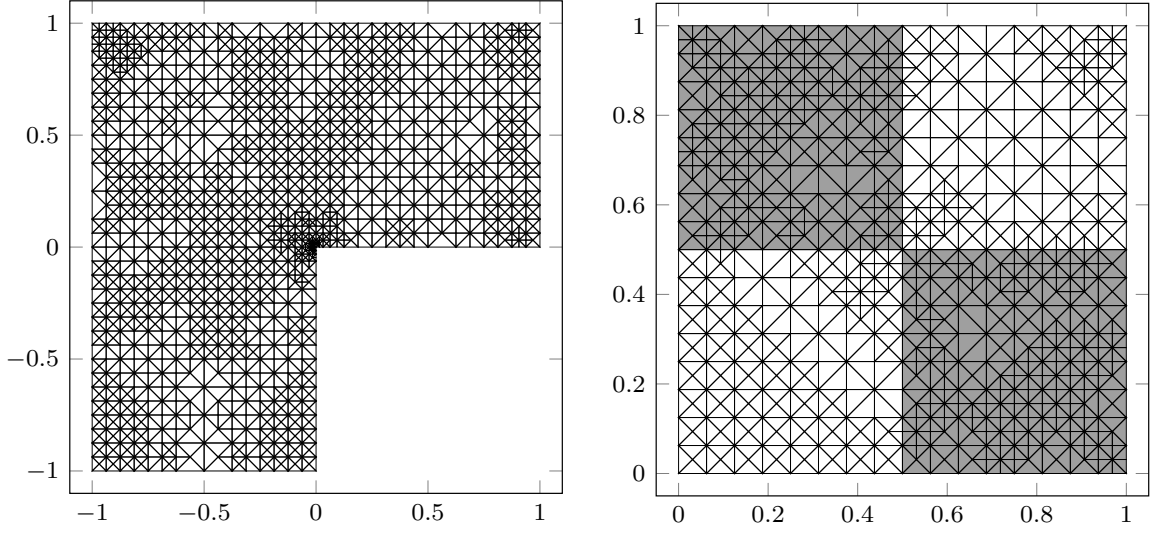


FIGURE 1. Adaptively refined meshes for the Poisson problem from Section 8.1 with  $\#\mathcal{T}_8 = 2490$  (left) and checkerboard problem from Section 8.1 with  $\#\mathcal{T}_8 = 1242$  (right) for adaptivity parameters  $\theta = 0.5$  and  $\lambda = 0.1$ .

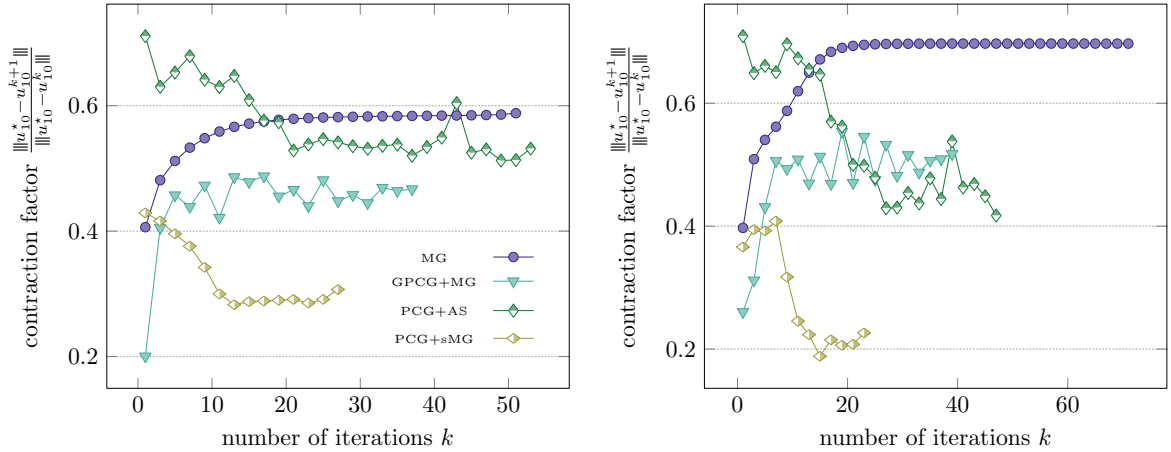


FIGURE 2. History plot of the contraction factor for the Poisson problem from Section 8.1 with  $\#\mathcal{T}_{10} = 9723$  for  $p = 1$  (left) and  $\#\mathcal{T}_{10} = 340$  for  $p = 4$  (right) and solver stopping criterion  $\|u_{10}^* - u_{10}^k\| < 10^{-13}$ .

- **PCG+AS**: PCG from Section 3 using the additive Schwarz preconditioner from (51), for which Theorem 19 guarantees  $h$ - and  $p$ -robust contraction.
- **PCG+sMG**: PCG employing the symmetric multigrid preconditioner from (41), for which Theorem 16 ensures  $h$ - and  $p$ -robust contraction.
- **PCG+MG**: PCG with the non-linear and non-symmetric multigrid preconditioner defined in (25). While no theoretical contraction properties are known, we test experimentally the performance.
- **PCG+nsMG**: PCG with the non-symmetric multigrid preconditioner  $\mathbf{B}_\ell^{\text{nsMG}}$  defined in Remark 14. While no theoretical contraction properties are known, we test experimentally the performance.

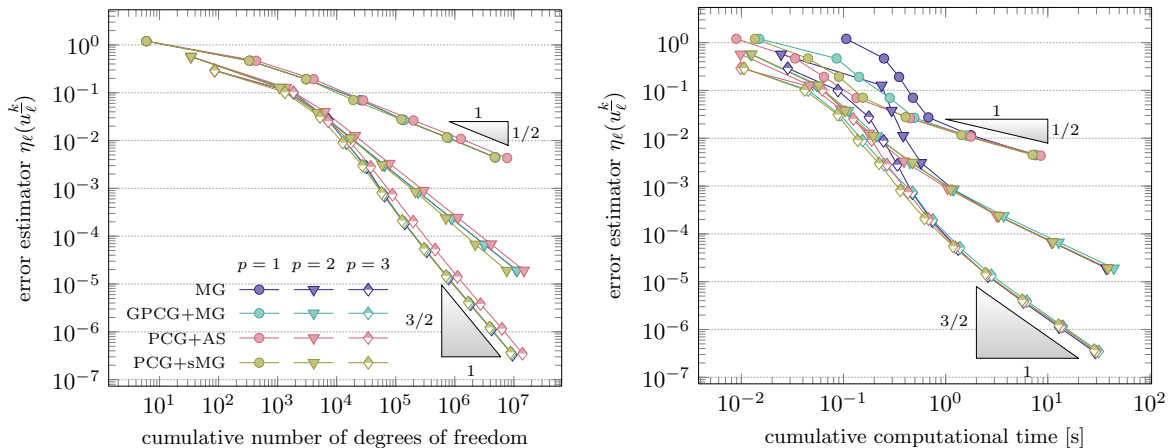


FIGURE 3. Convergence of the error estimator  $\eta_\ell(u_\ell^k)$  for the Poisson problem from Section 8.1 with  $\theta = 0.5$  and  $\lambda = 0.05$ .

**8.3. Solver contraction.** To calculate the experimental contraction factors of the different methods, we first precompute a mesh hierarchy with  $\ell = 10$  and corresponding polynomial degree for the Poisson problem from Section 8.1 using Algorithm E together with the geometric multigrid solver from [IMPS24] and parameters  $\theta = 0.5$  and  $\lambda = 0.1$ . As stopping criterion, we iterate until the algebraic error  $\|u_{10}^* - u_{10}^k\|$  drops below  $10^{-13}$ ; see Figure 2. While Theorem 8 provides the same analytical contraction factor for both the standalone multigrid method and GPCG employing the multigrid preconditioner. Figure 2 shows that the latter performs better. Among the tested methods, PCG+sMG attains the smallest contraction factor. However, one should interpret this comparison with caution, since each iteration of the symmetric preconditioner requires roughly twice the computational complexity of its non-symmetric counterpart. Another notable difference is that the contraction factor of GPCG+MG increases with the number of solver steps, whereas for PCG methods it decreases. This behavior can be attributed to PCG being a Krylov subspace method, where minimization occurs over an expanding subspace, while GPCG lacks this property. This could also help explain why the contraction factor for PCG+sMG improves over the iterations. Note that, as stated by theory, these contraction factors are indeed robust in the polynomial degree  $p$ .

**8.4. Overall solver performance within AFEM.** Figure 3 displays the error estimator  $\eta_\ell(u_\ell^k)$ , computed by Algorithm E, over the cumulative degrees of freedom and the cumulative time. Note that for the final iterate the error estimator  $\eta_\ell(u_\ell^k)$  is equivalent to the quasi-error  $H_\ell^k$  from Theorem 26. This follows from the axioms of adaptivity and the stopping criterion in Algorithm E; see, e.g., [BMP24]. After a pre-asymptotic phase, one can observe the optimal convergence rates  $-p/2$  in Figure 3 both with respect to the cumulative degrees of freedom and with respect to the cumulative time, across all optimal solvers GPCG+MG, PCG+AS, and PCG+sMG developed in this work. In Figure 3 (right), we see that the standalone MG is slower in the pre-asymptotic phase compared to the other solvers. Moreover, Figure 4 confirms numerically that the number of solver steps with respect to the degrees of freedom for  $p \in \{1, 3\}$  remains uniformly bounded for all solvers.

We now consider the checkerboard problem from Section 8.1 for AFEM with MG, GPCG+MG, PCG+AS, and PCG+sMG. The parameters are set to  $\lambda \in \{0.01, 0.05\}$ ,  $p \in \{1, 3\}$ , and  $\theta \in \{0.3, 0.5\}$ , and we study the decrease of the error estimator  $\eta_\ell(u_\ell^k)$  with respect to the cumulative number of degrees of freedom. In Figure 5 (left), all four solvers exhibit optimal convergence rates for  $\lambda = 0.01$  and  $\theta = 0.3$ . In contrast, when  $\lambda = 0.05$  and  $\theta = 0.5$ , the

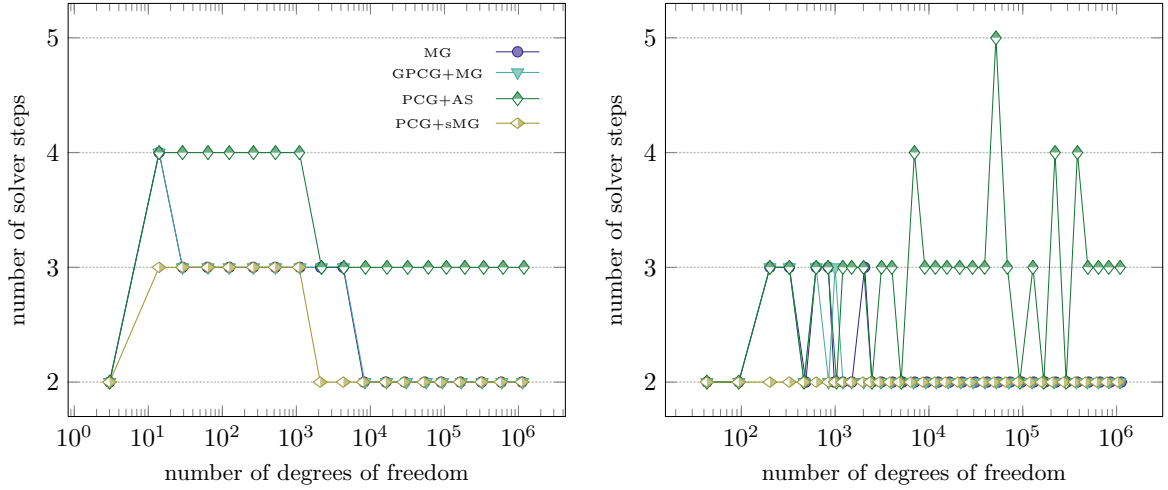


FIGURE 4. History plot of solver steps with respect to the degrees of freedom for the Poisson problem from Section 8.1 with  $p = 1$  (left) and  $p = 3$  (right).

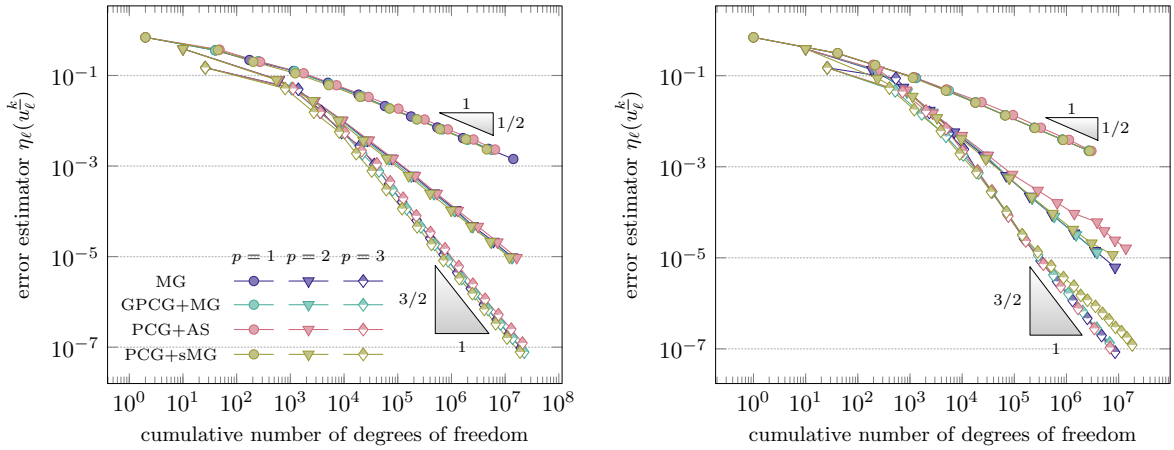


FIGURE 5. Convergence of the error estimator for the checkerboard problem from Section 8.1 with  $\lambda = 0.01$ ,  $\theta = 0.3$  (left) and  $\lambda = 0.05$ ,  $\theta = 0.5$  (right).

convergence rates for PCG+sMG degrade for both  $p = 2$  and  $p = 3$ . Additionally, PCG+AS shows a suboptimal rate for  $p = 2$ . While MG and GPCG+MG appear to maintain optimal rates across all configurations, a closer inspection for  $p = 3$  reveals a slightly suboptimal rate, as shown in Figure 5 (right). This does not contradict Theorem 26, since the assumptions require sufficiently small adaptivity parameters.

To gain a deeper insight into the solvers' behavior, we compute their experimental contraction factors using a pre-computed mesh hierarchy with  $\ell = 20$  levels for  $p = 2$  and  $\ell = 35$  levels for  $p = 3$ . Figure 6 shows the first 10 iterations, where PCG+AS and PCG+sMG exhibit larger contraction factors in the initial iterations compared to GPCG+MG. For this specific problem and parameter choices ( $\lambda = 0.01$  and  $\theta = 0.3$ ), the number of iterations per level required by any tested algebraic solver within the AFEM algorithm never exceeded 8 and was typically around 2. This observation may partially explain the inferior performance of PCG+AS and PCG+sMG. Additionally, both MG and GPCG+MG employ optimal step sizes (see step (ii) and (iii) of Algorithm C) within the multigrid scheme, whereas PCG+sMG relies on a fixed step size (see Algorithm D), which may further contribute to its suboptimal behavior.

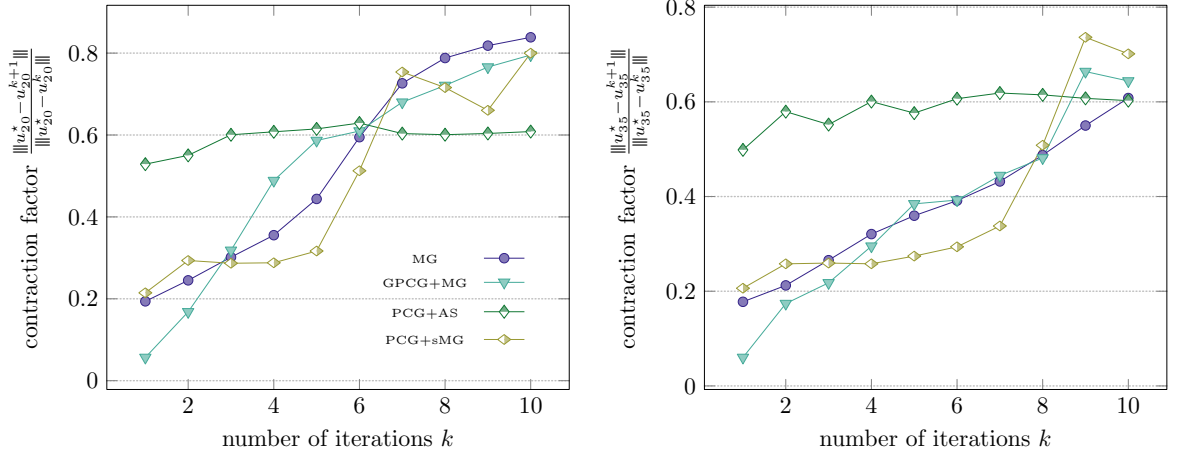


FIGURE 6. History plot of the contraction factor for the checkerboard problem from Section 8.1 with  $\ell = 20$  (with  $\#\mathcal{T}_\ell = 97136$ ) for  $p = 2$  (left) and  $\ell = 35$  (with  $\#\mathcal{T}_\ell = 77681$ ) for  $p = 3$  (right).

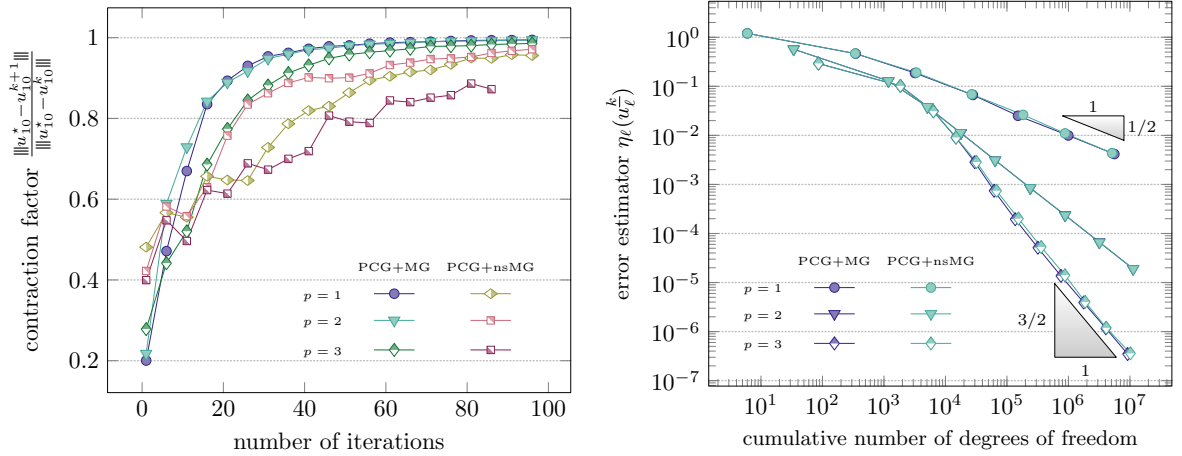


FIGURE 7. History plot of the contraction factor for the Poisson problem from Section 8.1 employing the algebraic solvers PCG+MG and PCG+nsMG until  $\|u_{10}^* - u_{10}^k\| < 10^{-13}$  or for a maximum of 100 iterations with  $\#\mathcal{T}_{10} = 9723$  for  $p = 1$ ,  $\#\mathcal{T}_{10} = 1096$  for  $p = 2$ , and  $\#\mathcal{T}_{10} = 405$  for  $p = 3$  (left). Convergence of the error estimator  $\eta_\ell(u_\ell^k)$  for AFEM with PCG+MG and PCG+nsMG using  $\theta = 0.5$  and  $\lambda = 0.05$  (right).

**8.5. Warning example.** Although the theory does not cover non-linear and/or non-symmetric preconditioners within PCG, we nevertheless test the solvers PCG+MG and PCG+nsMG from Section 8.2 numerically. Figure 7 (left) shows that the contraction factors deteriorate for PCG+MG for all polynomial degree and for PCG+nsMG when  $p \in \{1, 2\}$ . Since, in these cases the algebraic error  $\|u_{10}^* - u_{10}^k\|$  does not drop below  $10^{-13}$ , we stop the solver after 100 iterations and report the final errors in Table 1. Despite this poor algebraic performance, we also test PCG+MG and PCG+nsMG within AFEM. Figure 7 (right) shows optimal convergence rates with respect to the cumulative number of degrees of freedom. This is likely because AFEM typically requires only a small number of solver iterations per level (see Figure 4). Consequently, the deterioration of the contraction factors has little influence in this setting. However, for more

	PCG+MG		PCG+nsMG	
	final error	# solver steps	final error	# solver steps
$p = 1$	$1.373 \times 10^{-7}$	100	$4.562 \times 10^{-12}$	100
$p = 2$	$5.816 \times 10^{-7}$	100	$7.677 \times 10^{-10}$	100
$p = 3$	$1.127 \times 10^{-9}$	100	$9.338 \times 10^{-14}$	90

TABLE 1. Final algebraic error and number of solver steps of PCG+MG and PCG+nsMG for the Poisson problem from Section 8.1 applied to a pre-computed mesh hierarchy with  $\ell = 10$  levels after 100 iterations or the first error below  $10^{-13}$ . We stress that convergence of PCG+MG and PCG+nsMG remains theoretically open and can fail.

difficult problems, or for smaller values of the parameter  $\mu$ , AFEM requires more solver steps, so that optimal complexity will be affected by the degraded contraction behavior. Thus, both PCG+MG and PCG+nsMG are not suitable for use within AFEM.

#### REFERENCES

- [AFF<sup>+</sup>13] M. Aurada, M. Feischl, T. Führer, M. Karkulik, and D. Praetorius. Efficiency and optimality of some weighted-residual error estimator for adaptive 2D boundary element methods. *Comput. Methods Appl. Math.*, 13(3):305–332, 2013. DOI: [10.1515/cmam-2013-0010](https://doi.org/10.1515/cmam-2013-0010).
- [AV91] O. Axelsson and P. S. Vassilevski. A black box generalized conjugate gradient solver with inner iterations and variable-step preconditioning. *SIAM J. Matrix Anal. Appl.*, 12(4):625–644, 1991. DOI: [10.1137/0612048](https://doi.org/10.1137/0612048).
- [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. Dedicated to the memory of Vassil Popov on the occasion of his 60th birthday.
- [BEK93] F. Bornemann, B. Erdmann, and R. Kornhuber. Adaptive multilevel methods in three space dimensions. *Internat. J. Numer. Methods Engrg.*, 36(18):3187–3203, 1993. DOI: [10.1002/nme.1620361808](https://doi.org/10.1002/nme.1620361808).
- [BFM<sup>+</sup>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).
- [Bla02] R. Blaheta. GPCG-generalized preconditioned CG method and its use with non-linear and non-symmetric displacement decomposition preconditioners. In volume 9, number 6-7. 2002. DOI: [10.1002/nla.295](https://doi.org/10.1002/nla.295).
- [BMP24] P. Bringmann, A. Miraçi, and D. Praetorius. Iterative solvers in adaptive FEM: Adaptivity yields quasi-optimal computational runtime. In *Error Control, Adaptive Discretizations, and Applications, Part 2*. Volume 59. Elsevier, 2024. DOI: [10.1016/bs.aams.2024.08.002](https://doi.org/10.1016/bs.aams.2024.08.002).
- [BP93] J. H. Bramble and J. E. Pasciak. New estimates for multilevel algorithms including the V-cycle. *Math. Comp.*, 60(202):447–471, 1993. DOI: [10.2307/2153097](https://doi.org/10.2307/2153097).
- [BPWX91] J. H. Bramble, J. E. Pasciak, J. P. Wang, and J. Xu. Convergence estimates for product iterative methods with applications to domain decomposition. *Math. Comp.*, 57(195):1–21, 1991. DOI: [10.2307/2938660](https://doi.org/10.2307/2938660).
- [BY93] F. Bornemann and H. Yserentant. A basic norm equivalence for the theory of multilevel methods. *Numer. Math.*, 64(4):455–476, 1993. DOI: [10.1007/BF01388699](https://doi.org/10.1007/BF01388699).

- [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).
- [CG22] G. Ciaramella and M. J. Gander. *Iterative methods and preconditioners for systems of linear equations*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, 2022. DOI: [10.1137/1.9781611976908](https://doi.org/10.1137/1.9781611976908).
- [CKNS08] J. M. Cascon, C. Kreuzer, R. H. Nochetto, and K. G. Siebert. Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.*, 46(5):2524–2550, 2008. DOI: [10.1137/07069047X](https://doi.org/10.1137/07069047X).
- [CNX12] L. Chen, R. H. Nochetto, and J. Xu. Optimal multilevel methods for graded bisection grids. *Numer. Math.*, 120(1):1–34, 2012. DOI: [10.1007/s00211-011-0401-4](https://doi.org/10.1007/s00211-011-0401-4).
- [CW93] X.-C. Cai and O. B. Widlund. Multiplicative Schwarz algorithms for some nonsymmetric and indefinite problems. *SIAM J. Numer. Anal.*, 30(4):936–952, 1993. DOI: [10.1137/0730049](https://doi.org/10.1137/0730049).
- [CZ10] L. Chen and C. Zhang. A coarsening algorithm on adaptive grids by newest vertex bisection and its applications. *J. Comput. Math.*, 28(6):767–789, 2010. DOI: [10.4208/jcm.1004-m3172](https://doi.org/10.4208/jcm.1004-m3172).
- [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).
- [DHM<sup>+</sup>21] D. A. Di Pietro, F. Hülsemann, P. Matalon, P. Mycek, U. Råde, and D. Ruiz. An  $h$ -multigrid method for hybrid high-order discretizations. *SIAM J. Sci. Comput.*, 43(5):S839–S861, 2021. DOI: [10.1137/20M1342471](https://doi.org/10.1137/20M1342471).
- [FFPS17] M. Feischl, T. Führer, D. Praetorius, and E. P. Stephan. Optimal additive Schwarz preconditioning for hypersingular integral equations on locally refined triangulations. *Calcolo*, 54(1):367–399, 2017. DOI: [10.1007/s10092-016-0190-3](https://doi.org/10.1007/s10092-016-0190-3).
- [FHMP26] T. Führer, P. Hilbert, A. Miraçi, and D. Praetorius. Adaptive finite element methods with optimally preconditioned GMRES guarantee optimal complexity. *in preparation*, 2026.
- [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).
- [GV13] G. H. Golub and C. F. Van Loan. *Matrix computations*. Johns Hopkins University Press, Baltimore, fourth edition, 2013.
- [GY99] G. H. Golub and Q. Ye. Inexact preconditioned conjugate gradient method with inner-outer iteration. *SIAM J. Sci. Comput.*, 21(4):1305–1320, 1999. DOI: [10.1137/S1064827597323415](https://doi.org/10.1137/S1064827597323415).
- [Hil25] P. Hilbert. *Geometric Multigrid Method with  $hp$ -Robust Contraction*. Master’s thesis, Institut für Analysis und Scientific Computing, TU Wien, 2025. URL: <https://doi.org/10.34726/hss.2025.127742>.
- [HS52] M. R. Hestenes and E. Stiefel. Methods of conjugate gradients for solving linear systems. *J. Research Nat. Bur. Standards*, 49:409–436, 1952.
- [HWZ12] R. Hiptmair, H. Wu, and W. Zheng. Uniform convergence of adaptive multigrid methods for elliptic problems and Maxwell’s equations. *Numer. Math. Theory Methods Appl.*, 5(3):297–332, 2012. DOI: [10.4208/nmtma.2012.m1128](https://doi.org/10.4208/nmtma.2012.m1128).
- [HZ09] R. Hiptmair and W. Zheng. Local multigrid in  $\mathbf{H}(\mathbf{curl})$ . *J. Comput. Math.*, 27(5):573–603, 2009. DOI: [10.4208/jcm.2009.27.5.012](https://doi.org/10.4208/jcm.2009.27.5.012).
- [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).

- [IP23] M. Innerberger and D. Praetorius. MooAFEM: an object oriented Matlab code for higher-order adaptive FEM for (nonlinear) elliptic PDEs. *Appl. Math. Comput.*, 442:Paper No. 127731, 2023. DOI: [10.1016/j.amc.2022.127731](https://doi.org/10.1016/j.amc.2022.127731).
- [Kel75] R. B. Kellogg. On the Poisson equation with intersecting interfaces. *Appl. Anal.*, 4:101–129, 1975. DOI: [10.1080/00036817408839086](https://doi.org/10.1080/00036817408839086).
- [Lio88] P.-L. Lions. On the Schwarz alternating method. I. In *First International Symposium on Domain Decomposition Methods for Partial Differential Equations (Paris, 1987)*. SIAM, Philadelphia, 1988. ISBN: 0-89871-220-3.
- [MPV20] A. Miraçi, J. Papež, and M. Vohralík. A multilevel algebraic error estimator and the corresponding iterative solver with  $p$ -robust behavior. *SIAM J. Numer. Anal.*, 58(5):2856–2884, 2020.
- [MPV21] A. Miraçi, J. Papež, and M. Vohralík. A-posteriori-steered  $p$ -robust multigrid with optimal step-sizes and adaptive number of smoothing steps. *SIAM J. Sci. Comput.*, 43(5):S117–S145, 2021.
- [Not00] Y. Notay. Flexible conjugate gradients. *SIAM J. Sci. Comput.*, 22(4):1444–1460, 2000. DOI: [10.1137/S1064827599362314](https://doi.org/10.1137/S1064827599362314).
- [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).
- [SMPZ08] J. Schöberl, J. M. Melenk, C. Pechstein, and S. Zangl. Additive Schwarz preconditioning for  $p$ -version triangular and tetrahedral finite elements. *IMA J. Numer. Anal.*, 28(1):1–24, 2008. DOI: [10.1093/imanum/dr1046](https://doi.org/10.1093/imanum/dr1046).
- [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).
- [Tra97] C. T. Traxler. An algorithm for adaptive mesh refinement in  $n$  dimensions. *Computing*, 59(2):115–137, 1997. DOI: [10.1007/BF02684475](https://doi.org/10.1007/BF02684475).
- [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).
- [WC06] H. Wu and Z. Chen. Uniform convergence of multigrid V-cycle on adaptively refined finite element meshes for second order elliptic problems. *Sci. China Ser. A*, 49(10):1405–1429, 2006. DOI: [10.1007/s11425-006-2005-5](https://doi.org/10.1007/s11425-006-2005-5).
- [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).
- [XCN09] J. Xu, L. Chen, and R. H. Nochetto. Optimal multilevel methods for  $H(\text{grad})$ ,  $H(\text{curl})$ , and  $H(\text{div})$  systems on graded and unstructured grids. In *Multiscale, nonlinear and adaptive approximation*. Springer, Berlin, 2009. ISBN: 978-3-642-03412-1. DOI: [10.1007/978-3-642-03413-8\\_14](https://doi.org/10.1007/978-3-642-03413-8_14).
- [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).

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

*Email address:* [ani.miraci@sorbonne-universite.fr](mailto: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](mailto:paula.hilbert@asc.tuwien.ac.at) (corresponding author)

*Email address:* [dirk.praetorius@asc.tuwien.ac.at](mailto:dirk.praetorius@asc.tuwien.ac.at)