SIAM J. SCI. COMPUT.
Vol. 43, No. 5, pp. S117–S145

A-POSTERIORI-STEERED p-ROBUST MULTIGRID WITH OPTIMAL STEP-SIZES AND ADAPTIVE NUMBER OF SMOOTHING STEPS*

ANI MIRAÇI†, JAN PAPEŽ‡, AND MARTIN VOHRALÍK†

Abstract. We develop a multigrid solver steered by an a posteriori estimator of the algebraic error. We adopt the context of a second-order elliptic diffusion problem discretized by conforming finite elements of arbitrary polynomial degree $p \geq 1$. Our solver employs zero pre- and one postsmoothing by the overlapping Schwarz (block-Jacobi) method and features an optimal choice of the step-sizes in the smoothing correction on each level by line search. This leads to a simple Pythagorean formula of the algebraic error in the next step in terms of the current error and levelwise and patchwise error reductions. We show the following two results and their equivalence: the solver contracts the algebraic error independently of the polynomial degree p; and the estimator represents a two-sided p-robust bound on the algebraic error. The p-robustness results are obtained by carefully applying the results of [J. Schöberl et al., *IMA J. Numer. Anal.*, 28 (2008), pp. 1–24] for one mesh, combined with a multilevel stable decomposition for piecewise affine polynomials of [J. Xu, L. Chen, and R. H. Nochetto, *Optimal multilevel methods for* $H(\mathrm{grad})$*,* $H(\mathrm{curl})$*, and* $H(\mathrm{div})$ *systems on graded and unstructured grids*, in Multiscale, Nonlinear and Adaptive Approximation, Springer, Berlin, 2009, pp. 599–659]. We consider quasi-uniform or graded bisection simplicial meshes and prove at most linear dependence on the number of mesh levels for minimal H^1-regularity and complete independence for H^2-regularity. We also present a simple and effective way for the solver to adaptively choose the number of postsmoothing steps necessary at each individual level, yielding a yet improved error reduction. Numerical tests confirm p-robustness and show the benefits of the adaptive number of smoothing steps.

Key words. multigrid method, a posteriori error estimate, stable decomposition, p-robustness, optimal step-sizes, error representation formula

AMS subject classifications. 65N55, 65N30, 65F10, 65N15, 65N22

DOI. 10.1137/20M1349503

1. Introduction. Multilevel (multigrid) methods have shown their versatility as solvers and/or preconditioners of large sparse algebraic linear systems arising from numerical discretizations of partial differential equations. We refer to pioneering works such as Brandt, McCormick, and Ruge [9], Bramble, Pasciak, and Schatz [6], Bank, Dupont, and Yserentant [4], Ruge and Stüben [27], or Oswald [22], as well as to survey works that thoroughly treat subspace correction methods in Xu [34], robust multigrid methods with respect to nonsmooth coefficients in Chan and Wan [11], multigrid solvers for high-order discretizations in Sundar, Stadler, and Biros [31], and the references therein.

In this work, we develop a *multilevel solver* for algebraic linear systems arising from the discretization using conforming finite elements of *arbitrary polynomial degree*

*Received by the editors July 1, 2020; accepted for publication (in revised form) December 21, 2020; published electronically April 7, 2021.
https://doi.org/10.1137/20M1349503
Funding: This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement 647134 GATIPOR). The work of the second author was also supported by the Grant Agency of the Czech Republic under grant 20-01074S.
†Inria, 2 rue Simone Iff, 75589 Paris, France, and CERMICS, Ecole des Ponts, 77455 Marne-la-Vallée, France (ani.miraci@inria.fr, martin.vohralik@inria.fr).
‡Institute of Mathematics, Czech Academy of Sciences, Žitná 25, 115 67 Prague, Czech Republic (jan@papez.org).

$p \geq 1$. One iteration of our solver can be seen as a V-cycle employing *zero pre-* and *one postsmoothing step*, where the levelwise smoother is *overlapping additive Schwarz (block-Jacobi)* associated to the patches of elements sharing a common vertex. A crucial difference to the classical V-cycle is that on each level we use an *optimal step-size* at the error correction stage, yielding minimal algebraic error in the subsequent level.

The idea of an optimal step-size in the error correction is not new; in fact, a weighting of multigrid error corrections concept appears as early as in Brandt [8]. Then, this approach is used, e.g., in Canuto and Quarteroni [10], though not in a multigrid setting. The interest of an optimally weighted error correction in the context of multigrid has been also pointed out in Heinrichs [15], where this choice resulted in a better numerical performance of the solver. Another version of multigrid solvers with a changing step-size error correction can be found in the form of a scaled residual in Rüde [26]. A crucial immediate consequence of our present optimal step-sizes choice is that the error contraction becomes explicitly known. This allows one to obtain the following Pythagorean formula representing the *error decrease* from step i to step $i+1$:

$$\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^{i+1})\right\|^2 = \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^{i})\right\|^2 - \sum_{j=0}^{J}\left(\lambda_j^i\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|\right)^2. \tag{1.1}$$

Here, $\boldsymbol{\mathcal{K}}$ is the diffusion tensor, $j \in \{0,\dots,J\}$ is the level counter, u_J is the (unknown) exact algebraic solution, u_J^i denotes the available iterate, u_J^{i+1} is the next iterate, ρ_j^i are the computed levelwise smoothing corrections, and λ_j^i are the levelwise optimal step-sizes.

A salient feature of formula (1.1) is that the computable levelwise terms $\big\{\sum_{j=0}^{J}\big(\lambda_j^i\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\|\big)^2\big\}^{\frac{1}{2}}$ form an a posteriori estimator η_{alg}^i, representing a guaranteed lower bound for the algebraic error $\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|$. Thus our solver is actually driven by the information provided by the estimator, making the solver an *a-posteriori-steered* multigrid.

Our main results can be summarized as follows. First, we prove that our multilevel solver *contracts* the error in each iteration. Second, we show that the associated a posteriori estimator η_{alg}^i is *efficient* in that it also represents an upper bound of the error (up to a constant). These two claims are actually equivalent. Third, there holds

$$\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\right\|^2 \approx \underbrace{\sum_{j=0}^{J}\left(\lambda_j^i\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|\right)^2}_{\left(\eta_{\mathrm{alg}}^i\right)^2} = \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_0^i\right\|^2 + \sum_{j=1}^{J}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2, \tag{1.2}$$

so that the developed a posteriori error estimator actually *localizes* the algebraic error with respect to mesh levels and also with respect to patches of elements on each level. These results hold for quasi-uniform meshes as well as possibly highly graded ones. Importantly, all the results hold *p-robustly*, i.e., are robust with respect to the polynomial degree p.

Notable previous works in treating p-robustness include Quarteroni and Sacchi Landriani [25] for a specific domain configuration and Pavarino [24] for quadrilateral/hexahedral meshes, where the author introduced a p-robust additive Schwarz method. Later, Janssen and Kanschat [16] and Lucero Lorca and Kanschat [18] used multilevel preconditioners for rectangular/hexahedral meshes, and Antonietti and Pennesi [2] considered more general meshes. Therein, however, more smoothing steps are generally necessary, whereas, we recall, we only rely on a single postsmoothing step. A p-robust stable decompostion on triangular/tetrahedral meshes was presented in Schöberl et al. [29]. It leads to a (one-mesh) p-robust preconditioner and plays an important part in the analysis of our work.

Compared to our previous work [20], we can mention the following improvements: (1) In the solver of [20], a global optimal step-size was used, whereas we use here levelwise step-sizes. (2) We obtain here the powerful error decrease formula (1.1). (3) The solver proposed in this work does not need any damping, where tuning of the parameters can be cumbersome. (4) The current analysis gives at most linear dependence on the number of mesh levels J under minimal H^1-regularity. (5) The current analysis gives complete independence of J in an H^2-regularity setting.

Formula (1.1) is also the foundation of a simple and efficient *adaptive strategy* for the choice of the number of postsmoothing steps per level. The essence and particularity of our strategy relies on *a-posteriori-steered decision-making* of the number of smoothing steps. Following (1.1), after one mandatory smoothing step at each level, if the given decrease $\lambda_j^i \|\mathcal{K}^{\frac{1}{2}} \nabla \rho_j^i\|$ is higher than a user-prescribed portion of the decrease made by the previous levels, we decide to do another smoothing step before going to the next level. The idea of employing a *variable* number of smoothing steps per level has also been explored, e.g., in Bramble and Pasciak [5], where a generalized V-cycle uses more smoothing steps on coarser grids and fewer on finer ones. This decision is, however, taken a priori. Closely related to the subject is also the work of Thekale et al. [32], who suggest a variable number of multigrid cycles per level which optimizes the costs of the full multigrid method by formulating a nonlinear integer programming problem of small enough size to be solved exactly.

This manuscript is organized as follows. In section 2, we present the multilevel setting and notation we will be working with, and section 3 develops the motivation leading us to consider our particular multilevel solver. The solver is then presented in section 4, and the a posteriori error estimator is introduced in section 5. In section 6, we collect the main results of the manuscript. In section 7, we present the solver with the adaptive choice of number of postsmoothing steps, section 8 presents a simplified cost analysis, and section 9 collects the results of numerical experiments, which additionally show numerical robustness of our solver with respect to the jumps of the diffusion tensor for uniform mesh refinements. The proof of our main result is given in section 10, and we present our concluding remarks in section 11.

2. Setting.

This section presents the model problem and the multilevel setting with which we will be working.

2.1. Model problem, finite element discretization, and algebraic system.

We consider a second-order elliptic diffusion problem defined over $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, an open bounded polytope with a Lipschitz-continuous boundary. Let $f \in L^2(\Omega)$ be a source term and $\mathcal{K} \in [L^\infty(\Omega)]^{d\times d}$ a symmetric positive definite diffusion coefficient. The weak solution $u \in H_0^1(\Omega)$ is given by

$$(\mathcal{K}\nabla u, \nabla v) = (f, v) \quad \forall v \in H_0^1(\Omega), \tag{2.1}$$

where $(\cdot,\cdot)$ is the $L^2(\Omega)$ or $[L^2(\Omega)]^d$ scalar product.

We discretize the continuous problem (2.1) by fixing $\mathcal{T}_J$, a matching simplicial mesh of Ω, and an integer $p \geq 1$, in order to introduce the finite element space of continuous piecewise p-degree polynomials

$$V_J^p := \mathbb{P}_p(\mathcal{T}_J) \cap H_0^1(\Omega), \tag{2.2}$$

where $\mathbb{P}_p(\mathcal{T}_J) := \{v_J \in L^2(\Omega), v_J|_K \in \mathbb{P}_p(K)\ \forall K \in \mathcal{T}_J\}$. The discrete problem now consists of finding $u_J \in V_J^p$ such that

$$(\boldsymbol{\mathcal{K}}\nabla u_J, \nabla v_J) = (f, v_J) \quad \forall v_J \in V_J^p. \tag{2.3}$$

If one introduces a basis of V_J^p, then the discrete problem is equivalent to solving a system of linear algebraic equations whose matrix is symmetric and positive definite. However, such a linear system depends on the choice of the basis functions. To avoid this dependence, we work instead with a functional description of the problem. In particular, we define the algebraic residual functional on V_J^p, for any $u_J^i \in V_J^p$, by

$$v_J \mapsto (f, v_J) - (\boldsymbol{\mathcal{K}}\nabla u_J^i, \nabla v_J) \in \mathbb{R}, \quad v_J \in V_J^p. \tag{2.4}$$

2.2. A hierarchy of meshes and spaces. We work with a hierarchy of matching simplicial meshes $\{\mathcal{T}_j\}_{0\le j\le J}$, $J \ge 1$, where $\mathcal{T}_J$ has been introduced above, and where $\mathcal{T}_j$ is a refinement of $\mathcal{T}_{j-1}$, $1 \le j \le J$. We also introduce a hierarchy of finite element spaces associated to the mesh hierarchy. For this purpose, for $j \in \{0, \ldots, J\}$, fix p_j, the polynomial degree associated to mesh level j such that $1 = p_0 \le p_1 \le \cdots \le p_{J-1} \le p_J = p$. In particular, let

$$\text{for } j = 0: \qquad V_0^1 := \mathbb{P}_1(\mathcal{T}_0) \cap H_0^1(\Omega) \quad \text{(lowest-order space)}, \tag{2.5a}$$

$$\text{for } 1 \le j \le J: \qquad V_j^{p_j} := \mathbb{P}_{p_j}(\mathcal{T}_j) \cap H_0^1(\Omega) \quad (p_j\text{th-order spaces}), \tag{2.5b}$$

where $\mathbb{P}_{p_j}(\mathcal{T}_j) := \{v_j \in L^2(\Omega),\ v_j|_K \in \mathbb{P}_{p_j}(K) \ \ \forall K \in \mathcal{T}_j\}$. Note that $V_0^1 \subset V_1^{p_1} \subset \cdots \subset V_{J-1}^{p_{J-1}} \subset V_J^{p_J} = V_J^p$.

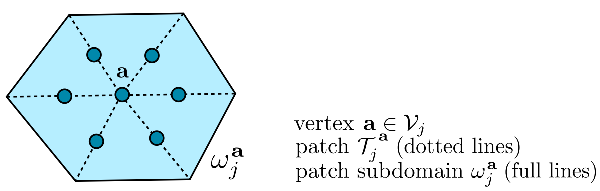

FIG. 1. *Illustration of degrees of freedom ($p_j = 2$) for the space $V_j^{\mathbf{a}}$ associated to the patch $\mathcal{T}_j^{\mathbf{a}}$.*

Let $\mathcal{V}_j$, $0 \le j \le J$, be the set of vertices of the mesh $\mathcal{T}_j$. In what follows, we need to define the notion of patches of elements illustrated in Figure 1. Given a vertex $\mathbf{a} \in \mathcal{V}_j$, we denote by $\mathcal{T}_j^{\mathbf{a}}$ all the mesh elements of $\mathcal{T}_j$ that share the vertex $\mathbf{a}$, $\mathcal{T}_j^{\mathbf{a}} := \{K \in \mathcal{T}_j, \mathbf{a} \in \mathcal{V}_K\}$, where $\mathcal{V}_K$ is the set of vertices of an element K. The corresponding open patch subdomain is denoted by $\omega_j^{\mathbf{a}}$. We also denote by $\psi_{j,\mathbf{a}}$ the standard hat function associated to the vertex $\mathbf{a} \in \mathcal{V}_j$, i.e., the piecewise affine function with respect to $\mathcal{T}_j$ taking value 1 at vertex $\mathbf{a}$ and vanishing in all other vertices of $\mathcal{V}_j$. Note that $\omega_j^{\mathbf{a}}$ is the support of $\psi_{j,\mathbf{a}}$. Finally, the local spaces $V_j^{\mathbf{a}}$ are defined by

$$V_j^{\mathbf{a}} := \mathbb{P}_{p_j}(\mathcal{T}_j) \cap H_0^1(\omega_j^{\mathbf{a}}); \tag{2.6}$$

cf. Figure 1 for the illustration of degrees of freedom when $p_j = 2$.

3. Motivation: Levelwise orthogonal decomposition of the error. It is known that a multilevel construction is required to correctly capture the behavior of the algebraic error; cf., e.g., Rüde [26], or the counterexample of Papež et al. [23, section 2.1]. Consider, for a given $u_J^i \in V_J^p$, the following (costly for practice but illustrative) hierarchical construction $\tilde{\rho}_{J,\mathrm{alg}}^i \in V_J^p$:

$$\tilde{\rho}_{J,\mathrm{alg}}^i := \rho_0^i + \sum_{j=1}^{J} \tilde{\rho}_j^i; \tag{3.1}$$

here, $\rho_0^i = \tilde{\rho}_0^i \in V_0^1$ is the solution to a global lowest-order residual problem on the coarsest mesh

$$(\boldsymbol{\mathcal{K}}\nabla \rho_0^i, \nabla v_0) = (f, v_0) - (\boldsymbol{\mathcal{K}}\nabla u_J^i, \nabla v_0) \quad \forall v_0 \in V_0^1, \tag{3.2}$$

and, moreover, for $j = 1 : J$, $\tilde{\rho}_j^i \in V_j^{p_j}$ are the solutions of

$$(\boldsymbol{\mathcal{K}}\nabla \tilde{\rho}_j^i, \nabla v_j) = (f, v_j) - (\boldsymbol{\mathcal{K}}\nabla u_J^i, \nabla v_j) - \sum_{k=0}^{j-1} (\boldsymbol{\mathcal{K}}\nabla \tilde{\rho}_k^i, \nabla v_j) \quad \forall v_j \in V_j^{p_j}. \tag{3.3}$$

This construction returns the algebraic error, i.e., $\tilde{\rho}_{J,\mathrm{alg}}^i = u_J - u_J^i$, or, equivalently,

$$u_J = u_J^i + \sum_{j=0}^{J} \tilde{\rho}_j^i. \tag{3.4}$$

This, in turn, means that $\tilde{\rho}_{J,\mathrm{alg}}^i$ satisfies

$$(\boldsymbol{\mathcal{K}}\nabla \tilde{\rho}_{J,\mathrm{alg}}^i, \nabla v_J) = (f, v_J) - (\boldsymbol{\mathcal{K}}\nabla u_J^i, \nabla v_J) \quad \forall v_J \in V_J^p. \tag{3.5}$$

Moreover, there holds $(\boldsymbol{\mathcal{K}}\nabla \tilde{\rho}_j^i, \nabla \tilde{\rho}_k^i) = 0$, for $0 \le k, j \le J,\ j \ne k$. These observations altogether lead to the orthogonal decomposition of the error between u_J^i and u_J as

$$\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla (u_J - u_J^i)\right\|^2 = \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \tilde{\rho}_{J,\mathrm{alg}}^i\right\|^2 = \sum_{j=0}^{J} \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \tilde{\rho}_j^i\right\|^2. \tag{3.6}$$

4. Multilevel solver. We now introduce our local constructions inspired by (3.1)–(3.3), producing levelwise approximations of the algebraic error components $\tilde{\rho}_j^i$ of (3.3). The construction relies on the inexpensive coarse residual solve (3.2) and on local contributions, defined on patches of elements on each level; see Figure 1. We go through the levels adding gradually levelwise updates $u_{J,j}^i$ to the current approximation u_J^i as described below. Hereafter, $(\cdot,\cdot)_{\omega_j^{\mathbf{a}}}$ stands for the $L^2(\omega_j^{\mathbf{a}})$ or $[L^2(\omega_j^{\mathbf{a}})]^d$ scalar product.

Definition 4.1 (multilevel solver).

1. *Initialize $u_J^0 \in V_J^p$ as the zero function and set $i := 0$.*
2. *Perform the following steps* (a)–(d)*:*
   (a) *Define ρ_0^i by* (3.2)*, impose $\lambda_0^i := 1$, and set*

$$u_{J,0}^i := u_J^i + \lambda_0^i \rho_0^i.$$

(b) *For $j = 1 : J$, define the local contributions $\rho^i_{j,\mathbf{a}} \in V^{\mathbf{a}}_j$ as solutions of patch problems, for all vertices $\mathbf{a} \in \mathcal{V}_j$,*

$$(\boldsymbol{\mathcal{K}}\nabla\rho^i_{j,\mathbf{a}}, \nabla v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_j} = (f, v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_j} - (\boldsymbol{\mathcal{K}}\nabla u^i_{J,j-1}, \nabla v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_j} \quad \forall v_{j,\mathbf{a}} \in V^{\mathbf{a}}_j, \tag{4.1}$$

and the descent direction $\rho^i_j \in V^{p_j}_j$ on the level j by

$$\rho^i_j := \sum_{\mathbf{a}\in\mathcal{V}_j} \rho^i_{j,\mathbf{a}}. \tag{4.2}$$

If $\rho^i_j \neq 0$, define the optimal step-size on level j by

$$\lambda^i_j := \frac{(f, \rho^i_j) - (\boldsymbol{\mathcal{K}}\nabla u^i_{J,j-1}, \nabla\rho^i_j)}{\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^i_j\|^2}; \tag{4.3}$$

otherwise, set $\lambda^i_j := 1$. Define the level update by

$$u^i_{J,j} := u^i_{J,j-1} + \lambda^i_j\rho^i_j. \tag{4.4}$$

(c) *Set the final update as $u^{i+1}_J := u^i_{J,J} \in V^p_J$.*

(d) *If $u^{i+1}_J = u^i_J$, then stop the solver. Otherwise, increase $i := i + 1$ and go to step* 2(a).

Note that by definition $\lambda^i_0 = 1$, and we thus have for $\rho^i_0 \neq 0$,

$$\frac{(f, \rho^i_0) - (\boldsymbol{\mathcal{K}}\nabla u^i_J, \nabla\rho^i_0)}{\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^i_0\|^2} \overset{(3.2)}{=} 1 = \lambda^i_0.$$

Remark 4.2 (compact writing of the iteration update). Let $u^i_J \in V^p_J$. It is easily noted that the level update (4.4) equivalently writes as

$$u^i_{J,j} = u^i_J + \sum_{k=0}^{j} \lambda^i_k\rho^i_k. \tag{4.5}$$

Thus, using the conventions $u^i_{J,-1} := u^i_J$ and $\frac{0}{0} = 0$, the new iterate after one step of the solver described in Definition 4.1 is, compared to (3.4),

$$u^{i+1}_J = u^i_J + \sum_{j=0}^{J} \lambda^i_j\rho^i_j = u^i_J + \sum_{j=0}^{J} \frac{(f, \rho^i_j) - (\boldsymbol{\mathcal{K}}\nabla u^i_{J,j-1}, \nabla\rho^i_j)}{\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^i_j\|^2}\rho^i_j. \tag{4.6}$$

The lemma below rigorously justifies the choice and use of the step-sizes (4.3).

Lemma 4.3 (levelwise optimal step-sizes). *Let $u^i_{J,j-1} \in V^p_J$ be arbitrary, let $j \in \{1, \ldots, J\}$, and let ρ^i_j and λ^i_j be given by* (4.2) *and* (4.3), *respectively. Then*

$$\lambda^i_j = \arg\min_{\lambda\in\mathbb{R}} \big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\big(u_J - (u^i_{J,j-1} + \lambda\rho^i_j)\big)\big\|.$$

Proof. We write the algebraic error associated to $u^i_{J,j-1} + \lambda\rho^i_j$ as a function of λ:

$$\begin{aligned}\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\big(u_J - (u^i_{J,j-1} + \lambda\rho^i_j)\big)\big\|^2 &= \big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u^i_{J,j-1})\big\|^2 \\ &\quad - 2\lambda\big(\boldsymbol{\mathcal{K}}\nabla(u_J - u^i_{J,j-1}), \nabla\rho^i_j\big) + \lambda^2\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^i_j\big\|^2.\end{aligned} \tag{4.7}$$

We realize that this function has a minimum, as given by (4.3), at

$$\lambda_j^i = \frac{(\mathcal{K}\nabla(u_J - u_{J,j-1}^i), \nabla\rho_j^i)}{\|\mathcal{K}^{\frac{1}{2}}\nabla\rho_j^i\|^2} \overset{(2.3)}{=} \frac{(f,\rho_j^i) - (\mathcal{K}\nabla u_{J,j-1}^i, \nabla\rho_j^i)}{\|\mathcal{K}^{\frac{1}{2}}\nabla\rho_j^i\|^2}. \qquad \square$$

Remark 4.4 (construction of the new iterate). The construction of u_J^{i+1} from u_J^i by the solver of Definition 4.1 can be seen as one iteration of a V-cycle multigrid, with no pre- and one postsmoothing step, with an optimal step-size at the error correction stage. The smoother on each level is additive Schwarz associated to patch subdomains where the local problems (4.1) are defined. Note that when $p_j = 1,\ j \in \{1,\dots,J\}$, the smoother is the diagonal Jacobi smoother, whereas when $p_j > 1$, the smoother is block-Jacobi. As detailed in [20, section 6.2], employing a weighted restricted additive Schwarz (wRAS) smoothing can offer a further speed-up of the solver, briefly addressed in section 9.4.

Remark 4.5 (connection of local contributions with levelwise updates). Note that for ρ_j^i given by (4.1)–(4.2), $j \in \{1,\dots,J\}$, we have

$$\sum_{\mathbf{a}\in\mathcal{V}_j} \|\mathcal{K}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}^2 \overset{\substack{(4.1)\\(4.2)}}{=} (f,\rho_j^i) - (\mathcal{K}\nabla u_{J,j-1}^i, \nabla\rho_j^i) \overset{(4.3)}{=} \lambda_j^i \|\mathcal{K}^{\frac{1}{2}}\nabla\rho_j^i\|^2. \tag{4.8}$$

Remark 4.6 (extension of the solver to hp-refinement hierarchy). The multilevel approach we take in this work can be easily extended to a setting where the mesh and space hierarchies are obtained by hp-refinement, since all we require in our multilevel construction of Definition 4.1 is nestedness of the meshes and finite element spaces. To obtain the theoretical results, one would need to adapt the stable decomposition results of Schöberl et al. [29] from a global fixed polynomial order to a variable one.

The optimal step-sizes also lead to the following important result, which can be compared to the orthogonal error decomposition (3.6).

THEOREM 4.7 (error representation of one solver step). *For $u_J^i \in V_J^p$, let $u_J^{i+1} \in V_J^p$ be given by Definition* 4.1*. Then*

$$\|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - u_J^{i+1})\|^2 = \|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|^2 - \sum_{j=0}^{J} \left(\lambda_j^i \|\mathcal{K}^{\frac{1}{2}}\nabla\rho_j^i\|\right)^2, \tag{4.9a}$$

$$= \|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|^2 - \|\mathcal{K}^{\frac{1}{2}}\nabla\rho_0^i\|^2 - \sum_{j=1}^{J} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{V}_j} \|\mathcal{K}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}^2. \tag{4.9b}$$

Proof. The second line (4.9b) follows immediately upon multiplying (4.8) by λ_j^i on both sides and summing over the mesh levels. We obtain the first line (4.9a) by going through the levels from finest to coarsest and using the relation of each level's update with its associated optimal step-size, similarly to (4.7):

$$
\begin{aligned}
\big\|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - u_J^{i+1})\big\|^2 &\overset{(4.4)}{=} \big\|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - (u_{J,J-1}^i + \lambda_J^i \rho_J^i))\big\|^2 \\
&\overset{\substack{(2.3)\\(4.3)}}{=} \big\|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - u_{J,J-1}^i)\big\|^2 - \big(\lambda_J^i \big\|\mathcal{K}^{\frac{1}{2}}\nabla\rho_J^i\big\|\big)^2 \\
&= \cdots = \big\|\mathcal{K}^{\frac{1}{2}}\nabla\big(u_J - (u_J^i + \lambda_0^i\rho_0^i)\big)\|^2 - \sum_{j=1}^{J}\big(\lambda_j^i\big\|\mathcal{K}^{\frac{1}{2}}\nabla\rho_j^i\big\|\big)^2 \\
&\overset{(3.2)}{=} \big\|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - u_J^i)\big\|^2 - \sum_{j=0}^{J}\big(\lambda_j^i\big\|\mathcal{K}^{\frac{1}{2}}\nabla\rho_j^i\big\|\big)^2. \qquad \square
\end{aligned}
$$

5. A posteriori estimator on the algebraic error. We now present how the solver introduced in section 4 induces an a posteriori estimator η_{alg}^i.

DEFINITION 5.1 (algebraic error estimator). *Let $u_J^i \in V_J^p$ be arbitrary, and let $u_J^{i+1} \in V_J^p$ be the update at the end of one step of the solver introduced in Definition 4.1. We define the algebraic error estimator*

$$
\eta_{\mathrm{alg}}^i := \left(\sum_{j=0}^{J}\big(\lambda_j^i\big\|\mathcal{K}^{\frac{1}{2}}\nabla\rho_j^i\big\|\big)^2\right)^{\frac{1}{2}}. \tag{5.1}
$$

Following Theorem 4.7, the estimator η_{alg}^i is immediately a guaranteed lower bound on the algebraic error.

LEMMA 5.2 (guaranteed lower bound on the algebraic error). *There holds*

$$
\big\|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - u_J^i)\big\| \geq \eta_{\mathrm{alg}}^i. \tag{5.2}
$$

6. Main results. In this section, we present the main results concerning our multilevel solver of Definition 4.1 and our a posteriori estimator η_{alg}^i of Definition 5.1. As in [20], these two results are equivalent. We first collect our assumptions.

6.1. Setting, mesh, and regularity assumptions. For any mesh level $j \in \{1, \dots, J\}$, we denote by $h_K := \mathrm{diam}(K)$ the diameter of the element $K \in \mathcal{T}_j$ and by $h_j = \max_{K\in\mathcal{T}_j} h_K$ the mesh size on level j. We shall always assume that our meshes are shape-regular.

Assumption 6.1 (mesh shape regularity). There exists $\kappa_{\mathcal{T}} > 0$ such that

$$
\max_{K\in\mathcal{T}_j} \frac{h_K}{\rho_K} \leq \kappa_{\mathcal{T}} \,\forall\, 0 \leq j \leq J, \tag{6.1}
$$

where ρ_K denotes the diameter of the largest ball contained in K.

Below, we work in one of the three following settings. In the first setting, the hierarchy consists of quasi-uniform meshes with a bounded refinement factor between consecutive levels.

Assumption 6.2 (refinement strength and mesh quasi-uniformity). There exists a fixed positive real number $0 < C_{\mathrm{ref}} \leq 1$ such that for all $j \in \{1, \dots, J\}$, for all $K \in \mathcal{T}_{j-1}$, and for any $K^* \in \mathcal{T}_j$ such that $K^* \subset K$, there holds

$$
C_{\mathrm{ref}} h_K \leq h_{K^*} \leq h_K. \tag{6.2}
$$

There further exists a fixed positive real number $0 < C_{\rm qu} \leq 1$ such that for all $j \in \{0,\dots,J\}$ and for all $K \in \mathcal{T}_j$, there holds

$$C_{\rm qu} h_j \leq h_K \leq h_j. \tag{6.3}$$

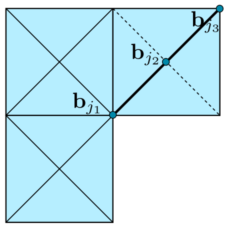

FIG. 2. *Illustration of the set $\mathcal{B}_j$; the refinement $\mathcal{T}_j$ (dotted lines) of the mesh $\mathcal{T}_{j-1}$ (full lines).*

In the second setting, we work with a hierarchy generated from a quasi-uniform coarse mesh by a series of bisections, e.g., newest vertex bisection; cf. Sewell [30] and Mitchell [21]. In this case, one refinement edge of $\mathcal{T}_{j-1}$, for $j \in \{1,\dots,J\}$, gives us a new finer mesh $\mathcal{T}_j$. We denote by $\mathcal{B}_j \subset \mathcal{V}_j$ the set consisting of the new vertex obtained after the bisection together with its two neighbors on the refinement edge; see Figure 2 for $d = 2$. We also denote by $h_{\mathcal{B}_j}$ the maximal diameter of elements having a vertex in the set $\mathcal{B}_j$, for $j \in \{1,\dots,J\}$. Here we assume the following.

Assumption 6.3 (local refinement strength and the coarsest mesh quasi-uniformity of bisection-generated meshes). The coarsest mesh $\mathcal{T}_0$ is a conforming quasi-uniform mesh in the sense of (6.3), with parameter $0 < C^0_{\rm qu} \leq 1$. The (possibly highly graded) conforming mesh $\mathcal{T}_J$ is generated from $\mathcal{T}_0$ by a series of bisections. There exists a fixed positive real number $0 < C_{\rm loc,qu} \leq 1$ such that for all $j \in \{1,\dots,J\}$, there holds

$$C_{\rm loc,qu} h_{\mathcal{B}_j} \leq h_K \leq h_{\mathcal{B}_j} \quad \forall K \in \mathcal{T}_j \text{ such that a vertex of } K \text{ belongs to } \mathcal{B}_j. \tag{6.4}$$

In the third setting, we assume the following.

Assumption 6.4 (refinement strength, mesh quasi-uniformity, and H^2-regularity). Let Assumption 6.2 hold. Moreover, let for each $g \in L^2(\Omega)$, $w_g \in H^1_0(\Omega)$ such that

$$(\nabla w_g, \nabla v) = (g, v) \quad \forall v \in H^1_0(\Omega)$$

belong to $H^2(\Omega)$.

6.2. Main results. We now present our main results, the proofs of which are given in section 10. For the solver, the following holds.

THEOREM 6.5 (p-robust error contraction of the multilevel solver). *Let $u_J \in V_J^p$ be the (unknown) finite element solution of* (2.3)*, and let $u_J^i \in V_J^p$ be arbitrary, $i \geq 0$. Take u_J^{i+1} to be constructed from u_J^i using one step of the multilevel solver of Definition* 4.1*. Under Assumption* 6.1 *and either Assumption* 6.2, 6.3*, or* 6.4*, there holds*

$$\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla(u_J - u_J^{i+1})\right\| \leq \alpha \left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla(u_J - u_J^{i})\right\|. \tag{6.5}$$

Here $0 < \alpha < 1$ depends on the space dimension d, the mesh shape regularity parameter $\kappa_{\mathcal{T}}$, the ratio of the largest and the smallest eigenvalue of the diffusion coefficient $\boldsymbol{\mathcal{K}}$, *and additionally on* (i) *the parameters C_{ref} and C_{qu} and at most linearly on the number of mesh levels J under Assumption* 6.2*;* (ii) *the parameters C^0_{qu} and $C_{\mathrm{loc,qu}}$ and at most linearly on the number of mesh levels J under Assumption* 6.3*;* (iii) *the parameters C_{ref} and C_{qu} under Assumption* 6.4*. In particular, α is independent of the polynomial degree p.*

In (6.5), α represents an upper bound on the algebraic error contraction factor at each step i. In particular, this means that the solver of Definition 4.1 contracts the algebraic error at each iteration step robustly with respect to the polynomial degree p. Moreover, under Assumption 6.4, the contraction is also robust with respect to the number of mesh levels J.

For the estimator, in turn, we have the following theorem.

THEOREM 6.6 (p-robust reliable and efficient bound on the algebraic error). *Let $u_J \in V_J^p$ be the (unknown) finite element solution of* (2.3)*, and let $u_J^i \in V_J^p$ be arbitrary, $i \geq 0$. Let η^i_{alg} be given by Definition* 5.1*. Let Assumption* 6.1 *and either Assumption* 6.2*,* 6.3*, or* 6.4 *hold. Then, in addition to $\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\| \geq \eta^i_{\mathrm{alg}}$ of* (5.2)*, there holds*

$$\eta^i_{\mathrm{alg}} \geq \beta\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|, \tag{6.6}$$

where $0 < \beta < 1$ is given by $\beta = \sqrt{1-\alpha^2}$ with α from (6.5)*.*

Theorem 6.6 allows one to write η^i_{alg} as a two-sided bound of the algebraic error (up to the constant β for the upper bound), meaning that the estimator is reliable and efficient, robustly with respect to the polynomial degree p.

6.3. Additional results. Theorems 6.5 and 6.6 are actually equivalent, similarly to [20, Corollary 5.4] (we thus only prove Theorem 6.6 in section 10).

COROLLARY 6.7 (equivalence of the p-robust solver contraction and p-robust estimator efficiency). *Let the assumptions of Theorems* 6.5 *and* 6.6 *be satisfied. Then* (6.5) *holds if and only if* (6.6) *holds, and $\alpha = \sqrt{1-\beta^2}$.*

Proof. We give the proof for completeness. Starting from (6.5), with $0 < \alpha < 1$,

$$\begin{aligned}
&\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^{i+1})\|^2 \leq \alpha^2\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|^2\\
&\overset{(4.9\mathrm{a})}{\Leftrightarrow} \|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|^2 - \sum_{j=0}^{J}(\lambda_j^i\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\|)^2 \leq \alpha^2\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|^2\\
&\overset{(5.1)}{\Leftrightarrow} \|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|^2(1-\alpha^2) \leq \left(\eta^i_{\mathrm{alg}}\right)^2.
\end{aligned}$$ ∎

Finally, the following corollary formulates a three-part equivalence (recall that the step-sizes are given by (4.3) and the local (patchwise) contributions by (4.1)).

COROLLARY 6.8 (equivalence error–estimator–localized contributions). *Let Assumption* 6.1 *hold, as well as either Assumption* 6.2*,* 6.3*, or* 6.4*. Then*

$$\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|^2 \approx \left(\eta^i_{\mathrm{alg}}\right)^2 = \|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_0^i\|^2 + \sum_{j=1}^{J}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{V}_j}\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\|^2_{\omega_j^{\mathbf{a}}}, \tag{6.7}$$

where the constant hidden in the equivalence is β from (6.6)*.*

Proof. Under Assumptions 6.2, 6.3, or 6.4, Theorem 6.6 together with (5.2) gives $\|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - u_J^i)\| \approx \eta_{\mathrm{alg}}^i$. The equality in (6.7) is easily obtained as in Theorem 4.7 upon multiplying (4.8) by λ_j^i on both sides and summing over the mesh levels. ∎

Remark 6.9 (localized a posteriori estimator of the algebraic error). The localization (6.7) is over vertex patches as in the a posteriori error estimators of the discretization error in the finite element method; see, e.g., Babuška and Rheinboldt [3] or Verfürth [33]. Therein, the construction also relies on solving local Dirichlet problems.

7. Adaptive number of smoothing steps.

We consider here a simple and practical way to make the solver described in Definition 4.1 choose autonomously and adaptively the number of smoothing steps on each mesh level. The idea of the adaptive version is to make more postsmoothing steps *if needed* on levels that contribute most to the algebraic error. This is decided relying on the a posteriori error estimate on the algebraic error we have at our disposal, relying on a Dörfler-type condition; cf. [12].

DEFINITION 7.1 (adaptive multilevel solver). *Let $\nu_{\max} \geq 1$ be a user-specified maximal number of smoothing steps, and let $0 < \theta < 1$ be a bulk-chasing parameter.*

1. *Initialize $u_J^0 \in V_J^p$ as the zero function and set $i := 0$.*
2. *Perform the following steps* (a)–(d):
   - (a) *Let ρ_0^i be constructed by* (3.2). *Set $\rho_{0,1}^i := \rho_0^i$, $\lambda_{0,1}^i := 1$, $\nu_0^i := 1$, and $u_{J,0}^i := u_J^i + \lambda_{0,1}^i \rho_{0,1}^i$.*
   - (b) *For $j = 1 : J$:*
     - i. *Set $\nu := 1$.*
     - ii. *From $u_{J,j-1}^i$, construct ρ_j^i and λ_j^i by* (4.1)–(4.3). *Set $\rho_{j,\nu}^i := \rho_j^i$, $\lambda_{j,\nu}^i := \lambda_j^i$, $u_{J,j,\nu}^i := u_{J,j-1}^i + \lambda_{j,\nu}^i \rho_{j,\nu}^i$, and*

`while` $\Big[\nu < \nu_{\max}$ `and`

$$\left(\lambda_{j,\nu}^i \|\mathcal{K}^{\frac{1}{2}}\nabla\rho_{j,\nu}^i\|\right)^2 \geq \theta^2 \Big(\sum_{k=0}^{j-1} \sum_{\ell=1}^{\nu_k^i} \left(\lambda_{k,\ell}^i \|\mathcal{K}^{\frac{1}{2}}\nabla\rho_{k,\ell}^i\|\right)^2 + \sum_{\ell=1}^{\nu-1} \left(\lambda_{j,\ell}^i \|\mathcal{K}^{\frac{1}{2}}\nabla\rho_{j,\ell}^i\|\right)^2 \Big)\Big]$$

`do` *Set $\nu := \nu + 1$.*
From $u_{J,j,\nu-1}^i$, construct $\bar\rho_j^i$ and $\bar\lambda_j^i$ by (4.1)–(4.3).
Set $\rho_{j,\nu}^i := \bar\rho_j^i$, $\lambda_{j,\nu}^i := \bar\lambda_j^i$, $u_{J,j,\nu}^i := u_{J,j,\nu-1}^i + \lambda_{j,\nu}^i \rho_{j,\nu}^i$.

`endwhile`

     - iii. *Set $\nu_j^i = \nu$ and $u_{J,j}^i := u_{J,j,\nu}^i$.*
   - (c) *Define the final update on step i as $u_J^{i+1} := u_{J,J}^i \in V_J^p$.*
   - (d) *If $u_J^{i+1} = u_J^i$, then stop the solver. Otherwise, increase $i := i+1$ and go to step* 2(a).

Remark 7.2 (adaptive substep). Note that if we skip the adaptive substep in 2(b) in Definition 7.1 by setting $\nu_{\max} = 1$, we obtain the nonadaptive version of the solver of Definition 4.1. Otherwise, we continue the smoothing iterations until the decrease of the algebraic error on mesh level j and solver iteration i, estimated by the left term in the while condition, is not important in comparison with the cumulated estimated decrease achieved so far on iteration i.

Remark 7.3 (optimal step-sizes and adaptive number of smoothing steps as a general approach). The main ideas of optimal step-size per level and adaptive number of smoothing steps we use in Definition 7.1 can be used in other geometric multigrid solvers. Implementationwise, these ideas are easy to add to existing codes and alleviate the task of choosing the number of smoothing steps arbitrarily.

Remark 7.4 (adaptivity criterion). The bulk-chasing (Dörfler's) marking criterion is not crucial above, so other criteria like the maximal one can be considered as well. We note that we do not analyze here the influence of the additional adaptive smoothing steps on the convergence speed.

8. Complexity of the solver. We wish to give some insights into the complexity of the solver of Definition 7.1 here. In particular, estimating the number of floating point operations after i_{s} iterations can be done by the formula

$$\text{nflops} := \frac{|\mathcal{V}_0|^3}{3} + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\frac{\text{ndof}(V_j^{\mathbf{a}})^3}{3} + \sum_{i=1}^{i_{\mathrm{s}}}\Bigg[2|\mathcal{V}_0|^2 + \sum_{j=1}^{J}\nu_j^i\sum_{\mathbf{a}\in\mathcal{V}_j}2\text{ndof}(V_j^{\mathbf{a}})^2\Bigg]$$
$$+\sum_{i=1}^{i_{\mathrm{s}}}\sum_{j=1}^{J}\Bigg[2\,\text{nnz}(\mathcal{I}_{j-1}^{j}) + 2\,\text{nnz}(\mathcal{I}_{j}^{j-1}) + 2\nu_j^i\,\text{nnz}(\mathbb{A}_j) + 3\nu_j^i(2\,\text{size}(\mathbb{A}_j))\Bigg]. \tag{8.1}$$

This formula is derived assuming (1) an initial Cholesky decomposition of the local matrices associated to each patch on each level except for the coarsest one, where the global stiffness matrix for piecewise affine functions is factorized (for a matrix of size n, this cost is estimated as $1/3n^3$); (2) local solves by forward and backward substitutions (cost $2n^2$); (3) intergrid operators $\mathcal{I}_{j-1}^{j} : V_{j-1}^{p_{j-1}} \to V_j^{p_j}$ with the cost estimated by twice the number of nonzeros of the associated interpolation matrix; and (4) evaluation of the optimal step-sizes λ_j as in formula (4.3) with a cost equal to twice the number of nonzeros of the stiffness matrix $\mathbb{A}_j$ on the given level and three inner products. Recall that ν_j^i is the number of smoothing steps on level j at iteration i.

We would like to point out that the above estimation (8.1) is a worst-case scenario. In fact, in the case of a structured initial mesh $\mathcal{T}_0$ containing an arbitrary number of simplices, or $\mathcal{T}_0$ only containing a few simplices and uniform or newest vertex bisection graded refinemement, most patches have the same geometry. Then the second (cubic, potentially dominant) term in (8.1) almost vanishes. Moreover, the developed solver and estimator are fully parallelizable on each mesh level and thus the discussion of complexity in floating point operations no longer has the same meaning in a parallel implementation; in particular, all the terms in (8.1) containing the sum over (all) vertices can be fully parallelized. On the other hand, formula (8.1) ignores the operations needed to evaluate the right-hand sides of local problems (4.1). Such evaluation may affect the overall flops count, but this is very dependent on the particular implementation.

9. Numerical experiments. In this section, we consider three test cases with the diffusion tensor constant in Ω, $\boldsymbol{\mathcal{K}} = I$, where the domains $\Omega \subset \mathbb{R}^2$ and the exact solutions u are given by

$$\text{Sine:} \quad u(x,y) := \sin(2\pi x)\sin(2\pi y),\ \Omega := (-1,1)^2, \tag{9.1}$$

$$\text{Peak:} \quad u(x,y) := x(x-1)y(y-1)\mathrm{e}^{-100((x-0.5)^2-(y-0.117)^2)},\ \Omega := (0,1)^2, \tag{9.2}$$

$$\text{L-shape:} \quad u(r,\varphi) := r^{2/3}\sin(2\varphi/3),\ \Omega := (-1,1)^2 \setminus ([0,1]\times[-1,0]). \tag{9.3}$$

We further consider two tests with piecewise constant diffusion tensor $\boldsymbol{\mathcal{K}} = c(x,y)\cdot I$ on the square domain. For each of these tests, we will vary $c(x,y)$ in order to study its influence on the solver's performance. The tests are described by

$$\text{Checkerboard:}\ u(r,\varphi) = r^{\gamma}\mu(\varphi),\ \Omega := (-1,1)^2, \tag{9.4}$$

where $\mu(\varphi)$ is constructed following Kellogg [17]. We consider the case $\gamma = 1$, $\boldsymbol{\mathcal{K}} = I$, and a singular solution with $\gamma = 0.0009$ and diffusion contrast 2001405.429972. For the latter, $c(x, y)$ varies across the domain as in Figure 3.

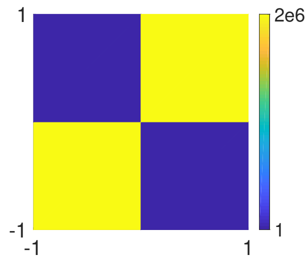

FIG. 3. *Variations of the coefficient* $c(x, y)$ *across the domain for the Checkerboard test case.*

$$\text{Skyscraper: unknown analytic solution, } \Omega := (0,1)^2. \tag{9.5}$$

The variations of $c(x, y)$ are shown in Figure 4. We take the source term $f = 1$ and Dirichlet boundary condition $u_D(x, y) = \sqrt{x}$ on $\partial\Omega$. We adjust $c(x, y)$ to obtain two tests: one with diffusion contrast proportional to 1, and another proportional to 10^7. An analogous test case is also described and used in Anciaux-Sedrakian et al. [1, section 5.3].

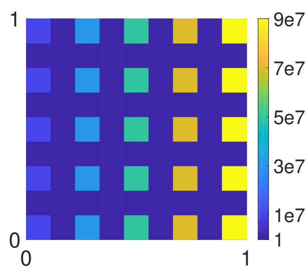

FIG. 4. *Variations of the coefficient* $c(x, y)$ *across the domain for the Skyscraper test case.*

In all tests, the exact solution of the algebraic systems is given by a direct solver.

9.1. Performance of the multilevel solver of Definition 4.1. We first consider mesh hierarchies obtained by J uniform refinements of an initial Delaunay triangulation of the domain Ω. We study the solver of Definition 4.1 stopped when the ℓ^2-norm of the algebraic residual vector drops below 10^{-5} times the initial one; then we expect for a p-robust solver that the number of iterations $i_{\rm s}$ needed to reach it will be similar for different polynomial degrees. We also numerically investigate J-robustness and robustness with respect to the jump in the diffusion coefficient, denoted henceforth by $\mathcal{J}(\boldsymbol{\mathcal{K}})$. Results presented in Table 1 confirm perfect p-robustness, as well as numerical $\boldsymbol{\mathcal{K}}$- and J-robustness even in low-regularity cases.

We now present some experiments for graded mesh hierarchies. The meshes were obtained by the newest vertex bisection algorithm; cf. Sewell [30] and Mitchell [21], and a Dörfler's bulk-chasing criterion [12] which uses the true discretizaton error and

TABLE 1

Number of iterations i_s for different polynomial degrees p, number of mesh levels J, space hierarchies with two different p_j, $j \in \{1, \ldots, J-1\}$, and jump in the diffusion coefficient $\mathcal{J}(\boldsymbol{\mathcal{K}})$.

			Sine		Peak		L-shape		Checkerboard				Skyscraper			
			$\boldsymbol{\mathcal{K}}=I$		$\boldsymbol{\mathcal{K}}=I$		$\boldsymbol{\mathcal{K}}=I$		$\boldsymbol{\mathcal{K}}=I$		$\mathcal{J}(\boldsymbol{\mathcal{K}})=O(10^6)$		$\mathcal{J}(\boldsymbol{\mathcal{K}})=O(1)$		$\mathcal{J}(\boldsymbol{\mathcal{K}})=O(10^7)$	
p_j			1	p	1	p	1	p	1	p	1	p	1	p	1	p
J	p	DoF	i_s	i_s	i_s	i_s	i_s	i_s	i_s	i_s	i_s	i_s	i_s	i_s	i_s	i_s
3	1	$2e^4$	19	19	19	19	21	21	18	18	18	18	19	19	19	19
	3	$1e^5$	29	13	28	14	29	11	27	11	28	11	31	13	31	13
	6	$6e^5$	30	13	30	14	26	9	24	9	25	10	28	11	28	11
	9	$1e^6$	31	14	30	14	23	9	23	9	23	9	26	10	26	10
4	1	$6e^4$	21	21	20	20	21	21	19	19	19	19	19	19	19	19
	3	$6e^5$	29	13	29	14	28	11	26	11	27	11	30	11	30	11
	6	$2e^6$	31	13	30	14	25	9	24	9	24	9	27	10	27	10
	9	$5e^6$	32	14	31	15	23	9	22	9	23	9	25	9	25	9

marking parameter 0.8. The true discretizaton error is used in the marking for refinement instead of an a posteriori discretization error estimator for the purpose of simplicity and result reproducibility: our main goal is to test the solver of Definition 4.1 in graded meshes that satisfy Assumption 6.3. The resulting meshes are depicted in Figure 5 for three different test cases, and the results are given in Table 2. We observe perfect p-robustness behavior of the solver of Definition 4.1, which is in agreement with our theoretical results also covering graded mesh hierarchies. Moreover, as expected from the theoretical results, the solver behaves perfectly J-robustly for the Peak test case with H^2-regular weak solution, and a linear increase of the number of iterations with respect to J appears in the Checkerboard test case for $p = 1$.

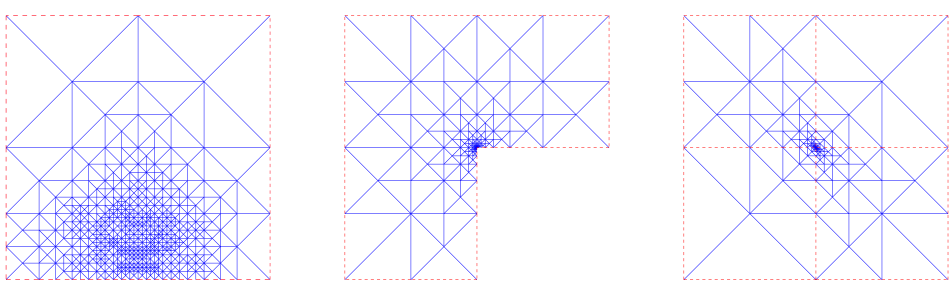

FIG. 5. *Graded meshes obtained by the newest-vertex bisection algorithm. Left: Peak problem, $J = 10$, and $p = 3$. Center: L-shape problem, $J = 10$, and $p = 3$. Right: Checkerboard $O(10^6)$, $J = 10$, and $p = 3$. The regions where the diffusion coefficient is constant are bordered by dashed lines.*

9.2. Adaptive number of smoothing steps using Definition 7.1. Now we will study the behavior of the solver described in Definition 7.1, where we set the maximum number of smoothing steps $\nu_{\max} = 5$. In order to do a comparison study of the solver's performance in different settings, we will use the estimated number of floating point operations (8.1), and we also introduce the number of global synchronizations

$$\mathrm{sync} := i_\mathrm{s} + \sum_{i=1}^{i_\mathrm{s}} \sum_{j=1}^{J} \nu_j^i. \tag{9.6}$$

TABLE 2

Number of iterations i_{s} for different polynomial degrees p, number of mesh levels J, space hierarchies given by $p_j = p$, $j \in \{1, \dots, J-1\}$, and graded mesh hierarchies.

Peak, $\boldsymbol{\mathcal{K}} = I$

J	p	i_{s}	J	p	i_{s}	J	p	i_{s}
5	1	14	10	1	16	15	1	17
	3	11		3	9		3	9
	6	8		6	8		6	8
	9	9		9	7		9	7

L-shape, $\boldsymbol{\mathcal{K}} = I$

J	p	i_{s}	J	p	i_{s}	J	p	i_{s}
5	1	16	10	1	15	15	1	17
	3	7		3	6		3	11
	6	6		6	5		6	5
	9	5		9	5		9	4

Checkerboard, $\mathcal{J}(\boldsymbol{\mathcal{K}}) = O(10^6)$

J	p	i_{s}	J	p	i_{s}	J	p	i_{s}
5	1	33	10	1	57	15	1	97
	3	15		3	23		3	32
	6	12		6	15		6	20
	9	11		9	12		9	15

9.2.1. Dependence on the parameter $\boldsymbol{\theta}$. In Figure 6 we report the cumulated number of smoothing steps employed at each level for different choices of θ. The nonadaptive variant of the solver of Definition 4.1 ($\nu_{\max} = 1$) is also plotted for comparison. Recall that this employs just one postsmoothing step and may lead to an increased number of iterations, whereas the solver of Definition 7.1 makes more smoothing steps and typically cuts the number of iterations. If in Figure 6 we find for a given θ that all numbers are consistently low for all levels, then this results in a cheaper procedure and gives us an idea of the best candidates for θ. Table 3 then gives the detailed numbers of smoothing steps per level and iteration for $\theta = 0.2$.

In Table 4 more results are presented together with the estimated costs in order to compare the performance of the solver for different values of θ. Most often, the costs are very close for different choices of θ and in practice the choice $\theta = 0.2$ is quite satisfactory. It typically brings the number of iterations down to 5–8 from 9–28, upon usually performing 2–4 postsmoothing steps on each level instead of just one. Note also that choosing θ in our setting somehow differs from typical bulk-chasing criteria, where larger θ means including more elements. Here instead, smaller θ make the condition of the while loop of Definition 7.1 more likely to be satisfied, thus leading to more smoothing steps and overall smaller iteration numbers, as seen in Table 4.

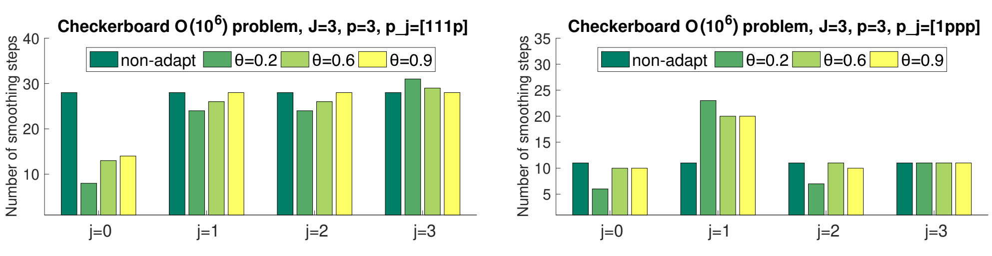

FIG. 6. *Number of smoothing steps per level for the Checkerboard case, polynomial degree $p = 3$, number of mesh levels $J = 3$, diffusion coefficient jump $\mathcal{J}(\boldsymbol{\mathcal{K}}) = O(10^6)$, and mesh hierarchies with $p_j = 1$ and $p_j = p$, $j \in \{1, \dots, J-1\}$.*

TABLE 3

Number of smoothing steps per level in each iteration it *for the Checkerboard case, $\theta = 0.2$, polynomial degree $p = 3$, number of mesh levels $J = 3$, diffusion coefficient jump $\mathcal{J}(\boldsymbol{\mathcal{K}}) = O(10^6)$, and mesh hierarchies with $p_j = 1$ and $p_j = p$, $j \in \{1, \ldots, J-1\}$. The numbers of iterations for the nonadaptive versions ($\nu_{\max} = 1$) are, respectively,* 28 *and* 11.

	$p_j = 1$								$p_j = p$					
	it=1	it=2	it=3	it=4	it=5	it=6	it=7	it=8	it=1	it=2	it=3	it=4	it=5	it=6
Level 0	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Level 1	3	3	3	3	3	3	3	3	3	4	4	4	4	4
Level 2	3	3	3	3	3	3	3	3	2	1	1	1	1	1
Level 3	3	4	4	4	4	4	4	4	2	2	2	2	2	1

TABLE 4

Estimated number of floating point operations given by (8.1) *and number of iterations i_s for two singular test cases, different polynomial degrees p, number of mesh levels J, and space hierarchies with p_j, $j \in \{0, \ldots, J\}$.*

		L-shape test case								Checkerboard O(10^6)							
		non-adapt		$\theta = 0.2$		$\theta = 0.6$		$\theta = 0.9$		non-adapt		$\theta = 0.2$		$\theta = 0.6$		$\theta = 0.9$	
J	p_j	i_s	nflops	i_s	nflops	i_s	nflops	i_s	nflops	i_s	nflops	i_s	nflops	i_s	nflops	i_s	nflops
3	1111	21	$2.17e^7$	7	$1.57e^7$	11	$1.75e^7$	11	$1.67e^7$	18	$2.01e^7$	8	$1.76e^7$	12	$1.91e^7$	11	$1.72e^7$
	1113	29	$6.05e^8$	7	$5.28e^8$	12	$5.75e^8$	15	$5.84e^8$	28	$6.05e^8$	8	$6.01e^8$	13	$5.80e^8$	14	$5.66e^8$
	1116	26	$1.20e^{10}$	7	$1.28e^{10}$	11	$1.22e^{10}$	13	$1.19e^{10}$	25	$1.21e^{10}$	8	$1.38e^{10}$	12	$1.23e^{10}$	13	$1.23e^{10}$
	1119	23	$9.08e^{10}$	6	$9.22e^{10}$	10	$9.23e^{10}$	12	$9.23e^{10}$	23	$9.39e^{10}$	7	$1.00e^{11}$	12	$9.54e^{10}$	12	$9.54e^{10}$
	1333	11	$3.90e^8$	6	$3.61e^8$	10	$4.07e^8$	10	$3.86e^8$	11	$4.04e^8$	6	$3.52e^8$	10	$4.04e^8$	10	$3.99e^8$
	1666	9	$9.49e^9$	6	$1.00e^{10}$	8	$9.53e^9$	8	$9.45e^9$	10	$1.03e^{10}$	6	$9.71e^9$	9	$1.04e^{10}$	8	$9.77e^9$
	1999	9	$9.18e^{10}$	6	$9.31e^{10}$	8	$9.21e^{10}$	8	$9.17e^{10}$	9	$9.48e^{10}$	6	$9.45e^{10}$	8	$9.51e^{10}$	8	$9.47e^{10}$
4	11111	21	$7.24e^7$	8	$5.61e^7$	11	$5.66e^7$	12	$6.00e^7$	19	$6.83e^7$	9	$6.29e^7$	11	$5.71e^7$	12	$5.92e^7$
	11113	28	$2.34e^9$	7	$2.04e^9$	12	$2.30e^9$	14	$2.19e^9$	27	$2.33e^9$	8	$2.40e^9$	12	$2.17e^9$	14	$2.26e^9$
	11116	25	$4.69e^{10}$	7	$5.00e^{10}$	11	$4.77e^{10}$	13	$4.78e^{10}$	24	$4.72e^{10}$	7	$5.04e^{10}$	12	$4.93e^{10}$	13	$4.93e^{10}$
	11119	23	$3.65e^{11}$	7	$3.97e^{11}$	10	$3.64e^{11}$	12	$3.71e^{11}$	23	$3.77e^{11}$	7	$4.03e^{11}$	11	$3.76e^{11}$	12	$3.83e^{11}$
	13333	11	$1.59e^9$	6	$1.43e^9$	9	$1.50e^9$	10	$1.61e^9$	11	$1.64e^9$	6	$1.48e^9$	9	$1.55e^9$	10	$1.59e^9$
	16666	9	$3.88e^{10}$	5	$3.65e^{10}$	8	$3.85e^{10}$	8	$3.81e^{10}$	9	$4.00e^{10}$	6	$3.99e^{10}$	9	$4.19e^{10}$	8	$3.94e^{10}$
	19999	9	$3.74e^{11}$	5	$3.64e^{11}$	8	$3.73e^{11}$	8	$3.71e^{11}$	9	$3.87e^{11}$	6	$3.78e^{11}$	8	$3.86e^{11}$	8	$3.83e^{11}$

9.2.2. Performance of the adaptive solver of Definition 7.1. In Figure 7, we fix $\theta = 0.2$ and compare our adaptive number of smoothing steps approach with the number of smoothing steps $\nu \geq 1$ being fixed to the same value on each mesh level. Relative to the adaptive approach, the solver using a fixed number of smoothing steps, whatever it is, is typically more costly, both in terms of nflops computed by (8.1) and of sync computed by (9.6). Note also that when using a fixed number of smoothing steps, the simplistic solver of Definition 4.1 ($\nu_{\max} = 1$) is often the cheapest to employ, although its number of iterations may seem rather increased at first sight. As for the adaptive solver, we also point out that the maximum number of smoothing steps $\nu_{\max} = 5$ is hardly ever reached in our experiments, endorsing our adaptive approach in two ways: a fixed number of smoothing steps is not the best way to take advantage of a multigrid solver; the criterion used for the while loop in Definition 7.1 successfully identifies the levels in which more smoothing steps are necessary, without oversmoothing.

9.3. Examples in three space dimensions. We consider now three test cases where $\Omega \subset \mathbb{R}^3$, $\boldsymbol{\mathcal{K}} = I$ except in areas of the domain explicitly specified below, and,

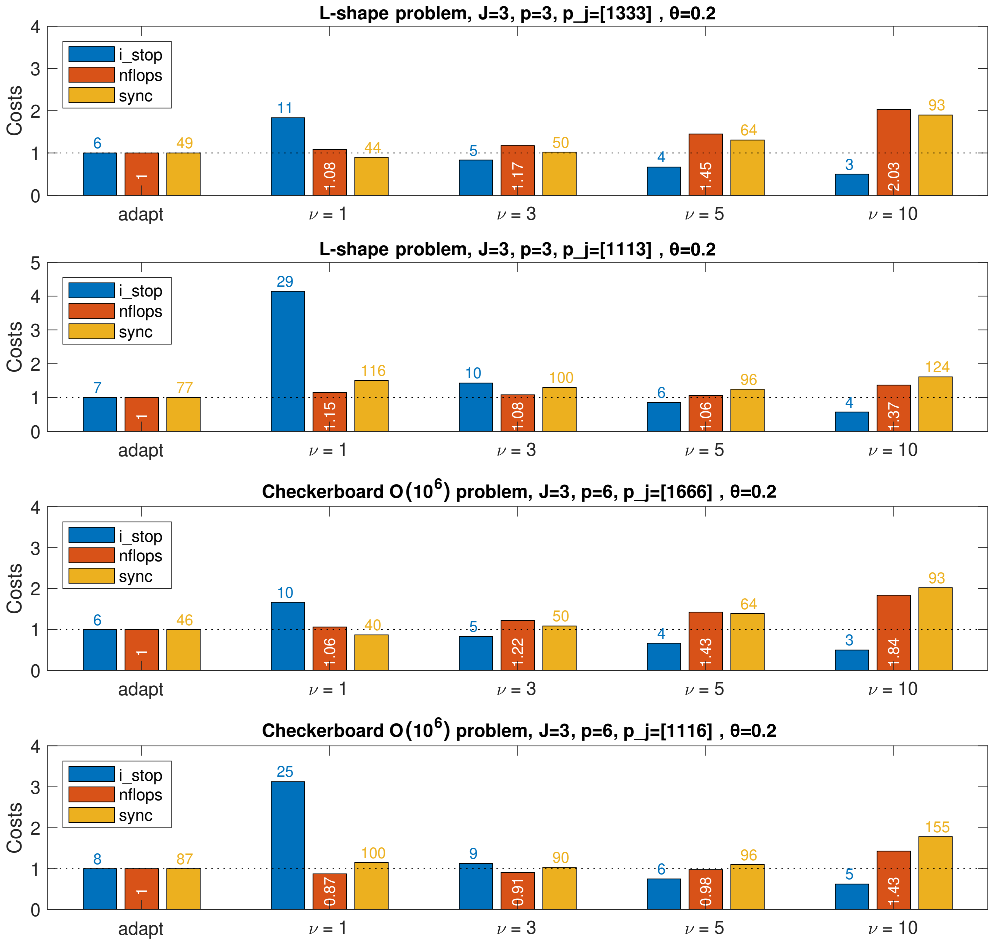

FIG. 7. *Comparison between a fixed number of (block-Jacobi) smoothing steps ν on all levels (Definition* 4.1 *and its obvious modification for $\nu \geq 1$) and the adaptive number of smoothing steps of Definition* 7.1. *Number of iterations i_s, floating point operations given by* (8.1) *relative with respect to Definition* 7.1, *and the number of global synchronizations by* (9.6).

when available, exact solution u:

$$u(x,y,z) := x(x-1)y(y-1)z(z-1), \;\; \Omega := (0,1)^3. \tag{9.7 Cube}$$

(9.7) Cube: $u(x,y,z) := x(x-1)y(y-1)z(z-1),\ \Omega := (0,1)^3.$

(9.8) Nested cubes: unknown analytic solution, $\Omega := (-1,1)^3$,
$\boldsymbol{\mathcal{K}} = 10^5 \cdot I$ in $(-0.5, 0.5)^3$.

(9.9) Checkers cubes: unknown analytic solution, $\Omega := (0,1)^3$,
$\boldsymbol{\mathcal{K}} = 10^6 \cdot I$ in $(0, 0.5)^3 \cup (0.5, 1)^3$.

In the case of nested cubes and checkers cubes, the source term is given by $f = 1$ in Ω and zero Dirichlet boundary conditions are prescribed on $\partial\Omega$.

We employ our solver of Definition 4.1 for polynomial degrees $p = 1, 2, 3, 4$, number of mesh levels $J = 4$, and hierarchies given by $p_j = 1$, $j \in \{1, \ldots, J-1\}$. The coarse mesh in all these test cases is unstructured, and the hierarchy is obtained by uniform refinement, where each tetrahedron is refined into eight new tetrahedra using the midpoints of edges in the initial tetrahedron. In Figures 8–10, we present the decay of the relative energy norm of the algebraic error and the relative ℓ^2-norm of the algebraic residual vector with respect to the iterations. Even in three space dimensions, in accordance with our theory, we see that the results are p-robust and in agreement with the more in-depth experiments of two space dimensions. Moreover, similarly to the previous tests in two space dimensions, we numerically observe that the behavior of our solver is not influenced by the magnitude of the diffusion coefficient jump. The implementation of the experiments in this section is done with NGSolve [28].

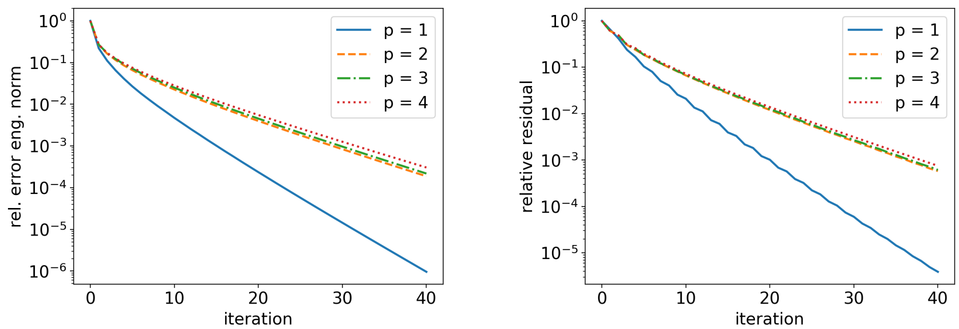

FIG. 8. *Cube case: decay of the relative algebraic error (left) and of the relative residual (right) for the hierarchy with* $p_j = 1$, $j \in \{1, \dots, J-1\}$, $J = 4$. *The solver of Definition* 4.1 *is stopped at iteration* $i = 40$. *nDoFs:* 5 501 *for* $p = 1$, 41 337 *for* $p = 2$, 136 693 *for* $p = 3$, 320 753 *for* $p = 4$.

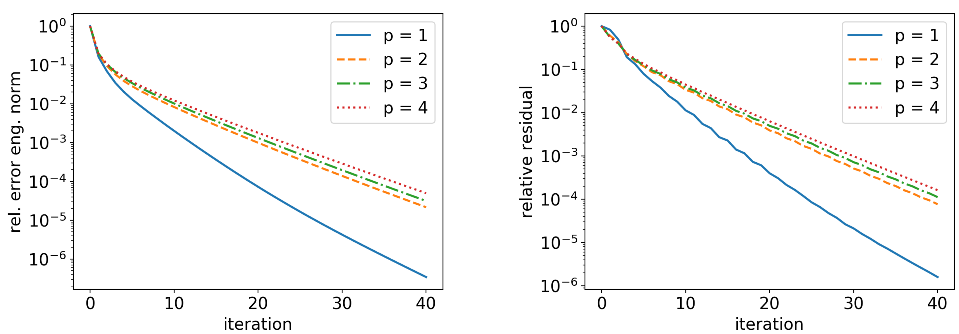

FIG. 9. *Nested cubes case: decay of the relative algebraic error (left) and of the relative residual (right) for the hierarchy with* $p_j = 1$, $j \in \{1, \dots, J-1\}$, $J = 4$. *The solver of Definition* 4.1 *is stopped at iteration* $i = 40$. *nDoFs:* 7 281 *for* $p = 1$, 55 649 *for* $p = 2$, 185 041 *for* $p = 3$, 435 393 *for* $p = 4$.

9.4. Comparison with solvers from literature. In Table 5, we finally compare our solver of Definition 4.1 (denoted as ∼MG(0,1)-bJ due to the similarity with the multigrid using only one postsmoothing step by block-Jacobi, the only difference being the use of the optimal step-size per level in the error correction stage), with different multigrid solvers used in the literature as solvers or preconditioners; see [20] for a more detailed discussion on these methods. The test case we choose has a poor

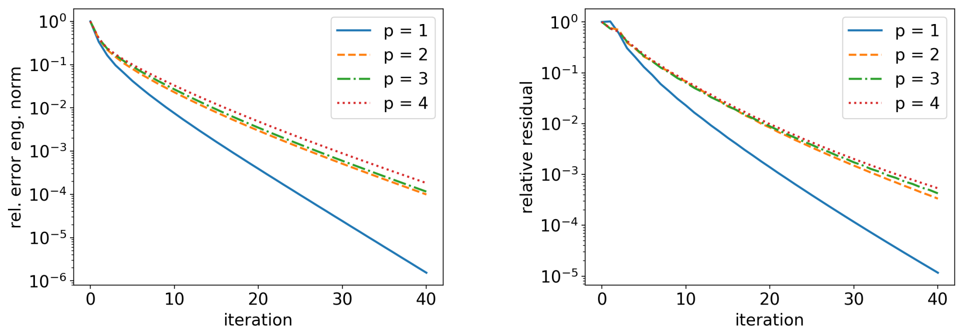

FIG. 10. *Checkers cubes: decay of the relative algebraic error (left) and of the relative residual (right) for the hierarchy with $p_j = 1$, $j \in \{1, \ldots, J-1\}$, $J = 4$. The solver of Definition* 4.1 *is stopped at iteration $i = 40$. nDoFs: 5 425 for $p = 1$, 40 033 for $p = 2$, 131 473 for $p = 3$, 307 393 for $p = 4$.*

regularity, and as we see in terms of both CPU timing[1] and iteration numbers, our solver performs well compared with the other methods, despite the more simplistic one postsmoothing step, while having the advantage of being naturally parallelizable on each level as the smoother is block-Jacobi. Importantly, note that other block smoothing methods, namely PCG(MG-bJ), which uses a symmetric multigrid with block-Jacobi smoothing as preconditioner, and MG-bGS, the multigrid using block Gauss–Seidel as smoother, also exhibit numerical p-robustness, whereas the classical MG-GS does not. In addition to the solver of Definition 7.1 with $\theta = 0.2$ and $\nu_{\max} = 5$ (denoted as $\sim$MG(0,adapt)-bJ), we also introduce its weighted restrictive additive Schwarz (wRAS) smoother variant, which outperforms the other methods while preserving numerical p-robustness. Smoothing by wRAS (see details in [20, section 6.2]) only differs from the additive Schwarz smoothing used in Definitions 4.1 and 7.1 by summing in (4.2) the local contributions $\rho^i_{j,\mathbf{a}}$ weighted by the corresponding hat functions $\psi_{j,\mathbf{a}}$ and then interpolated to the local spaces $V_j^{\mathbf{a}}$. Another important observation, as proven in, e.g., Bramble et al. [7], is that multigrid methods can perform robustly with respect to the diffusion contrast in two space dimensions. This is reflected by most methods of Table 5, having rather low iteration numbers for all diffusion tensors.

10. Proof of Theorem 6.6.

Our approach to proving Theorem 6.6 consists of studying levelwise the contributions $\tilde{\rho}^i_j$ of (3.3) of the uncomputable exact residual lifting $\tilde{\rho}^i_{J,\mathrm{alg}}$ given by (3.1). The polynomial-degree-robust stable decomposition result of Schöberl et al. [29] then allows us to exploit the similarities of the local computable contributions $\rho^i_{j,\mathbf{a}}$ (4.1) to the global inaccessible ones $\tilde{\rho}^i_j$ (3.3).

We will first present the proof of p-robust efficiency of the estimator stated in Theorem 6.6 under Assumption 6.2 or 6.3. Then we give the proof of p-robust and J-robust efficiency under Assumption 6.4. Let us start with some generalities.

10.1. Properties of the estimator η^i_{alg}.

We first present some general properties of the estimator η^i_{alg} of Definition 5.1 needed for the proof.

[1]The codes were prepared to benefit as much as possible from MATLAB's fast operations on matrices and vectors. The timings cover the solution time only, without the preparation phase of matrices assembly. The experiments were run on one Dell C6220 dual-Xeon E5-2650 node of Inria Sophia Antipolis – Méditerranée "NEF" computation cluster, in a sequential MATLAB script.

TABLE 5

Checkerboard $O(10^6)$ problem: Comparison of iteration numbers i_s and CPU times for different solvers. The horizontal/rising arrow denotes whether the polynomial degree per level remains the same/gradually increases. The number of pre- and postsmoothing steps are given in parantheses, and the smoothers are given by block-Jacobi (bJ), block Gauss–Seidel (bGS), pointwise Gauss–Seidel (GS), or PCG with incomplete Cholesky preconditioner (PCG(iChol)). The number of iterations is limited to 80.

		~MG(0,1)-bJ $1 \to 1, p$		~MG(0,1)-bJ $1, p \to p$		~MG(0, adapt)-bJ $1, p \to p$		~MG(0,adapt)-bJ (wRAS) $1 \nearrow p$		PCG(MG (3,3)-bJ) $p \to p$		MG(1,1)-PCG(iChol) $1 \nearrow p$		MG(0,1)-bGS $1 \to 1, p$		MG(3,3)-GS $1 \nearrow p$	
J	p	i_s	time	i_s	time	i_s	time	i_s	time	i_s	time	i_s	time	i_s	time	i_s	time
3	1	18	0.05 s	18	0.07 s	8	0.04 s	8	0.04 s	10	0.07 s	6	0.39 s	10	0.04 s	4	0.02 s
	3	28	0.96 s	11	0.50 s	6	0.43 s	6	0.41 s	3	0.57 s	22	3.43 s	11	2.62 s	6	0.34 s
	6	25	9.88 s	10	5.43 s	6	5.24 s	5	2.90 s	2	5.24 s	44	51.38 s	9	7.35 s	11	5.91 s
	9	23	45.87 s	9	27.01 s	6	25.25 s	4	13.86 s	2	36.95 s	>80	>5.22m	8	32.53 s	11	19.72 s
4	1	19	0.12 s	19	0.12 s	9	0.11 s	9	0.11 s	11	0.20 s	16	0.74 s	11	0.06 s	4	0.05 s
	3	27	3.85 s	11	2.07 s	6	1.89 s	7	1.62 s	3	2.34 s	44	27.48 s	10	9.64 s	5	1.37 s
	6	24	41.79 s	9	20.19 s	6	20.69 s	4	12.54 s	3	38.40 s	>80	>6.87m	9	34.78 s	6	14.44 s
	9	23	3.63m	9	2.13m	6	2.09m	3	49.84 s	2	2.24m	>80	>23.08m	8	1.72m	9	1.21m

LEMMA 10.1 (estimation of $\|\mathcal{K}^{\frac12}\nabla\rho_j^i\|$ by local contributions). *Let $\rho_{j,\mathbf{a}}^i$ and ρ_j^i for $j \in \{1,\dots,J\}$, $\mathbf{a}\in\mathcal{V}_j$, be given by* (4.1) *and* (4.2)*. Then there holds*

$$\left\|\mathcal{K}^{\frac12}\nabla\rho_j^i\right\|^2 \le (d+1)\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\right\|^2_{\omega_j^{\mathbf{a}}}. \tag{10.1}$$

Proof. Since $\rho_j^i = \sum_{\mathbf{a}\in\mathcal{V}_j}\rho_{j,\mathbf{a}}^i$, the inequality $\left|\sum_{k=1}^{d+1} a_k\right|^2 \le (d+1)\sum_{k=1}^{d+1}|a_k|^2$ leads to

$$\begin{aligned}\left\|\mathcal{K}^{\frac12}\nabla\rho_j^i\right\|^2 &= \sum_{K\in\mathcal{T}_j}\left\|\mathcal{K}^{\frac12}\nabla\rho_j^i\right\|_K^2 = \sum_{K\in\mathcal{T}_j}\left\|\sum_{\mathbf{a}\in\mathcal{V}_K}\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\right\|_K^2\\ &\le (d+1)\sum_{K\in\mathcal{T}_j}\sum_{\mathbf{a}\in\mathcal{V}_K}\left\|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\right\|_K^2 = (d+1)\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\right\|^2_{\omega_j^{\mathbf{a}}}. \qquad \square\end{aligned}$$

Remark 10.2 (lower bound on the optimal step-sizes). Note that (4.8) together with (10.1) and the definition $\lambda_j^i = 1$ when $\rho_j^i = 0$ or $j = 0$ lead to

$$\lambda_j^i \ge \frac{1}{d+1} \quad 0 \le j \le J. \tag{10.2}$$

10.2. Properties of the exact residual lifting $\tilde{\rho}^i_{J,\mathrm{alg}}$. Hereafter, we use two crucial properties of the levelwise error contributions of (3.3) $\tilde{\rho}_j^i$, $j\in\{1,\dots,J\}$: the orthogonality of $\tilde{\rho}_j^i$ with respect to previous levels and local properties of $\tilde{\rho}_j^i$ on level j.

LEMMA 10.3 (interlevel properties of $\tilde{\rho}_j^i$). *Consider the hierarchical construction of the error $\tilde{\rho}^i_{J,\mathrm{alg}}$ given in* (3.1)*. For $j\in\{1,\dots,J\}$, there holds*

$$(\mathcal{K}\nabla\tilde{\rho}_j^i, \nabla v_k) = 0 \quad \forall v_k \in V_k^{p_k}, \quad 0 \le k < j. \tag{10.3}$$

Proof. Take $v_k \in V_k^{p_k}$. Note that since $k \le j-1$, and by nestedness of the spaces, we have $v_k \in V_{j-1}^{p_{j-1}} \subset V_j^{p_j}$. The definition given in (3.3) applied to $\tilde{\rho}_j^i$ and $\tilde{\rho}_{j-1}^i$ allows us to write

$$
\begin{aligned}
(\mathcal{K}\nabla\tilde{\rho}_j^i, \nabla v_k) &= (f, v_k) - (\mathcal{K}\nabla u_J^i, \nabla v_k) - \sum_{l=0}^{j-2}(\mathcal{K}\nabla\tilde{\rho}_l^i, \nabla v_k) - (\mathcal{K}\nabla\tilde{\rho}_{j-1}^i, \nabla v_k) \\
&= (\mathcal{K}\nabla\tilde{\rho}_{j-1}^i, \nabla v_k) - (\mathcal{K}\nabla\tilde{\rho}_{j-1}^i, \nabla v_k) = 0. \qquad \square
\end{aligned}
$$

Now, we present the relation between $\tilde{\rho}_j^i$ and ρ_j^i locally on patches, more precisely when tested against functions of the local spaces $V_j^{\mathbf{a}}$ given by (2.6).

LEMMA 10.4 (local relation between $\tilde{\rho}_j^i$ and $\rho_{j,\mathbf{a}}^i$). *Let $j \in \{1, \ldots, J\}$. Let $\tilde{\rho}_j^i$, $\rho_{j,\mathbf{a}}^i$, and ρ_j^i be respectively given by* (3.3), (4.1), *and* (4.2). *For all vertices $\mathbf{a}\in\mathcal{V}_j$ and all functions $v_{j,\mathbf{a}}\in V_j^{\mathbf{a}}$, we have*

$$
(\mathcal{K}\nabla\tilde{\rho}_j^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} = (\mathcal{K}\nabla\rho_{j,\mathbf{a}}^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} - \sum_{k=1}^{j-1}(\mathcal{K}\nabla(\tilde{\rho}_k^i - \lambda_k^i\rho_k^i), \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}}. \tag{10.4}
$$

We use the convention that the sum in the relation above is zero when $j = 1$.

Proof. We take $v_{j,\mathbf{a}} \in V_j^{\mathbf{a}}$. This implies that $v_{j,\mathbf{a}}$ is zero on the boundary of the patch domain $\omega_j^{\mathbf{a}}$. Since $v_{j,\mathbf{a}} \in V_J^p$, we can use it as a test function in the definition of $\tilde{\rho}_j^i$ in (3.3) as well as in the definition of $\rho_{j,\mathbf{a}}^i$ in (4.1). We conclude by using (4.5) and subtracting the following two identities:

$$
\begin{aligned}
(\mathcal{K}\nabla\tilde{\rho}_j^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} &= (f, v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} - (\mathcal{K}\nabla u_J^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} - \sum_{k=0}^{j-1}(\mathcal{K}\nabla\tilde{\rho}_k^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}}, \\
(\mathcal{K}\nabla\rho_{j,\mathbf{a}}^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} &= (f, v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} - (\mathcal{K}\nabla u_J^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} - \sum_{k=0}^{j-1}\lambda_k^i(\mathcal{K}\nabla\rho_k^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}}. \qquad \square
\end{aligned}
$$

10.3. Proof of Theorem 6.6 under the minimal $H_0^1(\Omega)$-regularity assumption. We begin by presenting here a result given in [20, Proposition 7.6], obtained by a combination of a one-level p-robust stable decomposition proven in Schöberl et al. [29] and a multilevel stable decomposition for piecewise linear functions given in Xu, Chen, and Nochetto [35].

LEMMA 10.5 (p-robust multilevel stable decomposition). *Let $v_J \in V_J^p$. Under Assumption* 6.1 *and either Assumption* 6.2 *or* 6.3, *there exists a decomposition*

$$
v_J = v_0 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j} v_{j,\mathbf{a}}, \quad v_0 \in V_0^1,\ v_{j,\mathbf{a}} \in V_j^{\mathbf{a}}, \tag{10.5}
$$

stable as

$$
\|\nabla v_0\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\|\nabla v_{j,\mathbf{a}}\|_{\omega_j^{\mathbf{a}}}^2 \le C_{\mathrm{S}}^2\|\nabla v_J\|^2, \tag{10.6}
$$

where $C_{\mathrm{S}} \ge 1$ only depends on the space dimension d, the mesh shape regularity parameter $\kappa_{\mathcal{T}}$, and on the maximum strength of refinement parameter C_{ref} and quasi-uniformity parameter C_{qu} when Assumption 6.2 *is satisfied, or on the coarse and local quasi-uniformity parameters C_{qu}^0, $C_{\mathrm{loc,qu}}$ when Assumption* 6.3 *is satisfied.*

The previous results and properties allow us now to give concise proofs.

Proof of Theorem 6.6 (p-robust estimator efficiency under Assumption 6.2 or 6.3). Note that by (3.6), we have $\|\mathcal{K}^{\frac{1}{2}}\nabla(u_J - u_J^i)\| = \|\mathcal{K}^{\frac{1}{2}}\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\|$. Thus, we work with the exact algebraic residual lifting $\tilde{\rho}^i_{J,\mathrm{alg}}$. We begin by applying Lemma 10.5 to $\tilde{\rho}^i_{J,\mathrm{alg}}$, which allows to decompose it as

$$\tilde{\rho}^i_{J,\mathrm{alg}} = \tilde{c}^i_0 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\tilde{\rho}^i_{j,\mathbf{a}}, \quad \tilde{c}^i_0 \in V_0^1,\ \tilde{\rho}^i_{j,\mathbf{a}} \in V_j^{\mathbf{a}}, \tag{10.7}$$

$$\|\nabla\tilde{c}^i_0\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\|\nabla\tilde{\rho}^i_{j,\mathbf{a}}\|^2_{\omega_j^{\mathbf{a}}} \le C_{\mathrm{S}}^2\|\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\|^2. \tag{10.8}$$

Taking into account the variations of the diffusion coefficient $\mathcal{K}$, we have

$$\|\mathcal{K}^{\frac{1}{2}}\nabla\tilde{c}^i_0\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\mathcal{K}^{\frac{1}{2}}\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right\|^2_{\omega_j^{\mathbf{a}}} \le C_{\mathrm{S},\mathcal{K}}^2\left\|\mathcal{K}^{\frac{1}{2}}\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\right\|^2, \tag{10.9}$$

where the constant $C^2_{\mathrm{S},\mathcal{K}}$ additionally depends on the ratio of the largest and the smallest eigenvalue of the diffusion coefficient $\mathcal{K}$. Since $\max\left(1, C^2_{\mathrm{S},\mathcal{K}}\right)$ also satisfies (10.9), we can assume $C_{\mathrm{S},\mathcal{K}} \ge 1$. We use this decomposition to develop

$$\begin{aligned}
&\left\|\mathcal{K}^{\frac{1}{2}}\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\right\|^2 \overset{(10.7)}{=} \left(\mathcal{K}\nabla\tilde{\rho}^i_{J,\mathrm{alg}}, \nabla\tilde{c}^i_0 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right)\\
&\overset{(3.2)}{=} \left(\mathcal{K}\nabla\rho^i_0, \nabla\tilde{c}^i_0\right) + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left(\mathcal{K}\nabla\tilde{\rho}^i_{J,\mathrm{alg}}, \nabla\tilde{\rho}^i_{j,\mathbf{a}}\right)_{\omega_j^{\mathbf{a}}}\\
&\overset{(3.5)}{=} \left(\mathcal{K}\nabla\rho^i_0, \nabla\tilde{c}^i_0\right) + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left(\left(f, \tilde{\rho}^i_{j,\mathbf{a}}\right)_{\omega_j^{\mathbf{a}}} - \left(\mathcal{K}\nabla u^i_J, \nabla\tilde{\rho}^i_{j,\mathbf{a}}\right)_{\omega_j^{\mathbf{a}}}\right)\\
&\overset{(4.1)}{\underset{(4.5)}{=}} \left(\mathcal{K}\nabla\rho^i_0, \nabla\tilde{c}^i_0\right) + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left(\left(\mathcal{K}\nabla\rho^i_{j,\mathbf{a}}, \nabla\tilde{\rho}^i_{j,\mathbf{a}}\right)_{\omega_j^{\mathbf{a}}} + \sum_{k=0}^{j-1}\left(\lambda^i_k\mathcal{K}\nabla\rho^i_k, \nabla\tilde{\rho}^i_{j,\mathbf{a}}\right)_{\omega_j^{\mathbf{a}}}\right)\\
&= \left(\mathcal{K}\nabla\rho^i_0, \nabla\tilde{c}^i_0\right) + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left(\mathcal{K}\nabla\rho^i_{j,\mathbf{a}}, \nabla\tilde{\rho}^i_{j,\mathbf{a}}\right)_{\omega_j^{\mathbf{a}}} + \sum_{j=1}^{J}\sum_{k=0}^{j-1}\left(\lambda^i_k\mathcal{K}\nabla\rho^i_k, \sum_{\mathbf{a}\in\mathcal{V}_j}\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right).
\end{aligned}$$

We will now estimate each of the above three terms using Young's inequality and patch overlap arguments as done in the proof of Lemma 10.1. First, we have, using the fact that $\lambda^i_0 = 1$,

$$\left(\mathcal{K}\nabla\rho^i_0, \nabla\tilde{c}^i_0\right) \le \frac{C^2_{\mathrm{S},\mathcal{K}}}{2}\left(\lambda^i_0\left\|\mathcal{K}^{\frac{1}{2}}\nabla\rho^i_0\right\|\right)^2 + \frac{1}{2C^2_{\mathrm{S},\mathcal{K}}}\left\|\mathcal{K}^{\frac{1}{2}}\nabla\tilde{c}^i_0\right\|^2.$$

For the second term, we similarly obtain

$$\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left(\boldsymbol{\mathcal{K}}\nabla\rho^i_{j,\mathbf{a}},\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right)_{\omega^{\mathbf{a}}_j}\leq C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\rho^i_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j}+\frac{1}{4C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}}\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j}$$

$$\overset{\substack{(4.8)\\(10.2)}}{\leq} C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}(d+1)\sum_{j=1}^{J}\left(\lambda^i_j\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\rho^i_j\right\|\right)^2+\frac{1}{4C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}}\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j}.$$

Finally, for the third term, we have

$$\begin{aligned}&\sum_{j=1}^{J}\sum_{k=0}^{j-1}\left(\lambda^i_k\boldsymbol{\mathcal{K}}\nabla\rho^i_k,\sum_{\mathbf{a}\in\mathcal{V}_j}\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right)\\ &\leq\frac{2(d+1)C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}J}{2}\sum_{j=1}^{J}\sum_{k=0}^{j-1}\left(\lambda^i_k\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\rho^i_k\right\|\right)^2+\frac{\sum_{j=1}^{J}\sum_{k=0}^{j-1}\left\|\boldsymbol{\mathcal{K}}^{\frac12}\sum_{\mathbf{a}\in\mathcal{V}_j}\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right\|^2}{2(2(d+1)C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}J)}\\ &\leq(d+1)C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}J^2\sum_{k=0}^{J}\left(\lambda^i_k\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\rho^i_k\right\|\right)^2+\frac{1}{4C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}}\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j}.\end{aligned}$$

Summing these components together, we can now pursue our main estimate

$$\begin{aligned}\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\right\|^2&\leq 2(d+1)C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}J^2\sum_{j=0}^{J}\left(\lambda^i_j\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\rho^i_j\right\|\right)^2+\frac{\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\tilde{c}^i_0\right\|^2+\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\tilde{\rho}^i_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j}}{2C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}}\\ &\overset{\substack{(5.1)\\(10.9)}}{\leq} 2(d+1)C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}J^2\left(\eta^i_{\mathrm{alg}}\right)^2+\frac12\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\right\|^2.\end{aligned}$$

After subtracting $\frac12\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\right\|^2$ on both sides, we finally obtain the desired result

$$\left\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\right\|^2\leq 4(d+1)C^2_{\mathrm{S},\boldsymbol{\mathcal{K}}}J^2\left(\eta^i_{\mathrm{alg}}\right)^2. \tag{10.10}$$ ∎

10.4. Proof of Theorem 6.6 under the $H^2(\Omega)$-regularity assumption. Under Assumption 6.4, we now prove that the result of Theorem 6.6 holds not only p-robustly but also J-robustly. For this, we exhibit a different levelwise stable decomposition from that of section 10.3. Here, we will define the piecewise linear component of the stable decomposition via an H^1-orthogonal projection and then use a duality-type argument.

DEFINITION 10.6 (H^1-orthogonal lowest-order projection of error components). *For any $j\in\{1,\ldots,J\}$, let $\tilde{\rho}^i_j$ be given by* (3.3). *Then let $c^i_j\in V^1_j$ be the solution of*

$$(\nabla c^i_j,\nabla v_j)=(\nabla\tilde{\rho}^i_j,\nabla v_j)\quad\forall v_j\in V^1_j. \tag{10.11}$$

Remark 10.7 (orthogonality properties of c^i_j). For any $j\in\{1,\ldots,J\}$, c^i_j satisfies the following orthogonality with piecewise affine functions of previous levels:

$$(\nabla c^i_j,\nabla v_k)\overset{(10.11)}{=}(\nabla\tilde{\rho}^i_j,\nabla v_k)\overset{(10.3)}{=}0\quad\forall v_k\in V^1_k,\quad\forall 0\leq k<j. \tag{10.12}$$

LEMMA 10.8 (H^2-regularity result). *Under Assumption* 6.4, *for any* c_j^i *given by Definition* 10.6, $j \in \{1, \dots, J\}$, *there holds*

$$\|\tilde{\rho}_j^i\| \le \frac{C_{\mathrm{app}}}{C_{\mathrm{qu}} C_{\mathrm{ref}}} h_j \|\nabla \tilde{\rho}_j^i\|, \tag{10.13}$$

$$\|c_j^i\| \le \frac{C_{\mathrm{app}}}{C_{\mathrm{qu}} C_{\mathrm{ref}}} h_j \|\nabla c_j^i\|, \tag{10.14}$$

where the constant C_{app} *depends on the space dimension* d *and the mesh shape regularity parameter* $\kappa_{\mathcal{T}}$, *and* C_{qu} *and* C_{ref} *are the quasi-uniformity and refinement strength parameters from Assumption* 6.2.

Proof. To prove the first result, we proceed by a standard duality argument.

We consider the following problem: find $\xi_j \in H_0^1(\Omega)$ such that

$$(\nabla \xi_j, \nabla v) = (\tilde{\rho}_j^i, v) \quad \forall v \in H_0^1(\Omega). \tag{10.15}$$

Following Grisvard [14, Theorem 4.3.1.4], under Assumption 6.4, $\xi_j \in H^2(\Omega)$, and we have

$$|\xi_j|_{H^2(\Omega)} = \|\Delta \xi_j\| = \|\tilde{\rho}_j^i\|. \tag{10.16}$$

Consider $\mathrm{I}_{j-1}^1(\xi_j)$ the $\mathbb{P}^1$-Lagrange interpolation of ξ_j on mesh level $j-1$. Since $\xi_j \in H^2(\Omega)$, following, e.g., Ern and Guermond [13, Corollary 1.110], we obtain

$$\|\nabla(\xi_j - \mathrm{I}_{j-1}^1(\xi_j))\| \le C_{\mathrm{app}} h_{j-1} |\xi_j|_{H^2(\Omega)}. \tag{10.17}$$

In particular, $\mathrm{I}_{j-1}^1(\xi_j) \in V_{j-1}^{p_{j-1}}$, so by the orthogonality relation (10.3)

$$(\nabla \mathrm{I}_{j-1}^1(\xi_j), \nabla \tilde{\rho}_j^i) = 0. \tag{10.18}$$

We have now all the elements to conclude

$$\begin{aligned}
\|\tilde{\rho}_j^i\|^2 &\overset{(10.15)}{=} (\nabla \xi_j, \nabla \tilde{\rho}_j^i) \overset{(10.18)}{=} (\nabla(\xi_j - \mathrm{I}_{j-1}^1(\xi_j)), \nabla \tilde{\rho}_j^i) \;\le\; \|\nabla(\xi_j - \mathrm{I}_{j-1}^1(\xi_j))\| \|\nabla \tilde{\rho}_j^i\| \\
&\overset{(10.17)}{\le} C_{\mathrm{app}} h_{j-1} |\xi_j|_{H^2(\Omega)} \|\nabla \tilde{\rho}_j^i\| \overset{(10.16)}{=} C_{\mathrm{app}} h_{j-1} \|\tilde{\rho}_j^i\| \|\nabla \tilde{\rho}_j^i\| \overset{(6.2)}{\underset{(6.3)}{\le}} \frac{C_{\mathrm{app}}}{C_{\mathrm{qu}} C_{\mathrm{ref}}} h_j \|\tilde{\rho}_j^i\| \|\nabla \tilde{\rho}_j^i\|,
\end{aligned}$$

which gives us (10.13). To obtain (10.14), the same argument is used once the right-hand side of the dual problem (10.15) is modified to (c_j^i, v), and we replace the orthogonality relation (10.18) by (10.12). Note that at this point, it is important that $\mathrm{I}_{j-1}^1(\xi_j) \in V_{j-1}^1$. ∎

We can now present the stable decomposition used in the proof of Theorem 6.6.

LEMMA 10.9 (stable decomposition of the error levelwise components). *For* $\tilde{\rho}_j^i$ *given by* (3.3), c_j^i *given by Definition* 10.6, $j \in \{1, \dots, J\}$, *there exist* $\tilde{\rho}_{j,\mathbf{a}}^i \in V_j^{\mathbf{a}}$, *so that*

$$\tilde{\rho}_j^i = c_j^i + \sum_{\mathbf{a} \in \mathcal{V}_j} \tilde{\rho}_{j,\mathbf{a}}^i, \tag{10.19}$$

$$\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla c_j^i\right\|^2 + \sum_{\mathbf{a} \in \mathcal{V}_j} \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \tilde{\rho}_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2 \le C_{\mathrm{SD},\boldsymbol{\mathcal{K}}}^2 \|\boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \tilde{\rho}_j^i\|^2, \tag{10.20}$$

where $C^2_{\mathrm{SD},\boldsymbol{\mathcal{K}}} \geq 1$ only depends on the space dimension d, the mesh shape regularity parameter $\kappa_{\mathcal{T}}$, the quasi-uniformity parameter C_{qu}, the strength refinement parameter C_{ref}, and the ratio of the largest and the smallest eigenvalue of the diffusion coefficient $\boldsymbol{\mathcal{K}}$.

Proof. We now rely on the stable decomposition result of Schöberl et al. [29] for a one-level setting. We will first show, as in [29, Lemma 3.1], that the coarse contribution c^i_j satisfies

$$\|\nabla c^i_j\|^2 + \|\nabla(\tilde{\rho}^i_j - c^i_j)\|^2 + \sum_{K\in\mathcal{T}_j} h_K^{-2}\|(\tilde{\rho}^i_j - c^i_j)\|^2_K \leq \left(5 + \left(\frac{2C_{\mathrm{app}}}{C_{\mathrm{ref}}C^2_{\mathrm{qu}}}\right)^2\right)\|\nabla\tilde{\rho}^i_j\|^2. \tag{10.21}$$

Then, one can construct local contributions $\tilde{\rho}^i_{j,\mathbf{a}} \in V^{\mathbf{a}}_j$ as in [29, section 3], which by [29, Proof of Theorem 2.1] gives us

$$\|\nabla c^i_j\|^2 + \sum_{\mathbf{a}\in\mathcal{V}_j} \|\nabla\tilde{\rho}^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_j} \leq C^2_{\mathrm{SD}}\|\nabla\tilde{\rho}^i_j\|^2,$$

and the claim (10.20) follows by taking into consideration the variations of $\boldsymbol{\mathcal{K}}$.

To show (10.21), we first use Definition 10.6 of c^i_j

$$\|\nabla c^i_j\|^2 = (\nabla c^i_j, \nabla c^i_j) \overset{(10.11)}{=} (\nabla c^i_j, \nabla\tilde{\rho}^i_j) \leq \|\nabla c^i_j\|\|\nabla\tilde{\rho}^i_j\|. \tag{10.22}$$

This allows one to estimate the first and second terms (after using the triangle inequality) of (10.21). The third term is then estimated by

$$\begin{aligned}\sum_{K\in\mathcal{T}_j} h_K^{-2}\|(\tilde{\rho}^i_j - c^i_j)\|^2_K &\overset{(6.3)}{\leq} C^{-2}_{\mathrm{qu}} h_j^{-2}\sum_{K\in\mathcal{T}_j}\|(\tilde{\rho}^i_j - c^i_j)\|^2_K \leq 2C^{-2}_{\mathrm{qu}}h_j^{-2}\left(\|\tilde{\rho}^i_j\|^2 + \|c^i_j\|^2\right)\\ &\overset{\substack{(10.13)\\(10.14)}}{\leq} 2\left(\frac{C_{\mathrm{app}}}{C_{\mathrm{ref}}C^2_{\mathrm{qu}}}\right)^2\left(\|\nabla\tilde{\rho}^i_j\|^2 + \|\nabla c^i_j)\|^2\right) \overset{(10.22)}{\leq} 4\left(\frac{C_{\mathrm{app}}}{C_{\mathrm{ref}}C^2_{\mathrm{qu}}}\right)^2\|\nabla\tilde{\rho}^i_j\|^2. \qquad \square\end{aligned}$$

Remark 10.10 (localized writing of levelwise components). Note that we can decompose the piecewise linear $c^i_j \in V^1_j$ using the nodal basis functions. We can then write

$$\tilde{\rho}^i_j = c^i_j + \sum_{\mathbf{a}\in\mathcal{V}_j}\tilde{\rho}^i_{j,\mathbf{a}} = \sum_{\mathbf{a}\in\mathcal{V}_j}(c^i_{j,\mathbf{a}}\psi_{j,\mathbf{a}} + \tilde{\rho}^i_{j,\mathbf{a}}), \tag{10.23}$$

where $c^i_{j,\mathbf{a}}$ is the nodal value on vertex $\mathbf{a} \in \mathcal{V}_j$ of c^i_j, and $c^i_{j,\mathbf{a}}\psi_{j,\mathbf{a}} + \tilde{\rho}^i_{j,\mathbf{a}} \in V^{\mathbf{a}}_j$.

LEMMA 10.11 (L^2-stability of nodal decomposition). *For all $j \in \{1,\ldots,J\}$ and all $v_j \in V^1_j$ decomposed into the hat functions $v_j = \sum_{\mathbf{a}\in\mathcal{V}_j} v_{j,\mathbf{a}}\psi_{j,\mathbf{a}}$, we have*

$$\|v_j\|^2 \leq (d+1)\sum_{\mathbf{a}\in\mathcal{V}_j}\|v_{j,\mathbf{a}}\psi_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_j}, \quad \text{and} \quad \sum_{\mathbf{a}\in\mathcal{V}_j}\|v_{j,\mathbf{a}}\psi_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_j} \leq C^2_{\mathrm{nd}}\|v_j\|^2, \tag{10.24}$$

where $C_{\mathrm{nd}} \geq 1$ only depends on the space dimension d and the mesh shape regularity parameter $\kappa_{\mathcal{T}}$.

Proof. For the first estimate, we apply the usual overlapping argument as done for (10.1). As for the second estimate, consider a patch $\omega_j^{\mathbf{a}}$ and element K contained in the patch. Since $v_j \in V_j^1$ and by mesh shape regularity and equivalence of norms in finite dimension, we have

$$\begin{aligned}
\|v_{j,\mathbf{a}}\psi_{j,\mathbf{a}}\|_{\omega_j^{\mathbf{a}}} &\le C_{\kappa_{\mathcal{T}},d}\|v_{j,\mathbf{a}}\psi_{j,\mathbf{a}}\|_K \le C_{\kappa_{\mathcal{T}},d}\|v_{j,\mathbf{a}}\psi_{j,\mathbf{a}}\|_{\infty,K}|K|^{\frac{1}{2}}\\
&\le C_{\kappa_{\mathcal{T}},d}\Big\|\sum_{\mathbf{a}\in\mathcal{V}_K} v_{j,\mathbf{a}}\psi_{j,\mathbf{a}}\Big\|_{\infty,K}|K|^{\frac{1}{2}} = C_{\kappa_{\mathcal{T}},d}\|v_j\|_{\infty,K}|K|^{\frac{1}{2}}\\
&\le C_{\kappa_{\mathcal{T}},d}\widetilde{C}_{\kappa_{\mathcal{T}},d}\|v_j\|_K,
\end{aligned}$$

where $C_{\kappa_{\mathcal{T}},d} \ge 1$ and $\widetilde{C}_{\kappa_{\mathcal{T}},d} \ge 1$ only depend on the mesh shape regularity parameter $\kappa_{\mathcal{T}}$ and space dimension d. The result is obtained by summing both sides over all vertices. ∎

LEMMA 10.12 (levelwise estimation of c_j^i). *Let $j \in \{1,\dots,J\}$ and let $c_j^i = \sum_{\mathbf{a}\in\mathcal{V}_j} c_{j,\mathbf{a}}^i\psi_{j,\mathbf{a}}$ be given by Definition* 10.6. *Then there holds*

$$\sum_{\mathbf{a}\in\mathcal{V}_j}\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}c_{j,\mathbf{a}}^i\nabla\psi_{j,\mathbf{a}}\big\|_{\omega_j^{\mathbf{a}}}^2 \le C_{\mathrm{stab},\boldsymbol{\mathcal{K}}}^2\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla c_j^i\big\|^2, \tag{10.25}$$

where $C_{\mathrm{stab},\boldsymbol{\mathcal{K}}} \ge 1$ only depends on the space dimension d, the mesh shape regularity parameter $\kappa_{\mathcal{T}}$, the quasi-uniformity parameter C_{qu}, the strength refinement parameter C_{ref}, and the ratio of the largest and the smallest eigenvalue of the diffusion coefficient $\boldsymbol{\mathcal{K}}$.

Proof. We start by using an inverse inequality, denoting by $h_{\omega_j^{\mathbf{a}}}$ the diameter of patch $\omega_j^{\mathbf{a}}$, and then use the quasi-uniformity assumption (6.3)

$$\begin{aligned}
&\sum_{\mathbf{a}\in\mathcal{V}_j}\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}c_{j,\mathbf{a}}^i\nabla\psi_{j,\mathbf{a}}\|_{\omega_j^{\mathbf{a}}}^2 \le C_{\boldsymbol{\mathcal{K}}}^2\sum_{\mathbf{a}\in\mathcal{V}_j}\|c_{j,\mathbf{a}}^i\nabla\psi_{j,\mathbf{a}}\|_{\omega_j^{\mathbf{a}}}^2 \le C_{\boldsymbol{\mathcal{K}}}^2C_{\mathrm{inv}}^2\sum_{\mathbf{a}\in\mathcal{V}_j}h_{\omega_j^{\mathbf{a}}}^{-2}\|c_{j,\mathbf{a}}^i\psi_{j,\mathbf{a}}\|_{\omega_j^{\mathbf{a}}}^2\\
&\le C_{\boldsymbol{\mathcal{K}}}^2C_{\mathrm{qu}}^{-2}C_{\mathrm{inv}}^2h_j^{-2}\sum_{\mathbf{a}\in\mathcal{V}_j}\|c_{j,\mathbf{a}}^i\psi_{j,\mathbf{a}}\|_{\omega_j^{\mathbf{a}}}^2 \overset{(10.24)}{\le} C_{\boldsymbol{\mathcal{K}}}^2C_{\mathrm{qu}}^{-2}C_{\mathrm{inv}}^2C_{\mathrm{nd}}^2h_j^{-2}\|c_j^i\|^2\\
&\overset{(10.14)}{\le} \frac{C_{\boldsymbol{\mathcal{K}}}^2C_{\mathrm{inv}}^2C_{\mathrm{nd}}^2C_{\mathrm{app}}^2}{C_{\mathrm{qu}}^4C_{\mathrm{ref}}^2}\|\nabla c_j^i\|^2 \le \frac{C_{\boldsymbol{\mathcal{K}}}^2C_{\mathrm{inv}}^2C_{\mathrm{nd}}^2C_{\mathrm{app}}^2}{C_{\mathrm{qu}}^4C_{\mathrm{ref}}^2c_{\boldsymbol{\mathcal{K}}}^2}\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla c_j^i\|^2 \;=\; C_{\mathrm{stab},\boldsymbol{\mathcal{K}}}^2\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla c_j^i\|^2,
\end{aligned}$$

where $c_{\boldsymbol{\mathcal{K}}}^2$, $C_{\boldsymbol{\mathcal{K}}}^2$ are respectively constants that depend on the smallest and the largest eigenvalue of the diffusion coefficient $\boldsymbol{\mathcal{K}}$. Note that the resulting constant $C_{\mathrm{stab},\boldsymbol{\mathcal{K}}}$ can be safely assumed to be greater than 1; otherwise replace it by $\max(1, C_{\mathrm{stab},\boldsymbol{\mathcal{K}}})$. ∎

LEMMA 10.13 (p-robust levelwise error estimation). *Let $j \in \{1,\dots,J\}$ and let $\tilde{\rho}_j^i$ and ρ_j^i be defined by* (3.3) *and* (4.2), *respectively. Then there holds*

$$\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\tilde{\rho}_j^i\big\|^2 \le 2C_{\mathrm{SD},\boldsymbol{\mathcal{K}}}^2C_{\mathrm{stab},\boldsymbol{\mathcal{K}}}^2\sum_{\mathbf{a}\in\mathcal{V}_j}\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\big\|_{\omega_j^{\mathbf{a}}}^2. \tag{10.26}$$

Proof. We begin by using the splitting (10.19) in the form (10.23), which gives

$$
\begin{aligned}
\big\|\mathcal{K}^{\frac12}\nabla\tilde\rho_j^i\big\|^2 &\overset{(10.23)}{=} \sum_{\mathbf{a}\in\mathcal{V}_j}\big(\mathcal{K}\nabla\tilde\rho_j^i,\nabla(c_{j,\mathbf{a}}^i\psi_{j,\mathbf{a}}+\tilde\rho_{j,\mathbf{a}}^i)\big)_{\omega_j^{\mathbf{a}}}\\
&\overset{(10.4)}{=}\sum_{\mathbf{a}\in\mathcal{V}_j}\Big(\big(\mathcal{K}\nabla\rho_{j,\mathbf{a}}^i,\nabla(c_{j,\mathbf{a}}^i\psi_{j,\mathbf{a}}+\tilde\rho_{j,\mathbf{a}}^i)\big)_{\omega_j^{\mathbf{a}}}-\sum_{k=1}^{j-1}\big(\mathcal{K}\nabla(\tilde\rho_k^i-\lambda_k^i\rho_k^i),\nabla(c_{j,\mathbf{a}}^i\psi_{j,\mathbf{a}}+\tilde\rho_{j,\mathbf{a}}^i)\big)_{\omega_j^{\mathbf{a}}}\Big)\\
&\overset{(10.23)}{=}\sum_{\mathbf{a}\in\mathcal{V}_j}\big(\mathcal{K}\nabla\rho_{j,\mathbf{a}}^i,\nabla(c_{j,\mathbf{a}}^i\psi_{j,\mathbf{a}}+\tilde\rho_{j,\mathbf{a}}^i)\big)_{\omega_j^{\mathbf{a}}}-\sum_{k=1}^{j-1}\big(\mathcal{K}\nabla(\tilde\rho_k^i-\lambda_k^i\rho_k^i),\nabla\tilde\rho_j^i\big)\\
&\overset{(10.3)}{=}\sum_{\mathbf{a}\in\mathcal{V}_j}\big(\mathcal{K}\nabla\rho_{j,\mathbf{a}}^i,\nabla\big(c_{j,\mathbf{a}}^i\psi_{j,\mathbf{a}}+\tilde\rho_{j,\mathbf{a}}^i\big)\big)_{\omega_j^{\mathbf{a}}}-0\\
&\le C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2\sum_{\mathbf{a}\in\mathcal{V}_j}\big\|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\big\|_{\omega_j^{\mathbf{a}}}^2+\frac{\sum_{\mathbf{a}\in\mathcal{V}_j}\big\|\mathcal{K}^{\frac12}\nabla(c_{j,\mathbf{a}}^i\psi_{j,\mathbf{a}}+\tilde\rho_{j,\mathbf{a}}^i)\big\|_{\omega_j^{\mathbf{a}}}^2}{4C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2}\\
&\le C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2\sum_{\mathbf{a}\in\mathcal{V}_j}\|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}^2+\frac{\sum_{\mathbf{a}\in\mathcal{V}_j}\big(\big\|\mathcal{K}^{\frac12}c_{j,\mathbf{a}}^i\nabla\psi_{j,\mathbf{a}}\big\|_{\omega_j^{\mathbf{a}}}^2+\big\|\mathcal{K}^{\frac12}\nabla\tilde\rho_{j,\mathbf{a}}^i\big\|_{\omega_j^{\mathbf{a}}}^2\big)}{2C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2}\\
&\overset{(10.25)}{\le}C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2\sum_{\mathbf{a}\in\mathcal{V}_j}\big\|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\big\|_{\omega_j^{\mathbf{a}}}^2+\frac{C_{\mathrm{stab},\mathcal{K}}^2\big\|\mathcal{K}^{\frac12}\nabla c_j^i\big\|^2+\sum_{\mathbf{a}\in\mathcal{V}_j}\big\|\mathcal{K}^{\frac12}\nabla\tilde\rho_{j,\mathbf{a}}^i\big\|_{\omega_j^{\mathbf{a}}}^2}{2C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2}\\
&\overset{(10.20)}{\le}C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2\sum_{\mathbf{a}\in\mathcal{V}_j}\big\|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\big\|_{\omega_j^{\mathbf{a}}}^2+\frac12\big\|\mathcal{K}^{\frac12}\nabla\tilde\rho_j^i\big\|^2,
\end{aligned}
$$

which leads to the assertion (10.26). ∎

We can now give a concise proof of Theorem 6.6.

Proof of Theorem 6.6 (p- and J-robust estimator efficiency under Assumption 6.4). To estimate the algebraic error, we use the levelwise decomposition (3.6). Each level's contribution was estimated in Lemma 10.13. Summing over different levels,

$$
\begin{aligned}
\big\|\mathcal{K}^{\frac12}\nabla(u_J-u_J^i)\big\|^2 &\overset{(3.6)}{=}\sum_{j=0}^{J}\big\|\mathcal{K}^{\frac12}\nabla\tilde\rho_j^i\big\|^2\overset{(10.26)}{\le}\big\|\mathcal{K}^{\frac12}\nabla\rho_0^i\big\|^2+2C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\big\|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\big\|_{\omega_j^{\mathbf{a}}}^2\\
&\overset{(4.8),(10.2)}{\le}2C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2(d+1)\Big(\big(\lambda_0^i\big\|\mathcal{K}^{\frac12}\nabla\rho_0^i\big\|\big)^2+\sum_{j=1}^{J}\big(\lambda_j^i\big\|\mathcal{K}^{\frac12}\nabla\rho_j^i\big\|\big)^2\Big)\\
&\overset{(5.1)}{=}2C_{\mathrm{SD},\mathcal{K}}^2C_{\mathrm{stab},\mathcal{K}}^2(d+1)\big(\eta_{\mathrm{alg}}^i\big)^2.
\end{aligned}
$$

Thus we have showed $\eta_{\mathrm{alg}}^i\ge\beta\big\|\mathcal{K}^{\frac12}\nabla(u_J-u_J^i)\big\|$ for $\beta:=\frac{1}{\sqrt{2(d+1)}C_{\mathrm{SD},\mathcal{K}}C_{\mathrm{stab},\mathcal{K}}}>0$. ∎

11. Conclusions and future work. In this work we presented a multilevel algebraic solver whose construction is inherently interconnected with an a posteriori estimator of the algebraic error. The solver can be seen as a geometric multigrid relying on V-cycles with zero pre- and one postsmoothing, where the smoother is additive Schwarz associated to patches of elements (block-Jacobi). A crucial difference compared to classic multigrid solvers is the use of an optimal step-size in the error correction stage on each level of the mesh hierarchy. This significantly improves the behavior of the solver and conveniently enough, makes the analysis easier leading in particular to the Pythagorean error decrease formula (4.9a). We also presented a simple and efficient way for the solver to automatically increase the number of postsmoothing steps on each level to the amount needed, based on the a posteriori estimator of the algebraic error. We showed that the nonadaptive version of the solver (with only one postsmoothing step) contracts the error in each iteration robustly with respect to the polynomial degree p of the underlying finite element discretization; this result is equivalent to showing p-robust efficiency of the a posteriori algebraic error estimate. If we, additionally, assume H^2-regularity in the sense of Assumption 6.4, we can show that these results are also robust with respect to the number of mesh levels J. An interesting side property is that the error estimator is equivalent to a sum of level- and patchwise-localized computable contributions by formula (6.7). Future work [19] will explore how to incorporate this information in the solver so that it adaptively tackles only problematic regions contributing most to the algebraic error (local adaptive smoothing). Finally, numerical results indicate that even for singular test cases, for quasi-uniform meshes, the solver behaves robustly with respect to the polynomial degree p, the number of levels J, as well as the diffusion coefficient $\boldsymbol{\mathcal{K}}$.

Acknowledgment. The authors are grateful to Inria Sophia Antipolis – Méditerranée "NEF" computation cluster for providing resources and support.

REFERENCES

[1] A. Anciaux-Sedrakian, L. Grigori, Z. Jorti, J. Papež, and S. Yousef, *Adaptive solution of linear systems of equations based on a posteriori error estimators*, Numer. Algorithms, 84 (2020), pp. 331–364, https://doi.org/10.1007/s11075-019-00757-z.

[2] P. F. Antonietti and G. Pennesi, *V-cycle multigrid algorithms for discontinuous Galerkin methods on non-nested polytopic meshes*, J. Sci. Comput., 78 (2019), pp. 625–652, https://doi.org/10.1007/s10915-018-0783-x.

[3] I. Babuška and W. C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal., 15 (1978), pp. 736–754, https://doi.org/10.1137/0715049.

[4] R. E. Bank, T. F. Dupont, and H. Yserentant, *The hierarchical basis multigrid method*, Numer. Math., 52 (1988), pp. 427–458, https://doi.org/10.1007/BF01462238.

[5] J. H. Bramble and J. E. Pasciak, *New convergence estimates for multigrid algorithms*, Math. Comp., 49 (1987), pp. 311–329, https://doi.org/10.2307/2008314.

[6] J. H. Bramble, J. E. Pasciak, and A. H. Schatz, *The construction of preconditioners for elliptic problems by substructuring.* I, Math. Comp., 47 (1986), pp. 103–134, https://doi.org/10.2307/2008084.

[7] J. H. Bramble, J. E. Pasciak, J. P. Wang, and J. Xu, *Convergence estimates for multigrid algorithms without regularity assumptions*, Math. Comp., 57 (1991), pp. 23–45, https://doi.org/10.2307/2938661.

[8] A. Brandt, *Multi-level adaptive solutions to boundary-value problems*, Math. Comp., 31 (1977), pp. 333–390, https://doi.org/10.2307/2006422.

[9] A. Brandt, S. McCormick, and J. Ruge, *Algebraic multigrid (AMG) for sparse matrix equations*, in Sparsity and its Applications (Loughborough, 1983), Cambridge University Press, Cambridge, 1985, pp. 257–284.

[10] C. Canuto and A. Quarteroni, *Preconditioned minimal residual methods for Chebyshev spectral calculations*, J. Comput. Phys., 60 (1985), pp. 315–337, https://doi.org/10.1016/0021-9991(85)90010-5.

[11] T. F. Chan and W. L. Wan, *Robust multigrid methods for nonsmooth coefficient elliptic linear systems*, J. Comput. Appl. Math., 123 (2000), pp. 323–352, https://doi.org/10.1016/S0377-0427(00)00411-8.

[12] W. Dörfler, *A convergent adaptive algorithm for Poisson's equation*, SIAM J. Numer. Anal., 33 (1996), pp. 1106–1124, https://doi.org/10.1137/0733054.

[13] A. Ern and J.-L. Guermond, *Theory and Practice of Finite Elements*, Appl. Math. Sci. 159, Springer-Verlag, New York, 2004, https://doi.org/10.1007/978-1-4757-4355-5.

[14] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Monogr. Stud. Math. 24, Pitman (Advanced Publishing Program), Boston, 1985.

[15] W. Heinrichs, *Line relaxation for spectral multigrid methods*, J. Comput. Phys., 77 (1988), pp. 166–182, https://doi.org/10.1016/0021-9991(88)90161-1.

[16] B. Janssen and G. Kanschat, *Adaptive multilevel methods with local smoothing for H^1- and H^{curl}-conforming high order finite element methods*, SIAM J. Sci. Comput., 33 (2011), pp. 2095–2114, https://doi.org/10.1137/090778523.

[17] R. B. Kellogg, *On the Poisson equation with intersecting interfaces*, Appl. Anal., 4 (1975), pp. 101–129, https://doi.org/10.1080/00036817408839086.

[18] J. P. Lucero Lorca and G. Kanschat, *Multilevel Schwarz preconditioners for singularly perturbed symmetric reaction-diffusion systems*, Electron. Trans. Numer. Anal., 54 (2021), pp. 89–107, https://doi.org/10.1553/etna_vol54s89.

[19] A. Miraçi, J. Papež, and M. Vohralík, *Contractive local adaptive smoothing based on Dörfler's marking in a-posteriori-steered p-robust multigrid solvers*, Comput. Methods Appl. Math., 2021, https://doi.org/10.1515/cmam-2020-0024.

[20] 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 (2020), pp. 2856–2884, https://doi.org/10.1137/19M1275929.

[21] W. F. Mitchell, *Adaptive refinement for arbitrary finite-element spaces with hierarchical bases*, J. Comput. Appl. Math., 36 (1991), pp. 65–78, https://doi.org/10.1016/0377-0427(91)90226-A.

[22] P. Oswald, *Multilevel Finite Element Approximation. Theory and Applications*, Teubner Skripten zur Numerik [Teubner Scripts on Numerical Mathematics], B. G. Teubner, Stuttgart, 1994, https://doi.org/10.1007/978-3-322-91215-2.

[23] J. Papež, U. Rüde, M. Vohralík, and B. Wohlmuth, *Sharp algebraic and total a posteriori error bounds for h and p finite elements via a multilevel approach. Recovering mass balance in any situation*, Comput. Methods Appl. Mech. Engrg., 371 (2020), 113243, https://doi.org/10.1016/j.cma.2020.113243.

[24] L. F. Pavarino, *Additive Schwarz methods for the p-version finite element method*, Numer. Math., 66 (1994), pp. 493–515, https://doi.org/10.1007/BF01385709.

[25] A. Quarteroni and G. Sacchi Landriani, *Domain decomposition preconditioners for the spectral collocation method*, J. Sci. Comput., 3 (1988), pp. 45–76, https://doi.org/10.1007/BF01066482.

[26] U. Rüde, *Mathematical and Computational Techniques for Multilevel Adaptive Methods*, Frontiers Appl. Math. 13, SIAM, Philadelphia, 1993, https://doi.org/10.1137/1.9781611970968.

[27] J. W. Ruge and K. Stüben, *Algebraic multigrid*, in Multigrid Methods, Frontiers Appl. Math. 3, SIAM, Philadelphia, 1987, pp. 73–130, https://doi.org/10.1137/1.9781611971057.ch4.

[28] J. Schöberl, *C++11 Implementation of Finite Elements in NGsolve*, Technical Report, ASC Report 30/2014, Institute for Analysis and Scientific Computing, Vienna University of Technology, Vienna, Austria, 2014, https://ngsolve.org.

[29] 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), pp. 1–24, https://doi.org/10.1093/imanum/drl046.

[30] E. G. Sewell, *Automatic Generation of Triangulations for Piecewise Polynomial Approximation*, ProQuest LLC, Ann Arbor, MI, Ph.D. thesis, Purdue University, West Lafayette, IN, 1972.

[31] H. Sundar, G. Stadler, and G. Biros, *Comparison of multigrid algorithms for high-order continuous finite element discretizations*, Numer. Linear Algebra Appl., 22 (2015), pp. 664–680, https://doi.org/10.1002/nla.1979.

[32] A. Thekale, T. Gradl, K. Klamroth, and U. Rüde, *Optimizing the number of multigrid cycles in the full multigrid algorithm*, Numer. Linear Algebra Appl., 17 (2010), pp. 199–210, https://doi.org/10.1002/nla.697.

[33] R. Verfürth, *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Teubner-Wiley, Stuttgart, 1996.

[34] J. Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Rev., 34 (1992), pp. 581–613, https://doi.org/10.1137/1034116.

[35] J. Xu, L. Chen, and R. H. Nochetto, *Optimal multilevel methods for $H(\mathrm{grad})$, $H(\mathrm{curl})$, and $H(\mathrm{div})$ systems on graded and unstructured grids*, in Multiscale, Nonlinear and Adaptive Approximation, Springer, Berlin, 2009, pp. 599–659, https://doi.org/10.1007/978-3-642-03413-8_14.