hp-ROBUST MULTIGRID SOLVER ON LOCALLY REFINED MESHES FOR FEM DISCRETIZATIONS OF SYMMETRIC ELLIPTIC PDES

MICHAEL INNERBERGER⬤, ANI MIRAÇI⬤,
DIRK PRAETORIUS⬤ AND JULIAN STREITBERGER*⬤

Abstract. In this work, we formulate and analyze a geometric multigrid method for the iterative solution of the discrete systems arising from the finite element discretization of symmetric second-order linear elliptic diffusion problems. We show that the iterative solver contracts the algebraic error robustly with respect to the polynomial degree $p \geq 1$ and the (local) mesh size h. We further prove that the built-in algebraic error estimator which comes with the solver is hp-robustly equivalent to the algebraic error. The application of the solver within the framework of adaptive finite element methods with quasi-optimal computational cost is outlined. Numerical experiments confirm the theoretical findings.

1. INTRODUCTION

Numerical schemes for PDEs aim at approximating the solution $u^\star$ of the weak formulation with an error below a certain tolerance at minimal computational cost. Since the accuracy is spoiled by singularities, *e.g.*, in given data or domain geometry, adaptive finite element methods (AFEMs) employ the loop to obtain a sequence of meshes $\mathcal{T}_L$ that resolve such singularities. For a large class of problems, it is known that AFEM is *rate-optimal, i.e.*, one can construct an estimator $\eta_L(u_L^\star)$ from the exact Galerkin solution $u_L^\star$ for the discretization error $\|u^\star - u_L^\star\|$ that decreases with the largest possible rate with respect to the number elements in $\mathcal{T}_L$; see, *e.g.*, the seminal works [5, 13, 17, 26, 33] or the abstract overview [12] for h-adaptive FEM with fixed polynomial degree p.
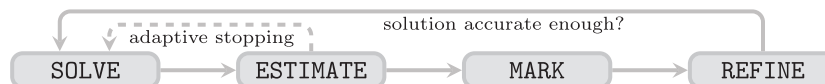
In practice, the SOLVE module may become computationally expensive (in contrast to all other modules) when employing a direct solver; see, *e.g.*, [18, 22, 29] for a discussion of implementational aspects. Thus, usually, an iterative solver is employed to compute an approximation u_L of $u_L^\star$ on each level, and the exact Galerkin solution $u_L^\star$ is not available. The question of whether the approximations u_L converge with optimal rate with respect to the *overall computational cost* was already treated in the seminal work [33] under some *realistic* assumptions about an abstract iterative solver. The recent work [18] employs nested iterations and an adaptive

---

stopping criterion to steer a uniformly contractive iterative solver, linking the `SOLVE` and `ESTIMATE` module in the above scheme by an inner loop. Then, it is shown that even the *full sequence of iterates* converges with optimal rates with respect to the overall computational cost. For this reason, the design of algebraic solvers that are uniformly contractive and robust with respect to the discretization parameters is of utmost importance.

The hierarchical structure of AFEM and the very nature of the arising linear systems suggest to use a *multilevel* solver; see, *e.g.*, [4, 6–8, 19, 27, 30, 38]. Different adaptive methods integrating a multilevel solver are possible; see, *e.g.*, [3] for generating local meshes, and [31] for a fully adaptive multigrid method that steers the local refinement process. In the context of AFEM, the adaptively constructed hierarchy of locally refined meshes calls for suitable *local* solvers. We refer to [14] for a multilevel preconditioner on a mesh hierarchy consisting of one bisection in each step and [21, 37] for multiplicative multigrid methods, all of which are robust with respect to the mesh size h. Though these works allow for higher-order FEM, an analytic and numerical study on the behavior with increasing polynomial degree was not presented. This aspect is treated, *e.g.*, in [1, 9, 28, 32], which design iterative solvers that are robust with respect to the polynomial degree p on various types of polyhedral meshes. The recent own work [25] proposes a p-robust geometric multigrid which comes with a *built-in algebraic error estimator* $\zeta_L(u_L)$, which is suited perfectly for *a posterori* steering (*i.e.*, adaptive termination of the algebraic solver). However, the employed patchwise smoothing associated to every vertex and every level causes a linear dependence on the number of adaptive mesh levels L.

In the present work, we modify the solver from [25] and overcome this dependence for locally refined meshes: we only apply *local* lowest-order smoothing on patches which change in the refinement step on intermediate levels, whereas a patchwise (and hence parallelizable) higher-order smoothing on all patches of the finest level is applied. This solver only needs one post-smoothing step, requires no symmetrization of the procedure (see also [15]), and, in particular, has no tunable parameters since it utilizes optimal step-sizes on the error-correction stage. As the main result of the present work, we show that the proposed solver *uniformly contracts* the algebraic error $\|u_L^\star - u_L\|$. Moreover, it comes with a built-in estimator $\zeta_L(u_L)$, which is shown to be *equivalent* to $\|u_L^\star - u_L\|$. Throughout, all involved estimates are robust in the discretization parameters h and p.

As one potential application, we formulate an AFEM algorithm in the spirit of [18] that naturally embeds the multigrid solver and leverages the solver's built-in algebraic error estimator $\zeta_L(u_L)$ to stop the solver as soon as the discretization and algebraic error are comparable. Adapting the arguments of [18], we prove that, for fixed polynomial degree p, the AFEM algorithm guarantees optimal convergence rates with respect to overall computational cost.

Using the open-source object-oriented 2D MATLAB code MooAFEM [22], we present a detailed numerical study of both the algebraic solver and the adaptive algorithm, including higher-order experiments and jumping coefficients.

The outline of this paper reads as follows: Section 2 first poses the model problem and introduces some notation. Then, we state the proposed multigrid solver (Algorithm 2.1) and formulate our main results on hp-robust contraction (Thm. 2.5) and algebraic error control (Cor. 2.6). As a potential application, Section 3 formulates an AFEM algorithm (Algorithm 3.1) which employs nested iteration and an adaptive stopping criterion for the iterative solver using the built-in *a posteriori* estimator for the algebraic error. Theorem 3.2 proves optimal computational complexity of the proposed AFEM algorithm. After we confirm the theoretical results by numerical examples in Section 4, we present proofs of the main results in Section 5. For better readability, we precede these proofs with three subsections presenting their core arguments: geometric properties of the meshes $\mathcal{T}_L$, an hp-robust stable decomposition combining a local lowest-order multilevel stable decomposition from [37] with a one-level p-robust decomposition from [32], and a strengthened Cauchy–Schwarz inequality in the spirit of [14, 21].

2. hp-ROBUST MULTIGRID SOLVER

In this section, we formulate the model problem, the proposed geometric multigrid method, and the main results, while the proofs are postponed to Section 5.

2.1. Model problem

For $d \in \{1, 2, 3\}$, let $\Omega \subset \mathbb{R}^d$ be a bounded Lipschitz domain with polygonal boundary $\partial\Omega$. Given $f \in L^2(\Omega)$ and $\boldsymbol{f} \in [L^2(\Omega)]^d$, we consider the second-order linear elliptic diffusion problem

$$-\operatorname{div}(\boldsymbol{K}\nabla u^\star) = f - \operatorname{div}\boldsymbol{f} \qquad \text{in } \Omega,$$
$$u^\star = 0 \qquad \text{on } \partial\Omega, \tag{1}$$

where $\boldsymbol{K} \in [L^\infty(\Omega)]_{\text{sym}}^{d \times d}$ is the symmetric and uniformly positive definite diffusion coefficient. More precisely, given a conforming simplicial triangulation $\mathcal{T}_h$ of Ω into compact simplices, we have $\boldsymbol{K}|_T \in [W^{1,\infty}(T)]^{d \times d}$ for all $T \in \mathcal{T}_h$. For $x \in \Omega$ we denote the maximal and minimal eigenvalue of $\boldsymbol{K}(x) \in \mathbb{R}_{\text{sym}}^{d \times d}$ by $\lambda_{\max}(\boldsymbol{K}(x))$ and $\lambda_{\min}(\boldsymbol{K}(x))$, respectively, and define $\Lambda_{\max} := \operatorname{ess\,sup}_{x \in \Omega} \lambda_{\max}(\boldsymbol{K}(x))$ as well as $\Lambda_{\min} := \operatorname{ess\,inf}_{x \in \Omega} \lambda_{\min}(\boldsymbol{K}(x))$. With $\langle \cdot, \cdot \rangle_\omega$ denoting the usual $L^2(\omega)$-scalar product for a measurable subset $\omega \subseteq \Omega$, the weak formulation of (1) seeks $u^\star \in \mathbb{V} := H_0^1(\Omega)$ solving

$$\langle\!\langle u^\star, v \rangle\!\rangle_\Omega := \langle \boldsymbol{K}\nabla u^\star, \nabla v \rangle_\Omega = \langle f, v \rangle_\Omega + \langle \boldsymbol{f}, \nabla v \rangle_\Omega =: F(v) \quad \text{for all } v \in \mathbb{V}. \tag{2}$$

We note that $\langle\!\langle \cdot, \cdot \rangle\!\rangle_\Omega$ is a scalar product and the induced semi-norm $\|u\|_\Omega^2 := \langle\!\langle u, u \rangle\!\rangle_\Omega$ is an equivalent norm on $\mathbb{V}$. Therefore, the Lax–Milgram lemma yields existence and uniqueness of the weak solution $u^\star \in \mathbb{V}$. For $\omega = \Omega$, we omit the index ω throughout.

To discretize (2), denote for a polynomial degree $p \geq 1$ and a triangle $T \in \mathcal{T}_h$ the space of all polynomials on T of degree at most p with $\mathbb{P}^p(T)$ and define

$$\mathbb{S}^q(\mathcal{T}_h) := \left\{ v_h \in C(\Omega) : v_h|_T \in \mathbb{P}^q(T) \text{ for all } T \in \mathcal{T}_h \right\} \quad \text{with } q \in \{1, p\}. \tag{3}$$

With the definition $\mathbb{V}_h^p := \mathbb{S}_0^p(\mathcal{T}_h) := \mathbb{S}^p(\mathcal{T}_h) \cap H_0^1(\Omega)$, the discrete problem consists of finding $u_h^\star \in \mathbb{V}_h^p$ such that

$$\langle\!\langle u_h^\star, v_h \rangle\!\rangle = F(v_h) \quad \text{for all } v_h \in \mathbb{V}_h^p. \tag{4}$$

Clearly, the formulation of the discrete problem (4) hinges on the choice of the mesh $\mathcal{T}_h$, which directly influences the quality of $u_h^\star$ as an approximation of $u^\star$. Note that (4) can be rewritten as a symmetric and positive definite linear system by introducing a basis of $\mathbb{V}_h^p$. However, we opt to work instead with the functional basis-independent description.

2.2. Mesh and space hierarchy

We suppose that the refinement strategy in the module REFINE is newest vertex bisection (NVB); see, *e.g.*, [34, 35] and Figure 1 for an illustration in 2D. Let $\mathcal{T}_0$ be the conforming initial mesh. We refer to [34] for NVB with admissible $\mathcal{T}_0$ and $d \geq 2$, to [23] for NVB with general $\mathcal{T}_0$ for $d = 2$, and to the recent work [16] for NVB with general $\mathcal{T}_0$ in any dimension. Throughout, we suppose that $\mathcal{T}_0$ is admissible. In the 1D case, Aurada *et al.* [2] splits each element into two children of half-length and additionally ensures that any two neighboring elements have uniformly comparable diameter. Let $\mathbb{T} := \mathbb{T}(\mathcal{T}_0)$ be the set of all refinements of $\mathcal{T}_0$ that can be obtained by arbitrarily many steps of NVB.

From now on, suppose that we are given a sequence $\{\mathcal{T}_\ell\}_{\ell=0}^L \subset \mathbb{T}$ of successively refined triangulations, *i.e.*, for all $\ell = 1, \dots, L$, it holds that $\mathcal{T}_\ell = \text{REFINE}(\mathcal{T}_{\ell-1}, \mathcal{M}_{\ell-1})$ is the coarsest conforming triangulation obtained by NVB, where all marked elements $\mathcal{M}_{\ell-1} \subseteq \mathcal{T}_{\ell-1}$ have been refined by (at least) one bisection. We note that NVB-refinement generates meshes that are uniformly γ-shape regular, *i.e.*,

$$\max_{\ell=0,\dots,L} \max_{T \in \mathcal{T}_\ell} \frac{\operatorname{diam}(T)}{|T|^{1/d}} \leq \gamma < \infty, \tag{5a}$$

and

$$\max_{\ell=0,\dots,L} \max_{T \in \mathcal{T}_\ell} \max_{\substack{T' \in \mathcal{T}_\ell \\ T \cap T' \neq \emptyset}} \frac{\operatorname{diam}(T)}{\operatorname{diam}(T')} \leq \gamma < \infty. \tag{5b}$$
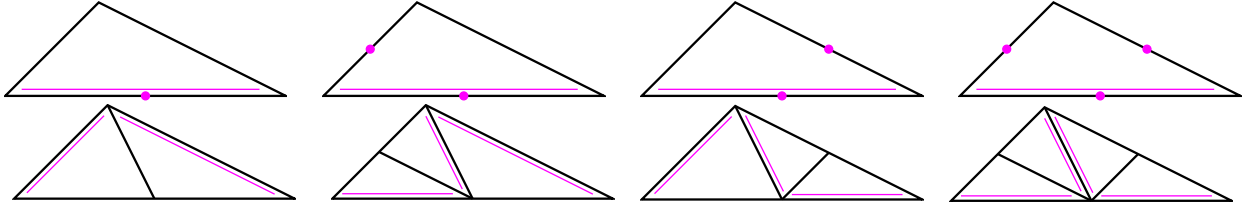
FIGURE 1. Schematic of 2D NVB refinement pattern: For each triangle $T \in \mathcal{T}$, there is one fixed *refinement edge* E_T indicated by the extra pink line. The pink dots indicate edges that are marked for refinement. If an element is marked for refinement, at least its refinement edge is marked for refinement (*top*). Iterated bisection refines a marked element into 2, 3, or 4 children (*bottom*).
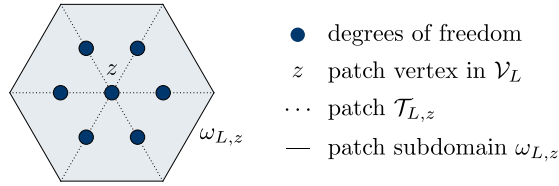

FIGURE 2. Illustration of degrees of freedom ($p = 2$) for the space $\mathbb{V}^p_{L,z}$ associated to the patch $\mathcal{T}_{L,z}$.

where γ depends only on $\mathcal{T}_0$ and is, in particular, independent of L and the meshes $\mathcal{T}_1, \ldots, \mathcal{T}_L$; see, *e.g.*, Theorem 2.1 of [34] for $d \geq 2$ or [2] for $d = 1$. We note that (5a) implies (5b) for $d \geq 2$, while (5a) is trivial with $\gamma = 1$ and independent of (5b) for $d = 1$. In addition, we define the quasi-uniformity constant

$$C_{\mathrm{qu}} \coloneqq \min \left\{ \mathrm{diam}(T)/\mathrm{diam}(T') \,:\, T, T' \in \mathcal{T}_0 \right\} \in (0, 1]. \tag{6}$$

For each mesh $\mathcal{T}_\ell$, let $\mathcal{V}_\ell$ denote the set of vertices. Given a vertex $z \in \mathcal{V}_\ell$, we denote by $\mathcal{T}_{\ell,z} \coloneqq \left\{ T \in \mathcal{T}_\ell \,:\, z \in T \right\}$ the patch of elements of $\mathcal{T}_\ell$ that share the vertex z. The corresponding (open) patch subdomain is denoted by $\omega_{\ell,z} \coloneqq \mathrm{interior}\left(\bigcup_{T \in \mathcal{T}_{\ell,z}} T \right)$ and its size by $h_{\ell,z} \coloneqq \max_{T \in \mathcal{T}_{\ell,z}} h_T \coloneqq \max_{T \in \mathcal{T}_{\ell,z}} |T|^{1/d}$. Finally, we denote by $\mathcal{V}^+_\ell$ the set of new vertices in $\mathcal{T}_\ell$ and the pre-existing vertices of $\mathcal{T}_{\ell-1}$ whose associated patches have shrunk in size in the refinement step ℓ, *i.e.*,

$$\mathcal{V}^+_0 \coloneqq \mathcal{V}_0 \quad \text{and} \quad \mathcal{V}^+_\ell \coloneqq \mathcal{V}_\ell \setminus \mathcal{V}_{\ell-1} \cup \left\{ z \in \mathcal{V}_\ell \cap \mathcal{V}_{\ell-1} \,:\, \omega_{\ell,z} \neq \omega_{\ell-1,z} \right\} \quad \text{for } \ell \geq 1.$$

While this notation is used in the analysis of the solver below, the presentation of Algorithm 2.1 is more compact with the abbreviation $\mathcal{N}_\ell = \mathcal{V}^+_\ell$ for $\ell = 1, \ldots, L-1$ and $\mathcal{N}_L \coloneqq \mathcal{V}^+_L$ for $p = 1$ and $\mathcal{N}_L \coloneqq \mathcal{V}_L$ otherwise, where we recall that $p \in \mathbb{N}$ is the fixed polynomial degree of the FEM ansatz functions.

From the definition of the discrete FEM spaces (3) and NVB-refinement, we see that there holds nestedness

$$\mathbb{V}^1_0 \subseteq \mathbb{V}^1_1 \subseteq \cdots \subseteq \mathbb{V}^1_{L-1} \subseteq \mathbb{V}^p_L. \tag{7}$$

Furthermore, we require the local spaces

$$\mathbb{V}^q_{\ell,z} \coloneqq \mathbb{S}^q_0(\mathcal{T}_{\ell,z}) \quad \text{for all vertices } z \in \mathcal{V}_\ell \text{ and } q \in \{1, p\}, \tag{8}$$

where we use $q = 1$ for $\ell = 0, \ldots, L-1$ and $q = p$ for $\ell = L$; see Figure 2 for the illustration of the degrees of freedom for $p = 2$.

2.3. Multigrid solver

In the following, we introduce a local geometric multigrid method, which will serve as iterative solver within the SOLVE module of an adaptive FEM algorithm. Each full step of the proposed multigrid method can be mathematically described by an iteration operator $\Phi\colon \mathbb{V}_L^p \to \mathbb{V}_L^p$, *i.e.*, given the current approximation $u_L \in \mathbb{V}_L^p$, the solver generates the new iterate $\Phi(u_L) \in \mathbb{V}_L^p$.

The main ingredients in the solver construction are an inexpensive global residual solve on $\mathcal{T}_0$ and local residual solves on all patches $\mathcal{T}_{\ell,z}$ for $z \in \mathcal{V}_\ell^+$ on the intermediate levels $\ell = 1, \ldots, L-1$ and all patches on the finest level $\mathcal{T}_L$ when $p > 1$. For ease of notation, we define the algebraic residual functional $R_L\colon \mathbb{V}_L^p \to \mathbb{R}$ by

$$v_L \in \mathbb{V}_L^p \mapsto R_L(v_L) := F(v_L) - \langle\!\langle u_L,\, v_L \rangle\!\rangle = \langle\!\langle u_L^\star - u_L,\, v_L \rangle\!\rangle \in \mathbb{R}. \tag{9}$$

To construct the new iterate $\Phi(u_L)$, levelwise residual liftings of the algebraic error are added to the current approximation u_L. The same levelwise residual liftings are used to define an *a posteriori* error estimator $\zeta_L(u_L)$ for the algebraic error, *i.e.*, the solver comes with a built-in estimator.

Algorithm 2.1 (One step of the optimal local multigrid solver). **Input:** current approximation $u_L \in \mathbb{V}_L^p$, meshes $\{\mathcal{T}_\ell\}_{\ell=0}^L$, polynomial degree $p \in \mathbb{N}$.
Solver step: perform the following steps (i) and (ii):

(i) **Global lowest-order residual problem on the coarsest level:**
   – Compute $\rho_0 \in \mathbb{V}_0^1$ by solving

$$\langle\!\langle \rho_0,\, v_0 \rangle\!\rangle = R_L(v_0) \qquad \text{for all } v_0 \in \mathbb{V}_0^1. \tag{10}$$

   – Define step-size $\lambda_0 := 1$.
   – Initialize algebraic lifting $\sigma_0 := \lambda_0 \rho_0$ and *a posteriori* estimator $\zeta_0^2 := \|\!|\lambda_0 \rho_0|\!\|^2$.

(ii) **Local residual-update:** for all $\ell = 1, \ldots, L$, do the following steps, where $q = 1$ for $\ell = 1, \ldots, L-1$ and $q = p$ for $\ell = L$:
   – For all $z \in \mathcal{N}_\ell$, compute $\rho_{\ell,z} \in \mathbb{V}_{\ell,z}^q$ by solving

$$\langle\!\langle \rho_{\ell,z},\, v_{\ell,z} \rangle\!\rangle = R_L(v_{\ell,z}) - \langle\!\langle \sigma_{\ell-1},\, v_{\ell,z} \rangle\!\rangle \qquad \text{for all } v_{\ell,z} \in \mathbb{V}_{\ell,z}^q. \tag{11}$$

   – Define the line-search step-size $s_\ell := (R_L(\rho_\ell) - \langle\!\langle \sigma_{\ell-1},\, \rho_\ell \rangle\!\rangle)/\|\!|\rho_\ell|\!\|^2$, with $\rho_\ell := \sum_{z \in \mathcal{N}_\ell} \rho_{\ell,z}$ and the understanding that $0/0 := 0$ if $\rho_\ell = 0$, and

$$\lambda_\ell := \begin{cases} s_\ell & \text{if } s_\ell \leq d+1 \ \text{ or } \ \big[\ell = L \text{ and } p > 1\big], \\ (d+1)^{-1} & \text{otherwise.} \end{cases}$$

   – Update $\sigma_\ell := \sigma_{\ell-1} + \lambda_\ell \rho_\ell$ and $\zeta_\ell^2 := \zeta_{\ell-1}^2 + \lambda_\ell \sum_{z \in \mathcal{N}_\ell} \|\!|\rho_{\ell,z}|\!\|^2$.

Output: improved approximation $\Phi(u_L) := u_L + \sigma_L \in \mathbb{V}_L^p$ and associated *a posteriori* estimator $\zeta_L(u_L) := \zeta_L$ of the algebraic error.

Remark 2.2 (Construction of the new iterate). The construction of $\Phi(u_L)$ from u_L by Algorithm 2.1 can be seen as one iteration of a V-cycle multigrid with no pre- and one post-smoothing step, and a step-size ar the error correction stage. The smoother on each level is additive Schwarz associated to patch subdomains where the local problems (11) are defined. This is equivalent to diagonal Jacobi smoothing for $p = 1$ (*e.g.*, on intermediate levels) and block-Jacobi smoothing for $p > 1$ (*e.g.*, on the finest level). The choice and use of the step-sizes λ_ℓ in Algorithm 2.1 (ii) comes from a line-search approach; see, *e.g.*, Lemma 4.3 of [25] and one of the earlier works [20]. However, if the step-size from the line-search is too large, we use instead a fixed damping parameter offsetting the $d+1$ patch overlaps. We note that this case never occurred in practice in any of our numerical experiments.

Remark 2.3 (Computational effort and speed of convergence)**.** We note that we apply a patchwise Cholesky factorization on the finest level. Hence, the computational effort for the local residual solve on the finest mesh $\mathcal{T}_L$ in dependence on the polynomial degree p is of order $\mathcal{O}(p^{3d}\#\mathcal{T}_L)$. The presented algorithm is a linear method. One could symmetrize the procedure by adding one pre-smoothing step to define a preconditioner in the hope of accelerating convergence with the help of conjugate gradients. However, in our experience, the patchwise pre-smoothing typically did not yield considerable algebraic error decrease; see, *e.g.*, [15], while still doubling the number of smoothing operations of a V-cycle. The remaining steps needed to compute the new approximation stem from classical multigrid solvers (such as intergrid operators). We stress that the overall effort does not depend on the number of levels L.

Remark 2.4 (Nested iterations)**.** In the context of adaptive FEM, the solver does not start from an arbitrary initial guess on each newly-refined mesh but from the final approximation of the previous level (see Algorithm 3.1 below). This will ensure *a posteriori* error control in each step after initialization as well as optimal computational cost. From the algebraic solver perspective, such an approach can be seen as a full multigrid method over the evolving hierarchy of meshes whose number of cycles is determined by the adaptive stopping criterion.

2.4. Main result

This subsection formulates the main results regarding the iterative solver stating the *contraction* of the multigrid solver and *reliability* of the built-in *a posteriori* estimator of the algebraic error. Both results hold *robustly* in the number of levels L and the polynomial degree p.

Theorem 2.5. *Let $u_L^\star \in \mathbb{V}_L^p$ be the (unknown) finite element solution of* (4) *and let $v_L \in \mathbb{V}_L^p$ be arbitrary. Let $\Phi(v_L) \in \mathbb{V}_L^p$ and $\zeta_L(v_L)$ be generated by Algorithm 2.1. Then, the solver iterates and the estimator are connected by*

$$\|\|u_L^\star - \Phi(v_L)\|\|^2 \leq \|\|u_L^\star - v_L\|\|^2 - \zeta_L(v_L)^2. \tag{12}$$

Moreover, the solver contracts the error, i.e., there exists $0 < q_{\mathrm{ctr}} < 1$ such that

$$\|\|u_L^\star - \Phi(v_L)\|\| \leq q_{\mathrm{ctr}} \|\|u_L^\star - v_L\|\|. \tag{13}$$

Finally, the estimator is a two-sided bound of the algebraic error, i.e., there exists $C_{\mathrm{rel}} > 1$ such that

$$\zeta_L(v_L) \leq \|\|u_L^\star - v_L\|\| \leq C_{\mathrm{rel}}\, \zeta_L(v_L). \tag{14}$$

The contraction and reliability constants q_{ctr} and C_{rel} depend only on the space dimension d, the γ-shape regularity (5)*, the quasi-uniformity constant C_{qu} from* (6)*, $\Lambda_{\max}/\Lambda_{\min}$, and $\max_{T \in \mathcal{T}_L} \|\mathrm{div}(\boldsymbol{K})\|_{L^\infty(T)}/\Lambda_{\min}$. In particular, q_{ctr} is independent of the polynomial degree p, the number of mesh levels L, and the meshes $\mathcal{T}_1, \ldots, \mathcal{T}_L$.*

Corollary 2.6. *The reliability of the estimator in* (14) *is equivalent to the solver contraction* (13)*. In particular, this also yields that*

$$\|\|u_L^\star - \Phi(v_L)\|\| \leq q_{\mathrm{ctr}}\, C_{\mathrm{rel}}\, \zeta_L(v_L). \tag{15}$$

Remark 2.7. We note that (12) holds with equality whenever the step-size criterion $s_\ell \leq d + 1$ in Algorithm 2.1(ii) are fulfilled and the construction is thus done by optimal-line search. In such a case, which was always satisfied in all our numerical tests, a Pythagoras identity in the spirit of Theorem 4.7 from [25] yielding exact algebraic error decrease is obtained.

3. APPLICATION TO ADAPTIVE FEM WITH INEXACT SOLVER

Given a coarse mesh $\mathcal{T}_0$, we use an adaptive finite element method (AFEM) to generate locally refined meshes $\{\mathcal{T}_L\}_{L\in\mathbb{N}}$ tailored to the behavior of the sought solution. In the spirit of [18], Algorithm 3.1 presents such an approach with an adaptively stopped iterative solver, where Step (Ii) exploits the built-in *a posteriori* estimator of the geometric multigrid solver from Section 2.

While we note that the present Algorithm 3.1 and the corresponding Theorem 3.2 are restricted to fixed polynomial degree p, the inclusion of the proposed hp-robust iterative solver into the hp-adaptive FEM algorithm of [11] remains for future research, since the mathematical understanding of hp-adaptive FEM is still widely open.

Algorithm 3.1 (AFEM with iterative solver).
Input: initial mesh $\mathcal{T}_0$, polynomial degree $p \in \mathbb{N}$, adaptivity parameters $0 < \theta \leq 1$, $C_{\mathrm{mark}} \geq 1$, and $\mu > 0$, initial guess $u_0^0 := 0$.
Adaptive loop: repeat the following steps (I)–(III) for all $L = 0, 1, 2, \dots$:

(I) `SOLVE & ESTIMATE`: repeat the following steps (i)–(iii) for all $k = 1, 2, 3, \dots$:
   (i) Do one step of the algebraic solver to obtain $u_L^k \in \mathbb{V}_L^p$ and an associated *a posteriori* estimator $\zeta_L(u_L^{k-1})$ for the algebraic error

$$\big[u_L^k, \zeta_L(u_L^{k-1})\big] := \mathtt{SOLVE}\Big(u_L^{k-1}, \{\mathcal{T}_\ell\}_{\ell=0}^L, p\Big).$$

   (ii) Compute *a posteriori* indicators for the elementwise discretization error

$$\big\{\eta_L(T, u_L^k)\big\}_{T \in \mathcal{T}_L} := \mathtt{ESTIMATE}\big(u_L^k, \mathcal{T}_L\big).$$

   (iii) If $\zeta_L(u_L^{k-1}) \leq \mu \eta_L(u_L^k)$, terminate the k-loop, set the index $\underline{k}[L] := k$ and define $u_L := u_L^{k[L]}$.
(II) `MARK`: determine a set of marked elements $\mathcal{M}_L \subseteq \mathcal{T}_L$ of (up to the multiplicative constant C_{mark}) minimal cardinality that satisfies

$$\theta \, \eta_L(u_L)^2 \leq \sum_{T \in \mathcal{M}_L} \eta_L(T, u_L)^2.$$

(III) `REFINE`: generate the new mesh $\mathcal{T}_{L+1} := \mathtt{REFINE}(\mathcal{M}_L, \mathcal{T}_L)$ and define $u_{L+1}^0 := u_L$.

Output: sequences of successively refined triangulations $\mathcal{T}_L$, discrete approximations u_L and corresponding error estimators $(\eta_L(u_L), \zeta_L(u_L))$.

Mesh-refinement is steered by the discretization error estimator. For all $T \in \mathcal{T}_h$, let $\eta_h(T; \cdot) \colon \mathbb{V}_h^p \to \mathbb{R}_{\geq 0}$ be the local contributions of the standard residual error estimator defined by

$$\eta_h^2(T; v_h) := h_T^2 \|f + \mathrm{div}(\boldsymbol{K}\nabla v_h - \boldsymbol{f})\|_T^2 + h_T \|[\![\boldsymbol{K}\nabla v_h - \boldsymbol{f}]\!] \cdot \boldsymbol{n}\|_{\partial T \cap \Omega}^2, \tag{16}$$

where $\|\cdot\|_\omega$ denote the appropriate $L^2(\omega)$-norms. We define

$$\eta_h(\mathcal{U}_h; v_h) := \left(\sum_{T \in \mathcal{U}_h} \eta_h(T; v_h)^2\right)^{1/2} \quad \text{for all} \quad \mathcal{U}_h \subseteq \mathcal{T}_h \text{ and } v_h \in \mathbb{V}_h^p.$$

To abbreviate notation, let $\eta_h(v_h) := \eta_h(\mathcal{T}_h; v_h)$.

One important consequence of Theorem 2.5 is optimal convergence of Algorithm 3.1 with respect to computational complexity. To formulate this mathematically, we define the ordered set

$$\mathcal{Q} := \big\{(L, k) \in \mathbb{N}_0^2 : \text{index tuple } (L, k) \text{ is used as loop variable in Algorithm 3.1 and } 1 \leq k \leq \underline{k}[L]\big\}.$$

On $\mathcal{Q}$, we define the ordering $\leq$ by

$$(L', k') \leq (L, k) \quad \iff \quad u_{L'}^{k'} \text{ is computed earlier than or equal to } u_L^k \text{ in Algorithm 3.1.}$$

Furthermore, we introduce the total step counter $|\cdot, \cdot|$, defined for all $(L, k) \in \mathcal{Q}$, by

$$|L, k| := \#\big\{(L', k') \in \mathcal{Q} \,:\, (L', k') \leq (L, k)\big\}.$$

Before we state the theorem, we introduce the notion of approximation classes. For $s > 0$, define

$$\|u\|_{\mathbb{A}_s} := \sup_{N \in \mathbb{N}_0}\left((N+1)^s \min_{\mathcal{T}_{\mathrm{opt}} \in \mathbb{T}_N(\mathcal{T}_0)}\big(\|u^\star - u_{\mathrm{opt}}^\star\| + \eta_{\mathrm{opt}}\big(u_{\mathrm{opt}}^\star\big)\big)\right), \tag{17}$$

with Galerkin solution $u_{\mathrm{opt}}^\star$ and estimator η_{opt} on the optimal triangulation $\mathcal{T}_{\mathrm{opt}} \in \mathbb{T}_N(\mathcal{T}_0)$, where $\mathbb{T}_N(\mathcal{T}_0) := \big\{\mathcal{T}_h \in \mathbb{T}(\mathcal{T}_0) \,:\, \#\mathcal{T}_h - \#\mathcal{T}_0 \leq N\big\}$. By reliability (A3) of the estimator, see, *e.g.*, [12], the sum on the right-hand side of (17) is equivalent to $\eta_{\mathrm{opt}}(u_{\mathrm{opt}}^\star)$. If $\|u\|_{\mathbb{A}_s} < \infty$, then we say that rate s is possible.

In [18], it is shown that in the case of a contractive solver, convergence rates with respect to degrees of freedom are equivalent to convergence rates with respect to computational complexity. We abbreviate with $\mathrm{cost}(L, k)$ the total costs of Algorithm 3.1 defined by

$$\mathrm{cost}(L, k) := \sum_{\substack{(L', k') \in \mathcal{Q} \\ (L', k') \leq (L, k)}} \#\mathcal{T}_{L'}.$$

Theorem 3.2. *Let $\{\mathcal{T}_L\}_{L \in \mathbb{N}_0}$ be the sequence generated by Algorithm 3.1 and define the quasi-error by*

$$\Delta_L^k := \big\|\|u^\star - u_L^k\|\big\| + \eta_L(u_L^k) \quad \text{for all } (L, k) \in \mathcal{Q}.$$

Then, for all parameters $0 < \theta \leq 1$ and $\mu > 0$, it holds that

$$\sup_{(L,k) \in \mathcal{Q}} (\#\mathcal{T}_L)^s \Delta_L^k \simeq \sup_{(L,k) \in \mathcal{Q}} \mathrm{cost}(L, k)^s \Delta_L^k \quad \text{and} \quad \Delta_L^k \to 0 \quad \text{as} \quad |L, k| \to \infty. \tag{18}$$

Furthermore, there exist $0 < \theta^\star \leq 1$, and $\mu^\star > 0$ such that, for sufficiently small parameters $0 < \theta < \theta^\star$ and $0 < \mu/\theta < \mu^\star$, and for all $s > 0$, it holds that

$$c_{\mathrm{opt}}\, \|u\|_{\mathbb{A}_s} \leq \sup_{(L,k) \in \mathcal{Q}} \mathrm{cost}(L, k)^s\, \Delta_L^k \leq C_{\mathrm{opt}}\, \max\big\{\|u\|_{\mathbb{A}_s}, \Delta_0^0\big\}. \tag{19}$$

The constants $c_{\mathrm{opt}}, C_{\mathrm{opt}} > 0$ depend only on the polynomial degree p, the initial triangulation $\mathcal{T}_0$, $\Lambda_{\max}/\Lambda_{\min}$, $\max_{T \in \mathcal{T}_L} \|\mathrm{div}(\boldsymbol{K})\|_{L^\infty(T)}/\Lambda_{\min}$, the rate s, the ratios $\theta/\theta^\star$ and $\mu/(\theta\mu^\star)$, and the properties of newest vertex bisection. In particular, this proves the equivalence

$$\|u\|_{\mathbb{A}_s} < \infty \quad \iff \quad \sup_{(L,k) \in \mathcal{Q}} \mathrm{cost}(L, k)^s\, \Delta_L^k < \infty, \tag{20}$$

which proves optimal complexity of Algorithm 3.1.

Remark 3.3. We note that in Theorem 8 of [18], the constant $c_{\mathrm{opt}} > 0$ additionally depends on the stopping index $\underline{L}$ in the case the algorithm terminates after a finite number of mesh levels $\underline{L} < \infty$ or the estimator satisfies $\eta_L(u_L) = 0$. The refined analysis in the recent work [10] removes this dependence.

Remark 3.4. We note that it is also possible to use the same stopping criterion for the algebraic solver as in Algorithm 2 of [18]. However, since the multigrid solver from Algorithm 2.1 has a built-in estimator for the algebraic error, we opt for its choice within Algorithm 3.1 instead.

Proof of Theorem 3.2. We show that Algorithm 3.1 satisfies the requirements of Theorems 4 and 8 from [18]. First note that the standard residual error estimator from (16) satisfies the axioms of adaptivity from [12] and thus satisfies the assumptions (A1)–(A4) from Theorem 8 of [18]. Furthermore, newest vertex bisection satisfies the assumptions (R1)–(R3) from Section 2.2 of [18]. For the present setting, the conditions (C1) and (C2) from Section 2.5 of [18] coincide and are satisfied.

Tracing the role of the stopping criterion for the case (C1) in the proof of Theorem 4 from [18], one sees that the stopping criterion needs to guarantee that, for all $(L, k) \in \mathcal{Q}$,

$$
\begin{aligned}
\left\|\left| u_L^k - u_L^{k-1} \right\|\right| &\leq \lambda_1 \, \eta_L\big(u_L^k\big) && \text{if } u_L^k = u_L, \\
\eta_L\big(u_L^k\big) &\leq \lambda_2^{-1} \left\|\left| u_L^k - u_L^{k-1} \right\|\right| && \text{else,}
\end{aligned}
\tag{21}
$$

for some $\lambda_1, \lambda_2 > 0$. The upper bound in (14) in Theorem 2.5 as well as contraction (13) show that, for all $(L, k) \in \mathcal{Q}$, our stopping criterion in Algorithm 3.1 Step (Iiii) leads to

$$
\left\|\left| u_L^k - u_L^{k-1} \right\|\right| \overset{(13)}{\leq} (1 + q_{\mathrm{ctr}}) \left\|\left| u_L^\star - u_L^{k-1} \right\|\right| \overset{(14)}{\leq} C_{\mathrm{rel}} (1 + q_{\mathrm{ctr}}) \, \zeta_L\big(u_L^{k-1}\big) \leq \mu \, C_{\mathrm{rel}} (1 + q_{\mathrm{ctr}}) \, \eta_L\big(u_L^k\big) \qquad \text{if } u_L^k = u_L.
$$

For the remaining case, the contraction (13) leads to

$$
\left\|\left| u_L^\star - u_L^k \right\|\right| \overset{(13)}{\leq} q_{\mathrm{ctr}} \left\|\left| u_L^\star - u_L^{k-1} \right\|\right| \leq q_{\mathrm{ctr}} \left\|\left| u_L^\star - u_L^k \right\|\right| + q_{\mathrm{ctr}} \left\|\left| u_L^k - u_L^{k-1} \right\|\right|.
$$

This implies

$$
\left\|\left| u_L^\star - u_L^k \right\|\right| \leq \frac{q_{\mathrm{ctr}}}{1 - q_{\mathrm{ctr}}} \left\|\left| u_L^k - u_L^{k-1} \right\|\right|.
\tag{22}
$$

The not met stopping criterion in Algorithm 3.1(Iiii), the lower bound in (14), and (22) show

$$
\begin{aligned}
\eta_L\big(u_L^k\big) < \mu^{-1} \, \zeta_L\big(u_L^{k-1}\big) &\overset{(14)}{\leq} \mu^{-1} \left\|\left| u_L^\star - u_L^{k-1} \right\|\right| \leq \mu^{-1} \left(\left\|\left| u_L^\star - u_L^k \right\|\right| + \left\|\left| u_L^k - u_L^{k-1} \right\|\right| \right) \\
&\overset{(22)}{\leq} \mu^{-1} \left(1 + \frac{q_{\mathrm{ctr}}}{1 - q_{\mathrm{ctr}}} \right) \left\|\left| u_L^k - u_L^{k-1} \right\|\right| = \mu^{-1} \left(\frac{1}{1 - q_{\mathrm{ctr}}} \right) \left\|\left| u_L^k - u_L^{k-1} \right\|\right|.
\end{aligned}
$$

Overall, (21) is satisfied with

$$
\lambda_1 = C_{\mathrm{rel}} (1 + q_{\mathrm{ctr}}) \, \mu \quad \text{and} \quad \lambda_2 = (1 - q_{\mathrm{ctr}}) \, \mu,
$$

and Theorem 4 of [18] proves full linear convergence, so that, in particular, equation (18) is fulfilled (see the proof of Theorem 8 from [18] or [10], Cor. 4.2).

The lower bound in (19) follows as in Theorem 8 from [18] or Theorem 4.3 from [10]. For the upper bound in (19), Theorem 8 from [18] requires that

$$
0 < \lambda_1/\theta < \lambda_{\mathrm{opt}} := (1 - q_{\mathrm{ctr}})/(q_{\mathrm{ctr}} \, C_{\mathrm{stab}})
$$

and

$$
0 < \theta' := \frac{\theta + \lambda_1/\lambda_{\mathrm{opt}}}{1 - \lambda_1/\lambda_{\mathrm{opt}}} < \theta_{\mathrm{opt}} := \big(1 + C_{\mathrm{stab}}^2 \, C_{\mathrm{drel}}^2\big)^{-1},
$$

where C_{stab} is the stability constant from (A1) and C_{drel} is the constant from discrete reliability (A4); see, *e.g.*, [18]. We define

$$
\mu^\star := \frac{\lambda_{\mathrm{opt}}}{C_{\mathrm{rel}} (1 + q_{\mathrm{ctr}})},
$$

and $\mu/\theta < \mu^\star$ thus implies $\lambda_1/\theta = C_{\mathrm{rel}} (1 + q_{\mathrm{ctr}}) \, \mu/\theta < \lambda_{\mathrm{opt}}$. Finally, we choose $\theta^\star$ such that any $0 < \widetilde{\theta} \leq \theta^\star$ satisfies $\frac{2\widetilde{\theta}}{1-\widetilde{\theta}} < \theta_{\mathrm{opt}}$. Then, $0 < \theta < \theta^\star$ yields $\theta' = \frac{\theta + \lambda_1/\lambda_{\mathrm{opt}}}{1 - \lambda_1/\lambda_{\mathrm{opt}}} < \frac{2\theta}{1-\theta} < \theta_{\mathrm{opt}}$ and optimal cost in Theorem 3.2 follows directly from Theorem 8 of [18]. □

4. Numerical experiments

This section investigates the numerical performance of the proposed multigrid solver of Algorithm 2.1 and the adaptive Algorithm 3.1. The MATLAB implementation of the multigrid solver is embedded into the MooAFEM[1] framework from [22]. Throughout, we choose the marking parameter $\theta = 0.5$ in the adaptive Algorithm 3.1 and $\boldsymbol{f} = (0,0)^\top$. We introduce the following test case:

– *L-shaped domain.* Let $\Omega = (-1,1)^2 \setminus \big([0,1] \times [-1,0]\big)$ with right-hand side $f = 1$ and $\boldsymbol{K} = \boldsymbol{I}$.

4.1. Contraction and performance of local multigrid solver

We confirm numerically our main results from Theorem 2.5. In order to study the algebraic solver and its built-in estimator with respect to different polynomial degrees, we take $\mu = 10^{-5}$ in Algorithm 3.1, thus *oversolving* the algebraic problem. Moreover, we stop the adaptive algorithm once the final mesh consists of revision 10^6 degrees of freedom. Note that thanks to Corollary 2.6 proving the equivalence of the reliability of the algebraic error estimator with the contraction of the algebraic solver, we indeed only need to investigate numerically the existence of the p-robust bound on the contraction of the solver. In Figure 3 (left), we present the maximal contraction factors on each level L of the adaptive algorithm from Algorithm 3.1. We see that the contraction factors are robust in the polynomial degree p with an upper bound of about 0.7 in all our experiments. In Figure 3 (right), we see that on a fixed number of levels ($L = 10$) even for higher-order polynomials their behavior is clustered around similar values. Moreover, from a purely solver-centric perspective, we see that the solver variant which employs higher-order smoothing also on the intermediate levels (and not only on the finest one) as studied in [25] only leads to slight improvements of the contraction constants. Adapting the arguments of [25], this modified construction can be guaranteed to be contractive with p-robust, but linearly L-dependent contraction bound on the algebraic error. However, this degradation with increasing L is not seen in practice, provided that the patchwise smoothing is done *everywhere* for *level $L = 1$* (as new degrees of freedom are added on all patches when the polynomial degree is $p > 1$) and *local* patchwise smoothing is employed in the remaining levels. We present a comparison of the resulting contraction factors of this approach to Algorithm 2.1 for a fixed number of level ($L = 10$) in Figure 3 (right).

4.2. Optimality of the adaptive algorithm

We take $\mu = 0.1$ in Algorithm 3.1 and study the decrease of the discretization error estimator $\eta_L(u_L)$, both in terms of number of degrees of freedom and timing. We remark that the error estimator $\eta_L(u_L)$ on the final iterates is equivalent to the quasi-error Δ_L. After a pre-asymptotic phase, we see in Figure 4 for different polynomial degrees p that the optimal convergence rate $-p/2$ is recovered both with respect to number of degrees of freedom and computational time, and the singularity at the reentrant corner $(0,0)$ is resolved through local mesh refinement. Furthermore, Figure 5 shows that the proposed multigrid solver behaves faster than the built-in direct solver (MATLAB backslash operator) concerning the time per dof. The displayed timings include the setup of the linear system, the time for the solver module, computation of estimator, and mesh refinement. Overall, the numerical experiments in Figure 5 validate the linear complexity of the suggested local multigrid solver from Algorithm 2.1.

4.3. Numerical performance and insights for jumping coefficients

We consider two additional test cases with jumps in the diffusion coefficient:

– *Checkerboard.* Let $\Omega = (0,1)^2$ be the unit square and $\boldsymbol{K}$ the 2×2 checkerboard diffusion with values 1 (white) and 10^k (grey) for fixed $k = 1,2,3$, see Figure 6 (left).
– *Striped diffusion.* Let $\Omega = (0,1)^2$ be the unit square split into 2^k stripes for $k = 1,2,3$. The value of $\boldsymbol{K}$ on the j-th stripe is 10^{j-1} with $j \in \{1, \ldots, 2^k\}$, see Figure 6 (right).

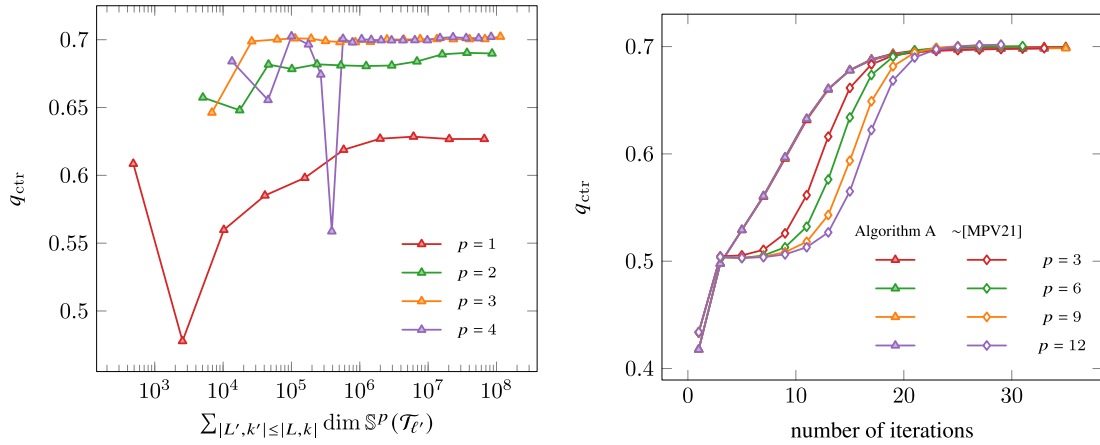[1]All experiments presented in this paper are reproducible with the openly available software package under https://www.tuwien.at/mg/asc/praetorius/software/mooafem.

FIGURE 3. *Contraction of the algebraic solver.* History plot of the contraction factors q_{ctr} from (13) for various polynomial degrees p with parameter $\mu = 10^{-5}$ for the presented polynomial hierarchy from (7) in the adaptive algorithm from Algorithm 3.1 stopping once the final mesh consists of 10^6 degrees of freedom (*left*) and the comparison with polynomial hierarchy motivated by Miraçi *et al.* [25] with localized smoothing for a fixed number of levels $L = 10$ (*right*).
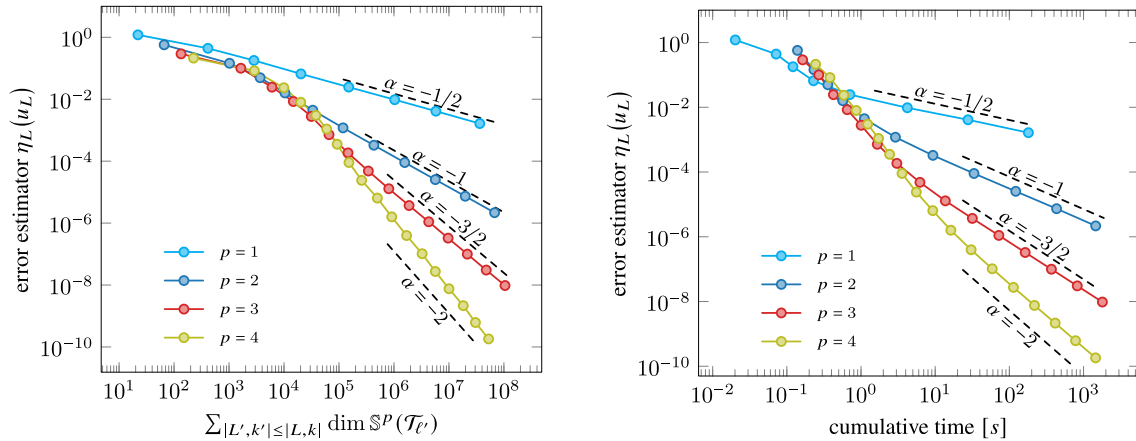

FIGURE 4. *Optimality of AFEM on L-shape.* The convergence history plot of the discretization error estimator $\eta_L(u_L)$ with respect to the total computational cost (*left*) and the cumulative computational time (*right*).

In Table 1 and Figure 7, we see the optimal convergence of the discretization estimator with the optimal rate $-1/2$ for $p = 1$ as well as -1 for $p = 2$ for both diffusion coefficients regardless of the jump size. We stress that the discontinuity in the diffusion coefficient does not affect the optimality of the proposed adaptive algorithm and the iteration numbers remain uniformly bounded as displayed in Table 2.

Both test cases exhibit singularities due to jumps in the diffusion coefficient; however, the jump can be much higher for two neighboring elements in the checkerboard case. In this case, near the cross point $(1/2, 1/2)$, the jump is of order 10^k from one element to the next, which coincides with the jump from the highest to the lowest value of $\boldsymbol{K}$ on the *whole* domain. For the striped test case, the jump between two neighboring elements
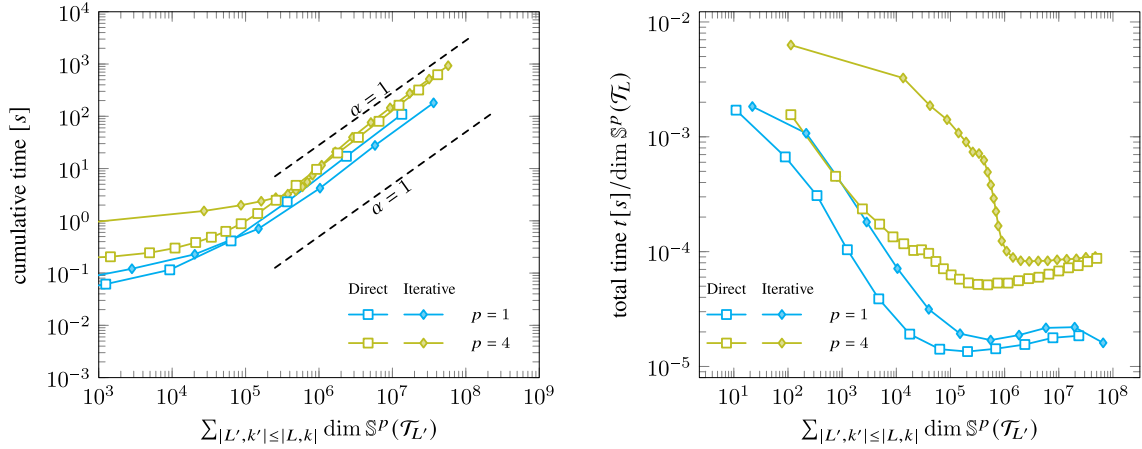
FIGURE 5. *Optimality of the local multigrid solver.* History plot of the cumulative computational time and the relative computational time per degree of freedom for the polynomial degrees $p = 1$ and $p = 4$. We compare the overall time with the direct solve (square) to the overall time of the AFEM algorithm with the multigrid solver (diamond). In particular, the displayed times include setup, marking, and mesh refinement.
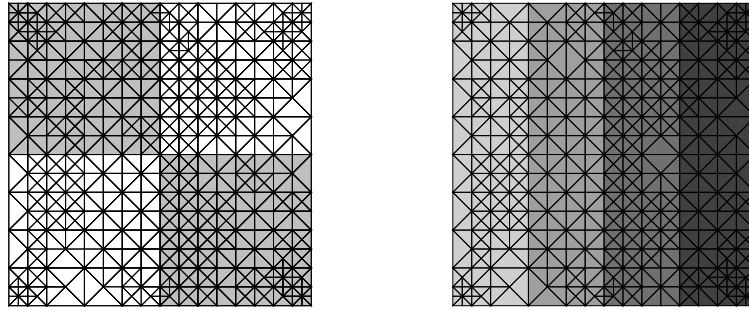

FIGURE 6. Adaptively refined meshes. *Left*: checkerboard diffusion with $k = 1$, polynomial degree $p = 1$ and $\#\mathcal{T}_8 = 603$. *Right*: stripe diffusion with $k = 2$, $p = 1$ and $\#\mathcal{T}_8 = 753$ (right).

belonging to different "stripes" is of order 10, even if the *global* jump in the diffusion (for non-neighboring elements) is of order 10^{2^k-1}.

This gives us the tools to observe numerically if the performance of our method only depends on *local* jumps in the diffusion coefficient.

5. PROOFS

Below we present proofs of intermediate results leading to our main Theorem 2.5 of L- and p-robust contraction of the multigrid solver and the L- and p-robust two-sided bound of the algebraic error by the built-in *a posteriori* estimator. We emphasize that this result improves the recent work [25] by removing the L-dependence. From an *algorithmic* point of view, this is done by applying *local* smoothing only on patches which change in the refinement step on lowest-order levels instead of on *every* patch as was the case in [25]. From an *analysis* point of view, L-robustness is achieved thanks to the strengthened Cauchy–Schwarz inequality on bisection-generated meshes (Prop. 5.7) building on the property that the levelwise overlap of the smoothed patches stays

TABLE 1. Mean value of experimental convergence rates of the discretization error estimator $\eta_L(u_L)$ over the cumulative cost in a log–log-plot for polynomial degrees $p = 1, 2$ and diffusion coefficient numbers $k = 1, 2, 3$.

	Checkerboard		Stripe	
	$p = 1$	$p = 2$	$p = 1$	$p = 2$
$k = 1$	-0.4961	-0.9877	-0.4956	-1.0116
$k = 2$	-0.4960	-0.9946	-0.4969	-0.9670
$k = 3$	-0.4960	-0.9826	-0.5095	-0.9766

TABLE 2. Mean and maximal iteration numbers for polynomial degrees $p = 1, 2$ and diffusion coefficient numbers $k = 1, 2, 3$.

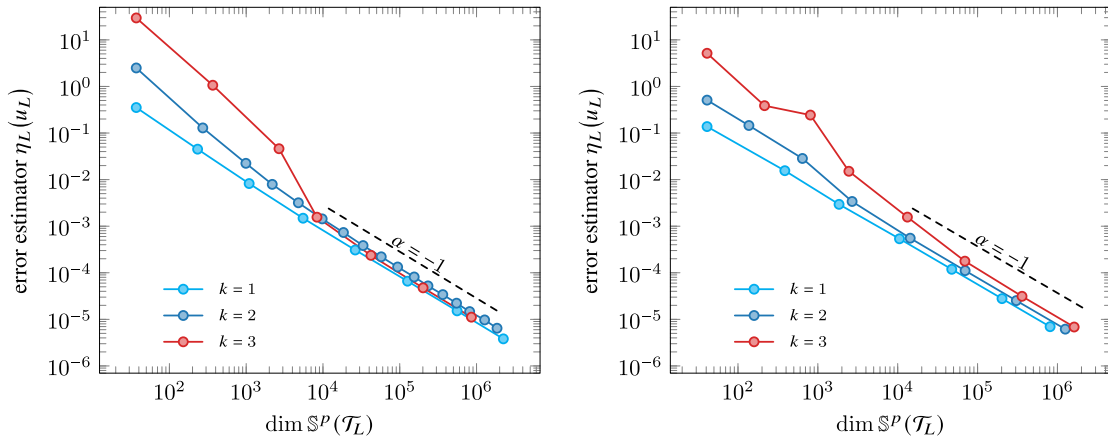|	Checkerboard		Stripe	
	$p = 1$	$p = 2$	$p = 1$	$p = 2$
$k = 1$	1	1.0455 (mean), 2 (max)	1	1.0455 (mean), 2 (max)
$k = 2$	1	2.3261 (mean), 5 (max)	1	1.0417 (mean), 2 (max)
$k = 3$	1	1.1818 (mean), 3 (max)	1	1.0833 (mean), 2 (max)

FIGURE 7. *Optimality of AFEM for jumping diffusion.* The convergence history plot of the discretization error estimator $\eta_L(u_L)$ for polynomial degree $p = 2$ with respect to the total computational cost for the checkerboard diffusion (*left*) and the stripe diffusion (*right*).

uniformly bounded. The next essential ingredient to prove the main result is an hp-stable decomposition on bisection generated meshes (Prop. 5.5), then one combines the results carefully together with the simple but crucial observation of uniform boundedness in the number of overlapping patches for a fixed level (Lem. 5.1) and bounds on the step-sizes and the levelwise solver update (Lem. 5.2).

5.1. Auxiliary results

We start with the simple observation that the number of overlapping patches is uniformly bounded.

Lemma 5.1 (Finite patch overlap)**.** *For all $T \in \mathcal{T}_\ell$, there holds*

$$\#(\mathcal{V}_\ell \cap T) = d + 1. \tag{23}$$

Therefore, for all $q \in \mathbb{N}$, it holds that

$$\left\|\!\left\|\sum_{z \in \mathcal{V}_\ell} v_{\ell,z}\right\|\!\right\|^2 \leq (d+1) \sum_{z \in \mathcal{V}_\ell} \|\!|v_{\ell,z}|\!\|^2 \qquad \text{for all } v_{\ell,z} \in \mathbb{V}_{\ell,z}^q. \tag{24}$$

Similar arguments show that

$$\left\|\!\left\|\nabla \sum_{z \in \mathcal{V}_\ell} v_{\ell,z}\right\|\!\right\|^2 \leq (d+1) \sum_{z \in \mathcal{V}_\ell} \|\nabla v_{\ell,z}\|^2 \qquad \text{for all } v_{\ell,z} \in \mathbb{V}_{\ell,z}^q. \tag{25}$$

Proof. The overlap (23) is clear from the geometry of the elements in the mesh. For all $\ell = 0, \ldots, L$, the discrete Cauchy–Schwarz inequality and (23) lead to

$$\left\|\!\left\|\sum_{z \in \mathcal{V}_\ell} v_{\ell,z}\right\|\!\right\|^2 = \sum_{T \in \mathcal{T}_\ell} \left\|\!\left\|\sum_{z \in \mathcal{V}_\ell \cap T} v_{\ell,z}\right\|\!\right\|_T^2 \leq (d+1) \sum_{z \in \mathcal{V}_\ell} \|\!|v_{\ell,z}|\!\|^2.$$

This concludes the proof. □

Next, we present bounds on the step-size and the levelwise solver update.

Lemma 5.2. *For all $\ell \in \{1, \ldots, L\}$, we have*

$$\|\!|\lambda_\ell \rho_\ell|\!\|^2 \leq \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\!|\rho_{\ell,z}|\!\|^2. \tag{26}$$

Moreover, we have upper and lower bounds for the step-sizes,

$$\frac{1}{d+1} \leq \lambda_\ell \leq d+1 \qquad \text{for all } \ell = 1, \ldots, L-1 \quad \text{and} \quad \frac{1}{d+1} \leq \lambda_L. \tag{27}$$

Proof. **Step 1.** Proof of (26) if $\ell = L$ or $(R_L(\rho_\ell) - \langle\!\langle \sigma_{\ell-1}, \rho_\ell \rangle\!\rangle)/\|\!|\rho_\ell|\!\|^2 \leq d+1$ for $\ell \in \{1, \ldots, L-1\}$. From Step (ii) of Algorithm 2.1, we have that $\lambda_\ell = (R_L(\rho_\ell) - \langle\!\langle \sigma_{\ell-1}, \rho_\ell \rangle\!\rangle)/\|\!|\rho_\ell|\!\|^2$ and thus

$$\|\!|\lambda_\ell \rho_\ell|\!\|^2 = \lambda_\ell \frac{R_L(\rho_\ell) - \langle\!\langle \sigma_{\ell-1}, \rho_\ell \rangle\!\rangle}{\|\!|\rho_\ell|\!\|^2} \|\!|\rho_\ell|\!\|^2 = \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} (R_L(\rho_{\ell,z}) - \langle\!\langle \sigma_{\ell-1}, \rho_{\ell,z} \rangle\!\rangle) \overset{(11)}{=} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\!|\rho_{\ell,z}|\!\|^2.$$

Step 2. Proof of (26) in the remaining cases. We use the finite overlap of the patches in Lemma 5.1 to obtain

$$\|\!|\lambda_\ell \rho_\ell|\!\|^2 = \frac{\lambda_\ell}{d+1} \|\!|\rho_\ell|\!\|^2 \overset{(24)}{\leq} \frac{\lambda_\ell}{d+1}(d+1) \sum_{z \in \mathcal{V}_\ell^+} \|\!|\rho_{\ell,z}|\!\|^2 = \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\!|\rho_{\ell,z}|\!\|^2.$$

Step 3. Proof of (27). For $\ell \in \{1, \ldots, L-1\}$, the upper bound is guaranteed by definition of λ_ℓ. The lower bound for $\ell \in \{1, \ldots, L\}$ is trivial if $\lambda_\ell = 1/(d+1)$. Otherwise, it is a consequence of the finite patch overlap:

$$\lambda_\ell = \frac{R_L(\rho_\ell) - \langle\!\langle \sigma_{\ell-1}, \rho_\ell \rangle\!\rangle}{\|\!|\rho_\ell|\!\|^2} \overset{(11)}{=} \frac{\sum_{z \in \mathcal{V}_\ell^+} \|\!|\rho_{\ell,z}|\!\|^2}{\|\!|\rho_\ell|\!\|^2} \overset{(24)}{\geq} \frac{1}{d+1}.$$

This concludes the proof.

□

In the next two subsections, we combine existing results from the literature to obtain a multilevel *hp*-robust stable decomposition and a strengthened Cauchy–Schwarz inequality for our setting of bisection-generated meshes. These will be crucial for the proofs of Theorem 2.5 and Corollary 2.6 in Section 5.4 below.

5.2. Multilevel *hp*-robust stable decomposition on NVB-generated meshes

We start by recalling the one-level *p*-robust stable decomposition from Sections 3.4 and 4.3 in [32] for $d = 2$ and $d = 3$, respectively.

Lemma 5.3 (*p*-robust one level decomposition). *Let $v_L \in \mathbb{V}_L^p$. Then, there exists a decomposition*

$$v_L = v_L^1 + \sum_{z \in \mathcal{V}_L} v_{L,z}^p \quad \text{with } v_L^1 \in \mathbb{V}_L^1 \text{ and } v_{L,z}^p \in \mathbb{V}_{L,z}^p, \tag{28}$$

which is stable in the sense of

$$\left\|\nabla v_L^1\right\|^2 + \sum_{z \in \mathcal{V}_L} \left\|\nabla v_{L,z}^p\right\|^2 \leq C_{\mathrm{OL}}^2 \|\nabla v_L\|^2. \tag{29}$$

The constant C_{OL} depends only on the space dimension d, the γ-shape regularity (5), and the quasi-uniformity constant C_{qu} from (6).

Similarly, we recall the *local* multilevel decomposition for piecewise affine functions proven in Lemma 3.1 from [37]. In order to present this stable decomposition in a form that is more suitable for our forthcoming analysis, we add a short proof for completeness.

Lemma 5.4 (*h*-robust local multilevel decomposition for lowest-order functions). *Let $v_L^1 \in \mathbb{V}_L^1$. Then, there exists a decomposition*

$$v_L^1 = \sum_{\ell=0}^L \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z}^1 \quad \text{with } v_{\ell,z}^1 \in \mathbb{V}_{\ell,z}^1, \tag{30}$$

which is stable in the sense of

$$\sum_{\ell=0}^L \sum_{z \in \mathcal{V}_\ell^+} \left\|\nabla v_{\ell,z}^1\right\|^2 \leq C_{\mathrm{ML}}^2 \left\|\nabla v_L^1\right\|^2. \tag{31}$$

The constant C_{ML} depends only on the space dimension d, the γ-shape regularity (5), and the quasi-uniformity constant C_{qu} from (6).

Proof. Let $v_L^1 \in \mathbb{V}_L^1$. Define $w_\ell^1 := (\Pi_\ell - \Pi_{\ell-1})v_L^1$ for $\ell \in \{0, \ldots, L\}$, where $\Pi_{-1} := 0$ and Π_ℓ is the projection to $\mathbb{V}_\ell^1$ from Section 3 of [37]. From Lemma 3.1 of [37], it holds that $w_\ell^1 \in \mathrm{span}\{\varphi_{\ell,z} : z \in \mathcal{V}_\ell^+\}$ with $\varphi_{\ell,z}$ being the $\mathbb{S}^1(\mathcal{T}_\ell)$ hat-function at vertex $z \in \mathcal{V}_\ell$. We decompose $w_\ell^1 = \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z}^1$ with $v_{\ell,z}^1 := w_\ell^1(z)\varphi_{\ell,z} \in \mathbb{V}_{\ell,z}^1$ and thus obtain

$$v_L^1 = \sum_{\ell=0}^L (\Pi_\ell - \Pi_{\ell-1})v_L^1 = \sum_{\ell=0}^L w_\ell^1 = \sum_{\ell=0}^L \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z}^1. \tag{32}$$

For fixed ℓ and $z \in \mathcal{V}_\ell^+$, the equivalence of norms on finite-dimensional spaces proves

$$\left\|v_{\ell,z}^1\right\|_{\omega_{\ell,z}} \leq \sum_{T \in \mathcal{T}_{\ell,z}} \left\|w_\ell^1(z)\varphi_{\ell,z}\right\|_T \leq \sum_{T \in \mathcal{T}_{\ell,z}} \left\|w_\ell^1\right\|_{L^\infty(T)} |T|^{1/2} \lesssim \sum_{T \in \mathcal{T}_{\ell,z}} \left\|w_\ell^1\right\|_T \simeq \left\|w_\ell^1\right\|_{\omega_{\ell,z}}, \tag{33}$$

where the hidden constants depend only on γ-shape regularity (5). To obtain stability of the decomposition (32), we use an inverse inequality on the patches and the stability proved in Lemma 3.7 of [37]:

$$\sum_{\ell=0}^L \sum_{z \in \mathcal{V}_\ell^+} \left\|\nabla v_{\ell,z}^1\right\|^2 \lesssim \sum_{\ell=0}^L \sum_{z \in \mathcal{V}_\ell^+} h_{\ell,z}^{-2} \left\|v_{\ell,z}^1\right\|_{\omega_{\ell,z}}^2 \overset{(33)}{\lesssim} \sum_{\ell=0}^L \sum_{z \in \mathcal{V}_\ell^+} h_{\ell,z}^{-2} \left\|w_\ell^1\right\|_{\omega_{\ell,z}}^2$$

$$= \sum_{\ell=0}^{L} \sum_{z \in \mathcal{V}_\ell^+} h_{\ell,z}^{-2} \big\| (\Pi_\ell - \Pi_{\ell-1}) v_L^1 \big\|_{\omega_{\ell,z}}^2 \overset{[37]}{\lesssim} \big\| \nabla v_L^1 \big\|^2.$$

This concludes the proof.   □

The combination of the two previous lemmas, done similarly in Proposition 7.6 from [24] for a *non-local* and hence not *h*-robust solver, leads to the following *hp*-robust decomposition.

Proposition 5.5 (*hp*-robust local multilevel decomposition)**.** *Let* $v_L \in \mathbb{V}_L^p$. *Then, there exist* $v_0 \in \mathbb{V}_0^1$, $v_{\ell,z} \in \mathbb{V}_{\ell,z}^1$, *and* $v_{L,z} \in \mathbb{V}_{L,z}^p$ *such that*

$$v_L = v_0 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z} + \sum_{z \in \mathcal{V}_L} v_{L,z}. \tag{34}$$

and this decomposition is stable in the sense of

$$\|v_0\|^2 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \|v_{\ell,z}\|^2 + \sum_{z \in \mathcal{V}_L} \|v_{L,z}\|^2 \le C_{\mathrm{SD}}^2 \, \|v_L\|^2. \tag{35}$$

The constant $C_{\mathrm{SD}} \ge 1$ *depends only on the space dimension d, γ-shape regularity* (5)*, the quasi-uniformity constant C_{qu} from* (6)*, and the ratio of $\Lambda_{\max}$ and $\Lambda_{\min}$.*

Proof. Let $v_L \in \mathbb{V}_L^p$. We begin with the decomposition of v_L by (28), then continue with the further decomposition of the lowest-order contribution v_L^1 in a multilevel way (30):

$$v_L \overset{(28)}{=} v_L^1 + \sum_{z \in \mathcal{V}_L} v_{L,z}^p \overset{(30)}{=} \sum_{\ell=0}^{L} \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z}^1 + \sum_{z \in \mathcal{V}_L} v_{L,z}^p$$

$$= \sum_{z \in \mathcal{V}_0} v_{0,z}^1 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z}^1 + \sum_{z \in \mathcal{V}_L^+} v_{L,z}^1 + \sum_{z \in \mathcal{V}_L} v_{L,z}^p.$$

By defining $v_0 := \sum_{z \in \mathcal{V}_0} v_{0,z}^1 \in \mathbb{V}_0^1$, $v_{\ell,z} := v_{\ell,z}^1 \in \mathbb{V}_{\ell,z}^1$ for $z \in \mathcal{V}_\ell^+$ and $1 \le \ell \le L-1$, and $v_{L,z} := v_{L,z}^1 + v_{L,z}^p \in \mathbb{V}_{L,z}^p$ for $z \in \mathcal{V}_L^+$ and $v_{L,z} := v_{L,z}^p \in \mathbb{V}_{L,z}^p$ for $z \in \mathcal{V}_L \setminus \mathcal{V}_L^+$, we obtain the decomposition (34). It remains to show that this decomposition is stable (35). First, we have for the coarsest level that

$$\|\nabla v_0\|^2 \overset{(25)}{\le} (d+1) \sum_{z \in \mathcal{V}_0} \big\| \nabla v_{0,z}^1 \big\|^2.$$

For the finest level, it holds that

$$\sum_{z \in \mathcal{V}_L} \|\nabla v_{L,z}\|^2 \le \sum_{z \in \mathcal{V}_L \setminus \mathcal{V}_L^+} \big\| \nabla v_{L,z}^p \big\|^2 + 2 \sum_{z \in \mathcal{V}_L} \left(\big\| \nabla v_{L,z}^1 \big\|^2 + \big\| \nabla v_{L,z}^p \big\|^2 \right)$$

$$\le (d+1) \sum_{z \in \mathcal{V}_L^+} \big\| \nabla v_{L,z}^1 \big\|^2 + (d+1) \sum_{z \in \mathcal{V}_L} \big\| \nabla v_{L,z}^p \big\|^2.$$

A combination of the two estimates shows that

$$\|\nabla v_0\|^2 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \|\nabla v_{\ell,z}\|^2 + \sum_{z \in \mathcal{V}_L} \|\nabla v_{L,z}\|^2$$

$$\leq (d+1) \left(\sum_{z \in \mathcal{V}_0} \left\| \nabla v_{0,z}^1 \right\|^2 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \left\| \nabla v_{\ell,z}^1 \right\|^2 + \sum_{z \in \mathcal{V}_L^+} \left\| \nabla v_{L,z}^1 \right\|^2 + \sum_{z \in \mathcal{V}_L} \left\| \nabla v_{L,z}^p \right\|^2 \right)$$

$$\leq (d+1) \sum_{\ell=0}^{L} \sum_{z \in \mathcal{V}_\ell^+} \left\| \nabla v_{\ell,z}^1 \right\|^2 + (d+1) \sum_{z \in \mathcal{V}_L} \left\| \nabla v_{L,z}^p \right\|^2$$

$$\overset{(31)}{\leq} C_{\mathrm{ML}}^2 (d+1) \left\| \nabla v_L^1 \right\|^2 + (d+1) \sum_{z \in \mathcal{V}_L} \left\| \nabla v_{L,z}^p \right\|^2$$

$$\overset{(29)}{\leq} \max\{1, C_{\mathrm{ML}}^2\} C_{\mathrm{OL}}^2 (d+1) \| \nabla v_L \|^2.$$

Hence, the decomposition (34) is stable with $(C'_{\mathrm{SD}})^2 := \max\{1, C_{\mathrm{ML}}^2\} C_{\mathrm{OL}}^2 (d+1)$ with respect to the $H^1(\Omega)$-seminorm. Taking into account the variations of the diffusion coefficient $\boldsymbol{K}$, we obtain (35) with the stability constant $C_{\mathrm{SD}} := C'_{\mathrm{SD}} \Lambda_{\max} / \Lambda_{\min}$. $\qquad\square$

5.3. Strengthened Cauchy–Schwarz inequality on NVB-generated meshes

The following results are proved in the spirit of [14, 21]. Note that the setting of this work is similar to [21], and unlike [14], the underlying adaptive meshes of the space hierarchy are not restricted to one bisection per level.

For analysis purposes, we introduce a sequence of uniformly refined triangulations indicated by $\{\widehat{\mathcal{T}}_j\}_{j=0}^M$ such that $\widehat{\mathcal{T}}_{j+1} := \texttt{refine}(\widehat{\mathcal{T}}_j, \widehat{\mathcal{T}}_j)$ and $\widehat{\mathcal{T}}_0 = \mathcal{T}_0$, where $\texttt{refine}$ enforces one bisection per element. According to [34], admissibility of $\mathcal{T}_0$ ensures that indeed each element $T \in \widehat{\mathcal{T}}_j$ is bisected only once into two children $T', T'' \in \widehat{\mathcal{T}}_{j+1}$. In the following, we will indicate the equivalent notation to Section 2 on uniform triangulations $\widehat{\mathcal{T}}_j$ with a hat, e.g., $\widehat{\mathbb{V}}_j^1$ is the equivalent of $\mathbb{V}_\ell^1$ on the uniformly refined mesh $\widehat{\mathcal{T}}_j$. The connection of the uniformly refined meshes and their adaptively generated counterpart requires further notation. For a given level $0 \leq \ell \leq L$ and a given node $z \in \mathcal{V}_\ell$, we define the generation $g_{\ell,z}$ of the patch by the maximum number of times an element of the patch has been bisected

$$g_{\ell,z} := \max_{T \in \mathcal{T}_{\ell,z}} \log_2(|T_0|/|T|) \in \mathbb{N}_0, \tag{36}$$

where $T_0 \in \mathcal{T}_0$ denotes the unique ancestor element of $T \in \mathcal{T}_\ell$. Define the maximal generation $M = \max_{z \in \mathcal{V}_L} g_{L,z}$.

First, we present the following result for uniformly refined meshes and then exploit this for our setting of adaptively refined meshes.

Lemma 5.6 (Strengthened Cauchy–Schwarz on nested uniform meshes). *Let $0 \leq i \leq j \leq M$, and $\widehat{u}_i \in \widehat{\mathbb{V}}_i^1$ as well as $\widehat{v}_j \in \widehat{\mathbb{V}}_j^1$. Then, it holds that*

$$\langle\!\langle \widehat{u}_i, \widehat{v}_j \rangle\!\rangle \leq \widehat{C}_{\mathrm{SCS}} \, \delta^{j-i} \, \widehat{h}_j^{-1} \| \nabla \widehat{u}_i \| \| \widehat{v}_j \|, \tag{37}$$

where $\delta = 2^{-1/2}$ and $\widehat{C}_{\mathrm{SCS}} > 0$ depends only on the domain Ω, the initial triangulation $\mathcal{T}_0$, $\Lambda_{\max}$, $\max_{T \in \widehat{\mathcal{T}}_M} \| \operatorname{div}(\boldsymbol{K}) \|_{L^\infty(T)}$, and γ-shape regularity from (5).

Proof. We begin by splitting the domain Ω into elementwise components, applying integration by parts, and using the Cauchy–Schwarz inequality. Note that the restriction of $\widehat{u}_i$ to any element $T \in \widehat{\mathcal{T}}_i$ is an affine function, and hence the second derivatives vanish. Thus, it holds with the outer normal $\boldsymbol{n}$ to ∂T that

$$\langle\!\langle \widehat{u}_i, \widehat{v}_j \rangle\!\rangle = \sum_{T \in \widehat{\mathcal{T}}_i} \int_T \boldsymbol{K} \nabla \widehat{u}_i \cdot \nabla \widehat{v}_j \, \mathrm{d}x$$

$$= \sum_{T \in \widehat{\mathcal{T}}_i} \left(- \int_T \mathrm{div}(\boldsymbol{K} \nabla \widehat{u}_i) \widehat{v}_j \, \mathrm{d}x + \int_{\partial T} \boldsymbol{K} \nabla \widehat{u}_i \cdot \boldsymbol{n} \, \widehat{v}_j \, \mathrm{d}x \right)$$

$$\leq \sum_{T \in \widehat{\mathcal{T}}_i} \left(\|(\mathrm{div}\boldsymbol{K}) \cdot \nabla \widehat{u}_i\|_{L^2(T)} \|\widehat{v}_j\|_{L^2(T)} + \|\boldsymbol{K} \nabla \widehat{u}_i\|_{L^2(\partial T)} \|\widehat{v}_j\|_{L^2(\partial T)} \right).$$

Due to $\boldsymbol{K} \in W^{1,\infty}(T)$, the fact that $\widehat{u}_i, \widehat{v}_j$ are piecewise affine, a discrete trace inequality, and $\widehat{h}_i^{-1} \gtrsim 1$, we get

$$\langle\!\langle \widehat{u}_i, \widehat{v}_j \rangle\!\rangle \lesssim \sum_{T \in \widehat{\mathcal{T}}_i} \left(\|\nabla \widehat{u}_i\|_{L^2(T)} \|\widehat{v}_j\|_{L^2(T)} + \|\nabla \widehat{u}_i\|_{L^2(\partial T)} \|\widehat{v}_j\|_{L^2(\partial T)} \right)$$

$$\lesssim \sum_{T \in \widehat{\mathcal{T}}_i} \left(\|\nabla \widehat{u}_i\|_{L^2(T)} \|\widehat{v}_j\|_{L^2(T)} + \left(\widehat{h}_i^{-1/2} \|\nabla \widehat{u}_i\|_{L^2(T)} \right) \left(\widehat{h}_i^{-1/2} \|\widehat{v}_j\|_{L^2(T)} \right) \right)$$

$$= \sum_{T \in \widehat{\mathcal{T}}_i} \left(1 + h_i^{-1} \right) \|\nabla \widehat{u}_i\|_{L^2(T)} \|\widehat{v}_j\|_{L^2(T)}$$

$$\lesssim \sum_{T \in \widehat{\mathcal{T}}_i} \widehat{h}_i^{-1} \|\nabla \widehat{u}_i\|_{L^2(T)} \|\widehat{v}_j\|_{L^2(T)}.$$

Moreover, note that due to uniform refinement, we have the equivalence $\delta^{j-i} = (2^{-1/2})^{j-i} \simeq \left(\widehat{h}_j / \widehat{h}_i \right)^{1/2}$ and $\widehat{h}_j \leq \widehat{h}_i$. Using the last equation multiplied by $1 = \widehat{h}_j^{1/2} \widehat{h}_j^{-1/2}$, we derive that

$$\langle\!\langle \widehat{u}_i, \widehat{v}_j \rangle\!\rangle \lesssim \sum_{T \in \widehat{\mathcal{T}}_i} \left(\frac{\widehat{h}_j}{\widehat{h}_i} \right)^{1/2} \widehat{h}_i^{-1/2} \widehat{h}_j^{-1/2} \|\nabla \widehat{u}_i\|_{L^2(T)} \|\widehat{v}_j\|_{L^2(T)}$$

$$\lesssim \sum_{T \in \widehat{\mathcal{T}}_i} \delta^{j-i} \widehat{h}_j^{-1} \|\nabla \widehat{u}_i\|_{L^2(T)} \|\widehat{v}_j\|_{L^2(T)} \leq \widehat{h}_j^{-1} \delta^{j-i} \|\nabla \widehat{u}_i\|_{L^2(\Omega)} \|\widehat{v}_j\|_{L^2(\Omega)}.$$

This concludes the proof. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\quad \square$

The last result enables us to tackle the setting of adaptively refined meshes.

Proposition 5.7 (Strengthened Cauchy–Schwarz on nested adaptive meshes)**.** *Consider levelwise functions* $v_\ell = \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z}^1 \in \mathbb{V}_\ell^1$ *with* $v_{\ell,z}^1 \in \mathbb{V}_{\ell,z}^1$ *for all* $1 \leq \ell \leq L - 1$. *Then, it holds that*

$$\sum_{\ell=1}^{L-1} \sum_{k=1}^{\ell-1} \langle\!\langle v_k, v_\ell \rangle\!\rangle \leq C_{\mathrm{SCS}} \left(\sum_{k=1}^{L-2} \sum_{w \in \mathcal{V}_k^+} \|\!|v_{k,w}^1|\!\|^2 \right)^{1/2} \left(\sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \|\!|v_{\ell,z}^1|\!\|^2 \right)^{1/2}, \tag{38}$$

where $C_{\mathrm{SCS}} > 0$ *depends only on* Ω, *the initial triangulation* $\mathcal{T}_0$, $\Lambda_{\max}/\Lambda_{\min}$, $\max_{T \in \mathcal{T}_L} \|\mathrm{div}(\boldsymbol{K})\|_{L^\infty(T)}/\Lambda_{\min}$, *and* γ-*shape regularity* (5).

Proof. Let $M \in \mathbb{N}$. The proof consists of five steps.

Step 1. First note that, for any $0 < \delta < 1$ and $x_i, y_i > 0$ with $0 \leq i \leq M$, there holds

$$\sum_{i=0}^{M} \sum_{j=i}^{M} \delta^{j-i} x_i y_j \leq \frac{1}{1-\delta} \left(\sum_{i=0}^{M} x_i^2 \right)^{1/2} \left(\sum_{j=0}^{M} y_j^2 \right)^{1/2}. \tag{39}$$

To see this, we change the summation order accordingly and use the Cauchy–Schwarz inequality to obtain

$$\sum_{i=0}^{M}\sum_{j=i}^{M}\delta^{j-i}x_iy_j = \sum_{i=0}^{M}\sum_{m=0}^{M-i}\delta^m x_i y_{m+i} = \sum_{m=0}^{M}\sum_{i=0}^{M-m}\delta^m x_i y_{m+i}$$

$$\leq \sum_{m=0}^{M}\delta^m\left[\left(\sum_{i=0}^{M-m}x_i^2\right)^{1/2}\left(\sum_{i=0}^{M-m}y_{m+i}^2\right)^{1/2}\right] \leq \left(\sum_{m=0}^{M}\delta^m\right)\left(\sum_{i=0}^{M}x_i^2\right)^{1/2}\left(\sum_{j=0}^{M}y_j^2\right)^{1/2}.$$

The geometric series then proves the claim (39).

Step 2. Let $z \in \mathcal{V}_L$ and $0 \leq j \leq M$ and recall the patch generation $g_{\ell,z}$ from (36). We introduce the set

$$\mathscr{L}_{\underline{\ell},\overline{\ell}}(z,j) := \left\{\ell \in \{\underline{\ell},\dots,\overline{\ell}\} : z \in \mathcal{V}_\ell^+ \text{ and } g_{\ell,z} = j\right\} \quad \text{with } 0 \leq \underline{\ell} \leq \overline{\ell} \leq L. \tag{40}$$

This set allows to track how large the levelwise overlap of patches with the same generation is. Crucially, the cardinality of these sets is uniformly bounded by

$$\max_{\substack{z\in\mathcal{V}_L \\ 0\leq j\leq M}} \#(\mathscr{L}_{0,L}(z,j)) \leq C_{\text{lev}} < \infty; \tag{41}$$

see, *e.g.*, Lemma 3.1 of [36] in the two-dimensional setting with arguments that transfer to three dimensions. The constant C_{lev} solely depends on γ-shape regularity (5).

Step 3. We introduce a way to reorder the patch contributions by generations (36). Note that, for any $0 \leq j \leq M$, $1 \leq \ell \leq L-1$, and $z \in \mathcal{V}_\ell^+$ such that $g_{\ell,z} = j$, the patch contribution $v_{\ell,z}^1 \in \mathbb{V}_{\ell,z}^1$ also belongs to $\widehat{\mathbb{V}}_j^1$. Once the generation constraint is introduced, one can shift the perspective from summing over "adaptive" levels and associated vertices to summing over "uniform" vertices and *only* the (finitely many, cf. (41)) levels where each vertex satisfies the generation constraint, *i.e.*, for $0 \leq \underline{\ell} \leq \overline{\ell} \leq L$ and $0 \leq j \leq M$, the two following sets coincide

$$\left\{(\ell,z)\in\mathbb{N}_0\times\mathcal{V}_L : \ell\in\{\underline{\ell},\dots,\overline{\ell}\}, z\in\mathcal{V}_\ell^+ \text{ with } g_{\ell,z}=j\right\} = \left\{(\ell,z)\in\mathbb{N}_0\times\mathcal{V}_L : z\in\widehat{\mathcal{V}}_j, \ell\in\mathscr{L}_{\underline{\ell},\overline{\ell}}(z,j)\right\}. \tag{42}$$

Step 4. According to γ-shape regularity (5), all elements in the patch have comparable size depending on C_{qu} from (6). If $g_{\ell,z} = j$, (at least) one element $T^\star \in \mathcal{T}_{\ell,z}$ satisfies $T^\star \in \widehat{\mathcal{T}}_j$ and it follows that $\widehat{h}_j \simeq |T^\star|^{1/d} \simeq |\omega_{\ell,z}|^{1/d} \simeq h_{\ell,z}$. In particular, there exists $C_{\text{eq}} > 0$ such that

$$\widehat{h}_j^{-1} \leq C_{\text{eq}} h_{\ell,z}^{-1}. \tag{43}$$

Step 5. We proceed to prove the main estimate (38). The central feature of the following approach is to introduce *additional* sums over the generations with generation constraints, *i.e.*, there holds for every admissible ℓ, k, that

$$\langle\!\langle v_k,\, v_\ell\rangle\!\rangle = \sum_{z\in\mathcal{V}_\ell^+}\sum_{w\in\mathcal{V}_k^+}\langle\!\langle v_{k,w}^1,\, v_{\ell,z}^1\rangle\!\rangle = \sum_{j=0}^{M}\sum_{i=0}^{M}\sum_{\substack{z\in\mathcal{V}_\ell^+ \\ g_{\ell,z}=j}}\sum_{\substack{w\in\mathcal{V}_k^+ \\ g_{k,w}=i}}\langle\!\langle v_{k,w}^1,\, v_{\ell,z}^1\rangle\!\rangle$$

$$= \sum_{j=0}^{M}\sum_{i=0}^{j}\sum_{\substack{z\in\mathcal{V}_\ell^+ \\ g_{\ell,z}=j}}\sum_{\substack{w\in\mathcal{V}_k^+ \\ g_{k,w}=i}}\langle\!\langle v_{k,w}^1,\, v_{\ell,z}^1\rangle\!\rangle + \sum_{j=0}^{M}\sum_{i=j+1}^{M}\sum_{\substack{z\in\mathcal{V}_\ell^+ \\ g_{\ell,z}=j}}\sum_{\substack{w\in\mathcal{V}_k^+ \\ g_{k,w}=i}}\langle\!\langle v_{k,w}^1,\, v_{\ell,z}^1\rangle\!\rangle.$$

We abbreviate the terms as $S_1(\ell,k)$ and $S_2(\ell,k)$, respectively. A change of the summation of order i and j yields for $S_1(\ell,k)$ that

$$S_1(\ell,k) = \sum_{i=0}^{M}\sum_{j=i}^{M}\sum_{\substack{z\in\mathcal{V}_\ell^+ \\ g_{\ell,z}=j}}\sum_{\substack{w\in\mathcal{V}_k^+ \\ g_{k,w}=i}}\langle\!\langle v_{k,w}^1,\, v_{\ell,z}^1\rangle\!\rangle.$$

Summing $S_2(\ell, k)$ over all ℓ and k and changing the order of summation, we obtain

$$\sum_{\ell=1}^{L-1}\sum_{k=1}^{\ell-1} S_2(\ell,k) = \sum_{j=0}^{M}\sum_{i=j+1}^{M}\sum_{k=1}^{L-2}\sum_{\ell=k+1}^{L-1}\sum_{\substack{z\in\mathcal{V}_\ell^+\\ g_{\ell,z}=j}}\sum_{\substack{w\in\mathcal{V}_k^+\\ g_{k,w}=i}}\langle\!\langle v_{k,w}^1,\, v_{\ell,z}^1\rangle\!\rangle.$$

Combining these two identities with (42), we see that

$$\sum_{\ell=1}^{L-1}\sum_{k=1}^{\ell-1}(S_1(\ell,k)+S_2(\ell,k)) = \sum_{i=0}^{M}\sum_{j=i}^{M}\sum_{\ell=1}^{L-1}\left\langle\!\!\left\langle \sum_{w\in\widehat{\mathcal{V}}_i}\sum_{k\in\mathscr{L}_{1,\ell-1}(w,i)} v_{k,w}^1,\, \sum_{\substack{z\in\mathcal{V}_\ell^+\\ g_{\ell,z}=j}} v_{\ell,z}^1 \right\rangle\!\!\right\rangle$$

$$+ \sum_{j=0}^{M}\sum_{i=j+1}^{M}\sum_{k=1}^{L-2}\left\langle\!\!\left\langle \sum_{\substack{w\in\mathcal{V}_k^+\\ g_{k,w}=i}} v_{k,w}^1,\, \sum_{z\in\widehat{\mathcal{V}}_j}\sum_{\ell\in\mathscr{L}_{k+1,L-1}(z,j)} v_{\ell,z}^1 \right\rangle\!\!\right\rangle.$$

We define the last two terms as S_1 and S_2, respectively. Since the second term S_2 is treated in the same way, we only present detailed estimations of the first term S_1. The strengthened Cauchy–Schwarz inequality (37) for functions defined on uniform meshes followed by the patch overlap (24) leads to

$$S_1 \le \widehat{C}_{\mathrm{SCS}}\sum_{i=0}^{M}\sum_{j=i}^{M}\delta^{j-i}\sum_{\ell=1}^{L-1}\left((d+1)\sum_{w\in\widehat{\mathcal{V}}_i}\left\|\sum_{k\in\mathscr{L}_{1,\ell-1}(w,i)}\nabla v_{k,w}^1\right\|^2\right)^{1/2}\sum_{\substack{z\in\mathcal{V}_\ell^+\\ g_{\ell,z}=j}}\widehat{h}_j^{-1}\left\|v_{\ell,z}^1\right\|.$$

The identity (42) and the finite levelwise overlap (41) show

$$\sum_{w\in\widehat{\mathcal{V}}_i}\left\|\sum_{k\in\mathscr{L}_{1,\ell-1}(w,i)}\nabla v_{k,w}^1\right\|^2 \le \sum_{k=1}^{\ell-1}\sum_{\substack{w\in\mathcal{V}_k^+\\ g_{k,w}=i}}\#(\mathscr{L}_{1,\ell-1}(w,i))\left\|\nabla v_{k,w}^1\right\|^2 \le C_{\mathrm{lev}}\sum_{k=1}^{L-2}\sum_{\substack{w\in\mathcal{V}_k^+\\ g_{k,w}=i}}\left\|\nabla v_{k,w}^1\right\|^2.$$

The equivalence of mesh sizes from (43) and a Poincaré-inequality prove

$$\sum_{\ell=1}^{L-1}\sum_{\substack{z\in\mathcal{V}_\ell^+\\ g_{\ell,z}=j}}\widehat{h}_j^{-1}\left\|v_{\ell,z}^1\right\| \le C_{\mathrm{eq}}C_{\mathrm{P}}\sum_{\ell=1}^{L-1}\sum_{\substack{z\in\mathcal{V}_\ell^+\\ g_{\ell,z}=j}}\left\|\nabla v_{\ell,z}^1\right\|.$$

A combination of (42) with (25) and (41), followed again by (42), yields

$$\left(\sum_{\ell=1}^{L-1}\sum_{\substack{z\in\mathcal{V}_\ell^+\\ g_{\ell,z}=j}}\left\|\nabla v_{\ell,z}^1\right\|\right)^2 = \left(\sum_{z\in\widehat{\mathcal{V}}_j}\sum_{\ell\in\mathscr{L}_{1,L-1}(z,j)}\left\|\nabla v_{\ell,z}^1\right\|\right)^2 \le (d+1)\,C_{\mathrm{lev}}\sum_{\ell=1}^{L-1}\sum_{\substack{z\in\mathcal{V}_\ell^+\\ g_{\ell,z}=j}}\left\|\nabla v_{\ell,z}^1\right\|^2.$$

Thus, we obtain the bound

$$\sum_{\ell=1}^{L-1}\sum_{\substack{z\in\mathcal{V}_\ell^+\\ g_{\ell,z}=j}}\left\|\nabla v_{\ell,z}^1\right\| \le (d+1)^{1/2}C_{\mathrm{lev}}^{1/2}\left(\sum_{\ell=1}^{L-1}\sum_{\substack{z\in\mathcal{V}_\ell^+\\ g_{\ell,z}=j}}\left\|\nabla v_{\ell,z}^1\right\|^2\right)^{1/2}.$$

Combining all estimates, together with the geometric series bound (39), confirms

$$S_1 \le \widehat{C}_{\mathrm{SCS}} \, (d+1) \, C_{\mathrm{lev}} \, C_{\mathrm{eq}} \, C_{\mathrm{P}} \frac{1}{1-\delta} \left(\sum_{k=1}^{L-2} \sum_{w \in \mathcal{V}_k^+} \left\| \nabla v_{k,w}^1 \right\|^2 \right)^{1/2} \left(\sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \left\| \nabla v_{\ell,z}^1 \right\|^2 \right)^{1/2}.$$

Finally, the result (38) is obtained after summing together with the analogous estimations coming from the remaining term S_2 and taking into consideration the variations of the diffusion coefficient $\boldsymbol{K}$ so that the result holds for the energy norm. This concludes the proof.

$\square$

5.4. Proof of the main results

For the sake of a concise presentation, we only consider the case $p > 1$. The case $p = 1$ is already covered in the literature [14, 37] and follows from our proof with only minor modifications.

Proof of Theorem 2.5, connection of solver and estimator (12). The proof consists of two steps.

Step 1. We show that there holds the identity

$$\left\| \left\| \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right\| \right\|^2 - 2 \left\langle\!\!\left\langle u_L^\star - v_L, \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right\rangle\!\!\right\rangle = -\|\rho_0\|^2 + \sum_{\ell=1}^{L-1} \|\lambda_\ell \rho_\ell\|^2 - 2 \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2. \qquad (44)$$

Indeed, note that $\sigma_\ell = \sum_{k=0}^\ell \lambda_k \rho_k$. By definition of the local lowest-order problems in (10) and (11) as well as the definition of $\rho_\ell = \sum_{z \in \mathcal{V}_\ell^+} \rho_{\ell,z}$, we have

$$\left\langle\!\!\left\langle u_L^\star - v_L, \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right\rangle\!\!\right\rangle \overset{(9)}{=} R_L(\rho_0) + \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} R_L(\rho_{\ell,z}) \overset{(10)}{=} \|\rho_0\|^2 + \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} R_L(\rho_{\ell,z})$$

$$\overset{(11)}{=} \|\rho_0\|^2 + \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \left(\|\rho_{\ell,z}\|^2 + \langle\!\langle \sigma_{\ell-1}, \rho_{\ell,z} \rangle\!\rangle \right)$$

$$= \|\rho_0\|^2 + \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \left(\|\rho_{\ell,z}\|^2 + \sum_{k=0}^{\ell-1} \langle\!\langle \lambda_k \rho_k, \rho_{\ell,z} \rangle\!\rangle \right)$$

$$= \|\rho_0\|^2 + \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2 + \sum_{\ell=1}^{L-1} \sum_{k=0}^{\ell-1} \langle\!\langle \lambda_k \rho_k, \lambda_\ell \rho_\ell \rangle\!\rangle.$$

Thus, by expanding the square, we have

$$\left\| \left\| \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right\| \right\|^2 - 2 \left\langle\!\!\left\langle u_L^\star - v_L, \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right\rangle\!\!\right\rangle = \sum_{\ell=0}^{L-1} \|\lambda_\ell \rho_\ell\|^2 - 2\|\rho_0\|^2 - 2 \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2$$

$$= -\|\rho_0\|^2 + \sum_{\ell=1}^{L-1} \|\lambda_\ell \rho_\ell\|^2 - 2 \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2.$$

This proves the identity (44).

Step 2. Recall that $\Phi(v_L) = v_L + \sigma_L = v_L + \sigma_{L-1} + \lambda_L \rho_L$. By definition of R_L in (9) and the choice of λ_L in Algorithm 2.1, we have

$$\|u_L^\star - \Phi(v_L)\|^2 = \|u_L^\star - (v_L + \sigma_{L-1})\|^2 - 2\,\lambda_L \,\langle\!\langle u_L^\star - (v_L + \sigma_{L-1}),\, \rho_L \rangle\!\rangle + \|\lambda_L \rho_L\|^2$$

$$= \|u_L^\star - (v_L + \sigma_{L-1})\|^2 - 2\lambda_L (R_L(\rho_L) - \langle\!\langle \sigma_{L-1},\, \rho_L \rangle\!\rangle) + \lambda_L \sum_{z \in \mathcal{V}_L} \|\rho_{L,z}\|^2$$

$$\overset{(11)}{=} \left\|\!\left| u_L^\star - \left(v_L + \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right) \right\|\!\right|^2 - \lambda_L \sum_{z \in \mathcal{V}_L} \|\rho_{L,z}\|^2.$$

For the first term it holds that

$$\left\|\!\left| u_L^\star - \left(v_L + \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right) \right\|\!\right|^2 = \|u_L^\star - v_L\|^2 + \left\|\!\left| \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right\|\!\right|^2 - 2 \left\langle\!\left\langle u_L^\star - v_L,\, \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right\rangle\!\right\rangle$$

$$\overset{(44)}{=} \|u_L^\star - v_L\|^2 - \|\rho_0\|^2 + \sum_{\ell=1}^{L-1} \|\lambda_\ell \rho_\ell\|^2 - 2 \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2$$

$$\overset{(26)}{\leq} \|u_L^\star - v_L\|^2 - \|\rho_0\|^2 - \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2.$$

Combining the last two estimates with the definition of $\zeta_L(v_L)$ in Algorithm 2.1, we obtain

$$\|u_L^\star - \Phi(v_L)\|^2 \leq \|u_L^\star - v_L\|^2 - \|\rho_0\|^2 - \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2 - \lambda_L \sum_{z \in \mathcal{V}_L} \|\rho_{L,z}\|^2$$

$$= \|u_L^\star - v_L\|^2 - \zeta_L(v_L)^2.$$

This concludes the proof of (12).

$\square$

Proof of Theorem 2.5, lower bound in (14). The relation between the solver and the estimator given in (12) shows that $\zeta_L(v_L) \leq \|u_L^\star - v_L\|$. $\square$

Proof of Corollary 2.6, equivalence of (13) *and* (14). We prove that the solver contraction (13) is equivalent to the upper bound of (14).

First, suppose that (13) holds. Then, we proceed similarly as in the proof of (12) to obtain

$$\|u_L^\star - v_L\|^2 = \|u_L^\star - \Phi(v_L)\|^2 - \left\|\!\left| \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right\|\!\right|^2 + 2 \left\langle\!\left\langle u_L^\star - v_L,\, \sum_{\ell=0}^{L-1} \lambda_\ell \rho_\ell \right\rangle\!\right\rangle + \lambda_L \sum_{z \in \mathcal{V}_L} \|\rho_{L,z}\|^2$$

$$\overset{(44)}{=} \|u_L^\star - \Phi(v_L)\|^2 + \|\rho_0\|^2 - \sum_{\ell=1}^{L-1} \|\lambda_\ell \rho_\ell\|^2 + 2 \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2 + \lambda_L \sum_{z \in \mathcal{V}_L} \|\rho_{L,z}\|^2$$

$$\overset{(13)}{\leq} q_{\text{ctr}}^2 \|u_L^\star - v_L\|^2 + 2\,\zeta_L(v_L)^2.$$

Rearranging this estimate proves the upper bound in (14) with $C_{\text{rel}}^2 = 2/(1 - q_{\text{ctr}}^2) > 1$.

Second, suppose the upper bound in (14). Then, it follows that

$$\| u_L^\star - \Phi(v_L) \|^2 \overset{(12)}{\leq} \| u_L^\star - v_L \|^2 - \zeta_L(v_L)^2 \overset{(14)}{\leq} \| u_L^\star - v_L \|^2 - C_{\mathrm{rel}}^{-2} \| u_L^\star - v_L \|^2.$$

This verifies the solver contraction (13) for $q_{\mathrm{ctr}}^2 = 1 - C_{\mathrm{rel}}^{-2} \in (0,1)$ and concludes the equivalence proof. $\qquad\square$

Proof of Theorem 2.5, upper bound in (14). We use the stable decomposition of Proposition 5.5 on the algebraic error $u_L^\star - v_L \in \mathbb{V}_L^p$ to obtain $v_0 \in \mathbb{V}_0^1$, $v_{\ell,z} \in \mathbb{V}_{\ell,z}^1$ and $v_{L,z} \in \mathbb{V}_{L,z}^p$ such that

$$u_L^\star - v_L = v_0 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z} + \sum_{z \in \mathcal{V}_L} v_{L,z}$$

$$\text{and} \quad \| v_0 \|^2 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \| v_{\ell,z} \|^2 + \sum_{z \in \mathcal{V}_L} \| v_{L,z} \|^2 \leq C_{\mathrm{SD}}^2 \| u_L^\star - v_L \|^2. \tag{45}$$

Note that $\sigma_\ell = \sum_{k=0}^\ell \lambda_k \rho_k$ for all $\ell = 0, \ldots, L$; see Algorithm 2.1. We use (45) to develop

$$\| u_L^\star - v_L \|^2 = \left\langle\!\left\langle u_L^\star - v_L, v_0 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z} + \sum_{z \in \mathcal{V}_L} v_{L,z} \right\rangle\!\right\rangle$$

$$\overset{\substack{(9)\\(10)}}{=} \langle\!\langle \rho_0, v_0 \rangle\!\rangle + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} R_L(v_{\ell,z}) + \sum_{z \in \mathcal{V}_L} R_L(v_{L,z})$$

$$\overset{(11)}{=} \langle\!\langle \rho_0, v_0 \rangle\!\rangle + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} (\langle\!\langle \rho_{\ell,z}, v_{\ell,z} \rangle\!\rangle + \langle\!\langle \sigma_{\ell-1}, v_{\ell,z} \rangle\!\rangle) + \sum_{z \in \mathcal{V}_L} (\langle\!\langle \rho_{L,z}, v_{L,z} \rangle\!\rangle + \langle\!\langle \sigma_{L-1}, v_{L,z} \rangle\!\rangle).$$

Expanding $\sigma_\ell = \rho_0 + \sum_{k=1}^\ell \lambda_\ell \rho_\ell$ and rearranging the terms finally leads to

$$\| u_L^\star - v_L \|^2 = \left\langle\!\left\langle \rho_0, v_0 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z} + \sum_{z \in \mathcal{V}_L} v_{L,z} \right\rangle\!\right\rangle + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \langle\!\langle \rho_{\ell,z}, v_{\ell,z} \rangle\!\rangle$$

$$+ \sum_{z \in \mathcal{V}_L} \langle\!\langle \rho_{L,z}, v_{L,z} \rangle\!\rangle + \sum_{\ell=1}^{L-1} \sum_{k=1}^{\ell-1} \left\langle\!\left\langle \lambda_k \rho_k, \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z} \right\rangle\!\right\rangle + \sum_{k=1}^{L-1} \left\langle\!\left\langle \lambda_k \rho_k, \sum_{z \in \mathcal{V}_L} v_{L,z} \right\rangle\!\right\rangle.$$

Note that, until this point, only equalities are used. In the following, we will estimate each of the constituting terms of the algebraic error using Young's inequality in the form $ab \leq (\alpha/2) a^2 + (2\alpha)^{-1} b^2$ with $\alpha = 4C_{\mathrm{SD}}^2$, the strengthened Cauchy–Schwarz inequality, and patch overlap arguments as done in the proof of Lemma 5.1. Using the fact that $\lambda_0 = 1$ and the decomposition of the error $u_L^\star - v_L = v_0 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z} + \sum_{z \in \mathcal{V}_L} v_{L,z}$, we see that the first term yields

$$\left\langle\!\left\langle \rho_0, v_0 + \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z} + \sum_{z \in \mathcal{V}_L} v_{L,z} \right\rangle\!\right\rangle \overset{(45)}{=} \langle\!\langle \rho_0, u_L^\star - v_L \rangle\!\rangle \leq \frac{1}{2} \| \lambda_0 \rho_0 \|^2 + \frac{1}{2} \| u_L^\star - v_L \|^2.$$

For the second term, we obtain that

$$\sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \langle\!\langle \rho_{\ell,z}, v_{\ell,z} \rangle\!\rangle \leq 2C_{\mathrm{SD}}^2 \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \| \rho_{\ell,z} \|^2 + \frac{1}{8C_{\mathrm{SD}}^2} \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \| v_{\ell,z} \|^2$$

$$\overset{(27)}{\leq} 2C_{\mathrm{SD}}^2 (d+1) \sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2 + \frac{1}{8C_{\mathrm{SD}}^2} \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \|v_{\ell,z}\|^2,$$

and similarly for the third term

$$\sum_{z \in \mathcal{V}_L} \langle\!\langle \rho_{L,z}, v_{L,z} \rangle\!\rangle \overset{(27)}{\leq} 2\,C_{\mathrm{SD}}^2 (d+1)\, \lambda_L \sum_{z \in \mathcal{V}_L} \|\rho_{L,z}\|^2 + \frac{1}{8C_{\mathrm{SD}}^2} \sum_{z \in \mathcal{V}_L} \|v_{L,z}\|^2.$$

For the fourth term, we have

$$\sum_{\ell=1}^{L-1} \sum_{k=1}^{\ell-1} \left\langle\!\!\left\langle \lambda_k \rho_k, \sum_{z \in \mathcal{V}_\ell^+} v_{\ell,z} \right\rangle\!\!\right\rangle \overset{(38)}{\leq} C_{\mathrm{SCS}} \left(\sum_{k=1}^{L-2} \sum_{w \in \mathcal{V}_k^+} \|\lambda_k \rho_{k,w}\|^2 \right)^{1/2} \left(\sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \|v_{\ell,z}\|^2 \right)^{1/2}$$

$$\leq 2\,C_{\mathrm{SCS}}^2 C_{\mathrm{SD}}^2 \sum_{k=0}^{L-2} \sum_{w \in \mathcal{V}_k^+} \|\lambda_k \rho_{k,w}\|^2 + \frac{1}{8C_{\mathrm{SD}}^2} \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \|v_{\ell,z}\|^2$$

$$\overset{(27)}{\leq} 2C_{\mathrm{SCS}}^2 C_{\mathrm{SD}}^2 (d+1) \sum_{k=0}^{L-2} \lambda_k \sum_{w \in \mathcal{V}_k^+} \|\rho_{k,w}\|^2 + \frac{1}{8C_{\mathrm{SD}}^2} \sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \|v_{\ell,z}\|^2.$$

Finally, to treat the last term where higher-order terms appear together with a sum over levels, we proceed similarly as in Proof of Theorem 4.8 from [14] and obtain

$$\sum_{k=1}^{L-1} \left\langle\!\!\left\langle \lambda_k \rho_k, \sum_{z \in \mathcal{V}_L} v_{L,z} \right\rangle\!\!\right\rangle = \sum_{z \in \mathcal{V}_L} \left\langle\!\!\left\langle \sum_{k=1}^{L-1} \lambda_k \rho_k, v_{L,z} \right\rangle\!\!\right\rangle$$

$$\leq 2C_{\mathrm{SD}}^2 \sum_{z \in \mathcal{V}_L} \left\|\sum_{k=1}^{L-1} \lambda_k \rho_k \right\|_{\omega_{L,z}}^2 + \frac{1}{8C_{\mathrm{SD}}^2} \sum_{z \in \mathcal{V}_L} \|v_{L,z}\|^2.$$

For the first term of the last bound, we have that

$$\sum_{z \in \mathcal{V}_L} \left\|\sum_{k=1}^{L-1} \lambda_k \rho_k \right\|_{\omega_{L,z}}^2 \lesssim \left\|\sum_{k=1}^{L-1} \lambda_k \rho_k \right\|^2 = \sum_{k=1}^{L-1} \|\lambda_k \rho_k\|^2 + 2 \sum_{\ell=1}^{L-1} \sum_{k=1}^{\ell-1} \langle\!\langle \lambda_k \rho_k, \lambda_\ell \rho_\ell \rangle\!\rangle$$

$$\overset{(38)}{\leq} \sum_{k=1}^{L-1} \|\lambda_k \rho_k\|^2 + 2\,C_{\mathrm{SCS}} \left(\sum_{k=1}^{L-2} \sum_{w \in \mathcal{V}_k^+} \|\lambda_k \rho_{k,w}\|^2 \right)^{1/2} \left(\sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \|\lambda_\ell \rho_{\ell,z}\|^2 \right)^{1/2}$$

$$\overset{\substack{(27)\\(26)}}{\leq} (1 + 2\,C_{\mathrm{SCS}} (d+1)) \left(\sum_{\ell=1}^{L-1} \lambda_\ell + \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2 \right).$$

Summing all the estimates of the algebraic error components and defining the constant $C_{\mathrm{rel}}^2 := \max\{1/2, C_{\mathrm{SD}}^2 (d+1) (2 + C_{\mathrm{SCS}}^2 + 2\,C_{\mathrm{SCS}} (d+1)^{1/2})\}$, we see that

$$\|u_L^\star - v_L\|^2 \leq \frac{1}{2} \|\lambda_0 \rho_0\|^2 + \frac{1}{2} \|u_L^\star - v_L\|^2 + 4C_{\mathrm{rel}}^2 \left(\sum_{\ell=1}^{L-1} \lambda_\ell \sum_{z \in \mathcal{V}_\ell^+} \|\rho_{\ell,z}\|^2 + \lambda_L \sum_{z \in \mathcal{V}_L} \|\rho_{L,z}\|^2 \right)$$

$$+ \frac{1}{4C_{\mathrm{SD}}^2} \left(\sum_{\ell=1}^{L-1} \sum_{z \in \mathcal{V}_\ell^+} \|v_{\ell,z}\|^2 + \sum_{z \in \mathcal{V}_L} \|v_{L,z}\|^2 \right)$$

$$\overset{(45)}{\leq} 4C_{\mathrm{rel}}^2 \zeta_L(v_L)^2 + \frac{3}{4} \|u_L^\star - v_L\|^2.$$

After rearranging the terms, we finally obtain that

$$\|u_L^\star - v_L\|^2 \leq C_{\mathrm{rel}}^2 \zeta_L(v_L)^2. \tag{46}$$

This proves the upper bound of (14) and thus concludes the proof of Theorem 2.5. $\qquad \square$

REFERENCES

[1] P.F. Antonietti, L. Mascotto and M. Verani, A multigrid algorithm for the *p*-version of the virtual element method. *ESAIM: Math. Modell. Numer. Anal.* **52** (2018) 337–364.

[2] M. Aurada, M. Feischl, T. Führer, M. Karkulik and D. Praetorius, Efficiency and optimality of some weighted-residual error estimator for adaptive 2D boundary element methods. *Comput. Methods Appl. Math.* **13** (2013) 305–332.

[3] D. Bai and A. Brandt, Local mesh refinement multilevel techniques. *SIAM J. Sci. Stat. Comput.* **8** (1987) 109–134.

[4] R.E. Bank, T.F. Dupont and H. Yserentant, The hierarchical basis multigrid method. *Numer. Math.* **52** (1988) 427–458.

[5] P. Binev, W. Dahmen and R. DeVore, Adaptive finite element methods with convergence rates. *Numer. Math.* **97** (2004): 219–268.

[6] J.H. Bramble, J.E. Pasciak and A.H. Schatz, The construction of preconditioners for elliptic problems by substructuring. I. *Math. Comput.* **47** (1986) 103–134.

[7] J.H. Bramble, J.E. Pasciak and J. Xu, Parallel multilevel preconditioners, in Numerical Analysis 1989 (Dundee, 1989). Vol. 228 of *Pitman Res. Notes Math. Ser. Longman Sci. Tech.*. Harlow (1990) 23–39.

[8] A. Brandt, S. McCormick and J. Ruge, Algebraic multigrid (AMG) for sparse matrix equations, in Sparsity and its Applications. Cambridge Univ. Press, Cambridge (1985) 257–284.

[9] P.D. Brubeck and P.E. Farrell, A scalable and robust vertex-star relaxation for high-order FEM. *SIAM J. Sci. Comput.* **44** (2022) 2991.

[10] M. Brunner, M. Innerberger, A. Miraçi, D. Praetorius, J. Streitberger and P. Heid, Adaptive FEM with quasi-optimal overall cost for nonsymmetric linear elliptic PDEs. *IMA J. Numer. Anal.* (2023).

[11] C. Canuto, R.H. Nochetto, R. Stevenson and M. Verani, Convergence and optimality of *hp*-AFEM. *Numer. Math.* **135** (2017) 1073–1119.

[12] C. Carstensen, M. Feischl, M. Page and D. Praetorius, Axioms of adaptivity. *Comput. Math. Appl.* **67** (2014) 1195–1253.

[13] J.M. Cascón, C. Kreuzer, R.H. Nochetto and K.G. Siebert, Quasi-optimal convergence rate for an adaptive finite element method. *SIAM J. Numer. Anal.* **46** (2008) 2524–2550.

[14] L. Chen, R.H. Nochetto and J. Xu, Optimal multilevel methods for graded bisection grids. *Numer. Math.* **120** (2012) 1–34.

[15] D.A. Di Pietro, F. Hülsemann, P. Matalon, P. Mycek, U. Rüde and D. Ruiz, An *h*-multigrid method for hybrid high-order discretizations. *SIAM J. Sci. Comput.* **43** (2021) S839–S861.

[16] L. Diening, L. Gehring and J. Storn, Adaptive mesh refinement for arbitrary initial triangulations. Preprint arXiv:2306.02674 (2023).

[17] W. Dörfler, A convergent adaptive algorithm for Poisson's equation. *SIAM J. Numer. Anal.* **33** (1996) 1106–1124.

[18] G. Gantner, A. Haberl, D. Praetorius and S. Schimanko, Rate optimality of adaptive finite element methods with respect to overall computational costs. *Math. Comput.* **90** (2021) 2011–2040.

[19] W. Hackbusch, Multigrid Methods and Applications. Vol. 4 of *Springer Series in Computational Mathematics*. Springer-Verlag, Berlin (1985) xiv+377.

[20] W. Heinrichs, Line relaxation for spectral multigrid methods. *J. Comput. Phys.* **77** (1988) 166–182.

[21] 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** (2012) 297–332.

[22] M. Innerberger and D. Praetorius, MooAFEM: an object oriented matlab code for higher-order adaptive FEM for (nonlinear) elliptic PDEs. *Appl. Math. Comput.* **442** (2023) 127731.

[23] M. Karkulik, D. Pavlicek and D. Praetorius, On 2D newest vertex bisection: optimality of mesh-closure and H^1-stability of L_2-projection. *Constr. Approx.* **38** (2013) 213–234.

[24] 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** (2019) 2856–2884.

[25] 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** (2021) S117–S145.

[26] P. Morin, R.H. Nochetto and K.G. Siebert, Data oscillation and convergence of adaptive FEM. *SIAM J. Numer. Anal.* **38** (2000) 466–488.

[27] P. Oswald, Multilevel Finite Element Approximation: Theory and Applications. *Teubner Skripten zur Numerik* [*Teubner Scripts on Numerical Mathematics*]. B.G. Teubner Stuttgart (1994).

[28] L.F. Pavarino, Additive Schwarz methods for the p-version finite element method. *Numer. Math.* **66** (1994) 493–515.

[29] C.-M. Pfeiler and D. Praetorius, Dörfler marking with minimal cardinality is a linear complexity problem. *Math. Comput.* **89** (2020) 2735–2752.

[30] U. Rüde, Mathematical and Computational Techniques for Multilevel Adaptive Methods. Vol. 13 of *Frontiers in Applied Mathematics*. SIAM, Philadelphia, PA (1993) xii+140.

[31] U. Rüde, Fully adaptive multigrid methods. *SIAM J. Numer. Anal.* **30** (1993) 230–248.

[32] J. Schöberl, J.M. Melenk, C. Pechstein and S. Zaglmayr, Additive Schwarz preconditioning for p-version triangular and tetrahedral finite elements. *IMA J. Numer. Anal.* **28** (2008) 1–24.

[33] R. Stevenson, Optimality of a standard adaptive finite element method. *Found. Comput. Math.* **7** (2007) 245–269.

[34] R. Stevenson, The completion of locally refined simplicial partitions created by bisection. *Math. Comput.* **77** (2008) 227–241.

[35] C.T. Traxler, An algorithm for adaptive mesh refinement in n dimensions. *Computing* **59** (1997) 115–137.

[36] 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.* **49** (2006) 1405–1429.

[37] J. Wu and H. Zheng, Uniform convergence of multigrid methods for adaptive meshes. *Appl. Numer. Math.* **113** (2017) 109–123.

[38] X. Zhang, Multilevel Schwarz methods. *Numer. Math.* **63** (1992) 521–539.