Contractive local adaptive smoothing based on Dörfler marking in a-posteriori-steered p-robust multigrid solvers

Ani Miraçi, Jan Papež, Martin Vohralík

Contractive local adaptive smoothing based on Dörfler marking in a-posteriori-steered p-robust multigrid solvers*

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

September 7, 2020

Abstract

In this work we study a local adaptive smoothing algorithm for a-posteriori-steered p-robust multigrid methods. The solver tackles a linear system which is generated by the discretization of a second-order elliptic diffusion problem using conforming finite elements of polynomial order $p \geq 1$. After one V-cycle ("full-smoothing" substep) of the solver of [HAL Preprint 02494538, 2020], we dispose of a reliable, efficient, and localized estimation of the algebraic error. We use this existing result to develop our new adaptive algorithm: thanks to the information of the estimator and based on a bulk-chasing criterion, cf. Dörfler [SIAM J. Numer. Anal., 33 (1996), pp. 1106–1124], we mark patches of elements and levels with increased estimated error. Then, we proceed by a modified and cheaper V-cycle ("adaptive-smoothing" substep), which only applies smoothing in the marked regions. The proposed adaptive multigrid solver picks autonomously and adaptively the optimal step-size per level as in our previous work but also the type of smoothing per level (weighted restricted additive or additive Schwarz) and concentrates smoothing to marked regions with high error. We prove that each substep (full and adaptive) contracts the error p-robustly, which is confirmed by numerical experiments. Moreover, the proposed algorithm behaves numerically robustly with respect to the number of levels as well as to the diffusion coefficient jump.

Key words: finite element method, multigrid method, Schwarz method, block-Jacobi smoother, a posteriori estimate, adaptive smoothing, stable decomposition, p-robustness

1 Introduction

The finite element method is a widespread and versatile discretization method for partial differential equations, see e.g. Ciarlet [9], Ern and Guermond [12], or Brenner and Scott [4]. In particular, the use of high-order methods has shown numerous advantages in terms of accuracy, see e.g. Szabó and Babuška [29], Bernardi and Maday [3], Šolín *et al.* [30], and the references therein. The implementation of these methods however, leads to a linear system that is abundantly bigger than for low-order discretizations. This appeals to the use of an appropriate iterative solver. Moreover, since the conditioning degrades with increasing order, commonly used solvers begin to suffer. Amongst the most efficient solvers we mention multigrid solvers, see e.g. Hackbusch [14], Briggs *et al.* [6], more generally multilevel methods e.g. Zhang [32], Oswald [22], Griebel and Oswald [13], and the closely related domain decomposition methods, e.g. Quarteroni and Valli [25] or Dolean *et al.* [10]. Note that the above methods can be used in their own right as fixed point methods, or as a preconditioner (possibly after making them symmetric).

The idea of defining an *adaptive* algebraic solver is rather old. On the subject of local smoothing methods, we refer, e.g., to Bai and Brandt [2], McCormick [19], Rüde [26], Lötzbeyer and Rüde [18], and

*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 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

[§]Institute of Mathematics, Czech Academy of Sciences, Žitná 25, 115 67 Prague, Czech Republic

more recently Xu *et al.* [31], Janssen and Kanschat [16], or Chen *et al.* [8]. Here the smoothing is typically localized to parts where the adaptive mesh refinement was performed, but it is not adaptive per se. Adaptive smoothed aggregation aiming at building a coarser linear system by determining near-kernel components was proposed in the context of algebraic multigrid, see e.g. Brezina *et al.* [5] and the references therein. More recently, an aggregation based on path covers was proposed by Hu *et al.* [15]. Another interesting approach consists in applying an adaptive construction of preconditioners, see, e.g., the recent work of Anciaux-Sedrakian *et al.* [1], where the adaptivity relies on a posteriori error estimates of the algebraic error, cf. Papež *et al.* [23, 24], combined with a bulk-chasing criterion in the spirit of Dörfler [11]. To the best of the authors' knowledge, this is the first time a bulk-chasing criterion is used in an algebraic solver adaptivity (and not mesh refinement) setting. However, the results therein are mainly numerical, whereas mathematical analysis is not really developed.

The subject of this work is to propose a multigrid solver with *local adaptive smoothing* based on rigorous a posteriori error estimates of the algebraic error and a bulk-chasing criterion, and to prove its convergence. We rely on the polynomial-degree-robust solver introduced in Miraçi *et al.* [21], which is a geometric multigrid whose iteration consists of a V-cycle with zero pre- and one post-smoothing step, where the smoothing is overlapping additive Schwarz (block-Jacobi) associated to patches of elements. This solver already contains a first adaptive step, since the error correction update from one level to the next, in contrast to a standard multigrid, picks the optimal (adaptive) step-size that reduces the algebraic error in the best possible way. The results of [21] also give us a *reliable* and *efficient a posteriori estimator* on the algebraic error and equivalence of the algebraic error with *localized* (levelwise/patchwise) computable estimators that serve as a starting point for our current contribution.

In this work, after implementing one step of the original solver of [21] (one full-smoothing V-cycle), we obtain a fairly good indication of where (levelwise/patchwise) the algebraic error is concentrated. We then use a bulk-chasing criterion to mark the highest contribution regions, and then perform a cheaper step (one adaptive-smoothing V-cycle) only smoothing in these problematic regions. Additionally, based on numerical performance and literature results, see, e.g., Cai and Sarkis [7] or Loisel *et al.* [17], we give the solver the option to pick adaptively the type of smoothing, be it additive Schwarz or (the typically better performing) weighted restricted additive Schwarz. We focus on quasi-uniform meshes, but our theory also applies to possibly highly graded bisection grids, where smoothing is already local around the refinement edges.

We prove that the algorithm we present contracts the error in *each of the substeps*, the full-smoothing and the adaptive-smoothing, *robustly* with respect to the *polynomial degree* p of the underlying finite element discretization. The results on the full-smoothing substep rely on [21], where a p-robust stable decomposition for one level by Schöberl *et al.* [27], and a multilevel stable decomposition for piecewise affine polynomials on quasi-uniform/bisection grids by Xu *et al.* [31] are crucial. Numerically, we additionally observe robustness with respect to the number of levels in the mesh hierarchy as well as the jumps in the diffusion coefficient.

Compared to [21], the novelties of this work are: 1) Development of a new kind of adaptivity that is local in patches with increased algebraic error, whereas the adaptivity in [21] chooses the number of post-smoothing steps globally per level. 2) Localization in space relying on Dörfler's marking. 3) Proof that the new adaptive sub-step contracts the error p-robustly, despite it only smoothes in marked patches (no convergence proof of the adaptive scheme is given in [21]). 4) Adaptive decision on which smoothing (additive Schwarz or weighted restricted additive Schwarz) variant to employ per level and inclusion of the weighted restricted additive Schwarz in the analysis, which was not done in [21].

The manuscript is organized as follows. In Section 2 we introduce the model problem and the notation we will be working with. Section 3 presents in detail the algorithmic description of the solver with each of its modules, as well as the rigorous mathematical definition of the solver. In Section 4 we define the algebraic error estimator. The main results are collected in Section 5, and the numerical tests are showcased in Section 6. Section 7 gives the proofs of our main results. Finally, some concluding remarks are given in Section 8.

2 Setting

In this section we present the model problem we will be studying and the notation needed for the multilevel setting we work on.

2.1 Model problem and its finite element discretization

We work with a second-order elliptic problem defined over $\Omega \subset \mathbb{R}^d$, $d \in \{1,2,3\}$, an open bounded polytope with a Lipschitz-continuous boundary. In the weak formulation, we search for $u \in H_0^1(\Omega)$ such that

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

where $f \in L^2(\Omega)$ is a source term and $\boldsymbol{\mathcal{K}} \in [L^\infty(\Omega)]^{d\times d}$ is a symmetric and positive definite diffusion coefficient.

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

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

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

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

2.2 A hierarchy of meshes and spaces

We rely in this contribution on a hierarchy of matching simplicial meshes $\{\mathcal{T}_j\}_{0\leq j\leq J}$, where $\mathcal{T}_J$ has been introduced in Section 2.1, and where $\mathcal{T}_j$ is a refinement of $\mathcal{T}_{j-1}$, $1 \leq j \leq J$. We also introduce a hierarchy of finite element spaces associated to the mesh hierarchy. For this purpose, fix p_j the polynomial degree associated to mesh level $j \in \{1,\dots,J\}$ such that $1 \leq p_1 \leq \dots \leq p_{J-1} \leq p_J = p$. We then introduce

$$\text{for } j = 0: \qquad V_0^1 := \mathbb{P}_1(\mathcal{T}_0) \cap H_0^1(\Omega) \quad \text{(lowest-order space)}, \tag{2.4a}$$

$$\text{for } 1 \leq j \leq J-1: \qquad V_j^{p_j} := \mathbb{P}_{p_j}(\mathcal{T}_j) \cap H_0^1(\Omega) \quad (p_j\text{-th order spaces}), \tag{2.4b}$$

where $\mathbb{P}_{p_j}(\mathcal{T}_j) := \{v_j \in L^2(\Omega), v_j|_K \in \mathbb{P}_{p_j}(K) \ \ \forall K \in \mathcal{T}_j\}$. Note that $V_0^1 \subset V_1^{p_1} \subset \dots \subset V_{J-1}^{p_{J-1}} \subset V_J^p$, so that the spaces are nested. We also formally set $p_0 = 1$. 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 $\mathbf{a} \in \mathcal{V}_j$, $0 \leq j \leq J$; 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 vertices of $\mathcal{V}_j$.

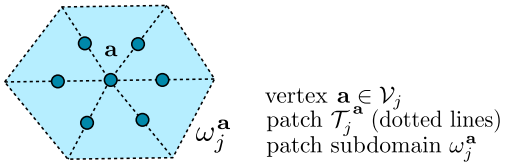


Figure 1: Illustration of a patch $\mathcal{T}_j^{\mathbf{a}}$, the patch subdomain $\omega_j^{\mathbf{a}}$, and of the degrees of freedom for the space $V_j^{\mathbf{a}}$ with $p_j = 2$.

For the following, we need to define the notion of patches of elements, illustrated in Figure 1. Let $j \in \{1,\dots,J\}$. For any element $K \in \mathcal{T}_j$, we denote by $\mathcal{V}_K$ the set of its vertices. Then, given an arbitrary vertex $\mathbf{a} \in \mathcal{V}_j$, we denote by $\mathcal{T}_j^{\mathbf{a}}$ the patch formed by all elements of the mesh $\mathcal{T}_j^{\mathbf{a}}$ sharing the vertex $\mathbf{a}$, i.e., $\mathcal{T}_j^{\mathbf{a}} := \{K \in \mathcal{T}_j, \mathbf{a} \in \mathcal{V}_K\}$. Then we denote by $\omega_j^{\mathbf{a}}$ the open patch subdomain corresponding to $\mathcal{T}_j^{\mathbf{a}}$. Finally, the associated local space is $V_j^{\mathbf{a}}$

$$V_j^{\mathbf{a}} := \mathbb{P}_{p_j}(\mathcal{T}_j^{\mathbf{a}}) \cap H_0^1(\omega_j^{\mathbf{a}}),\ j \in \{1,\dots,J\}. \tag{2.5}$$

Larger subdomains can also be considered, cf. [20]. Finally, denote by $\mathcal{I}_j^{p_j}$ the $\mathbb{P}^{p_j}$ Lagrange interpolation operator on the mesh level j, i.e. $\mathcal{I}_j^{p_j} : C^0(\overline{\Omega}) \to V_j^{p_j}$, $\mathcal{I}_j^{p_j}(v)$ preseves the values of v in the nodes corresponding to the Lagrange degrees of freedom. This will play an important role in the adaptive choice of smoothing of the solver presented below in Section 3.

3 Adaptive multilevel solver

The basic idea of our adaptive solver is illustrated by Figure 2. In Section 3.1, we give an algorithmic description of the solver, followed by the explanation of its constituting modules. Then in Section 3.2, we provide a mathematical description of the solver, lengthier but better suited for the forthcoming theoretical analysis.

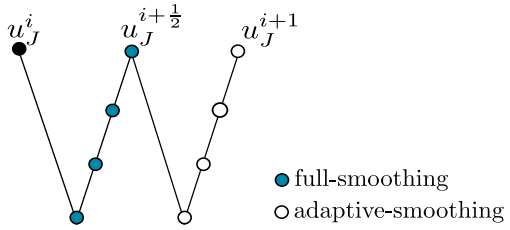


Figure 2: Illustration of the full-smoothing and adaptive-smoothing V-cycle substeps, $J = 3$.

3.1 Algorithmic description of the solver

The adaptive solver we propose can be written in an algorithmic description:

Algorithm 1: A-posteriori-steered multigrid with local adaptive smoothing

Input: [polynomial degree p, number of levels J, bulk-chasing parameter θ, adaptivity-decision parameter γ, requested tolerance tol]

$i := 0$; $u_J^i := 0$; $\eta_{\text{alg}}^i := 10\text{tol}$;

while $\eta_{\text{alg}}^i \geq \text{tol}$ **do**

  $i := i + 1$; $u_J^i := u_J^{i-1}$;

  $u_J^i := u_J^i + \texttt{COARSE\_SOLVE}$; $\left(\eta_{\text{alg}}^i\right)^2 := \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla(\texttt{COARSE\_SOLVE}) \right\|^2$;

  for $j = 1, ..., J$ **do**

    for $\mathbf{a} \in \mathcal{V}_j$ **do**

      $\rho_{j,\mathbf{a}}^i := \texttt{LOCAL\_SOLVE}(j, \mathbf{a})$;

    end

    $\rho_j^i := \texttt{ADAPT\_SMOOTH}(j, \mathcal{V}_j)$; $\lambda_j^i := \texttt{OPTIMAL\_STEPSIZE}(\rho_j^i)$;

    $u_J^i := u_J^i + \lambda_j^i \rho_j^i$; $\left(\eta_{\text{alg}}^i\right)^2 := \left(\eta_{\text{alg}}^i\right)^2 + \left(\lambda_j^i \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_j^i \right\|\right)^2$;

  end

  if $\eta_{\text{alg}}^i < \text{tol}$ **break while loop**;

  $\left(\mathcal{M}, \{\mathbf{a} \in \mathcal{M}_j\}_{j \in \mathcal{M}}\right) := \texttt{DÖRFLER\_MARKING}\left(\rho_0^i, \{\{\rho_{j,\mathbf{a}}^i\}_{j=1}^J\}_{\mathbf{a} \in \mathcal{V}_j}, \theta\right)$;

  if [$\texttt{TEST\_ADAPT}(\gamma)$] **then**

    if $0 \in \mathcal{M}$ **then**

      $u_J^i := u_J^i + \texttt{COARSE\_SOLVE}$; $\left(\eta_{\text{alg}}^i\right)^2 := \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla(\texttt{COARSE\_SOLVE}) \right\|^2$;

    end

    for $j \in \mathcal{M} \setminus \{0\}$ **do**

      for $\mathbf{a} \in \mathcal{M}_j$ **do**

        $\rho_{j,\mathbf{a}}^i := \texttt{LOCAL\_SOLVE}(j, \mathbf{a})$;

      end

      $\rho_j^i := \texttt{ADAPT\_SMOOTH}(j, \mathcal{M}_j)$; $\lambda_j^i := \texttt{OPTIMAL\_STEPSIZE}(\rho_j^i)$;

      $u_J^i := u_J^i + \lambda_j^i \rho_j^i$; $\left(\eta_{\text{alg}}^i\right)^2 := \left(\eta_{\text{alg}}^i\right)^2 + \left(\lambda_j^i \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_j^i \right\|\right)^2$;

    end

  end

end

$i_{\text{stop}} := i$;

Output: [$u_J^{i_{\text{stop}}}$, $\eta_{\text{alg}}^{i_{\text{stop}}}$]

3.1.1 Module COARSE_SOLVE (coarse grid solution)

Input: - ; Output: global $\mathbb{P}_1$-lifting ρ_0^i of the current algebraic residual.

Given the latest approximation $u_J^i \in V_J^p$, define $\rho_0^i \in V_0^1$ by

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

3.1.2 Module LOCAL_SOLVE (block-Jacobi solution)

Input: level j, vertex $\mathbf{a}$; Output: local $\mathbb{P}_{p_j}$-lifting $\rho_{j,\mathbf{a}}^i$ of the current algebraic residual.

Given the latest approximation $u_J^i \in V_J^p$, define the local contribution $\rho_{j,\mathbf{a}}^i \in V_j^{\mathbf{a}}$ by

$$(\boldsymbol{\mathcal{K}}\nabla\rho_{j,\mathbf{a}}^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} = (f, v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} - (\boldsymbol{\mathcal{K}}\nabla u_J^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} \quad \forall v_{j,\mathbf{a}} \in V_j^{\mathbf{a}}.$$

3.1.3 Module ADAPT_SMOOTH (descent direction)

Input: level j, set of vertices $\mathcal{V}(j)$; Output: descent direction ρ_j^i.

The following test verifies if the weighted restricted additive Schwarz smoothing is compatible with the convergence analysis of the solver.

Given the latest approximation $u_J^i \in V_J^p$, if the following conditions hold

- $\displaystyle\sum_{\mathbf{a}\in\mathcal{V}(j)} \mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i) \neq 0,$
- $\displaystyle\left(\frac{\sum_{\mathbf{a}\in\mathcal{V}(j)} \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2}{d+1}\right)^{\frac{1}{2}} \leq \frac{\sum_{\mathbf{a}\in\mathcal{V}(j)} \left((f, \mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i))_{\omega_j^{\mathbf{a}}} - (\boldsymbol{\mathcal{K}}\nabla u_J^i, \nabla\mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i))_{\omega_j^{\mathbf{a}}}\right)}{\left\|\sum_{\mathbf{a}\in\mathcal{V}(j)} \boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i)\right\|},$

and, if the module is in the full-smoothing substep, if

- $\displaystyle\sum_{\mathbf{a}\in\mathcal{V}_j} \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i)\right\|_{\omega_j^{\mathbf{a}}}^2 \leq \sum_{\mathbf{a}\in\mathcal{V}_j} \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2,$

then the solver employs weighted restricted additive Schwarz smoothing, by defining the descent direction on level j, $\rho_j^i \in V_j^{p_j}$, as

$$\rho_j^i := \sum_{\mathbf{a}\in\mathcal{V}(j)} \mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i).$$

Otherwise, additive Schwarz smoothing is employed and

$$\rho_j^i := \sum_{\mathbf{a}\in\mathcal{V}(j)} \rho_{j,\mathbf{a}}^i.$$

3.1.4 Module OPTIMAL_STEPSIZE (optimal level step-size)

Input: descent direction ρ_j^i on level j; Output: optimal step-size λ_j^i on level j.

Given the latest approximation $u_J^i \in V_J^p$, if $\rho_j^i = 0$, set $\lambda_j^i := 1$, otherwise define the optimal step-size on level j, as

$$\lambda_j^i := \frac{(f, \rho_j^i) - (\boldsymbol{\mathcal{K}}\nabla u_J^i, \nabla\rho_j^i)}{\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|^2}.$$

3.1.5 Module DÖRFLER_MARKING (bulk choice of levels/patches for smoothing)

Input: liftings ρ_0^i, $\rho_{j,\mathbf{a}}^i$ for $1 \le j \le J$, $\mathbf{a} \in \mathcal{V}_j$, bulk-chasing parameter θ;
Output: set of marked levels $\mathcal{M}$, set of marked vertices per level $\mathcal{M}_j$, $j \in \mathcal{M}$.

For $\theta \in (0,1)$, we sort all patchwise contributions on all levels and select for marking the smallest cardinality set of the coarsest level and vertex indices, $1 \le j \le J$, by the following bulk-chasing criterion, cf. Dörfler [11],

$$\theta^2 \left(\left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_0^i \right\|^2 + \sum_{j=1}^{J} \lambda_j^i \sum_{\mathbf{a} \in \mathcal{V}_j} \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_{j,\mathbf{a}}^i \right\|_{\omega_j^{\mathbf{a}}}^2 \right) \le \sum_{j \in \mathcal{M}} \lambda_j^i \sum_{\mathbf{a} \in \mathcal{M}_j} \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_{j,\mathbf{a}}^i \right\|_{\omega_j^{\mathbf{a}}}^2,$$

where only $\left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_0^i \right\|$ appears on the coarsest level $j = 0$ if it is marked.

Here and below, we will always use the shorthand notation "$j \in \mathcal{M}$" for accessing the set $\mathcal{M}$ in ascending order.

3.1.6 Module TEST_ADAPT (deciding whether adaptivity will pay-off)

Input: User-prescribed parameter γ; Output: bool.

For $\gamma \in (0,1)$, if the following (analysis-driven) conditions hold, the solver will proceed to the adaptive-smoothing substep.

- $\displaystyle\sum_{j \in \mathcal{M}} \lambda_j^i \sum_{\mathbf{a} \in \mathcal{M}_j} \left(\sum_{k=j}^{J} \lambda_k^i \boldsymbol{\mathcal{K}} \nabla \rho_k^i, \nabla \rho_{j,\mathbf{a}}^i \right)_{\omega_j^{\mathbf{a}}} \le \gamma^2 \sum_{j \in \mathcal{M}} \lambda_j^i \sum_{\mathbf{a} \in \mathcal{M}_j} \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_{j,\mathbf{a}}^i \right\|_{\omega_j^{\mathbf{a}}}^2,$
- $\lambda_j^i \le 2(d+1) \quad \forall j \in \{0, \dots, J\}.$

In practice, one needs to verify mainly the first condition, whereas the second one seems much less restrictive.

3.2 Mathematical description of the solver

We now present the adaptive solver in a rigorous mathematical notation. This notation will be used for the remainder of the manuscript. Below we describe in detail one iteration of the adaptive solver. The initialization is given by $u_J^0 := 0 \in V_J^p$.

1. **Full-smoothing substep**

   (a) Define $\rho_0^i \in V_0^1$ by
   $$(\boldsymbol{\mathcal{K}} \nabla \rho_0^i, \nabla v_0) = (f, v_0) - (\boldsymbol{\mathcal{K}} \nabla u_J^i, \nabla v_0) \quad \forall v_0 \in V_0^1 \tag{3.1}$$
   and set $\lambda_0^i := 1$ and $u_{J,0}^i := u_J^i + \lambda_0^i \rho_0^i$.

   (b) For all $j \in \{1, \dots, J\}$, $\mathbf{a} \in \mathcal{V}_j$, define the local contributions $\rho_{j,\mathbf{a}}^i \in V_j^{\mathbf{a}}$ by
   $$(\boldsymbol{\mathcal{K}} \nabla \rho_{j,\mathbf{a}}^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} = (f, v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} - (\boldsymbol{\mathcal{K}} \nabla u_{J,j-1}^i, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} \quad \forall v_{j,\mathbf{a}} \in V_j^{\mathbf{a}}. \tag{3.2}$$

      i. **Test (adaptive smoothing choice)**: If the following conditions hold

      $$\sum_{\mathbf{a} \in \mathcal{V}_j} \mathcal{I}_j^{p_j} (\psi_j^{\mathbf{a}} \rho_{j,\mathbf{a}}^i) \neq 0, \tag{3.3a}$$

      $$\left(\frac{\sum_{\mathbf{a} \in \mathcal{V}_j} \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_{j,\mathbf{a}}^i \right\|_{\omega_j^{\mathbf{a}}}^2}{d+1} \right)^{\frac{1}{2}} \le \frac{\sum_{\mathbf{a} \in \mathcal{V}_j} \left[\left(f, \mathcal{I}_j^{p_j} (\psi_j^{\mathbf{a}} \rho_{j,\mathbf{a}}^i) \right)_{\omega_j^{\mathbf{a}}} - \left(\boldsymbol{\mathcal{K}} \nabla u_{J,j-1}^i, \nabla \mathcal{I}_j^{p_j} (\psi_j^{\mathbf{a}} \rho_{j,\mathbf{a}}^i) \right)_{\omega_j^{\mathbf{a}}} \right]}{\left\| \sum_{\mathbf{a} \in \mathcal{V}_j} \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \mathcal{I}_j^{p_j} (\psi_j^{\mathbf{a}} \rho_{j,\mathbf{a}}^i) \right\|}, \tag{3.3b}$$

      $$\sum_{\mathbf{a} \in \mathcal{V}_j} \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \mathcal{I}_j^{p_j} (\psi_j^{\mathbf{a}} \rho_{j,\mathbf{a}}^i) \right\|_{\omega_j^{\mathbf{a}}}^2 \le \sum_{\mathbf{a} \in \mathcal{V}_j} \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_{j,\mathbf{a}}^i \right\|_{\omega_j^{\mathbf{a}}}^2, \tag{3.3c}$$

then define the level j descent direction $\rho_j^i \in V_j^{p_j}$ as

$$\rho_j^i := \sum_{\mathbf{a}\in\mathcal{V}_j} \mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i), \tag{3.4}$$

otherwise define

$$\rho_j^i := \sum_{\mathbf{a}\in\mathcal{V}_j} \rho_{j,\mathbf{a}}^i. \tag{3.5}$$

If $\rho_j^i = 0$, set $\lambda_j^i := 1$, otherwise define the optimal step-size on level j

$$\lambda_j^i := \frac{(f,\rho_j^i) - (\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^i, \nabla\rho_j^i)}{\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|^2}. \tag{3.6}$$

The level update is given by

$$u_{J,j}^i := u_{J,j-1}^i + \lambda_j^i\rho_j^i, \tag{3.7}$$

and the update after the full-smoothing substep is $u_J^{i+\frac{1}{2}} := u_{J,J}^i \in V_J^p$.

2. **Marking** We mark the patches and/or the coarse level by the following bulk-chasing criterion [11], for a parameter $\theta \in (0,1)$

$$\theta^2\left(\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_0^i\right\|^2 + \sum_{j=1}^{J}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2\right) \le \sum_{j\in\mathcal{M}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2, \tag{3.8}$$

with the convention that if $0\in\mathcal{M}$, we write $\sum_{\mathbf{a}\in\mathcal{M}_0}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{0,\mathbf{a}}^i\right\|_{\omega_0^{\mathbf{a}}}^2$ to mean $\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_0^i\right\|^2$.

3. **Test (adaptive substep):** If the two following (analysis-driven) conditions are satisfied, proceed to the adaptive-smoothing substep:

$$\sum_{j\in\mathcal{M}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\Big(\sum_{k=j}^{J}\lambda_k^i\boldsymbol{\mathcal{K}}\nabla\rho_k^i, \nabla\rho_{j,\mathbf{a}}^i\Big)_{\omega_j^{\mathbf{a}}} \le \gamma^2\sum_{j\in\mathcal{M}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2, \tag{3.9}$$

$$\lambda_j^i \le 2(d+1) \quad \forall j\in\{0,\dots,J\}, \tag{3.10}$$

where $\gamma\in(0,1)$ is a user-prescribed parameter. If these conditions do not hold, then let $u_J^{i+1} := u_J^{i+\frac{1}{2}}$ and ignore the adaptive-smoothing substep.

4. **Adaptive-smoothing substep**

   (a) If $0\notin\mathcal{M}$, define $\rho_0^{i+\frac{1}{2}} := 0$ and $\lambda_0^{i+\frac{1}{2}} := 1$.

   Otherwise, when $0\in\mathcal{M}$, set $\lambda_0^{i+\frac{1}{2}} := 1$ and define $\rho_0^{i+\frac{1}{2}} \in V_0^1$ by

$$(\boldsymbol{\mathcal{K}}\nabla\rho_0^{i+\frac{1}{2}}, \nabla v_0) = (f, v_0) - (\boldsymbol{\mathcal{K}}\nabla u_J^{i+\frac{1}{2}}, \nabla v_0) \quad \forall v_0 \in V_0^1. \tag{3.11}$$

   Define the coarsest level update $u_{J,0}^{i+\frac{1}{2}} := u_J^{i+\frac{1}{2}} + \lambda_0^{i+\frac{1}{2}}\rho_0^{i+\frac{1}{2}}$.

   (b) Let $j\in\{1,\dots,J\}$. If j is not a marked level ($j\notin\mathcal{M}$), define $\rho_j^{i+\frac{1}{2}} := 0$, $\lambda_j^{i+\frac{1}{2}} := 1$, and $u_{J,j}^{i+\frac{1}{2}} := u_{J,j-1}^{i+\frac{1}{2}}$. Otherwise, when j is a marked level ($j\in\mathcal{M}$), define $\rho_{j,\mathbf{a}}^{i+\frac{1}{2}} \in V_j^{\mathbf{a}}$ for all marked vertices $\mathbf{a}\in\mathcal{M}_j$ by

$$(\boldsymbol{\mathcal{K}}\nabla\rho_{j,\mathbf{a}}^{i+\frac{1}{2}}, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} = (f, v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} - (\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^{i+\frac{1}{2}}, \nabla v_{j,\mathbf{a}})_{\omega_j^{\mathbf{a}}} \quad \forall v_{j,\mathbf{a}} \in V_j^{\mathbf{a}}. \tag{3.12}$$

i. **Test (adaptive smoothing choice)**: If the following conditions hold

$$\sum_{\mathbf{a}\in\mathcal{M}_j} \mathcal{I}_j^{p_j}\left(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^{i+\frac{1}{2}}\right) \neq 0, \tag{3.13a}$$

$$\left(\frac{\sum_{\mathbf{a}\in\mathcal{M}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^{i+\frac{1}{2}}\right\|_{\omega_j^{\mathbf{a}}}^2}{d+1}\right)^{\frac{1}{2}} \leq \frac{\sum_{\mathbf{a}\in\mathcal{M}_j}\left[\left(f,\mathcal{I}_j^{p_j}\left(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^{i+\frac{1}{2}}\right)\right)_{\omega_j^{\mathbf{a}}} - \left(\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^{i+\frac{1}{2}},\nabla\mathcal{I}_j^{p_j}\left(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^{i+\frac{1}{2}}\right)\right)_{\omega_j^{\mathbf{a}}}\right]}{\left\|\sum_{\mathbf{a}\in\mathcal{M}_j}\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\mathcal{I}_j^{p_j}\left(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^{i+\frac{1}{2}}\right)\right\|}, \tag{3.13b}$$

then define the level j descent direction $\rho_j^{i+\frac{1}{2}} \in V_j^{p_j}$ as

$$\rho_j^{i+\frac{1}{2}} := \sum_{\mathbf{a}\in\mathcal{M}_j} \mathcal{I}_j^{p_j}\left(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^{i+\frac{1}{2}}\right), \tag{3.14}$$

otherwise define

$$\rho_j^{i+\frac{1}{2}} := \sum_{\mathbf{a}\in\mathcal{M}_j} \rho_{j,\mathbf{a}}^{i+\frac{1}{2}}. \tag{3.15}$$

If $\rho_j^{i+\frac{1}{2}} = 0$, set $\lambda_j^{i+\frac{1}{2}} := 1$, otherwise define the optimal step-size on level j

$$\lambda_j^{i+\frac{1}{2}} := \frac{(f,\rho_j^{i+\frac{1}{2}}) - (\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^i,\nabla\rho_j^{i+\frac{1}{2}})}{\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^{i+\frac{1}{2}}\right\|^2}. \tag{3.16}$$

The level update is given by

$$u_{J,j}^{i+\frac{1}{2}} := u_{J,j-1}^{i+\frac{1}{2}} + \lambda_j^{i+\frac{1}{2}}\rho_j^{i+\frac{1}{2}}, \tag{3.17}$$

and the final update is $u_J^{i+1} := u_{J,J}^{i+\frac{1}{2}} \in V_J^p$.

Remark 3.1 (Compact writing of the iteration updates)**.** *Let $u_J^i \in V_J^p$. After the full-smoothing substep of the solver introduced above, we have*

$$u_J^{i+\frac{1}{2}} = u_J^i + \sum_{j=0}^{J}\lambda_j^i\rho_j^i, \tag{3.18}$$

and after the adaptive-smoothing substep we have

$$u_J^{i+1} = u_J^{i+\frac{1}{2}} + \sum_{j\in\mathcal{M}}\lambda_j^{i+\frac{1}{2}}\rho_j^{i+\frac{1}{2}}. \tag{3.19}$$

Analogously to [21, Theorem 4.6], due to the optimal step-sizes (3.6),(3.16), the error after each substep of the solver can be represented conveniently:

Lemma 3.2 (Error representation of each substep of the solver)**.** *For $u_J^i \in V_J^p$, let $u_J^{i+\frac{1}{2}} \in V_J^p$, $u_J^{i+1} \in V_J^p$ be constructed from u_J^i by the full-smoothing and the adaptive-smoothing substep of the solver of Section 3, respectively. Then*

$$\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\left(u_J - u_J^{i+\frac{1}{2}}\right)\right\|^2 = \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\left(u_J - u_J^i\right)\right\|^2 - \sum_{j=0}^{J}\left(\lambda_j^i\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|\right)^2, \tag{3.20}$$

$$\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\left(u_J - u_J^{i+1}\right)\right\|^2 = \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\left(u_J - u_J^{i+\frac{1}{2}}\right)\right\|^2 - \sum_{j\in\mathcal{M}}\left(\lambda_j^{i+\frac{1}{2}}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^{i+\frac{1}{2}}\right\|\right)^2. \tag{3.21}$$

4 A posteriori estimator on the algebraic error

The solver we introduced in Section 3 is inherently linked to an a posteriori estimator η_{alg}^i for the full-smoothing substep and $\eta_{\mathrm{alg}}^{i+\frac{1}{2}}$ for the adaptive-smoothing substep.

Definition 4.1 (Algebraic error estimator)**.** *Let $u_J^i \in V_J^p$ be arbitrary, let $u_J^{i+\frac{1}{2}} \in V_J^p$ be the update at the end of the full-smoothing substep, and let $u_J^{i+1} \in V_J^p$ be the update at the end of the adaptive substep. We define the algebraic error estimators*

$$\eta_{\mathrm{alg}}^i := \left(\sum_{j=0}^{J} \left(\lambda_j^i \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_j^i \right\| \right)^2 \right)^{\frac{1}{2}}, \tag{4.1}$$

$$\eta_{\mathrm{alg}}^{i+\frac{1}{2}} := \left(\sum_{j \in \mathcal{M}} \left(\lambda_j^{i+\frac{1}{2}} \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho_j^{i+\frac{1}{2}} \right\| \right)^2 \right)^{\frac{1}{2}}. \tag{4.2}$$

The following result is immediate from Lemma 3.2:

Lemma 4.2 (Guaranteed lower bound on the algebraic error per substep)**.** *Under the assumptions of Lemma 3.2 and Definition 4.1, the estimators are guaranteed lower bounds on the algebraic error for the respective substeps of the solver*

$$\left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \left(u_J - u_J^i \right) \right\| \geq \eta_{\mathrm{alg}}^i, \tag{4.3}$$

$$\left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \left(u_J - u_J^{i+\frac{1}{2}} \right) \right\| \geq \eta_{\mathrm{alg}}^{i+\frac{1}{2}}. \tag{4.4}$$

5 Main results

We present here our main result for the solver introduced in Section 3. Similarly to [20, 21], we show for each substep that the error contraction of the solver is equivalent to the efficiency of the associated a posteriori error estimator.

5.1 Mesh assumptions

For $j \in \{1, \dots, J\}$, we denote in the following $h_K := \mathrm{diam}(K)$ for $K \in \mathcal{T}_j$ and $h_j = \max_{K \in \mathcal{T}_j} h_K$. We shall always assume that our meshes are shape-regular:

Assumption 5.1 (Shape regularity)**.** *There exists $\kappa_{\mathcal{T}} > 0$ such that*

$$\max_{K \in \mathcal{T}_j} \frac{h_K}{\rho_K} \leq \kappa_{\mathcal{T}} \text{ for all } 0 \leq j \leq J, \tag{5.1}$$

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

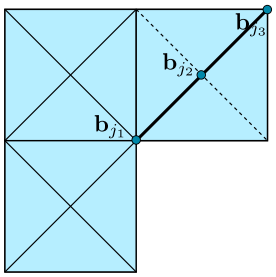
We mainly work with a hierarchy of quasi-uniform meshes with a bounded refinement factor between consecutive levels. This setting is described by:

Assumption 5.2 (Refinement strength and mesh quasi-uniformity)**.** *There exists $0 < C_{\mathrm{ref}} \leq 1$, a fixed positive real number such that for any $j \in \{1, \dots, J\}$, for all $K \in \mathcal{T}_{j-1}$, and for any $K^* \in \mathcal{T}_j$ such that $K^* \subset K$, there holds*

$$C_{\mathrm{ref}} h_K \leq h_{K^*} \leq h_K. \tag{5.2}$$

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

$$C_{\mathrm{qu}} h_j \leq h_K \leq h_j. \tag{5.3}$$



$\mathcal{T}_j$ 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 3: Illustration of the set $\mathcal{B}_j$; the refinement $\mathcal{T}_j$ (dotted lines) of mesh $\mathcal{T}_{j-1}$ (full lines).

The forthcoming main result also covers the setting of graded bisection grids, e.g. the newest vertex bisection, cf. Sewell [28], that we present here for completeness. In this case, one refinement of an edge of $\mathcal{T}_{j-1}$, for $j \in \{1, \dots, J\}$, gives us a new finer mesh $\mathcal{T}_j$. We denote by $\mathcal{B}_j \subset \mathcal{V}_j$ the set consisting of the new vertex obtained after the bisection together with its two neighbors on the refinement edge, cf. Figure 3 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:

Assumption 5.3 (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 $C_{\mathrm{loc,qu}}$ such that for any $j \in \{1, \dots, J\}$, there holds*

$$C_{\mathrm{loc,qu}} h_{\mathcal{B}_j} \leq h_K \leq h_{\mathcal{B}_j} \quad \forall K \in \mathcal{T}_j \text{ such that a vertex of } K \text{ belongs to } \mathcal{B}_j. \tag{5.4}$$

5.2 Main result

We now present the main result of this manuscript.

Theorem 5.4 (p-robust error contraction of the adaptive multilevel solver)**.** *Let Assumption 5.1 hold, together with either Assumption 5.2 or Assumption 5.3. Let $u_J \in V_J^p$ be the (unknown) solution of (2.3) and let $u_J^i \in V_J^p$ be arbitrary, $i \geq 0$. Let $u_J^{i+\frac{1}{2}} \in V_J^p$ be the update at the end of the full-smoothing substep of the solver described in Section 3. Then*

$$\left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \left(u_J - u_J^{i+\frac{1}{2}} \right) \right\| \leq \alpha \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \left(u_J - u_J^{i} \right) \right\|. \tag{5.5}$$

When tests (3.9)–(3.10) are satisfied, let $u_J^{i+1} \in V_J^p$ be the update at the end of the adaptive substep. Then

$$\left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \left(u_J - u_J^{i+1} \right) \right\| \leq \tilde{\alpha} \left\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \left(u_J - u_J^{i+\frac{1}{2}} \right) \right\|. \tag{5.6}$$

Here $0 < \alpha < 1$, $0 < \tilde{\alpha} < 1$ depend on the space dimension d, the mesh shape regularity parameter $\kappa_{\mathcal{T}}$, the number of mesh levels J, and the ratio of the largest and the smallest eigenvalues of the diffusion coefficient $\boldsymbol{\mathcal{K}}$, as well as on the mesh refinement parameter C_{ref} and quasi-uniformity parameter C_{qu} if Assumption 5.2 holds, or the coarse grid/local quasi-uniformity parameters C^0_{qu} and $C_{\mathrm{loc,qu}}$ if Assumption 5.3 holds. The dependence of the number of levels J is at most linear for α and cubic for $\tilde{\alpha}$. The factor $\tilde{\alpha}$ depends additionally on the marking parameter θ and the adaptivity tests parameter γ from (3.9).

Tests (3.9)–(3.10) are analysis-driven checks, that, if satisfied, ensure at the end of the full-smoothing substep, based on pre-computed quantities, that the adaptive-smoothing substep will also contract the error.

5.3 Additional results

There is a strong link between the solver defined in Section 3 and the a posteriori estimators defined in Section 4. Similarly to [20, 21], we have:

Theorem 5.5 (Equivalence estimator efficiency–solver contraction)**.** *Let the assumptions of Theorem 5.4 be satisfied. Then (5.5) holds if and only if*

$$\eta^i_{\mathrm{alg}} \geq \beta \big\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla (u_J - u^i_J) \big\| \tag{5.7}$$

holds with $\beta = \sqrt{1-\alpha^2}$*. Similarly, (5.6) holds if and only if*

$$\eta^{i+\frac{1}{2}}_{\mathrm{alg}} \geq \tilde{\beta} \big\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla (u_J - u_J^{i+\frac{1}{2}}) \big\| \tag{5.8}$$

holds with $\tilde{\beta} = \sqrt{1-\tilde{\alpha}^2}$*.*

The following result can be seen as the main motivation for our adaptive algorithm.

Corollary 5.6 (Equivalence error–estimator–localized contributions)**.** *Let the assumptions of Theorem 5.4 be satisfied. There holds*

$$\big\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla (u_J - u^i_J) \big\|^2 \approx \left(\eta^i_{\mathrm{alg}} \right)^2 \approx \big\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho^i_0 \big\|^2 + \sum_{j=1}^{J} \lambda^i_j \sum_{\mathbf{a} \in \mathcal{V}_j} \big\| \boldsymbol{\mathcal{K}}^{\frac{1}{2}} \nabla \rho^i_{j,\mathbf{a}} \big\|^2_{\omega^{\mathbf{a}}_j}, \tag{5.9}$$

where the constants involved in the equivalences "≈" have the same dependency as α *in (5.5), see (7.6) below for details.*

6 Numerical experiments

We consider four test cases: "Peak" (smooth solution with source term dominating in a part of a square domain), "L-shape" (problem with a singularity due to the L-shaped domain with a re-entrant corner), and "Skyscraper" (a problem we consider in two variants: with diffusion tensor having a jump of order 10^2 and 10^5), also described in detail in [21, Section 8]. We point out that test (3.10) is always satisfied in practice. In order to see numerical evidence of p-robustness, the stopping criterion is given by the relative residual dropping below 10^{-5}.

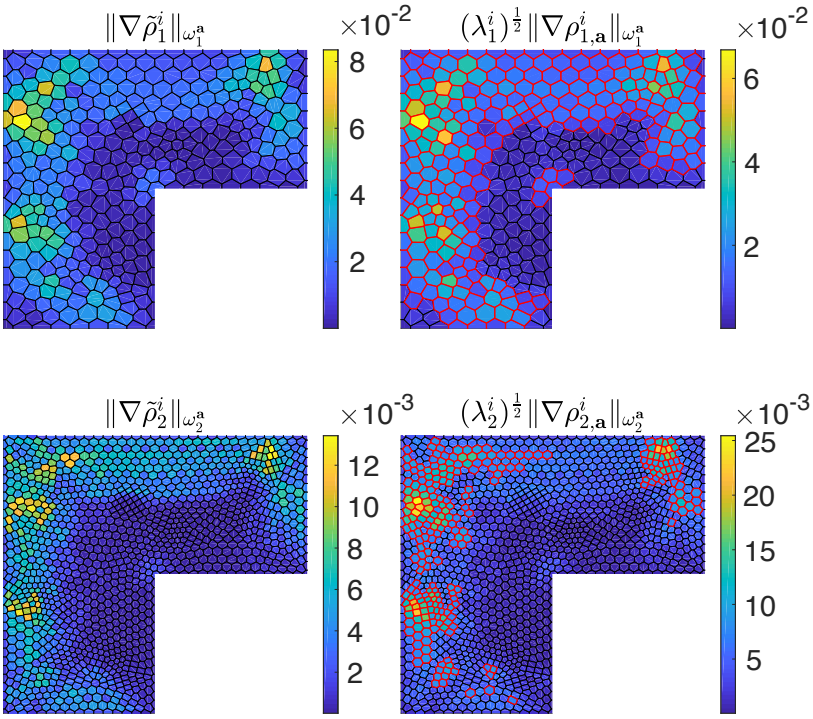


Figure 4: [L-shape, J=2, p_0=1, p_1=p_2=3, θ=0.95, γ=0.7] Comparing algebraic error distribution (left) to local error indicators (right) (levels $j=1$ top, $j=2$ bottom). Voronoi cells correspond to patch values, and the ones with the red border are marked for local smoothing.

6.1 Can we predict the distribution of the algebraic error?

We provide in Figures 4–5 an illustration on how the distribution of the algebraic error $\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\right\|$ is locally estimated using our algebraic error indicators. For this purpose, we consider the L-shape and Peak problems on a mesh hierarchy with $J = 2$ and $p_1 = p_2 = 3$, respectively $p_1 = p_2 = 6$ (recall that $p_0 = 1$ in our setting). In the figures, we compare, for a single iteration ($i = 3$ for L-shape, $i = 4$ for Peak), our algebraic error indicators $\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}$ with the local algebraic error distribution $\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\tilde{\rho}_j^i\|_{\omega_j^{\mathbf{a}}}$, where $\tilde{\rho}_j^i \in V_j^{p_j}$ is the levelwise orthogonal decomposition of the algebraic error with $\tilde{\rho}_0^i = \rho_0^i$ and, for $j \in \{1, \dots, J\}$,

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

see, e.g., [21, Section 3]. We highlight by a red border patches marked for local smoothing in the adaptive-smoothing substep, with the choice of the Dörfler marking parameter $\theta = 0.95$ in (3.8).

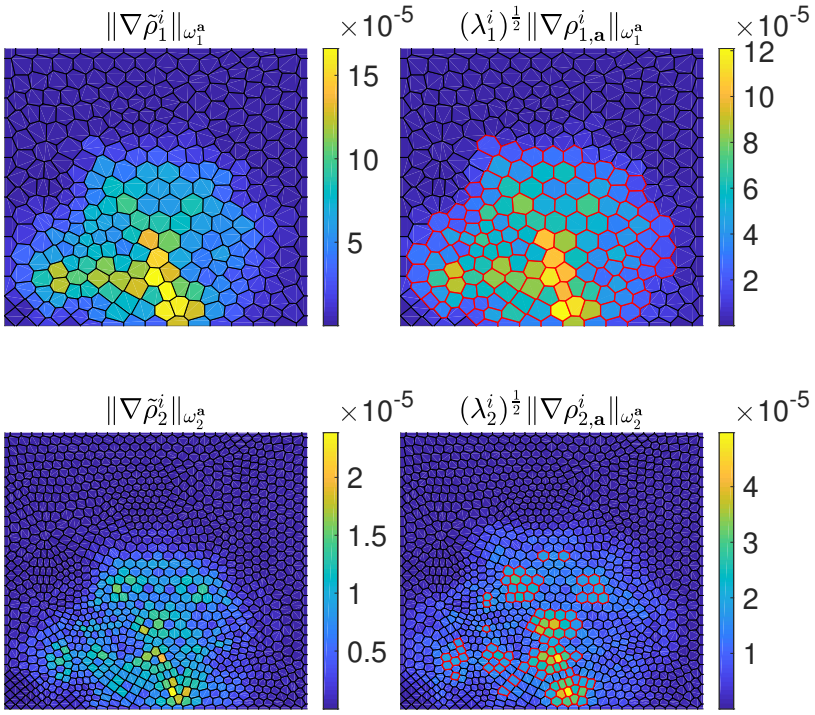


Figure 5: [Peak, J=2, p_0=1, p_1=p_2=6, θ=0.95, γ=0.7] Comparing algebraic error distribution (left) to local error indicators (right) (levels $j = 1$ top, $j = 2$ bottom). Voronoi cells correspond to patch values, and the ones with the red border are marked for local smoothing.

One can see that the local error indicators provide indeed a quite accurate information about the error distribution over the levels and patches in these tests. We note that one obtains similar results also for the other test cases, higher number of mesh levels J, different polynomial degrees, and different choices of the marking parameter θ. Thus the considered adaptivity indeed targets the problematic regions. It is important to note that the region with increased error could be dynamically changing from iteration to iteration. Our localized a posteriori estimator is designed in such a way that it will dynamically adjust to the new region with increased error. In all our experiments, the regions of increased algebraic error were rather stable, but we note that when periodic flipping occurred, the overall efficiency of the adaptive local smoothing Algorithm 1 may be spoiled.

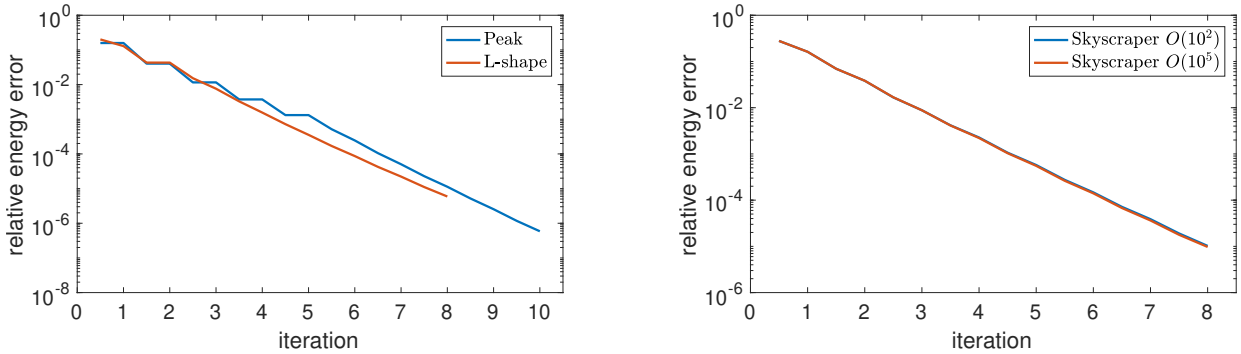


Figure 6: [All tests, $J=3$, $p_0=1$, $p_1=1$, $p_2=2$, $p_3=3$, $\theta=0.95$, $\gamma=0.7$] Convergence of the Algorithm 1 in the relative energy norm of the algebraic error $\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J - u_J^i)\|/\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla u_J\|$.

6.2 Does the adaptivity pay off?

Next, we investigate the performace of the adaptive Algorithm 1. We focus on convergence in the energy norm of the algebraic error during the iterations and the percentage of patches marked for local adaptive smoothing. For this purpose, we consider the four test cases and $J = 3$, $p_j = 1, 1, 2, 3$, $0 \le j \le J$, $\gamma = 0.7$, and the marking parameter θ fixed to 0.95; one obtains similar results also for other polynomial degrees. The results are summarized in Figure 6. One can see the decrease in each full-smoothing substep and that the adaptive substeps indeed also yield a decrease of the energy norm of the error; the adaptive-smoothing substeps actually yield nearly the same decrease as the full substeps – the convergence curve is nearly linear (in log scale) in the iterations where the adaptive smoothing is performed. Figures 7–8 then confirm that only a small portion of patches is marked for local adaptive smoothing, which suggest that Algorithm 1 may also be computationally beneficial.

Next, we test if the adaptive substeps provide a speed-up with respect to the variant without the adaptive substep. In Table 1, we compare, for varying polynomial degrees and number of levels, the results of Algorithm 1 when varying the parameter γ from test (3.9). We consider choices $\gamma = 0$, which corresponds to not using the adaptive substep at all, $\gamma = 0.7$, and, formally, $\gamma = \infty$, which stands for skipping the evaluation of (3.9), (3.10) and using the adaptive substep in every iteration. The latter choice is motivated by the fact that one would want to avoid evaluating the terms in test (3.9) if possible.

In Table 1, we in particular provide the number of iterations i with the number of adaptive-smoothing substeps in the brackets. For example "6(4)" means that the solver took 6 iterations to reach the stopping criterion, and the tests (3.9)–(3.10) were passed four times, i.e., 4 adaptive-smoothing substeps were performed in addition to the 6 full-smoothing substeps. For $p = 1$, test (3.9) is typically not verified, but otherwise Algorithm 1 with $\gamma = 0.7$ usually passes the adaptivity test (3.9) and leads to a reduction of the total number of iterations for the price of only employing a few local-adaptive-smoothing substeps. By always employing the adaptive substep ($\gamma = \infty$), we may cut the iteration count by nearly a half also for $p = 1$.

For comparison of the associated computational cost, we also provide, as in [21], an estimated number of floating point operations. This number is given by the formula

$$\begin{aligned}\text{nflops} := \frac{|\mathcal{V}_0|^3}{3} + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\frac{\text{ndof}(V_j^{\mathbf{a}})^3}{3} + \sum_{i=1}^{i_{\text{stop}}}\left[2\delta_0^i|\mathcal{V}_0|^2 + \sum_{j\in\mathcal{M}\setminus\{0\}}\sum_{\mathbf{a}\in\mathcal{M}_j} 2\text{ndof}(V_j^{\mathbf{a}})^2\right] \\ + \sum_{i=1}^{i_{\text{stop}}}\sum_{j=1}^{J}\left[2\,\text{nnz}(\mathcal{I}_{j-1}^{j}) + 2\,\text{nnz}(\mathcal{I}_{j}^{j-1}) + 2\,\text{nnz}(\mathbb{A}_j) + 3(2\,\text{size}(\mathbb{A}_j))\right].\end{aligned}$$

This formula is derived assuming 1) an initial Cholesky decomposition of local matrices associated to each patch on each level except for the coarsest one, where the global stiffness matrix for piecewise affine functions is factorized (for a matrix of size n, this cost is estimated as $1/3n^3$); 2) local solves by forward and backward substitutions (cost $2n^2$); 3) $\mathcal{I}_{j-1}^{j} : V_{j-1}^{p_{j-1}} \to V_j^{p_j}$ with the cost estimated by two-times the number of nonzeros of the associated interpolation matrix; and 4) evaluation of the optimal step-sizes λ_j as

Peak test case

J	p_j	$\gamma=0$ niter	$\gamma=0$ nflops	$\gamma=0.7$ niter	$\gamma=0.7$ nflops	$\gamma=\infty$ niter	$\gamma=\infty$ nflops
3	1 1 1 1	19(0)	2.11×10^7	19(0)	2.11×10^7	11(11)	2.22×10^7
	1 1 2 3	15(0)	4.26×10^8	10(5)	3.70×10^8	8(8)	3.63×10^8
	1 2 4 6	12(0)	8.81×10^9	9(4)	8.15×10^9	7(7)	7.74×10^9
	1 3 6 9	13(0)	8.17×10^{10}	9(7)	7.69×10^{10}	8(8)	7.54×10^{10}
4	1 1 1 1 1	20(0)	7.17×10^7	20(0)	7.17×10^7	12(12)	8.20×10^7
	1 1 2 2 3	13(0)	1.51×10^9	10(4)	1.43×10^9	8(8)	1.46×10^9
	1 2 3 5 6	11(0)	3.78×10^{10}	9(4)	3.68×10^{10}	7(7)	3.52×10^{10}
	1 3 5 7 9	13(0)	3.46×10^{11}	9(7)	3.28×10^{11}	8(8)	3.21×10^{11}

L-shape test case

J	p_j	$\gamma=0$ niter	$\gamma=0$ nflops	$\gamma=0.7$ niter	$\gamma=0.7$ nflops	$\gamma=\infty$ niter	$\gamma=\infty$ nflops
3	1 1 1 1	21(0)	2.17×10^7	21(0)	2.17×10^7	11(11)	2.11×10^7
	1 1 2 3	13(0)	3.63×10^8	8(7)	3.43×10^8	7(7)	3.19×10^8
	1 2 4 6	8(0)	7.02×10^9	5(5)	6.50×10^9	5(5)	6.50×10^9
	1 3 6 9	8(0)	6.94×10^{10}	5(5)	6.59×10^{10}	5(5)	6.59×10^{10}
4	1 1 1 1 1	21(0)	7.24×10^7	21(0)	7.24×10^7	11(11)	7.29×10^7
	1 1 2 2 3	9(0)	1.06×10^9	8(5)	1.24×10^9	6(6)	1.10×10^9
	1 2 3 5 6	7(0)	2.95×10^{10}	5(5)	2.92×10^{10}	5(5)	2.92×10^{10}
	1 3 5 7 9	6(0)	2.75×10^{11}	5(5)	2.78×10^{11}	5(5)	2.78×10^{11}

Skyscraper test case diff. contrast $O(10^2)$

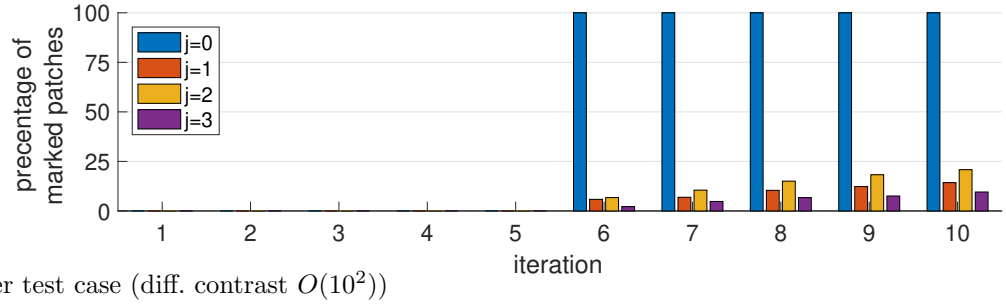
J	p_j	$\gamma=0$ niter	$\gamma=0$ nflops	$\gamma=0.7$ niter	$\gamma=0.7$ nflops	$\gamma=\infty$ niter	$\gamma=\infty$ nflops
3	1 1 1 1	19(0)	1.90×10^7	19(0)	1.90×10^7	12(12)	2.18×10^7
	1 1 2 3	15(0)	4.10×10^8	8(8)	3.50×10^8	8(8)	3.50×10^8
	1 2 4 6	9(0)	7.36×10^9	6(6)	6.94×10^9	6(6)	6.94×10^9
	1 3 6 9	9(0)	7.11×10^{10}	6(6)	6.80×10^{10}	6(6)	6.80×10^{10}
4	1 1 1 1 1	19(0)	6.31×10^7	19(0)	6.31×10^7	12(12)	7.61×10^7
	1 1 2 2 3	11(0)	1.26×10^9	8(7)	1.35×10^9	7(7)	1.25×10^9
	1 2 3 5 6	8(0)	3.11×10^{10}	6(6)	3.15×10^{10}	6(6)	3.15×10^{10}
	1 3 5 7 9	8(0)	2.91×10^{11}	5(5)	2.77×10^{11}	5(5)	2.77×10^{11}

Skyscraper test case diff. contrast $O(10^5)$

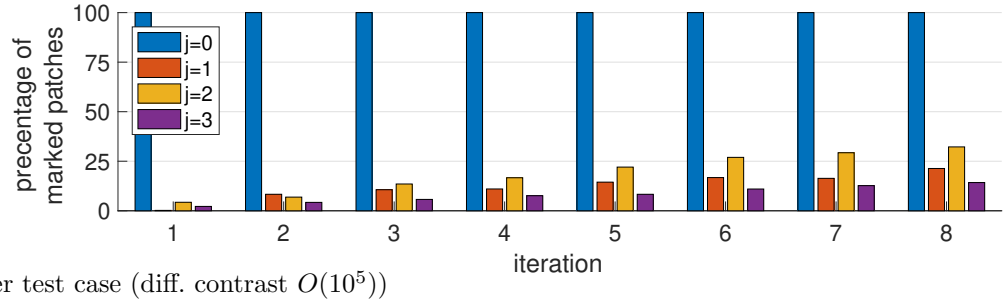
J	p_j	$\gamma=0$ niter	$\gamma=0$ nflops	$\gamma=0.7$ niter	$\gamma=0.7$ nflops	$\gamma=\infty$ niter	$\gamma=\infty$ nflops
3	1 1 1 1	19(0)	1.90×10^7	19(0)	1.90×10^7	13(13)	2.33×10^7
	1 1 2 3	15(0)	4.10×10^8	8(8)	3.48×10^8	8(8)	3.48×10^8
	1 2 4 6	9(0)	7.36×10^9	6(6)	6.93×10^9	6(6)	6.93×10^9
	1 3 6 9	9(0)	7.11×10^{10}	6(6)	6.79×10^{10}	6(6)	6.79×10^{10}
4	1 1 1 1 1	19(0)	6.31×10^7	19(0)	6.31×10^7	12(12)	7.60×10^7
	1 1 2 2 3	11(0)	1.26×10^9	8(7)	1.35×10^9	7(7)	1.25×10^9
	1 2 3 5 6	8(0)	3.11×10^{10}	6(6)	3.15×10^{10}	6(6)	3.15×10^{10}
	1 3 5 7 9	8(0)	2.91×10^{11}	5(5)	2.77×10^{11}	5(5)	2.77×10^{11}

Table 1: Number of iterations (number of adaptive-smoothing substeps in brackets) for various choices of the parameter γ in (3.9). The marking parameter in (3.8) is set as $\theta = 0.95$

Peak test case



Skyscraper test case (diff. contrast $O(10^2)$)



Skyscraper test case (diff. contrast $O(10^5)$)

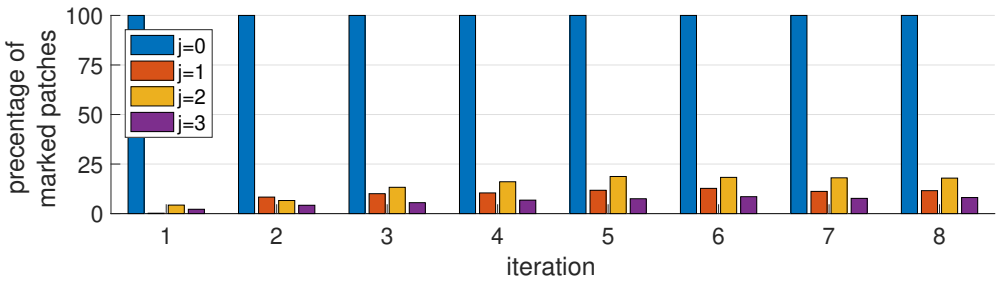


Figure 7: [Different tests, $J = 3$, $p_0 = 1$, $p_1 = 1$, $p_2 = 2$, $p_3 = 3$, $\theta = 0.95$, $\gamma = 0.7$] Local adaptive smoothing: coarsest level marked or not and percentages of patches marked for each level $1 \le j \le J$ (Y-axis). Iterations of Algorithm 1 (X-axis). Results for the L-shape test case are given in the separate Figure 8.

in formulas (3.6), (3.16) involving multiplication with the stiffness matrix $\mathbb{A}_j$ on the given level (cost equal to two-times the number of nonzeros) and three inner products. From the above tests, we see that adaptivity is of interest. Not only does it provide error contraction on the adaptive substep of almost the same quality as the full-smoothing substep with just local smoothing in a relatively small percentage of marked patches, cf. Figures 6–8, but in numerous cases, the adaptive variant is cheaper than the non-adaptive one in terms of the nflops formula. Note that the nflops only represent *one* way of estimating the costs and the interest in adaptivity is not solely determined by it.

6.3 Dependence on the marking parameter

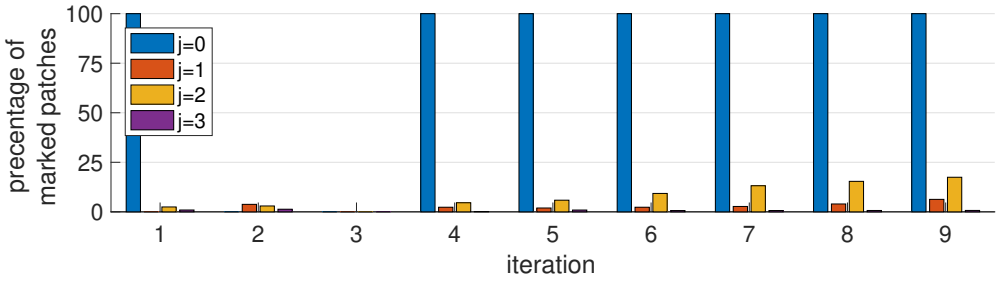
We finally vary the Dörfler marking parameter θ from (3.8), setting $\theta = 0.7, 0.9, 0.95, 0.99$. The results are given in Figure 9 and in Table 2, where we consider $\gamma = 0.7$.

One can see that the choice $\theta = 0.7$ is often not sufficiently efficient. For this choice, the number of iterations is not reduced sufficiently and the cost of intergrid operation then dominates over the cost of local smoothings. The best choice of θ seems to differ, but $\theta = 0.95$ reveals quite satisfactory in most of the cases.

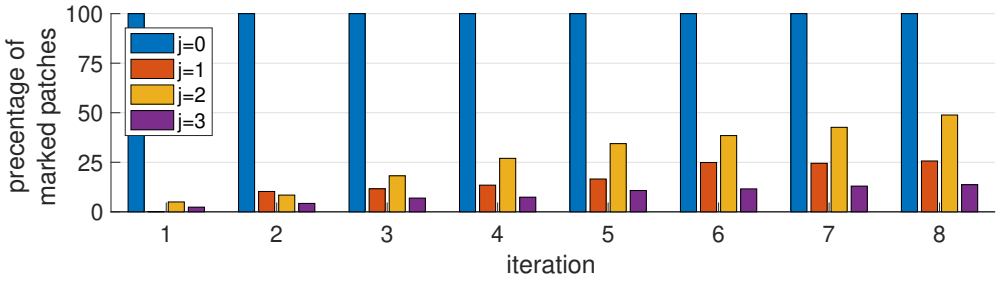
Remark 6.1 (Dependence on the shape regularity parameter)**.** *We would like to point out how the performance of the solver depends on the parameters of the Assumptions 5.1–5.3. As an example, we present in Table 3 the number of iterations required when the shape regularity parameter $\kappa_{\mathcal{T}}$ degrades. One can see an overall degradation, but the polynomial degree robustness is preserved as expected.*

L-shape test case

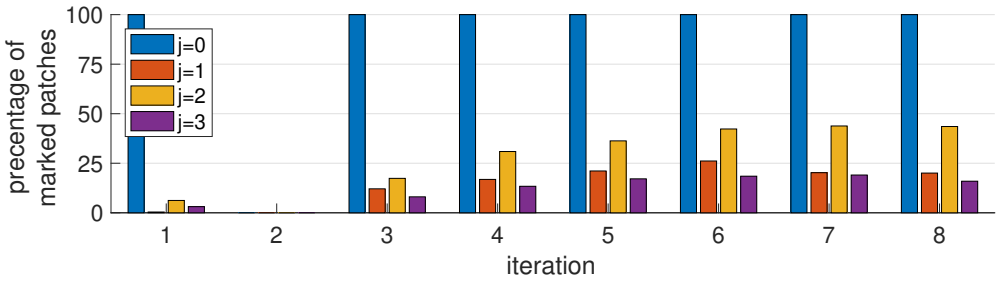
$\theta = 0.7$



$\theta = 0.9$



$\theta = 0.95$



$\theta = 0.99$

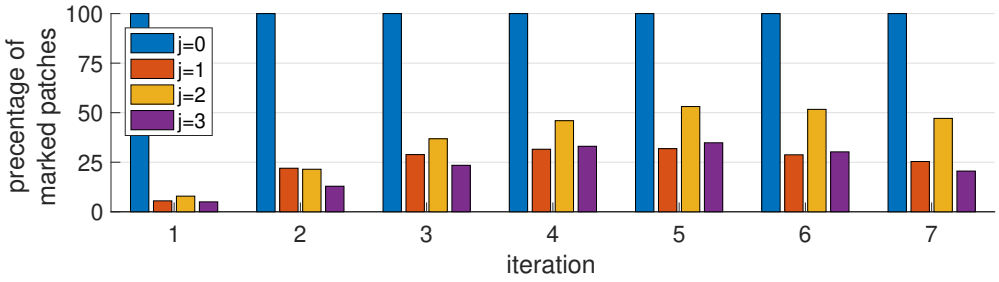


Figure 8: [L-shape, $J = 3$, $p_0 = 1$, $p_1 = 1$, $p_2 = 2$, $p_3 = 3$, $\gamma = 0.7$, varying θ] Local adaptive smoothing: coarsest level marked or not and percentages of patches marked for each level $1 \leq j \leq J$ (Y-axis). Iterations of Algorithm 1 (X-axis).

Peak test case

J	p_j	$\theta = 0.7$ niter	$\theta = 0.7$ nflops	$\theta = 0.9$ niter	$\theta = 0.9$ nflops	$\theta = 0.95$ niter	$\theta = 0.95$ nflops	$\theta = 0.99$ niter	$\theta = 0.99$ nflops
4	1 1 1 1 1	20(0)	7.17×10^7	20(0)	7.17×10^7	20(0)	7.17×10^7	20(0)	7.17×10^7
	1 1 2 2 3	12(2)	1.52×10^9	11(3)	1.47×10^9	10(4)	1.43×10^9	10(4)	1.44×10^9
	1 2 3 5 6	11(0)	3.78×10^{10}	10(3)	3.80×10^{10}	9(4)	3.68×10^{10}	8(4)	3.52×10^{10}
	1 3 5 7 9	12(8)	3.57×10^{11}	10(8)	3.39×10^{11}	9(7)	3.28×10^{11}	8(6)	3.17×10^{11}

L-shape test case

J	p_j	$\theta = 0.7$ niter	$\theta = 0.7$ nflops	$\theta = 0.9$ niter	$\theta = 0.9$ nflops	$\theta = 0.95$ niter	$\theta = 0.95$ nflops	$\theta = 0.99$ niter	$\theta = 0.99$ nflops
4	1 1 1 1 1	21(0)	7.24×10^7	21(0)	7.24×10^7	21(0)	7.24×10^7	21(0)	7.24×10^7
	1 1 2 2 3	9(4)	1.28×10^9	8(5)	1.24×10^9	8(5)	1.24×10^9	6(5)	1.06×10^9
	1 2 3 5 6	6(3)	2.97×10^{10}	6(4)	3.03×10^{10}	5(5)	2.92×10^{10}	4(4)	2.70×10^{10}
	1 3 5 7 9	6(6)	2.90×10^{11}	5(5)	2.78×10^{11}	5(5)	2.78×10^{11}	4 (4)	2.68×10^{11}

Skyscraper test case (diff. contrast $O(10^2)$)

J	p_j	$\theta = 0.7$ niter	$\theta = 0.7$ nflops	$\theta = 0.9$ niter	$\theta = 0.9$ nflops	$\theta = 0.95$ niter	$\theta = 0.95$ nflops	$\theta = 0.99$ niter	$\theta = 0.99$ nflops
4	1 1 1 1 1	19(0)	6.31×10^7	19(0)	6.31×10^7	19(0)	6.31×10^7	19(0)	6.31×10^7
	1 1 2 2 3	10(4)	1.38×10^9	8(7)	1.34×10^9	8(7)	1.35×10^9	6(6)	1.10×10^9
	1 2 3 5 6	8(4)	3.38×10^{10}	6(6)	3.15×10^{10}	6(6)	3.15×10^{10}	5(5)	2.92×10^{10}
	1 3 5 7 9	7(7)	2.99×10^{11}	6(6)	2.88×10^{11}	5(5)	2.77×10^{11}	5(5)	2.77×10^{11}

Skyscraper test case (diff. contrast $O(10^5)$)

J	p_j	$\theta = 0.7$ niter	$\theta = 0.7$ nflops	$\theta = 0.9$ niter	$\theta = 0.9$ nflops	$\theta = 0.95$ niter	$\theta = 0.95$ nflops	$\theta = 0.99$ niter	$\theta = 0.99$ nflops
4	1 1 1 1 1	19(0)	6.31×10^7	19(0)	6.31×10^7	19(0)	6.31×10^7	19(0)	6.31×10^7
	1 1 2 2 3	11(5)	1.53×10^9	8(7)	1.34×10^9	8(7)	1.35×10^9	7(7)	1.26×10^9
	1 2 3 5 6	8(4)	3.38×10^{10}	6(6)	3.15×10^{10}	6(6)	3.15×10^{10}	5(5)	2.91×10^{10}
	1 3 5 7 9	7(7)	2.99×10^{11}	6(6)	2.88×10^{11}	5(5)	2.77×10^{11}	5(5)	2.77×10^{11}

Table 2: Number of iterations (number of adaptive-smoothing substeps in brackets) for various choices of marking parameter θ in (3.8). The parameter γ from (3.9) is set as $\gamma = 0.7$

p_j	DoF	minimal angle: 32.1° niter	minimal angle: 21.4° niter	minimal angle: 12.0° niter
1 1 2 3	1e5	8(7)	9(9)	17(17)
1 2 4 6	6e5	5(5)	6(6)	11(11)
1 3 6 9	1e6	5(5)	6(6)	10(10)

Table 3: [L-shape, J=3, θ=0.95, γ=0.7] Study of sensitivity with respect to the shape regularity of the mesh (minimal angle of mesh elements) for the local adaptive smoothing solver.

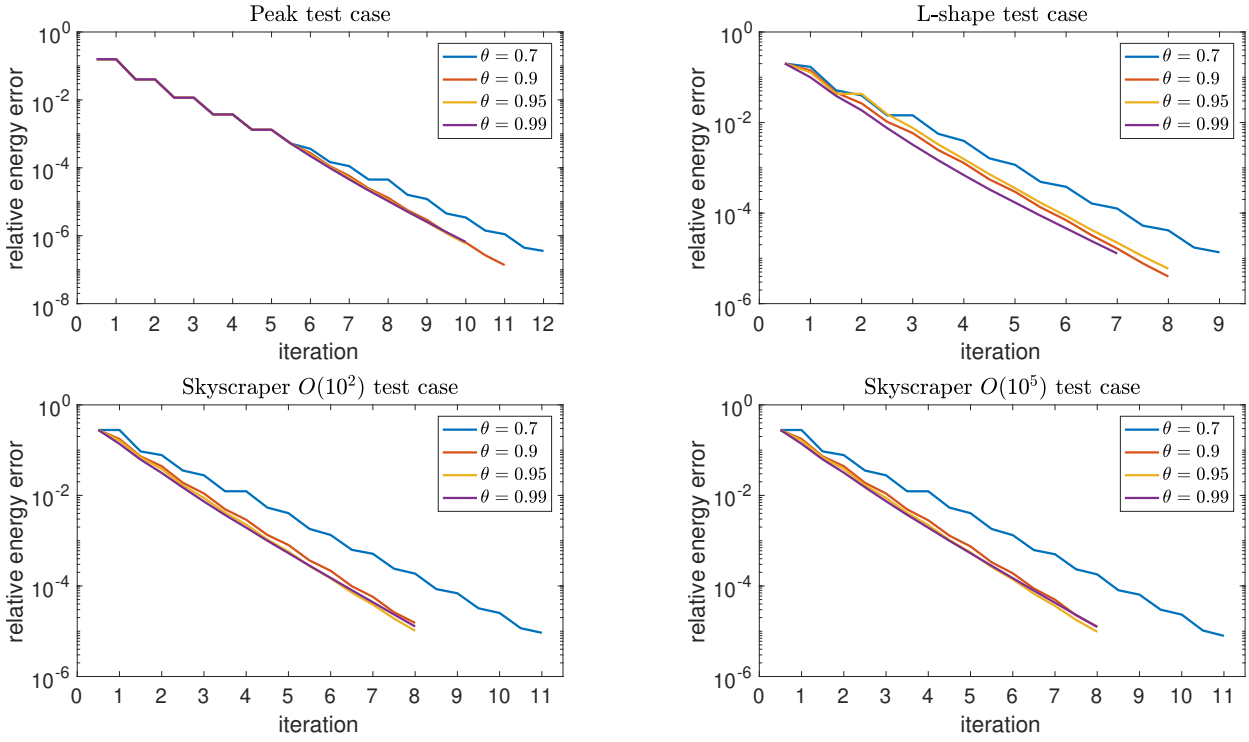


Figure 9: [All tests, $J=3$, $p_0=1$, $p_1=1$, $p_2=2$, $p_3=3$, $\gamma=0.7$, varying θ] Convergence of Algorithm 1 in the relative energy norm of the algebraic error $\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla(u_J-u_J^i)\|/\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla u_J\|$.

7 Proofs of the main results

In this section we present the proofs of the results stated in Section 5. We start with noting that Theorem 5.5 can be proven exactly along the lines of [21, Corollary 6.7].

7.1 Proof of contraction: full-smoothing substep

We start with a generalization of the properties given in [21] covering the test (3.3), in order to extend the results to the case of weighted restricted additive Schwarz smoothing.

Lemma 7.1 (Lower bound on levelwise updates by patchwise contributions)**.** *Let $u_J^i \in V_J^p$ be arbitrary. Let $j \in \{1,\dots,J\}$ and let ρ_j^i, λ_j^i be constructed from u_J^i by the full-smoothing substep of the solver described in Section 3. Then*

$$\sum_{\mathbf{a}\in\mathcal{V}_j} \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2 \leq (d+1)\left(\lambda_j^i\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|\right)^2 \quad \forall 1\leq j\leq J, \tag{7.1}$$

where for each vertex $\mathbf{a}\in\mathcal{V}_j$, $\rho_{j,\mathbf{a}}^i$ is the solution of the local problem (3.2).

Proof. Depending if test (3.3) of the solver in Section 3 is satisfied or not, ρ_j^i will be constructed differently. We show that (7.1) holds for either outcome of test (3.3).

Case test (3.3) is satisfied: Then ρ_j^i is constructed by (3.4) and the outcome of Test (3.3a),(3.3b) ensures

on the one hand that $\rho_j^i \neq 0$ and on the other hand that

$$\left(\frac{\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2}{d+1}\right)^{\frac{1}{2}} \leq \frac{(f,\rho_j^i)-(\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^i,\nabla\rho_j^i)}{\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|}.$$

Using (3.6), this leads to: $\left(\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2\right)^{\frac{1}{2}} \leq \sqrt{d+1}\lambda_j^i\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|$.

Case test (3.3) *is not satisfied:* Then ρ_j^i is constructed by (3.5). First, note that

$$\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2 \overset{(3.2),(3.5)}{=} (f,\rho_j^i)-(\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^i,\nabla\rho_j^i). \tag{7.2}$$

Thus, if $\rho_j^i = 0$, then the result (7.1) holds trivially. To treat the remaining case $\rho_j^i \neq 0$, we use the expression of λ_j^i together with [21, Lemma 9.1] to obtain

$$\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2 \overset{(3.6)}{=} \lambda_j^i\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|^2 \leq \lambda_j^i\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|\left((d+1)\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2\right)^{\frac{1}{2}}.$$

□

The second important property we will need is given below.

Lemma 7.2 (Upper bound on levelwise updates by patchwise contributions)**.** *Let $u_J^i \in V_J^p$ be arbitrary. Let $j \in \{1,\dots,J\}$ and let ρ_j^i, λ_j^i be constructed from u_J^i by the full-smoothing substep of the solver described in Section 3. Then*

$$\left(\lambda_j^i\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|\right)^2 \leq \lambda_j^i\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2 \quad \forall 1\leq j\leq J, \tag{7.3}$$

where for each vertex $\mathbf{a}\in\mathcal{V}_j$, $\rho_{j,\mathbf{a}}^i$ is the solution of the local problem (3.2).

Proof. We only need to show (7.3) when $\rho_j^i \neq 0$, otherwise the result is trivial.

Case test (3.3) *is satisfied:* Then ρ_j^i is constructed by (3.4) and by using a Young inequality together with test (3.3c), we obtain

$$\begin{aligned}(f,\rho_j^i)-(\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^i,\nabla\rho_j^i) &= \sum_{\mathbf{a}\in\mathcal{V}_j}\left(\left(f,\mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i)\right)_{\omega_j^{\mathbf{a}}}-\left(\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^i,\nabla\mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i)\right)_{\omega_j^{\mathbf{a}}}\right)\\ &\overset{(3.2)}{=}\sum_{\mathbf{a}\in\mathcal{V}_j}\left(\boldsymbol{\mathcal{K}}\nabla\rho_{j,\mathbf{a}}^i,\nabla\mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i)\right)_{\omega_j^{\mathbf{a}}} \leq \sum_{\mathbf{a}\in\mathcal{V}_j}\frac{1}{2}\left(\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2+\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\mathcal{I}_j^{p_j}(\psi_j^{\mathbf{a}}\rho_{j,\mathbf{a}}^i)\right\|_{\omega_j^{\mathbf{a}}}^2\right)\\ &\overset{(3.3c)}{\leq}\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2\end{aligned}$$

Case test (3.3) *is not satisfied:* The above estimate is in fact an equality, by (7.2).

As we see, for both possible outcomes of test (3.3), we obtain the desired result

$$\left(\lambda_j^i\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|\right)^2 \overset{(3.6)}{=} \lambda_j^i\frac{(f,\rho_j^i)-(\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^i,\nabla\rho_j^i)}{\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|^2}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_j^i\right\|^2 \leq \lambda_j^i\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\right\|_{\omega_j^{\mathbf{a}}}^2.$$

□

Remark 7.3 (Lower bound on the optimal step-sizes)**.** *As in* [21, Remark 9.2]*, by putting together the results of Lemmas 7.1, 7.2, and since $\lambda_j^i = 1$ when $\rho_j^i = 0$ or $j = 0$, we have*

$$\lambda_j^i \geq \frac{1}{d+1} \quad 0\leq j\leq J. \tag{7.4}$$

We can now present the proof of contraction of the solver for the full-smoothing substep. The proof follows as the proof of [21, Theorem 6.6].

Proof of part 1 of Theorem 5.4. Even though the results in [21] are given for the case of additive Schwarz smoothing only, we will use here the three main estimates established in the proof of [21, Theorem 6.6] under minimal H^1-regularity. This is possible because the estimates only use the levelwise and patchwise contributions $\rho^i_{j,\mathbf{a}}$ which are constructed in the same way here, allowing us to extend the proof for case of the weighted restricted additive Schwarz smoothing. This yields $C_{\mathrm{S},1} := \sqrt{2(d+1)}C_{\mathrm{S},\boldsymbol{\mathcal{K}}}J$, $C_{\mathrm{S},2} := \sqrt{2(d+1)}C_{\mathrm{S},\boldsymbol{\mathcal{K}}}$, for $C_{\mathrm{S},\boldsymbol{\mathcal{K}}} \geq 1$ of [21] having the same dependencies as α, such that

$$\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\left(u_J - u^i_J\right)\right\|^2 \leq C^2_{\mathrm{S},1}\left(\eta^i_{\mathrm{alg}}\right)^2 + C^2_{\mathrm{S},2}\sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^i_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j} \overset{(7.1)}{\leq} C^2_{\mathrm{S}}\left(\eta^i_{\mathrm{alg}}\right)^2, \tag{7.5}$$

with $C^2_{\mathrm{S}} := 2\max(C_{\mathrm{S},1}, (d+1)C_{\mathrm{S},2})$.

By Theorem 5.5, this is equivalent to (5.5) with $\alpha = \sqrt{1 - C^2_{\mathrm{S}}}$. □

Proof of Corollary 5.6. First, note that this result extends [21, Corollary 6.8] to the weighted restricted additive Schwarz smoothing case. In the case when additive Schwarz smoothing is employed, the second equivalence in (5.9) is in fact an equality as given in [21]. We obtain the desired equivalences by using Lemmas 7.1, 7.2 together with Remark 7.3, (3.10) and $\lambda^i_0 = 1$ in a closed chain of estimates

$$\begin{aligned}
\left(\eta^i_{\mathrm{alg}}\right)^2 &\overset{(4.3)}{\leq} \left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\left(u_J - u^i_J\right)\right\|^2 \\
&\overset{\substack{(7.5)\\(7.3)}}{\leq} C^2_{\mathrm{S}}\left(\left(\lambda^i_0\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^i_0\right\|\right)^2 + \sum_{j=1}^{J}\lambda^i_j\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^i_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j}\right) \\
&\overset{(3.10)}{\leq} 2(d+1)C^2_{\mathrm{S}}\left(\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^i_0\right\|^2 + \sum_{j=1}^{J}\sum_{\mathbf{a}\in\mathcal{V}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^i_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j}\right) \\
&\overset{(7.1)}{\leq} 2(d+1)^2C^2_{\mathrm{S}}\left(\eta^i_{\mathrm{alg}}\right)^2.
\end{aligned} \tag{7.6}$$

□

7.2 Proof of contraction: adaptive-smoothing substep

Let tests (3.9)–(3.10) be satisfied. We introduce the notation $\delta_j = 1$ if the level j is marked (when $j \in \mathcal{M}$), otherwise $\delta_j = 0$. Firstly, we present the generalization of Lemma 7.1, obtained by only working with the marked vertices.

Lemma 7.4 (Lower bound on levelwise updates by patchwise contributions)**.** *Let $u^i_J \in V^p_J$ be arbitrary. Let $j \in \mathcal{M}\setminus\{0\}$, and $\rho^{i+\frac{1}{2}}_j$, $\lambda^{i+\frac{1}{2}}_j$ be constructed from $u^{i+\frac{1}{2}}_J$ by the adaptive-smoothing substep of the solver described in Section 3. There holds*

$$\sum_{\mathbf{a}\in\mathcal{M}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^{i+\frac{1}{2}}_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j} \leq (d+1)\left(\lambda^{i+\frac{1}{2}}_j\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^{i+\frac{1}{2}}_j\right\|\right)^2 \quad \forall 1 \leq j \leq J, \tag{7.7}$$

where for each vertex $\mathbf{a} \in \mathcal{V}_j$, $\rho^{i+\frac{1}{2}}_{j,\mathbf{a}}$ is the solution of the local problem (3.12).

Summing over all mesh levels and since $d+1 \geq 1$ (on $j=0$), (7.7) gives:

Corollary 7.5 (Lower bound on the estimator by localized contributions)**.** *There holds*

$$\sum_{j\in\mathcal{M}}\sum_{\mathbf{a}\in\mathcal{M}_j}\left\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho^{i+\frac{1}{2}}_{j,\mathbf{a}}\right\|^2_{\omega^{\mathbf{a}}_j} \leq (d+1)\left(\eta^{i+\frac{1}{2}}_{\mathrm{alg}}\right)^2. \tag{7.8}$$

The following result is crucial in the proof of contraction of the adaptive-smoothing substep. Since the marking takes place at the end of the full-smoothing substep, which determines where the adaptive-smoothing takes place, a connection between the two substeps is needed. This is the goal of the tests (3.9)–(3.10).

Lemma 7.6 (Link between full- and adaptive-smoothing substeps)**.** *Under the adaptivity tests* (3.9)–(3.10)*, there holds*

$$\sum_{j\in\mathcal{M}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\big\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\big\|_{\omega_j^{\mathbf{a}}}^2\le\frac{4(d+1)^2(|\mathcal{M}|^2+1)}{(1-\gamma^2)^2}\Big(\eta_{\mathrm{alg}}^{i+\frac12}\Big)^2. \tag{7.9}$$

Proof. We first make the connection between the two substeps, then we arrange together the terms given by the adaptive substep. The remaining full-smoothing substep terms are then treated by (3.9) and finally, we apply a Young's inequality. The main term we want to estimate can be split in the two quantities below

$$\sum_{j\in\mathcal{M}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\big\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\big\|_{\omega_j^{\mathbf{a}}}^2=\delta_0(\boldsymbol{\mathcal{K}}\nabla\rho_0^i,\nabla\rho_0^i)+\sum_{j\in\mathcal{M}\setminus\{0\}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}(\boldsymbol{\mathcal{K}}\nabla\rho_{j,\mathbf{a}}^i,\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}.$$

First,

$$\begin{aligned}
\delta_0(\boldsymbol{\mathcal{K}}\nabla\rho_0^i,\nabla\rho_0^i)&\overset{(3.1),(3.18)}{=}\delta_0\Big((f,\rho_0^i)-(\boldsymbol{\mathcal{K}}\nabla u_J^{i+\frac12},\nabla\rho_0^i)+\sum_{j=0}^J\lambda_j^i(\boldsymbol{\mathcal{K}}\nabla\rho_j^i,\nabla\rho_0^i)\Big)\\
&\overset{(3.11)}{=}\delta_0\Big((\boldsymbol{\mathcal{K}}\nabla\rho_0^{i+\frac12},\nabla\rho_0^i)+\sum_{j=0}^J\lambda_j^i(\boldsymbol{\mathcal{K}}\nabla\rho_j^i,\nabla\rho_0^i)\Big)\\
&\le\ \delta_0\frac{1}{2(1-\gamma^2)}\big\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\rho_0^{i+\frac12}\big\|^2+\delta_0\frac{1-\gamma^2}{2}\big\|\boldsymbol{\mathcal{K}}^{\frac12}\nabla\rho_0^i\big\|^2+\delta_0\sum_{j=0}^J\lambda_j^i(\boldsymbol{\mathcal{K}}\nabla\rho_j^i,\nabla\rho_0^i).
\end{aligned}$$

Second,

$$\begin{aligned}
\sum_{j\in\mathcal{M}\setminus\{0\}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}(\boldsymbol{\mathcal{K}}\nabla\rho_{j,\mathbf{a}}^i,\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}&\overset{(3.2)}{=}\sum_{j\in\mathcal{M}\setminus\{0\}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\Big((f,\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}-(\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^i,\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}\Big)\\
&\overset{(3.7)}{=}\sum_{j\in\mathcal{M}\setminus\{0\}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\Big((f,\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}-(\boldsymbol{\mathcal{K}}\nabla u_J^i,\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}-\sum_{k=0}^{j-1}\lambda_k^i(\boldsymbol{\mathcal{K}}\nabla\rho_k^i,\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}\Big)\\
&\overset{(3.18)}{=}\sum_{j\in\mathcal{M}\setminus\{0\}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\Big((f,\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}-(\boldsymbol{\mathcal{K}}\nabla u_J^{i+\frac12},\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}\\
&\qquad+\sum_{k=0}^J\lambda_k^i(\boldsymbol{\mathcal{K}}\nabla\rho_k^i,\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}-\sum_{k=0}^{j-1}\lambda_k^i(\boldsymbol{\mathcal{K}}\nabla\rho_k^i,\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}\Big)\\
&\overset{(3.17)}{=}\sum_{j\in\mathcal{M}\setminus\{0\}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\Big((f,\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}-(\boldsymbol{\mathcal{K}}\nabla u_{J,j-1}^{i+\frac12},\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}\\
&\quad+\sum_{\substack{l=0\\l\in\mathcal{M}}}^{j-1}\lambda_l^{i+\frac12}(\boldsymbol{\mathcal{K}}\nabla\rho_l^{i+\frac12},\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}+\sum_{k=j}^J\lambda_k^i(\boldsymbol{\mathcal{K}}\nabla\rho_k^i,\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}\Big)\\
&\overset{(3.12)}{=}\sum_{j\in\mathcal{M}\setminus\{0\}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\Big((\boldsymbol{\mathcal{K}}\nabla\rho_{j,\mathbf{a}}^{i+\frac12},\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}\\
&\quad+\sum_{l=0}^{j-1}\delta_l\lambda_l^{i+\frac12}(\boldsymbol{\mathcal{K}}\nabla\rho_l^{i+\frac12},\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}+\sum_{k=j}^J\lambda_k^i(\boldsymbol{\mathcal{K}}\nabla\rho_k^i,\nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}}\Big)
\end{aligned}$$

$$\overset{(3.12)}{\leq} \sum_{j\in\mathcal{M}\setminus\{0\}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \Big(\frac{1}{1-\gamma^2} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^{i+\frac12}\|_{\omega_j^{\mathbf{a}}}^2 + \frac{1-\gamma^2}{4} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^{i}\|_{\omega_j^{\mathbf{a}}}^2$$
$$+ \frac{1}{1-\gamma^2} \Big\| \sum_{l=0}^{j-1} \delta_l \lambda_l^{i+\frac12} \mathcal{K}^{\frac12}\nabla\rho_l^{i+\frac12} \Big\|_{\omega_j^{\mathbf{a}}}^2 + \frac{1-\gamma^2}{4} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^{i}\|_{\omega_j^{\mathbf{a}}}^2 + \sum_{k=j}^{J} \lambda_k^i (\mathcal{K}\nabla\rho_k^i, \nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}} \Big)$$

We return to the main estimate by summing the two estimates and using the result of Test (3.9)

$$\sum_{j\in\mathcal{M}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}^2 \leq \frac{1}{1-\gamma^2} \sum_{j\in\mathcal{M}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^{i+\frac12}\|_{\omega_j^{\mathbf{a}}}^2$$
$$+ \frac{1-\gamma^2}{2} \sum_{j\in\mathcal{M}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}^2 + \frac{1}{1-\gamma^2} \sum_{j\in\mathcal{M}\setminus\{0\}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \Big\| \sum_{\substack{l=0\\ l\in\mathcal{M}}}^{j-1} \lambda_l^{i+\frac12} \mathcal{K}^{\frac12}\nabla\rho_l^{i+\frac12} \Big\|_{\omega_j^{\mathbf{a}}}^2$$
$$+ \sum_{j\in\mathcal{M}\setminus\{0\}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \Big(\sum_{k=j}^{J} \lambda_k^i (\mathcal{K}\nabla\rho_k^i, \nabla\rho_{j,\mathbf{a}}^i)_{\omega_j^{\mathbf{a}}} \Big) + \delta_0 \sum_{j=0}^{J} \lambda_j^i (\mathcal{K}\nabla\rho_j^i, \nabla\rho_0^i)$$

$$\overset{(3.9)}{\leq} \frac{1}{1-\gamma^2} \sum_{j\in\mathcal{M}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^{i+\frac12}\|_{\omega_j^{\mathbf{a}}}^2 + \frac{1-\gamma^2}{2} \sum_{j\in\mathcal{M}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}^2$$
$$+ \frac{1}{1-\gamma^2} \sum_{j\in\mathcal{M}\setminus\{0\}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \Big\| \sum_{\substack{l=0\\ l\in\mathcal{M}}}^{j-1} \lambda_l^{i+\frac12} \mathcal{K}^{\frac12}\nabla\rho_l^{i+\frac12} \Big\|_{\omega_j^{\mathbf{a}}}^2 + \gamma^2 \sum_{j\in\mathcal{M}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}^2.$$

Rearranging the terms, we have

$$\frac{1-\gamma^2}{2} \sum_{j\in\mathcal{M}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}^2 \leq \frac{1}{1-\gamma^2} \sum_{j\in\mathcal{M}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^{i+\frac12}\|_{\omega_j^{\mathbf{a}}}^2$$
$$+ \frac{1}{1-\gamma^2} \sum_{j\in\mathcal{M}\setminus\{0\}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \Big\| \sum_{\substack{l=0\\ l\in\mathcal{M}}}^{j-1} \lambda_l^{i+\frac12} \mathcal{K}^{\frac12}\nabla\rho_l^{i+\frac12} \Big\|_{\omega_j^{\mathbf{a}}}^2,$$

leading to

$$\sum_{j\in\mathcal{M}} \lambda_j^i \sum_{\mathbf{a}\in\mathcal{M}_j} \|\mathcal{K}^{\frac12}\nabla\rho_{j,\mathbf{a}}^i\|_{\omega_j^{\mathbf{a}}}^2 \overset{\substack{(3.10)\\(7.8)}}{\leq} \frac{4(d+1)^2}{(1-\gamma^2)^2} \Big(\Big(\eta_{\mathrm{alg}}^{i+\frac12}\Big)^2 + \sum_{j\in\mathcal{M}\setminus\{0\}} \Big\| \sum_{\substack{l=0\\ l\in\mathcal{M}}}^{j-1} \lambda_l^{i+\frac12} \mathcal{K}^{\frac12}\nabla\rho_l^{i+\frac12} \Big\|^2 \Big)$$
$$\leq \frac{4(d+1)^2}{(1-\gamma^2)^2} \Big(\Big(\eta_{\mathrm{alg}}^{i+\frac12}\Big)^2 + \sum_{j\in\mathcal{M}\setminus\{0\}} |\mathcal{M}| \sum_{\substack{l=0\\ l\in\mathcal{M}}}^{j-1} \Big\| \lambda_l^{i+\frac12} \mathcal{K}^{\frac12}\nabla\rho_l^{i+\frac12} \Big\|^2 \Big)$$
$$\leq \frac{4(d+1)^2}{(1-\gamma^2)^2} \Big(\Big(\eta_{\mathrm{alg}}^{i+\frac12}\Big)^2 + |\mathcal{M}|^2 \sum_{l\in\mathcal{M}} \Big(\lambda_l^{i+\frac12} \|\mathcal{K}^{\frac12}\nabla\rho_l^{i+\frac12}\| \Big)^2 \Big) \overset{(4.2)}{=} \frac{4(d+1)^2(|\mathcal{M}|^2+1)}{(1-\gamma^2)^2} \Big(\eta_{\mathrm{alg}}^{i+\frac12}\Big)^2.$$

where $|\mathcal{M}|$ denotes the number of marked levels. □

We can now prove the contraction of the adaptive-smoothing substep below.

Proof of part 2 of Theorem 5.4. Step 1. We prove that there holds:

$$\|\mathcal{K}^{\frac12}\nabla(u_J - u_J^{i+\frac12})\|^2 \leq \tilde{\beta}^2 \Big(\eta_{\mathrm{alg}}^{i+\frac12}\Big)^2. \tag{7.10}$$

By Theorem 5.5, the efficiency of the estimator η^i_{alg} is equivalent to error contraction of the full-smoothing substep. Using the equivalence error–localized contributions of Corollary 5.6/(7.6), the bulk-chasing criterion (3.8), and the result of Lemma 7.6,

$$\begin{aligned}\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\big(u_J-u_J^{i+\frac{1}{2}}\big)\big\|^2 &\overset{\text{Theorem 5.5}}{\leq} \alpha^2\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\big(u_J-u_J^{i}\big)\big\|^2\\ &\overset{(7.6)}{\leq} \alpha^2 C_{\mathrm{S}}^2\Big(\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_0^i\big\|^2+\sum_{j=1}^{J}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{V}_j}\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\big\|^2_{\omega_j^{\mathbf{a}}}\Big)\\ &\overset{(3.8)}{\leq} \frac{\alpha^2 C_{\mathrm{S}}^2}{\theta^2}\sum_{j\in\mathcal{M}}\lambda_j^i\sum_{\mathbf{a}\in\mathcal{M}_j}\big\|\boldsymbol{\mathcal{K}}^{\frac{1}{2}}\nabla\rho_{j,\mathbf{a}}^i\big\|^2_{\omega_j^{\mathbf{a}}}\overset{(7.9)}{\leq}\frac{4\alpha^2C_{\mathrm{S}}^2(d+1)^2(|\mathcal{M}|^2+1)}{\theta^2(1-\gamma^2)^2}\Big(\eta_{\mathrm{alg}}^{i+\frac{1}{2}}\Big)^2,\end{aligned}$$

giving the desired result for $\tilde{\beta}^2=\dfrac{4\alpha^2C_{\mathrm{S}}^2(d+1)^2(|\mathcal{M}|^2+1)}{\theta^2(1-\gamma^2)^2}$. Thus, the estimator $\eta_{\mathrm{alg}}^{i+\frac{1}{2}}$ (guaranteed lower bound (4.4)), is p-robustly efficient.

Step 2. By Theorem 5.5, (7.10) is equivalent to (5.6) with $\tilde{\alpha}=\sqrt{1-\tilde{\beta}^2}$. □

8 Conclusions

In this work we presented an adaptive multilevel solver whose adaptive process is supervised by an a posteriori estimator of the algebraic error. We showed that both full-smoothing and adaptive-smoothing substeps of the solver contract the error robustly with respect to the polynomial degree of approximation p, under the decision tests (3.9)–(3.10) for the latter. To the best of the authors' knowledge, this is the first work where adaptive smoothing not necessarily everywhere in the meshes is proven to contract the algebraic error, and moreover does so in a p-robust way. Numerical experiments indicate that the adaptivity can provide an interesting speed-up and is worth considering in practice. Furthermore, the solver appears numerically robust with respect to the number of levels in the hierarchy as well as the jump in the diffusion coefficient. Further work would explore how this can be rigorously proven.

References

[1] A. Anciaux-Sedrakian, L. Grigori, Z. Jorti, J. Papež, and S. Yousef, *Adaptive solution of linear systems of equations based on a posteriori error estimators*, Numerical Algorithms, 84 (2020), pp. 331–364.

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

[3] C. Bernardi and Y. Maday, *Spectral methods*, in Handbook of numerical analysis, Vol. V, Handb. Numer. Anal., V, North-Holland, Amsterdam, 1997, pp. 209–485.

[4] 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.

[5] M. Brezina, R. Falgout, S. MacLachlan, T. Manteuffel, S. McCormick, and J. Ruge, *Adaptive algebraic multigrid*, SIAM J. Sci. Comput., 27 (2006), pp. 1261–1286.

[6] W. L. Briggs, V. E. Henson, and S. F. McCormick, *A multigrid tutorial*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, second ed., 2000.

[7] 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.

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

[9] 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.

[10] V. Dolean, P. Jolivet, and F. Nataf, *An introduction to domain decomposition methods*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2015. Algorithms, theory, and parallel implementation.

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

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

[13] M. Griebel and P. Oswald, *On the abstract theory of additive and multiplicative Schwarz algorithms*, Numer. Math., 70 (1995), pp. 163–180.

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

[15] X. Hu, J. Lin, and L. T. Zikatanov, *An adaptive multigrid method based on path cover*, SIAM J. Sci. Comput., 41 (2019), pp. S220–S241.

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

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

[18] H. Lötzbeyer and U. Rüde, *Patch-adaptive multilevel iteration*, BIT, 37 (1997), pp. 739–758. Direct methods, linear algebra in optimization, iterative methods (Toulouse, 1995/1996).

[19] S. F. McCormick, *Multilevel adaptive methods for partial differential equations*, vol. 6 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1989.

[20] A. Miraçi, J. Papež, and M. Vohralík, *A multilevel algebraic error estimator and the corresponding iterative solver with p-robust behavior.* SIAM J. Numer. Anal. (2020). DOI 10.1137/19M1275929.

[21] A. Miraçi, J. Papež, and M. Vohralík, *A-posteriori-steered p-robust multigrid with optimal step-sizes and adaptive number of smoothing steps.* HAL preprint 02494538, 2020. URL https://hal.inria.fr/hal-02494538.

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

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

[24] 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.

[25] A. Quarteroni and A. Valli, *Domain decomposition methods for partial differential equations*, Numerical Mathematics and Scientific Computation, The Clarendon Press Oxford University Press, New York, 1999. Oxford Science Publications.

[26] U. Rüde, *Mathematical and computational techniques for multilevel adaptive methods*, vol. 13 of Frontiers in Applied Mathematics, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1993.

[27] 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.

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

[29] B. Szabó and I. Babuška, *Finite element analysis*, A Wiley-Interscience Publication, John Wiley & Sons Inc., New York, 1991.

[30] P. Šolín, K. Segeth, and I. Doležel, *Higher-order finite element methods*, Studies in Advanced Mathematics, Chapman & Hall/CRC, Boca Raton, FL, 2004. With 1 CD-ROM (Windows, Macintosh, UNIX and LINUX).

[31] 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.

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