2014

A Posteriori Estimates Using Auxiliary Subspace Techniques

Harri Hakula
Aalto University

Michael Neilan
University of Pittsburgh,

Jeffrey S. Ovall
Portland State University, jovall@pdx.edu

A POSTERIORI ESTIMATES USING AUXILIARY SUBSPACE TECHNIQUES

HARRI HAKULA[†], MICHAEL NEILAN[‡], AND JEFFREY S. OVALL[§]

Abstract. A posteriori error estimators based on auxiliary subspace techniques for second order elliptic problems in $\mathbb{R}^d$ ($d \geq 2$) are considered. In this approach, the solution of a global problem is utilized as the error estimator. As the continuity and coercivity of the problem trivially leads to an efficiency bound, the main focus of this paper is to derive an analogous effectivity bound and to determine the computational complexity of the auxiliary approximation problem. With a carefully chosen auxiliary subspace, we prove that the error is bounded above by the error estimate up to oscillation terms. In addition, we show that the stiffness matrix of the auxiliary problem is spectrally equivalent to its diagonal. Several numerical experiments are presented verifying the theoretical results.

1. Introduction.

1.1. Problem Statement and Background. Let $\Omega \subset \mathbb{R}^d$ ($d \geq 2$) be a bounded polytope, having boundary $\partial\Omega = \Gamma_N \cup \Gamma_D$, a disjoint union with Γ_D closed in the relative topology on $\partial\Omega$. We define the space

$$H^1_{0,D}(\Omega) = \{v \in H^1(\Omega) : \ v = 0 \text{ on } \Gamma_D \text{ in the sense of trace}\} \ ,$$

and adopt the following notation for norms and semi-norms on Hilbert spaces $H^k(\omega)$ ($k \geq 0$) for $\omega \subset \Omega$,

$$\|v\|^2_{k,\omega} = \sum_{|\alpha|\leq k} \|D^\alpha v\|^2_{L^2(\omega)} \quad , \quad |v|^2_{k,\omega} = \sum_{|\alpha|=k} \|D^\alpha v\|^2_{L^2(\omega)} \ .$$

When $\omega = \Omega$, we omit it from the subscript. We also employ these Sobolev spaces and norms on subsets of Ω having lower dimension.

We consider variational problems of the form

(1.1)

$$\text{Find } u \in H^1_{0,D}(\Omega) \ \ni: \ \underbrace{\int_\Omega A\nabla u \cdot \nabla v + (\boldsymbol{b}\cdot\nabla u + cu)v}_{B(u,v)} = \underbrace{\int_\Omega fv + \int_{\Gamma_N} gv}_{F(v)} \ \forall v \in H^1_{0,D}(\Omega) \ ,$$

where the data $A : \Omega \to \mathbb{R}^{d\times d}$, $\boldsymbol{b} : \Omega \to \mathbb{R}^d$, $c, f : \Omega \to \mathbb{R}$ and $g : \Gamma_N \to \mathbb{R}$ are piecewise smooth with respect to some polyhedral partition of Ω. The matrix A is symmetric and uniformly positive definite a.e. in Ω, $A(x)\boldsymbol{z}\cdot\boldsymbol{z} \geq \alpha|\boldsymbol{z}|^2$ for all $\boldsymbol{z} \in \mathbb{R}^d$ and a.e. $x \in \Omega$. We further assume conditions on the coefficients so that B is continuous and coercive,

$$|B(v,w)| \leq \mathfrak{C}\|v\|_1\|w\|_1 \quad , \quad B(v,v) \geq \mathfrak{c}\|v\|^2_1 \quad \text{for all } v, w \in H^1_{0,D}(\Omega) \ . \tag{1.2}$$

[†]Department of Mathematics and Systems Analysis, Aalto University, harri.hakula@aalto.fi. Partially supported by the Mathematisches Forschungsinstitut Oberwolfach Research-In-Pairs program.

[‡]Department of Mathematics, University of Pittsburgh, neilan@pitt.edu. Partially supported by the National Science Foundation through grant DMS–1417980.

[§]Fariborz Maseeh Department of Mathematics and Statistics, Portland State University, jovall@pdx.edu. Partially supported by the National Science Foundation through grant DMS–1414365, and the Mathematisches Forschungsinstitut Oberwolfach Research-In-Pairs program.

Under these assumptions, the problem (1.1) is well-posed. We refer to $\mathfrak{C}$ and $\mathfrak{c}$, respectively, as the continuity constant and the coercivity constant.

Given a family $\{\mathcal{T}\}$ of conforming, shape-regular simplicial partitions of Ω, we define the standard piecewise polynomial finite element spaces on $\mathcal{T}$:

$$V = V_p = \{v \in H^1_{0,D}(\Omega) : \ v|_T \in \mathcal{P}_p(T) \text{ for each } T \in \mathcal{T}\}, \tag{1.3}$$

where $\mathcal{P}_p(T)$ is the space of polynomials of total degree $\leq p$ on T. More generally $\mathcal{P}_p(S)$ is taken to be the space of polynomials of total degree $\leq p$ having domain $S \subset \mathbb{R}^j$ for some $0 \leq j \leq d$. Given an auxiliary "error space" $W \subset H^1_{0,D}(\Omega)$ such that $V \cap W = \{0\}$, we consider the approximation problem

$$\text{Find } \hat{u} \in V \ \ni: \ B(\hat{u}, v) = F(v) \ \ (= B(u, v)) \ \forall v \in V \ , \tag{1.4}$$

and the error problem

$$\text{Find } \varepsilon \in W \ \ni: \ B(\varepsilon, v) = F(v) - B(\hat{u}, v) \ \ (= B(u - \hat{u}, v)) \ \forall v \in W \ . \tag{1.5}$$

Since V and W inherit the continuity and coercivity of B, these problems are also well-posed. In the present work, W is also a piecewise polynomial space with respect to $\mathcal{T}$, but we postpone its definition to later sections. Throughout, we implicitly assume that $\mathcal{T}$ is subordinate to the polyhedral partition of Ω on which the data is piecewise smooth—i.e., the data is smooth on each simplex $T \in \mathcal{T}$.

The computation of an approximate error function $\varepsilon \approx u - \hat{u}$ via (1.5) has historically been called *hierarchical basis error estimation*, and its origins can be traced back to [27, 12, 13], with what we will call the traditional analysis first presented in [7]. We refer interested readers to [5, 2] for more detailed discussion of the traditional analysis of hierarchical bases in error estimation and linear solvers, though we will mention a few basic results below. We have opted to use *auxiliary subspace error estimation* to describe (1.5) in the present work, instead of *hierarchical basis error estimation*, primarily for two reasons. The first is that, in the traditional approach, W is generally chosen so that $V \oplus W$ is a natural finite element space—for example, piecewise polynomials of one higher degree than V on the same triangulation, or piecewise polynomials of the same degree as V on a refined (nested) triangulation. Our choice of W is motivated by different considerations, and becomes more obviously different from what would be considered natural choices under the traditional approach when $d \geq 3$. The second reason for choosing a different descriptor is that *hierarchical basis error estimation* is also being used in the literature (cf. [26, 20]) to describe an approach which is quite different from (1.5), though it bears some superficial similarities in terms of basic components (bubble-functions), in its development. The most obvious distinction between the two is that (1.5) is clearly an implicit method, whereas the approach put forth in [26] is an explicit method, which can be shown to be equivalent to the standard residual method. In fact, it is this equivalence which is used in [20] to assert that local indicators based on such a method can be used to drive a provably convergent adaptive algorithm. At present, there are no known results proving that marking strategies based on local indicators computed from ε lead to convergent adaptive algorithms, even in the energy norm setting, though there is a wealth of empirical evidence that they work at least as well as any other local indicators. Here we have used the term *implicit method*, in the manner of [2], to describe methods which require the solution of local or global problems after having solved for $\hat{u}$.

We briefly rehearse the key ideas and result of the traditional analysis, and then make a few remarks on other work which might most readily be compared with our

own. If B is an inner-product on $H^1_{0,D}(\Omega)$, with corresponding energy norm $\|\!| \cdot |\!\|$, it is clear that $\|\!|\varepsilon|\!\| \leq \|\!|u - \hat{u}|\!\|$, so ε provides (globally) efficient error estimation. There is a constant $\gamma = \gamma(B, V, W) \in [0, 1)$ such that a strong Cauchy-Schwarz inequality (cf. [15]), $B(v, w) \leq \gamma \|\!|v|\!\| \|\!|w|\!\|$, holds for all $v \in V$ and $w \in W$. If the local spaces $V(T)$ and $W(T)$ do not vary in type as the mesh is refined, e.g. they consist of some fixed subspaces of polynomials on T, then it can be shown that γ does not deteriorate (approach 1) as the mesh is refined, as long as shape-regularity is maintained. If a saturation assumption,

$$\exists \beta \in [0, 1) \ \ni: \ \|\!|u - \hat{u}|\!\| = \inf_{v \in V} \|\!|u - v|\!\| \leq \beta \inf_{v \in V} \|\!|u - v|\!\| \tag{1.6}$$

also holds, one obtains a complementary reliability result. We summarize both efficiency and reliability with the two-sided bound

$$\|\!|\varepsilon|\!\| \leq \|\!|u - \hat{u}|\!\| \leq [(1 - \gamma^2)(1 - \beta^2)]^{-1/2} \|\!|\varepsilon|\!\| \ . \tag{1.7}$$

Clearly (1.6) depends the specifics of u, and although it is generally expected to hold on sufficiently fine meshes for problems likely to be encountered in practice, it is not difficult to construct counter-examples (cf. [9, 14]). In [9, 14], notions of *data oscillation* are used to replace the saturation assumption with a quantity which is at least directly measurable in principle, even it if is not measured in practice. We use a related notion of *residual oscillation* in the present work in a similar fashion, although the approach to the analysis is quite different, and a potential link between residual oscillation and the saturation assumption is not apparent. Most treatments of hierarchical basis error estimation do not consider non-self-adjoint B. A few that do are [7, 3], and some of our previous work [6, 17] concerning linear finite elements. In both [7] and [3] an energy norm derived from the symmetric part of B plays a key role in the construction and/or analysis of the derived estimator, and the analysis in both cases is quite different from our own.

1.2. Main Results. In order to describe our basic approach to constructing and analyzing ε, we introduce some basic notation. Let $\mathcal{F}$ denote the set of $(d-1)$-dimensional subsimplices, the "faces" of $\mathcal{T}$, and further decompose this as $\mathcal{F} = \mathcal{F}_I \cup \mathcal{F}_D \cup \mathcal{F}_N$, where $\mathcal{F}_I$ comprises those faces in the interior of Ω, and $\mathcal{F}_D$ and $\mathcal{F}_N$ comprise those faces in Γ_D and Γ_N, respectively. The starting point of our analysis is the following error identity, which follows directly from (1.4) and (1.5) and elementwise integration-by-parts as used in the derivation of residual methods:

PROPOSITION 1.1. *For any $v \in H^1_{0,D}(\Omega)$, $w \in W$ and $\hat{v} \in V$, it holds that*

$$B(u - \hat{u}, v) = B(\varepsilon, w) + \mathcal{R}(v - \hat{v} - w) \ ,$$

where

$$\mathcal{R}(\phi) = F(\phi) - B(\hat{u}, \phi) = \sum_{T \in \mathcal{T}} \int_T R_T \phi + \sum_{F \in \mathcal{F}_I \cup \mathcal{F}_N} \int_F r_F \phi \ ,$$

and

$$R_T = f - (-\nabla \cdot A \nabla \hat{u} + \boldsymbol{b} \cdot \nabla \hat{u} + c\hat{u})|_T \ ,$$
$$r_F = \begin{cases} g - A \nabla \hat{u} \cdot \boldsymbol{n} & , F \in \mathcal{F}_N \\ (-A \nabla \hat{u} \cdot \boldsymbol{n}_T)|_T - (A \nabla \hat{u} \cdot \boldsymbol{n}_{T'})|_{T'} & , F \in \mathcal{F}_I \end{cases} \ .$$

Here, T and T' are the simplices sharing the face F, and $\boldsymbol{n}_T$ and $\boldsymbol{n}_{T'}$ are their outward unit normals; for $F \in \mathcal{F}_N$, the outward normal to $\boldsymbol{n}$ for $\partial\Omega$ is used.

REMARK 1.2. *The identity $B(u-\hat{u},v) = \mathcal{R}(v-\hat{v})$ for $v \in H^1_{0,D}(\Omega)$ and $\hat{v} \in V$ is the starting point for residual error estimates, which are obtained by choosing $\hat{v}$ to be a suitable interpolant of v, and deriving corresponding bounds on the weak residual, $|\mathcal{R}(v-\hat{v})| \leq C\eta\|v\|_1$. Here η is comprised of appropriate weights, involving the local mesh size, on the volumetric and face residuals, $\|R_T\|_{0,T}$ and $\|r_F\|_{0,F}$. We note that reliability bounds for residual estimators of this sort are very naturally obtained, and it is efficiency bounds, involving oscillation terms, which require more ingenuity to establish. This is the opposite of the situation for the auxiliary subspace error estimators discussed here.*

With an appropriate choice of error space $W = W_{p+d}$, described in detail later, we obtain our key error theorem, the upper bound of which is proved in Section 2.3—the lower bound is a trivial consequence of the coercivity and continuity conditions (1.2).

THEOREM 1.3. *There is a constant C depending only on the dimension d, polynomial degree p, continuity and coercivity constants $\mathfrak{C}$ and $\mathfrak{c}$, and the shape-regularity of $\mathcal{T}$ such that*

$$\frac{\mathfrak{c}}{\mathfrak{C}}\|\varepsilon\|_1 \leq \|u-\hat{u}\|_1 \leq C\left(\|\varepsilon\|_1 + \operatorname{osc}(R,r,\mathcal{T})\right) \ ,$$

where the residual oscillation is defined by

$$\operatorname{osc}(R,r,\mathcal{T})^2 = \sum_{T\in\mathcal{T}} h_T^2 \inf_{\kappa\in\mathcal{P}_{p-1}(T)} \|R_T-\kappa\|^2_{0,T} + \sum_{F\in\mathcal{F}_I\cup\mathcal{F}_N} h_F \inf_{\kappa\in\mathcal{P}_{p-1}(F)} \|r_F-\kappa\|^2_{0,F} \ .$$

Here and following, h_T is the diameter of T and h_F is the diameter of F. The space W_{p+d} will be spanned by appropriate "face bubble functions" supported in the two (or one) simplices sharing a face, and "interior bubble functions" supported in a simplex. In [17, 6] it was shown that interior bubbles are not needed for lowest order elements when $d = 2, 3$. A very different sort of analysis was used in [23, 21] for lowest order elements and $d = 2$ to show that the (H^1 or energy) error estimates based on ε are often asymptotically identical to the actual error.

As stated, the computation of ε requires the formation and solution of a global system, so one might naturally be concerned that the approach is too expensive for practical consideration. However, even those implicit methods which are based on the solution local (elementwise or patchwise) problems require the computation of local stiffness matrices. In Section 3 we argue that the size and sparsity structure of the system for computing ε is comparable to that of setting up all of the local systems for other implicit methods. So when comparing the cost of this and other implicit methods, the real issue is whether or not it is more expensive to solve a single global system or a collection of local systems. Our key result in this regard is

THEOREM 1.4. *The global stiffness matrix for W_{p+d} is spectrally-equivalent to its diagonal.* Although this result implies that we could get away with solving a diagonal system, and that the modified $\tilde{\varepsilon}$ would still provide two-sided bounds as in Theorem 1.3, with suitably adjusted constants, we instead advocate (approximately) solving the full system using a few steps of a Krylov solver (CG, BiCG-Stab, GMRES) either with no preconditioning (e.g. when $d = 2$) or a simple Jacobi or Gauss-Seidel preconditioner.

We offer a few more remarks concerning the solution of local or global problems in the construction of error estimates. An approximate error function $\varepsilon \approx u - \hat{u}$ is

very naturally obtained through the solution of the global problem (1.5), and can be used for driving anisotropic h-refinement or r-refinement (mesh smoothing). Although local problems might also be used in this regard, empirical evidence [18, 19] suggests that the solution of global problems are better suited for this purpose. A point in which some approaches based on local problems currently have a theoretical advantage over the approach presented here is that they are provably robust with respect to polynomial degree [10, 16], a property which is known not to hold for standard residual-based error estimates [22]. The efficiency (lower) bound in Theorem 1.3 is clearly independent of the polynomial degree p, but the proof presented here for the reliability bound suggests that the constant C could depend on p, which is not ideal. Extensive numerical experiments, as reported in Section 4, provide empirical evidence that our estimate is robust with respect to p, and we hope to prove this in future work.

Before outlining the rest of the paper, we finally provide some motivation for the development in Section 2 by considering the residual oscillation term. We define the local residual oscillation for each $T \in \mathcal{T}$ by

$$\begin{aligned}\operatorname{osc}(R,r,T)^2 = h_T^2 \inf_{\kappa\in\mathcal{P}_{p-1}(T)} \|R_T-\kappa\|_{0,T}^2 + \frac{1}{2}\sum_{F\in\mathcal{F}_{I,T}} h_F \inf_{\kappa\in\mathcal{P}_{p-1}(F)} \|r_F-\kappa\|_{0,F}^2 \\ + \sum_{F\in\mathcal{F}_{N,T}} h_F \inf_{\kappa\in\mathcal{P}_{p-1}(F)} \|r_F-\kappa\|_{0,F}^2 \ ,\end{aligned}$$

where $\mathcal{F}_{I,T}$ are the faces of T in $\mathcal{F}_I$ and $\mathcal{F}_{N,T}$ are the faces of T in $\mathcal{F}_N$. By definition,

$$\operatorname{osc}(R,r,\mathcal{T})^2 = \sum_{T\in\mathcal{T}} \operatorname{osc}(R,r,T)^2 \ .$$

The choice of W_{p+d} is such that, if the data is piecewise smooth (for example), then the local oscillation is of higher order than the local best-approximation error:

$$\frac{\operatorname{osc}(R,r,T)}{\inf_{v\in\mathcal{P}_p(T)}\|u-v\|_{1,T}} \to 0 \text{ as } |T|\to 0 \ . \tag{1.8}$$

This is illustrated more explicitly in the following example.

EXAMPLE 1.5. *Suppose that A and $\boldsymbol{b}$ are piecewise constant and $c=0$. In this case it holds that $(-\nabla\cdot A\nabla\hat{u}+\boldsymbol{b}\cdot\nabla\hat{u})|_T \in \mathcal{P}_{p-1}(T)$ and $A\nabla\hat{u}|_T \in [\mathcal{P}_{p-1}(T)]^d$, so we have*

$$\operatorname{osc}(R,r,T)^2 = h_T^2 \inf_{\kappa\in\mathcal{P}_{p-1}(T)} \|f-\kappa\|_{0,T}^2 + \sum_{F\in\mathcal{F}_{N,T}} h_F \inf_{\kappa\in\mathcal{P}_{p-1}(F)} \|g-\kappa\|_{0,F}^2 \ .$$

If $f\in H^p(T)$ and $g\in H^p(F)$, then $\operatorname{osc}(R,r,T)=\mathcal{O}(h_T^{p+1/2})$. If $f\in H^p(T)$ and $g\in\mathcal{P}_{p-1}(F)$, then $\operatorname{osc}(R,r,T)=\mathcal{O}(h_T^{p+1})$. Finally, if $f\in\mathcal{P}_{p-1}(T)$ and $g\in\mathcal{P}_{p-1}(F)$, then $\operatorname{osc}(R,r,T)=0$.

1.3. Outline of Paper. The rest of the paper is organized as follows. In Section 2 we provide additional notation, define the auxiliary finite element space W, and state and prove some crucial properties of this space. With these results established, we prove Theorem 1.3 in Section 2.3. In Section 3 we discuss the computational complexity of the auxiliary problem, including the size and structure of the resulting system as well as its spectral properties. The proof of Theorem 1.4 is presented here. Finally, in Section 4, we give several numerical experiments verifying the theoretical results.

2. Reliability Analysis.

2.1. Local and Global Polynomial Spaces for Estimating Error. Given a (non-degenerate) simplex $T \subset \mathbb{R}^d$ of diameter h_T, we define $\mathcal{S}_j(T)$, $0 \le j \le d$ to be the set of sub-simplices of T of dimension j; its cardinality is $\mathbb{S}_j(T) = \binom{d+1}{j+1}$. We denote by $\mathcal{S}_j$ the set of subsub-simplex-simplices of the triangulation of dimension j, and point out the overlap of notation $\mathcal{S}_{d-1} = \mathcal{F}_I \cup \mathcal{F}_D \cup \mathcal{F}_N$ and $\mathcal{S}_d = \mathcal{T}$.

Recall that $\mathcal{P}_m(S)$ is the set of polynomials of total degree $\le m$ with domain S, and note that $\dim \mathcal{P}_m(S) = \binom{m+j}{j}$ for $S \in \mathcal{S}_j(T)$. Taking the vertices of T to be $\{z_0, \ldots, z_d\}$; we let $\lambda_i \in \mathcal{P}_1(\mathbb{R}^d)$, $0 \le i \le d$, be the corresponding barycentric coordinates, uniquely defined by the relations $\lambda_i(z_j) = \delta_{ij}$. We let the *face* $F_j \in \mathcal{S}_{d-1}(T)$ be the sub-simplex not containing z_j, and n_j be the outward-pointing unit normal to F_j.

Definition 2.1 (Element and Face Bubbles). *The fundamental element and face bubbles for T are given by ($j = 0, 1, \ldots, d$)*

$$b_T = \prod_{k=0}^{d} \lambda_k \in \mathcal{P}_{d+1}(T) \quad , \quad b_{F_j} = \prod_{\substack{k=0 \\ k \neq j}}^{d} \lambda_k \in \mathcal{P}_d(T) \ . \tag{2.1}$$

We also define general volume and face bubbles of degree m,

$$Q_{m,T} = \{v = b_T w \in \mathcal{P}_m(T) : \ w \in \mathcal{P}_{m-d-1}(T)\} \tag{2.2}$$

$$Q_{m,F_j} = \{v = b_{F_j} w \in \mathcal{P}_m(T) : \ w \in \mathcal{P}_{m-d}(T)\} \ominus Q_{m,T} \ . \tag{2.3}$$

The functions in $Q_{m,T}$ are precisely those in $\mathcal{P}_m(T)$ which vanish on ∂T; and the functions in Q_{m,F_j} are precisely those in $\mathcal{P}_m(T)$ which vanish on $\partial T \setminus F_j$, with the additional constraint that, if $v \in Q_{m,F_j}$ and v vanishes on F_j, then v vanishes on T. It is evident from their definitions that

(2.4)

$$Q_{m,T} \cap Q_{m,F_j} = \{0\} \quad , \quad Q_{m,F_i} \cap Q_{m,F_j} = \{0\} \text{ for } i \neq j \ ,$$

$$\dim Q_{m,T} = \dim \mathcal{P}_{m-d-1}(T) = \binom{m-1}{d} \ , \tag{2.5}$$

$$\dim Q_{m,F_j} = \dim \left(\mathcal{P}_{m-d}(T) \ominus \mathcal{P}_{m-d-1}(T)\right) = \binom{m}{d} - \binom{m-1}{d} = \binom{m-1}{d-1} \ . \tag{2.6}$$

Here and elsewhere, we use the conventions that $\binom{n}{k} = 0$ when $k > n$, and $\mathcal{P}_n = \{0\}$ when $n < 0$. It will be useful to characterize the volume and face bubbles in terms of moments, as we do in the following lemma.

Lemma 2.2. *A function $v \in Q_{m,T}$ is uniquely determined by the moments*

$$\int_T v\kappa \quad , \quad \forall \kappa \in \mathcal{P}_{m-d-1}(T) \ , \tag{2.7}$$

and a function $v \in Q_{m,F_j}$ is uniquely determined by the moments

$$\int_{F_j} v\kappa \quad , \quad \forall \kappa \in \mathcal{P}_{m-d}(F_j) \ . \tag{2.8}$$

Proof. As is shown, for example, in [4], a function $v \in \mathcal{P}_m(T)$ is uniquely determined by the moments

$$\int_S v\kappa \quad , \quad \forall \kappa \in \mathcal{P}_{m-\ell-1}(S) \ , \ \forall S \in \mathcal{S}_\ell(T) \ , \ 0 \le \ell \le d \ , \tag{2.9}$$

where $\int_S v\kappa$ with $S \in \mathcal{S}_0(T)$ is understood to be the evaluation of v at the vertex S. Since $v \in Q_{m,T}$ vanishes on S for $S \in \mathcal{S}_j(T)$ and $j < d$, v is determined by the moments on T alone. Similarly, any $v \in \{v = b_{F_j} w \in \mathcal{P}_m(T) : \ w \in \mathcal{P}_{m-d}(T)\}$ is uniquely determined by its moments on T and F_j, so any $v \in Q_{m,F_j}$ is uniquely determined by its moments on F_j alone. ∎

Definition 2.3 (Local Error Space). *Given $m \in \mathbb{N}$, we define the local space*

$$R_m(T) = Q_{m,T} \oplus \left(\bigoplus_{j=0}^{d} Q_{m-1,F_j} \right) \ . \tag{2.10}$$

Given $p \in \mathbb{N}$, we define the local error space

$$W_{p+d}(T) = (Q_{p+d,T} \ominus Q_{p,T}) \oplus \left(\bigoplus_{j=0}^{d} (Q_{p+d-1,F_j} \ominus Q_{p,F_j}) \right) = R_{p+d}(T) \ominus R_p(T) \ , \tag{2.11}$$

so that $\mathcal{P}_p(T) + R_{p+d}(T) = \mathcal{P}_p(T) \oplus W_{p+d}(T)$. The dimension of $W_{p+d}(T)$ is readily deduced from (2.4)–(2.6),

$$\dim W_{p+d}(T) = \binom{p+d-1}{d} - \binom{p-1}{d} + (d+1)\left(\binom{p+d-2}{d-1} - \binom{p-1}{d-1} \right) \ . \tag{2.12}$$

We further note that, by Lemma 2.2 and its proof, a function $v \in R_{p+d}(T)$ is uniquely determined by the values

$$\int_S v\kappa \qquad \forall \kappa \in \mathcal{P}_{p-1}(S), \ \forall S \in \mathcal{S}_\ell(T), \ d-1 \le \ell \le d. \tag{2.13}$$

Remark 2.4. *Starting with the standard basis for $\mathcal{P}_1(T)$, $\{\lambda_j : 0 \le j \le d\}$, a p-hierarchical basis for $\mathcal{P}_m(T)$, $m > 1$, is built from a p-hierarchical basis for $\mathcal{P}_{m-1}(T)$ by adding basis functions of degree m. Three approaches to such constructions, at least in $d = 2, 3$, are described in [25, 11, 8], with a useful summary of the constructions from [11, 25] provided in [1]. In these constructions, hierarchical basis functions are associated with each sub–simplex $S \in \mathcal{S}_\ell$, $0 \le \ell \le d$, so it is simple in this setting to construct a basis for $W_{p+d}(T)$.*

The corresponding global finite element spaces, defined by the degrees of freedom and local spaces, are given by

$$\begin{aligned} R_{p+d} &:= \{v \in H^1_{0,D} : \ v|_T \in R_{p+d}(T) \text{ for each } T \in \mathcal{T}\}, \\ W_{p+d} &:= \{w \in H^1_{0,D}(\Omega) : \ v|_T \in W_{p+d}(T) \text{ for each } T \in \mathcal{T}\}, \end{aligned}$$

and we recall that the Lagrange finite element space V_p is defined by (1.3). Similar to the local setting, the global spaces satisfy the relation $R_{p+d} = R_p \oplus W_{p+d}$ so that $V_p + R_{p+d} = V_p \oplus W_{p+d}$.

2.2. A Quasi-Interpolant Based on Moment Conditions. LEMMA 2.5. *Given $v \in H^1(\Omega)$, there exits a $\hat{v} \in V_p$ and $\hat{w} \in W_{p+d}$ such that*

(i) $\int_T (v - \hat{v} - \hat{w})\kappa = 0$ *for all* $\kappa \in \mathcal{P}_{p-1}(T)$ *and* $T \in \mathcal{T}$.
(ii) $\int_F (v - \hat{v} - \hat{w})\kappa = 0$ *for all* $\kappa \in \mathcal{P}_{p-1}(F)$ *and* $F \in \mathcal{F}_I \cup \mathcal{F}_N$.
(iii) $|v - \hat{v} - \hat{w}|_{m,T} \leq C h_T^{1-m} |v|_{1,\Omega_T}$, *where* Ω_T *is a local patch of elements containing* T.
(iv) $|v - \hat{v} - \hat{w}|_{0,F} \leq C h_F^{1/2} |v|_{1,\Omega_F}$, *where* h_F *is the diameter of* $F \in \mathcal{F}$, *and* $\Omega_F = \Omega_T$ *for some* $T \in \mathcal{T}$ *with* $F \subset \partial T$.
(v) $|\hat{w}|_{1,T} \leq C|v|_{1,\Omega_T}$ *for each* $T \in \mathcal{T}$.

Proof. Since functions in $R_{p+d}(T)$ are uniquely determined by the values (2.13), the function $|\!|\!| \cdot |\!|\!|_{m,T} : R_{p+d}(T) \to \mathbb{R}^+$ defined by

$$|\!|\!|\phi|\!|\!|_{m,T} = \max_{\substack{S \in \mathcal{S}_\ell(T) \\ d-1 \leq \ell \leq d}} \sup_{\kappa \in \mathcal{P}_{p-1}(S)} \frac{h_T^{d/2-\ell/2-m}}{\|\kappa\|_{0,S}} \int_S \phi\kappa \tag{2.14}$$

is a norm on $R_{p+d}(T)$.

Let $\tilde{T} = \{y = h_T^{-1}x : \ x \in T\}$, and for each $\psi : T \to \mathbb{R}$, define $\tilde{\psi} : \tilde{T} \to \mathbb{R}$ by $\psi(y) = \psi(h_T x)$. Analogous definitions are given for the sub-simplices of T and $\tilde{T}$ and functions defined on them. It is clear that $|\phi|_{j,T} = h_T^{d/2-j} |\tilde{\phi}|_{j,\tilde{T}}$. We also have for any $S \in \mathcal{S}_\ell(T)$

$$\frac{h_T^{d/2-\ell/2-m}}{\|\kappa\|_{0,S}} \int_S \phi\kappa = \frac{h_T^{d/2-\ell/2-m}}{h_T^{\ell/2}\|\tilde{\kappa}\|_{0,\tilde{S}}} \int_{\tilde{S}} \tilde{\phi}\tilde{\kappa} h_T^\ell = \frac{h_T^{d/2-m}}{\|\tilde{\kappa}\|_{0,\tilde{S}}} \int_{\tilde{S}} \tilde{\phi}\tilde{\kappa}.$$

Since $h_{\tilde{T}} = 1$, we see that $|\!|\!|\phi|\!|\!|_{m,T} = h_T^{d/2-m} |\!|\!|\tilde{\phi}|\!|\!|_{m,\tilde{T}}$. Therefore there exists a scale-invariant constant $C > 0$ which depends solely on p, d and m such that

$$|\phi|_{m,T} = h_T^{d/2-m} |\tilde{\phi}|_{m,\tilde{T}} \leq C h_T^{d/2-m} |\!|\!|\tilde{\phi}|\!|\!|_{m,\tilde{T}} = C |\!|\!|\phi|\!|\!|_{m,T}. \tag{2.15}$$

Next, denote by $\hat{v}_1 \in V_p$ the Scott-Zhang interpolant of v satisfying [24]

$$\|v - \hat{v}\|_{m,T} \leq C h_T^{1-m} |\hat{v}|_{1,\Omega_T} \quad (m = 0, 1) \ , \tag{2.16a}$$

$$\|v - \hat{v}\|_{0,\partial T} \leq C h_T^{1/2} |\hat{v}|_{1,\Omega_T} \ , \tag{2.16b}$$

on each $T \in \mathcal{T}$. Set $\hat{v}_2 \in R_{p+d}$ such that

$$\int_S \hat{v}_2 \kappa = \int_S (v - \hat{v}_1)\kappa \qquad \forall \kappa \in \mathcal{P}_{p-1}(S), \ \forall S \in \mathcal{S}_\ell, \ d - 1 \leq \ell \leq d.$$

By (2.15) and (2.16) we find

$$\begin{aligned} |\hat{v}_2|_{H^m(T)} &\leq C \max_{\substack{S \in \mathcal{S}_\ell(T) \\ d-1 \leq \ell \leq d}} \sup_{\kappa \in \mathcal{P}_{p-1}(S)} \frac{h_T^{d/2-\ell/2-m}}{\|\kappa\|_{0,S}} \int_S \hat{v}_2 \kappa \\ &= C \max_{\substack{S \in \mathcal{S}_\ell(T) \\ d-1 \leq \ell \leq d}} \sup_{\kappa \in \mathcal{P}_{p-1}(S)} \frac{h_T^{d/2-\ell/2-m}}{\|\kappa\|_{0,S}} \int_S (v - \hat{v}_1)\kappa \\ &\leq C\big(h_T^{1/2-m} \|v - \hat{v}_1\|_{0,\partial T} + h_T^{-m} \|v - \hat{v}_1\|_{0,T}\big) \leq C h_T^{1-m} |v|_{1,\Omega_T}. \end{aligned}$$

Uniquely decomposing $\hat{v}_2$ as $\hat{v}_2 = \hat{v}_3 + \hat{w}$ with $\hat{v}_3 \in R_p$ and $\hat{w} \in W_{p+d}$, and setting $\hat{v} := \hat{v}_1 + \hat{v}_3$ so that $\hat{v} + \hat{w} = \hat{v}_1 + \hat{v}_2$, we see that properties (i)–(ii) clearly hold, and

$$\|v - \hat{v} - \hat{w}\|_{m,T} \leq \|v - \hat{v}_1\|_{m,T} + \|\hat{v}_2\|_{m,T} \leq C h_T^{1-m} |v|_{1,\Omega_T}.$$

Therefore by standard trace inequalities and the shape regularity of the mesh, we also have on $F \subset \partial T$

$$\|v - \hat{v} - \hat{w}\|_{0,F} \leq C\big(h_F^{-1/2}\|v - \hat{v} - \hat{w}\|_{0,T} + h_F^{1/2}|v - \hat{v} - \hat{w}|_{1,T}\big) \leq C h_F^{1/2}|v|_{1,\Omega_F}.$$

Hence, properties (iii)–(iv) are satisfied.

Finally, since $R_p(T) \cap W_{p+d}(T) = \{0\}$, the strengthened Cauchy–Schwarz inequality [2] gives the existence of a constant $\gamma \in [0,1)$ such that $\int_T \nabla\hat{w}\cdot\nabla\hat{v}_3 \leq \gamma|\hat{w}|_{1,T}|\hat{v}_3|_{1,T}$ Consequently, we have

$$\begin{aligned} |\hat{v}_2|_{1,T}^2 &= |\hat{w}|_{1,T}^2 + |\hat{v}_3|_{1,T}^2 + 2\int_T \nabla\hat{w}\cdot\nabla\hat{v}_3 \\ &\geq |\hat{w}|_{1,T}^2 + |\hat{v}_3|_{1,T}^2 - 2\gamma|\hat{w}|_{1,T}|\hat{v}_3|_{1,T} \geq (1-\gamma^2)|\hat{w}|_{1,T}^2. \end{aligned}$$

Therefore we find $|\hat{w}|_{H^1(T)} \leq \sqrt{(1-\gamma^2)^{-1}}|\hat{v}_2|_{H^1(T)} \leq C|v|_{H^1(\Omega_T)}$. ∎

Remark 2.6. *The moment conditions (i)-(ii) of Lemma 2.5 imply the conditions*

$$\int_T \nabla(v - \hat{v} - \hat{w})\cdot\phi = 0 \text{ for all } \phi \in RT_{p-1}(T) \text{ and all } T \in \mathcal{T}\ , \tag{2.17}$$

where $RT_{p-1}(T) = x\mathcal{P}_{p-1}(T) + [\mathcal{P}_{p-1}(T)]^d = \{\phi = \sum_{j=0}^d (x - z_j)\kappa_j : \kappa_j \in \mathcal{P}_{p-1}(T)\}$ *is the local Raviart-Thomas space. Recalling the vertex, face and normal vector notation above, this equivalence is most readily seen through the following simple consequence of integration-by-parts on a simplex:*

$$\text{For } f \in H^1(T),\ \int_T (x - z_j)\cdot\nabla f = a_j \int_{F_j} f - d\int_T f\ , \tag{2.18}$$

where a_j *is the distance (altitude) between* z_j *and* F_j*. Choosing* $f = (v - \hat{v} - \hat{w})\kappa_j$ *for* $\kappa_j \in \mathcal{P}_{p-1}(T)$*, and combining results for each* j*, makes the comparison between* (2.17) *and (i)-(ii) apparent. The conditions* (2.17) *are not independent, so they do not impose* $\dim(RT_{p-1}(T)) = d\binom{p+d-1}{d} + \binom{p+d-2}{d-1}$ *independent constraints on* $R_{p+d}(T)$*, whose dimension,* $\binom{p+d-1}{d} + (d+1)\binom{p+d-2}{d-1}$*, is generally smaller.*

2.3. Proof of Theorem 1.3. *Proof.* [Proof of Theorem 1.3] Combining Proposition 1.1 and Lemma 2.5, we determine that

$$\begin{aligned} |B(u - \hat{u}, v)| &\leq |B(\varepsilon, \hat{w})| + \sum_{T\in\mathcal{T}} \inf_{\kappa\in\mathcal{P}_{p-1}(T)} \|R_T - \kappa\|_{0,T}\|v - \hat{v} - \hat{w}\|_{0,F} \\ &\qquad + \sum_{F\in\mathcal{F}_I\cup\mathcal{F}_N} \inf_{\kappa\in\mathcal{P}_{p-1}(F)} \|r_F - \kappa\|_{0,F}\|v - \hat{v} - \hat{w}\|_{0,F} \\ &\lesssim \|\varepsilon\|_1\|\hat{w}\|_1 + \sum_{T\in\mathcal{T}} h_T\|v\|_{1,\Omega_T} \inf_{\kappa\in\mathcal{P}_{p-1}(T)} \|R_T - \kappa\|_{0,T} \\ &\qquad + \sum_{F\in\mathcal{F}_I\cup\mathcal{F}_N} h_F^{1/2}\|v\|_{1,\Omega_T} \inf_{\kappa\in\mathcal{P}_{p-1}(F)} \|r_F - \kappa\|_{0,F} \\ &\lesssim \|\varepsilon\|_1\|v\|_1 + \mathrm{osc}(R, r, \mathcal{T})\|v\|_1\ . \end{aligned}$$

For the final inequality, we have used Lemma 2.5 (v), the (discrete) Cauchy-Schwarz Inequality and the bounded overlap of the patches Ω_T and Ω_F (which is also a consequence of shape-regularity). Finally, we choose $v = u - \hat{u}$ and use the coercivity of B to complete the proof. ∎

REMARK 2.7. *We note that the continuity constant enters in the bound* $|B(\varepsilon, \hat{w})| \leq \mathfrak{C}\|\varepsilon\|_{1,\Omega}\|\hat{w}\|_1$, *and only affects the term* $\|\varepsilon\|_1$ *in the reliability bound of Theorem 1.3. The coercivity constant* $\mathfrak{c}$ *affects both terms in the reliability bound.*

REMARK 2.8. *Although our approach is analyzed as an h-method with global fixed p, the general approach is very naturally adjusted to both p and hp-methods. As indicated in the introduction, the driving motivation for the choice of* $W_{p+d}(T)$ *is to make sure that the local oscillation is of higher order than the local best approximation error. The development suggests that, if the local approximation space is* $V(T) = \mathcal{P}_{p_T}(T)$, *then the local error space* $W(T)$ *should be spanned by face bubbles of degree* $p_T + d - 1$ *and interior bubbles of degree* $p_T + d$ *which are not already represented in* $\mathcal{P}_{p_T}(T)$. *Again, although our approach is analyzed for simplicial elements, the shapes of the elements are irrelevant for much of our development. In particular, it is straight-forward to apply the prescription above for choosing* $W(T)$ *on tensor-product elements such as quadrilaterals or bricks. In Section 4 we investigate our approach as a p-method on meshes which include tensorial elements.*

3. Computational Considerations. As presented above, the computation of ε requires the solution of a global system involving the stiffness matrix associated with W_{p+d}. At first glance this would seem to rule out the approach as too expensive for practical computations, but we argue herein that this is not the case. Our argument is based on considerations of sparsity structure and size of the linear systems, and on their spectral properties. Using standard (p-hierarchical) bases for the spaces V_p and W_{p+d}, we compare and contrast the corresponding global and element stiffness matrices. We assume that global stiffness matrices are assembled by summing contributions from element stiffness matrices computed on each simplex $T \in \mathcal{T}$.

3.1. Size and Sparsity Structure. We begin by comparing the sizes of the element stiffness matrices for $V_p(T)$ and $W_{p+d}(T)$, as well as the amount of information which must be transferred to the global stiffness matrices in each case if static condensation is used locally to eliminate interior degrees of freedom. Letting $n = n(p,d)$ and $m = m(p,d)$ be the number of degrees of freedom associated with $V_p(T)$ and $W_{p+d}(T)$, respectively, and $\hat{n} = \hat{n}(p,d)$ and $\hat{m} = \hat{m}(p,d)$ denote the analogous quantities after interior degrees of freedom have been eliminated, we have

(3.1)
$$n = \binom{p+d}{d} \quad , \quad m = \binom{p+d-1}{d} - \binom{p-1}{d} + (d+1)\left(\binom{p+d-2}{d-1} - \binom{p-1}{d-1}\right) ,$$

(3.2)
$$\hat{n} = \binom{p+d}{d} - \binom{p-1}{d} \quad , \quad \hat{m} = (d+1)\left(\binom{p+d-2}{d-1} - \binom{p-1}{d-1}\right) .$$

We note that n is a polynomial of degree d in p and m is a polynomial of degree $d-1$ in p, so it is clear that $n > m$ when p is large enough, for any fixed d. The polynomial degrees for $\hat{n}$ and $\hat{m}$ are of degrees $d-1$ and $d-2$ in p, respectively. In Table 1 we list values of the the four quantities (3.1)-(3.2) for $1 \leq p \leq 7$ and $d = 2, 3$.

Recall that $\mathcal{S}_j$ denotes the set of subsimplices of dimension j in $\mathcal{T}$, and $\mathbb{S}_j$ denotes its cardinality, $0 \leq j \leq d$. Without static condensation to eliminate the degrees of

TABLE 1
Size of the local stiffness matrices for $V_p(T)$ and $W_{p+d}(T)$ with and without static condensation, for $d = 2, 3$.

	$d=2$				$d=3$			
p	n	m	$\hat{n}$	$\hat{m}$	n	m	$\hat{n}$	$\hat{m}$
1	3	4	3	3	4	5	4	4
2	6	6	6	3	10	16	10	12
3	10	8	9	3	20	30	20	20
4	15	10	12	3	35	47	34	28
5	21	12	15	3	56	67	52	36
6	28	14	18	3	84	90	74	44
7	36	16	21	3	120	116	100	52

freedom associated with the interiors of each $T \in \mathcal{T}$, the sizes of the global stiffness matrices for V_p and W_{p+d} are, respectively,

$$N = \sum_{j=0}^{d} \mathbb{S}_j \binom{p-1}{j}\ ,\ M = \mathbb{S}_d \left(\binom{p+d-1}{d} - \binom{p-1}{d} \right) + \mathbb{S}_{d-1} \left(\binom{p+d-2}{d-1} - \binom{p-1}{d-1} \right)\ .$$

When the interior degrees of freedom are eliminated, the sizes become

$$\hat{N} = \sum_{j=0}^{d-1} \mathbb{S}_j \binom{p-1}{j} \quad , \quad \hat{M} = \mathbb{S}_{d-1} \left(\binom{p+d-2}{d-1} - \binom{p-1}{d-1} \right)\ .$$

The formulas count degrees of freedom on Γ_D, though these are not truly unknowns in the problem, because many practical implementations proceed in this way when assembling global matrices, and encode Dirichlet boundary conditions in the system as a final step. Recognizing that $\binom{p-1}{d}$ and $\binom{p-1}{d-1}$ are polynomials of degree d and $d-1$ in p, respectively, we see again that, for any fixed d, $N > M$ and $\hat{N} > \hat{M}$ for sufficiently large p. To illustrate this, consider a standard uniform triangulation of the unit square by isosceles right triangles (half-squares) with side-length $1/s$. For such triangulations, $N = (p(s-1)+1)^2$ and $M = 4p(s-1)^2 + (s^2-1)$, so $N > M$ for all $s \geq 2$ when $p \geq 5$. For such triangulations we also have $\hat{N} = (3p-2)s^2 - 4(p-1)s + p - 1$ and $\hat{M} = 3s^2 - 4s + 1$, so $\hat{N} > \hat{M}$ for all $s \geq 2$ when $p \geq 2$.

We now turn to the discussion of sparsity for the global matrices for V_p and W_{p+d}. Given $S \in \mathcal{S}_j$, let $\mathcal{T}_S$ be the set of simplices $T \in \mathcal{T}$ which have S as a sub-simplex. We also define $\mathcal{S}_i(\mathcal{T}_S) = \cup_{T \in \mathcal{T}_S} \mathcal{S}_i(T)$ and denote its cardinality by $\mathbb{S}_i(\mathcal{T}_S)$. If $S = T \in \mathcal{S}_d = \mathcal{T}$, then $\mathcal{T}_S = \{T\}$ and $\mathbb{S}_i(\mathcal{T}_S) = \binom{d+1}{i+1}$. If $S = F \in \mathcal{S}_{d-1}$, then $\mathcal{T}_S$ consists of the one or two simplices which have F as a face; in the first case $\mathbb{S}_i(\mathcal{T}_S) = \binom{d+1}{i+1}$ as before, and in the second $\mathbb{S}_i(\mathcal{T}_S) = 2\binom{d+1}{i+1} - \binom{d}{i+1}$. For $j < d-1$, the cardinalities of $\mathcal{T}_S$ and $\mathcal{S}_i(\mathcal{T}_S)$ for a given $S \in \mathcal{S}_j$ cannot be determined *a priori* for general unstructured meshes. To compute the sparsity structure of the global stiffness matrix, the sets $\mathcal{S}_i(\mathcal{T}_S)$ for each $S \in \mathcal{S}_j$ (and each i and j) must be determined, at least indirectly. For W_{p+d} this task is greatly simplified by the fact that we need only consider $\mathcal{S}_i(\mathcal{T}_S)$ for each $S \in \mathcal{S}_j$ with $i, j \in \{d-1, d\}$—these are the two cases which are easiest to resolve! More specifically, let $\phi \in W_{p+d}$ be a basis function, and let S be the (sub-)simplex of minimal dimension $j \in \{d-1, d\}$ on which ϕ does not vanish identically. The number of possible non-zeros in the the row of the matrix

corresponding to ϕ is

$$\mathbb{S}_{d-1}(\mathcal{T}_S)\left(\binom{p+d-2}{d-1}-\binom{p-1}{d-1}\right)+\mathbb{S}_{d-1}(\mathcal{T}_S)\left(\binom{p+d-1}{d}-\binom{p-1}{d}\right)\ .$$

If static condensation is used, the number of non-zeros in a row for ϕ associated with an interior face is

$$(2d+1)\left(\binom{p+d-2}{d-1}-\binom{p-1}{d-1}\right)\ .$$

For a boundary face, $2d+1$ is replaced by $d+1$. We see that, in the case of W_{p+d}, the sparsity structure is known in advance. For example, when $d=2$ and static condensation is used, the number of non-zeros in any row does not exceed 5, regardless of p and the mesh topology. When $d=3$ and static condensation is used, the number of non-zeros in any row does not exceed $7(2p-1)$.

For comparison, we briefly discuss the situation for V_p. Let $\phi\in V_p$ be a basis function, and let S be the (sub-)simplex of minimal dimension j on which ϕ does not vanish identically. The number of possible non-zeros in the the row of B corresponding to ϕ, and the total number of possible non-zeros are, respectively,

$$\sum_{i=0}^{d}\mathbb{S}_i(\mathcal{T}_S)\binom{p-1}{i}\quad,\quad\sum_{j=0}^{d}\sum_{S\in\mathcal{S}_j}\sum_{i=0}^{d}\mathbb{S}_i(\mathcal{T}_S)\binom{p-1}{i}\ .$$

If static condensation is used to eliminate interior degrees of freedom, the sums are terminated at $d-1$ instead instead of d.

3.2. Spectral Behavior of the Stiffness Matrix for W_{p+d}. We argue in Theorem 1.4 and Remark 3.4 below that the spectral behavior of the global stiffness matrix for W_{p+d}, with or without static condensation, makes it amenable to solution techniques which are simpler/faster than those for V_p. In brief, the conditioning of the stiffness matrix for W_{p+d}, perhaps after simple diagonal rescaling, does not deteriorate as the triangulation is refined, unlike that for V_p. Before specifically commenting on the spectral behavior of the global stiffness matrix for W_{p+d}, we first make comparison with matrices arising from the H^1-inner-product for a general finite dimensional subspace $X\subset H^1_{0,D}(\Omega)$, having basis $\{\phi_i:1\le i\le N\}$. We define the stiffness matrices

$$B_{ij}=B(\phi_j,\phi_i)\quad,\quad\hat{B}_{ij}=(\phi_j,\phi_i)_1=\int_\Omega\nabla\phi_j\cdot\nabla\phi_i+\phi_j\phi_i\ .$$

Making the obvious identification between $\boldsymbol{v}\in\mathbb{R}^N$ and $v\in X$, we see that

$$B(v,w)=\boldsymbol{w}^tB\boldsymbol{v}\quad,\quad(v,w)_1=\boldsymbol{w}^t\hat{B}\boldsymbol{v}\ .$$

Stated in terms of the matrices B and $\hat{B}$, the continuity and coercivity of the bilinear form B are

$$|\boldsymbol{w}^tB\boldsymbol{v}|\le\mathfrak{C}\left(\boldsymbol{v}^t\hat{B}\boldsymbol{v}\right)^{1/2}\left(\boldsymbol{w}^t\hat{B}\boldsymbol{w}\right)^{1/2}\quad,\quad\boldsymbol{v}^tB\boldsymbol{v}\ge\mathfrak{c}\boldsymbol{v}^t\hat{B}\boldsymbol{v}\quad\forall\boldsymbol{v},\boldsymbol{w}\in\mathbb{R}^N\ .$$

Proposition 3.1. *Let $\mu=\mu_1+i\mu_2$, $\mu_1,\mu_2\in\mathbb{R}$, be an eigenvalue of B. Then*

$$\mathfrak{c}\lambda_{\min}(\hat{B})\le\mu_1\le\mathfrak{C}\lambda_{\max}(\hat{B})\quad,\quad|\mu_2|\le\mathfrak{C}\lambda_{\max}(\hat{B})\ .$$

Proof. Let $\boldsymbol{v} = \boldsymbol{v}_1 + i\boldsymbol{v}_2$, $\boldsymbol{v}_1, \boldsymbol{v}_2 \in \mathbb{R}^N$, be an eigenvector for μ; and assume, without loss of generality, that $\|\boldsymbol{v}\| = 1$, where $\|\cdot\|$ is the Euclidean norm on $\mathbb{C}^N$. It is straightforward to show that

$$\boldsymbol{v}_1^t B \boldsymbol{v}_1 + \boldsymbol{v}_2^t B \boldsymbol{v}_2 = \mu_1 \quad , \quad \boldsymbol{v}_1^t B \boldsymbol{v}_2 - \boldsymbol{v}_2^t B \boldsymbol{v}_1 = \mu_2 \ .$$

So we see that

$$\mathfrak{c}\lambda_{\min}(\hat{B}) \leq \mathfrak{c}\left(\boldsymbol{v}_1^t \hat{B} \boldsymbol{v}_1 + \boldsymbol{v}_2^t \hat{B} \boldsymbol{v}_2\right) \leq \mu_1 \leq \mathfrak{C}\left(\boldsymbol{v}_1^t \hat{B} \boldsymbol{v}_1 + \boldsymbol{v}_2^t \hat{B} \boldsymbol{v}_2\right) \leq \mathfrak{C}\lambda_{\max}(\hat{B}) \ .$$

Furthermore,

$$|\mu_2| \leq 2|\boldsymbol{v}_2^t B \boldsymbol{v}_1| \leq 2\mathfrak{C}\left(\boldsymbol{v}_1^t \hat{B} \boldsymbol{v}_1\right)^{1/2}\left(\boldsymbol{v}_2^t \hat{B} \boldsymbol{v}_2\right)^{1/2} \leq 2\mathfrak{C}\lambda_{\max}(\hat{B})\|\boldsymbol{v}_1\|\|\boldsymbol{v}_2\| \leq \mathfrak{C}\lambda_{\max}(\hat{B}) \ .$$

We have used the Cauchy Inequality, $2ab \leq a^2 + b^2$ for the final inequality above. □

To get a better handle on the spectral properties of $\hat{B}$, we we now consider element stiffness matrices for X. Given $T \in \mathfrak{T}$, let $I(T) = \{j : T \cap \mathrm{supp}(\phi_j) \neq \emptyset\}$ and $n = n(T) = |I(T)|$. We define $\hat{B}_T \in \mathbb{R}^{n\times n}$ via

$$(\hat{B}_T)_{ij} = (\phi_j, \phi_i)_{1,T} = \int_T \nabla\phi_j \cdot \nabla\phi_i + \phi_j\phi_i \text{ for } i, j \in I(T) \ .$$

Again making the obvious association between $\boldsymbol{v} \in \mathbb{R}^N$ and $v \in X$, we define $\boldsymbol{v}_T \in \mathbb{R}^n$ such that $\|v\|_{1,T}^2 = \boldsymbol{v}_T^t \hat{B}_T \boldsymbol{v}_T$; it is clear that $\boldsymbol{v}_T$ consists of the components of $\boldsymbol{v}$ whose indices are in $I(T)$. We also define $\hat{D} = \mathrm{diag}(\hat{B})$ and $\hat{D}_T = \mathrm{diag}(\hat{B}_T)$. It is apparent from these definitions that

$$\boldsymbol{v}^t \hat{B} \boldsymbol{v} = \sum_{T\in\mathfrak{T}} \boldsymbol{v}_T^t \hat{B}_T \boldsymbol{v}_T \quad , \quad \boldsymbol{v}^t \hat{D} \boldsymbol{v} = \sum_{T\in\mathfrak{T}} \boldsymbol{v}_T^t \hat{D}_T \boldsymbol{v}_T \ .$$

The next result follows immediately from this discussion.

PROPOSITION 3.2. *Suppose there are constants $c, C > 0$ such that $c \leq \frac{\boldsymbol{w}^t \hat{B}_T \boldsymbol{w}}{\boldsymbol{w}^t \hat{D}_T \boldsymbol{w}} \leq C$ for all non-zero $\boldsymbol{w} \in \mathbb{R}^n$. Then $c \leq \frac{\boldsymbol{w}^t \hat{B} \boldsymbol{w}}{\boldsymbol{w}^t \hat{D} \boldsymbol{w}} \leq C$ for all non-zero $\boldsymbol{w} \in \mathbb{R}^N$. As a consequence, the spectrum of $\hat{D}^{-1/2}\hat{B}\hat{D}^{-1/2}$ is contained in $[c, C]$.*

Suppose $X = W_{p+d}$ and we use a hierarchical basis (cf. Remark 2.4). Fixing $T \in \mathfrak{T}$ and using the corresponding basis for $W_{p+d}(T)$, we may use simple scaling arguments to see that that $\hat{B}_T$ can be expressed in the form

$$\hat{B}_T = h_T^{d-2} B_1 + h_T^d B_2 \ ,$$

where B_1, B_2 depend only on p, d and the shape-regularity of T. The matrix B_1, whose entries are $h_T^{2-d}\int_T \nabla\phi_j \cdot \nabla\phi_i$, has full-rank because $(\cdot,\cdot)_{1,T}$ is an inner-product on $W_{p+d}(T)$. The matrix B_2, whose entries are $h_T^{-d}\int_T \phi_j\phi_i$ is clearly a full-rank Gram matrix. This implies that there are constants $c_T, C_T > 0$ depending only on p, d and the shape-regularity of T for which

$$c_T \boldsymbol{w}^t \hat{D}_T \boldsymbol{w} \leq \boldsymbol{w}^t \hat{B}_T \boldsymbol{w} \leq C_T \boldsymbol{w}^t \hat{D}_T \boldsymbol{w} \text{ for all } \boldsymbol{w} \in \mathbb{R}^n \ . \tag{3.3}$$

Invoking the shape-regularity of the family $\{\mathfrak{T}\}$, we can replace the local constants c_T, C_T with universal constants c, C and apply Proposition 3.2. We are now ready to prove Theorem 1.4 our key result concerning the spectral properties of B for W_{p+d}:

Proof. [Proof of Theorem 1.4] Letting $\hat{B}$ and $\hat{D}$ be as in the discussion above, and D be the diagonal of B, we have already seen in Proposition 3.1 that B and $\hat{B}$ are spectrally equivalent to each other. It is trivial to see that D and $\hat{D}$ are spectrally equivalent to each other. So, to prove that B and D are spectrally equivalent to each other, we need merely show that $\hat{B}$ and $\hat{D}$ are spectrally equivalent to each other. But this was established by Proposition 3.2 and the discussion that followed. ∎

REMARK 3.3. *Had we chosen $X = V_p$, the corresponding matrix B_1 has a one-dimensional nullspace spanned by the vector $\boldsymbol{e} \in \mathbb{R}^n$ of ones, corresponding to the constant functions in $V_p(T)$. We deduce that $\boldsymbol{e}^T \hat{B}_T \boldsymbol{e} = \int_T 1 = |T| \sim h_T^d$, whereas $\boldsymbol{e}^T \hat{B}_T \boldsymbol{e} \sim h_T^{d-2}$. For any other non-zero $\boldsymbol{w} \in \mathbb{R}^n$, $\boldsymbol{w}^T \hat{B}_T \boldsymbol{w}$ and $\boldsymbol{w}^T \hat{D}_T \boldsymbol{w}$ scale in precisely the same way, so there are no scale-invariant c_T, C_T for which* (3.3) *holds. Therefore, Proposition 3.2 cannot be applied.*

REMARK 3.4 (Effect of Static Condensation). *To analyze the effect of static condensation on the global stiffness matrix B for W_{p+d}, we split the space as $W_{p+d} = W_{p+d,1} \oplus W_{p+d,2}$, where $W_{p+d,1}$ is spanned by the "interior" basis functions—those supported on a single element. This splitting of the space induces the natural 2×2 block structure on B*

$$B = \begin{pmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{pmatrix} ,$$

and we must investigate the spectral properties of the Schur complement $S = B_{22} - B_{21} B_{11}^{-1} B_{12}$. Given $\boldsymbol{z} \in \mathbb{R}^{\hat{M}}$, we extend it to a vector $\tilde{\boldsymbol{z}} \in \mathbb{R}^M$ by appending it to the vector $-B_{11}^{-1} B_{12} \boldsymbol{z} \in \mathbb{R}^{M-\hat{M}}$. For $\boldsymbol{v}, \boldsymbol{w} \in \mathbb{R}^{\hat{M}}$, we have $\boldsymbol{w}^t S \boldsymbol{v} = \tilde{\boldsymbol{w}}^t B \tilde{\boldsymbol{v}}$, so our analysis above suffices to show that the spectral properties of S cannot be worse than those of B.

4. Numerical Experiments. We recall that our results concerning the reliability and computational cost of our estimator were obtained for fixed p and (adaptive) h-refinement on simplicial meshes, with reliability shown in the H^1-norm. In the first subsection, we numerically illustrate these results on a standard test problem in $\mathbb{R}^2$ for modest p. The second subsection is devoted to extensive testing of the robustness of the estimator with respect to polynomial degree. Here we consider the behavior of the error estimator under uniform p-refinement on fixed (adapted) meshes of quadrilaterals/bricks and/or simplices for several different types of problems, one of which is in $\mathbb{R}^3$. We did not use static condensation for any of the linear systems.

A key measure of the quality of the estimator is its *effectivity* in a norm of interest,

$$\text{EFF} = \|\varepsilon\| / \|u - \hat{u}\| .$$

In most cases, we will report effectivities in the global H^1 or appropriate energy norm, because our theory deals with such cases. But as a matter of interest, for the h-refinement study we also report global L^2-effectivity and local H^1-effectivity—the latter of which provides a good measure of the efficiency of local indicators $\|\varepsilon\|_{1,T}$ for driving an adaptive algorithm.

4.1. Verification of Properties of the Estimator Under Adaptive h-Refinement. We consider a prototypical problem on the L-shaped domain

$$-\Delta u = f \text{ in } \Omega = (-1,1)^2 \backslash (0,1) \times (-1,0) \quad , \quad u = 0 \text{ on } \partial\Omega ,$$

with f chosen so that the exact solution is given by $u = r^{2/3} \sin(\frac{2}{3}\theta)(x_1^2 - 1)(x_2^2 - 1)$. This solution exhibits the typical singular behavior at the origin for generic f. We

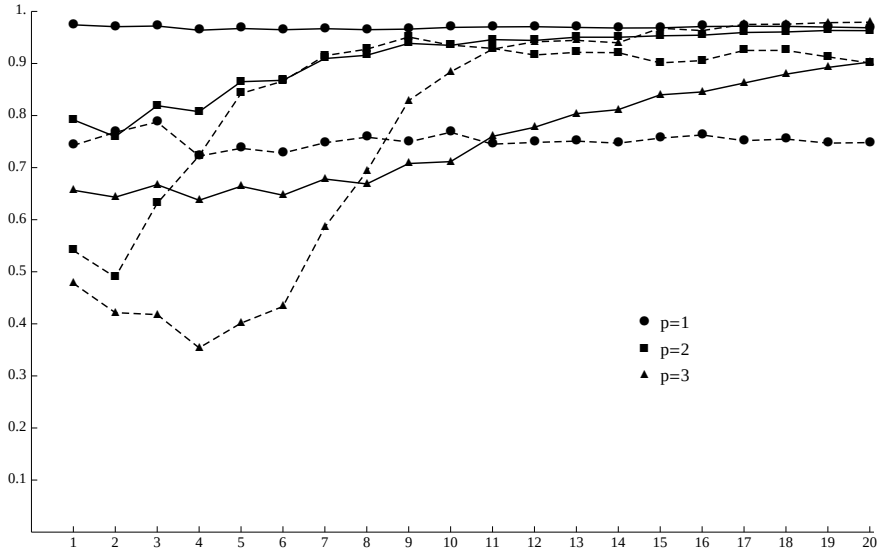

FIG. 1. *Global H^1 (solid) and L^2 (dashed) effectivities for the L-shaped domain on a sequence of 20 adaptively-refined meshes.*

note that, in this case, the oscillation term in the reliability bound reduces to purely data oscillation which has the local form $\mathrm{osc}(R, r, T) = h_T \inf_{\kappa \in \mathcal{P}_{p-1}(T)} \|f - \kappa\|_{0,T}$.

We study the performance of the a posteriori error estimator with an h-refinement algorithm for fixed $p = 1, 2, 3$ on a starting uniform mesh with $h = 1/8$. The marking strategy for refinement is performed in the following manner: if we denote by $T_{\max} \in \mathcal{T}$ the simplex with the largest estimated error, i.e, $\|\varepsilon\|_{H^1(T_{\max})} \geq \|\varepsilon\|_{H^1(T)}\ \forall T \in \mathcal{T}$, then we set T for refinement if $\|\varepsilon\|_{H^1(T)} \geq \gamma \|\varepsilon\|_{H^1(T_{\max})}$ for some user–defined parameter $\gamma \in [0, 1]$. In all of the tests below we take $\gamma = 0.3$.

For each p, relevant data was collected for a sequence of nested 20 meshes obtained by the adaptive scheme described above. Global H^1 and L^2 effectivities are given in Figure 1. The global effectivities in both norms are quite good, with some indication of asymptotic exactness (or at least effectivities very near 1) in H^1 for each p, and in L^2 when $p \geq 2$. In terms of local H^1 effectivities, we observe that the maximum local effectivities range from $[1.02, 2.17]$ for all tested polynomial degrees and for all meshes, which bodes well for their efficiency as local indicators for driving adaptive refinement.

As a verification of the claims of Section 3 we briefly summarize the ratios $\dim W_{p+d} / \dim V_p$ and the condition numbers of the diagonally-rescaled stiffness matrices for W_{p+d}, $B \to D^{-1/2} B D^{-1/2}$. In all cases the largest value of the dimension ratio corresponds to the coarsest mesh, and the smallest ratio to the finest mesh. For $p = 1$, we have $\dim W_{p+d} / \dim V_p \in [5.09, 5.77]$, for $p = 2$ the ratios were in $[2.32, 2.41]$, and for $p = 3$ the ratios were in $[1.49, 1.51]$. Again, for each p and all meshes the computed condition numbers for W_{p+d} remained in a relatively narrow range, neither monotonically increasing nor decreasing as the mesh was refined. For $p = 1$ the range of condition numbers was $[28.4, 37.3]$, for $p = 2$ this range was $[15.0, 16.9]$, and for

$p = 3$ this range was $[30.1, 33.2]$.

4.2. Investigation of Properties of the Estimator Under Uniform p-Refinement. In the experiments that follow, we investigate the behavior of our estimator with respect to uniform p refinement on fixed (adapted) meshes which may consist of quadrilaterals (or bricks), triangles, or a combination of the two. In the case of quadrilateral or brick elements, we use the full tensor-product space indexed by maximal degree in each variable, not a reduced space indexed by total degree. The choice of full tensor-product space more naturally fits with our theoretical development of the error estimator, and it provides better convergence for some of the more challenging problems below. The auxiliary space W_{p+d} for the tensor elements still consists of the interior bubbles of degree up to $p + d$ and face bubbles of degree up to $p + d - 1$ which were not already present in V_p. The problems are chosen to illustrate the behavior of the estimator in a variety of situations in which certain problem-dependent parameters might reasonably affect performance.

In nearly all cases below, we observe that the error estimates stay within a factor of two of the actual errors, and the one case in which the factor reaches roughly 2.5 is where both quantities are smaller than 10^{-15}. The conditioning varied widely between problems due to problem parameters and the use of an integrated Legendre basis for tensor elements versus a standard Legendre basis for triangular elements, but the ratio of condition numbers (W_{p+d} over V_p) indicates that the cost of computing ε is acceptable. For example, for all choices of β in Subsection 4.2.1, the condition number ratios for rectangular elements remained $\mathcal{O}(1)$ and the condition numbers themselves remained $\mathcal{O}(10)$ for all p. For the same problem on triangular elements this ratio decreased steadily to reach $\mathcal{O}(10^{-4})$ when $p = 8$, with the condition number for W_{p+d} at $\mathcal{O}(100)$. The size of the stiffness matrix and number of non-zeros for W_{p+d} tended to drop below that of V_p at either $p = 4$ or $p = 5$ for all 2D problems.

4.2.1. Discontinuous and Anisotropic Diffusion on the Square. Letting $\Omega = (-1, 1) \times (-1, 1)$, we consider problems of the form

$$-\nabla \cdot (A\nabla u) = f \text{ in } \Omega \quad , \quad u = 0 \text{ on } \partial\Omega \quad , \quad A = \begin{pmatrix} \alpha & 0 \\ 0 & 1 \end{pmatrix} \quad , \quad \alpha = \begin{cases} 1 & x < 0 \\ \beta & x > 0 \end{cases} ,$$

for various choices of $\beta > 1$. Because the jump discontinuity in the diffusion matrix happens along a straight line, one does not expect singularities in u for generic f. This allows us to isolate potential effects of varying β on the effectivity of the estimator from those which might arise due to singularities in u—singular solutions are considered in the two subsequent problems. The function f is chosen so that the solution is given by

$$u = \cos(\pi y/2) \begin{cases} \left(e^{-1} - e^{x} + \frac{(e-1)(\beta+e)}{e(\beta+1)}(x+1)\right) & x < 0 \\ \beta^{-1}\left(e - e^{x} + \frac{(e-1)(\beta+e)}{e(\beta+1)}(x-1)\right) & x > 0 \end{cases} .$$

We note that $u = \cos(\pi y/2)w(x)$, where w is the solution of the 1D problem $-(\alpha w')' = e^x$ in $(-1, 1)$ with $w(-1) = w(1) = 0$, so u exhibits the typical behavior of having relatively small magnitude where β is large.

We report convergence and effectivity for $\beta = 10, 100, 1000$ on two different meshes—the first consisting of two rectangles obtained by dividing the domain along the line $x = 0$, and the second consisting of four triangles obtained by dividing the

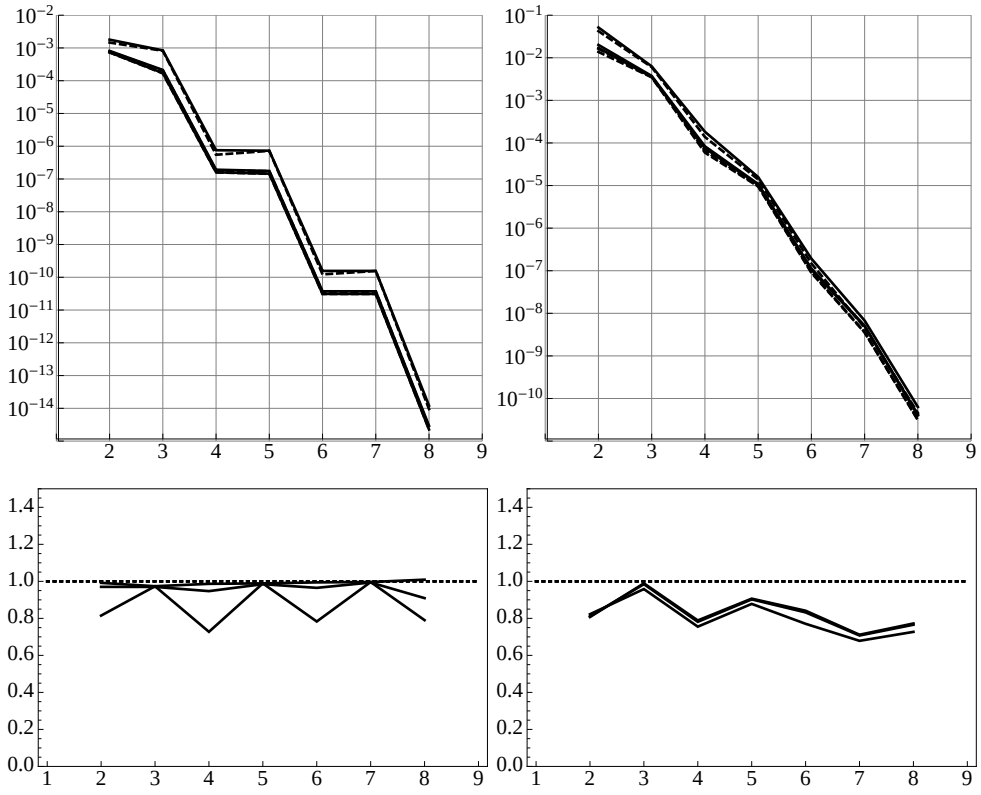

FIG. 2. *Convergence of the error (solid) and error estimates (dashed) with respect to p for both rectangular elements (left) and triangular elements(right) and* $\beta = 10, 100, 100$. *Global effectivities (solid) for both types of elements are given below their respective convergence plots.*

two rectangles along their diagonals. Convergence and effectivity plots, for the energy norm, are provided in Figure 2 for both types of elements. To save space, each of the four plots contain graphs for all three values of β. The convergence and effectivity behavior for $\beta = 100$ and $\beta = 1000$ is nearly identical, so their graphs are almost indistinguishable—the case $\beta = 10$ is more clearly distinguishable from the other for both types of elements. The effectivities stay within the range $[0.7, 1)$ in all cases.

4.2.2. Slit Disk. Let Ω be the unit disk with a slit along the positive x-axis, with Γ_1 consisting of the boundary of the disk ($r = 1$) and the top of the slit ($\theta = 0^+$, $0 \leq r \leq 1$), and Γ_2 consisting of the bottom of the slit ($\theta = 2\pi^-$, $0 < r < 1$); see Figure 3. We consider the problem

$$-\Delta u = f = (4 - \sigma^2)\sin(\sigma\theta) \text{ in } \Omega \quad , \quad u = 0 \text{ on } \Gamma_1 \quad , \quad \text{condition on } \Gamma_2 \ ,$$

for two choices of σ. If $u = 0$ on Γ_2, we take $\sigma = 1/2$ and refer to the problem as the *Dirichlet-Dirichlet* slit; and if $\partial u/\partial n = 0$ on Γ_2 we take $\sigma = 1/4$ call this the *Dirichlet-Neumann* slit. In both cases, the solution is given by

$$u = (r^\sigma - r^2)\sin(\sigma\theta) \ ,$$

and it exhibits the typical singularities present for generic f. In Figure 3 we see the mesh and a close-up of the central portion of the mesh. It is clear from these images

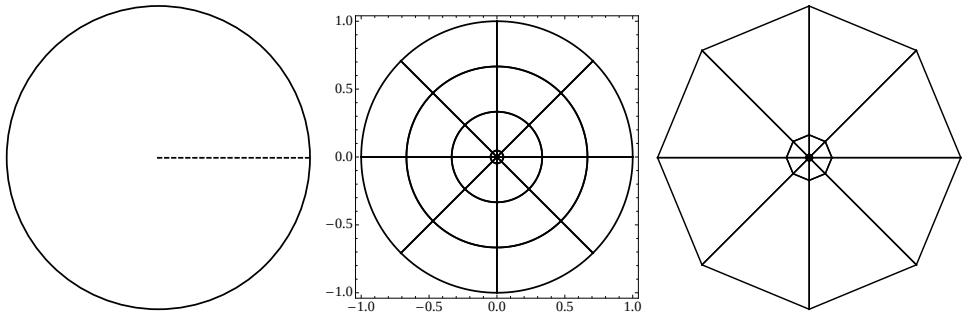

FIG. 3. *The slit disk, together with its mesh and a close-up of the central portion.*

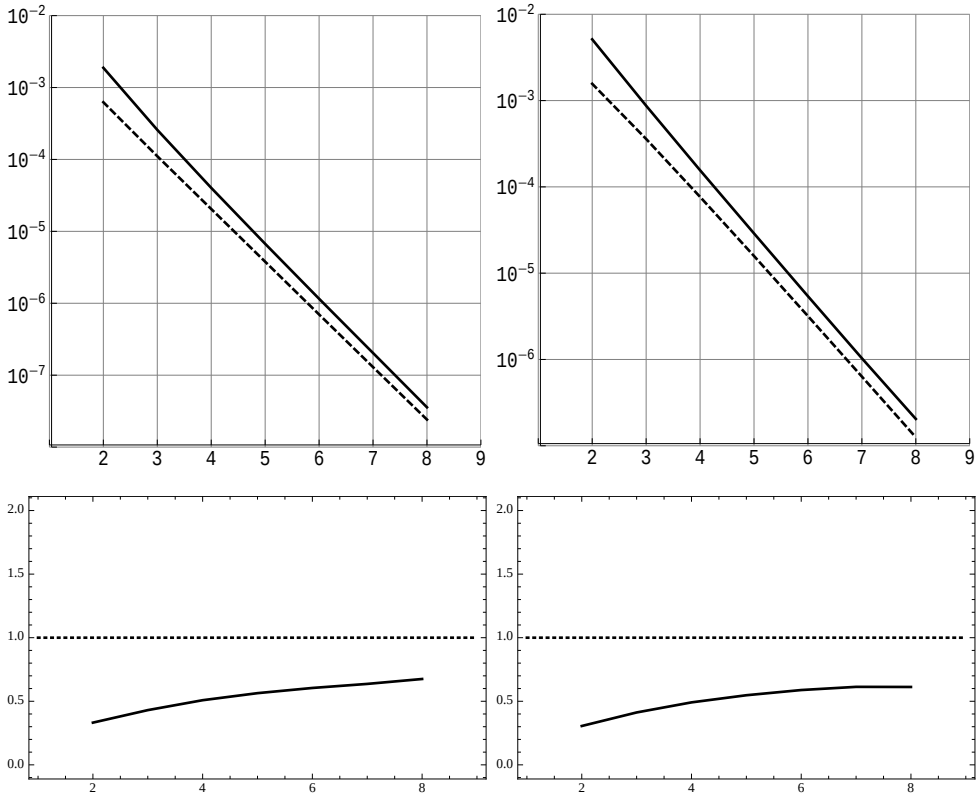

FIG. 4. *Convergence of the error (solid) and error estimates (dashed) with respect to p for both the Dirichlet-Dirichlet case (left) and the Dirichlet-Neumann case (right). Global effectivities (solid) for both problems are given below their respective convergence plots.*

that the mesh includes both curved and straight quadrilaterals, but it also includes triangles touching the origin. Despite the difference in singularity strength for the two types of boundary conditions, the same mesh is used in both cases. Convergence and effectivity plots are given in Figure 4 for both problems, with respect to polynomial degree p. We emphasize that the effectivities in both cases do not deteriorate with p, and indicate that the error estimate is generally within a factor of two of the actual error.

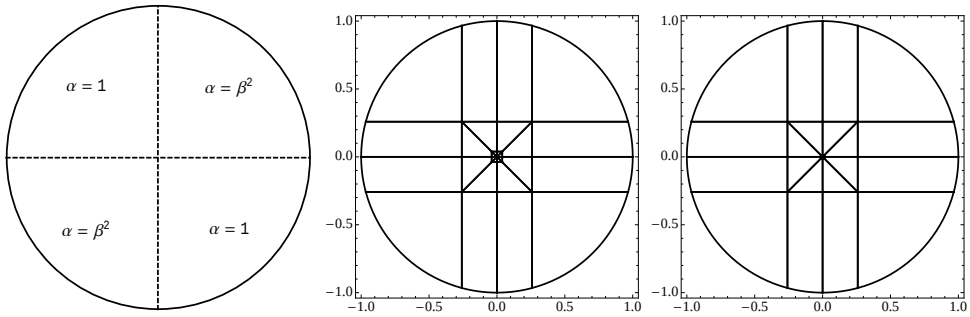

FIG. 5. *The Kellogg problem, together with its mesh for $\beta = 5$ (center) and $\beta = 10$ (right).*

4.2.3. Kellogg Problem. Let Ω be the unit disk and $\beta > 1$, and define $\sigma = \arctan(\beta^{-1})/(\pi/4)$. We consider the problem

$$-\nabla \cdot (\alpha \nabla u) = f = (4 - \sigma^2)\, \alpha\, g \text{ in } \Omega \quad , \quad u = 0 \text{ on } \partial\Omega \ ,$$

where

$$g(\theta) = \begin{cases} -\cos(\sigma(\pi/4 - \theta))/\beta & , \ \theta \in [0, \pi/2) \\ -\sin(\sigma(3\pi/4 - \theta)) & , \ \theta \in [\pi/2, \pi) \\ \cos(\sigma(5\pi/4 - \theta))/\beta & , \ \theta \in [\pi, 3\pi/2) \\ \sin(\sigma(7\pi/4 - \theta)) & , \ \theta \in [3\pi/2, 2\pi) \end{cases} \quad , \quad \alpha(\theta) = \begin{cases} \beta^2 & , \ \theta \in [0, \pi/2) \cup [\pi, 3\pi/2) \\ 1 & , \ \theta \in [\pi/2, \pi) \cup [3\pi/2, 2\pi) \end{cases} \ ,$$

and we require that both u and $\alpha\, \partial u/\partial n$ are continuous across the interfaces between the four quadrants (see Figure 5). We may naturally think of α and g as functions on $\mathbb{R}$ via 2π-periodic extension. The solution is given by

$$u = (r^\sigma - r^2) g(\theta) \ ,$$

and it exhibits the typical leading singularity present for generic f. By increasing β, we can make $\sigma > 0$ as small as we like, thereby generating an increasingly strong singularity at the origin. For our experiments we consider the cases $\beta = 5$ and $\beta = 10$, for which the solution has leading singularities $r^{0.251332}$ and $r^{0.126902}$, respectively. Again, the meshes have a mix of curved and straight triangles and quadrilaterals, as seen in Figure 5. Convergence and effectivity plots are given in Figure 6 for both $\beta = 5$ and $\beta = 10$, with respect to polynomial degree p. As before, we see that the effectivities in both cases do not deteriorate with p, and indicate that the error estimate is generally within a factor of two of the actual error.

4.2.4. Boundary Layers. Letting Ω be either the unit square or the unit cube, we consider the problem

$$-\epsilon \Delta u + \frac{\partial u}{\partial x} + 2u/\alpha = 1 \text{ in } \Omega \ ,$$

with homogenous Dirichlet conditions at $x = 0$ and $x = 1$, and homogeneous Neumann conditions on the rest of the boundary. The solution is given by

$$u = \frac{\alpha}{2} \left(1 + \left(\frac{e^{r^-} - 1}{e^{r^+} - e^{r^-}} \right) e^{r^+ x} - \left(\frac{e^{r^+} - 1}{e^{r^+} - e^{r^-}} \right) e^{r^- x} \right) \quad , \quad r^\pm = \frac{1 \pm \sqrt{1 + 8\epsilon/\alpha}}{2\epsilon} \ .$$

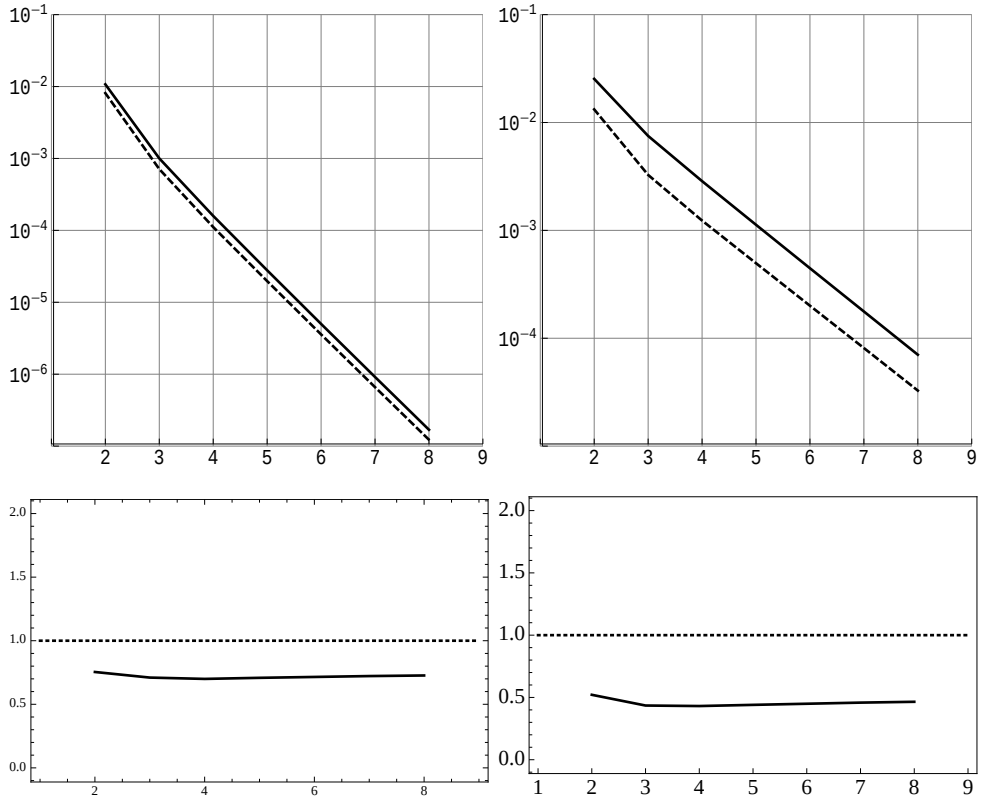

FIG. 6. *Convergence of the error (solid) and error estimates (dashed) with respect to p for the Kellogg problem with $\beta = 5$ (left) and $\beta = 10$ (right). Global effectivities (solid) in the energy norm for both problems are given below their respective convergence plots.*

Such solutions exhibit boundary layers near both $x = 0$ and $x = 1$ when $0 < \varepsilon \ll 1$ and $0 < \alpha \ll 1$. The quadrilateral meshes for $\epsilon = 10^{-1}, \alpha = 10^{-2}$ and $\epsilon = 10^{-2}, \alpha = 10^{-2}$ are given in Figure 7. The convergence and effectivity plots for these problems are

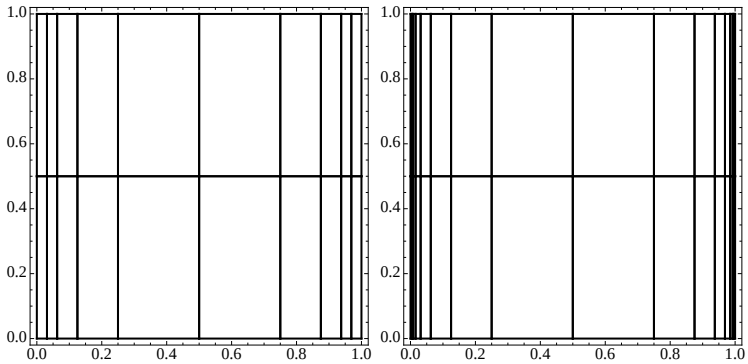

FIG. 7. *Rectangular meshes for the cases $\epsilon = 10^{-2}, \alpha = 10^{-3}$ and $\epsilon = 10^{-3}, \alpha = 10^{-3}$ of the Boundary Layer problem.*

given in Figure 8. These are given in the energy-norm,

$$\|v\|^2 = \int_\Omega \epsilon|\nabla v|^2 + 2u^2/\alpha \, dx \ ,$$

derived from the symmetric part of the associated bilinear form. Since the errors are

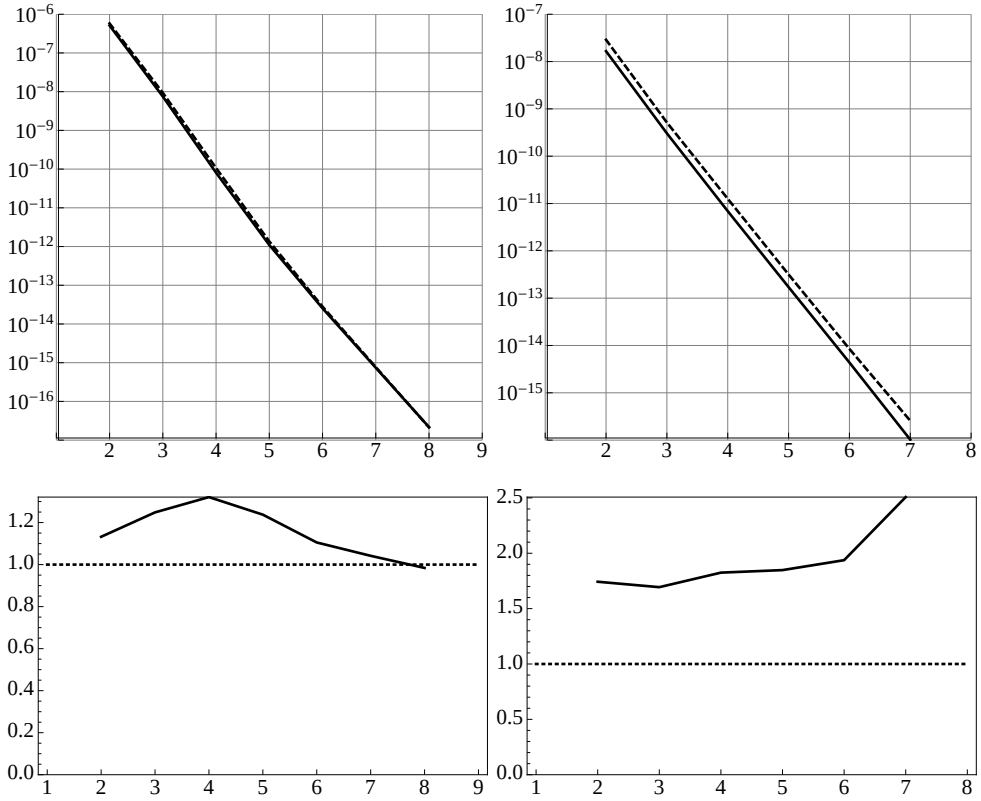

FIG. 8. *Convergence of the error (solid) and error estimates (dashed) with respect to p for the Boundary Layer problem with* $\epsilon = 10^{-1}, \alpha = 10^{-2}$ *(left) and* $\epsilon = 10^{-2}, \alpha = 10^{-2}$ *(right). Global effectivities (solid) for both problems are given below their respective convergence plots.*

near machine-precision for $p \geq 7$, it is expected that the reported effectivities may not be as accurate in that range. Finally, we consider the case $\epsilon = \alpha = 10^{-2}$ in 3D with hexahedral bricks with an appropriate x-grading, and whose yz-aspect ratio is 1 for each brick. The convergence and effectivity information are given in Figure 9.

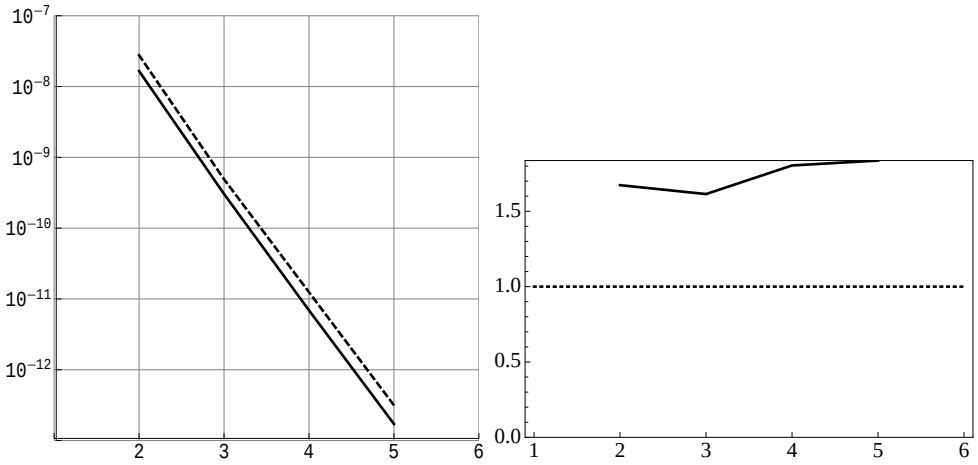

FIG. 9. *At left, convergence of the error (solid) and error estimates (dashed) with respect to p for the 3D Boundary Layer problem with* $\epsilon = \alpha = 10^{-2}$. *Global effectivities (solid) are given at right.*

REFERENCES

[1] S. Adjerid, M. Aiffa, and J. E. Flaherty. Hierarchical finite element bases for triangular and tetrahedral elements. *Comput. Methods Appl. Mech. Engrg.*, 190(22-23):2925 – 2941, 2001.

[2] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*. Pure and Applied Mathematics (New York). Wiley-Interscience [John Wiley & Sons], New York, 2000.

[3] R. Araya, G. R. Barrenechea, and A. Poza. An adaptive stabilized finite element method for the generalized stokes problem. *Journal of Computational and Applied Mathematics*, 214(2):457 – 479, 2008.

[4] D. N. Arnold. Spaces of finite element differential forms. In F. Brezzi, P. Colli-Franzone, U. P. Gianazza, and G. Gilardi, editors, *Analysis and Numerics of Partial Differential Equations*, volume 4 of *Springer INdAM Series*, pages 117–140. Springer, Milan, 2013.

[5] R. E. Bank. Hierarchical bases and the finite element method. In *Acta numerica, 1996*, volume 5 of *Acta Numer.*, pages 1–43. Cambridge Univ. Press, Cambridge, 1996.

[6] R. E. Bank, L. Grubišić, and J. S. Ovall. A framework for robust eigenvalue and eigenvector error estimation and ritz value convergence enhancement. *Applied Numerical Mathematics*, 66(0):1 – 29, 2013.

[7] R. E. Bank and R. K. Smith. A posteriori error estimates based on hierarchical bases. *SIAM J. Numer. Anal.*, 30(4):921–935, 1993.

[8] S. Beuchler and J. Schöberl. New shape functions for triangular p-FEM using integrated Jacobi polynomials. *Numer. Math.*, 103(3):339–366, 2006.

[9] F. A. Bornemann, B. Erdmann, and R. Kornhuber. A posteriori error estimates for elliptic problems in two and three space dimensions. *SIAM J. Numer. Anal.*, 33(3):1188–1204, 1996.

[10] D. Braess, V. Pillwein, and J. Schöberl. Equilibrated residual error estimates are p-robust. *Computer Methods in Applied Mechanics and Engineering*, 198(1314):1189 – 1197, 2009. {HOFEM07} International Workshop on High-Order Finite Element Methods, 2007.

[11] P. Carnevali, R. B. Morris, Y. Tsuji, and G. Taylor. New basis functions and computational procedures for p-version finite element analysis. *International Journal for Numerical Methods in Engineering*, 36(22):3759–3779, 1993.

[12] A. W. Craig, J. Z. Zhu, and O. C. Zienkiewicz. A posteriori error estimation, adaptive mesh refinement and multigrid methods using hierarchical finite element bases. In *The mathematics of finite elements and applications, V (Uxbridge, 1984)*, pages 587–594. Academic Press, London, 1985.

[13] P. Deuflhard, P. Leinen, and H. Yserentant. Concepts of an adaptive hierarchical finite element code. *IMPACT Comput. Sci. Eng.*, 1(1):3–35, 1989.

[14] W. Dörfler and R. H. Nochetto. Small data oscillation implies the saturation assumption. *Numer. Math.*, 91(1):1–12, 2002.

[15] V. Eijkhout and P. Vassilevski. The role of the strengthened Cauchy-Buniakowskiĭ-Schwarz inequality in multilevel methods. *SIAM Rev.*, 33(3):405–419, 1991.

[16] A. Ern and M. Vohralík. Polynomial-degree-robust a posteriori estimates in a unified setting for conforming, nonconforming, discontinuous Galerkin, and mixed discretizations. HAL (Inria) Preprint 00921583, Dec. 2013.

[17] M. Holst, J. S. Ovall, and R. Szypowski. An efficient, reliable and robust error estimator for elliptic problems in R^3. *Applied Numerical Mathematics*, 61(5):675 – 695, 2011.

[18] W. Huang, L. Kamenski, and J. Lang. A new anisotropic mesh adaptation method based upon hierarchical a posteriori error estimates. *J. Comput. Phys.*, 229(6):2179–2198, 2010.

[19] L. Kamenski. A study on using hierarchical basis error estimates in anisotropic mesh adaptation for the finite element method. *Engineering with Computers*, 28(4):451–460, 2012.

[20] C. Kreuzer and K. G. Siebert. Decay rates of adaptive finite elements with Dörfler marking. *Numer. Math.*, 117(2):679–716, 2011.

[21] H. Li and J. Ovall. A posteriori error estimation of hierarchical type for the Schrödinger operator with inverse square potential. *Numer. Math.*, pages 1–34, 2014.

[22] J. M. Melenk and B. I. Wohlmuth. On residual-based a posteriori error estimation in hp-FEM. *Adv. Comput. Math.*, 15(1-4):311–331 (2002), 2001. A posteriori error estimation and adaptive computational methods.

[23] J. S. Ovall. Function, gradient, and Hessian recovery using quadratic edge-bump functions. *SIAM J. Numer. Anal.*, 45(3):1064–1080 (electronic), 2007.

[24] L. R. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[25] B. Szabó and I. Babuška. *Finite element analysis*. A Wiley-Interscience Publication. John

Wiley & Sons Inc., New York, 1991.

[26] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques*. Wiley-Teubner [John Wiley & Sons and B.G. Teubner], New York/Stuttgart, 1996.

[27] O. C. Zienkiewicz, D. W. Kelly, J. Gago, and I. Babuška. Hierarchical finite element approaches, error estimates and adaptive refinement. In *The mathematics of finite elements and applications, IV (Uxbridge, 1981)*, pages 313–346. Academic Press, London, 1982.