Electronic Transactions on Numerical Analysis.
Volume 63, pp. 566–608, 2025.
© 2025, the Author(s), licensed via CC BY 4.0.
ISSN 1068–9613.
DOI: 10.1553/etna_vol63s566

ETNA
Kent State University and
Johann Radon Institute (RICAM)

A POSTERIORI ERROR ESTIMATES BASED ON MULTILEVEL DECOMPOSITIONS WITH AN ITERATIVE SOLVER ON THE COARSEST LEVEL[*]

PETR VACEK[†‡], JAN PAPEŽ[§], AND ZDENĚK STRAKOŠ[‡]

Abstract. Multilevel methods represent a powerful approach to the numerical solution of partial differential equations. The multilevel structure can also be used to construct estimates for the total and algebraic errors of the computed approximations. This paper deals with residual-based error estimates that rely on properties of quasi-interpolation operators, stable splittings, or frames. We focus on the settings where the system matrix on the coarsest level is still large and the associated terms in the estimates can only be approximated. We show that the way in which the error term associated with the coarsest level is approximated is crucial. It can significantly affect both the efficiency (accuracy) of the overall error estimates and their robustness with respect to the size of the coarsest-level problem. We propose a new approximation of the coarsest-level term based on using the conjugate gradient method with an appropriate stopping criterion. We prove that the resulting estimates are efficient and robust with respect to the size of the coarsest-level problem. Numerical experiments illustrate the theoretical findings.

Key words. a posteriori estimates, multilevel hierarchy, residual-based error estimator, large coarsest-level problem, iterative computation

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

1. Introduction.

Multilevel methods [7, 10, 22, 43] are frequently used for solving systems of linear equations obtained from the discretization of partial differential equations (PDEs). They are applied either as standalone iterative solvers or as preconditioners. In *geometric* multigrid methods, the hierarchy of systems is obtained by the discretization of an infinite-dimensional problem for a sequence of nested meshes. In *algebraic* multigrid methods, the coarse systems are constructed using algebraic properties of the matrix. Each multigrid cycle involves smoothing on the fine levels, prolongation, and solving a system of linear equations on the coarsest level. Smoothing is typically done by a few iterations of a stationary iterative method. If the size permits, the coarsest-level problem is typically solved using a direct method based on an LU or a Cholesky decomposition. Although this does not provide a computed result with zero error, many theoretical results for multigrid methods are proved under the assumption that the coarsest-level problem is solved exactly; see, e.g., [47, 49].

Multilevel methods can in practice also use hierarchies where the coarsest-level problem is large. This arises for problems on complicated domains or for large-scale problems solved on modern parallel computers; see, e.g., [11, 18]. When using a very large number of computing units (either CPUs or GPUs), usually very few equations are assigned to the processes at the end of the coarsening. Then the only feasible way of employing a *direct (distributed) solver* on the coarsest-level is to pool resources restricting the number of involved processes [13] or to replicate computations [4]. The coarsest-level problem is therefore often solved *iteratively*, i.e., only approximately *to a suitably prescribed accuracy*, e.g., by Krylov subspace methods (such as in [18]) or direct methods with low-rank matrix approximations; see, e.g., [11]. Effects of approximate coarsest-level solves on the convergence of multigrid method were analyzed, e.g., in [32, 45, 48].

[*] Received June 23, 2025. Accepted November 7, 2025. Published online on December 19, 2025. Recommended by Marco Donatelli.

[†] Corresponding author. IFP Energies Nouvelles, Rueil-Malmaison, France (petr.vacek@ifpen.fr).

[‡] Department of Numerical Mathematics, Faculty of Mathematics and Physics, Charles University, Czech Republic.

[§] Institute of Mathematics, Czech Academy of Sciences.

This work is published by ETNA and licensed under the Creative Commons license CC BY 4.0.

The multilevel structure can also be used to construct estimates for the total and algebraic errors; see, e.g., [5, 23, 26, 31, 34, 37]. The estimates of [5, 23, 26, 31, 34, 37] are, however, not suited for multilevel hierarchies with large coarsest-level problems that are used for complicated domains and/or in parallel implementations. They either assume that the coarsest-level problem is solved exactly [5, 31, 34], or they require a computation of the term $\mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0$ associated with the coarsest level, where $\mathbf{A}_0$ is the coarsest-level system matrix and $\mathbf{r}_0$ a projection of the finest-level residual to the coarsest level [23, 26, 37]. The term $\mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0$ can be approximated, e.g., using the conjugate gradient (CG) method as in [26] or by replacing the system matrix with a diagonal matrix as in [23]. Then proving efficiency and robustness of the estimates becomes an important challenge.

In this work, we discuss properties of the error estimates in multilevel settings where the system matrix on the coarsest level is large and the associated terms are only approximated. We consider several a posteriori estimates for the total and algebraic errors, based on decomposing the error into a sequence of finite element subspaces and using either approximation properties of quasi-interpolation operators [5], stable splittings [26, 37], or so-called frames [23]. The main contribution of this paper is a new procedure for approximating the term $\mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0$ associated with the coarsest level that is based on using the conjugate gradient method with an appropriate stopping criterion. We prove that the resulting estimates are *efficient and robust* with respect to the size of the coarsest-level problem.

The text is organized as follows. First, we present a model problem, its discretization, and the notation used in the text. Derivations of error estimates for the total and algebraic errors are presented in Section 3 following the literature or standard techniques used in the literature. In Section 4, we recall results on the efficiency of the bounds. In Section 5 we first describe how to replace the (uncomputable) terms in the estimates by a computable approximation. Then we present the new adaptive procedure for approximating the coarsest-level term. Numerical illustrations are given in Section 6 and conclusions in Section 7. To avoid interrupting the presentation, we provide detailed theoretical results that are used in the derivation of the estimates in the appendices. Appendix A summarizes standard results from the analysis of PDE and the finite element method (FEM). Appendix B discusses properties of quasi-interpolation operators, while Appendix C presents results for stable splittings and frames. Its purpose is to facilitate an easy comparison of various findings that are scattered in the literature. The results presented in Appendix C describe the dependence of the constants in the estimates on properties of the mesh.

2. Model problem, setting, and notation.

The estimates will be studied for a standard model problem, a prototype for elliptic equations, namely the Poisson's problem with homogeneous Dirichlet boundary conditions. Let $\Omega \subset \mathbb{R}^d$, $d = 2, 3$, be an open bounded polytope with a Lipschitz-continuous boundary. Given $f \in L^2(\Omega)$, the weak form reads:

Find $u \in H_0^1(\Omega)$ such that

$$\int_\Omega \nabla u \cdot \nabla v = \int_\Omega fv, \qquad \forall v \in H_0^1(\Omega). \tag{2.1}$$

In this section, we introduce some notation for meshes and finite element spaces and the multilevel framework. Further, we present the Galerkin finite element discretization of the model problem on a particular level, define its approximate solution, the error, and the (scaled) residuals associated with the individual levels of the multilevel hierarchy.

Similarly to the standard literature, we introduce some simplifying assumptions, e.g., on the model problem or the mesh hierarchies. This is done in order to reduce the complexity of proofs (that are already quite technical) and to allow us to refer to particular results in the literature. We use a standard notation for Lebesgue and Sobolev (Hilbert) spaces, norms, and seminorms; see, e.g., [9].

2.1. Notation for a single level. Throughout the paper, we consider simplicial meshes of Ω that are matching in the sense that for two distinct elements of a mesh $\mathcal{T}$ (triangles in 2D or tetrahedra in 3D), their intersection is either an empty set or a common node (vertex), edge, or face. By $\mathcal{E}_\mathcal{T}$ and $\mathcal{N}_\mathcal{T}$ we denote the set of $(d-1)$-dimensional faces and the set of nodes in the mesh $\mathcal{T}$, respectively. By $\mathcal{E}_{\mathcal{T},\mathrm{int}}$ we denote the set of all faces that are not on the boundary $\partial\Omega$. By $\mathcal{K}_\mathcal{T} \subset \mathcal{N}_\mathcal{T}$ we denote the set of all nodes in the mesh $\mathcal{T}$ that are not on the boundary, i.e., the *free nodes*. For any element (simplex) $K \in \mathcal{T}$, the symbol $\mathcal{E}_K \subset \mathcal{E}_\mathcal{T}$ denotes the set of faces of the element K, the symbol $\mathcal{N}_K \subset \mathcal{N}_\mathcal{T}$ denotes the set of nodes of the element K, and $\mathcal{E}_{K,\mathrm{int}} = \mathcal{E}_K \cap \mathcal{E}_{\mathcal{T},\mathrm{int}}$ and $\mathcal{K}_K = \mathcal{N}_K \cap \mathcal{K}_\mathcal{T}$. We use a hash to denote the cardinality of a set, for example, $\#\mathcal{K}_\mathcal{T}$ means the number of free nodes in the mesh $\mathcal{T}$. For the ease of presentation, we will assume that the nodes in $\mathcal{N}_\mathcal{T}$ are ordered such that the nodes $1, \ldots, \#\mathcal{K}_\mathcal{T}$ belong to $\mathcal{K}_\mathcal{T}$, i.e., the first indices correspond to the free nodes followed by the nodes on the boundary.

By h_K we denote the diameter of $K \in \mathcal{T}$ and define a meshsize $h_\mathcal{T} \in L^\infty(\Omega)$ as

$$h_\mathcal{T}(x) = h_K, \qquad x \in K, \quad \forall K \in \mathcal{T}.$$

Similarly h_ω denotes the diameter of a domain ω. We in particular use h_Ω to indicate the diameter of the domain Ω. By $|\omega|$ we denote the Lebesgue measure of a domain ω.

For any element $K \in \mathcal{T}$, the symbol ω_K denotes the patch of elements that share at least one common vertex with K, i.e.,

$$\omega_K = \bigcup_{K' \in \mathcal{T}; K' \cap K \neq \emptyset} K'.$$

By ρ_K we denote the diameter of the largest ball inscribed in the element K.

For every node $z \in \mathcal{N}_\mathcal{T}$, let ϕ_z be the continuous piecewise linear function (*hat function*) that has the value one at the node z and vanishes at all other nodes in $\mathcal{N}_\mathcal{T}$. Let $S_\mathcal{T}$ denote the space of continuous, piecewise linear functions,

$$S_\mathcal{T} = \{v \in H^1(\Omega), v|_K \in \mathbb{P}^1(K), \ \forall K \in \mathcal{T}\} = \mathrm{span}\{\phi_z, \ z \in \mathcal{N}_\mathcal{T}\},$$

and $V_\mathcal{T} \subset S_\mathcal{T}$ the subspace of functions vanishing on the boundary $\partial\Omega$,

$$V_\mathcal{T} = \{v \in H^1_0(\Omega), v|_K \in \mathbb{P}^1(K), \ \forall K \in \mathcal{T}\} = \mathrm{span}\{\phi_z, \ z \in \mathcal{K}_\mathcal{T}\}.$$

We write the basis of $V_\mathcal{T}$ as $\Phi_\mathcal{T} = (\phi_1, \ldots, \phi_{\#\mathcal{K}_\mathcal{T}})$.

One of the key properties of a mesh that affects the size of the constants in the estimates derived below is the so-called *shape regularity* of the mesh. This can be quantified by the shape-regularity constant, i.e., the smallest $\gamma_\mathcal{T} > 0$ satisfying

$$\frac{h_K}{\rho_K} \leq \gamma_\mathcal{T}, \qquad \forall K \in \mathcal{T}; \tag{2.2}$$

see, e.g., [38, p. 484].

2.2. Multilevel framework. As the title of the paper suggests, we will work with a sequence of levels $j = 0, 1, \ldots, J$. For some parts of the theory, we will also consider infinite sequences of levels $j = 0, 1, \ldots, J, \ldots$ To simplify the previously introduced notation, we will replace in subscripts the symbol $\mathcal{T}_j$ by j, hereby denoting objects associated with the mesh $\mathcal{T}_j$ on the jth level.

Let $\mathcal{T}_0$ be an initial mesh of Ω. We consider a sequence of meshes $\mathcal{T}_1, \mathcal{T}_2, \ldots$ obtained by successive uniform dyadic refinements of $\mathcal{T}_0$, i.e., each element is refined into 2^d elements (congruent triangles in 2D; for a proper nondegenerating 3D mesh refinement, see, e.g., [50]). We recall that S_j and V_j, $j = 0, 1, \ldots$, are the finite element spaces of continuous piecewise linear functions on $\mathcal{T}_j$, respectively the spaces of continuous piecewise linear functions on $\mathcal{T}_j$ that vanish on the boundary $\partial\Omega$. These spaces are nested, i.e.,

$$S_0 \subset S_1 \subset \cdots \subset H^1(\Omega), \qquad V_0 \subset V_1 \subset \cdots \subset H^1_0(\Omega).$$

On each level j, we consider a quasi-interpolation operator

$$I_{V_j} : L^1(\Omega) \to V_j$$

with the definition and properties described in detail in Appendix B.

Due to the uniform refinement, the mesh sizes h_j of $\mathcal{T}_j$, $j \geq 0$, satisfy $h_j = 2^{-j} h_0$. Moreover, the uniform refinement assures that the shape-regularity constants γ_j of the meshes are the same on all levels in two dimensions, i.e., $\gamma_0 = \gamma_j$, $j \in \mathbb{N}$, and that in three dimensions there exists a constant $C_{3\mathrm{D}} > 0$ such that $\gamma_j \leq C_{3\mathrm{D}} \gamma_0$, $j \in \mathbb{N}$; see [50].

2.3. Discretization, approximate solution, and residuals. Discretizing the model problem (2.1) on the subspace V_J, for some $J \geq 0$, using the Galerkin method reads as:

Find $u_J \in V_J$ such that

$$\int_\Omega \nabla u_J \cdot \nabla w_J = \int_\Omega f w_J, \qquad \forall w_J \in V_J. \tag{2.3}$$

Let $v_J \in V_J$ be a (computed) approximation of the discrete solution u_J. Our goal is to bound the energy norm of the total error $e = u - v_J$ using computable quantities involving v_J and f. The squared energy norm of the error $\|\nabla e\|^2$ can be expressed as

$$\|\nabla e\|^2 = \|\nabla(u - v_J)\|^2 = \int_\Omega \nabla(u - v_J) \cdot \nabla(u - v_J) = \int_\Omega f(u - v_J) - \nabla v_J \cdot \nabla(u - v_J).$$

Denote by $\left(H^1_0(\Omega)\right)^\star$ the dual space to $H^1_0(\Omega)$, and define the residual $r \in \left(H^1_0(\Omega)\right)^\star$ as

$$\langle r, w \rangle = \int_\Omega f w - \nabla v_J \cdot \nabla w, \qquad \forall w \in H^1_0(\Omega). \tag{2.4}$$

Then (2.4) yields the so-called residual equation

$$\|\nabla e\|^2 = \langle r, e \rangle, \tag{2.5}$$

which is the key formula for the development of the error bounds presented below. Moreover, it can be shown (see, e.g., [46, Section 1.4.1]) that

$$\|\nabla e\| = \|r\|_{\left(H^1_0(\Omega)\right)^\star}.$$

In order to derive computable estimates, we consider Riesz representations of the infinite-dimensional residual r in the finite-dimensional spaces V_j, $j = 0, 1, \ldots$ In particular, let $r_j \in V_j$, $j = 1, \ldots$, be the Riesz representation of r in the space V_j with the scaled L^2-inner product, i.e.,

$$\langle r, w_j \rangle = \int_\Omega h_j^{-2} r_j w_j, \qquad \forall w_j \in V_j, \tag{2.6}$$

and let $r_0 \in V_0$ be the Riesz representation of the residual r in the space V_0 with the H_0^1-inner product, i.e.,

$$\langle r, w_0 \rangle = \int_\Omega \nabla r_0 \cdot \nabla w_0, \qquad \forall w_0 \in V_0. \tag{2.7}$$

These definitions are used in [37, Section 2.6], where r_j are called *scaled residuals*. In [5, Section 5] the authors use Riesz representations of r in the spaces V_j, $j = 1, \dots, J$, with the classical L^2-inner products, and they refer to them as discrete residuals. The different definition that we use results in a slightly different form of the estimates below in comparison to [5, Section 5].

2.4. Table of constants.

For the clarity of presentation and the reader's convenience, we present in Table 2.1 a list of constants used in the later development, together with a brief description and reference to their definition or appearance, which is mostly in the appendices of the paper. All these constants only depend on the dimension d and the shape regularity parameter γ_0 of the initial mesh.

TABLE 2.1
List of constants used in Sections 3 to 5.

Constant	Brief description
C_{cls}	Residual-based error estimate (3.1).
$C_{I_{V_j},\ell}$, $C_{I,2\mathrm{lvl}}$	Properties of the quasi-interpolation operator, Theorems B.5 and B.6; we also write $C_{I_j,\ell}$ for $C_{I_{V_j},\ell}$, $\ell = 1, 2, 3, 4$.
c_{S,I_V}, C_{S,I_V}	Stability of splitting of $H_0^1(\Omega)$ using the quasi-interpolation operators, Theorem C.9.
c_s, C_S	Stability of splitting of $H_0^1(\Omega)$ into subspaces of piecewise linear functions, Theorem C.10.
c_B, C_B	Stability of basis functions, Lemma C.11.
$\overline{c}_B$, $\overline{C}_B$	Defined as $\overline{c}_B = \min\{1, c_B\}$, $\overline{C}_B = \max\{1, C_B\}$.
C_{HS}	Estimate (3.16), also appearing in [23, Proof of Thm. 5.1].
c_M, C_M	Spectral equivalence of mass matrix and stiffness matrix diagonal (C.19).
C_F	Friedrich's inequality, Lemma A.2.
C_{INV}	Inverse inequality, Lemma A.4.

3. Residual-based error estimates.

In this section we derive several error estimates following the techniques in the literature. We first recall, in a single-level setting, the standard residual-based error estimator for the discretization error that assumes exact algebraic computations.

Consider the model problem (2.1) discretized on a level $J \geq 0$ of a multilevel hierarchy as in Section 2.2. The classical residual-based estimator (see, e.g., [1, Section 3], [46, Section 1.4]) is, for a (computed) approximation $v_J \in V_J$, defined as

$$\begin{aligned}
\eta_J^2 &= \left(\eta_J^{\mathrm{RHS}}\right)^2 + \left(\eta_J^{\mathrm{JUMP}}\right)^2 + \left(\mathrm{osc}_J\right)^2, \\
\left(\eta_J^{\mathrm{RHS}}\right)^2 &= \sum_{K\in\mathcal{T}_J} h_K^2 \|f_K\|_K^2, \\
\left(\eta_J^{\mathrm{JUMP}}\right)^2 &= \frac{1}{2}\sum_{K\in\mathcal{T}_J} h_K \sum_{E\in\mathcal{E}_{K,\mathrm{int}}} \|\left[\nabla v_J\right]\|_E^2, \\
\left(\mathrm{osc}_J\right)^2 &= \sum_{K\in\mathcal{T}_J} h_K^2 \|f - f_K\|_K^2,
\end{aligned}$$

where $[\cdot]$ denotes the jump of a piecewise constant function over the $(d-1)$-dimensional faces (faces in 3D and edges in 2D) and f_K is the mean value of f on K. Other choices of f_K are also possible; see, e.g., [23].

The following result (see, e.g., [5, Lemma 3], [39, Section 4], or [46, Section 1.4]) will be useful below. There exists a constant $C_{\mathrm{cls}} > 0$ depending only on the dimension d and the shape-regularity parameter γ_0 such that

$$\langle r, w - I_{V_J} w\rangle \leq C_{\mathrm{cls}}\eta_J\|\nabla w\|, \qquad \forall w \in H_0^1(\Omega). \tag{3.1}$$

Note that if v_J is equal to the Galerkin solution u_J, then the associated residual $r = r(u_J)$ satisfies the Galerkin orthogonality on the finest level, i.e.,

$$\langle r, w_J\rangle = 0, \qquad \forall w_J \in V_J.$$

Then,

$$\|\nabla(u-u_J)\|^2 = \langle r, (u-u_J) - I_{V_J}(u-u_J)\rangle,$$

and using (3.1) for $w = u - u_J$ yields the standard bound for the discretization error:

$$\|\nabla(u-u_J)\| \leq C_{\mathrm{cls}}\eta_J(u_J).$$

3.1. Estimates of Becker, Johnson & Rannacher. The following derivation is motivated by [5] and uses a decomposition of the error via quasi-interpolation operators. Considering the residual equation (2.5) and writing the error $e = u - v_J$ as

$$e = e - I_{V_J}e + \sum_{j=1}^{J}\left(I_{V_j}e - I_{V_{j-1}}e\right) + I_{V_0}e$$

yields

$$\|\nabla e\|^2 = \langle r, e\rangle = \langle r, e - I_{V_J}e\rangle + \sum_{j=1}^{J}\langle r, I_{V_j}e - I_{V_{j-1}}e\rangle + \langle r, I_{V_0}e\rangle. \tag{3.2}$$

The first term on the right-hand side of (3.2) can be bounded using (3.1) as

$$\langle r, e - I_{V_J}e\rangle \leq C_{\mathrm{cls}}\eta_J\|\nabla e\|. \tag{3.3}$$

The second and the third term on the right-hand side of (3.2) can be rewritten using the scaled residuals (2.6), (2.7) and subsequently bounded as

$$\begin{aligned}
&\sum_{j=1}^{J}\langle r, I_{V_j}e - I_{V_{j-1}}e\rangle + \langle r, I_{V_0}e\rangle \\
&\qquad = \sum_{j=1}^{J}\int_\Omega h_j^{-2} r_j (I_{V_j}e - I_{V_{j-1}}e) + \int_\Omega \nabla r_0 \cdot \nabla I_{V_0} e \qquad (3.4)\\
&\qquad \le \sum_{j=1}^{J}\|h_j^{-1} r_j\| \cdot \|h_j^{-1}(I_{V_j}e - I_{V_{j-1}}e)\| + \|\nabla r_0\| \cdot \|\nabla I_{V_0} e\|.
\end{aligned}$$

Further, using the bound for the difference of the quasi-interpolants on two consecutive levels (Appendix B, Theorem B.6) and the stability of the quasi-interpolation operator on the coarsest level in the $H_0^1(\Omega)$-norm (Appendix B, Theorem B.5, inequality (B.10)), we get

$$\begin{aligned}
&\sum_{j=1}^{J}\|h_j^{-1} r_j\| \cdot \|h_j^{-1}(I_{V_j}e - I_{V_{j-1}}e)\| + \|\nabla r_0\| \cdot \|\nabla I_{V_0} e\| \qquad (3.5)\\
&\qquad \le C_{I,\mathrm{2lvl}}\left(\sum_{j=1}^{J}\|h_j^{-1} r_j\|\right)\|\nabla e\| + \|\nabla r_0\| \cdot C_{I_{V_0},4} \cdot \|\nabla e\|.
\end{aligned}$$

Combining (3.2)–(3.5) yields the following:

ESTIMATE FOR THE TOTAL ERROR 1.

$$\|\nabla e\| \le C_{\mathrm{cls}}\eta_J + C_{I,\mathrm{2lvl}}\sum_{j=1}^{J}\|h_j^{-1} r_j\| + C_{I_{V_0},4}\|\nabla r_0\|. \qquad (3.6)$$

In [5] the authors assume that the approximation v_J is computed by a multigrid scheme without post-smoothing and with the exact solution of the problem on the coarsest level. This yields the Galerkin orthogonality on the coarsest level, i.e.,

$$\langle r, w_0\rangle = 0, \qquad \forall w_0 \in V_0.$$

As a consequence, their estimate for the energy norm of the error (see [5, Theorem 1]) does not contain the term corresponding to the coarsest level. Another difference between (3.6) and the estimate in [5, Theorem 1] is due to the difference in the definitions of the scaled/discrete residuals described in Section 2.3.

Instead of using the bound for the difference of the quasi-interpolants on two consecutive levels (Appendix B, Theorem B.6) and the stability of the quasi-interpolation operator on the coarsest level (Appendix B, Theorem B.5, inequality (B.10)), we can use the stability of the decomposition of the space $H_0^1(\Omega)$ via the quasi-interpolation operators I_{V_j} (Appendix C, Theorem C.9). In particular,

$$\begin{aligned}
&\sum_{j=1}^{J}\|h_j^{-1} r_j\| \cdot \|h_j^{-1}(I_{V_j}e - I_{V_{j-1}}e)\| + \|\nabla r_0\| \cdot \|\nabla I_{V_0} e\| \\
&\qquad \le \left(\sum_{j=1}^{J}\|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2\right)^{\frac{1}{2}}\left(\sum_{j=1}^{J}\|h_j^{-1}(I_{V_j}e - I_{V_{j-1}}e)\|^2 + \|\nabla I_{V_0} e\|^2\right)^{\frac{1}{2}}
\end{aligned}$$

$$\leq \left(\sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \right)^{\frac{1}{2}} C_{S,I_V}{}^{\frac{1}{2}} \|\nabla e\|.$$

Combining this inequality with (3.2)–(3.4) and using $\sqrt{a} + \sqrt{b} \leq \sqrt{2}\sqrt{a+b}$ leads to the following:

ESTIMATE FOR THE TOTAL ERROR 2.

$$\|\nabla e\| \leq \sqrt{2} \left(C_{\text{cls}}^2 \eta_J^2 + C_{S,I_V} \left(\sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \right) \right)^{\frac{1}{2}}.$$

As we will see in Section 4, this estimate is efficient with an efficiency constant independent of the number of levels in the hierarchy, i.e., independent of J.

Observing that

$$\begin{aligned} \|\nabla(u_J - v_J)\|^2 &= \int_\Omega f(u_J - v_J) - \int_\Omega \nabla v_J \cdot \nabla(u_J - v_J) = \langle r, u_J - v_J \rangle \\ &= \sum_{j=1}^{J} \langle r, I_{V_j}(u_J - v_J) - I_{V_{j-1}}(u_J - v_J) \rangle + \langle r, I_{V_0}(u_J - v_J) \rangle, \end{aligned}$$

analogous steps can be applied to show that the following "algebraic parts" of the presented estimates provide upper bounds for the algebraic error:

ESTIMATE FOR THE ALGEBRAIC ERROR 1.

$$\|\nabla(u_J - v_J)\| \leq C_{I,2\text{lvl}} \sum_{j=1}^{J} \|h_j^{-1} r_j\| + C_{I_0,3} \|\nabla r_0\|, \tag{3.7}$$

ESTIMATE FOR THE ALGEBRAIC ERROR 2.

$$\|\nabla(u_J - v_J)\| \leq C_{S,I_V}{}^{\frac{1}{2}} \left(\sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \right)^{\frac{1}{2}}. \tag{3.8}$$

3.2. Estimates of Rüde & Huber. The following derivation is motivated by [37, Section 2.6] and [26, Sections 4.1–4.3]. Considering the residual equation (2.5) and decomposing the error using the quasi-interpolation operator on the finest level I_{V_J} yields

$$\|\nabla e\|^2 = \langle r, e - I_{V_J} e \rangle + \langle r, I_{V_J} e \rangle. \tag{3.9}$$

The first term can be bounded as in (3.3). Rewriting the second term using the exact solution of the discrete problem u_J gives

$$\begin{aligned} \langle r, I_{V_J} e \rangle &= \int_\Omega \nabla(u - v_J) \nabla I_{V_J} e \\ &= \int_\Omega \nabla(u - u_J) \nabla I_{V_J} e + \int_\Omega \nabla(u_J - v_J) \nabla I_{V_J} e. \end{aligned}$$

The Galerkin orthogonality on the finest level yields that $\int_\Omega \nabla(u-u_J)\nabla I_{V_J} e$ vanishes, and thus,

$$\langle r, I_{V_J} e\rangle = \int_\Omega \nabla(u_J - v_J)\nabla I_{V_J} e \leq \|\nabla(u_J - v_J)\|\, \|\nabla I_{V_J} e\|. \tag{3.10}$$

After establishing a bound for the term $\|\nabla I_{V_J} e\|$ by using the stability property of the quasi-interpolation operator (Appendix B, Theorem B.5, and inequality (B.10)) of the form

$$\|\nabla I_{V_J} e\| \leq C_{I_{V_J},4}\|\nabla e\|, \tag{3.11}$$

it remains to bound the energy norm of the algebraic error $\|\nabla(u_J - v_J)\|$. This can be done using a stable splitting of the function space of piecewise linear functions; see Appendix C, Theorem C.13, or [37, Theorem 2.6.2]. Consider an arbitrary decomposition of the algebraic error $u_J - v_J$ into the subspaces V_j, i.e.,

$$u_J - v_J = \sum_{j=0}^{J} e_j, \qquad e_j \in V_j, \quad j = 0, 1, \ldots, J. \tag{3.12}$$

Then

$$\begin{aligned}
\|\nabla(u_J - v_J)\|^2 = \langle r, u_J - v_J\rangle = \sum_{j=0}^{J}\langle r, e_j\rangle \\
&\leq \|\nabla r_0\| \cdot \|\nabla e_0\| + \sum_{j=1}^{J} \|h_j^{-1} r_j\| \cdot \|h_j^{-1} e_j\| \\
&\leq \left(\|\nabla r_0\|^2 + \sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 \right)^{\frac{1}{2}} \cdot \left(\|\nabla e_0\|^2 + \sum_{j=1}^{J} \|h_j^{-1} e_j\|^2 \right)^{\frac{1}{2}}.
\end{aligned}$$

Taking the infimum over all possible decompositions (3.12) and using Appendix C, Theorem C.13 yields the following:

ESTIMATE FOR THE ALGEBRAIC ERROR 3.

$$\|\nabla(u_J - v_J)\| \leq C_S^{\frac{1}{2}} \left(\sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \right)^{\frac{1}{2}}. \tag{3.13}$$

Combining (3.9)–(3.11), the estimate (3.13) for the algebraic error, and using the inequality $\sqrt{a} + \sqrt{b} \leq \sqrt{2}\sqrt{a+b}$, we obtain:

ESTIMATE FOR THE TOTAL ERROR 3.

$$\|\nabla e\| \leq \sqrt{2} \left(C_{\mathrm{cls}}^2 \eta_J^2 + C_{I_{V_J},4}^2 C_S \left(\sum_{j=0}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \right) \right)^{\frac{1}{2}}. \tag{3.14}$$

3.3. Estimates of Harbrecht & Schneider. In this section we present error estimates motivated by [23], which are based on the fact that the basis functions provide a frame in

$\left(H_0^1(\Omega)\right)^\star$; see Appendix C, Theorem C.15. Recall that $\left(H_0^1(\Omega)\right)^\star$ is the dual space to $H_0^1(\Omega)$. Using the upper bound for the residual yields

$$\|\nabla e\| = \|r\|_{\left(H_0^1(\Omega)\right)^\star} \le C_S^{\frac{1}{2}} \overline{C}_B^{\frac{1}{2}} \left(\|\nabla r_0\|^2 + \sum_{j=1}^{+\infty} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle r, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} \right)^{\frac{1}{2}}. \tag{3.15}$$

Following the derivation in [23, Proof of Theorem 5.1], it can be shown that the sum of the terms corresponding to levels $j > J$, i.e.,

$$\sum_{j=J+1}^{+\infty} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle r, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2},$$

can be bounded by the classic residual-based estimator on the Jth level up to a constant $C_{\mathrm{HS}} > 0$ depending only on d and γ_0, i.e.,

$$\sum_{j=J+1}^{+\infty} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle r, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} \le C_{\mathrm{HS}} \eta_J^2. \tag{3.16}$$

Combining (3.15) and (3.16) yields the following:

ESTIMATE FOR THE TOTAL ERROR 4.

$$\|\nabla e\| \le C_S^{\frac{1}{2}} \overline{C}_B^{\frac{1}{2}} \left(C_{\mathrm{HS}} \eta_J^2 + \sum_{j=1}^{J} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle r, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} + \|\nabla r_0\|^2 \right)^{\frac{1}{2}}. \tag{3.17}$$

Considering the residual r as a functional on V_J, which is possible since $\left(H_0^1(\Omega)\right)^\star \subset V_J^\star$, one can show that

$$\|\nabla (u_J - v_J)\| = \|r\|_{V_J^\star}.$$

From Appendix C, Theorem C.16, it follows that a part of the total error estimator (3.17) serves as an upper bound for the algebraic error:

ESTIMATE FOR THE ALGEBRAIC ERROR 4.

$$\|\nabla (u_J - v_J)\| \le C_S^{\frac{1}{2}} \overline{C}_B^{\frac{1}{2}} \left(\sum_{j=1}^{J} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle r, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} + \|\nabla r_0\|^2 \right)^{\frac{1}{2}}. \tag{3.18}$$

4. Efficiency of the estimates.

The efficiency of the estimates (upper bounds) is described by the constant C_{eff} such that

$$\text{estimate} \le C_{\mathrm{eff}} \cdot \|\text{error}\|.$$

Here we focus in particular on whether C_{eff} depends on the number of levels J, the quasi-uniformity of the coarsest mesh, and/or on the ratio $h_\Omega / \min_{K \in \mathcal{T}_0} h_K$, which is related to the size of the coarsest-level problem.

4.1. Efficiency of the estimates for the algebraic error. We will first discuss estimates of the form

$$\|\nabla(u_J - v_J)\| \leq C \left(\sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \right)^{1/2},$$

where either $C = C_{S,I_V}{}^{\frac{1}{2}}$ or $C = C_S{}^{\frac{1}{2}}$ is a constant depending only on the dimension d and the shape-regularity parameter γ_0; see (3.8) and (3.13). Using the definition of scaled residuals (2.6)–(2.7), the Cauchy–Schwarz inequality, and the lower bound from Appendix C, Theorem C.13, we have (see also the proof of Theorem 2.6.2 in [37])

$$\begin{aligned}
\sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 &= \sum_{j=0}^{J} \langle r, r_j \rangle \\
&= \int_\Omega \nabla(u_J - v_J) \cdot \nabla\Big(\sum_{j=0}^{J} r_j \Big) \leq \|\nabla(u_J - v_J)\| \cdot \left\| \nabla\Big(\sum_{j=0}^{J} r_j \Big) \right\| \\
&\leq \|\nabla(u_J - v_J)\| \cdot c_S{}^{-\frac{1}{2}} \left(\sum_{j=1}^{J} \|h_j^{-2} r_j\|^2 + \|\nabla r_0\|^2 \right)^{1/2}.
\end{aligned}$$

Consequently,

$$\left(\sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \right)^{1/2} \leq c_S{}^{-\frac{1}{2}} \|\nabla(u_J - v_J)\|, \tag{4.1}$$

i.e., the efficiency constant depends only on the dimension d and the shape-regularity parameter γ_0.

The efficiency of the estimate (3.18),

$$\|\nabla(u_J - v_J)\| \leq C_S^{\frac{1}{2}} \overline{C}_B^{\frac{1}{2}} \left(\sum_{j=1}^{J} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle r, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} + \|\nabla r_0\|^2 \right)^{1/2},$$

can be shown by using $\|\nabla(u_J - v_J)\| = \|r\|_{(V_J)^\star}$ and the lower bound from Appendix C, Theorem C.16, yielding

$$\left(\sum_{j=1}^{J} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle r, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} + \|\nabla r_0\|^2 \right)^{1/2} \leq c_S^{-\frac{1}{2}} \overline{c}_B^{-\frac{1}{2}} \|\nabla(u_J - v_J)\|.$$

Again, the efficiency constant depends only on the dimension d and the shape-regularity parameter γ_0.

Finally, for the estimate (3.7),

$$\|\nabla(u_J - v_J)\| \leq C_{I,\mathrm{2lvl}} \sum_{j=1}^{J} \|h_j^{-1} r_j\| + C_{I_0,3} \|\nabla r_0\|,$$

the equivalence of the Euclidean and ℓ^1-norm,

$$\sum_{j=1}^{J} \|h_j^{-1} r_j\| + \|\nabla r_0\| \leq \sqrt{J} \left(\sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \right)^{1/2},$$

and (4.1) yield

$$C_{I,2\mathrm{lvl}} \sum_{j=1}^{J} \|h_j^{-1} r_j\| + C_{I_0,3} \|\nabla r_0\| \leq \sqrt{J} \max\{C_{I,2\mathrm{lvl}}, C_{I_0,3}\}\, c_S^{-\frac{1}{2}} \|\nabla(u_J - v_J)\|.$$

This shows the efficiency of (3.7) with $C_{\mathrm{eff}} = \sqrt{J}\widetilde{C}_{\mathrm{eff}}$, where $\widetilde{C}_{\mathrm{eff}}$ depends only on d and γ_0. This result does not necessarily imply a dependence of (3.7) on the number of levels J. However, below in Section 6.1, we present a numerical experiment indicating such a behaviour.

4.2. Efficiency of the estimates for the total error. The efficiency of the total error estimates follows from the standard result on the efficiency of the classical (one-level) residual-based error estimator. There exists a positive constant $\overline{C}_{\mathrm{eff}}$ depending on the shape regularity of $\mathcal{T}_J$ such that

$$\left(\left(\eta_J^{\mathrm{RHS}}\right)^2 + \left(\eta_J^{\mathrm{JUMP}}\right)^2 \right)^{\frac{1}{2}} \leq \overline{C}_{\mathrm{eff}} \left(\|\nabla e\| + \mathrm{osc}_J \right); \tag{4.2}$$

see, e.g., [46, Section 1.4]. Since $\|\nabla(u_J - v_J)\| \leq \|\nabla e\|$, we can use the efficiency of the algebraic error estimates together with (4.2) to show the efficiency of the estimates for the total error (up to the oscillation term). The resulting efficiency constants depend on the same quantities as the efficiency constants for the algebraic error estimates.

For example, for the estimate (3.14) associated with the algebraic error estimate (3.13), we obtain

$$\sqrt{2} \left(C_{\mathrm{cls}}^2 \eta_J^2 + C_{I_{V_J},4}^2 C_S \left(\sum_{j=0}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \right) \right)^{\frac{1}{2}} \leq C \left(\|\nabla e\| + \mathrm{osc}_J \right),$$

with $C^2 = 2(C_{\mathrm{cls}}^2(\overline{C}_{\mathrm{eff}}^2 + 1) + C_{I_{V_J},4}^2 C_S c_S^{-1})$.

5. Computability of the error estimates. In this section we address several ways in which the scaled residual norms from the estimates presented in Section 3 can be evaluated or bounded. When the scaled residual norms are replaced by their bounds, proving the efficiency of the estimates from Section 3 becomes a nontrivial task.

We first state an algebraic formulation of the problem (2.3). Then we present the algebraic representation of the scaled residual norms and some of their bounds from the literature. Section 5.4 provides a new approach for approximating the scaled residual norm on the coarsest level using an adaptive number of conjugate gradient iterations. This yields total and algebraic error estimates which are *provably* efficient and robust with respect to the size of the coarsest-level problem.

5.1. Algebraic formulation of the problem, residual vectors. Given a basis Φ_J of V_J, the problem (2.3) can be algebraically formulated as finding the vector of coefficients $\mathbf{u}_J \in \mathbb{R}^{\#\mathcal{K}_J}$ of the function u_J in the basis Φ_J such that

$$\mathbf{A}_J \mathbf{u}_J = \mathbf{f}_J,$$

where $\mathbf{A}_J$ is the *stiffness matrix* on the finest level J,

$$[\mathbf{A}_J]_{mn} = \int_\Omega \nabla\phi_n^{(J)} \cdot \nabla\phi_m^{(J)},$$

and

$$[\mathbf{f}_J]_m = \int_\Omega f\phi_m^{(J)}, \qquad m, n = 1, \ldots, \#\mathcal{K}_J.$$

Recall that $\#\mathcal{K}_J$ is the cardinality of the basis Φ_J. We use the standard assumption that the right-hand side vector $\mathbf{f}_J$ can be computed exactly using a numerical quadrature. If $\mathbf{f}_J$ is only known approximately, then an additional term must be added to the error bounds presented above; see, e.g., the discussion in [39, Section 6].

Let v_J be an approximation of the solution u_J of (2.3) and $\mathbf{v}_J$ be the vector of coefficients of v_J in the basis Φ_J. Let r be the residual (2.4) associated with v_J. Consider the residual vectors $\mathbf{r}_j \in \mathbb{R}^{\#\mathcal{K}_j}$, $j = 0, \ldots, J$,

$$[\mathbf{r}_j]_m = \langle r, \phi_m^{(j)} \rangle, \qquad m = 1, \ldots, \#\mathcal{K}_j. \tag{5.1}$$

The vector $\mathbf{r}_J$ corresponding to the finest level can be computed as

$$\mathbf{r}_J = \mathbf{f}_J - \mathbf{A}_J\mathbf{v}_J.$$

The residual vectors corresponding to coarser levels can be computed from $\mathbf{r}_J$ by restriction. Let $\mathbf{P}_j^J \in \mathbb{R}^{\#\mathcal{K}_J \times \#\mathcal{K}_j}$ be the prolongation matrix associated with the (nested) finite element spaces V_j, V_J and the bases Φ_j, Φ_J. This means that for an arbitrary $v_j \in V_j \subset V_J$ with $\mathbf{v}_j$ being its coefficients with respect to the basis Φ_j, the vector $\mathbf{P}_j^J\mathbf{v}_j$ provides the associated coefficients of v_j with respect to the basis Φ_J. Then,

$$\mathbf{r}_j = (\mathbf{P}_j^J)^\top \mathbf{r}_J;$$

see, e.g., [37, Section 3.2] or [44, Section 2.4] for a more detailed explanation.

5.2. The terms associated with the fine levels. In this section we present an algebraic form of the term $\|h_j^{-1}r_j\|^2$, $j = 1, \ldots, J$, and several methods adapted from the literature to bound it using computable quantities.

Let $\mathbf{c}_j$ be the vector of coefficients of r_j in the basis Φ_j. The definitions (5.1) of $\mathbf{r}_j$ and (2.6) of r_j give, for all $m = 1, \ldots, \#\mathcal{K}_j$,

$$[\mathbf{r}_j]_m = \langle r, \phi_m^{(j)} \rangle = \int_\Omega h_j^{-2} r_j \phi_m^{(j)} = \sum_n \int_\Omega h_j^{-2} \left[\mathbf{c}_j\right]_n \phi_n^{(j)} \phi_m^{(j)}. \tag{5.2}$$

Let $\mathbf{M}_j^{\mathrm{S}}$ be a *scaled mass matrix* defined as

$$\left[\mathbf{M}_j^{\mathrm{S}}\right]_{m,n} = \int_\Omega h_j^{-2} \phi_n^{(j)} \phi_m^{(j)}, \qquad \forall m, n = 1, \ldots, \#\mathcal{K}_j.$$

Equation (5.2) can then be expressed as $\mathbf{r}_j = \mathbf{M}_j^{\mathrm{S}}\mathbf{c}_j$, and therefore

$$\|h_j^{-1}r_j\|^2 = \int_\Omega h_j^{-2} \Phi_j \mathbf{c}_j \cdot \Phi_j \mathbf{c}_j = \mathbf{c}_j^* \mathbf{M}_j^{\mathrm{S}} \mathbf{c}_j = \mathbf{r}_j^* (\mathbf{M}_j^{\mathrm{S}})^{-1} \mathbf{r}_j. \tag{5.3}$$

The evaluation of the term (5.3) thus involves the solution of a system with a possibly large matrix $\mathbf{M}_j^{\mathrm{S}}$. Instead of computing this quantity, one can seek a computable upper bound.

Let $\mathbf{D}_j$ be a diagonal matrix $[\mathbf{D}_j]_{m,m} = \int_\Omega \nabla\phi_m^{(j)} \cdot \nabla\phi_m^{(j)}$, $m = 1, \dots, \#\mathcal{K}_j$. The stability of basis functions (Appendix C, Theorem C.11) and (C.18) give

$$c_B \mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{r}_j \le \|h_j^{-1} r_j\|^2 = \mathbf{r}_j^* (\mathbf{M}_j^{\mathrm{S}})^{-1} \mathbf{r}_j \le C_B \mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{r}_j. \tag{5.4}$$

The upper bound in (5.4) is used in [26, 37] to bound the algebraic error as

$$\begin{aligned} \|\nabla(u_J - v_J)\| &\le C_S^{\frac{1}{2}} \Big(C_B \sum_{j=1}^{J} \mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{r}_j + \|\nabla r_0\|^2 \Big)^{\frac{1}{2}} \\ &\le C_S^{\frac{1}{2}} \overline{C}_B^{\frac{1}{2}} \Big(\sum_{j=1}^{J} \mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{r}_j + \|\nabla r_0\|^2 \Big)^{\frac{1}{2}}, \end{aligned} \tag{5.5}$$

where $\overline{C}_B = \max\{1, C_B\}$. For $\overline{c}_B = \min\{1, c_B\}$, using the lower bound in (5.4) and (4.1) yields

$$\begin{aligned} \Big(\sum_{j=1}^{J} \mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{r}_j + \|\nabla r_0\|^2 \Big)^{\frac{1}{2}} &\le \Big(c_B^{-1} \sum_{j=1}^{J} \|h_j^{-1} r_j\|^2 + \|\nabla r_0\|^2 \Big)^{\frac{1}{2}} \\ &\le \overline{c}_B^{-\frac{1}{2}} c_S^{-\frac{1}{2}} \|\nabla(u_J - v_J)\|, \end{aligned} \tag{5.6}$$

which proves the efficiency of the bound (5.5). Recall that c_B, C_B, c_S, and C_S only depend on d and γ_0.

Noting that

$$\mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{r}_j = \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle r, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2},$$

we see that the algebraic error bounds (3.18) and (5.5) are identical.

The term (5.3) can also be bounded by using other techniques, e.g., using the so-called mass lumping (suggested in [5, Section 4]) or the multigrid smoothing routines (see the discussion in [26, Section 4.5.2]). By using these techniques, however, we introduce another unknown constant into the overall estimate and possibly weaken its efficiency.

In order to get a fully computable bound for (5.3) (i.e., a bound without any unknown constant) and to avoid solving an algebraic problem with a large matrix, we can proceed similarly to [35]. Define $\bar{r}_j \in L^2(\Omega)$ to be the (discontinuous) piecewise affine functions on $\mathcal{T}_j$ such that for all $K \in \mathcal{T}_j$,

$$\int_K h_j^{-2} \bar{r}_j \phi_m^{(j)} = [\mathbf{r}_j]_m \cdot \frac{1}{\#\left\{\bar{K} \in \mathcal{T}_j; m \text{ is vertex of } \bar{K}\right\}} =: [\mathbf{r}_{j,K}]_m \qquad \forall \phi_m^{(j)}.$$

This ensures that

$$\int_\Omega h_j^{-2} \bar{r}_j \phi_m^{(j)} = [\mathbf{r}_j]_m = \langle r, \phi_m^{(j)} \rangle.$$

Since $\bar{r}_j$ is piecewise affine on elements, the norms $\|h_j^{-1}\bar{r}_j\|_K^2$ can be computed using the solutions of systems with local scaled mass matrices, i.e., $\|h_j^{-1}\bar{r}_j\|_K^2 = \mathbf{r}_{j,K}^* (\mathbf{M}_{j,K}^{\mathrm{S}})^{-1} \mathbf{r}_{j,K}$, where

$$[\mathbf{M}_{j,K}^S]_{m,n} = \int_K h_j^{-2} \phi_n^{(j)} \phi_m^{(j)}, \qquad \forall m, n \in \mathcal{N}_K.$$

For the whole term $\|h_j^{-1} r_j\|$, we have

$$\|h_j^{-1} r_j\|^2 \leq \|h_j^{-1} \bar{r}_j\|^2 = \sum_{K \in \mathcal{T}_j} \mathbf{r}_{j,K}^* (\mathbf{M}_{j,K}^{\mathrm{S}})^{-1} \mathbf{r}_{j,K};$$

cf. [35, Eq. (5.9)].

5.3. The term associated with the coarsest level.

In this section we present the algebraic form of the term $\|\nabla r_0\|$ and several ways of bounding it adapted from the literature.

Let $\mathbf{c}_0$ be the vector of coefficients of r_0 in the basis Φ_0. Analogously to (5.3), using the definitions (5.1) of $\mathbf{r}_0$ and (2.7) of r_0, we have

$$[\mathbf{r}_0]_m = \langle r, \phi_m^{(0)} \rangle = \int_\Omega \nabla r_0 \cdot \nabla \phi_m^{(0)} = \sum_n \int_\Omega [\mathbf{c}_0]_n \nabla \phi_n^{(0)} \cdot \nabla \phi_m^{(0)}, \quad \forall m = 1, \ldots, \#\mathcal{K}_0.$$

Let $\mathbf{A}_0$ be the stiffness matrix associated with the coarsest level,

$$[\mathbf{A}_0]_{mn} = \int_\Omega \nabla \phi_n^{(0)} \cdot \nabla \phi_m^{(0)}, \qquad m, n = 1, \ldots, \#\mathcal{K}_0.$$

The vector of coefficients $\mathbf{c}_0$ then satisfies $\mathbf{A}_0 \mathbf{c}_0 = \mathbf{r}_0$. This leads to

$$\|\nabla r_0\|^2 = \mathbf{c}_0^* \mathbf{A}_0 \mathbf{c}_0 = \mathbf{r}_0^* \mathbf{A}_0^{-1} \mathbf{r}_0. \tag{5.7}$$

The evaluation of the term $\|\nabla r_0\|^2$ thus requires the solution of the system with the stiffness matrix associated with the coarsest level. For problems where the stiffness matrix is large, this can be too costly and in some settings even unfeasible.

An approximate solution $\widetilde{\mathbf{c}}_0$ of $\mathbf{A}_0 \mathbf{c}_0 = \mathbf{r}_0$ computed by the (preconditioned) conjugate gradient method with a fixed number of iterations was used in [26, Section 4.5.2]. The resulting term $\widetilde{\mathbf{c}}_0^* \mathbf{r}_0$ might not be, however, an upper bound for $\|\nabla r_0\|^2$. Therefore, the resulting value may neither lead to an upper bound for the algebraic nor the total error.

The term (5.7) can be bounded using a quantity involving only the inverse of a diagonal matrix. Friedrich's inequality (Appendix A, Theorem A.2) implies that

$$\|w_0\|^2 \leq C_F^2 h_\Omega^2 \|\nabla w_0\|^2, \qquad \forall w_0 \in V_0. \tag{5.8}$$

Let $\mathbf{M}_0$ be the mass matrix associated with the coarsest level, i.e., $[\mathbf{M}_0]_{mn} = \int_\Omega \phi_n^{(0)} \phi_m^{(0)}$, $m, n = 1, \ldots, \#\mathcal{K}_0$. The inequality (5.8) can be equivalently expressed algebraically as

$$\mathbf{w}^* \mathbf{M}_0 \mathbf{w} \leq C_F^2 h_\Omega^2 \mathbf{w}^* \mathbf{A}_0 \mathbf{w}, \qquad \forall \mathbf{w} \in \mathbb{R}^{\#\mathcal{K}_0}.$$

Since $\mathbf{A}_0$ and $\mathbf{M}_0$ are symmetric positive definite matrices, we have

$$\mathbf{w}^* \mathbf{A}_0^{-1} \mathbf{w} \leq C_F^2 h_\Omega^2 \mathbf{w}^* \mathbf{M}_0^{-1} \mathbf{w}, \qquad \forall \mathbf{w} \in \mathbb{R}^{\#\mathcal{K}_0}.$$

This bound may possibly be a large overestimation; see the discussion in [35, Sections 3.1 and 5.2]. Define the diagonal matrix $\mathbf{D}_0$ as $[\mathbf{D}_0]_{m,m} = \int_\Omega \nabla \phi_m^{(0)} \cdot \nabla \phi_m^{(0)}$, $m = 1, \ldots, \#\mathcal{K}_0$. The term on the right-hand side can be further simplified using the spectral equivalence of the mass matrix $\mathbf{M}_0$ with $\mathbf{D}_0$; see inequality (C.19) in Appendix C. Altogether we have

$$\|\nabla r_0\|^2 = \mathbf{r}_0^* \mathbf{A}_0^{-1} \mathbf{r}_0 \leq C_F^2 h_\Omega^2 \mathbf{r}_0^* \mathbf{M}_0^{-1} \mathbf{r}_0 \leq C_M C_F^2 \frac{h_\Omega^2}{\min_{K \in \mathcal{T}_0} h_K^2} \mathbf{r}_0^* \mathbf{D}_0^{-1} \mathbf{r}_0. \tag{5.9}$$

As for the efficiency, this allows to prove, using the inverse inequality (Appendix A, Theorem A.4) and Appendix C, (C.19),

$$\mathbf{r}_0^*\mathbf{D}_0^{-1}\mathbf{r}_0 \leq \frac{C_{\mathrm{INV}}^2}{c_M}\frac{\max_{K\in\mathcal{T}_0} h_K^2}{\min_{K\in\mathcal{T}_0} h_K^2}\|\nabla r_0\|^2,$$

which indicates that the bound (5.9) may not be robust with respect to $h_\Omega^2/\min_{K\in\mathcal{T}_0} h_K^2$. Numerical experiments in Section 6 illustrate this deficiency.

5.4. Adaptive approximation of the coarsest-level term. In order to overcome the deficiencies described above, we now present a new approach for approximating the term (5.7). It consists of applying the preconditioned conjugate gradient (PCG) method to $\mathbf{A}_0\mathbf{c}_0 = \mathbf{r}_0$ and using lower and upper bounds for the error in the PCG method. A number of PCG iterations is determined adaptively in order to ensure the efficiency of the resulting bounds for the total and algebraic errors.

Let $\mathbf{c}_0^{(i)}$ be the approximation of $\mathbf{c}_0 = \mathbf{A}_0^{-1}\mathbf{r}_0$ computed at the i-th iteration of PCG with a zero initial guess. Let $\|\cdot\|_{\mathbf{A}_0}$ be the norm generated by the matrix $\mathbf{A}_0$, i.e., $\|\mathbf{v}\|_{\mathbf{A}_0}^2 = \mathbf{v}^*\mathbf{A}_0\mathbf{v}$, for all $\mathbf{v} \in \mathbb{R}^{\#\mathcal{K}_0}$. The term (5.7) can be expressed using the decomposition

$$\mathbf{c}_0^*\mathbf{A}_0\mathbf{c}_0 = \underbrace{\sum_{m=0}^{i-1}\|\mathbf{c}_0^{(m+1)} - \mathbf{c}_0^{(m)}\|_{\mathbf{A}_0}^2}_{=:\mu_i^2} + \|\mathbf{c}_0 - \mathbf{c}_0^{(i)}\|_{\mathbf{A}_0}^2, \tag{5.10}$$

which is a consequence of the *local orthogonality* in PCG. This formula was already shown for CG in the seminal paper [25, Theorem 6:1, Eq. (6:2)]. The terms $\|\mathbf{c}_0^{(m)} - \mathbf{c}_0^{(m+1)}\|_{\mathbf{A}_0}^2$ can be computed at minimal cost from the scalars available during the computations. It is crucial to note that the local orthogonality in CG and PCG computations is preserved up to machine precision. Therefore, (5.10) is valid, up to a negligible error, also in finite-precision computations; see the derivation and proofs in [40] (respectively in [41] for the preconditioned variant).

Let ζ_i^2 be an upper bound for the squared $\mathbf{A}_0$-norm of the error in the PCG computation, i.e., for $\|\mathbf{c}_0 - \mathbf{c}_0^{(i)}\|_{\mathbf{A}_0}^2$. Such a bound can be derived using the interpretation of PCG as a procedure for computing the Gauss-quadrature approximation to a Riemann–Stieltjes integral. A detailed explanation in [21, 30] and the references therein[1] consider the unpreconditioned CG method, but the approach can be easily extended also for PCG. This requires a lower bound for the smallest eigenvalue of the preconditioned stiffness matrix in PCG. If a Gauss-quadrature-based upper bound is not available, then the $\mathbf{A}_0$-norm of the error $\|\mathbf{c}_0 - \mathbf{c}_0^{(i)}\|_{\mathbf{A}_0}^2$ can be bounded, for CG or PCG, using the ideas presented in Section 5.3 or the technique from [35, Section 3.2].

The approach for bounding (5.7) then consists of running PCG for the coarsest-level problem $\mathbf{A}_0\mathbf{c}_0 = \mathbf{r}_0$ until

$$\zeta_i^2 \leq \theta\left(\sum_{j=1}^{J}\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j + \mu_i^2\right), \tag{5.11}$$

[1] Strictly speaking, numerical stability for the upper bounds of the A-norm of the error in CG computations has not been rigorously proved. Well-justified heuristics supported by numerical experiments, however, suggest their validity also in finite-precision computations; see [21, 30].

where $\theta > 0$ is a chosen parameter. Then we consider the bound

$$\mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0 \le \mu_i^2 + \zeta_i^2, \tag{5.12}$$

which can be combined, e.g., with (5.5), to get an upper bound for the algebraic error

$$\|\nabla(u_J - v_J)\| \le C_S^{\frac{1}{2}}\overline{C}_B^{\frac{1}{2}}\left(\sum_{j=1}^{J}\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j + \mu_i^2 + \zeta_i^2\right)^{\frac{1}{2}}. \tag{5.13}$$

The criterion (5.11) guarantees that

$$\mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0 \le \mu_i^2 + \zeta_i^2 \le \theta\sum_{j=1}^{J}\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j + (1+\theta)\mu_i^2, \tag{5.14}$$

which allows us to prove the efficiency of (5.13). Indeed, using (5.14), $\mu_i^2 \le \|\nabla r_0\|^2$ (see (5.10)), and (5.6),

$$\begin{aligned}\left(\sum_{j=1}^{J}\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j + \mu_i^2 + \zeta_i^2\right)^{\frac{1}{2}} &\le (1+\theta)^{\frac{1}{2}}\left(\sum_{j=1}^{J}\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j + \|\nabla r_0\|^2\right)^{\frac{1}{2}}\\ &\le (1+\theta)^{\frac{1}{2}}\, c_S^{-\frac{1}{2}} c_B^{-\frac{1}{2}}\|\nabla(u_J - v_J)\|\,.\end{aligned}$$

The proposed strategy follows the ideas of [35, Section 3.2]. In principle, the possible overestimation in $\|\mathbf{c}_0 - \mathbf{c}_0^{(i)}\|_{\mathbf{A}_0}^2 \le \zeta_i^2$ is controlled by (5.11), and it is compensated for within the procedure by performing extra iterations. This allows us to prove the efficiency even if the estimate ζ_i is not very tight. However, when the convergence is slow, the number of extra iterations might be quite large; see [35, Section 7.1].

We note that $\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j$ in (5.11) can be replaced by any (efficient) bound for $\|h_j^{-1}r_j\|^2$. Then the algebraic error bound (5.13) should be changed accordingly, replacing $\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j$ and $\overline{C}_B$.

The cost of estimating the coarsest-level term (5.7) depends, in general, on the preconditioner for $\mathbf{A}_0\mathbf{c}_0 = \mathbf{r}_0$ and on the overestimation of ζ_i. The number of PCG iterations needed in the adaptive approximation is one of the quantities investigated in the numerical experiments below. In the experiment in Section 6.3, we also use a heuristic upper bound for the error $\|\mathbf{c}_0 - \mathbf{c}_0^{(i)}\|_{\mathbf{A}_0}^2$ from [29], which has no theoretical guarantee but seems to perform well in practice.

6. Numerical experiments.

The experiments focus on the efficiency of the error estimates for the algebraic error. In particular, we consider the estimate

$$C\left(\sum_{j=1}^{J}\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j + \mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0\right)^{\frac{1}{2}} \tag{6.1}$$

and variants, where $\mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0 = \|\nabla r_0\|^2$ is replaced by computable approximations. This prototype covers most of the algebraic error estimates from Section 3, where the scaled residual norms $\|h_j^{-1}r_j\|^2$ on the fine levels are efficiently approximated by $\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j$ using (5.4). As shown in the previous sections, approximating the coarsest-level term $\|\nabla r_0\|^2$ while preserving the efficiency is more subtle.

For the experiments, we consider a 3D Poisson problem on a unit cube, $\Omega = (0,1)^3$, with the exact solution

$$u(x,y,z) = x(x-1)y(y-1)z(z-1)e^{-100\left((x-\frac{1}{2})^2+(y-\frac{1}{2})^2+(z-\frac{1}{2})^2\right)}.$$

The problem is discretized by the standard Galerkin finite element method with piecewise affine polynomials on a sequence of six uniformly refined meshes with the same shape regularity (2.2). The associated matrices are generated in the FE software FEniCS [2, 27], and the computations are done in MATLAB 2023a. The codes for the experiments are available from `https://github.com/vacek-petr/inMLEstimate`. The repository contains a collection of experiments with two-dimensional problems as well as a three-dimensional example on a more complex geometry that cannot be easily included in this text.

Given the mesh $\mathcal{T}_J$ (the finest mesh varies in the experiments), the associated Galerkin solution u_J of (2.3) is—for the purpose of the evaluation of the efficiency of the estimates—obtained (with a negligible inaccuracy) by using the MATLAB backslash, or, for very large problems, by using the multigrid V-cycle with an excessive number (30) of V-cycle repetitions. The approximation v_J to u_J is given by a multigrid solver starting with a zero approximation and repeating V-cycles until the relative energy norm of the (algebraic) error $u_J - v_J$ drops below 10^{-11}. Each multigrid V-cycle uses 3 pre- and 3 post-Gauss–Seidel smoothing iterations. The problem on the coarsest level is solved using the CG method, where the stopping criterion is based on the relative residual with the tolerance 10^{-1}. In order to monitor the efficiency for varying the algebraic error, we also provide intermediate results after completing each multigrid V-cycle.

6.1. Robustness with respect to the number of levels.

The first experiment studies the efficiency of the estimates while varying the number of levels, $J = 1, 2, \dots, 5$, in the hierarchy. We fix the size of the problem on the coarsest level and, consequently, the size of the finest problem grows; see Table 6.1.

TABLE 6.1
Size of the problems for the experiment in Section 6.1.

coarsest-level DoFs	finest-level DoFs
125	1 331
125	12 167
125	103 823
125	857 375
125	6 967 871

For the prototype estimate (6.1), the efficiency index

$$I_1 = \frac{C_{\mathrm{numexp}} \left(\sum_{j=1}^{J} \mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{r}_j + \mathbf{r}_0^* \mathbf{A}_0^{-1} \mathbf{r}_0 \right)^{\frac{1}{2}}}{\|\nabla(u_J - v_J)\|}, \tag{6.2}$$

is evaluated for every J, v_J, and also for intermediate results after each V-cycle. The factor C_{numexp} accounts for $C_S^{\frac{1}{2}} \overline{C}_B^{\frac{1}{2}}$; see (5.5). For the purpose of the experiment, it is chosen as the minimal value such that the efficiency indices I_1 are, for all J and in all V-cycle repetitions, above or equal to one; $C_{\mathrm{numexp}} = 1.28$. In order to examine the difference, we also evaluate

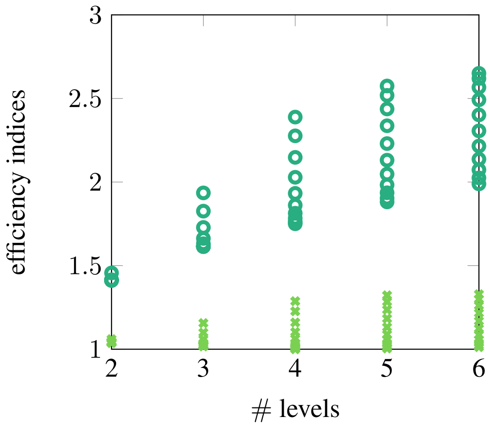

FIG. 6.1. *Efficiency indices I_1 (✖) and I_2 (○), (6.2) and (6.3), for varying number of levels J. We plot the efficiency for the approximations v_J and for the associated intermediate results after each V-cycle; each instance corresponds to a single mark.*

the index

$$I_2 = \frac{C_{\text{numexp}} \left(\sum_{j=1}^{J} \left(\mathbf{r}_j^* \mathbf{D}_j^{-1} \mathbf{r}_j \right)^{\frac{1}{2}} + \left(\mathbf{r}_0^* \mathbf{A}_0^{-1} \mathbf{r}_0 \right)^{\frac{1}{2}} \right)}{\|\nabla(u_J - v_J)\|}, \tag{6.3}$$

which corresponds to the algebraic error bound (3.7).

The index I_1 (6.2) corresponds to the estimate (5.5) that is proved to be robust with respect to the number of levels J and consequently also to the size of the finest problem; see Section 4.1 or the original papers [23, 37]. This is what the experiment confirms; see Figure 6.1. Contrary to that, I_2 (6.3) deteriorates with increasing J. This is in alignment with the discussion at the end of of Section 4.1, where we proved the efficiency of the estimate with a factor depending on $\sqrt{J}$.

6.2. Robustness with respect to the size of the coarsest-level problem. The second experiment describes the effect of the size of the coarsest-level problem on the efficiency of the estimates. We fix the number of levels to two ($J = 1$) and vary the coarse- and fine-level problems; see Table 6.2. For the approximation v_1 and the intermediate results computed after each V-cycle, we display the efficiency index

$$I_3 = \frac{C_{\text{numexp}} \left(\mathbf{r}_1^* \mathbf{D}_1^{-1} \mathbf{r}_1 + \eta \right)^{\frac{1}{2}}}{\|\nabla(u_1 - v_1)\|}, \tag{6.4}$$

where η denotes the following approximations to $\mathbf{r}_0^* \mathbf{A}_0^{-1} \mathbf{r}_0 = \|\nabla r_0\|^2$:

(i) $\eta = \mathbf{r}_0^* \bar{\mathbf{c}}_0$, where $\bar{\mathbf{c}}_0$ is computed using a direct solver for $\mathbf{A}_0 \mathbf{c}_0 = \mathbf{r}_0$;

(ii) $\eta = \mathbf{r}_0^* \tilde{\mathbf{c}}_0$, where $\tilde{\mathbf{c}}_0$ is computed by 4 iterations of CG for the system $\mathbf{A}_0 \mathbf{c}_0 = \mathbf{r}_0$ with a zero initial approximation;

(iii) $\eta = \dfrac{h_\Omega^2}{\min_{K \in \mathcal{T}_0} h_K^2} \mathbf{r}_0^* \mathbf{D}_0^{-1} \mathbf{r}_0$;

(iv) $\eta = \mu_i^2 + \zeta_i^2$; see (5.12) and the adaptive approach from Section 5.4 using PCG. Here ζ_i^2 is the upper bound for the $\mathbf{A}_0$-norm in the PCG method from [30, 2nd inequality in (3.5) with the updating formula for a coefficient in (3.3)]. For evaluating

ζ_i^2, an estimate of the smallest eigenvalue of $\mathbf{A}_0$ is computed by the MATLAB `eigs` function for the first four problems and extrapolated for the largest problem. In (5.11) we set $\theta = 0.1$.

TABLE 6.2
Size of the problems for the experiment in Section 6.2. The table also gives the squared ratios of the diameter of the computational domain and the coarsest-level meshsize.

coarsest-level DoFs	finest-level DoFs	$h_\Omega^2/\min_{K\in\mathcal{T}_0} h_K^2$
125	1 331	36
1 331	12 167	144
12 167	103 823	576
103 823	857 375	2 304
857 375	6 967 871	9 216

The factor $C_{\mathrm{numexp}} = 1.28$ accounts for $C_S^{\frac{1}{2}}\overline{C}_B^{\frac{1}{2}}$ and was set to a minimal value such that the efficiency index (6.4) for the variant (i) with the direct solver is above or equal to one. The results are displayed in Figure 6.2.

The variant (i), where the coarsest-level term is computed using a direct solver, exhibits only a very mild increase of the efficiency index I_3 (6.4). Recall, however, that using a direct solver is in practice unfeasible for large problems.

The variant (ii), which uses four iterations of the CG method to approximate the term on the coarsest level, provides no longer an upper bound for the algebraic error. It is not surprising that a fixed number of CG iterations is not sufficient for problems with increasing size. In the newly proposed adaptive approach, the number of CG iteration varies and is determined automatically.

For the variant (iii), where the stiffness matrix on the coarsest level is replaced by its scaled diagonal (see (5.9)), the efficiency indices deteriorate with the increasing ratio $h_\Omega^2/\min_{K\in\mathcal{T}_0} h_K^2$; see Table 6.2. The experiment illustrates that the estimate is not robust with respect to this ratio; see the discussion at the end of Section 5.3.

When the term $\|\nabla r_0\|$ is approximated using the adaptive computation (iv) proposed in Section 5.4, the efficiency behaves as in the case (i). Unlike in (i), the approximation in (iv) is computable even for very large problems on the coarsest level. The adaptively chosen number of used CG iterations within the new procedure is displayed in Figure 6.3.

6.3. Robustness with respect to the size of the coarsest-level problem—nonhomogeneous diffusion. In order to illustrate that the observed phenomena will likely be pronounced in practical applications, we replicate the previous experiment in a setting with nonhomogeneous diffusion. More specifically, we solve

$$\int_\Omega k\,\nabla u \cdot \nabla v = \int_\Omega f v, \qquad \forall v \in H_0^1(\Omega),$$

where $\Omega = (0,1)^3$, $f = 1$, and the diffusivity factor k is

$$k(x,y,z) = \begin{cases} 1024 & \text{when} \quad (x-0.5)(y-0.5)(z-0.5) > 0, \\ 1 & \text{elsewhere.} \end{cases}$$

Then, the presented results remain in principle valid apart from an additional multiplicative factor given by the ratio of the maximal and minimal values of the diffusivity factor (here equal to 1024) in some formulas. The discretization meshes are the same as in the experiment

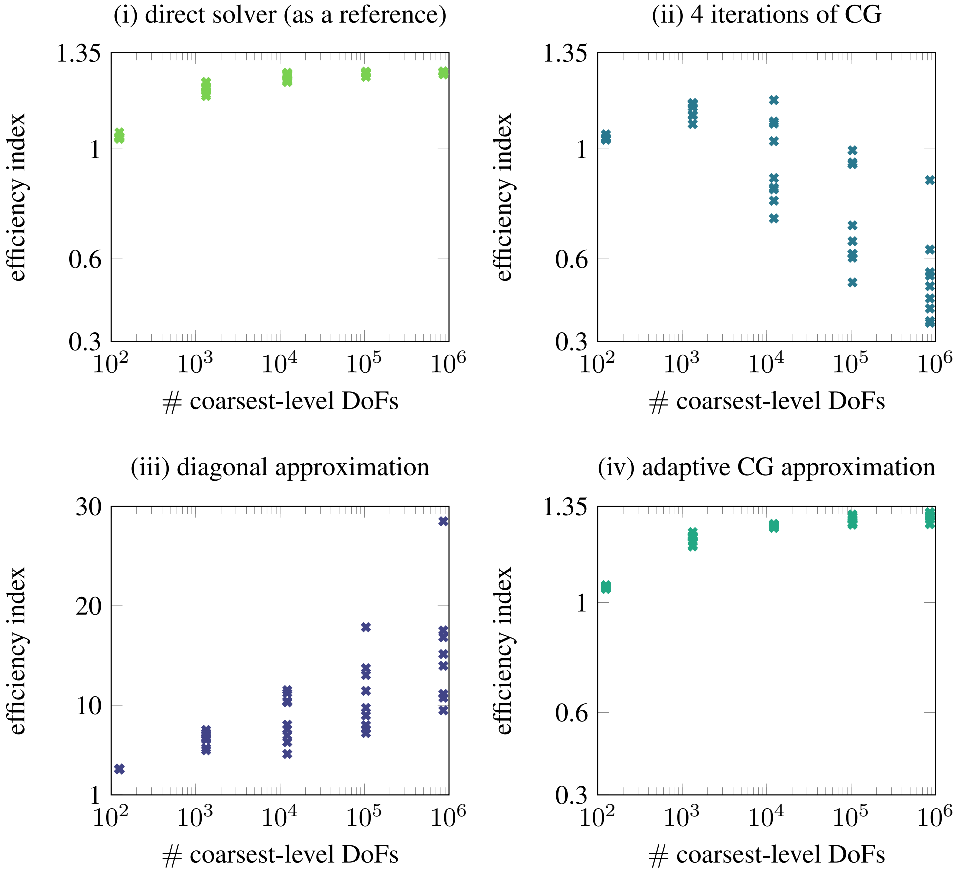

FIG. 6.2. *Efficiency indices I_3 (6.4) for the experiment in Section 6.2. The estimates differ in the way of approximating the coarsest-level term $\|\nabla r_0\|^2 = \mathbf{r}_0^* \mathbf{A}_0^{-1} \mathbf{r}_0$. This term is: computed by a direct solver for the coarsest problem (i), approximated using four iterations of the CG solver (ii), approximated by replacing the stiffness matrix by its scaled diagonal approximation (iii), computed using the adaptive CG approximation (iv).*

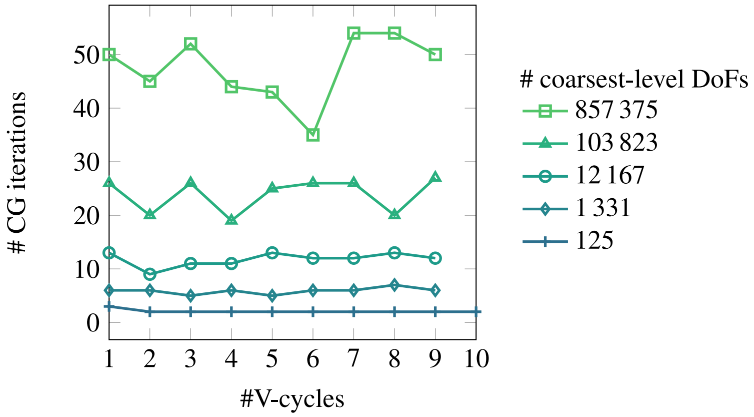

FIG. 6.3. *Number of CG iterations determined by the adaptive approach described in Section 5.4, which is used to estimate the residual norm $\|\nabla r_0\|$ associated with the coarsest level. The horizontal axis indicates the number of V-cycles used in computing the approximation v_J.*

of Section 6.2; see Table 6.2. The energy norm associated with the problem is now $\|k^{1/2}\nabla\cdot\|$ instead of $\|\nabla\cdot\|$. However, the algebraic quantities $\mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0$ and $\mathbf{r}_j^*\mathbf{D}_j^{-1}\mathbf{r}_j$, $j=0,1$, remain unchanged in the formulas with the nonhomogenous diffusion reflected in the stiffness matrices $\mathbf{A}_0$, $\mathbf{A}_1$.

The approximations v_J are generated by a V-cycle with 3 pre- and 3 post-Gauss–Seidel smoothing iterations. The V-cycle method is stopped when the relative $\mathbf{A}$-norm of the error is less than 10^{-11}. The problem on the coarsest level is now solved by the PCG method with a preconditioner based on the Cholesky factorization with zero-fill in (PCG-IC(0)). The coarsest-level solver is stopped when the relative residual drops below 0.1.

We use PCG-IC(0) also when computing the approximation of the coarsest-level term in the multilevel estimator using the adaptive approach presented in Section 5.4; see (5.10). Note that the derivation of the adaptive approach is valid also in this case.

As upper bound ζ_i^2 for $\|\mathbf{c}_0-\mathbf{c}_0^{(i)}\|^2_{\mathbf{A}_0}$ in (5.10), we use the heuristic upper bound from [29], which does not require an estimation of the smallest eigenvalue of the (preconditioned) matrix. The parameter τ prescribing the relative accuracy of the bound ζ_i (see [29, Eq. (8)]) is set to 0.1. Analogously to the previous experiment, the constant C_{numexp} accounting for $C_S^{1/2}\overline{C}_B^{1/2}$ is chosen as 1.11.

The results, analogous to Figure 6.2 and Figure 6.3, respectively, are displayed in Figure 6.4 and Figure 6.5. In particular, one can again observe the deterioration of the estimates with the fixed number of PCG iterations and with the scaled diagonal matrix (the variants (ii) and (iii), respectively). The adaptive approach from Section 5.4 again provides the estimate (iv) with the efficiency close to the reference solution from a direct solver, variant (i). As expected, the efficiency is controlled by the parameter θ from (5.11) that is here set as 0.1. In other words, the efficiency of the variant (iv) does not differ from the efficiency of the variant (i) by more then 10 percent. However, the associated estimate is computed using only a decent number of PCG iteration (see Figure 6.5) instead of using a direct solver that may not be available for large coarse-level problems.

7. Conclusions.

This paper presents residual-based a posteriori error estimates for the total and algebraic errors in multilevel frameworks inspired by several derivations from the literature. The estimates for the algebraic error contain a sum of the (scaled) residual norms over the levels, including the coarsest level. The estimates for the total error additionally incorporate the standard residual-based estimator evaluated on the finest level. For several estimates of this type, the efficiency and robustness with respect to the number of levels and the size of the algebraic problem on the coarsest level were proved in literature. However, estimates that involve scaled residual norms cannot be directly used in practice, as the norms themselves are not readily computable, and they can only be approximated.

Approximating the scaled residual norms on the fine levels, i.e., the terms $\mathbf{r}_j^*(\mathbf{M}_j^{\mathrm{S}})^{-1}\mathbf{r}_j$, where $\mathbf{M}_j^{\mathrm{S}}$ denotes the scaled mass matrix and $\mathbf{r}_j$ the algebraic residual at the level j, does not represent a significant difficulty. The term $\mathbf{r}_j^*(\mathbf{M}_j^{\mathrm{S}})^{-1}\mathbf{r}_j$ can be bounded from above by the simpler term $\mathbf{r}_j^*(\mathbf{D}_j)^{-1}\mathbf{r}_j$, where $\mathbf{D}_j$ is an appropriate diagonal matrix, without affecting the efficiency and robustness.

Dealing with the residual norm $\|\nabla r_0\|^2=\mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0$ associated with the coarsest level, where $\mathbf{A}_0$ is the stiffness matrix, is more subtle. When using bounds or techniques to approximate $\mathbf{r}_0^*\mathbf{A}_0^{-1}\mathbf{r}_0$ presented in the literature, the resulting (multilevel) estimates for the total and algebraic errors are no longer guaranteed to be independent of the size of the coarsest-level problem. This behavior is illustrated by numerical experiments.

The approach proposed in this paper approximates the coarsest-level term $\|\nabla r_0\|^2$ using the preconditioned conjugate gradient iterates. A number of PCG iterations is determined

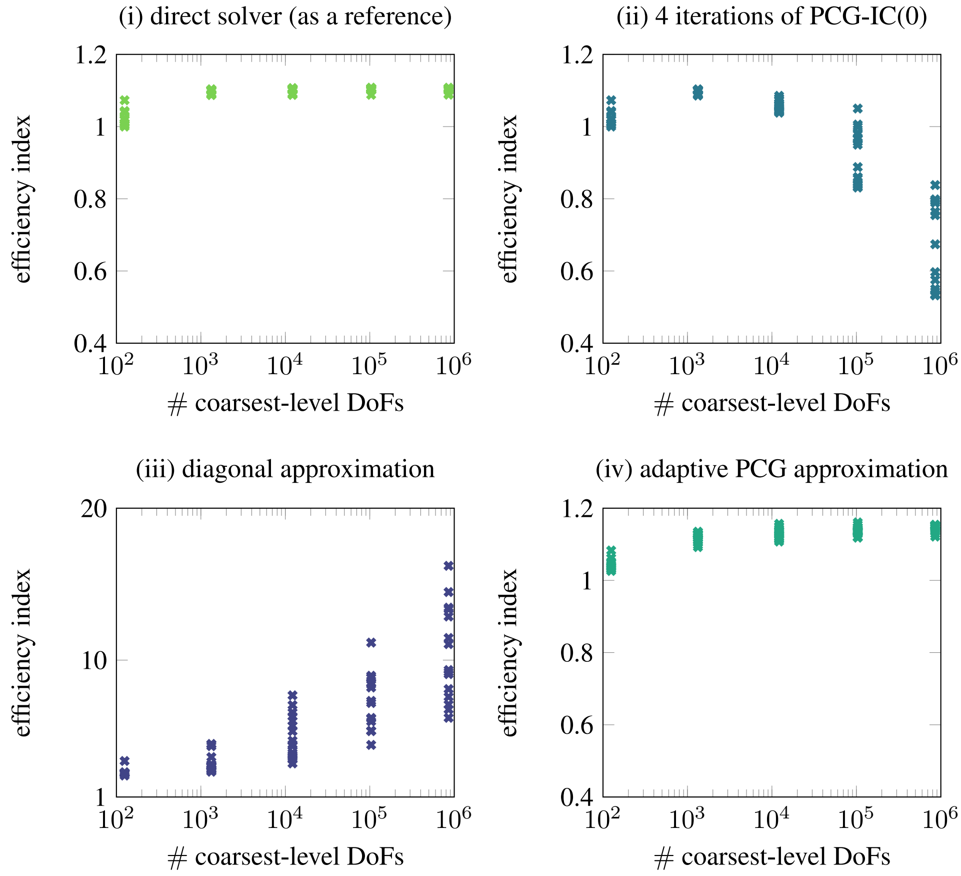

FIG. 6.4. *Efficiency indices I_3 (6.4) for the experiment in Section 6.3 with nonhomogeneous diffusion. The estimates differ in the way of approximating the coarsest-level term $\|k^{1/2}\nabla r_0\|^2 = \mathbf{r}_0^* \mathbf{A}_0^{-1}\mathbf{r}_0$. This term is: computed by a direct solver for the coarsest problem (i), approximated using four iterations of the PCG-IC(0) solver (ii), approximated by replacing the stiffness matrix by its scaled diagonal approximation (iii), computed using the adaptive PCG approximation, here with the IC(0) preconditioner (iv).*

adaptively such that the efficiency of the bound does not deteriorate with an increasing size of the coarsest-level problem, and the efficiency and robustness of the multilevel error estimates is preserved. Numerical results support the theoretical findings.

The estimates for the total and algebraic errors involve some constants that must be approximately determined, which involves heuristics. For residual-based error estimates, the constants can be determined for smaller problems with the same or analogous geometry, where an approximation with a very small algebraic error can be computed; see, e.g., the discussion in [5, Section 7]. Since the new result in Section 5.4 proves the robustness of the adaptive estimate with respect to the size of the coarsest-level problem, this justifies an extrapolation of the estimated values of the constants from smaller to larger problems.

In view of a recent trend on using multiple precision in multigrid algorithms (see, e.g., [28, 42]), it is worth considering an extension of the presented results to include effects of inexact (limited-precision) operations. This will require substantial further analysis. We plan to address this topic in the future.

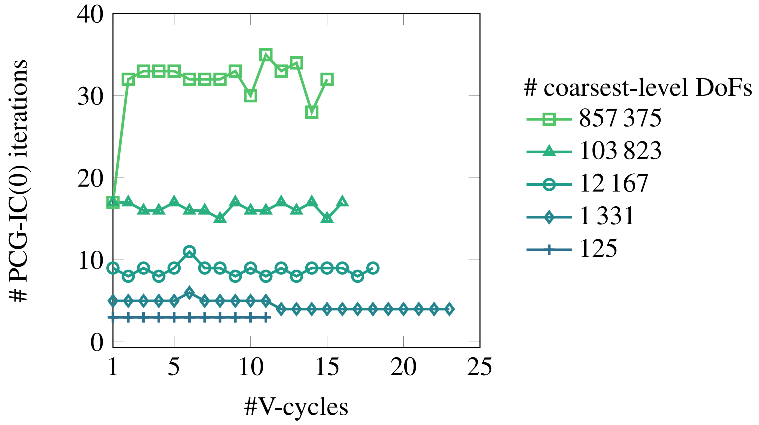

FIG. 6.5. *Number of PCG-IC(0) iterations determined by the adaptive approach described in Section 5.4, which is used to estimate the residual norm $\|k^{1/2}\nabla r_0\|$ associated with the coarsest level. The horizontal axis indicates the number of V-cycles used in computing the approximation v_J.*

Supplementary material. The code and the data for the numerical experiments are available at the GitHub repository
`https://github.com/vacek-petr/inMLEstimate`.

Acknowledgments. The authors wish to thank Peter Oswald for his help on proving Theorem C.5 and to Erin C. Carson for valuable comments. We also greatly appreciate the suggestions of the referee that improved the text, including a pointer to the references [4, 13, 18].

Jan Papež and Zdeněk Strakoš are members of the Nečas Center for Mathematical Modeling. The work of Petr Vacek was supported by the Charles University PRIMUS project no. PRIMUS/19/SCI/11, the Exascale Computing Project (17-SC-20-SC), a collaborative effort of the U.S. Department of Energy Office of Science and the National Nuclear Security Administration, and by the European Union (ERC, inEXASCALE, 101075632). The work of Jan Papež has been supported by the Czech Academy of Sciences (RVO 67985840) and by the Grant Agency of the Czech Republic (grant no. 23-06159S). Views and opinions expressed are those of the authors only and do not necessarily reflect those of the European Union or the European Research Council. Neither the European Union nor the granting authority can be held responsible for them.

Appendix A. Auxiliary results from the theory of PDEs and FEM.

The following results are standard in PDE and FEM analysis. They are presented in various forms and sometimes with different names. We provide them in forms suitable for our development, with some standard references to where the proofs can be found.

LEMMA A.1 (Bramble–Hilbert lemma). *There exists a constant $C_{\mathrm{BH}}(\mathcal{T}) > 0$ depending only on d and $\gamma_{\mathcal{T}}$ such that for all $K \in \mathcal{T}$*

$$\inf_{c\in\mathbb{R}} \|w - c\|_{\omega_K} \leq C_{\mathrm{BH}}(\mathcal{T}) h_K \|\nabla w\|_{\omega_K}, \qquad \forall w \in H^1(\omega_K),$$
$$\inf_{p\in\mathbb{P}^1(\omega_K)} \|w - p\|_{\omega_K} \leq C_{\mathrm{BH}}(\mathcal{T}) h_K^2 |w|_{H^2(\omega_K)}, \qquad \forall w \in H^2(\omega_K).$$

For the proof, see, e.g., [38, p. 490] and the references therein.

LEMMA A.2 (Friedrich's inequality). *Let $\omega \subset \mathbb{R}^d$ be a bounded domain. There exists a constant $C_F(\omega) > 0$ such that for all $w \in H^1(\omega)$ that have a zero trace on a part of the* *boundary $\partial\omega$ of nonzero measure, it holds that*

$$\|w\|_\omega \leq C_F(\omega) h_\omega \|\nabla w\|_\omega.$$

When using Friedrich's inequality on patches associated with the elements of the triangulation $\mathcal{T}$, there exists a constant $C_F(\mathcal{T})$ depending only on d and $\gamma_\mathcal{T}$ such that for all $K \in \mathcal{T}$

$$C_F(\omega_K) \leq C_F(\mathcal{T});$$

see, e.g., [36, Chapter 18].

LEMMA A.3 (Trace inequality). *There exists a constant $C_{\mathrm{TR}}(\mathcal{T}) > 0$ depending only on d and $\gamma_\mathcal{T}$ such that for all $K \in \mathcal{T}$ and all $w \in H^1(K)$ it holds that*

$$\|w\|^2_{\partial K} \leq C_{\mathrm{TR}}(\mathcal{T}) \left(h_K^{-1} \|w\|_K^2 + h_K \|\nabla w\|_K^2 \right).$$

For the proof, see, e.g., [12, Proposition 4.1].

LEMMA A.4 (Inverse inequality). *There exists a constant $C_{\mathrm{INV}}(\mathcal{T}) > 0$ depending only on d and $\gamma_\mathcal{T}$ such that for all $K \in \mathcal{T}$ and all $w_\mathcal{T} \in S_\mathcal{T}$,*

$$\|\nabla w_\mathcal{T}\|_K \leq C_{\mathrm{INV}}(\mathcal{T}) h_K^{-1} \|w_\mathcal{T}\|_K.$$

For the proof, see, e.g., [19, Lemma 1.27].

Appendix B. Quasi-interpolation operators.

A quasi-interpolation operator is not explicitly used in the construction of the estimators, but it is a crucial tool for proving the bounds. In this section we present the quasi-interpolation operator as a generalization of nodal interpolation to integrable functions. We consider the quasi-interpolation operator used in [33], which is closely related to the operator from [38]. Other, slightly different quasi-interpolation operators can be found, e.g., in [12, 15, 46]. We list and prove some of the properties of the operator to be used later. The proofs of the properties are based on standard techniques. To keep the text self-contained and formally accurate, we provide most of the proofs below.

The results in this section are mostly derived for a single mesh $\mathcal{T}$. We show that the constants only depend on the dimension d and the shape-regularity $\gamma_\mathcal{T}$, and therefore we can again use them in the mesh hierarchy with the dependence on d and γ_0.

B.1. Nodal interpolation and its generalization.

For a node $z \in \mathcal{N}_\mathcal{T}$, let $\Psi_z : C(\overline{\Omega}) \to \mathbb{R}$ denote the linear functional evaluation at point z, i.e.,

$$\Psi_z(w) = w(z), \qquad \forall w \in C(\overline{\Omega}).$$

The standard nodal interpolation operator $\mathcal{I} : C(\overline{\Omega}) \to S_\mathcal{T}$ for continuous functions is defined as (see, e.g., [8, 14])

$$\mathcal{I}w = \sum_{z \in \mathcal{N}_\mathcal{T}} \Psi_z(w) \phi_z, \qquad \forall w \in C(\overline{\Omega}).$$

Recall that ϕ_z is the continuous piecewise linear basis function taking the value one at the node z and vanishing at all the other nodes. In order to construct an analogy of the operator $\mathcal{I}$ for functions from $L^1(\Omega)$, the point evaluation is replaced by an appropriate average of the approximated function. We will consider the quasi-interpolation operator defined in [33] and [39].

For a node $z \in \mathcal{N}_{\mathcal{T}}$, let K_z be a fixed element having z as its vertex, i.e., $z \in K_z$. Let $\mathbb{P}^1(K_z)$ denote the space of linear polynomials on K_z, and denote by $\widetilde{\Psi}_z$ the restriction of the linear functional Ψ_z to functions from $\mathbb{P}^1(K_z)$. Since $\mathbb{P}^1(K_z)$ is a finite-dimensional space, the linear functional $\widetilde{\Psi}_z$ is bounded, and it therefore belongs to the dual space $(\mathbb{P}^1(K_z))^\star$. Considering the space $\mathbb{P}^1(K_z)$ equipped with the L^2-inner product, the Riesz representation theorem (see, e.g., [8, Section 2.4]) yields the existence of a function $\psi_z \in \mathbb{P}^1(K_z)$ such that

$$\widetilde{\Psi}_z(w) = w(z) = \int_{K_z} w\psi_z, \qquad \forall w \in \mathbb{P}^1(K_z).$$

Since ψ_z is the Riesz representation of the point evaluation at z, it holds for all $z_1, z_2 \in \mathcal{N}_{\mathcal{T}}$ (recall that ϕ_{z_2} is the hat function associated with z_2) that

$$\int_{K_{z_1}} \phi_{z_2}\psi_{z_1} = \phi_{z_2}(z_1) = \begin{cases} 1 & z_1 = z_2, \\ 0 & z_1 \neq z_2. \end{cases} \tag{B.1}$$

We will consider the quasi-interpolation operators defined as follows:

$$\begin{aligned} I_{S_{\mathcal{T}}} &: L^1(\Omega) \to S_{\mathcal{T}}, \quad I_{S_{\mathcal{T}}} w = \sum_{z\in\mathcal{N}_{\mathcal{T}}} \left(\int_{K_z} w\psi_z\right)\phi_z, \\ I_{V_{\mathcal{T}}} &: L^1(\Omega) \to V_{\mathcal{T}}, \quad I_{V_{\mathcal{T}}} w = \sum_{z\in\mathcal{K}_{\mathcal{T}}} \left(\int_{K_z} w\psi_z\right)\phi_z. \end{aligned}$$

These definitions and relation (B.1) imply that $I_{S_{\mathcal{T}}}$ and $I_{V_{\mathcal{T}}}$ are projections onto $S_{\mathcal{T}}$ and $V_{\mathcal{T}}$, respectively. Further, $I_{S_{\mathcal{T}}}$ preserves linear polynomials on Ω, and $I_{V_{\mathcal{T}}}$ preserves linear polynomials on ω_K for any element $K \in \mathcal{T}$ whose patch ω_K does not intersect with the boundary of Ω, i.e., $\overline{\omega_K} \cap \partial\Omega = \emptyset$.

B.2. Local estimates. We now present local (elementwise) bounds for an interpolant $I_{S_{\mathcal{T}}} w$ and the interpolation error $w - I_{S_{\mathcal{T}}} w$.

THEOREM B.1. *There exist positive constants $\widehat{C}_{I_{S_{\mathcal{T}}},\ell}$, $\ell = 1,2,3,4$, depending only on d and $\gamma_{\mathcal{T}}$ such that for all elements $K \in \mathcal{T}$,*

$$\|I_{S_{\mathcal{T}}} w\|_K \leq \widehat{C}_{I_{S_{\mathcal{T}}},1}\|w\|_{\omega_K}, \qquad \forall w \in L^2(\omega_K), \tag{B.2}$$

$$\|w - I_{S_{\mathcal{T}}} w\|_K \leq \widehat{C}_{I_{S_{\mathcal{T}}},2} h_K \|\nabla w\|_{\omega_K}, \qquad \forall w \in H^1(\omega_K), \tag{B.3}$$

$$\|w - I_{S_{\mathcal{T}}} w\|_K \leq \widehat{C}_{I_{S_{\mathcal{T}}},3} h_K^2 |w|_{H^2(\omega_K)}, \qquad \forall w \in H^2(\omega_K), \tag{B.4}$$

$$\|\nabla I_{S_{\mathcal{T}}} w\|_K \leq \widehat{C}_{I_{S_{\mathcal{T}}},4}\|\nabla w\|_{\omega_K}, \qquad \forall w \in H^1(\omega_K). \tag{B.5}$$

Proof. The steps in the proof are inspired by [33, pp. 17–18] and [38, Sections 3–4]. Using a standard affine transformation to a reference element, it can be shown that there exists a constant $C_\psi > 0$ depending only on d and $\gamma_{\mathcal{T}}$ such that for all $z \in \mathcal{N}_{\mathcal{T}}$,

$$\|\psi_z\|_{L^\infty(K_z)} \leq C_\psi |K_z|^{-1}, \tag{B.6}$$

and that there exists a constant $C_\phi > 0$ depending only on d and $\gamma_{\mathcal{T}}$ such that for all $K \in \mathcal{T}$ and all $z \in \mathcal{K}_K$,

$$\|\nabla\phi_z\|_{L^\infty(K)} \leq C_\phi \rho_K^{-1}; \tag{B.7}$$

see, e.g., [38, pp. 487–488].

Using Hölder's inequality and (B.6), we can show that for all $z \in \mathcal{N}_{\mathcal{T}}$ and all $w \in L^2(K_z)$

$$\begin{aligned} \left| \int_{K_z} w\psi_z \right|^2 &\le \|\psi_z\|^2_{L^\infty(K_z)} \left(\int_{K_z} |w| \right)^2 \\ &\le C_\psi^2 |K_z|^{-2} |K_z| \|w\|^2_{K_z} = C_\psi^2 |K_z|^{-1} \|w\|^2_{K_z}. \end{aligned} \tag{B.8}$$

We now proceed to prove the inequality (B.2). Using that $0 \le \phi_z \le 1$ gives

$$\begin{aligned} \|I_{S_\mathcal{T}} w\|^2_K &= \left\| \sum_{z \in \mathcal{N}_K} \left(\int_{K_z} w\psi_z \right) \phi_z \right\|^2_K \le \left| \sum_{z \in \mathcal{N}_K} \int_{K_z} w\psi_z \right|^2 |K| \\ &\le (\#\mathcal{N}_K)|K| \sum_{z \in \mathcal{N}_K} \left| \int_{K_z} w\psi_z \right|^2. \end{aligned}$$

The inequality (B.8) and the fact that $\#\mathcal{N}_K \le d+1$ yields

$$\begin{aligned} \|I_{S_\mathcal{T}} w\|^2_K &\le (d+1)|K| \sum_{z \in \mathcal{K}_K} C_\psi^2 |K_z|^{-1} \|w\|^2_{K_z} \le (d+1)|K| C_\psi^2 \max_{z \in \mathcal{N}_K} |K_z|^{-1} \|w\|^2_{\omega_K} \\ &\le (d+1) C_\psi^2 \frac{|K|}{\min_{z \in \mathcal{N}_K} |K_z|} \|w\|^2_{\omega_K}. \end{aligned}$$

Since $|K|$ and $|K_z|$, $z \in \mathcal{N}_K$, are comparable up to a constant depending on d and $\gamma_\mathcal{T}$ (in a shape-regular mesh, we can compare the size of any neighboring elements), inequality (B.2) follows.

To prove the inequalities (B.3) and (B.4), let p be a constant or linear polynomial on ω_K. Using the fact that $I_{S_\mathcal{T}}$ reproduces linear polynomials and (B.2), we get

$$\begin{aligned} \|w - I_{S_\mathcal{T}} w\|_K &= \|w - p - I_{S_\mathcal{T}}(w-p)\|_K \le \|w-p\|_K + \widehat{C}_{I_{S_\mathcal{T}},1} \|w-p\|_{\omega_K} \\ &\le (\widehat{C}_{I_{S_\mathcal{T}},1} + 1) \|w-p\|_{\omega_K}. \end{aligned}$$

Using the Bramble–Hilbert lemma (Theorem A.1) gives

$$\|w - I_{S_\mathcal{T}} w\|_K \le (\widehat{C}_{I_{S_\mathcal{T}},1} + 1) C_{\mathrm{BH}}(\mathcal{T}) h_K \|\nabla w\|_{\omega_K}$$

or

$$\|w - I_{S_\mathcal{T}} w\|_K \le (\widehat{C}_{I_{S_\mathcal{T}},1} + 1) C_{\mathrm{BH}}(\mathcal{T}) h_K^2 |w|_{H^2(\omega_K)}.$$

It remains to verify the inequality (B.5). Using the fact that $I_{S_\mathcal{T}}$ reproduces constants, we have, for arbitrary $c \in \mathbb{R}$,

$$\begin{aligned} \|\nabla I_{S_\mathcal{T}} w\|^2_K = \|\nabla I_{S_\mathcal{T}}(w-c)\|^2_K &= \int_K \left| \sum_{z \in \mathcal{N}_K} \left(\int_{K_z} (w-c)\psi_z \right) \nabla\phi_z \right|^2 \\ &\le (\#\mathcal{N}_K) \sum_{z \in \mathcal{N}_K} \|\nabla\phi_z\|^2_{L^\infty(K)} \int_K \left| \int_{K_z} (w-c)\psi_z \right|^2 \\ &\le (d+1) \sum_{z \in \mathcal{N}_K} \|\nabla\phi_z\|^2_{L^\infty(K)} \left| \int_{K_z} (w-c)\psi_z \right|^2 |K| \end{aligned}$$

$$\leq (d+1)C_\phi^2 \rho_K^{-2} |K| \sum_{z \in \mathcal{N}_K} \left| \int_{K_z} (w-c)\psi_z \right|^2 ,$$

where we also used (B.7). Then, from (B.8), we get

$$\|\nabla I_{S_\mathcal{T}} w\|_K^2 \leq (d+1)C_\phi^2 \rho_K^{-2} |K| C_\psi^2 \max_{z \in \mathcal{N}_K} |K_z|^{-1} \|w-c\|_{\omega_K}^2 .$$

Using the Bramble–Hilbert lemma (Theorem A.1) and rearranging yields

$$\|\nabla I_{S_\mathcal{T}} w\|_K^2 \leq (d+1)C_\phi^2 C_\psi^2 \big(C_{\mathrm{BH}}(\mathcal{T})\big)^2 \frac{|K|}{\min_{z\in\mathcal{K}_K} |K_z|} \cdot \frac{h_K^2}{\rho_K^2} \|\nabla w\|_{\omega_K}^2 . \qquad \square$$

For the interpolation operator $I_{V_\mathcal{T}}$, we can derive bounds analogous to those of Theorem B.1. For the "inner" elements, i.e., the elements $K \in \mathcal{T}$ such that patch ω_K does not intersect with the boundary of Ω, i.e., $\overline{\omega_K} \cap \partial\Omega = \emptyset$, the forms of the bounds and their proofs are analogous to Theorem B.1, because $I_{V_\mathcal{T}}$ also reproduces constants on ω_K. For the elements whose patch intersects with the boundary of Ω, one cannot use this property, and the Bramble–Hilbert lemma (Theorem A.1) must be replaced by Friedrich's inequality (Theorem A.2) in the proofs.

THEOREM B.2. *There exist positive constants $\widehat{C}_{I_{V_\mathcal{T}},\ell}$, $\ell = 1, 2, 4$, depending only on d and $\gamma_\mathcal{T}$ such that for all elements $K \in \mathcal{T}$,*

$$\|I_{V_\mathcal{T}} w\|_K \leq \widehat{C}_{I_{V_\mathcal{T}},1} \|w\|_{\omega_K}, \qquad \forall w \in L^2(\omega_K),$$

and for all $w \in H^1(\omega_K)$, if $\overline{\omega_K} \cap \partial\Omega = \emptyset$, or otherwise for all $w \in H^1(\omega_K) \cap H_0^1(\Omega)$,

$$\begin{aligned} \|w - I_{V_\mathcal{T}} w\|_K &\leq \widehat{C}_{I_{V_\mathcal{T}},2} h_K \|\nabla w\|_{\omega_K}, \\ \|\nabla I_{V_\mathcal{T}} w\|_K &\leq \widehat{C}_{I_{V_\mathcal{T}},4} \|\nabla w\|_{\omega_K}. \end{aligned}$$

For the local interpolation error over the faces, we have the following bound:

THEOREM B.3. *There exists a positive constant $\widehat{C}_{I_{V_\mathcal{T}},5}$ depending only on d and $\gamma_\mathcal{T}$ such that for all elements $K \in \mathcal{T}$,*

$$\|w - I_{V_\mathcal{T}} w\|_{\partial K}^2 \leq \widehat{C}_{I_{V_\mathcal{T}},5} h_K \|\nabla w\|_{\omega_K}^2 .$$

Proof. Using the trace inequality (Theorem A.3) and the properties of $I_{V_\mathcal{T}}$ from Theorem B.2 yields

$$\begin{aligned} \|w - I_{V_\mathcal{T}} w\|_{\partial K} &\leq C_{\mathrm{TR}}(\mathcal{T}) \left[h_K^{-1} \|w - I_{V_\mathcal{T}} w\|_K^2 + h_K \|\nabla (w - I_{V_\mathcal{T}} w)\|_K^2 \right] \\ &\leq C_{\mathrm{TR}}(\mathcal{T}) \left[h_K^{-1} \|w - I_{V_\mathcal{T}} w\|_K^2 + 2h_K \left(\|\nabla w\|_K^2 + \|\nabla I_{V_\mathcal{T}} w\|_K^2 \right) \right] \\ &\leq C_{\mathrm{TR}}(\mathcal{T}) \Bigg[h_K^{-1} \big(\widehat{C}_{I_{V_\mathcal{T}},2}\big)^2 h_K^2 \|\nabla w\|_{\omega_K}^2 + \\ &\qquad\qquad + 2h_K \left(1 + \big(\widehat{C}_{I_{V_\mathcal{T}},4}\big)^2 \right) \|\nabla w\|_{\omega_K}^2 \Bigg] . \qquad \square \end{aligned}$$

B.3. Global estimates. We now state global variants of estimates for the quasi-interpolants and interpolation errors. For $K \in \mathcal{T}$, let $C_{\mathrm{ovrlp}}(K)$ denote the number of patches that this element is contained in, i.e.,

$$C_{\mathrm{ovrlp}}(K) = \# \left\{ K' \in \mathcal{T}; K \subset \omega_{K'} \right\} .$$

The constant $C_{\mathrm{ovrlp}}(K)$ depends only on the geometry of the mesh $\mathcal{T}$, i.e., d and the shape regularity $\gamma_\mathcal{T}$.

THEOREM B.4. *There exist positive constants $C_{I_{S_\mathcal{T}},\ell}$, $\ell = 1, 2, 4$, depending only on d and $\gamma_\mathcal{T}$ such that*

$$\begin{aligned} \|I_{S_\mathcal{T}} w\| &\le C_{I_{S_\mathcal{T}},1}\|w\|, && \forall w \in L^2(\Omega),\\ \left(\sum_{K\in\mathcal{T}} h_K^{-2}\|w - I_{S_\mathcal{T}} w\|_K^2\right)^{\frac{1}{2}} = \|h_\mathcal{T}^{-1}(w - I_{S_\mathcal{T}} w)\| &\le C_{I_{S_\mathcal{T}},2}\|\nabla w\|, && \forall w \in H^1(\Omega),\\ \|\nabla(I_{S_\mathcal{T}} w)\| &\le C_{I_{S_\mathcal{T}},4}\|\nabla w\|, && \forall w \in H^1(\Omega). \end{aligned}$$

Proof. Using Theorem B.1,

$$\begin{aligned} \|I_{S_\mathcal{T}} w\|^2 &= \sum_{K\in\mathcal{T}} \|I_{S_\mathcal{T}} w\|_K^2 \le \sum_{K\in\mathcal{T}} \big(\widehat{C}_{I_{S_\mathcal{T}},1}\big)^2 \|w\|_{\omega_K}^2 \\ &\le \big(\widehat{C}_{I_{S_\mathcal{T}},1}\big)^2 \sum_{K\in\mathcal{T}} C_{\mathrm{ovrlp}}(K)\|w\|_K^2 \le \big(\widehat{C}_{I_{S_\mathcal{T}},1}\big)^2 \max_{K\in\mathcal{T}} C_{\mathrm{ovrlp}}(K) \sum_{K\in\mathcal{T}} \|w\|_K^2. \end{aligned}$$

The proofs of the other three inequalities are analogous. □

THEOREM B.5. *There exist positive constants $C_{I_{V_\mathcal{T}},\ell}$, $\ell = 1, 2, 4, 5$, depending only on d and the shape-regularity constant $\gamma_\mathcal{T}$ such that*

$$\|I_{V_\mathcal{T}} w\| \le C_{I_{V_\mathcal{T}},1}\|w\|, \quad \forall w \in L^2(\Omega),$$

$$\left(\sum_{K\in\mathcal{T}} h_K^{-2}\|w - I_{V_\mathcal{T}} w\|_K^2\right)^{\frac{1}{2}} = \|h_\mathcal{T}^{-1}(w - I_{V_\mathcal{T}} w)\| \le C_{I_{V_\mathcal{T}},2}\|\nabla w\|, \ \forall w \in H_0^1(\Omega), \tag{B.9}$$

$$\|\nabla(I_{V_\mathcal{T}} w)\| \le C_{I_{V_\mathcal{T}},4}\|\nabla w\|, \ \forall w \in H_0^1(\Omega), \tag{B.10}$$

$$\left(\sum_{K\in\mathcal{T}} h_K^{-1}\|w - I_{V_\mathcal{T}} w\|_{\partial K}^2\right)^{\frac{1}{2}} \le C_{I_{V_\mathcal{T}},5}\|\nabla w\|, \ \forall w \in H_0^1(\Omega).$$

Let us now consider the mesh hierarchy as in Section 2.2. Since the constants $C_{I_{S_j},\ell}$ and $C_{I_{V_j},\ell}$ depend only on d and γ_j, they can be bounded by constants $C_{I_S,\ell}$ and $C_{I_V,\ell}$ depending only on d and the shape regularity γ_0 of the initial mesh $\mathcal{T}_0$.

Finally, we bound the difference of quasi-interpolates on two consecutive levels.

THEOREM B.6. *There exists a constant $C_{I,\mathrm{2lvl}} > 0$ depending only on d and γ_0 such that for all $j \ge 1$ and all $w \in H_0^1(\Omega)$,*

$$\|h_j^{-1}(I_{V_j} w - I_{V_{j-1}} w)\| \le C_{I,\mathrm{2lvl}}\|\nabla w\|.$$

Proof. Using the fact that $h_j^{-1} = 2h_{j-1}^{-1}$ and the estimate (B.9) from Theorem B.5,

$$\begin{aligned} \|h_j^{-1}(I_{V_j} w - I_{V_{j-1}} w)\| &\le \|h_j^{-1}(w - I_{V_j} w)\| + \|h_j^{-1}(w - I_{V_{j-1}} w)\| \\ &= \|h_j^{-1}(w - I_{V_j} w)\| + 2\|h_{j-1}^{-1}(w - I_{V_{j-1}} w)\| \\ &\le (C_{I_V,2} + 2C_{I_V,2})\|\nabla w\|. \end{aligned}$$

Taking $C_{I,\mathrm{2lvl}}$ as $C_{I,\mathrm{2lvl}} = C_{I_V,2} + 2C_{I_V,2}$ finishes the proof. □

Appendix C. Stable splitting. This section presents several results on the splitting (decomposing) of a $H_0^1(\Omega)$-function or a piecewise polynomial function into a sum of piecewise polynomial functions. Let a sequence of uniformly refined meshes $\mathcal{T}_j$, $j = 0, 1, \ldots$, as in Section 2.2 be given.

C.1. Splitting of $H_0^1(\Omega)$ into subspaces of piecewise linear functions. To make the text easier to follow, we first state the main result of this section and subsequently provide auxiliary results and proofs. We will show that any function $w \in H_0^1(\Omega)$ can uniquely be decomposed using the quasi-interpolation operators I_{V_j}, $j \in \mathbb{N}_0$, as

$$w = I_{V_0} w + \sum_{j=1}^{+\infty} (I_{V_j} - I_{V_{j-1}}) w;$$

the convergence of the sum is understood in the space $H_0^1(\Omega)$ with the norm $\|\nabla \cdot\|$. This decomposition is stable, meaning that there exist positive constants c_{S,I_V}, C_{S,I_V} such that for all $w \in H_0^1(\Omega)$,

$$c_{S,I_V}\|\nabla w\|^2 \leq \|\nabla I_{V_0} w\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1}(I_{V_j} w - I_{V_{j-1}} w)\|^2 \leq C_{S,I_V}\|\nabla w\|^2. \tag{C.1}$$

We will also show that the splitting of the space $H_0^1(\Omega)$ into subspaces V_j, $j \in \mathbb{N}_0$, is stable in the sense that there exist positive constants c_S, C_S such that for all $w \in H_0^1(\Omega)$,

$$c_S\|\nabla w\|^2 \leq \inf_{w_j \in V_j;\ w = \sum_{j=0}^{+\infty} w_j} \|\nabla w_0\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1} w_j\|^2 \leq C_S\|\nabla w\|^2. \tag{C.2}$$

The infimum is taken over all $(H_0^1(\Omega), \|\nabla \cdot\|)$-convergent decompositions.

We will show that the stability constants c_{S,I_V}, C_{S,I_V}, and c_S, C_S depend *only* on d and the shape regularity γ_0 of the initial mesh. In particular, the constants do not depend on the quasi-uniformity of the initial mesh or the ratio $h_\Omega / \min_{K \in \mathcal{T}_0} h_K$. This result is important when considering settings where the problem associated with the coarsest level is difficult to solve and where it can only be solved approximately in practice.

Variants of these results can be found, e.g., in [6, 16, 17, 33, 37] and the references therein. Our form is, however, to the best of our knowledge, not presented in the literature. The results in [16, 17, 33, 37] are derived under the assumption that the initial mesh is quasi-uniform, and the authors do not track the dependence of the constants on $h_\Omega / \min_{K \in \mathcal{T}_0} h_K$. The results of [6] are derived without the assumption on the quasi-uniformity of the initial mesh. The authors however consider only the splitting of piecewise linear functions. We combine the approaches from [33] and [6]. We first focus on showing the upper bound from (C.1), then continue with the lower bound, and later generalize it to show (C.2).

First, consider the K-functional in analogy to [6, Section 7, Eq. (7.4)]. For $\omega \subset \mathbb{R}^d$, $w \in L^2(\omega)$, it is defined as

$$K(t, w, \omega) = \inf_{g \in H^2(\omega)} \left\{ \|w - g\|^2_{L^2(\omega)} + t^2 |g|^2_{H^2(\omega)} \right\}^{\frac{1}{2}}, \qquad t > 0.$$

LEMMA C.1. *There exists a constant $C > 0$ such that for all $w \in H^1(\mathbb{R}^d)$ that have compact support in $\mathbb{R}^d$, it holds that*

$$\sum_{j=0}^{+\infty} 2^{2j} K(2^{-2j}, w, \mathbb{R}^d)^2 \leq C\|\nabla w\|^2_{L^2(\mathbb{R}^d)}.$$

Proof. A short proof for $d = 2$ is given in [6, Lemma 7.3]. We present its key part in more detail and for $d = 2, 3$. We will show that the K-functional can be expressed in terms of the Fourier transform (here denoted by $F[\cdot]$) as

$$K(t,w,\mathbb{R}^d)^2 = \frac{1}{(2\pi)^{\frac{d}{2}}}\int_{\mathbb{R}^d} \frac{t^2|\xi|^4}{1+t^2|\xi|^4}\big|F[w](\xi)\big|^2 d\xi. \tag{C.3}$$

For $w \in H^1(\mathbb{R}^d)$, $g \in H^2(\mathbb{R}^d)$, using the properties of the Fourier transform,

$$\begin{aligned} &\|w-g\|^2_{L^2(\mathbb{R}^d)} + t^2|g|^2_{H^2(\mathbb{R}^d)} \\ &\quad = \frac{1}{(2\pi)^{\frac{d}{2}}}\int_{\mathbb{R}^d} \frac{t^2|\xi|^4}{1+t^2|\xi|^4}\big|F[w](\xi)\big|^2 + (1+t^2|\xi|^4)\left(F[g](\xi) - \frac{F[w](\xi)}{1+t^2|\xi|^4}\right)^2 d\xi. \end{aligned} \tag{C.4}$$

By simple manipulations, one can show that the minimum is attained for

$$\widetilde{g}(x) = w(x) * F^{-1}\left[\frac{1}{1+t^2|\xi|^4}\right](x),$$

and it remains to show that $\widetilde{g} \in H^2(\mathbb{R}^d)$. First, note that

$$\int_{\mathbb{R}^d} (1+|\xi|)^2 \left(\frac{1}{1+t^2|\xi|^4}\right)^2 < \infty,$$

and therefore, due to the characterization of Sobolev spaces using the Fourier transformation (see, e.g., [33, Section 3.1.1]), it holds that $F^{-1}[(1+t^2|\xi|^4)^{-1}](x) \in H^2(\mathbb{R}^d)$. Then use Young's inequality for a convolution (recall that by assumption, w is compactly supported, and therefore $w \in L^1(\mathbb{R}^d)$) and the fact that $\partial/\partial\xi_i(f * h) = (\partial f/\partial \xi_i * h)$ to show that the H^2-norm of $\widetilde{g}$ is bounded.

The equality (C.3) then follows by plugging in the expression for $\widetilde{g}$ into (C.4) and performing algebraic manipulations. The rest of the proof of the lemma follows as in [6, Lemma 7.3]. □

LEMMA C.2. *Let $\omega \subset \mathbb{R}^d$ be a domain with a Lipschitz-continuous boundary. There exists a constant $C_\alpha(\omega) > 0$ depending on the shape of ω such that for all $w \in H^1(\omega)$,*

$$\sum_{j=0}^{+\infty} 2^{2j} K(2^{-2j}, w, \omega)^2 \leq C_\alpha(\omega)\|\nabla w\|^2_{L^2(\omega)}.$$

Proof. The proof for $d = 2$ is given in [6, Lemma 7.4]. It is based on the use of an extension operator and Theorem C.1. For the three-dimensional case, the proof is analogous, since Theorem C.1 is also valid for $d = 3$. □

LEMMA C.3. *There exists a constant $C_\beta > 0$ depending only on d and γ_0 such that for all $K \in \mathcal{T}_0$ and all $w \in H^1_0(\Omega)$,*

$$h_K^{-2}\sum_{j=0}^{+\infty} 2^{2j}\|w - I_{S_j}w\|^2_K \leq C_\beta \|\nabla w\|^2_{\omega_K}.$$

Proof. The steps in the proof are inspired by the development in [33, Section 2.3] and [6, Section 7]. We use a scaling argument to consider an element $\widetilde{K}$ with $h_{\widetilde{K}} = 1$. This is done by a transformation $x = h_K\widetilde{x}$, where $x \in K$, $\widetilde{x} \in \widetilde{K}$. We denote $\widetilde{f}(\widetilde{x}) := f(x)$ for any function f defined on ω_K. Then

$$\|w - I_{S_j}w\|^2_K = h_K^d\|\widetilde{w} - \widetilde{I_{S_j}w}\|^2_{\widetilde{K}}. \tag{C.5}$$

From the definition of the interpolation operator one can write $\widetilde{I_{S_j} w} = \widetilde{I_{S_j}}\,\widetilde{w}$. In words, one can either consider the transformation of the interpolant $I_{S_j} w$ or transform the function w to the element $\widetilde{K}$ first and then consider the quasi-interpolation $\widetilde{I_{S_j}}$ associated with the transformed mesh.

We will show that there exists a constant $C_\delta > 0$ depending only on d and γ_0 such that

$$\|\widetilde{w} - \widetilde{I_{S_j}}\widetilde{w}\|_{\widetilde{K}}^2 \le C_\delta \cdot \left(K(2^{-2j}, \widetilde{w}, \omega_{\widetilde{K}})\right)^2. \tag{C.6}$$

Let $\widetilde{g} \in H^2(\omega_{\widetilde{K}})$. Then,

$$\|\widetilde{w} - \widetilde{I_{S_j}}\widetilde{w}\|_{\widetilde{K}} \le \|\widetilde{w} - \widetilde{g}\|_{\widetilde{K}} + \|\widetilde{g} - \widetilde{I_{S_j}}\widetilde{g}\|_{\widetilde{K}} + \|\widetilde{I_{S_j}}(\widetilde{g} - \widetilde{w})\|_{\widetilde{K}}. \tag{C.7}$$

Let $\widetilde{K}_j \in \widetilde{\mathcal{T}}_j$ such that $\widetilde{K}_j \subset \widetilde{K}$. Then (thanks to the uniform refinement and $h_{\widetilde{K}} = 1$, $h_{\widetilde{K}_j} = 2^{-j}$), from Theorem B.1 (inequalities (B.2) and (B.4)),

$$\|\widetilde{I_{S_j}}(\widetilde{g} - \widetilde{w})\|_{\widetilde{K}_j} \le \widehat{C}_{\widetilde{I_{S_j}},1}\|\widetilde{g} - \widetilde{w}\|_{\omega_{\widetilde{K}_j}}, \tag{C.8}$$

$$\|\widetilde{g} - \widetilde{I_{S_j}}\widetilde{g}\|_{\widetilde{K}_j} \le \widehat{C}_{\widetilde{I_{S_j}},3}2^{-2j}|\widetilde{g}|_{H^2(\omega_{\widetilde{K}_j})}. \tag{C.9}$$

Define

$$U(\widetilde{K}, j) = \left\{\bigcup \omega_{\widetilde{K}_j}; \widetilde{K}_j \in \widetilde{\mathcal{T}}_j, \widetilde{K}_j \subset \widetilde{K}\right\}.$$

The term on the right-hand side of (C.8) can be bounded as

$$\begin{aligned}
\|\widetilde{I_{S_j}}(\widetilde{g} - \widetilde{w})\|_{\widetilde{K}} &= \sum_{\widetilde{K}_j \in \widetilde{\mathcal{T}}_j, \widetilde{K}_j \subset \widetilde{K}} \|\widetilde{I_{S_j}}(\widetilde{g} - \widetilde{w})\|_{\widetilde{K}_j} \le \sum_{\widetilde{K}_j \in \widetilde{\mathcal{T}}_j, \widetilde{K}_j \subset \widetilde{K}} \widehat{C}_{\widetilde{I_{S_j}},1}\|\widetilde{g} - \widetilde{w}\|_{\omega_{\widetilde{K}_j}} \\
&\le \widehat{C}_{\widetilde{I_{S_j}},1} \max_{\widetilde{K}_j \in \widetilde{\mathcal{T}}_j; \widetilde{K}_j \in U(\widetilde{K},j)} C_{\mathrm{ovrlp}}(\widetilde{K}_j)\|\widetilde{g} - \widetilde{w}\|_{U(\widetilde{K},j)} \\
&\le \widehat{C}_{\widetilde{I_{S_j}},1} \max_{\widetilde{K}_j \in \widetilde{\mathcal{T}}_j; \widetilde{K}_j \in U(\widetilde{K},j)} C_{\mathrm{ovrlp}}(\widetilde{K}_j)\|\widetilde{g} - \widetilde{w}\|_{\omega_{\widetilde{K}}} \\
&\le C_{I_S,1}\|\widetilde{g} - \widetilde{w}\|_{\omega_{\widetilde{K}}},
\end{aligned} \tag{C.10}$$

where the last inequality follows from the fact that $\widehat{C}_{\widetilde{I_{S_j}},1} = \widehat{C}_{I_{S_j},1}$ (scaling does not change the geometry and the shape regularity) and from the definition of $C_{I_S,1}$. The term on the right-hand side of (C.9) can be bounded as

$$\begin{aligned}
\|\widetilde{g} - \widetilde{I_{S_j}}\widetilde{g}\|_{\widetilde{K}}^2 &= \sum_{\widetilde{K}_j \in \widetilde{\mathcal{T}}_j, \widetilde{K}_j \subset \widetilde{K}} \|\widetilde{g} - \widetilde{I_{S_j}}\widetilde{g}\|_{\widetilde{K}_j}^2 \le \sum_{\widetilde{K}_j \in \widetilde{\mathcal{T}}_j, \widetilde{K}_j \subset \widetilde{K}} \left(\widehat{C}_{\widetilde{I_{S_j}},3}2^{-2j}|\widetilde{g}|_{\omega_{\widetilde{K}_j}}\right)^2 \\
&\le \left(\widehat{C}_{\widetilde{I_{S_j}},3}\right)^2 \max_{\widetilde{K}_j \in \widetilde{\mathcal{T}}_j, \widetilde{K}_j \in U(\widetilde{K},j)} C_{\mathrm{ovrlp}}(\widetilde{K}_j)2^{-4j}|\widetilde{g}|_{H^2(U(\widetilde{K},j))}^2 \\
&\le \left(\widehat{C}_{\widetilde{I_{S_j}},3}\right)^2 \max_{\widetilde{K}_j \in \widetilde{\mathcal{T}}_j, \widetilde{K}_j \subset \widetilde{K}} C_{\mathrm{ovrlp}}(\widetilde{K}_j)2^{-4j}|\widetilde{g}|_{H^2(\omega_{\widetilde{K}})}^2 \\
&\le \left(C_{I_S,3} \cdot 2^{-2j}|\widetilde{g}|_{H^2(\omega_{\widetilde{K}})}\right)^2.
\end{aligned} \tag{C.11}$$

Combining (C.7)–(C.11) yields

$$\|\widetilde{w} - \widetilde{I_{S_j}}\widetilde{w}\|_{\widetilde{K}} \le \max_{\ell=1,3} C_{I_S,\ell}\left(\|\widetilde{g} - \widetilde{w}\|_{\omega_{\widetilde{K}}} + 2^{-2j}|\widetilde{g}|_{H^2(\omega_{\widetilde{K}})}\right).$$

From the definition of the K-functional it follows that

$$\begin{aligned}\|\widetilde{w} - \widetilde{I_{S_j}}\widetilde{w}\|^2_{\widetilde{K}} &\leq \max_{\ell=1,3} C^2_{I_S,\ell}\left(\|\widetilde{g}-\widetilde{w}\|_{\omega_{\widetilde{K}}} + 2^{-2j}|\widetilde{g}|_{H^2(\omega_{\widetilde{K}})}\right)^2\\ &\leq 2\cdot \max_{\ell=1,3} C^2_{I_S,\ell}\left(\|\widetilde{g}-\widetilde{w}\|^2_{\omega_{\widetilde{K}}} + \left(2^{-2j}\right)^2|\widetilde{g}|^2_{H^2(\omega_{\widetilde{K}})}\right)\\ &= 2\cdot \max_{\ell=1,3} C^2_{I_S,\ell}\cdot\left(K(2^{-2j},\widetilde{w},\omega_{\widetilde{K}})\right)^2.\end{aligned}$$

In the notation above, $C_\delta = 2\cdot\max_{\ell=1,3} C^2_{I_S,\ell}$.

Using (C.5), (C.6), and Theorem C.2 yields

$$\begin{aligned}\sum_{j=0}^{+\infty} 2^{2j}\|w - I_{S_j}w\|^2_K &\leq h^d_K C_\delta \sum_{j=0}^{+\infty} 2^{2j}\left(K(2^{-2j},\widetilde{w},\omega_{\widetilde{K}})\right)^2\\ &\leq h^d_K C_\delta C_\alpha(\omega_{\widetilde{K}})\|\nabla\widetilde{w}\|^2_{\omega_{\widetilde{K}}}.\end{aligned}$$

Rescaling back to K gives

$$\sum_{j=0}^{+\infty} 2^{2j}\|w - I_{S_j}w\|^2_K \leq h^d_K C_\delta C_\alpha(\omega_{\widetilde{K}}) h^2_K h^{-d}_K \|\nabla w\|^2_{\omega_K}.$$

Finally, note that the shape of $\omega_{\widetilde{K}}$ depends on the shape regularity of the initial mesh and therefore $C_\alpha(\omega_{\widetilde{K}})$ can be bounded, for all $K \in \mathcal{T}_0$, by a constant C_α depending only on d and γ_0. □

THEOREM C.4. *There exists a constant $C_{S,I_S} > 0$ depending only on d and γ_0 such that for all $w \in H^1_0(\Omega)$,*

$$\|\nabla I_{S_0}w\|^2 + \sum_{j=1}^{+\infty}\|h_j^{-1}(I_{S_j}w - I_{S_{j-1}}w)\|^2 \leq C_{S,I_S}\|\nabla w\|^2.$$

Proof. From Theorem B.4,

$$\|\nabla I_{S_0}w\|^2 \leq C^2_{I_{S_0},4}\|\nabla w\|^2.$$

For the rest of the sum,

$$\begin{aligned}\sum_{j=1}^{+\infty}\|h_j^{-1}(I_{S_j} - I_{S_{j-1}})w\|^2 &= \sum_{j=1}^{+\infty}\|h_j^{-1}(I_{S_j}w - w + w - I_{S_{j-1}}w)\|^2\\ &\leq 2\sum_{j=1}^{+\infty}\left(\|h_j^{-1}(w - I_{S_j}w)\|^2 + \|h_j^{-1}(w - I_{S_{j-1}}w)\|^2\right)\\ &\leq 2\sum_{j=1}^{+\infty}\left(\|h_j^{-1}(w - I_{S_j}w)\|^2 + 4\|h_{j-1}^{-1}(w - I_{S_{j-1}}w)\|^2\right)\\ &\leq 2\left(\sum_{j=1}^{+\infty}\|h_j^{-1}(w - I_{S_j}w)\|^2 + 4\sum_{j=0}^{+\infty}\|h_j^{-1}(w - I_{S_j}w)\|^2\right)\\ &\leq 2\cdot 5\sum_{j=0}^{+\infty}\|h_j^{-1}(w - I_{S_j}w)\|^2\end{aligned}$$

$$
\begin{aligned}
&= 10 \sum_{K \in \mathcal{T}_0} \sum_{j=0}^{+\infty} 2^{2j} h_K^{-2} \|w - I_{S_j} w\|_K^2 \\
&\le 10 \sum_{K \in \mathcal{T}_0} C_\beta \|\nabla w\|_{\omega_K}^2 \le 10 \cdot C_\beta \cdot \max_{K \in \mathcal{T}_0} C_{\mathrm{ovrlp}}(K) \|\nabla w\|^2,
\end{aligned}
$$

where we have used Theorem C.3 in the second to last inequality. □

THEOREM C.5. *There exists a constant $C_{S,I_V} > 0$ depending only on d and γ_0 such that for all $w \in H_0^1(\Omega)$,*

$$
\|\nabla I_{V_0} w\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1}(I_{V_j} w - I_{V_{j-1}} w)\|^2 \le C_{S,I_V} \|\nabla w\|^2.
$$

Proof. The key steps of the following proof of the upper bound were provided to us by Prof. P. Oswald in personal communications. From Theorem B.5,

$$
\|\nabla I_{V_0} w\|^2 \le C_{I_{V_0},4}^2 \|\nabla w\|^2.
$$

To bound $\sum_{j=1}^{+\infty} \left\| h_j^{-1}(I_{V_j} w - I_{V_{j-1}} w) \right\|^2$, consider the sequence $w_j = (I_{S_j} - I_{S_{j-1}})w$, $j \in \mathbb{N}$. Let $K \in \mathcal{T}_0$. We will first show that there exists a constant $C_\epsilon > 0$ depending only on d and γ_0 such that

$$
\sum_{j=1}^{+\infty} 2^{2j} \|(I_{V_j} - I_{V_{j-1}})w\|_K^2 \le C_\epsilon \sum_{i=1}^{+\infty} 2^{2i} \|w_i\|_{\omega_K}^2. \tag{C.12}
$$

Since I_{V_j} are projections onto V_j, it holds that

$$
(I_{V_j} - I_{V_{j-1}}) w_i = 0, \qquad j > i.
$$

Then for all $j \ge 1$, using the Cauchy–Schwarz inequality for sums and Theorem B.2,

$$
\begin{aligned}
\|(I_{V_j} - I_{V_{j-1}})w\|_K^2 &= \int_K \left(\sum_{i=j}^{+\infty} (I_{V_j} - I_{V_{j-1}}) w_i \right)^2 \\
&\le \int_K \left(\sum_{i=j}^{+\infty} 2^{-i} \right) \left(\sum_{i=j}^{+\infty} 2^i \left((I_{V_j} - I_{V_{j-1}}) w_i \right)^2 \right) \\
&\le 2 \cdot 2^{-j} \sum_{i=j}^{+\infty} 2^i \|(I_{V_j} - I_{V_{j-1}}) w_i\|_K^2 \\
&\le 2 \cdot 2^{-j} \sum_{i=j}^{+\infty} 2^i \cdot 2 \left(\|I_{V_j} w_i\|_K^2 + \|I_{V_{j-1}} w_i\|_K^2 \right) \\
&\le 2 \cdot 2^{-j} \sum_{i=j}^{+\infty} 2^i \cdot 2 \cdot 2 \cdot \left(\widehat{C}_{I_V,1} \right)^2 \|w_i\|_{\omega_K}^2.
\end{aligned}
$$

Consequently,

$$
\sum_{j=1}^{+\infty} 2^{2j} \|(I_{V_j} - I_{V_{j-1}})w\|_K^2 \le \underbrace{8 \cdot \left(\widehat{C}_{I_{V_\mathcal{T}},1} \right)^2}_{C_\epsilon} \sum_{j=1}^{+\infty} 2^j \sum_{i=j}^{+\infty} 2^i \|w_i\|_{\omega_K}^2.
$$

Changing the order of summation yields

$$\sum_{j=1}^{+\infty} 2^j \sum_{i=j}^{+\infty} 2^i \|w_i\|^2_{\omega_K} = \sum_{i=1}^{+\infty} 2^i \|w_i\|^2_{\omega_K} \sum_{j=1}^{i-1} 2^j \leq \sum_{i=1}^{+\infty} 2^{2i} \|w_i\|^2_{\omega_K}.$$

For the sum $\sum_{j=1}^{+\infty} \|h_j^{-1}(I_{V_j} - I_{V_{j-1}})w\|^2$, using (C.12), we obtain

$$\begin{aligned}
&\sum_{j=1}^{+\infty} \|h_j^{-1}(I_{V_j} - I_{V_{j-1}})w\|^2 \\
&\quad = \sum_{K\in\mathcal{T}_0} h_K^{-2} \sum_{j=1}^{+\infty} 2^{2j} \|(I_{V_j} - I_{V_{j-1}})w\|^2_K \\
&\quad \leq C_\epsilon \sum_{K\in\mathcal{T}_0} h_K^{-2} \sum_{i=1}^{+\infty} 2^{2i} \|w_i\|^2_{\omega_K} \\
&\quad \leq C_\epsilon \sum_{i=1}^{+\infty} 2^{2i} \sum_{K\in\mathcal{T}_0} h_K^{-2} \|w_i\|^2_{\omega_K} \\
&\quad \leq C_\epsilon \sum_{i=1}^{+\infty} 2^{2i} \sum_{K\in\mathcal{T}_0} C_{\mathrm{ovrlp}}(K) \max_{\bar{K}\subset\omega_K} h_{\bar{K}}^{-2} \|w_i\|^2_K \\
&\quad \leq \underbrace{C_\epsilon \max_{K\in\mathcal{T}_0} \left[C_{\mathrm{ovrlp}}(K) \frac{\max_{\bar{K}\subset\omega_K} h_{\bar{K}}^{-2}}{h_K^{-2}} \right]}_{C_\zeta} \sum_{i=1}^{+\infty} 2^{2i} \sum_{K\in\mathcal{T}_0} h_K^{-2} \|w_i\|^2_K.
\end{aligned}$$

Finally, Theorem C.4 gives

$$\sum_{j=1}^{+\infty} \|h_j^{-1}(I_{V_j} - I_{V_{j-1}})w\|^2 \leq C_\zeta C_{S,I_S} \|\nabla w\|^2. \qquad \square$$

Now we proceed by bounding the norm of the splittings from below by a H^1-seminorm. We start with some auxiliary lemmas.

LEMMA C.6. *There exists a constant $c_S > 0$ depending only on d and γ_0 such that for any $N \in \mathbb{N}_0$ and any sequence $(w_j)_{j=0}^N$, $w_j \in S_j$, $j = 0, \ldots, N$, it holds that*

$$\left\| \nabla \Big(\sum_{j=0}^N w_j \Big) \right\|^2 \leq \frac{1}{c_S} \left(\|\nabla w_0\|^2 + \sum_{j=1}^N \|h_j^{-1} w_j\|^2 \right).$$

Proof. The proof for $d = 2$ is given in [6, Lemma 3.4]. It is based on the so-called strengthened Cauchy–Schwarz inequality. As the strengthened Cauchy–Schwarz inequality is valid also for $d = 3$ (see, e.g., [47, Lemma 6.1]), the proof of the theorem in the three-dimensional case is analogous to the two-dimensional one. $\square$

LEMMA C.7. *Let $(w_j)_{j=0}^{+\infty}$, $w_j \in V_j$, $j \in \mathbb{N}_0$, be a sequence which satisfies*

$$\|\nabla w_0\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1} w_j\|^2 < +\infty.$$

Then $\sum_{j=0}^{+\infty} w_j$ converges in $\big(H_0^1(\Omega), \|\nabla \cdot\|\big)$.

Proof. We will use Lemma C.6 to show that $(\sum_{j=0}^{N} w_j)_{N=0}^{+\infty}$ is a Cauchy sequence in $\big(H_0^1(\Omega), \|\nabla \cdot \|\big)$. Let $\epsilon > 0$. Since $\|\nabla w_0\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1} w_j\|^2$ converges in $\mathbb{R}$, there exists $M \in \mathbb{N}$ such that for all $m > n > M$, it holds that

$$\sum_{j=n}^{m} \|h_j^{-1} w_j\|^2 < c_S \epsilon^2.$$

Using Lemma C.6 for w_j, $j = n, \ldots, m$, and the previous inequality,

$$\left\|\nabla\Big(\sum_{j=n}^{m} w_j\Big)\right\|^2 \leq \frac{1}{c_S} \sum_{j=n}^{m} \|h_j^{-1} w_j\|^2 < \epsilon^2.$$

Hence, the partial sum $(\sum_{j=0}^{N} w_j)_{N=0}^{+\infty}$ is a Cauchy sequence in $\big(H_0^1(\Omega), \|\nabla \cdot \|\big)$, and thus $\sum_{j=0}^{+\infty} w_j$ converges in $\big(H_0^1(\Omega), \|\nabla \cdot \|\big)$. □

LEMMA C.8. *Let c_S be the constant from Theorem C.6. Let $(w_j)_{j=0}^{+\infty}$, $w_j \in V_j$, $j \in \mathbb{N}_0$, be a sequence such that $\sum_{j=0}^{+\infty} w_j$ converges in $\big(H_0^1(\Omega), \|\nabla \cdot \|\big)$. Then,*

$$\left\|\nabla\Big(\sum_{j=0}^{+\infty} w_j\Big)\right\|^2 \leq \frac{1}{c_S}\left(\|\nabla w_0\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1} w_j\|^2\right).$$

Proof. For any $N \in \mathbb{N}_0$, Theorem C.6 gives

$$\left\|\nabla\Big(\sum_{j=0}^{N} w_j\Big)\right\|^2 \leq \frac{1}{c_S}\left(\|\nabla w_0\|^2 + \sum_{j=1}^{N} \|h_j^{-1} w_j\|^2\right).$$

Since $\sum_{j=0}^{+\infty} w_j$ converges in $\big(H_0^1(\Omega), \|\nabla \cdot\|\big)$, we may switch the following limit and the norm giving

$$\begin{aligned}
\left\|\nabla\Big(\lim_{N\to+\infty} \sum_{j=0}^{N} w_j\Big)\right\|^2 &= \lim_{N\to+\infty} \left\|\nabla\Big(\sum_{j=0}^{N} w_j\Big)\right\|^2 \\
&\leq \lim_{N\to+\infty} \frac{1}{c_S}\left(\|\nabla w_0\|^2 + \sum_{j=1}^{N} \|h_j^{-1} w_j\|^2\right) \\
&= \frac{1}{c_S}\left(\|\nabla w_0\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1} w_j\|^2\right). \qquad □
\end{aligned}$$

THEOREM C.9. *Any function $w \in H_0^1(\Omega)$ can uniquely be decomposed as*

$$w = I_{V_0} w + \sum_{j=1}^{+\infty} (I_{V_j} - I_{V_{j-1}}) w$$

(the convergence of the sum is understood in the space $(H_0^1(\Omega), \|\nabla \cdot\|)$). Let c_S be the constant from Lemma C.6 and C_{S,I_V} the constant from Theorem C.5. Then for all $w \in H_0^1(\Omega)$,

$$c_S \|\nabla w\|^2 \leq \|\nabla I_{V_0} w\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1}(I_{V_j} w - I_{V_{j-1}} w)\|^2 \leq C_{S,I_V} \|\nabla w\|^2.$$

Proof. The upper bound is proven in Theorem C.5. Now we will prove the lower bound. Having the upper bound at hand, we can use Theorem C.7 to show that the sum $I_{V_0} w + \sum_{j=1}^{+\infty} (I_{V_j} - I_{V_{j-1}}) w$ converges in $\big(H_0^1(\Omega), \|\nabla \cdot \|\big)$, and consequently, from Theorem C.8 with $w_0 := I_{V_0} w$ and $w_j := (I_{V_j} - I_{V_{j-1}}) w$,

$$c_S \left\| \nabla \left(I_{V_0} w + \sum_{j=1}^{+\infty} (I_{V_j} - I_{V_{j-1}}) w \right) \right\|^2 \leq \|\nabla I_{V_0} w\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1}(I_{V_j} w - I_{V_{j-1}} w)\|^2.$$

It remains to show that $I_{V_0} w + \sum_{j=0}^{+\infty} (I_{V_j} - I_{V_{j-1}}) w = w$ in $\big(H_0^1(\Omega), \|\nabla \cdot \|\big)$. Since, for arbitrary $N \in \mathbb{N}$ (see Theorem B.5),

$$\left\| w - \left(I_{V_0} w + \sum_{j=1}^{N} (I_{V_j} - I_{V_{j-1}}) w \right) \right\| = \|w - I_{V_N} w\| \leq C_{I_V,2} \max_{K \in \mathcal{T}_N} h_K \|\nabla w\|,$$

and $\max_{K \in \mathcal{T}_N} h_K \to 0$, we have $I_{V_0} w + \sum_{j=1}^{+\infty} (I_{V_j} - I_{V_{j-1}}) w = w$ in $L^2(\Omega)$. We will show by contradiction that $I_{V_0} w + \sum_{j=1}^{+\infty} (I_{V_j} - I_{V_{j-1}}) w = w$ also in $\big(H_0^1(\Omega), \|\nabla \cdot \|\big)$. Let the sequence $I_{V_0} w + \sum_{j=1}^{N} (I_{V_j} - I_{V_{j-1}}) w$ converge in $\big(H_0^1(\Omega), \|\nabla \cdot \|\big)$ to $\bar{w} \neq w$. Then, thanks to Friedrich's inequality (Theorem A.2), the sequence converges to $\bar{w}$ in $L^2(\Omega)$, which is a contradiction with the uniqueness of the limit. □

THEOREM C.10. *Let c_S be the constant from Theorem C.8. There exists a constant $C_S > 0$ depending only on d and γ_0 such that for all $w \in H_0^1(\Omega)$,*

$$c_S \|\nabla w\|^2 \leq \inf_{w_j \in V_j;\; w = \sum_{j=0}^{+\infty} w_j} \|\nabla w_0\|^2 + \sum_{j=1}^{+\infty} \|h_j^{-1} w_j\|^2 \leq C_S \|\nabla w\|^2. \tag{C.13}$$

Proof. From Theorem C.9 we know that for any $w \in H_0^1(\Omega)$ there exists a decomposition $w = \sum_{j=0}^{+\infty} w_j$, $w_j \in V_j$, $j \in \mathbb{N}_0$, for which the upper bound holds with the constant C_{S,I_V}, so that we can take an infimum over all possible decompositions giving $C_S \leq C_{S,I_V}$. The lower bound in (C.13) follows from Theorem C.8. □

C.2. Splitting of $H_0^1(\Omega)$ into basis function spaces. This section presents a result on the splitting of a $H_0^1(\Omega)$-function into basis function spaces. Denote by $V_{j,i}$, $j = 1, 2 \ldots$, $i = 1, \ldots, \#\mathcal{K}_j$, the space spanned by the basis function $\phi_i^{(j)}$, $V_{j,i} \subset V_j$.

First we will show that splitting a function $w_j \in V_j$ into the basis function spaces $V_{j,i}$, $i = 1, \ldots, \#\mathcal{K}_j$, is stable. This property is called the stability of basis functions in the literature; see, e.g., [37, Definition 2.5.5] and [33, Assumption (A1), p. 17]. We present this property in a form which suits our further development.

LEMMA C.11 (Stability of basis functions). *There exist positive constants c_B and C_B depending only on d and γ_0 such that for all $j \in \mathbb{N}_0$ and all*

$$w_j = \sum_{i=1}^{\#\mathcal{K}_j} w_{j,i} \in V_j, \qquad w_{j,i} \in V_{j,i}, \quad i = 1, \ldots, \#\mathcal{K}_j,$$

it holds that

$$c_B \|h_j^{-1} w_j\|^2 \leq \sum_{i=1}^{\#\mathcal{K}_j} \|\nabla w_{j,i}\|^2 \leq C_B \|h_j^{-1} w_j\|^2. \tag{C.14}$$

Let $\mathbf{M}_j^{\mathrm{S}}$ be the so-called scaled mass matrix and $\mathbf{D}_j$ the diagonal matrix defined as

$$\left[\mathbf{M}_j^{\mathrm{S}}\right]_{m,n} = \int_\Omega h_j^{-2} \phi_n^{(j)} \phi_m^{(j)}, \quad [\mathbf{D}_j]_{m,m} = \int_\Omega \nabla \phi_m^{(j)} \cdot \nabla \phi_m^{(j)}, \qquad \forall m, n = 1, \ldots, \#\mathcal{K}_j.$$

Let $\mathbf{w}_j$ be the vector of coefficients of a function $w_j \in V_j$ in the basis Φ_j. Then, $w_j = \sum_{i=1}^{\#\mathcal{K}_j} w_{j,i}$, $w_{j,i} = [\mathbf{w}_j]_i \phi_i^{(j)}$ and (C.14) is equivalent to

$$c_B \mathbf{w}_j^* \mathbf{M}_j^S \mathbf{w}_j \leq \mathbf{w}_j^* \mathbf{D}_j \mathbf{w}_j \leq C_B \mathbf{w}_j^* \mathbf{M}_j^S \mathbf{w}_j. \tag{C.15}$$

That is, the matrices $\mathbf{M}_j^{\mathrm{S}}$ and $\mathbf{D}_j$ are spectrally equivalent with constants c_B and C_B.

Proof. The proof is inspired by [19, Proposition 1.30, Problem 1.35]; see also [20]. We prove the spectral equivalence of the local matrices associated with a mesh element. The assertion of the theorem for global matrices then follows by summing the local inequalities over the elements and taking into account the overlap.

Let $\mathbf{M}_{j,K}^{\mathrm{S}}$ be a local scaled mass matrix corresponding to an element $K \in \mathcal{T}_j$ defined as

$$\left[\mathbf{M}_{j,K}^{\mathrm{S}}\right]_{m,n} = \int_K h_K^{-2} \phi_n^{(j)} \phi_m^{(j)}, \qquad \forall m, n \in \mathcal{N}_K,$$

and let $\mathbf{M}_{\hat{K}}^{\mathrm{S}}$ be the local scaled mass matrix on a reference element $\hat{K}$, which does not depend on j, K, or $\mathcal{T}_j$. Using standard arguments of an affine transformation to a reference element, it holds that

$$\mathbf{M}_{j,K}^{\mathrm{S}} = \frac{|K|}{h_K^2} \mathbf{M}_{\hat{K}}^{\mathrm{S}}.$$

If we denote by $c_{\hat{K}}$ and $C_{\hat{K}}$ the smallest and the largest eigenvalues of $\mathbf{M}_{\hat{K}}^{\mathrm{S}}$, respectively, then the eigenvalues of $\mathbf{M}_{j,K}^{\mathrm{S}}$ can be bounded by $c_{\hat{K}}|K|/h_K^2$ and $C_{\hat{K}}|K|/h_K^2$. Consequently,

$$\frac{c_{\hat{K}}|K|}{h_K^2} \mathbf{x}^* \mathbf{x} \leq \mathbf{x}^* \mathbf{M}_{j,K}^{\mathrm{S}} \mathbf{x} \leq \frac{C_{\hat{K}}|K|}{h_K^2} \mathbf{x}^* \mathbf{x}, \qquad \forall \mathbf{x} \in \mathbb{R}^{(d+1)}. \tag{C.16}$$

By choosing $\mathbf{x}$ as the mth column of the identity matrix of size $d+1$,

$$\frac{c_{\hat{K}}|K|}{h_K^2} \leq \left[\mathbf{M}_{j,K}^{\mathrm{S}}\right]_{m,m} = \frac{\|\phi_m^{(j)}\|_K^2}{h_K^2} \leq \frac{C_{\hat{K}}|K|}{h_K^2}. \tag{C.17}$$

Let $\mathbf{D}_{j,K}$ be the local variant of $\mathbf{D}_j$, i.e.,

$$[\mathbf{D}_{j,K}]_{m,m} = \int_K \nabla\phi_m^{(j)}\nabla\phi_m^{(j)} = \|\nabla\phi_m^{(j)}\|_K^2.$$

Using the inverse inequality (Theorem A.4) and (C.17),

$$\|\nabla\phi_m^{(j)}\|_K^2 \le C_{\mathrm{INV}}^2 h_K^{-2}\|\phi_m^{(j)}\|_K^2 \le C_{\mathrm{INV}}^2 C_{\hat{K}}\frac{|K|}{h_K^2}.$$

Similarly, using Friedrich's inequality (Theorem A.2),

$$\frac{c_{\hat{K}}|K|}{C_F^2 h_K^2} \le \frac{1}{C_F^2 h_K^2}\|\phi_m^{(j)}\|_K^2 \le \|\nabla\phi_m^{(j)}\|_K^2.$$

Thus the matrix $\mathbf{D}_{j,K}$ is spectrally equivalent to the identity matrix times $|K|h_K^{-2}$. From (C.16), we conclude that $\mathbf{D}_{j,K}$ is also spectrally equivalent to $\mathbf{M}_{j,K}^{\mathrm{S}}$ with the equivalency constants involving C_{INV}, C_F, $c_{\hat{K}}$, and $C_{\hat{K}}$, i.e., depending only on d and the shape regularity γ_j. □

Since $\mathbf{M}_j^{\mathrm{S}}$ and $\mathbf{D}_j$ are spectrally equivalent matrices and they are symmetric positive definite, we can use the generalized Hermitian eigenvalue decomposition (see, e.g., [3, Eq. (5.3)]) and algebraic manipulations to show that $\left(\mathbf{M}_j^{\mathrm{S}}\right)^{-1}$ and $\mathbf{D}_j^{-1}$ are also spectrally equivalent,

$$\text{(C.18)}\qquad \frac{1}{C_B}\mathbf{w}^*\left(\mathbf{M}_j^{\mathrm{S}}\right)^{-1}\mathbf{w} \le \mathbf{w}^*\mathbf{D}_j^{-1}\mathbf{w} \le \frac{1}{c_B}\mathbf{w}^*\left(\mathbf{M}_j^{\mathrm{S}}\right)^{-1}\mathbf{w}, \qquad \forall \mathbf{w}\in\mathbb{R}^{\#\mathcal{K}_j}.$$

Let $\mathbf{M}_j$ denote the mass matrix associated with the jth level, i.e., $[\mathbf{M}_j]_{mn} = \int_\Omega \phi_n^{(j)}\phi_m^{(j)}$, $m,n = 1,\ldots,\#\mathcal{K}_j$. Analogously to (C.15) we can show the spectral equivalence of the mass matrix $\mathbf{M}_j$ with the diagonal matrix $\mathbf{D}_j$ in the following form: There exist positive constants c_M, C_M depending only on d and γ_0 such that

$$\text{(C.19)}\quad c_M \min_{K\in\mathcal{T}_j} h_K^{-2}\mathbf{w}^*\mathbf{M}_j\mathbf{w} \le \mathbf{w}^*\mathbf{D}_j\mathbf{w} \le C_M \max_{K\in\mathcal{T}_j} h_K^{-2}\mathbf{w}^*\mathbf{M}_j\mathbf{w}, \qquad \forall\mathbf{w}\in\mathbb{R}^{\#\mathcal{K}_j}.$$

Combining Theorem C.10 and Theorem C.11 yields the following theorem on splitting an $H_0^1(\Omega)$-function into basis function spaces. It can be proven by the same technique as in [37, Theorem 2.3.1]:

THEOREM C.12. *Let c_S, C_S be the constants from Theorem C.10, c_B, C_B the constants from Theorem C.11, and let $\overline{c}_B = \min\{1, c_B\}$ and $\overline{C}_B = \max\{1, C_B\}$. Then, for all $w \in H_0^1(\Omega)$,*

$$c_S\overline{c}_B\|\nabla w\|^2 \le \inf_{\substack{w_0\in V_0, w_{j,i}\in V_{j,i}\\ w=w_0+\sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j} w_{j,i}}} \|\nabla w_0\|^2 + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j}\|\nabla w_{j,i}\|^2 \le C_S\overline{C}_B\|\nabla w\|^2.$$

C.3. Splitting of spaces of piecewise linear functions. We now present consequences of the previous theorems for finite-dimensional piecewise linear functions from V_J, $J \ge 0$. The following theorems can be proven by the same techniques as the results in [37, Section 2.4]:

THEOREM C.13. *Let c_S and C_S be the constants from Theorem C.10. Let $J \ge 0$. For all $w_J \in V_J$,*

$$c_S\|\nabla w_J\|^2 \le \inf_{w_j\in V_j;\; w_J=\sum_{j=0}^J w_j} \|\nabla w_0\|^2 + \sum_{j=1}^J \|h_j^{-1}w_j\|^2 \le C_S\|\nabla w_J\|^2.$$

THEOREM C.14. *Let c_S and C_S be the constants from Theorem C.10 and $\overline{c}_B, \overline{C}_B$ the constants from Theorem C.12. Let $J \geq 0$. For all $w_J \in V_J$,*

$$c_S\overline{c}_B\|\nabla w_J\|^2 \leq \inf_{\substack{w_0\in V_0, w_{j,i}\in V_{j,i} \\ w_J = w_0 + \sum_{j=1}^{J}\sum_{i=1}^{\#\mathcal{K}_j} w_{j,i}}} \|\nabla w_0\|^2 + \sum_{j=1}^{J}\sum_{i=1}^{\#\mathcal{K}_j}\|\nabla w_{j,i}\|^2 \leq C_S\overline{C}_B\|\nabla w_J\|^2.$$

C.4. Frame. Finally, we present a consequence of the stability of the splittings presented in Theorem C.12, which is closely related to the fact that the normalized basis functions form a so-called *frame in* $\left(H_0^1(\Omega)\right)^\star$; see, e.g., [23, Section 3], [24].

THEOREM C.15. *Let c_S and C_S be the constants from Theorem C.10 and $\overline{c}_B, \overline{C}_B$ the constants from Theorem C.12. For all $g \in \left(H_0^1(\Omega)\right)^\star$,*

$$c_S\overline{c}_B\left(\|\nabla g_0\|^2 + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j}\frac{\langle g, \phi_i^{(j)}\rangle^2}{\|\nabla\phi_i^{(j)}\|^2}\right) \leq \|g\|^2_{\left(H_0^1(\Omega)\right)^\star}$$
$$\leq C_S\overline{C}_B\left(\|\nabla g_0\|^2 + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j}\frac{\langle g, \phi_i^{(j)}\rangle^2}{\|\nabla\phi_i^{(j)}\|^2}\right),$$

where $g_0 \in V_0$ is the Riesz representation of the functional g in the space V_0 with respect to the inner product $(u_0, v_0)_0 = \int_\Omega \nabla v_0 \cdot \nabla u_0$, $\forall u_0, v_0 \in V_0$.

Proof. The proof is inspired by the proof of [37, Theorem 2.6.2]. We will start with the upper bound. Let $w \in H_0^1(\Omega)$ and consider an arbitrary decomposition of w as $w = w_0 + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j} w_{j,i}$, $w_0 \in V_0$, $w_{j,i} \in V_{j,i}$. Using the fact that

$$w_{i,j} = \frac{\operatorname{sign}(w_{i,j})\,\|\nabla w_{i,j}\|}{\|\nabla\phi_i^{(j)}\|}\,\phi_i^{(j)},$$

we have

$$\begin{aligned}|\langle g, w\rangle| &\leq |\langle g, w_0\rangle| + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j}|\langle g, w_{i,j}\rangle| \\ &\leq \|g_0\|\cdot\|\nabla w_0\| + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j}\left|\left\langle g, \frac{\phi_i^{(j)}}{\|\nabla\phi_i^{(j)}\|}\right\rangle\right|\cdot\|\nabla w_{i,j}\| \\ &\leq \left(\|\nabla g_0\|^2 + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j}\frac{\langle g, \phi_i^{(j)}\rangle^2}{\|\nabla\phi_i^{(j)}\|^2}\right)^{\frac{1}{2}}\cdot\left(\|\nabla w_0\|^2 + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j}\|\nabla w_{j,i}\|^2\right)^{\frac{1}{2}}.\end{aligned}$$

Taking the infimum over all decompositions $w = w_0 + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j} w_{j,i}$, $w_0 \in V_0$, $w_{j,i} \in V_{j,i}$, and using the stability of the decomposition into spaces defined by basis functions (Theorem C.12) yields

$$|\langle g, w\rangle| \leq \left(\|\nabla g_0\|^2 + \sum_{j=1}^{+\infty}\sum_{i=1}^{\#\mathcal{K}_j}\frac{\langle g, \phi_i^{(j)}\rangle^2}{\|\nabla\phi_i^{(j)}\|^2}\right)^{\frac{1}{2}}\cdot C_S^{\frac{1}{2}}\overline{C}_B^{\frac{1}{2}}\|\nabla w\|.$$

Taking the supremum over all $w \in H_0^1(\Omega)$ such that $\|\nabla w\| = 1$ gives the upper bound.

Proving the lower bound is more subtle. We will first show that for any $N \in \mathbb{N}$,

$$\|\nabla g_0\|^2 + \sum_{j=1}^{N} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle g, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} \leq \frac{1}{c_S \bar{c}_B} \|g\|^2_{(H_0^1(\Omega))^\star}. \tag{C.20}$$

First, it holds that

$$\begin{aligned} \|\nabla g_0\|^2 + \sum_{j=1}^{N} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle g, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} &= \langle g, g_0 \rangle + \sum_{j=1}^{N} \sum_{i=1}^{\#\mathcal{K}_j} \left\langle g, \frac{\langle g, \phi_i^{(j)} \rangle}{\|\nabla \phi_i^{(j)}\|^2} \phi_i^{(j)} \right\rangle \\ &= \left\langle g, g_0 + \sum_{j=1}^{N} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle g, \phi_i^{(j)} \rangle}{\|\nabla \phi_i^{(j)}\|^2} \phi_i^{(j)} \right\rangle. \end{aligned}$$

Let $g_{j,i} = \frac{\langle g, \phi_i^{(j)} \rangle}{\|\nabla \phi_i^{(j)}\|^2} \phi_i^{(j)} \in V_{j,i}$. Then using Theorem C.14,

$$\begin{aligned} \left\langle g, g_0 + \sum_{j=1}^{N} \sum_{i=1}^{\#\mathcal{K}_j} g_{j,i} \right\rangle &= \|g\|_{(H_0^1(\Omega))^\star} \left\| \nabla \left(g_0 + \sum_{j=1}^{N} \sum_{i=1}^{\#\mathcal{K}_j} g_{j,i} \right) \right\| \\ &\leq \|g\|_{(H_0^1(\Omega))^\star} \frac{1}{c_S^{\frac{1}{2}} \bar{c}_B^{\frac{1}{2}}} \left(\|\nabla g_0\|^2 + \sum_{j=1}^{N} \sum_{i=1}^{\#\mathcal{K}_j} \|\nabla g_{j,i}\|^2 \right)^{\frac{1}{2}} \\ &= \|g\|_{(H_0^1(\Omega))^\star} \frac{1}{c_S^{\frac{1}{2}} \bar{c}_B^{\frac{1}{2}}} \left(\|\nabla g_0\|^2 + \sum_{j=1}^{N} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle g, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} \right)^{\frac{1}{2}}. \end{aligned}$$

This yields (C.20). Taking N to infinity in (C.20) completes the proof. □

THEOREM C.16. *Let c_S and C_S be the constants from Theorem C.10 and $\bar{c}_B, \overline{C}_B$ the constants from Theorem C.12. Let $J \geq 0$, and consider the space V_J with the norm $\|\nabla \cdot \|$. For all $g_J \in V_J^\star$,*

$$\begin{aligned} c_S \bar{c}_B \left(\|\nabla g_0\|^2 + \sum_{j=1}^{J} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle g, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} \right) &\leq \|g_J\|^2_{V_J^\star} \\ &\leq C_S \overline{C}_B \left(\|\nabla g_0\|^2 + \sum_{j=1}^{J} \sum_{i=1}^{\#\mathcal{K}_j} \frac{\langle g, \phi_i^{(j)} \rangle^2}{\|\nabla \phi_i^{(j)}\|^2} \right), \end{aligned}$$

where $g_0 \in V_0$ is the Riesz representation function of the functional g in the space V_0 with respect to the inner product $(u_0, v_0) = \int_\Omega \nabla v_0 \cdot \nabla u_0$, $\forall u_0, v_0 \in V_0$.

Proof. The proof is analogous to the proof of Theorem C.15. □

REFERENCES

[1] M. AINSWORTH AND J. T. ODEN, *A posteriori error estimation in finite element analysis*, Comput. Methods Appl. Mech. Engrg., 142 (1997), pp. 1–88.

[2] M. S. ALNAES, J. BLECHTA, J. HAKE, A. JOHANSSON, B. KEHLET, A. LOGG, C. RICHARDSON, J. RING, M. E. ROGNES, AND G. N. WELLS , *The FEniCS project version 1.5*, Arch. Numer. Software, 3 (2015), pp. 9–23.

[3] Z. BAI, J. DEMMEL, J. DONGARRA, A. RUHE, AND H. VAN DER VORST, eds., *Templates for the Solution of Algebraic Eigenvalue Problems*, SIAM, Philadelphia, 2000.

[4] A. H. BAKER, R. D. FALGOUT, H. GAHVARI, T. GAMBLIN, W. GROPP, T. V. KOLEV, K. E. JORDAN, M. SCHULZ, AND U. M. YANG, *Preparing Algebraic Multigrid for Exascale*, Tech. Rep. LLNL-TR-533076, Lawrence Livermore National Laboratory, Livermore, 2012.

[5] R. BECKER, C. JOHNSON, AND R. RANNACHER, *Adaptive error control for multigrid finite element methods*, Computing, 55 (1995), pp. 271–288.

[6] F. BORNEMANN AND H. YSERENTANT, *A basic norm equivalence for the theory of multilevel methods*, Numer. Math., 64 (1993), pp. 455–476.

[7] A. BRANDT, *Multigrid Technique—1984 Guide with Applications to Fluid Dynamics*, SIAM, Philadelphia, 2011.

[8] S. C. BRENNER AND L. R. SCOTT, *The Mathematical Theory of Finite Element Methods*, 3rd ed., Springer, New York, 2008.

[9] H. BREZIS, *Functional Analysis, Sobolev Spaces and Partial Differential Equations*, Springer, New York, 2011.

[10] W. L. BRIGGS, V. E. HENSON, AND S. F. MCCORMICK, *A Multigrid Tutorial*, 2nd ed., SIAM, Philadelphia, 2000.

[11] A. BUTTARI, M. HUBER, P. LELEUX, T. MARY, U. RÜDE, AND B. WOHLMUTH, *Block low-rank single precision coarse grid solvers for extreme scale multigrid methods*, Numer. Linear Algebra Appl., 29 (2022), Paper No. e2407, 15 pages.

[12] C. CARSTENSEN, *Quasi-interpolation and a posteriori error analysis in finite element methods*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 1187–1202.

[13] E. CHOW, R. D. FALGOUT, J. J. HU, R. S. TUMINARO, AND U. M. YANG, *A survey of parallelization techniques for multigrid solvers*, in Parallel Processing for Scientific Computing, M. A. Heroux, P. Raghavan, and H. D. Simon, eds., SIAM, Philadelphia, 2006, pp. 179–201.

[14] P. G. CIARLET, *The Finite Element Method for Elliptic Problems*, North-Holland, Amsterdam, 1978.

[15] P. CLÉMENT, *Approximation by finite element functions using local regularization*, Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér., 9 (1975), pp. 77–84.

[16] W. DAHMEN, *Wavelet and multiscale methods for operator equations*, in Acta Numerica 6, A. Iserles, ed., Cambridge University Press, Cambridge, 1997, pp. 55–228.

[17] W. DAHMEN AND A. KUNOTH, *Multilevel preconditioning*, Numer. Math., 63 (1992), pp. 315–344.

[18] P. D'AMBRA, F. DURASTANTE, AND S. FILIPPONE, *AMG preconditioners for linear solvers towards extreme scale*, SIAM J. Sci. Comput., 43 (2021), pp. S679–S703.

[19] H. C. ELMAN, D. J. SILVESTER, AND A. J. WATHEN, *Finite Elements and Fast Iterative Solvers: With Applications in Incompressible Fluid Dynamics*, Oxford University Press, New York, 2005.

[20] I. FRIED, *Bounds on the extremal eigenvalues of the finite element stiffness and mass matrices and their spectral condition number*, J. Sound Vibr., 22 (1972), pp. 407–418.

[21] G. H. GOLUB AND Z. STRAKOŠ, *Estimates in quadratic formulas*, Numer. Algorithms, 8 (1994), pp. 241–268.

[22] W. HACKBUSCH, *Iterative Solution of Large Sparse Systems of Equations*, 2nd ed., Springer, Cham, 2016.

[23] H. HARBRECHT AND R. SCHNEIDER, *A note on multilevel based error estimation*, Comput. Methods Appl. Math., 16 (2016), pp. 447–458.

[24] H. HARBRECHT, R. SCHNEIDER, AND C. SCHWAB, *Multilevel frames for sparse tensor product spaces*, Numer. Math., 110 (2008), pp. 199–220.

[25] M. R. HESTENES AND E. STIEFEL, *Methods of conjugate gradients for solving linear systems*, J. Research Nat. Bur. Standards, 49 (1952), pp. 409–436.

[26] M. HUBER, *Massively Parallel and Fault-Tolerant Multigrid Solvers on Peta-Scale Systems*, PhD. Thesis, TU Munich, Munich, 2019.

[27] A. LOGG, K.-A. MARDAL, AND G. N. WELLS, eds., *Automated Solution of Differential Equations by the Finite Element Method*, Springer, Heidelberg, 2012.

[28] S. F. MCCORMICK, J. BENZAKEN, AND R. TAMSTORF, *Algebraic error analysis for mixed-precision multigrid solvers*, SIAM J. Sci. Comput., 43 (2021), pp. S392–S419.

[29] G. MEURANT, J. PAPEŽ, AND P. TICHÝ, *Accurate error estimation in CG*, Numer. Algorithms, 88 (2021), pp. 1337–1359.

[30] G. MEURANT AND P. TICHÝ, *The behaviour of the Gauss-Radau upper bound of the error norm in CG*, Numer. Algorithms, 94 (2023), pp. 847–876.

[31] A. MIRAÇI, J. PAPEŽ, AND M. VOHRALÍK, *A-posteriori-steered p-robust multigrid with optimal step-sizes and adaptive number of smoothing steps*, SIAM J. Sci. Comput., 43 (2021), pp. S117–S145.

[32] Y. NOTAY, *Convergence analysis of perturbed two-grid and multigrid methods*, SIAM J. Numer. Anal., 45 (2007), pp. 1035–1044.

[33] P. OSWALD, *Multilevel Finite Element Approximation*, B. G. Teubner, Stuttgart, 1994.

[34] 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), Paper No. 113243, 39 pages.

[35] 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.

[36] K. Rektorys, *Variational Methods in Mathematics, Science and Engineering*, 2nd ed., D. Reidel, Dordrecht-Boston, 1980.

[37] U. Rüde, *Mathematical and Computational Techniques for Multilevel Adaptive Methods*, SIAM, Philadelphia, 1993.

[38] L. R. Scott and S. Zhang, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Math. Comp., 54 (1990), pp. 483–493.

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

[40] Z. Strakoš and P. Tichý, *On error estimation in the conjugate gradient method and why it works in finite precision computations*, Electron. Trans. Numer. Anal., 13 (2002), pp. 56–80.
https://etna.ricam.oeaw.ac.at/vol.13.2002/pp56-80.dir/pp56-80.pdf

[41] ———, *Error estimation in preconditioned conjugate gradients*, BIT, 45 (2005), pp. 789–817.

[42] R. Tamstorf, J. Benzaken, and S. F. McCormick, *Discretization-error-accurate mixed-precision multigrid solvers*, SIAM J. Sci. Comput., 43 (2021), pp. S420–S447.

[43] U. Trottenberg, C. W. Oosterlee, and A. Schuller, *Multigrid*, Academic Press, San Diego, 2001.

[44] P. Vacek, *Multilevel Methods*, Master Thesis, Charles University, Prague, 2020.

[45] P. Vacek, E. Carson, and K. M. Soodhalter, *The effect of approximate coarsest-level solves on the convergence of multigrid V-cycle methods*, SIAM J. Sci. Comput., 46 (2024), pp. A2634–A2659.

[46] R. Verfürth, *A Posteriori Error Estimation Techniques for Finite Element Methods*, Oxford University Press, Oxford, 2013.

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

[48] X. Xu and C.-S. Zhang, *Convergence analysis of inexact two-grid methods: a theoretical framework*, SIAM J. Numer. Anal., 60 (2022), pp. 133–156.

[49] H. Yserentant, *Old and new convergence proofs for multigrid methods*, in Acta Numerica, 1993, A. Iserles, ed., Cambridge University Press, Cambridge, 1993, pp. 285–326.

[50] S. Zhang, *Successive subdivisions of tetrahedra and multigrid methods on tetrahedral meshes*, Houston J. Math., 21 (1995), pp. 541–556.