A multilevel algebraic error estimator and the corresponding iterative solver with p-robust behavior

Ani Miraçi, Jan Papež, Martin Vohralík

A multilevel algebraic error estimator and the corresponding iterative solver with p-robust behavior*

Ani Miraçi[†‡] Jan Papež[†‡] Martin Vohralík[†‡]

October 14, 2020

Abstract

In this work, we consider conforming finite element discretizations of arbitrary polynomial degree $p \geq 1$ of the Poisson problem. We propose a multilevel a posteriori estimator of the algebraic error. We prove that this estimator is reliable and efficient (represents a two-sided bound of the error), with a constant independent of the degree p. We next design a multilevel iterative algebraic solver from our estimator and show that this solver contracts the algebraic error on each iteration by a factor bounded independently of p. Actually, we show that these two results are equivalent. The p-robustness results rely on the work of Schöberl et al. [IMA J. Numer. Anal., 28 (2008), pp. 1–24] for one given mesh. We combine this with the design of an algebraic residual lifting constructed over a hierarchy of nested unstructured, possibly highly graded, simplicial meshes. The lifting includes a global coarse-level solve with the lowest polynomial degree one together with local contributions from the subsequent mesh levels. These contributions, of the highest polynomial degree p on the finest mesh, are given as solutions of mutually independent local Dirichlet problems posed over overlapping patches of elements around vertices. The construction of this lifting can be seen as one geometric V-cycle multigrid step with zero pre- and one postsmoothing by (damped) additive Schwarz (block Jacobi). One particular feature of our approach is the optimal choice of the step-size generated from the algebraic residual lifting. Numerical tests are presented to illustrate the theoretical findings.

Key words: finite element method, stable decomposition, multilevel method, Schwarz method, a posteriori estimate, p-robustness

1 Introduction

The finite element method (FEM) is a widespread approach for discretizing problems given in the form of partial differential equations, and has been used in engineering for more than fifty years. For a thorough overview on the topic, we refer the reader to, e.g., Ciarlet [18], Ern and Guermond [20], and Brenner and Scott [14]. Many iterative methods have been suggested to treat the linear systems arising from finite element discretizations; see e.g., Bramble et al. [11] and [12], Hackbusch [24], Bank, Dupont, and Yserentant [6], Brandt, McCormick, and Ruge [13], Oswald [37], or Zhang [51], and the references therein. A systematic description of iterative solvers is given by Xu in [49]. For convergence results on unstructured and graded meshes, we refer the reader to, e.g., Wu and Chen [48], Hiptmair, Wu, and Zheng [26], Chen, Nochetto, and Xu [17], and Xu, Chen, and Nochetto [50]. The convergence of these methods is typically robust with respect to the size of the mesh (h-robustness). In fact, this is one of the key advantages of multigrid methods. For the conjugate gradient method on the other hand, h-robustness is not intrinsic; this problem can be bypassed with the development of appropriate preconditioners.

*This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation program (grant agreement No 647134 GATIPOR). The work of Jan Papež was supported by the NLAFET project as part of European Union's Horizon 2020 research and innovation program under grant 671633. The authors are grateful to Inria Sophia Antipolis - Méditerranée "NEF" computation cluster for providing resources and support.

[†]Inria, 2 rue Simone Iff, 75589 Paris, France

[‡]Université Paris-Est, CERMICS (ENPC), 77455 Marne-la-Vallée, France

If we are to consider methods of arbitrary approximation polynomial degree, an additional question arises: how does the polynomial degree p affect the performance of the method? In this regard, results for p-version FEM include Foresti et al. [21] for two-dimensional domains, Mandel [33] for three-dimensional domains, and Babuška *et al.* [5] for two-dimensional domains. For the latter, the condition number of the preconditioned system grows at most by $1 + \log^2(p)$, and a generalization of this work for hp-FEM is given by Ainsworth [1], where the p-dependence is still present. An early version of a polynomial-degree robust (p-robust) solver was introduced by Quarteroni and Sacchi Landriani [42] for a specific domain configuration (decomposable into rectangles without internal cross points). A notable development on p-robustness was later made by Pavarino [41] for quadrilateral/hexahedral meshes, where the author introduced a p-robust additive Schwarz method. The generalization of this result for triangular/tetrahedral meshes was achieved by Schöberl et al. [43], once more by introducing an additive Schwarz preconditioner. More recent works were carried out based on these approaches. In Antonietti et al. [3] (see also the references therein), the p-robust approach for rectangular/hexahedral meshes was used for high-order discontinuous Galerkin (DG) methods; moreover the spectral bounds of the preconditioner are also robust with respect to the method's penalization coefficient. We also mention the introduction of multilevel preconditioners yielded by block Gauss–Seidel smoothers in Kanschat [29] for rectangular/hexahedral meshes and DG discretizations. Further multilevel approaches for rectangular/hexahedral meshes using overlapping or nonoverlapping Schwarz smoothers can be found in, e.g., Janssen and Kanschat [27] and Lucero Lorca and Kanschat [32]. For a study on more general meshes, see, e.g., Antonietti and Pennesi [2], where a multigrid approach behaves p-robustly under the condition that the number of smoothing steps (depending on p) is chosen sufficiently large. Another notable contribution is the design of algebraic multigrid methods (AMG) via aggregation techniques; see, e.g., Notay and Napov [36], Bastian, Blatt, and Scheichl [7], and the references therein. The numerical results of the latter give a satisfactory indication of p-robustness.

An associated topic is the development of estimates of the algebraic error. In this regard, a posteriori tools have primarily been used to estimate the algebraic error for existing solvers. One particular goal is the development of reliable stopping criteria, allowing one to avoid unnecessary iterations. This is achieved with a combination of a posteriori error estimators for the discretization error. Some contributions on this matter (see also references therein) include Becker, Johnson, and Rannacher [8] where adaptive error control is achieved for a multigrid solver, Bornemann and Deuflhard [9], where a one-way multigrid method is presented by integrating an adaptive stopping criterion based on a posteriori tools. Further developments were made in Meidner, Rannacher, and Vihharev [34], where goal-oriented error estimates are used in the framework of the dual weighted residual (DWR) method. In Jiránek, Strakoš, and Vohralík [28] and later in Papež, Strakoš, and Vohralík [39], upper and lower bounds for both the algebraic and total errors are computed, which allow one to derive guaranteed upper and lower bounds on the discretization error, and consecutively construct safe stopping criteria for iterative algebraic solvers. Arioli, Georgoulis, and Loghin [4] propose practical stopping criteria which guarantee that the considered inexact adaptive FEM algorithm converges for inexact solvers of Krylov subspace type. To the best of the authors' knowledge, though, dedicated proofs of efficiency of a posteriori estimators of the algebraic error have so far not been presented.

In this work, we present an a posteriori algebraic error estimator and a multilevel iterative solver associated to it. The cornerstone of their definitions lies in the multilevel construction of a residual algebraic lifting, motivated partly by the approach of Papež et al. [38]. The lifting can be seen as an approximation of the algebraic error by continuous piecewise polynomials of degree p, obtained by a V-cycle multigrid method with no presmoothing step and a single postsmoothing step. The coarse correction is given by a lowest polynomial degree (piecewise affine) function. Our smoothing is chosen in the family of damped additive Schwarz (block Jacobi) methods applied to overlapping subdomains composed of patches of elements (two options for defining the patches will be given in due time) and corresponds to local Dirichlet problems with the highest p-degree on the finest mesh. Note that additive Schwarz-type smoothing allows for a parallelizable implementation at each level of the mesh hierarchy. Once this lifting is built, the a posteriori estimator is easily derived as a natural guaranteed lower bound on the algebraic error, following [38] and the references therein. As our first main result, we show that up to a p-robust constant, the estimator is also an upper bound on the error.

Our solver is then defined as a linear iterative method. Because we have at hand the residual lifting, which approximates the algebraic error, we use it as a descent direction (the asymmetric, since no presmoothing is used, approach in defining the lifting will not be a problem for the analysis). The step-size is then chosen

by a line search in the direction of the lifting. Our choice presents a resemblance to the conjugate gradient method, in that we choose the step-size that ensures the best error contraction in the energy norm at the next iteration. Other precedents of the use of optimal step-sizes include, e.g., Canuto and Quarteroni [16], and in the multigrid setting Heinrichs [25]. As our second main result, we prove that this solver contracts the error at each iteration by a p-robust constant. Actually, we also show that the p-robust efficiency of the estimator is equivalent to the p-robust convergence of the solver. All these results are defined for a general hierarchy of nested, unstructured, possibly highly refined (graded) matching simplicial meshes, and no assumption beyond $u \in H^1_0(\Omega)$ is imposed on the weak solution.

The work is structured as follows. In section 2, we introduce the setting in which we will be working as well as the notation employed throughout the paper. Then, we introduce our multilevel residual lifting construction in section 3, following Papež et al. [38]. In section 4, we present the a posteriori estimator on the algebraic error and the corresponding multilevel solver based on the residual lifting. Our main results are presented in the form of two theorems in section 5, together with a corollary establishing their equivalence. Another important corollary is the equivalence of the algebraic error with a computable estimator which is localized levelwise as well as patchwise. We provide numerical experiments in section 6, focusing mainly on showcasing p-robustness, in agreement with our theoretical results, and on a comparison with several existing approaches. We also introduce a weighted restricted additive Schwarz variant of our solver. The proofs of our main results are given in section 7. In particular, for the stable decomposition estimate, the p-robust result on one level introduced by Schöberl et al. [43] is crucial. We also rely on the multilevel stable splitting of Xu, Chen, and Nochetto [50] for $p = 1$ to obtain acceptable estimates with respect to the number of levels. Finally, section 8 brings forth our conclusions and outlook for future work.

2 Setting

We will consider in this work the Poisson problem defined over $\Omega \subset \mathbb{R}^d$, $d \in \{1, 2, 3\}$, an open bounded polytope with a Lipschitz-continuous boundary.

2.1 Model problem

Let $f \in L^2(\Omega)$ be the source term. We consider the following problem: find $u : \Omega \to \mathbb{R}$ such that

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2.1}$$

In the weak formulation, we search for $u \in H^1_0(\Omega)$ such that

$$(\nabla u, \nabla v) = (f, v) \quad \forall v \in H^1_0(\Omega), \tag{2.2}$$

where $(\cdot, \cdot)$ is the $L^2(\Omega)$ or $[L^2(\Omega)]^d$ scalar product. The existence and uniqueness of the solution of (2.2) follows from the Riesz representation theorem.

2.2 Finite element discretization

Let $\mathcal{T}_J$ be a given simplicial mesh of Ω. Fixing an integer $p \geq 1$, we introduce the finite element space of continuous piecewise p-degree polynomials

$$V_J^p := \mathbb{P}_p(\mathcal{T}_J) \cap H^1_0(\Omega), \tag{2.3}$$

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

$$(\nabla u_J, \nabla v_J) = (f, v_J) \quad \forall v_J \in V_J^p. \tag{2.4}$$

2.3 Algebraic system, approximate solution, and algebraic residual

If one introduces ψ_J^l, $1 \le l \le N_J$, a basis of V_J^p, then problem (2.4) is equivalent to solving a system of linear algebraic equations. Denoting by $(\mathbb{A}_J)_{lm} := (\nabla\psi_J^m, \nabla\psi_J^l)$ the symmetric, positive definite (stiffness) matrix and by $(\mathrm{F}_J)_l := (f, \psi_J^l)$ the right-hand side (load) vector, one obtains $u_J = \sum_{m=1}^{N_J} (\mathrm{U}_J)_m \psi_J^m$, where $\mathrm{U}_J \in \mathbb{R}^{N_J}$ is the solution of

$$\mathbb{A}_J \mathrm{U}_J = \mathrm{F}_J.$$

For any approximation $\mathrm{U}_J^i \in \mathbb{R}^{N_J}$ of U_J given by an arbitrary algebraic solver at iteration step $i \ge 0$, the associated continuous piecewise polynomial of degree p is $u_J^i = \sum_{m=1}^{N_J} (\mathrm{U}_J^i)_m \psi_J^m \in V_J^p$. The associated algebraic residual vector is given by

$$\mathrm{R}_J^i := \mathrm{F}_J - \mathbb{A}_J \mathrm{U}_J^i.$$

Note, however, that R_J^i depends on the choice of the basis functions ψ_J^l, $1 \le l \le N_J$. To avoid this dependence, we work instead with a residual functional on V_J^p given by

$$v_J \mapsto (f, v_J) - (\nabla u_J^i, \nabla v_J) \in \mathbb{R}, \quad v_J \in V_J^p. \tag{2.5}$$

We emphasize that the forthcoming results are independent of the choice of the basis.

2.4 A hierarchy of meshes

We consider a hierarchy of nested matching simplicial meshes $\{\mathcal{T}_j\}_{0\le j\le J}$, $J \ge 1$, where $\mathcal{T}_J$ was introduced in section 2.2, and where $\mathcal{T}_j$ is a refinement of $\mathcal{T}_{j-1}$, $1 \le j \le J$. For any element K on a given mesh, we denote $h_K := \mathrm{diam}(K)$ and by $\mathcal{V}_K$ the set of its vertices. We also denote $h_j := \max_{K\in\mathcal{T}_j} h_K$ for $0 \le j \le J$. Hereafter, we shall always assume that our meshes are shape regular.

Assumption 2.1 (Shape regularity)**.** *There exists $\kappa_{\mathcal{T}} > 0$ such that*

$$\max_{K\in\mathcal{T}_j} \frac{h_K}{\rho_K} \le \kappa_{\mathcal{T}} \text{ for all } 0 \le j \le J, \tag{2.6}$$

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

Additionally to the above assumption, we will treat below two specific cases. In the first one, we suppose quasi-uniformity of the meshes in the hierarchy and that the strength of refinement is bounded. In the second case, we suppose that the meshes are generated by a series of bisections, e.g., the newest vertex bisection; cf. Sewell [45].

2.4.1 A hierarchy of quasi-uniform meshes

We assume quasi-uniformity and that the hierarchy of meshes is such that the size of each parent element is comparable to the size of each of its children.

Assumption 2.2 (Maximum refinement strength and mesh quasi-uniformity)**.** *There exists $0 < C_{\mathrm{ref}} \le 1$, a fixed positive real number such that for any $j \in \{1, \dots, J\}$, $\forall K \in \mathcal{T}_{j-1}$, and for any $K^* \in \mathcal{T}_j$ such that $K^* \subset K$, there holds*

$$C_{\mathrm{ref}} h_K \le h_{K^*} \le h_K. \tag{2.7}$$

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

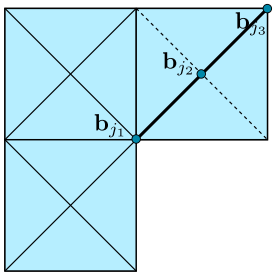
$$C_{\mathrm{qu}} h_j \le h_K \le h_j. \tag{2.8}$$

2.4.2 A hierarchy of graded bisection meshes

In the case of graded mesh hierarchies obtained by bisection, one refinement of an edge of $\mathcal{T}_{j-1}$, for $j \in \{1,\dots,J\}$, gives 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, cf. Figure 1 for an illustration when $d = 2$. We denote by $h_{\mathcal{B}_j}$ the maximal diameter of elements having a vertex in $\mathcal{B}_j$. This setting is described by the following.

Assumption 2.3 (Local quasi-uniformity of bisection-generated meshes)**.** *$\mathcal{T}_0$ is a conforming quasi-uniform mesh with parameter C^0_{qu}. The graded conforming mesh $\mathcal{T}_J$ is generated from $\mathcal{T}_0$ by a series of bisections. There exists a fixed positive real number $C^{\mathrm{loc}}_{\mathrm{qu}}$ such that for any $j \in \{1,\dots,J\}$, there holds*

$$C^{\mathrm{loc}}_{\mathrm{qu}} h_{\mathcal{B}_j} \le h_K \le h_{\mathcal{B}_j} \quad \forall K \in \mathcal{T}_j \text{ such that a vertex of } K \text{ belongs to } \mathcal{B}_j. \tag{2.9}$$



$\mathcal{T}_j$ obtained by a bisection of $\mathcal{T}_{j-1}$
new vertex after refinement $\mathbf{b}_{j_2}$
neighboring vertices on the refinement edge $\mathbf{b}_{j_1}, \mathbf{b}_{j_3}$
$\mathcal{B}_j = \{\mathbf{b}_{j_1}, \mathbf{b}_{j_2}, \mathbf{b}_{j_3}\} \subset \mathcal{V}_j$

Figure 1: Illustration of the set $\mathcal{B}_j$. The mesh $\mathcal{T}_{j-1}$ and its refinement $\mathcal{T}_j$ are given by full and dotted lines, respectively.

2.5 A hierarchy of spaces

In the following, we will need to introduce a hierarchy of finite element spaces associated to the mesh hierarchy. For this purpose, let $p' \in \{1,\dots,p\}$ be a polynomial degree between 1 and p that we employ for the intermediate levels. In particular, let the following hold:

$$\text{for } 1 \le j \le J-1, \qquad V_j^{p'} := \mathbb{P}_{p'}(\mathcal{T}_j) \cap H^1_0(\Omega) \quad (p'\text{th order spaces}), \tag{2.10a}$$

$$\text{for } j = 0, \qquad V_0^1 = \mathbb{P}_1(\mathcal{T}_0) \cap H^1_0(\Omega) \quad (\text{lowest-order space}), \tag{2.10b}$$

where $\mathbb{P}_{p'}(\mathcal{T}_j) := \{v_j \in L^2(\Omega), v_j|_K \in \mathbb{P}_{p'}(K) \ \ \forall K \in \mathcal{T}_j\}$ for $1 \le j \le J-1$. Note that $V_0^1 \subset V_1^{p'} \subset \cdots \subset V_{J-1}^{p'} \subset V_J^p$. Let $\mathcal{V}_j$ be the set of vertices of the mesh $\mathcal{T}_j$. We denote by $\psi_{j,\mathbf{a}}$ the standard hat function associated to the vertex $\mathbf{a} \in \mathcal{V}_j$, $0 \le j \le J$. This is the piecewise affine function with respect to the mesh $\mathcal{T}_j$ that takes value 1 in the vertex $\mathbf{a}$ and vanishes in all other jth level vertices of $\mathcal{V}_j$.

2.6 Two types of patches

For the following, we define two types of patches of elements. In order to facilitate the work with both, we introduce a switching parameter $s \in \{0,1\}$. First, given a vertex $\mathbf{a} \in \mathcal{V}_{j-s}$, $j \in \{1,\dots,J\}$, we denote by $\mathcal{T}^{\mathbf{a}}_{j,s}$ the patch formed by all elements of the mesh $\mathcal{T}_{j-s}$ sharing the vertex $\mathbf{a}$, i.e.,

$$\mathcal{T}^{\mathbf{a}}_{j,s} := \{K \in \mathcal{T}_{j-s}, \mathbf{a} \in \mathcal{V}_K\}. \tag{2.11}$$

We also denote by $\omega^{\mathbf{a}}_{j,s}$ the open patch subdomain corresponding to $\mathcal{T}^{\mathbf{a}}_{j,s}$. An illustration is given in Figure 2 (left) for "small" patches $s = 0$ and (right) for "large" patches $s = 1$. Then the associated local space $V^{\mathbf{a}}_{j,s}$ is given by

$$V^{\mathbf{a}}_{j,s} := \mathbb{P}_{p'}(\mathcal{T}_j) \cap H^1_0(\omega^{\mathbf{a}}_{j,s}), \ j \in \{1,\dots,J-1\} \text{ and } V^{\mathbf{a}}_{J,s} := \mathbb{P}_p(\mathcal{T}_J) \cap H^1_0(\omega^{\mathbf{a}}_{J,s}). \tag{2.12}$$

Note that $V^{\mathbf{a}}_{j,s}$ are continuous piecewise polynomial spaces with respect to the mesh $\mathcal{T}_j$ for both $s = 0$ and $s = 1$, the support being bigger in the latter case. An illustration is also given in Figure 2.

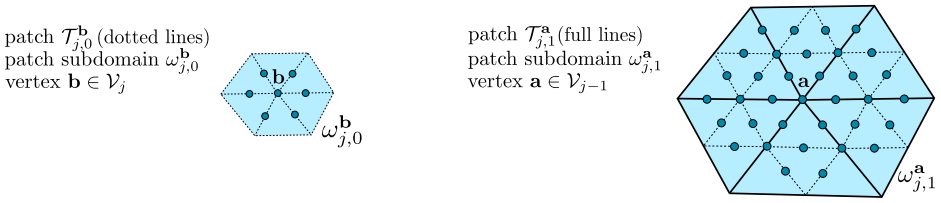


Figure 2: Illustration of degrees of freedom ($p' = p = 2$) for the space $V_{j,0}^{\mathbf{b}}$ associated to the "small" patch $\mathcal{T}_{j,0}^{\mathbf{b}}$ (left) and for the space $V_{j,1}^{\mathbf{a}}$ associated to the "large" patch $\mathcal{T}_{j,1}^{\mathbf{a}}$ (right). The mesh $\mathcal{T}_{j-1}$ and its refinement $\mathcal{T}_j$ are defined in bold and dotted lines, respectively.

3 Multilevel lifting of the algebraic residual

In the spirit of Papež et al. [38], we design a multilevel lifting of the algebraic residual given by (2.5). This lifting will lead to the construction of an a posteriori error estimator; it will also serve as a descent direction for the solver we introduce in the next section.

3.1 Exact algebraic residual lifting

For illustration and theoretical analysis later, we introduce the following definition.

Definition 3.1 (Exact residual lifting)**.** *Let $u_J^i \in V_J^p$ be arbitrary. We introduce $\tilde{\rho}_{J,\mathrm{alg}}^i \in V_J^p$ as the solution of the residual problem*

$$(\nabla \tilde{\rho}_{J,\mathrm{alg}}^i, \nabla v_J) = (f, v_J) - (\nabla u_J^i, \nabla v_J) \quad \forall v_J \in V_J^p, \tag{3.1}$$

so that

$$\tilde{\rho}_{J,\mathrm{alg}}^i = u_J - u_J^i. \tag{3.2}$$

3.2 Coarse solve

The first step of our construction is to solve a global lowest-order problem on the coarsest mesh. Let $u_J^i \in V_J^p$ be given. Recalling that $V_0^1 = \mathbb{P}_1(\mathcal{T}_0) \cap H_0^1(\Omega)$, we define $\rho_0^i \in V_0^1$ by

$$(\nabla \rho_0^i, \nabla v_0) = (f, v_0) - (\nabla u_J^i, \nabla v_0) \quad \forall v_0 \in V_0^1. \tag{3.3}$$

Note that due to (3.3) and (3.1), we have

$$(\nabla \rho_0^i, \nabla v_0) = (\nabla \tilde{\rho}_{J,\mathrm{alg}}^i, \nabla v_0) \quad \forall v_0 \in V_0^1, \tag{3.4}$$

so that ρ_0^i is the orthogonal projection of $\tilde{\rho}_{J,\mathrm{alg}}^i$ onto the coarsest space V_0^1.

3.3 Multilevel algebraic residual lifting

Let us now introduce our hierarchical construction of the algebraic residual lifting $\rho_{J,\mathrm{alg}}^i \in V_J^p$ that is hopefully close to $\tilde{\rho}_{J,\mathrm{alg}}^i$. The construction relies on the use of a coarse solution of (3.3) and on local contributions arising from all the finer mesh levels. These local contributions are defined on patch subdomains $\omega_{j,s}^{\mathbf{a}}$. We denote by $(\cdot,\cdot)_{\omega_{j,s}^{\mathbf{a}}}$ the $L^2(\omega_{j,s}^{\mathbf{a}})$ or $[L^2(\omega_{j,s}^{\mathbf{a}})]^d$ scalar product. Since we consider two definitions of patches with switching parameter $s \in \{0,1\}$ (see section 2.6), two constructions of $\rho_{J,\mathrm{alg}}^i$ are implied.

Definition 3.2 (Construction of the algebraic residual lifting)**.** *Let $w_1, w_2 \in \mathbb{R} \cup \{\infty\}$ be damping weights satisfying the conditions*

$$1 \le w_1 < 6J(d+1) \quad \textit{and} \quad w_2 \ge \max\left(1, \frac{5J^2(d+1)^2}{w_1(6J(d+1) - w_1)}\right). \tag{3.5}$$

Let $u_J^i \in V_J^p$ be arbitrary. We introduce $\rho_{J,\mathrm{alg}}^i \in V_J^p$ by

$$\rho_{J,\mathrm{alg}}^i := \rho_0^i + \sum_{j=1}^{J} \rho_j^i, \tag{3.6}$$

where $\rho_0^i \in V_0^1$ solves (3.3) and $\rho_j^i \in V_j^{p'}$, for $j \in \{1, \dots, J-1\}$, and $\rho_J^i \in V_J^p$ are given by

$$\rho_j^i := \frac{1}{w_1} \sum_{\mathbf{a} \in \mathcal{V}_{j-s}} \rho_{j,\mathbf{a}}^i, \ 1 \le j \le J, \tag{3.7}$$

with the local contributions $\rho_{j,\mathbf{a}}^i \in V_{j,s}^{\mathbf{a}}$ given by patch problems, $\forall v_{j,\mathbf{a}} \in V_{j,s}^{\mathbf{a}}$

$$(\nabla \rho_{j,\mathbf{a}}^i, \nabla v_{j,\mathbf{a}})_{\omega_{j,s}^{\mathbf{a}}} = (f, v_{j,\mathbf{a}})_{\omega_{j,s}^{\mathbf{a}}} - (\nabla u_J^i, \nabla v_{j,\mathbf{a}})_{\omega_{j,s}^{\mathbf{a}}} - \frac{1}{w_2} \sum_{k=0}^{j-1} (\nabla \rho_k^i, \nabla v_{j,\mathbf{a}})_{\omega_{j,s}^{\mathbf{a}}}. \tag{3.8}$$

Remark 3.3 (Construction of $\rho_{J,\mathrm{alg}}^i$)**.** *The construction (3.6)–(3.8) of $\rho_{J,\mathrm{alg}}^i$ can be seen as an approximation of $\tilde{\rho}_{J,\mathrm{alg}}^i$ from (3.1) by one iteration of a V-cycle multigrid, with no presmoothing and a single postsmoothing step, corresponding to a "damped" additive Schwarz iteration, with the damping factor determined by the weights w_1 and w_2. The subdomains for the Schwarz method correspond to the patch domains where the local problems in (3.8) are defined. Two patch options as in Figure 2 are considered. In particular, for $p = 1$ and "small" patches, $s = 0$ (Figure 2, left), this corresponds to one step of the Jacobi (diagonal) smoother, whereas when $p' = p > 1$, the smoother is block Jacobi. A weighted variant of Definition 3.2 is tested in section 6.*

Remark 3.4 (Value of the damping parameter)**.** *Condition (3.5) is based on the proofs in section 7 below, where the use of appropriate damping seems crucial. This is what is also indicated numerically to be needed in our approach. Possible combinations of the damping weights satisfying (3.5) include, for example,*

$$w_1 = J(d+1) \text{ and } w_2 = 1, \tag{3.9a}$$
$$w_1 = d+1 \text{ and } w_2 = J, \tag{3.9b}$$
$$w_1 = w_2 = \sqrt{J(d+1)}, \tag{3.9c}$$
$$w_1 = 1 \text{ and } w_2 = \infty, \tag{3.9d}$$
$$w_1 = 4\sqrt{J} \text{ and } w_2 = \infty. \tag{3.9e}$$

Examples (3.9a)–(3.9c) above result in a procedure that is additive patchwise and multiplicative levelwise. Examples (3.9d)–(3.9e), in turn, result in a completely additive patchwise and levelwise procedure, which is fully parallelizable. We also note that when the intermediate polynomial degree is $p' = 1$ and for any choice with $w_2 = 1$, the smoothing resulting from Definition 3.2 is local with respect to mesh $\mathcal{T}_0$ for graded meshes; it is actually only performed there where the meshes $\mathcal{T}_j$, $j \ge 1$, are different from $\mathcal{T}_0$.

4 An a posteriori estimator on the algebraic error and a multilevel solver

We now present how the residual lifting $\rho_{J,\mathrm{alg}}^i$ of Definition 3.2 can be used to define an a posteriori estimator as well as a multilevel solver.

4.1 A posteriori estimate on the algebraic error

We begin by introducing η_{alg}^i, an a posteriori estimator defined using the residual lifting $\rho_{J,\mathrm{alg}}^i$.

Definition 4.1 (Lower bound algebraic error estimator)**.** *Let $u_J^i \in V_J^p$ be arbitrary and let $\rho_{J,\mathrm{alg}}^i$ be the algebraic residual lifting given by Definition 3.2. If $\rho_{J,\mathrm{alg}}^i = 0$, we define the lower bound algebraic error estimator $\eta_{\mathrm{alg}}^i := 0$. Otherwise, set*

$$\eta_{\mathrm{alg}}^i := \frac{(f, \rho_{J,\mathrm{alg}}^i) - (\nabla u_J^i, \nabla \rho_{J,\mathrm{alg}}^i)}{\|\nabla \rho_{J,\mathrm{alg}}^i\|}. \tag{4.1}$$

The estimator η_{alg}^i is immediately a guaranteed lower bound on the algebraic error; cf., e.g., [38, Theorem 5.3].

Lemma 4.2 (Guaranteed lower bound on the algebraic error)**.** *There holds*

$$\|\nabla(u_J - u_J^i)\| \geq \eta_{\mathrm{alg}}^i. \tag{4.2}$$

Proof. Note that if $\rho_{J,\mathrm{alg}}^i = 0$, then $\|\nabla(u_J - u_J^i)\| \geq 0 = \eta_{\mathrm{alg}}^i$. Otherwise

$$\|\nabla(u_J - u_J^i)\| = \max_{\substack{v_J \in V_J^p, \\ \|\nabla v_J\| \neq 0}} \frac{(\nabla(u_J - u_J^i), \nabla v_J)}{\|\nabla v_J\|} \geq \frac{(\nabla(u_J - u_J^i), \nabla \rho_{J,\mathrm{alg}}^i)}{\|\nabla \rho_{J,\mathrm{alg}}^i\|} \overset{\substack{(2.4) \\ (4.1)}}{=} \eta_{\mathrm{alg}}^i.$$

□

4.2 Multilevel solver

We will now reuse the construction of $\rho_{J,\mathrm{alg}}^i$ given in Definition 3.2 to obtain an approximation of u_J on a next step in view of constructing a multilevel solver. Note that for any $u_J^i \in V_J^p$, the lifting $\rho_{J,\mathrm{alg}}^i$ is built to approximate the algebraic error $\tilde{\rho}_{J,\mathrm{alg}}^i$ given in (3.1), where we recall that $u_J = u_J^i + \tilde{\rho}_{J,\mathrm{alg}}^i$. Thus, it seems reasonable to consider a linear iterative solver of the form

$$u_J^{i+1} := u_J^i + \lambda \rho_{J,\mathrm{alg}}^i, \tag{4.3}$$

where $\lambda \in \mathbb{R}$ is a real parameter. The optimal choice of λ is given below.

Lemma 4.3 (Optimal step-size)**.** *Consider a solver of the form (4.3) and suppose $\rho_{J,\mathrm{alg}}^i \neq 0$. Then the choice $\lambda := [(f, \rho_{J,\mathrm{alg}}^i) - (\nabla u_J^i, \nabla \rho_{J,\mathrm{alg}}^i)]/\|\nabla \rho_{J,\mathrm{alg}}^i\|^2$ leads to minimal algebraic error with respect to the energy norm.*

Proof. We write the algebraic error of the next iteration as a function of λ,

$$\|\nabla(u_J - u_J^{i+1})\|^2 = \|\nabla(u_J - u_J^i)\|^2 - 2\lambda(\nabla(u_J - u_J^i), \nabla \rho_{J,\mathrm{alg}}^i) + \lambda^2 \|\nabla \rho_{J,\mathrm{alg}}^i\|^2, \tag{4.4}$$

and realize that this function has a minimum at

$$\lambda_{\min} = \frac{(\nabla(u_J - u_J^i), \nabla \rho_{J,\mathrm{alg}}^i)}{\|\nabla \rho_{J,\mathrm{alg}}^i\|^2} \overset{(2.4)}{=} \frac{(f, \rho_{J,\mathrm{alg}}^i) - (\nabla u_J^i, \nabla \rho_{J,\mathrm{alg}}^i)}{\|\nabla \rho_{J,\mathrm{alg}}^i\|^2} \overset{(4.1)}{=} \frac{\eta_{\mathrm{alg}}^i}{\|\nabla \rho_{J,\mathrm{alg}}^i\|}. \tag{4.5}$$

□

We are now ready to define our multilevel solver.

Definition 4.4 (Multilevel solver)**.** 1. *Initialize $u_J^0 \in V_0^1$ as the solution of $(\nabla u_J^0, \nabla v_0) = (f, v_0)$ $\forall v_0 \in V_0^1$.*

2. *Let $i \geq 0$ be the iteration counter and let $\rho_{J,\mathrm{alg}}^i$ be constructed from u_J^i following Definition 3.2. When $\rho_{J,\mathrm{alg}}^i = 0$, set $u_J^{i+1} := u_J^i$ and stop; then actually $u_J^{i+1} = u_J^i = u_J$. Otherwise, let*

$$u_J^{i+1} := u_J^i + \frac{(f, \rho_{J,\mathrm{alg}}^i) - (\nabla u_J^i, \nabla \rho_{J,\mathrm{alg}}^i)}{\|\nabla \rho_{J,\mathrm{alg}}^i\|^2} \rho_{J,\mathrm{alg}}^i. \tag{4.6}$$

Remark 4.5 (Multilevel solver). *Note that the solver of Definition 4.4 is not initialized randomly but via a coarse solve. The descent direction is the residual lifting $\rho^i_{J,\mathrm{alg}}$, constructed via a single V-cycle iteration with no presmoothing and one postsmoothing step, and the step-size is optimized via the line search (4.5). This minimalist and asymmetrical procedure will not be an issue in the forthcoming analysis.*

Remark 4.6 (Cost of one iteration). *On each iteration of the developed solver, there are costs which correspond to those of standard multigrid methods: coarse solve (here with the lowest polynomial degree) and interlevel transfer operations. The crucial difference is in the smoothing cost. While we prove below that our solver is p-robust and only mildly depends on h (since $J \sim |\log h|$), meaning the number of iterations will not degrade when p increases, the sizes of the local matrices used to solve the local problems (3.8) increase (in 2D approximately as p^2). This induces a significant computational, but perfectly parallelizable, cost for higher p. Other cheaper options may be developed to bypass the local problems, for example, in the spirit of Papež and Vohralík [40]. Recall, however, that there is only one smoothing per iteration in our approach.*

5 Main results

In this section, we present the main results concerning our a posteriori estimator η^i_{alg} of Definition 4.1 and our multilevel solver of Definition 4.4. We shall also see how these two main results are related.

For the estimator the following holds.

Theorem 5.1 (p-robust reliable and efficient bound on the algebraic error). *Let Assumption 2.1 hold, together with either Assumption 2.2 or 2.3. Let $u_J \in V_J^p$ be the (unknown) solution of (2.4) and let $u_J^i \in V_J^p$ be arbitrary, $i \geq 0$. Let η^i_{alg} be given by Definition 4.1. Then, in addition to $\|\nabla(u_J - u_J^i)\| \geq \eta^i_{\mathrm{alg}}$ of (4.2), there holds*

$$\eta^i_{\mathrm{alg}} \geq \beta \|\nabla(u_J - u_J^i)\|, \tag{5.1}$$

where $0 < \beta < 1$ only depends on the space dimension d, the mesh shape regularity parameter $\kappa_{\mathcal{T}}$, and the number of mesh levels J, as well as on the mesh refinement parameter C_{ref} and the quasi-uniformity parameter C_{qu} if Assumption 2.2 holds, or on the coarse mesh and the local quasi-uniformity parameters C^0_{qu} and $C^{\mathrm{loc}}_{\mathrm{qu}}$ if Assumption 2.3 holds. For all weights satisfying (3.5), there holds $\beta \geq J^{-5/2}\beta^*$ with β^* independent of the number of levels J. Better bounds hold for the weights of Remark 3.4; see Example 7.8 below for details.

The theorem allows us to write η^i_{alg} as a two-sided bound of the algebraic error (up to the generic constant β for the upper bound), meaning that the estimator is robustly efficient with respect to the polynomial degree p. We can also reinterpret this result as follows.

Remark 5.2 (Angle between the error and the descent direction). *Note that if we rewrite (5.1) by plugging in the Definition 4.1 of η^i_{alg}, when $u_J - u_J^i \neq 0$ and $\rho^i_{J,\mathrm{alg}} \neq 0$, we have*

$$\frac{\eta^i_{\mathrm{alg}}}{\|\nabla(u_J - u_J^i)\|} \overset{(4.1)}{=} \frac{(f, \rho^i_{J,\mathrm{alg}}) - (\nabla u_J^i, \nabla \rho^i_{J,\mathrm{alg}})}{\|\nabla(u_J - u_J^i)\| \|\nabla \rho^i_{J,\mathrm{alg}}\|} \overset{(2.4)}{=} \frac{(\nabla(u_J - u_J^i), \nabla \rho^i_{J,\mathrm{alg}})}{\|\nabla(u_J - u_J^i)\| \|\nabla \rho^i_{J,\mathrm{alg}}\|} \overset{(5.1)}{\geq} \beta > 0.$$

This can be compared to classical results in line search methods (see, e.g., *Nocedal and Wright* [35, Chapter 3.2]*) of boundedness away from zero of the cosine of the angle between the vector to be minimized (here $u_J - u_J^i$) and the descent direction (here the lifting $\rho^i_{J,\mathrm{alg}}$).*

For the solver, in turn, we have the following.

Theorem 5.3 (p-robust error contraction of the multilevel solver). *Let Assumption 2.1 hold, together with either Assumption 2.2 or 2.3. Let $u_J \in V_J^p$ be the (unknown) solution of (2.4) 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 4.4 by (4.6). Then there holds*

$$\|\nabla(u_J - u_J^{i+1})\| \leq \alpha \|\nabla(u_J - u_J^i)\|, \tag{5.2}$$

where $0 < \alpha < 1$ is given by $\alpha = \sqrt{1 - \beta^2}$ with β the constant from (5.1).

In the above theorem, α is a bound on the algebraic error contraction factor at each step i. Looking at the dependencies of α, we see that the solver of Definition 4.4 contracts the algebraic error at each iteration step in a robust way with respect to the polynomial degree p.

Theorems 5.1 and 5.3 are connected as follows.

Corollary 5.4 (Equivalence of the p-robust estimator efficiency and p-robust solver contraction)**.** *Let the assumptions of Theorems 5.1 and 5.3 be satisfied. Then (5.1) holds if and only if (5.2) holds, and* $\beta = \sqrt{1-\alpha^2}$.

Proof. Let $u_J \in V_J^p$ be the solution of (2.4), let $u_J^i \in V_J^p$ be arbitrary, and let $u_J^{i+1} \in V_J^p$ be constructed from u_J^i by our multilevel solver of Definition 4.4. First, we write the relation between the algebraic errors associated to u_J^{i+1} and u_J^i.

Case $\rho^i_{J,\mathrm{alg}} \neq 0$. Using (4.4) and (4.5), we see

$$\|\nabla(u_J - u_J^{i+1})\|^2 = \|\nabla(u_J - u_J^i)\|^2 - \left(\eta^i_{\mathrm{alg}}\right)^2. \tag{5.3}$$

Case $\rho^i_{J,\mathrm{alg}} = 0$. By Definitions 4.4 and 4.1, we have $u_J^{i+1} = u_J^i$ and $\eta^i_{\mathrm{alg}} = 0$. In particular, this means that $\|\nabla(u_J - u_J^{i+1})\| = \|\nabla(u_J - u_J^i)\|$, so that (5.3) still holds.

The above observations allow us to write, in any case, starting from (5.2) with $0 < \alpha < 1$,

$$\begin{aligned}\|\nabla(u_J - u_J^{i+1})\|^2 \leq \alpha^2\|\nabla(u_J - u_J^i)\|^2 \overset{(5.3)}{\Leftrightarrow} \|\nabla(u_J - u_J^i)\|^2 - \left(\eta^i_{\mathrm{alg}}\right)^2 \leq \alpha^2\|\nabla(u_J - u_J^i)\|^2 \\ \Leftrightarrow \ \|\nabla(u_J - u_J^i)\|^2(1-\alpha^2) \leq \left(\eta^i_{\mathrm{alg}}\right)^2,\end{aligned}$$

which is (5.1) with $\beta^2 = 1 - \alpha^2$. □

In view of Corollary 5.4, we will prove in section 7 below only Theorem 5.1.

Importantly, the following also holds.

Corollary 5.5 (Equivalence of vanishing algebraic lifting with the solver reaching the solution)**.** *Let the assumptions of Theorems 5.1 and 5.3 be satisfied. Then* $\rho^i_{J,\mathrm{alg}} = 0$ *if and only if* $u_J^{i+1} = u_J^i = u_J$.

Finally, by the proofs in section 7, the algebraic error is also equivalent to a localized a posteriori error estimate.

Corollary 5.6 (p-robust localized reliable and efficient a posteriori estimate on the algebraic error)**.** *Let the assumptions of Theorem 5.1 be satisfied. Let* $\rho^i_{J,\mathrm{alg}}$ *be the algebraic residual lifting constructed in Definition 3.2. Then*

$$\|\nabla(u_J - u_J^i)\|^2 \leq C_1^2\left(\|\nabla\rho_0^i\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}\right) \leq C_2^2\|\nabla(u_J - u_J^i)\|^2, \tag{5.4}$$

where $C_2 = \frac{1}{\beta}$ *and* C_1 *is identified in section 7.6.*

Equivalence (5.4) gives us an idea where the algebraic error is situated levelwise and patchwise. This information can be exploited to tackle problematic areas adaptively, which is the subject of forthcoming works.

6 Numerical experiments

In this section we report some numerical illustrations of the theoretical results of section 5. In particular, we focus on the p-robustness. In the following tests, we consider the model problem (2.1) with three different choices of the domain $\Omega \subset \mathbb{R}^2$ and of the exact solution u:

$$\text{Sine:} \quad u(x,y) := \sin(2\pi x)\sin(2\pi y), \ \Omega := (-1,1)^2. \tag{6.1}$$

$$\text{Peak:} \quad u(x,y) := x(x-1)y(y-1)\mathrm{e}^{-100((x-0.5)^2-(y-0.117)^2)}, \ \Omega := (0,1)^2. \tag{6.2}$$

$$\text{L-shape:} \quad u(r,\theta) := r^{2/3}\sin(2\theta/3), \ \Omega := (-1,1)^2 \setminus ([0,1]\times[-1,0]). \tag{6.3}$$

For the L-shape problem (6.3), we impose an inhomogeneous Dirichlet boundary condition corresponding to the exact solution, which is expressed here in polar coordinates. For each of the test cases, we start with an initial Delaunay triangulation of Ω. Then we consider J uniform refinements where all triangles are decomposed into four congruent subtriangles. Implementation-wise, we opt for Lagrange basis functions with nonuniformly distributed nodes because of their better behavior with respect to high-order methods; see Warburton [47]. Recall that this choice has no influence on the theoretical results of section 5 as well as presented numerical results (in exact arithmetic). Though it is not the focus of this work, we also remark that our solver can be implemented in a matrix-free way and can also be parallelized.

The contraction factor of the solver of Definition 4.4 on each step i is given by $\|\nabla(u_J - u_J^{i+1})\|/\|\nabla(u_J - u_J^i)\|$, and, as stated in Corollary 5.4, it reveals the efficiency of the a posteriori estimator η^i_{alg} of Definition 4.1. Keeping this in mind, we only focus on the solver and the contraction factor. We will follow a common choice for the stopping criterion, with the notation of section 2.3:

$$\frac{\|\mathrm{F}_J - \mathbb{A}_J \mathrm{U}_J^{i_\mathrm{s}}\|}{\|\mathrm{F}_J\|} \leq 10^{-5} \frac{\|\mathrm{F}_J - \mathbb{A}_J \mathrm{U}_J^{0}\|}{\|\mathrm{F}_J\|}. \tag{6.4}$$

We also introduce the average error contraction factor

$$\bar{\alpha} := \frac{1}{i_\mathrm{s}} \sum_{i=0}^{i_\mathrm{s}-1} \frac{\|\nabla(u_J - u_J^{i+1})\|}{\|\nabla(u_J - u_J^{i})\|}. \tag{6.5}$$

We expect a p-robust solver to converge in a similar number of iterations and have similar error contraction factors at all iterations for different polynomial degrees p. The tests below cover different numbers of mesh levels $J = 3, 4, 5$, polynomial degrees $p = 1, 3, 6, 9$, and the "small" as well as the "large" patches as in Figure 2.

6.1 Performance of the damped additive Schwarz (dAS) construction of the solver

A crucial component in the definition of our a posteriori estimator and multilevel solver is the construction of the residual lifting $\rho^i_{J,\mathrm{alg}}$ of Definition 3.2, where we have used damped additive Schwarz (dAS) to cope with overlapping:

$$\text{dAS:} \quad \rho^i_{J,\mathrm{alg}} := \sum_{j=0}^{J} \rho_j^i \quad \text{and} \quad \rho_j^i := \frac{1}{w_1} \sum_{\mathbf{a} \in \mathcal{V}_{j-s}} \rho^i_{j,\mathbf{a}}, \ 1 \leq j \leq J, \tag{6.6}$$

$$(\nabla \rho^i_{j,\mathbf{a}}, \nabla v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}} = (f, v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}} - (\nabla u_J^i, \nabla v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}} - \frac{1}{w_2} \sum_{k=0}^{j-1} (\nabla \rho_k^i, \nabla v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}}.$$

For the three test cases we consider three different choices of the damping weights which satisfy condition (3.5) (see Remark 3.4):

$$\begin{aligned} &\text{for problem (6.1):} && w_1 = J(d+1) \text{ and } w_2 = 1; \\ &\text{for problem (6.2):} && w_1 = 4\sqrt{J} \text{ and } w_2 = \infty; \\ &\text{for problem (6.3):} && w_1 = d+1 \text{ and } w_2 = J. \end{aligned}$$

Recall that the choice $w_2 = \infty$ means that the construction of the lifting $\rho^i_{J,\mathrm{alg}}$ can be implemented completely in parallel, levelwise as well as patchwise.

The results are presented in Figures 3–5 and in Table 1. They confirm the expected complete independence of the polynomial degree p for our multilevel solver which uses the construction dAS (6.6) of the lifting. Actually, we observe better contraction factors for higher polynomial degrees.

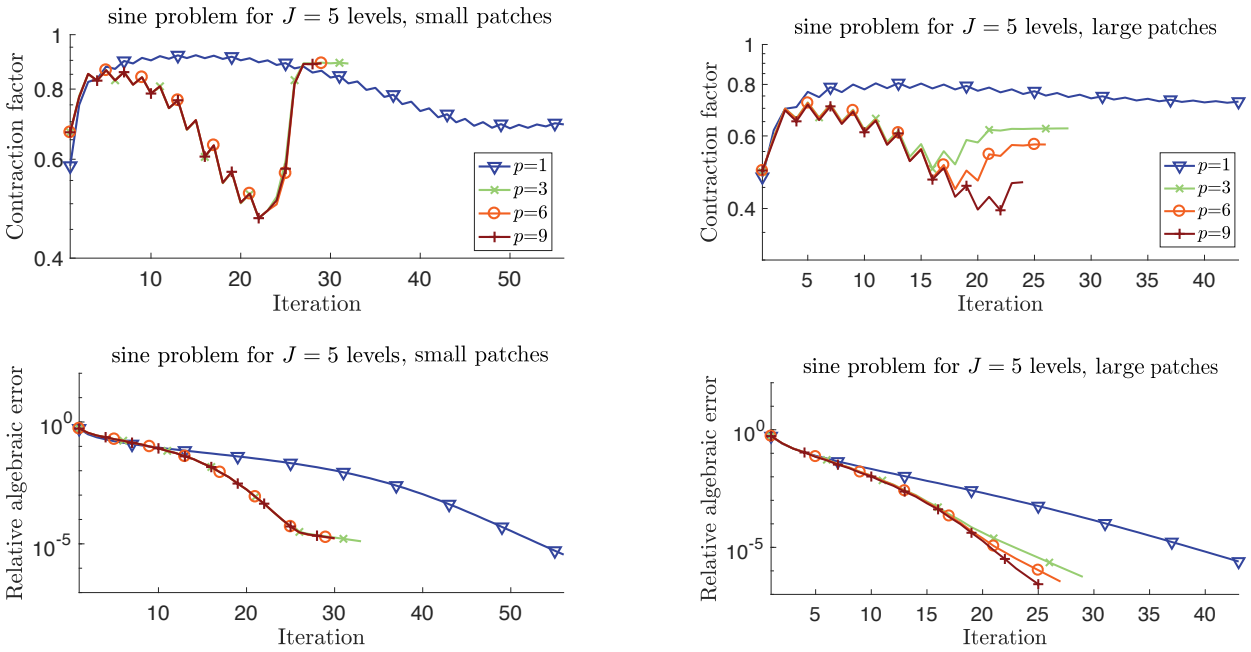


Figure 3: Sine problem (6.1), $w_1 = J(d+1)$, $w_2 = 1$: results of the solver (4.6) for $p' = p$ in (2.10a), "small" (left) and "large" (right) patches, and stopping criterion (6.4). Top: error contraction factors $\|\nabla(u_J - u_J^{i+1})\|/\|\nabla(u_J - u_J^i)\|$. Bottom: relative algebraic error $\|\nabla(u_J - u_J^i)\|/\|\nabla u_J\|$.

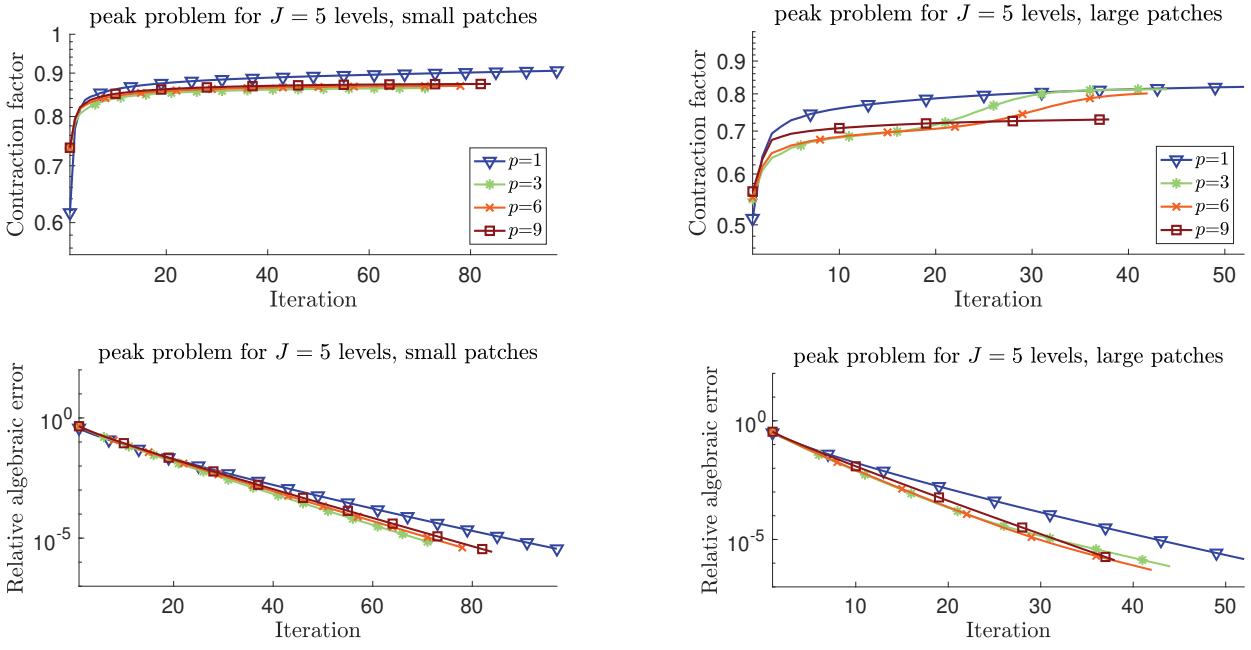


Figure 4: Peak problem (6.2), $w_1 = 4\sqrt{J}$, $w_2 = \infty$: results of the solver (4.6) for $p' = p$ in (2.10a), "small" (left) and "large" (right) patches, and stopping criterion (6.4). Top: error contraction factors $\|\nabla(u_J - u_J^{i+1})\|/\|\nabla(u_J - u_J^i)\|$. Bottom: relative algebraic error $\|\nabla(u_J - u_J^i)\|/\|\nabla u_J\|$.

An inferior quality of the contraction factors for the case of $p = 1$ and the use of damping factors $w_1 = J(d+1)$ and $w_2 = 1$ appears. This is in line with some precedents in literature, where numerically p-robust solvers also perform worse for order 1 approximations; we mention, for example, Griebel, Oswald, and

			Sine problem (6.1) $w_1 = J(d+1)$, $w_2 = 1$				Peak problem (6.2) $w_1 = 4\sqrt{J}$, $w_2 = \infty$				L-shape problem (6.3) $w_1 = d+1$, $w_2 = J$			
			"small"		"large"		"small"		"large"		"small"		"large"	
J	p	DoF	i_s	$\bar{\alpha}$	i_s	$\bar{\alpha}$	i_s	$\bar{\alpha}$	i_s	$\bar{\alpha}$	i_s	$\bar{\alpha}$	i_s	$\bar{\alpha}$
3	1	5e^3	48	0.79	34	0.70	74	0.85	43	0.75	38	0.75	20	0.56
	3	4e^4	23	0.63	24	0.59	60	0.83	36	0.70	28	0.68	18	0.53
	6	2e^5	23	0.63	22	0.55	58	0.82	34	0.68	27	0.69	16	0.49
	9	4e^5	23	0.63	19	0.50	58	0.82	31	0.65	25	0.69	14	0.46
4	1	2e^4	52	0.80	40	0.74	87	0.87	48	0.77	39	0.76	23	0.60
	3	2e^5	27	0.68	26	0.60	66	0.84	41	0.72	28	0.70	23	0.60
	6	1e^6	26	0.66	24	0.57	68	0.84	38	0.70	29	0.72	20	0.58
	9	2e^5	26	0.67	21	0.53	70	0.84	33	0.67	28	0.72	18	0.55
5	1	1e^5	56	0.81	43	0.75	97	0.88	52	0.78	40	0.76	25	0.63
	3	1e^6	32	0.73	28	0.61	72	0.85	44	0.74	30	0.72	27	0.65
	6	3e^6	29	0.71	26	0.58	78	0.86	42	0.72	31	0.74	25	0.63
	9	6e^6	29	0.71	24	0.56	84	0.86	38	0.71	30	0.74	21	0.59

Table 1: dAS construction (6.6): problems (6.1)–(6.3), $p' = p$ in (2.10a), "small" and "large" patches. i_s: the number of iterations needed to reach the stopping criterion (6.4). $\bar{\alpha}$: average error contraction factor given by (6.5).

Schweitzer [23, Table 1] and Kronbichler and Wall [30, Table 1]. Recall that we consider no presmoothing and only one postsmoothing step; an important drop in the number of iterations appears if more smoothing steps are employed, which will be explored below. Another observation is that the number of iterations depends on the number of mesh levels J, in accordance with the theoretical result of section 7, even though rather mildly.

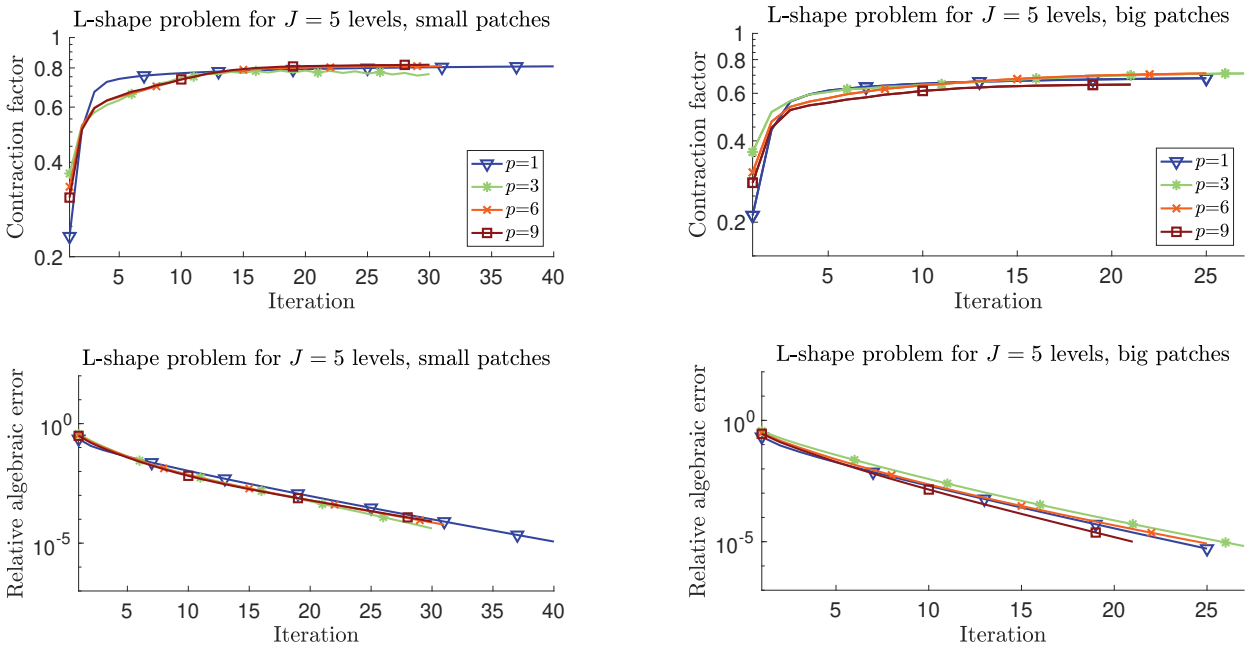


Figure 5: L-shape problem (6.3), $w_1 = d+1$, $w_2 = J$: results of the solver (4.6) for $p' = p$ in (2.10a), "small" (left) and "large" (right) patches, and stopping criterion (6.4). Top: error contraction factors $\|\nabla(u_J - u_J^{i+1})\|/\|\nabla(u_J - u_J^i)\|$. Bottom: relative algebraic error $\|\nabla(u_J - u_J^i)\|/\|\nabla u_J\|$.

The behavior of the contraction factor in each iteration in Figures 3–5 appears quite different. This seems to be related partly to the smoothness of the problem and partly to choice of the damping weights. We explore this in more detail in Figure 6 by using different choices of the weights and number of postsmoothing steps ν. In particular the degradation of the contraction factors observed in Figure 3 disappears when employing more smoothing steps.

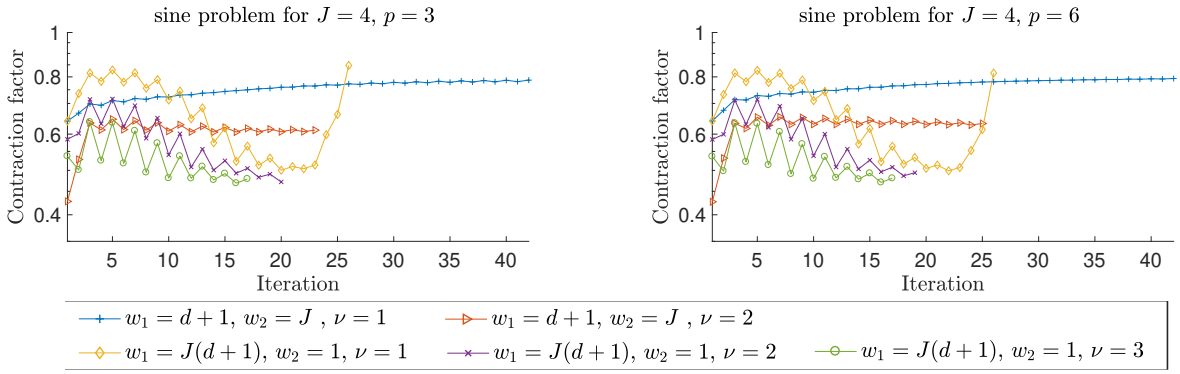


Figure 6: Sine problem (6.1), "small" patches, $p' = p$: study of the contraction factor behavior with respect to the number of postsmoothing steps and damping weights for the solver (4.6).

6.2 Performance of the weighted restrictive additive Schwarz (wRAS) construction of the solver

As observed in the literature, replacing the damping with parameter w_1 in (3.7) by hat function weighting via a restrictive additive Schwarz often performs better; cf. Cai and Sarkis [15], Efstathiou and Gander [19], or Loisel, Nabben, and Szyld [31]. Thus, in addition to the dAS construction (6.6), we now numerically also explore the weighted restricted additive Schwarz (wRAS) construction of the lifting $\rho^i_{J,\mathrm{alg}}$:

$$\text{wRAS:}\quad \rho^i_{J,\mathrm{alg}} := \sum_{j=0}^{J} \rho^i_j \quad \text{and} \quad \rho^i_j := \sum_{\mathbf{a}\in\mathcal{V}_{j-s}} \mathcal{I}^q_j(\psi^{\mathbf{a}}_{j-s}\rho^i_{j,\mathbf{a}}),\ 1 \le j \le J, \tag{6.7}$$

$$(\nabla\rho^i_{j,\mathbf{a}}, \nabla v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}} = (f, v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}} - (\nabla u^i_J, \nabla v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}} - \sum_{k=0}^{j-1} (\nabla\rho^i_k, \nabla v_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}}$$

with $q = p'$ except for $j = J$ where $q = p$, we denote by $\mathcal{I}^q_j$ the $\mathbb{P}_q$ Lagrange interpolation operator on the mesh level j, i.e., $\mathcal{I}^q_j : C^0(\overline{\Omega}) \to V^q_j$, $\mathcal{I}^q_j(v)$ preserves the values of v in the nodes corresponding to the Lagrange degrees of freedom. No damping weights are to be chosen here.

We summarize the results obtained for each of the problems (6.1)–(6.3) in Table 2. In addition to one postsmoothing step, $\nu = 1$, we also present the results for $\nu = 3$ postsmoothing steps. In both cases, no presmoothing has been employed. In the last two columns for each problem, we present a comparison of our solver of Definition 4.4 employing (6.7) with two standard smoothers for multigrid, namely the Jacobi (J) and the Gauss–Seidel (GS) ones. Here, we employ no presmoothing step, one postsmoothing step, and a coarse solve with polynomials of order 1 as in (3.3) to compare with our approach.

The results using the wRAS (6.7) construction of the lifting indicate an improvement in the error contraction factors with respect to dAS (6.6) of section 6.1 and, moreover, present a complete numerical independence of the number of levels J. Furthermore, the iteration numbers drop by at least half when three postsmoothing steps are employed. In contrast to these results, we see that the multigrid with standard smoothers degrades violently with respect to the polynomial degree p. Note in this respect that for $p = 1$, the only difference between wRAS of (6.7) with small patches and $\nu = 1$ and standard Jacobi lies in the optimally chosen step-size of Lemma 4.3. This gives a spectacular gain in the number of iterations, and makes the method convergent even when the standard Jacobi fails.

		Sine problem (6.1)						Peak problem (6.2)						L-shape problem (6.3)					
		wRAS				MG		wRAS				MG		wRAS				MG	
		"small"		"large"		$\nu=1$		"small"		"large"		$\nu=1$		"small"		"large"		$\nu=1$	
J	p	$\nu=1$	$\nu=3$	$\nu=1$	$\nu=3$	J	GS	$\nu=1$	$\nu=3$	$\nu=1$	$\nu=3$	J	GS	$\nu=1$	$\nu=3$	$\nu=1$	$\nu=3$	J	GS
3	1	21	10	9	4	-	10	19	9	9	4	68	8	17	9	8	4	44	9
	3	15	5	6	3	-	81	15	6	6	3	-	70	12	4	5	3	-	49
	6	13	5	6	3	-	470	14	6	6	3	-	462	10	4	5	2	-	228
	9	13	5	6	3	-	+600	14	6	5	3	-	+600	10	4	5	2	-	586
4	1	23	11	9	4	-	11	20	9	9	4	-	10	18	9	8	4	-	9
	3	15	5	6	3	-	81	15	6	5	3	-	79	12	4	5	3	-	42
	6	13	5	6	3	-	468	14	6	5	3	-	460	10	4	5	2	-	186
	9	13	5	5	3	-	+600	14	6	5	3	-	+600	9	4	5	2	-	454
5	1	22	11	9	4	-	11	20	11	9	4	-	11	17	9	8	4	-	8
	3	15	5	6	3	-	81	15	6	5	3	-	80	12	4	5	3	-	35
	6	13	5	6	3	-	470	14	6	5	3	-	461	9	4	5	2	-	147
	9	13	5	5	3	-	+600	13	6	5	3	-	+600	8	3	4	2	-	333

Table 2: Number of iterations needed to reach the stopping criterion (6.4): wRAS construction (6.7), problems (6.1)–(6.3), $p' = p$ in (2.10a), "small" and "large" patches, ν postsmoothing steps, and standard multigrid method with piecewise affine coarse solve (3.3), initialized by the coarse grid solution, no presmoothing, one postsmoothing step, and Jacobi (J) and Gauss–Seidel (GS) smoothers.

6.3 Comparison with other multilevel solvers

Some recent comparisons of state-of-the-art solvers for Poisson problems with multigrid methods in the high-order setting include Gholami et al. [22], Sundar, Stadler, and Biros [46], and Kronbichler and Wall [30]. In [46], it was in particular reported that none of the methods considered behaves fully independently of the polynomial degree. In this subsection, we compare our developments with four well-established methods. We focus on the number of iterations, but we also indicate CPU times of our vectorized MATLAB implementation[1], trusting the reader to understand the trickiness inherent in such implementation- and machine-dependent measurements. The timings below involve the solution time only; i.e., they do not include the assembly time of the matrices. The methods we consider for the comparison are as follows:

wRAS, $\nu=1$ (∼MG(0,1)-bJ): Definition 4.4, small patches, $p'=p$ in (2.10a) (to illustrate the associated space hierarchy, we write "$1, p \to p$"), wRAS construction (6.7).

wRAS, $\nu=3$ (∼MG(0,3)-bJ): Definition 4.4, small patches, $p' = p$ in (2.10a) ("$1, p \to p$"), wRAS construction (6.7), three postsmoothing steps employed.

wRAS, $\nu=1$ (∼MG(0,1)-bJ): Definition 4.4, small patches, $p' = 1$ in (2.10a) ("$1 \to 1, p$"), wRAS construction (6.7).

wRAS, $\nu=3$ (∼MG(0,3)-bJ): Definition 4.4, small patches, $p' = 1$ in (2.10a) ("$1 \to 1, p$"), wRAS construction (6.7), three postsmoothing steps employed.

PCG(MG(3,3)-bJ): Preconditioned conjugate gradient solver; the preconditioner is multigrid V-cycle(3,3) with weighted restrictive additive Schwarz (block Jacobi) smoother associated to small patches; the space hierarchy relies on order p discretization, including the coarsest space ("$p \to p$"); the iterations start with the zero vector. This choice of solver is motivated by Antonietti and Pennesi [2], adapted to the conforming finite elements setting.

[1]The codes were prepared to benefit as much as possible from MATLAB's fast operations on matrices and vectors. 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.

MG(1,1)-PCG(iChol): Multigrid solver V-cycle(1,1); the smoother is PCG with incomplete zero level fill-in Cholesky preconditioner; the space hierarchy is of increasing order: from order 1 for the coarsest level to order p for the finest level ("$1 \nearrow p$"); the iterations start with the zero vector. This choice of solver is motivated by Botti, Colombo, and Bassi [10], adapted for a symmetric setting.

MG(0,1)-bGS: Multigrid solver V-cycle(0,1); the smoother is block Gauss–Seidel associated to small patches; the space hierarchy consists of order 1 for all levels except the finest level, which is of order p ("$1 \to 1, p$"), i.e., as in (2.10a) with p'=1; the iterations start with the zero vector. This choice of the solver is motivated by NGSolve [44]; however, the multigrid is used here as a solver instead of a preconditioner.

MG(0,3)-bGS: Multigrid solver analogous to MG(0,1)-bGS where now three postsmoothing steps are employed.

MG(3,3)-GS: Multigrid solver V-cycle(3,3); the smoother is standard Gauss–Seidel; the space hierarchy is of increasing order: from order 1 for coarse level to order p for the finest level ("$1 \nearrow p$"); the iterations start with the zero vector.

		wRAS $\nu=1$ $1, p \to p$		wRAS $\nu=3$ $1, p \to p$		wRAS $\nu=1$ $1 \to 1, p$		wRAS $\nu=3$ $1 \to 1, p$		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(0,3)-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	i_s	time
3	1	17	0.0 s	9	0.0 s	17	0.0 s	9	0.0 s	7	0.0 s	4	0.0 s	9	0.0 s	4	0.0 s	3	0.0 s
	3	12	0.2 s	4	0.1 s	18	0.2 s	6	0.1 s	3	0.2 s	14	0.6 s	8	0.6 s	4	0.8 s	4	0.1 s
	6	10	1.8 s	4	1.7 s	15	1.9 s	6	2.0 s	2	2.0 s	21	8.6 s	7	1.8 s	4	2.7 s	9	1.5 s
	9	10	9.9 s	4	10.2 s	14	9.7 s	6	11.2 s	2	10.1 s	63	1.2m	6	6.9 s	3	8.7 s	9	5.3 s
4	1	18	0.0 s	9	0.0 s	18	0.0 s	9	0.0 s	8	0.1 s	7	0.1 s	9	0.0 s	4	0.0 s	3	0.0 s
	3	12	0.8 s	4	0.6 s	18	0.8 s	6	0.6 s	3	0.7 s	29	5.6 s	8	2.4 s	4	3.4 s	4	0.3 s
	6	10	7.3 s	4	7.4 s	15	7.8 s	6	7.9 s	3	10.9 s	49	1.2 m	7	8.6 s	3	9.4 s	5	3.5 s
	9	9	34.7 s	4	40.7 s	13	37.2 s	5	37.4 s	2	39.3 s	167	12.5m	6	28.3 s	3	36.7 s	8	20.7 s
5	1	17	0.1 s	9	0.1 s	17	0.1 s	9	0.1 s	8	0.2 s	19	1.2 s	8	0.1 s	4	0.1 s	3	0.1 s
	3	12	3.2 s	4	2.3 s	17	3.4 s	6	2.6 s	3	3.1 s	77	57.7 s	8	10.7 s	4	15.7 s	4	1.5 s
	6	9	27.6 s	4	30.3 s	14	32.0 s	6	33.8 s	3	45.6 s	129	11.6m	7	30.8 s	3	33.7 s	4	12.8 s
	9	8	2.3m	3	2.1m	12	2.3m	5	2.5m	2	3.1m	+200	+1.0 h	6	2.2m	3	2.7m	8	1.3m

Table 3: Comparison of various multilevel solvers (described in section 6.3) for the L-shape case (6.3), i_s is the number of iterations to reach the stopping criterion (6.4).

As one can see from Table 3, the presented methods split into two groups: numerically p-robust (wRAS, PCG(MG-bJ), MG-bGS) and not (MG-PCG(iChol), MG-GS). Note that the choice of three pre- and three postsmoothing steps makes every iteration of the methods PCG(MG(3,3)-bJ) and MG(3,3)-GS considerably more expensive than those of the methods wRAS and MG-bGS with $\nu = 1$, where the minimalist (0,1) choice is sufficient. The variants wRAS and MG-bGS with $\nu = 3$ are also cheaper. In addition, in PCG(MG(3,3)-bJ), the coarse grid correction is more expensive as it uses order p approximations. The inversion of the Jacobi blocks in PCG(MG(3,3)-bJ) on the finest level J, corresponds to solving the patch problems of order p as in (3.8), so that its cost is the same as for the local problems of wRAS. As for MG(1,1)-PCG(iChol), we find the method to be quite satisfactory for lower-order approximations and small J, but as soon as p and J increase, the number of iterations degrades considerably. In contrast to wRAS, MG-bGS is a multiplicative Schwarz method and is thus less suitable for parallelization. Finally, the classical MG(3,3)-GS is a combination of h- and p-multigrid and gives the best timings in our experiments. The numbers of pre- and postsmoothing steps, however, remain parameters, and their tuning might not be straightforward in order to get an efficient and numerically robust multigrid solver in general (cf. the poor results of the very similar—up to the different number of pre- and postsmoothing steps and a stronger hierarchy—MG(0,1)-GS version in Table 2). The Gauss–Seidel smoother used therein again makes the method harder to parallelize.

7 Proofs of the main results

As shown in Corollary 5.4, the results of Theorems 5.1 and 5.3 are equivalent. Therefore it suffices to prove the first one. Our approach to proving Theorem 5.1 consists in studying the uncomputable exact residual lifting $\tilde{\rho}^i_{J,\mathrm{alg}}$ given by (3.1) and its approximation $\rho^i_{J,\mathrm{alg}}$ given by Definition 3.2. In particular, we will estimate p-robustly the quantities $\|\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\|$, $\|\nabla\rho^i_{J,\mathrm{alg}}\|$, and $(f,\rho^i_{J,\mathrm{alg}})-(\nabla u^i_J,\nabla\rho^i_{J,\mathrm{alg}})$ by local contributions $\rho^i_{j,\mathbf{a}}$ of (3.8) used to construct $\rho^i_{J,\mathrm{alg}}$, and we show that

$$\begin{aligned}\frac{(f,\rho^i_{J,\mathrm{alg}})-(\nabla u^i_J,\nabla\rho^i_{J,\mathrm{alg}})}{\|\nabla\rho^i_{J,\mathrm{alg}}\|}\geq\beta\|\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\| &\qquad \text{when } \rho^i_{J,\mathrm{alg}}\neq 0,\\ \tilde{\rho}^i_{J,\mathrm{alg}}=0 &\qquad \text{when } \rho^i_{J,\mathrm{alg}}=0,\end{aligned}$$

which also establishes Corollary 5.5.

7.1 Upper bound on $\|\nabla\rho^i_{J,\mathrm{alg}}\|$

We present here properties of the constructed residual lifting $\rho^i_{J,\mathrm{alg}}$ and its levelwise components ρ^i_j, where $1\leq j\leq J$.

Lemma 7.1 (Estimate on $\|\nabla\rho^i_{J,\mathrm{alg}}\|$ and $\|\nabla\rho^i_j\|$ by patchwise contributions)**.** *Let $\rho^i_{J,\mathrm{alg}}$ and ρ^i_j for $j\in\{1,\dots,J\}$ be given by Definition 3.2. Then*

$$\|\nabla\rho^i_j\|^2\leq\frac{d+1}{w_1^2}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}},\tag{7.1}$$

$$\|\nabla\rho^i_{J,\mathrm{alg}}\|^2\leq C^2_{\max}(w_1)\Big(\|\nabla\rho^i_0\|^2+\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}\Big),\tag{7.2}$$

where

$$2\leq C^2_{\max}(w_1):=2\max\left(1,\frac{J(d+1)}{w_1^2}\right)\leq 2J(d+1).\tag{7.3}$$

Proof. Definition 3.2 and inequality $|\sum_{k=1}^{d+1}a_k|^2\leq(d+1)\sum_{k=1}^{d+1}|a_k|^2$ lead to

$$\begin{aligned}\|\nabla\rho^i_j\|^2=\sum_{K\in\mathcal{T}_{j-s}}\|\nabla\rho^i_j\|^2_K&=\sum_{K\in\mathcal{T}_{j-s}}\Big\|\frac{1}{w_1}\sum_{\mathbf{a}\in\mathcal{V}_K}\nabla\rho^i_{j,\mathbf{a}}\Big\|^2_K\\&\leq\frac{d+1}{w_1^2}\sum_{K\in\mathcal{T}_{j-s}}\sum_{\mathbf{a}\in\mathcal{V}_K}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_K=\frac{d+1}{w_1^2}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}.\end{aligned}$$

Note that this allows us to write

$$\Big\|\sum_{j=1}^{J}\nabla\rho^i_j\Big\|^2\leq J\sum_{j=1}^{J}\|\nabla\rho^i_j\|^2\leq\frac{J(d+1)}{w_1^2}\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}.\tag{7.4}$$

This property together with some simple manipulations, gives the second estimate:

$$\begin{aligned}\|\nabla\rho^i_{J,\mathrm{alg}}\|^2&\leq 2\|\nabla\rho^i_0\|^2+2\Big\|\sum_{j=1}^{J}\nabla\rho^i_j\Big\|^2\overset{(7.4)}{\leq}2\|\nabla\rho^i_0\|^2+\frac{2J(d+1)}{w_1^2}\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}\\&\leq 2\max\left(1,\frac{J(d+1)}{w_1^2}\right)\Big(\|\nabla\rho^i_0\|^2+\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}\Big).\end{aligned}$$

The bounds (7.3) on $C^2_{\max}(w_1)$ are easily obtained by using $w_1\geq 1$ requested in Definition 3.2. □

7.2 Lower bound on $(f, \rho^i_{J,\mathrm{alg}}) - (\nabla u^i_J, \nabla \rho^i_{J,\mathrm{alg}})$

While studying the term $(f, \rho^i_{J,\mathrm{alg}}) - (\nabla u^i_J, \nabla \rho^i_{J,\mathrm{alg}})$, the interaction of different level contributions ρ^i_j of the lifting $\rho^i_{J,\mathrm{alg}}$ arises naturally. In order to estimate these terms, the damping parameters w_1, w_2 used in the construction (3.7) of our lifting prove to be essential.

Lemma 7.2 (Estimate on $(f, \rho^i_{J,\mathrm{alg}}) - (\nabla u^i_J, \nabla \rho^i_{J,\mathrm{alg}})$ from below by patchwise contributions)**.** *Let $\rho^i_{J,\mathrm{alg}}$ be given by Definition 3.2. Then*

$$(f, \rho^i_{J,\mathrm{alg}}) - (\nabla u^i_J, \nabla \rho^i_{J,\mathrm{alg}}) \geq C^2_{\min}(w_1, w_2) \left(\|\nabla \rho^i_0\|^2 + \sum_{j=1}^{J} \sum_{\mathbf{a} \in \mathcal{V}_{j-s}} \|\nabla \rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}} \right), \tag{7.5}$$

where

$$\frac{1}{6J(d+1)} \leq C^2_{\min}(w_1, w_2) := \min\left(\frac{1}{4}\,,\ \frac{1}{w_1} - \frac{J(1+\frac{2}{3w_2})(d+1)}{2w_2 w_1^2} \right) \leq \frac{1}{4}. \tag{7.6}$$

Proof. We begin by using the construction of $\rho^i_{J,\mathrm{alg}}$ given in Definition 3.2 to write

$$\begin{aligned}
(f, \rho^i_{J,\mathrm{alg}}) - (\nabla u^i_J, \nabla \rho^i_{J,\mathrm{alg}}) &= (f, \rho^i_0) - (\nabla u^i_J, \nabla \rho^i_0) + \sum_{j=1}^{J} \Big((f, \rho^i_j) - (\nabla u^i_J, \nabla \rho^i_j) \Big) \\
&\overset{(3.3)}{\underset{(3.7)}{=}} \|\nabla \rho^i_0\|^2 + \frac{1}{w_1} \sum_{j=1}^{J} \sum_{\mathbf{a} \in \mathcal{V}_{j-s}} \Big((f, \rho^i_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}} - (\nabla u^i_J, \nabla \rho^i_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}} \Big) \\
&\overset{(3.8)}{=} \|\nabla \rho^i_0\|^2 + \frac{1}{w_1} \sum_{j=1}^{J} \sum_{\mathbf{a} \in \mathcal{V}_{j-s}} \left(\|\nabla \rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}} + \frac{1}{w_2} \sum_{k=0}^{j-1} (\nabla \rho^i_k, \nabla \rho^i_{j,\mathbf{a}})_{\omega^{\mathbf{a}}_{j,s}} \right) \\
&\overset{(3.7)}{=} \|\nabla \rho^i_0\|^2 + \frac{1}{w_1} \sum_{j=1}^{J} \sum_{\mathbf{a} \in \mathcal{V}_{j-s}} \|\nabla \rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}} + \frac{1}{w_2} \sum_{j=1}^{J} \sum_{k=0}^{j-1} (\nabla \rho^i_k, \nabla \rho^i_j).
\end{aligned}$$

The first two terms above are of the right form to prove the result, but one needs to be a bit more careful with the third one. We estimate it using Young's inequality and the sum interchange $\sum_{j=2}^{J} \sum_{k=1}^{j-1} = \sum_{k=1}^{J-1} \sum_{j=k+1}^{J}$; Young's parameter μ is picked later to control the dependence on J of the final estimate. We have

$$\begin{aligned}
&\frac{1}{w_2} \sum_{j=1}^{J} \sum_{k=0}^{j-1} (\nabla \rho^i_k, \nabla \rho^i_j) = \frac{1}{w_2} \left(\sum_{j=2}^{J} \sum_{k=1}^{j-1} (\nabla \rho^i_k, \nabla \rho^i_j) + \sum_{j=1}^{J} (\nabla \rho^i_0, \nabla \rho^i_j) \right) \\
&\geq \frac{1}{w_2} \sum_{j=2}^{J} \sum_{k=1}^{j-1} \left(-\frac{1}{2} \|\nabla \rho^i_k\|^2 - \frac{1}{2} \|\nabla \rho^i_j\|^2 \right) + \frac{1}{w_2} \sum_{j=1}^{J} \left(-\frac{1}{2\mu} \|\nabla \rho^i_0\|^2 - \frac{\mu}{2} \|\nabla \rho^i_j\|^2 \right) \\
&= -\frac{1}{2w_2} \sum_{j=1}^{J} (J-j) \|\nabla \rho^i_j\|^2 - \frac{1}{2w_2} \sum_{j=1}^{J} (j-1) \|\nabla \rho^i_j\|^2 - \frac{J}{2\mu w_2} \|\nabla \rho^i_0\|^2 - \frac{\mu}{2w_2} \sum_{j=1}^{J} \|\nabla \rho^i_j\|^2,
\end{aligned}$$

where we added the terms in the sum corresponding to $k = J$ and $j = 1$ since they are zero, and then renamed the summation index when there is no confusion. Picking Young's inequality parameter $\mu = \frac{2J}{3w_2}$, a few more manipulations on the right-hand side give us

$$\begin{aligned}
\frac{1}{w_2} \sum_{j=1}^{J} \sum_{k=0}^{j-1} (\nabla \rho^i_k, \nabla \rho^i_j) &\geq -\frac{3}{4} \|\nabla \rho^i_0\|^2 - \frac{J - 1 + \frac{2J}{3w_2}}{2w_2} \sum_{j=1}^{J} \|\nabla \rho^i_j\|^2 \\
&\overset{(7.1)}{\geq} -\frac{3}{4} \|\nabla \rho^i_0\|^2 - \frac{J(1+\frac{2}{3w_2})(d+1)}{2w_2 w_1^2} \sum_{j=1}^{J} \sum_{\mathbf{a} \in \mathcal{V}_{j-s}} \|\nabla \rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}.
\end{aligned}$$

We return to the main estimate and obtain the result by using definition (7.6)

$$(f,\rho^i_{J,\mathrm{alg}})-(\nabla u^i_J,\nabla\rho^i_{J,\mathrm{alg}})\geq\frac{1}{4}\|\nabla\rho^i_0\|^2+\left(\frac{1}{w_1}-\frac{J(1+\frac{2}{3w_2})(d+1)}{2w_2w_1^2}\right)\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}$$
$$\overset{(7.6)}{\geq} C^2_{\min}(w_1,w_2)\left(\|\nabla\rho^i_0\|^2+\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}\right). \tag{7.7}$$

The upper bound on $C^2_{\min}(w_1,w_2)$ in (7.6) is immediate due to the minimum in its expression, while the lower bound is obtained by rewriting condition (3.5) on w_2

$$w_2\geq\frac{5J^2(d+1)^2}{w_1 6J(d+1)\big(1-\frac{w_1}{6J(d+1)}\big)}\Leftrightarrow\frac{1}{w_1}-\frac{5}{6}\frac{J(d+1)}{w_2w_1^2}\geq\frac{1}{6J(d+1)}. \tag{7.8}$$

As also $w_2\geq 1$,

$$\frac{1}{w_1}-\frac{J(1+\frac{2}{3w_2})(d+1)}{2w_2w_1^2}\geq\frac{1}{w_1}-\frac{J(1+\frac{2}{3})(d+1)}{2w_2w_1^2}=\frac{1}{w_1}-\frac{5}{6}\frac{J(d+1)}{w_2w_1^2}\overset{(7.8)}{\geq}\frac{1}{6J(d+1)}.$$

□

7.3 Polynomial-degree-robust multilevel stable decomposition

Now, we devise a p-robust multilevel stable decomposition. This decomposition relies on the one level p-robust stable decomposition given in Schöberl et al. [43, Proof of Theorem 2.1] and the piecewise affine multilevel decomposition in the spirit of Xu, Chen, and Nochetto [50, Theorems 3.1 and 4.3]. These results are presented below in the form of lemmas. Note that in the decomposition, only "small" patches are used, which will be sufficient for our purposes. Recall also the definition of the local spaces (2.12), which will be useful below. Hereafter, we always assume that Assumption 2.1 is satisfied.

By [43, Proof of Theorem 2.1], we have the following.

Lemma 7.3 (One-level p-robust stable decomposition)**.** *For all $v_J\in V^p_J$, there exists a finest-level decomposition $v_J=v^\#_J+\sum_{\mathbf{b}\in\mathcal{V}_J}v^p_{J,\mathbf{b}}$, where $v^\#_J\in V^1_J$ and $v^p_{J,\mathbf{b}}\in V^{\mathbf{b}}_{J,0}$, $\mathbf{b}\in\mathcal{V}_J$, and this decomposition is stable in the sense*

$$\|\nabla v^\#_J\|^2+\sum_{\mathbf{b}\in\mathcal{V}_J}\|\nabla v^p_{J,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{J,0}}\leq C^2_{\mathrm{SD}}\|\nabla v_J\|^2, \tag{7.9}$$

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

Similarly to [50, Lemma 3.1 and Theorem 3.1], in the case of quasi-uniform meshes with bounded refinement strength, we have the following.

Lemma 7.4 ($\mathbb{P}_1$-multilevel stable decomposition for quasi-uniform meshes)**.** *For all $v^\#_J\in V^1_J$, there exists a multilevel piecewise affine decomposition $v^\#_J=v^1_0+\sum_{j=1}^J\sum_{\mathbf{b}\in\mathcal{V}_j}v^1_{j,\mathbf{b}}$ with $v^1_0\in V^1_0$ and $v^1_{j,\mathbf{b}}\in V^{\mathbf{b}}_{j,0}=\mathbb{P}_1(\mathcal{T}^{\mathbf{b}}_{j,0})\cap H^1_0(\omega^{\mathbf{b}}_{j,0})$. Under Assumption 2.2, this decomposition is stable as*

$$\|\nabla v^1_0\|^2+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}\|\nabla v^1_{j,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{j,0}}\leq C^2_{\mathrm{MD}}\|\nabla v^\#_J\|^2, \tag{7.10}$$

where $C_{\mathrm{MD}}\geq 1$ only depends on the space dimension d, the mesh shape regularity parameter $\kappa_{\mathcal{T}}$, the maximum strength of refinement parameter C_{ref}, and the quasi-uniformity parameter C_{qu}.

Proof. Let $v^\#_J\in V^1_J$. We first apply a levelwise decomposition that follows from Xu, Chen, and Nochetto in [50, Lemma 3.1] by keeping the gradient on level zero. This gives us $v^\#_J=\sum_{j=0}^J v^1_j$ with $v^1_j\in V^1_j$ such that

$$\|\nabla v^1_0\|^2+\sum_{j=1}^{J}h_j^{-2}\|v^1_j\|^2\leq C^2_{\mathrm{ml}}\|\nabla v^\#_J\|^2, \tag{7.11}$$

where $C_{\mathrm{ml}} \geq 1$ has the same dependencies as C_{MD}.

We further decompose each of the above $v_j^1 \in V_j^1$, $j \geq 1$, into patchwise components. For this purpose, we use the standard nodal decomposition $v_j^1 = \sum_{\mathbf{b}\in\mathcal{V}_j} v_{j,\mathbf{b}}^1$, where $v_{j,\mathbf{b}}^1 = v_j^1(\mathbf{b})\psi_{j,\mathbf{b}}$ belongs to the local space $V_{j,0}^{\mathbf{b}}$ for $p = p' = 1$. By stability of the $\mathbb{P}_1$ nodal decomposition,

$$\sum_{\mathbf{b}\in\mathcal{V}_j} \|v_{j,\mathbf{b}}^1\|_{\omega_{j,0}^{\mathbf{b}}}^2 \leq C_{\mathrm{nd}}^2 \|v_j^1\|^2, \tag{7.12}$$

where C_{nd} only depends on the space dimension d and the mesh shape regularity parameter $\kappa_{\mathcal{T}}$. This can, for instance, be shown by considering a patch $\omega_{j,0}^{\mathbf{b}}$ and an element K contained in the patch. Since $v_{j,\mathbf{b}}^1 = v_j^1(\mathbf{b})\psi_{j,\mathbf{b}} \in V_{j,0}^{\mathbf{b}}$ and by equivalence of norms in finite dimension, we have

$$\begin{aligned}\|v_j^1(\mathbf{b})\psi_{j,\mathbf{b}}\|_{\omega_{j,0}^{\mathbf{b}}} &\approx \|v_j^1(\mathbf{b})\psi_{j,\mathbf{b}}\|_K \leq \|v_j^1(\mathbf{b})\psi_{j,\mathbf{b}}\|_{\infty,K}|K|^{\frac{1}{2}} \\ &\leq \Big\|\sum_{\mathbf{b}\in\mathcal{V}_K} v_j^1(\mathbf{b})\psi_{j,\mathbf{b}}\Big\|_{\infty,K}|K|^{\frac{1}{2}} = \|v_j^1\|_{\infty,K}|K|^{\frac{1}{2}} \lesssim \|v_j^1\|_K.\end{aligned} \tag{7.13}$$

The result (7.12) is obtained by summing both sides over all vertices.

Now, the claim (7.10) follows by using an inverse inequality on patches, the quasi-uniformity of the meshes, the above decompositions as

$$\begin{aligned}\sum_{j=1}^J \sum_{\mathbf{b}\in\mathcal{V}_j} \|\nabla v_{j,\mathbf{b}}^1\|_{\omega_{j,0}^{\mathbf{b}}}^2 \leq C_{\mathrm{inv}}^2 \sum_{j=1}^J \sum_{\mathbf{b}\in\mathcal{V}_j} h_{\omega_{j,0}^{\mathbf{b}}}^{-2} \|v_{j,\mathbf{b}}^1\|_{\omega_{j,0}^{\mathbf{b}}}^2 &\overset{(2.8)}{\leq} C_{\mathrm{inv}}^2 C_{\mathrm{qu}}^{-2} \sum_{j=1}^J h_j^{-2} \sum_{\mathbf{b}\in\mathcal{V}_j} \|v_{j,\mathbf{b}}^1\|_{\omega_{j,0}^{\mathbf{b}}}^2 \\ &\overset{(7.12)}{\leq} C_{\mathrm{inv}}^2 C_{\mathrm{qu}}^{-2} C_{\mathrm{nd}}^2 \sum_{j=1}^J h_j^{-2} \|v_j^1\|^2 \overset{(7.11)}{\leq} C_{\mathrm{inv}}^2 C_{\mathrm{qu}}^{-2} C_{\mathrm{nd}}^2 C_{\mathrm{ml}}^2 \|\nabla v_J^{\#}\|^2\end{aligned}$$

and by summing the left-hand side with $\|\nabla v_0^1\|^2$, which satisfies a similar bound from (7.11). We set $C_{\mathrm{MD}}^2 := C_{\mathrm{inv}}^2 C_{\mathrm{qu}}^{-2} C_{\mathrm{nd}}^2 C_{\mathrm{ml}}^2 + C_{\mathrm{ml}}^2$ to obtain the result. □

By [50, Theorem 4.3], in the case of graded meshes, we have the following.

Lemma 7.5 ($\mathbb{P}_1$-multilevel stable decomposition for graded meshes)**.** *For all $v_J^{\#} \in V_J^1$, there exists a multilevel piecewise affine decomposition $v_J^{\#} = v_0^1 + \sum_{j=1}^J \sum_{\mathbf{b}\in\mathcal{V}_j} v_{j,\mathbf{b}}^1$ with $v_0^1 \in V_0^1$ and $v_{j,\mathbf{b}}^1 \in V_{j,0}^{\mathbf{b}} = \mathbb{P}_1(\mathcal{T}_{j,0}^{\mathbf{b}}) \cap H_0^1(\omega_{j,0}^{\mathbf{b}})$. Under Assumption 2.3, this decomposition is stable as*

$$\|\nabla v_0^1\|^2 + \sum_{j=1}^J \sum_{\mathbf{b}\in\mathcal{V}_j} \|\nabla v_{j,\mathbf{b}}^1\|_{\omega_{j,0}^{\mathbf{b}}}^2 \leq C_{\mathrm{MD}}^2 \|\nabla v_J^{\#}\|^2, \tag{7.14}$$

where $C_{\mathrm{MD}} \geq 1$ only depends on the space dimension d, the mesh shape regularity parameter $\kappa_{\mathcal{T}}$, the coarse mesh quasi-uniformity parameter C_{qu}^0, and the local quasi-uniformity parameter $C_{\mathrm{qu}}^{\mathrm{loc}}$.

Proof. Let $v_J^{\#} \in V_J^1$. We apply the results on stable decomposition on graded meshes of Xu, Chen, and Nochetto in [50, Theorem 4.3] adjusted to our setting. On the one hand, this gives us the decomposition

$$v_J^{\#} = v_0^1 + \sum_{j=1}^J \sum_{\mathbf{b}\in\mathcal{B}_j} v_{j,\mathbf{b}}^1 + \sum_{\mathbf{b}\in\mathcal{V}_J} v_{J,\mathbf{b}}^1,$$

where $v_0^1 \in V_0^1$, $\forall \mathbf{b} \in \mathcal{B}_j$, $v_{j,\mathbf{b}}^1 \in V_{j,0}^{\mathbf{b}}$ for $p = p' = 1$, and $\forall \mathbf{b} \in \mathcal{B}_J$, $v_{J,\mathbf{b}}^1 \in V_{J,0}^{\mathbf{b}}$ for $p = 1$. On the other hand, the result also gives us the following stability inequality:

$$\|\nabla v_0^1\|^2 + \sum_{j=1}^J h_{\mathcal{B}_j}^{-2} \Big\|\sum_{\mathbf{b}\in\mathcal{B}_j} v_{j,\mathbf{b}}^1\Big\|^2 + \sum_{\mathbf{b}\in\mathcal{V}_J} h_{\omega_{J,0}^{\mathbf{b}}}^{-2} \|v_{J,\mathbf{b}}^1\|_{\omega_{J,0}^{\mathbf{b}}}^2 \leq C_{\mathrm{gra}}^2 \|\nabla v_J^{\#}\|^2, \tag{7.15}$$

where C_{gra} has the same dependencies as C_{MD}.

First, since the mesh hierarchy is created via bisections, we have the local quasi-uniformity property (2.9). This, together with an inverse inequality and L^2-stability as in (7.13), gives us

$$\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{B}_j}\|\nabla v^1_{j,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{j,0}}\leq C^2_{\mathrm{inv}}\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{B}_j}h^{-2}_{\omega^{\mathbf{b}}_{j,0}}\|v^1_{j,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{j,0}}\overset{(2.9)}{\leq} C^2_{\mathrm{inv}}(C^{\mathrm{loc}}_{\mathrm{qu}})^{-2}C^2_{\mathrm{nd}}\sum_{j=1}^{J}h^{-2}_{\mathcal{B}_j}\Big\|\sum_{\mathbf{b}\in\mathcal{B}_j}v^1_{j,\mathbf{b}}\Big\|^2.$$

Second, we only need an inverse inequality to obtain

$$\sum_{\mathbf{b}\in\mathcal{V}_J}\|\nabla v^1_{J,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{J,0}}\leq C^2_{\mathrm{inv}}\sum_{\mathbf{b}\in\mathcal{V}_J}h^{-2}_{\omega^{\mathbf{b}}_{J,0}}\|v^1_{J,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{J,0}}.$$

Third, we can sum together the above estimations and use (7.15) to obtain

$$\|\nabla v^1_0\|^2+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{B}_j}\|\nabla v^1_{j,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{j,0}}+\sum_{\mathbf{b}\in\mathcal{V}_J}\|\nabla v^1_{J,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{J,0}}\leq C^2_{\mathrm{MD}}\|\nabla v^{\#}_J\|^2,$$

where $C_{\mathrm{MD}}:=C_{\mathrm{gra}}\cdot\max(C_{\mathrm{inv}},C_{\mathrm{inv}}C_{\mathrm{nd}}(C^{\mathrm{loc}}_{\mathrm{qu}})^{-1},1)$. Finally, since $\mathcal{B}_j\subset\mathcal{V}_j$, we can set $v^1_{j,\mathbf{b}}:=0$ for $\mathbf{b}\in\mathcal{V}_j\backslash\mathcal{B}_j$ and have a new decomposition $v^{\#}_J=v^1_0+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}v^1_{j,\mathbf{b}}$ (reusing the notation) such that (7.14) holds. □

Proposition 7.6 (p-robust multilevel stable decomposition)**.** *Let $v_J\in V^p_J$. Under either Assumption 2.2 or 2.3, there exists a decomposition*

$$v_J=v^1_0+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}v_{j,\mathbf{b}},\quad v^1_0\in V^1_0,\ v_{j,\mathbf{b}}\in V^{\mathbf{b}}_{j,0}\tag{7.16}$$

stable as

$$\|\nabla v^1_0\|^2+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}\|\nabla v_{j,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{j,0}}\leq C^2_{\mathrm{SMD}}\|\nabla v_J\|^2,\tag{7.17}$$

where $C_{\mathrm{SMD}}:=\sqrt{2}C_{\mathrm{SD}}C_{\mathrm{MD}}\geq 1$ only depends on the space dimension d, the mesh shape regularity parameter $\kappa_{\mathcal{T}}$, and, depending on whether Assumption 2.2 or 2.3 is satisfied, on either the quasi-uniformity parameter C_{qu} and the maximum strength of refinement parameter C_{ref} or the coarse mesh and local quasi-uniformity parameters C^0_{qu},$C^{\mathrm{loc}}_{\mathrm{qu}}$, respectively.

Proof. Let $v_J\in V^p_J$ and let us begin by applying the decomposition of Lemma 7.3. This gives $v_J=v^{\#}_J+\sum_{\mathbf{b}\in\mathcal{V}_J}v^p_{J,\mathbf{b}}$ with $v^{\#}_J\in V^1_J$ and $v^p_{J,\mathbf{b}}\in V^{\mathbf{b}}_{J,0}\ \forall\mathbf{b}\in\mathcal{V}_J$. Then, we further decompose $v^{\#}_J$ using either Lemma 7.4 or 7.5, depending on whether Assumption 2.2 or 2.3 is satisfied. We obtain $v^{\#}_J=v^1_0+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}v^1_{j,\mathbf{b}}$, where $v^1_0\in V^1_0$ and $v^1_{j,\mathbf{b}}\in V^{\mathbf{b}}_{j,0}$ (actually $V^{\mathbf{b}}_{j,0}\cap\mathbb{P}_1(\mathcal{T}^{\mathbf{b}}_{j,0})$). We set $v_{j,\mathbf{b}}:=v^1_{j,\mathbf{b}}\ \forall\mathbf{b}\in\mathcal{V}_j$, $j\in\{1,\dots,J-1\}$ and $v_{J,\mathbf{b}}:=v^1_{J,\mathbf{b}}+v^p_{J,\mathbf{b}}$. Thus, we have $v_J=v^1_0+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}v_{j,\mathbf{b}}$ with $v^1_0\in V^1_0$ and $v_{j,\mathbf{b}}\in V^{\mathbf{b}}_{j,0}$. The stable decomposition results presented in the previous lemmas allow us to write

$$\begin{aligned}\|\nabla v^1_0\|^2+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}\|\nabla v_{j,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{j,0}}&\leq 2\Big(\|\nabla v^1_0\|^2+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}\|\nabla v^1_{j,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{j,0}}+\sum_{\mathbf{b}\in\mathcal{V}_J}\|\nabla v^p_{J,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{J,0}}\Big)\\&\overset{(7.10)\text{ or }(7.14)}{\leq} 2C^2_{\mathrm{MD}}\Big(\|\nabla v^{\#}_J\|^2+\sum_{\mathbf{b}\in\mathcal{V}_J}\|\nabla v^p_{J,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{J,0}}\Big)\overset{(7.9)}{\leq} 2C^2_{\mathrm{SD}}C^2_{\mathrm{MD}}\|\nabla v_J\|^2.\end{aligned}$$

□

7.4 Upper bound on $\|\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\|$

Recall that $\tilde{\rho}^i_{J,\mathrm{alg}}$, introduced in (3.1), is the unknown exact algebraic error. We now estimate $\|\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\|$ from above. We introduce some helpful notation first. For all $\mathbf{a}\in\mathcal{V}_{j-s}$, let $\mathcal{I}_{\mathbf{a}}\subset\mathcal{V}_j$ be a set containing (fine-mesh) vertices of the interior of the patch $\omega^{\mathbf{a}}_{j,s}$ such that $\{\mathcal{I}_{\mathbf{a}}\}_{\mathbf{a}\in\mathcal{V}_{j-s}}$ cover $\mathcal{V}_j$ and are mutually disjoint; if $s=0$, we have $\mathcal{I}_{\mathbf{a}}=\{\mathbf{a}\}$. This allows us to write $\sum_{\mathbf{b}\in\mathcal{V}_j}=\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}$. Moreover, since the indices of $\mathcal{I}_{\mathbf{a}}$ are localized in the interior of the patch $\omega^{\mathbf{a}}_{j,s}$, we have $\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}v_{j,\mathbf{b}}\in V^{\mathbf{a}}_{j,s}$ when $v_{j,\mathbf{b}}\in V^{\mathbf{b}}_{j,0}$. Writing it this way will help us to apply the results on the p-robust stable decomposition of Lemma 7.6 given for "small" patches only to the "large"-patch setting as well.

Lemma 7.7 (Estimating $\|\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\|$ by local contributions)**.** *Let $\tilde{\rho}^i_{J,\mathrm{alg}}\in V^p_J$ be defined by (3.1). We have*

$$\|\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\|^2\le C^2_{\max}(w_1,w_2)\left(\|\nabla\rho^i_0\|^2+\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}\right),\tag{7.18}$$

where

$$2(d+1)\le C^2_{\max}(w_1,w_2):=2(d+1)C^2_{\mathrm{SMD}}\left(1+\frac{J^2}{w_2}\max\left(1,\frac{d+1}{w_1^2}\right)\right)\le 4(d+1)^2C^2_{\mathrm{SMD}}J^2.\tag{7.19}$$

Proof. The main ingredient of the proof is to replace locally the uncomputable $\tilde{\rho}^i_{J,\mathrm{alg}}=u_J-u^i_J$ by the constructed local contributions $\rho^i_{j,\mathbf{a}}$ using the problems they solve on patches. We begin by using Proposition 7.6 applied to $\tilde{\rho}^i_{J,\mathrm{alg}}\in V^p_J$, writing $\tilde{\rho}^i_{J,\mathrm{alg}}=e_0+\sum_{j=1}^J\sum_{\mathbf{b}\in\mathcal{V}_j}e_{j,\mathbf{b}}$. Then

$$\begin{aligned}
\|\nabla\tilde{\rho}^i_{J,\mathrm{alg}}\|^2&=\Big(\nabla\tilde{\rho}^i_{J,\mathrm{alg}},\nabla e_0+\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}\nabla e_{j,\mathbf{b}}\Big)\overset{(3.4)}{=}\Big(\nabla\rho^i_0,\nabla e_0\Big)+\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\Big(\nabla\tilde{\rho}^i_{J,\mathrm{alg}},\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}\nabla e_{j,\mathbf{b}}\Big)_{\omega^{\mathbf{a}}_{j,s}}\\
&\overset{(3.1)}{=}\Big(\nabla\rho^i_0,\nabla e_0\Big)+\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\left(\Big(f,\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}e_{j,\mathbf{b}}\Big)_{\omega^{\mathbf{a}}_{j,s}}-\Big(\nabla u_J,\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}\nabla e_{j,\mathbf{b}}\Big)_{\omega^{\mathbf{a}}_{j,s}}\right)\\
&\overset{(3.8)}{=}\Big(\nabla\rho^i_0,\nabla e_0\Big)+\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\left(\Big(\nabla\rho^i_{j,\mathbf{a}},\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}\nabla e_{j,\mathbf{b}}\Big)_{\omega^{\mathbf{a}}_{j,s}}+\frac{1}{w_2}\sum_{k=0}^{j-1}\Big(\nabla\rho^i_k,\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}\nabla e_{j,\mathbf{b}}\Big)_{\omega^{\mathbf{a}}_{j,s}}\right)\\
&=\Big(\nabla\rho^i_0,\nabla e_0\Big)+\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\Big(\nabla\rho^i_{j,\mathbf{a}},\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}\nabla e_{j,\mathbf{b}}\Big)_{\omega^{\mathbf{a}}_{j,s}}+\frac{1}{w_2}\sum_{j=1}^{J}\sum_{k=0}^{j-1}\Big(\nabla\rho^i_k,\sum_{\mathbf{b}\in\mathcal{V}_j}\nabla e_{j,\mathbf{b}}\Big).
\end{aligned}$$

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 7.1. First, we have

$$\Big(\nabla\rho^i_0,\sum_{\mathbf{b}\in\mathcal{V}_0}\nabla e_{0,\mathbf{b}}\Big)\le\frac{C^2_{\mathrm{SMD}}}{2}\|\nabla\rho^i_0\|^2+\frac{1}{2C^2_{\mathrm{SMD}}}\|\nabla e_0\|^2.$$

For the second term, we similarly obtain

$$\begin{aligned}
\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\Big(\nabla\rho^i_{j,\mathbf{a}},\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}\nabla e_{j,\mathbf{b}}\Big)_{\omega^{\mathbf{a}}_{j,s}}&\le\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\left(\frac{2(d+1)C^2_{\mathrm{SMD}}}{2}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}+\frac{\Big\|\sum_{\mathbf{b}\in\mathcal{I}_{\mathbf{a}}}\nabla e_{j,\mathbf{b}}\Big\|^2_{\omega^{\mathbf{a}}_{j,s}}}{2(2(d+1)C^2_{\mathrm{SMD}})}\right)\\
&\le(d+1)C^2_{\mathrm{SMD}}\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}}+\frac{1}{4C^2_{\mathrm{SMD}}}\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}\|\nabla e_{j,\mathbf{b}}\|^2_{\omega^{\mathbf{b}}_{j,0}}.
\end{aligned}$$

Finally, for the third term we additionally use the property $w_2 \geq 1$ and rename summation indices when there is no confusion:

$$\frac{1}{w_2}\sum_{j=1}^{J}\sum_{k=0}^{j-1}\Big(\nabla\rho_k^i, \sum_{\mathbf{b}\in\mathcal{V}_j}\nabla e_{j,\mathbf{b}}\Big) \leq \frac{2(d+1)C_{\mathrm{SMD}}^2 J}{2w_2}\sum_{j=1}^{J}\sum_{k=0}^{j-1}\|\nabla\rho_k^i\|^2 + \frac{\sum_{j=1}^{J}\sum_{k=0}^{j-1}\Big\|\sum_{\mathbf{b}\in\mathcal{V}_j}\nabla e_{j,\mathbf{b}}\Big\|^2}{2w_2(2(d+1)C_{\mathrm{SMD}}^2 J)}$$
$$\leq \frac{(d+1)C_{\mathrm{SMD}}^2 J^2}{w_2}\sum_{k=0}^{J}\|\nabla\rho_k^i\|^2 + \frac{1}{4C_{\mathrm{SMD}}^2}\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}\|\nabla e_{j,\mathbf{b}}\|_{\omega_{j,0}^{\mathbf{b}}}^2$$
$$\overset{(7.1)}{\leq} \frac{(d+1)C_{\mathrm{SMD}}^2 J^2}{w_2}\Big(\|\nabla\rho_0^i\|^2 + \frac{d+1}{w_1^2}\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_{j,s}^{\mathbf{a}}}^2\Big) + \frac{1}{4C_{\mathrm{SMD}}^2}\sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}\|\nabla e_{j,\mathbf{b}}\|_{\omega_{j,0}^{\mathbf{b}}}^2.$$

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

$$\|\nabla\tilde{\rho}_{J,\mathrm{alg}}^i\|^2 \overset{(7.19)}{\leq} \frac{C_{\max}^2(w_1,w_2)}{2}\Big(\|\nabla\rho_0^i\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_{j,s}^{\mathbf{a}}}^2\Big) + \frac{\|\nabla e_0\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{b}\in\mathcal{V}_j}\|\nabla e_{j,\mathbf{b}}\|_{\omega_{j,0}^{\mathbf{b}}}^2}{2C_{\mathrm{SMD}}^2}$$
$$\overset{(7.17)}{\leq} \frac{C_{\max}^2(w_1,w_2)}{2}\Big(\|\nabla\rho_0^i\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_{j,s}^{\mathbf{a}}}^2\Big) + \frac{1}{2}\|\nabla\tilde{\rho}_{J,\mathrm{alg}}^i\|^2.$$

After subtracting $\frac{1}{2}\|\nabla\tilde{\rho}_{J,\mathrm{alg}}^i\|^2$ on both sides, we finally obtain the desired result.

The lower bound on $C_{\max}^2(w_1,w_2)$ in (7.19) is obtained by using $C_{\mathrm{SMD}} \geq 1$ from Proposition 7.6. To derive the upper bound, we use the fact that weights of Definition 3.2 satisfy $w_1 \geq 1$, $w_2 \geq 1$. This gives $\frac{J^2}{w_2} \leq J^2$ and $\frac{d+1}{w_1^2} \leq d+1$, leading to the desired result:

$$C_{\max}^2(w_1,w_2) \leq 2(d+1)C_{\mathrm{SMD}}^2\left(1 + J^2(d+1)\right) \leq 2(d+1)C_{\mathrm{SMD}}^2\left(2J^2(d+1)\right).$$

□

7.5 Proof of Theorem 5.1

The results of the previous subsections allow us now to give a concise proof of Theorem 5.1.

Proof of Theorem 5.1. Case $\rho_{J,\mathrm{alg}}^i = 0$. By Definition 4.1 this means $\eta_{\mathrm{alg}}^i = 0$, so that it suffices to show that $u_J = u_J^i$ in this case. We do this by using Lemmas 7.2 and 7.7, which lead to

$$\|\nabla(u_J - u_J^i)\|^2 \overset{(3.2)}{=} \|\nabla\tilde{\rho}_{J,\mathrm{alg}}^i\|^2 \overset{(7.18)}{\leq} C_{\max}^2(w_1,w_2)\left(\|\nabla\rho_0^i\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_{j,s}^{\mathbf{a}}}^2\right)$$
$$\overset{(7.5)}{\leq} \frac{C_{\max}^2(w_1,w_2)}{C_{\min}^2(w_1,w_2)}\Big((f,\rho_{J,\mathrm{alg}}^i) - (\nabla u_J^i, \nabla\rho_{J,\mathrm{alg}}^i)\Big) = 0. \tag{7.20}$$

Case $\rho_{J,\mathrm{alg}}^i \neq 0$. In this case, we combine the results of Lemmas 7.1, 7.2, and 7.7

$$\eta_{\mathrm{alg}}^i = \frac{(f,\rho_{J,\mathrm{alg}}^i) - (\nabla u_J^i, \nabla\rho_{J,\mathrm{alg}}^i)}{\|\nabla\rho_{J,\mathrm{alg}}^i\|} \overset{(7.2)}{\underset{(7.7)}{\geq}} \frac{C_{\min}^2(w_1,w_2)}{C_{\max}(w_1)}\left(\|\nabla\rho_0^i\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_{j,s}^{\mathbf{a}}}^2\right)^{\frac{1}{2}}$$
$$\overset{(7.18)}{\geq} \frac{C_{\min}^2(w_1,w_2)}{C_{\max}(w_1)C_{\max}(w_1,w_2)}\|\nabla\tilde{\rho}_{J,\mathrm{alg}}^i\| \overset{(3.2)}{=} \beta\|\nabla(u_J - u_J^i)\| \tag{7.21}$$

for

$$\frac{1}{12\sqrt{2}C_{\mathrm{SMD}}J^{\frac{5}{2}}(d+1)^{\frac{5}{2}}} \leq \beta := \frac{C_{\min}^2(w_1,w_2)}{C_{\max}(w_1)C_{\max}(w_1,w_2)} \leq \frac{1}{8\sqrt{d+1}}.$$

The bounds on β follow from (7.3), (7.6), and (7.19). □

Example 7.8 (Specific choices of weights)**.** *We illustrate here a bound on the efficiency factor β in (5.1) for different choices of the damping weights satisfying the compatibility condition (3.5) from Remark 3.4:*

$$w_1 = J(d+1) \text{ and } w_2 = 1: \qquad \frac{1}{12C_{\mathrm{SMD}}J^2\sqrt{2(d+1)^3}} \leq \beta.$$

$$w_1 = d+1 \text{ and } w_2 = J: \qquad \frac{1}{12C_{\mathrm{SMD}}J\sqrt{2(d+1)^3}} \leq \beta.$$

$$w_1 = w_2 = \sqrt{J(d+1)}: \qquad \frac{1}{12\sqrt{2}C_{\mathrm{SMD}}J^{\frac{5}{4}}(d+1)} \leq \beta.$$

$$w_1 = 1 \text{ and } w_2 = \infty: \qquad \frac{1}{8C_{\mathrm{SMD}}\sqrt{J}(d+1)} \leq \beta.$$

$$w_1 = 4\sqrt{J} \text{ and } w_2 = \infty: \qquad \frac{1}{8C_{\mathrm{SMD}}\sqrt{J(d+1)}} \leq \beta.$$

7.6 Proof of Corollary 5.6

If $\rho^i_{J,\mathrm{alg}} = 0$, as a result of (7.21), we have $\|\nabla(u_J - u^i_J)\| = \|\nabla\rho^i_0\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_{j-s}}\|\nabla\rho^i_{j,\mathbf{a}}\|^2_{\omega^{\mathbf{a}}_{j,s}} = 0$. Otherwise by (7.18) we set $C_1 := C_{\max}(w_1, w_2)$ and by (4.2) and (7.22), $C_2 = \dfrac{1}{\beta} = \dfrac{C_{\max}(w_1)C_{\max}(w_1,w_2)}{C^2_{\min}(w_1,w_2)}$ gives the result.

8 Conclusions and outlook

In this work, we presented a hierarchical construction of the algebraic residual lifting in the spirit of Papež et al. [38]. This lifting approximates the algebraic error by one iteration of a V-cycle multigrid with no presmoothing step, a single damped additive Schwarz postsmoothing step, and a coarse solve of the lowest polynomial degree. The lifting leads us to an a posteriori estimator on the algebraic error and to a linear iterative solver. We showed that the two following results are equivalent: the (reliable) a posteriori estimator is p-robustly efficient, and the solver contracts p-robustly the error at each iteration. The provided numerical tests agree with these theoretical findings. Moreover, we also presented numerical results for a modified solver corresponding to a weighted restricted additive Schwarz smoothing. In accordance with the literature, this modified solver provides a further speed-up compared to the damped Schwarz smoothing. Although we currently cannot show that our p-robust theoretical result also applies to this construction, the use of high-degree polynomials does not seem to cause a degradation of the solver. So far, our theory involves estimates depending algebraically on the number of mesh levels J, which we do not observe in the numerical results for the weighted restricted variant. In forthcoming works, we plan to develop adaptivity based on the property (5.4), i.e., a computable splitting equivalent to the error and localized not only levelwise but also patchwise. Applications to more involved problems are also on our work list.

References

[1] M. Ainsworth, *A preconditioner based on domain decomposition for h-p finite-element approximation on quasi-uniform meshes*, SIAM J. Numer. Anal., 33 (1996), pp. 1358–1376.

[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.

[3] P. F. Antonietti, M. Sarti, M. Verani, and L. T. Zikatanov, *A uniform additive Schwarz preconditioner for high-order discontinuous Galerkin approximations of elliptic problems*, J. Sci. Comput., 70 (2017), pp. 608–630.

[4] M. Arioli, E. H. Georgoulis, and D. Loghin, *Stopping criteria for adaptive finite element solvers*, SIAM J. Sci. Comput., 35 (2013), pp. A1537–A1559.

[5] I. Babuška, A. Craig, J. Mandel, and J. Pitkäranta, *Efficient preconditioning for the p-version finite element method in two dimensions*, SIAM J. Numer. Anal., 28 (1991), pp. 624–661.

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

[7] P. Bastian, M. Blatt, and R. Scheichl, *Algebraic multigrid for discontinuous Galerkin discretizations of heterogeneous elliptic problems*, Numer. Linear Algebra Appl., 19 (2012), pp. 367–388.

[8] R. Becker, C. Johnson, and R. Rannacher, *Adaptive error control for multigrid finite element methods*, Computing, 55 (1995), pp. 271–288.

[9] F. A. Bornemann and P. Deuflhard, *The cascadic multigrid method for elliptic problems*, Numer. Math., 75 (1996), pp. 135–152.

[10] L. Botti, A. Colombo, and F. Bassi, *h-multigrid agglomeration based solution strategies for discontinuous Galerkin discretizations of incompressible flow problems*, J. Comput. Phys., 347 (2017), pp. 382–415.

[11] 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.

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

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

[14] S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, vol. 15 of Texts in Applied Mathematics, Springer, New York, third ed., 2008.

[15] X.-C. Cai and M. Sarkis, *A restricted additive Schwarz preconditioner for general sparse linear systems*, SIAM J. Sci. Comput., 21 (1999), pp. 792–797.

[16] C. Canuto and A. Quarteroni, *Preconditioned minimal residual methods for Chebyshev spectral calculations*, J. Comput. Phys., 60 (1985), pp. 315–337.

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

[18] P. G. Ciarlet, *The finite element method for elliptic problems*, North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978. Studies in Mathematics and its Applications, Vol. 4.

[19] E. Efstathiou and M. J. Gander, *Why restricted additive Schwarz converges faster than additive Schwarz*, BIT, 43 (2003), pp. 945–959.

[20] A. Ern and J.-L. Guermond, *Theory and practice of finite elements*, vol. 159 of Applied Mathematical Sciences, Springer-Verlag, New York, 2004.

[21] S. Foresti, G. Brussino, S. Hassanzadeh, and V. Sonnad, *Multilevel solution method for the p-version of finite elements*, Computer Physics Communications, 53 (1989), pp. 349 – 355.

[22] A. Gholami, D. Malhotra, H. Sundar, and G. Biros, *FFT, FMM, or multigrid? A comparative study of state-of-the-art Poisson solvers for uniform and nonuniform grids in the unit cube*, SIAM J. Sci. Comput., 38 (2016), pp. C280–C306.

[23] M. Griebel, P. Oswald, and M. A. Schweitzer, *A particle-partition of unity method. VI. A p-robust multilevel solver*, in Meshfree methods for partial differential equations II, vol. 43 of Lect. Notes Comput. Sci. Eng., Springer, Berlin, 2005, pp. 71–92.

[24] W. Hackbusch, *Multi-grid methods and applications*, vol. 4 of Springer Series in Computational Mathematics, Springer, Berlin, 2003.

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

[26] R. Hiptmair, H. Wu, and W. Zheng, *Uniform convergence of adaptive multigrid methods for elliptic problems and Maxwell's equations*, Numer. Math. Theory Methods Appl., 5 (2012), pp. 297–332.

[27] 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.

[28] P. Jiránek, Z. Strakoš, and M. Vohralík, *A posteriori error estimates including algebraic error and stopping criteria for iterative solvers*, SIAM J. Sci. Comput., 32 (2010), pp. 1567–1590.

[29] G. Kanschat, *Robust smoothers for high-order discontinuous Galerkin discretizations of advection-diffusion problems*, J. Comput. Appl. Math., 218 (2008), pp. 53–60.

[30] M. Kronbichler and W. A. Wall, *A performance comparison of continuous and discontinuous Galerkin methods with fast multigrid solvers*, SIAM J. Sci. Comput., 40 (2018), pp. A3423–A3448.

[31] S. Loisel, R. Nabben, and D. B. Szyld, *On hybrid multigrid-Schwarz algorithms*, J. Sci. Comput., 36 (2008), pp. 165–175.

[32] J. P. Lucero Lorca and G. Kanschat, *Multilevel Schwarz preconditioners for singularly perturbed symmetric reaction-diffusion systems*. arXiv:1811.03839 preprint, 2018. URL https://arxiv.org/abs/1811.03839v1.

[33] J. Mandel, *Two-level domain decomposition preconditioning for the p-version finite element method in three dimensions*, Internat. J. Numer. Methods Engrg., 29 (1990), pp. 1095–1108.

[34] D. Meidner, R. Rannacher, and J. Vihharev, *Goal-oriented error control of the iterative solution of finite element equations*, J. Numer. Math., 17 (2009), pp. 143–172.

[35] J. Nocedal and S. J. Wright, *Numerical optimization*, Springer Series in Operations Research and Financial Engineering, Springer, New York, second ed., 2006.

[36] Y. Notay and A. Napov, *A massively parallel solver for discrete Poisson-like problems*, J. Comput. Phys., 281 (2015), pp. 237–250.

[37] P. Oswald, *Multilevel finite element approximation*, Teubner Skripten zur Numerik. [Teubner Scripts on Numerical Mathematics], B. G. Teubner, Stuttgart, 1994. Theory and applications.

[38] 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*, Comput. Methods Appl. Mech. Engrg., 371 (2020), p. 113243.

[39] J. Papež, Z. Strakoš, and M. Vohralík, *Estimating and localizing the algebraic and total numerical errors using flux reconstructions*, Numer. Math., 138 (2018), pp. 681–721.

[40] J. Papež and M. Vohralík, *Inexpensive guaranteed and efficient upper bounds on the algebraic error in finite element discretizations*. HAL preprint 02422851, 2019. URL https://hal.inria.fr/hal-02422851.

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

[42] A. Quarteroni and G. Sacchi Landriani, *Domain decomposition preconditioners for the spectral collocation method*, J. Sci. Comput., 3 (1988), pp. 45–76.

[43] 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.

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

[45] E. G. Sewell, *Automatic generation of triangulations for piecewise polynomial approximation*, ProQuest LLC, Ann Arbor, MI, 1972. Thesis (Ph.D.)–Purdue University.

[46] 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.

[47] T. Warburton, *An explicit construction of interpolation nodes on the simplex*, J. Engrg. Math., 56 (2006), pp. 247–262.

[48] H. Wu and Z. Chen, *Uniform convergence of multigrid V-cycle on adaptively refined finite element meshes for second order elliptic problems*, Sci. China Ser. A, 49 (2006), pp. 1405–1429.

[49] J. Xu, *Iterative methods by space decomposition and subspace correction*, SIAM Rev., 34 (1992), pp. 581–613.

[50] 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.

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