Max-Planck-Institut
für Mathematik
in den Naturwissenschaften
Leipzig

An efficient, reliable and robust error estimator for elliptic problems in $\mathbb{R}^3$

by

Michael Holst, Jeffrey Ovall, and Ryan Szypowski

Preprint no.: 38 2010

AN EFFICIENT, RELIABLE AND ROBUST ERROR ESTIMATOR FOR ELLIPTIC PROBLEMS IN $\mathbb{R}^3$

MICHAEL HOLST, JEFFREY S. OVALL, RYAN SZYPOWSKI

Abstract. In this article, we develop and analyze error estimators for a general class of second-order linear elliptic boundary value problems in bounded three-dimensional domains. We first describe the target class of problems, and assemble some basic mathematical facts and tools. We then briefly examine discretizations based on tetrahedral partitions and conforming finite element subspaces, introduce notation, and subsequently define an error estimator based on the use of piecewise cubic face-bump functions that satisfy a residual equation. We show that this type of indicator automatically satisfies a global lower bound inequality thereby giving efficiency, without regularity assumptions beyond those giving well-posedness of the continuous and discrete problems. The main focus of the paper is then to establish the reverse inequality: a global upper bound on the error in terms of the error estimate (plus an oscillation term), again without regularity assumptions, thereby giving also reliability. To prove this result, we first derive some basic geometrical identities for conforming discretizations based on tetrahedral partitions, and then develop some interpolation results together with a collection of scale-invariant inequalities for the residual that are critical for establishing the global upper bound. After establishing the main result, we give an analysis of the computational costs of actually computing the error indicator. Through a sequence of spectral equivalence inequalities, we show that the cost to evaluate the indicator (involving the solution of a linear system) is linear in the number of degrees of freedom. We finish the article with a sequence of numerical experiments to illustrate the behavior predicted by the theoretical results, including: a Poisson problem on a 3D L-shaped domain, a jump coefficient problem in a cube, a convection-diffusion problem, and a strongly anisotropic diffusion problem.

Key words. finite elements, a posteriori estimates, reliability

AMS subject classifications. 65N15,65N30,65N50

1. Introduction.

Adaptive meshing (local refinement and coarsening) based on *a posteriori* error estimation is an essential component of efficient and robust finite element algorithms. To be actually useful in computation, the error estimates should be *reliable* (never under-estimating the error by too much) and *efficient* (never over-estimating the error by too much). Together, these properties imply the equivalence of the error estimate and the true error. Depending on the type of estimator under consideration, either reliability or efficiency will be more difficult to prove. In this article, we develop and analyze such error estimates for a general class of second-order linear elliptic boundary value problems in bounded three-dimensional domains. Our focus is on error estimates based on the use of piecewise cubic face-bump functions that satisfy a residual equation. We show that this type of error estimator automatically satisfies a global lower bound inequality thereby giving efficiency, without regularity assumptions beyond those giving well-posedness of the continuous and discrete problems. The main focus of the paper is then to establish the reverse inequality, namely a global upper bound on the error in terms of the error estimate (plus an oscillation term), again without regularity assumptions, thereby giving also reliability. We also give an analysis of the computational costs of actually computing the error indicator to ensure it is a practical technique, and we then present a sequence of numerical experiments to illustrate the behavior predicted by the theoretical results.

In [7], hierarchical estimators were examined for both two- and three-dimensional problems; the use cubic face-bubbles was considered, but the estimator was based on the solution of local problems rather than on a global residual. However, the estimator and the analysis in [7] are not easily extended to non-symmetric problems with convection terms. A more recent work [3] was apparently the first to consider

hierarchical estimators for problems with convection. The focus was only on two-dimensional problems, and the analysis was both significantly different from that appearing here, and also the extension to three dimensions appears problematic. The convection-dominated case was one of the primary focuses of [3]; the analysis employs a small constant diffusion parameter ϵ, similar to the convection-diffusion example appearing later in this article. Related relevant work includes [13] and several references contained therein, whereby data data oscillation is managed in the analysis, as opposed to residual oscillation used in the analysis in this article.

Outline of the article. The remainder of this paper is structured as follows. In Section 2, we describe the target class of problems, and assemble some basic mathematical facts and tools. We briefly examine discretizations based on tetrahedral partitions and conforming finite element subspaces, introduce notation, and subsequently define an error estimator based on the use of piecewise cubic face-bump functions that satisfy a residual equation. We show that this type of indicator automatically satisfies a global lower bound inequality thereby giving efficiency, without regularity assumptions beyond those giving well-posedness of the continuous and discrete problems. The main focus of the paper then shifts to establishing the reverse inequality: a global upper bound on the error in terms of the error estimate (plus an oscillation term), again without regularity assumptions, thereby giving also reliability. In Section 3, we begin the analysis by developing some basic geometrical identities for conforming discretizations based on tetrahedral partitions. Section 4 contains the main theoretical results of the paper. We begin by stating some quasi-interpolation results, the proofs of which we delay until Section 5. We then establish a collection of scale-invariant inequalities for the residual that are critical for establishing the global upper bound. The main result (the global upper bound in Theorem 4.6) is then established. In Section 5, we give the technical proofs of the quasi-interpolant results needed for the main result. In Section 6, we give an analysis of the computational costs of actually computing the error indicator. Through a sequence of spectral equivalence inequalities, we show that the cost to evaluate the indicator (involving the solution of a linear system) is linear in the number of degrees of freedom. In Section 7, we give a sequence of numerical experiments to illustrate the behavior predicted by the theoretical results, including: a Poisson problem on a 3D L-shaped domain, a jump coefficient problem in a cube, a convection-diffusion problem, and a strongly anisotropic diffusion problem. In Section 8, we reflect on the analysis and results, and consider possible extensions.

2. Model Problem and Basic Theory. Let $\Omega \subset \mathbb{R}^3$ be polyhedral and possibly non-convex, with boundary $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$ — a disjoint union with $\partial\Omega_D$ relatively closed. We take $\mathcal{H} = H^1_{0D}(\Omega) = \{u \in H^1(\Omega) : \ u_{|\partial\Omega_D} = 0\}$. Let $\omega \subset \Omega$. We take the standard Sobolev norms and semi-norms:

$$\|v\|^2_{k,\omega} = \sum_{|\alpha|\leq k} \|D^\alpha v\|^2_{L^2(\omega)} \qquad\qquad |v|^2_{k,\omega} = \sum_{|\alpha|=k} \|D^\alpha v\|^2_{L^2(\omega)} \tag{2.1}$$

When $\omega = \Omega$ we omit it from the subscript.

We are interested in problems of the form:

$$\text{Find } u \in \mathcal{H} \text{ such that } B(u,v) = G(v) \text{ for all } v \in \mathcal{H} \ , \tag{2.2}$$

where

$$B(u,v) = \int_\Omega A\nabla u \cdot \nabla v + (\mathbf{b}\cdot\nabla u + cu)v\,dx \quad , \quad G(v) = \int_\Omega fv\,dx + \int_{\partial\Omega_N} gv\,ds \ .$$

We will assume that $A \in [L^\infty(\Omega)]^{3\times 3}$, $\mathbf{b} \in [L^\infty(\Omega)]^3$ and $c \in L^\infty(\Omega)$ are piecewise-smooth on some polyhedral partition of Ω, and $f \in L^2(\Omega)$ and $g \in L^2(\partial\Omega_N)$. Furthermore, we assume that A is symmetric and positive-definite almost everywhere and that there exists a constants $m, M > 0$ such that

$$m\|u\|_1^2 \le B(u,u) \quad , \quad |B(u,v)| \le M\|u\|_1\|v\|_1 \ . \tag{2.3}$$

The coercivity condition (lower bound) is sufficient for (2.2) to be well-posed.

Given a conforming tetrahedral partition $\mathcal{T}$ of Ω, having: tetrahedra $T \in \mathcal{T}$, faces $F \in \bar{\mathcal{F}}$, edges $E \in \bar{\mathcal{E}}$ and vertices $z \in \bar{\mathcal{V}}$, we consider the finite element spaces $V \subset \mathcal{H}$ and $W \subset \mathcal{H}$ given by

$$V = \{v \in \mathcal{H} \cap C(\Omega) : \ v|_T \in \mathbb{P}_1 \text{ for every } T \in \mathcal{T}\} \tag{2.4}$$

$$W = \{v \in \mathcal{H} \cap C(\Omega) : \ v|_T \in \mathbb{P}_3 \text{ for every } T \in \mathcal{T} \text{ and } v_{|_E} = 0 \text{ for every } E \in \bar{\mathcal{E}}\} \tag{2.5}$$

Here $\mathbb{P}_k$ denotes the collection of polynomials of (total) degree no greater than k. We will implicitly assume that $\mathcal{T}$ lines up perfectly with any discontinuities in the data.

We will approximate the solution u of (2.2) by a piecewise linear function $\hat{u} \in V$ satisfying

$$\text{Find } \hat{u} \in V \text{ such that } B(\hat{u},v) = G(v) \text{ for all } v \in V \ , \tag{2.6}$$

and we will approximate the error $u - \hat{u}$ by a piecewise cubic "face-bump" function ε satisfying

$$\text{Find } \varepsilon \in W \text{ such that } B(\varepsilon,v) = G(v) - B(\hat{u},v) \text{ for all } v \in W \ . \tag{2.7}$$

The coercivity of B on $\mathcal{H}$ implies that problems (2.6) and (2.7) are also well-posed. From (2.7) we automatically get an efficiency estimate on the error and error estimate:

$$m\|\varepsilon\|_1^2 \le B(\varepsilon,\varepsilon) = B(u-\hat{u},\varepsilon) \le M\|u-\hat{u}\|_1\|\varepsilon\|_1 \tag{2.8}$$

$$\frac{m}{M}\|\varepsilon\|_1 \le \|u-\hat{u}\|_1 \ . \tag{2.9}$$

In other words, $\|\varepsilon\|_1$ does not over-estimate the actual error $\|u-\hat{u}\|_1$ by "too much"—in fact, it is nearly always the case that we obtain a slight under-estimate of the true error.

A key objective of this paper is to provide an efficiency (upper) estimate to complement (2.9), in which all quantities are explicitly computable (or estimable), making no further regularity assumptions on u. Our estimate is of the form:

$$\|u-\hat{u}\|_1 \le K_1\|\varepsilon\|_1 + K_2\,\mathrm{osc}(R,r,\mathcal{T}) \ , \tag{2.10}$$

where the oscillation term involves deviations of the element- and face-residuals, R and r, from constants on patches in the triangulation, and the constants are independent of the f, g, u and the sizes of the elements—though they do depend on their shapes. We can see very clearly in (2.10) that the only thing that could ever destroy the

effectivity of $\|\varepsilon\|_1$ as an estimator of $\|u - \hat{u}\|_1$ is if the data oscillation is relatively large; but that is something that we can assess and control directly if we choose, because it involves no unknown quantities.

Hierarchical error estimators, such as the one we propose, are traditionally analyzed only in the the setting in which the bilinear form B is an inner-product, with induced "energy norm" $|\!|\!| \cdot |\!|\!|$ (cf. [5, 1]). Such an analysis makes use of a *strong Cauchy inequality* between the spaces V and W,

$$|B(v,w)| \leq \gamma |\!|\!| v |\!|\!| |\!|\!| w |\!|\!| \text{ for } v \in V \text{ and } w \in W, \text{ where } \gamma = \gamma(B, \mathcal{T}) < 1 \ ,$$

as well as a *saturation assumption*,

$$\inf_{v \in V \oplus W} |\!|\!| u - v |\!|\!| \leq \beta \inf_{v \in V} |\!|\!| u - v |\!|\!|, \text{ where } \beta = \beta(B, G, \mathcal{T}) < 1 \ .$$

From this the following equivalence result between the error and error estimate is obtained

$$|\!|\!| \varepsilon |\!|\!| \leq |\!|\!| u - \hat{u} |\!|\!| \leq \frac{|\!|\!| \varepsilon |\!|\!|}{\sqrt{(1-\gamma^2)(1-\beta^2)}} \ .$$

A criticism which is sometimes made about such an analysis is that counterexamples to the saturation assumption are readily constructed for a given mesh and finite dimensional spaces V and W (cf. [10])—although it could be argued that such cases rarely occur in practical computations. Regardless of how one views this point, such analysis cannot be applied for the more general bilinear forms such as those considered here.

REMARK 2.1. *The traditional analysis referred to above, particularly the saturation assumption, generally leads one to choose an auxiliary space W for which $V \oplus W$ is a standard approximation space—for example, the piecewise polynomial space of next higher degree on the same mesh, or the piecewise polynomial space of the same degree but on a uniformly refined mesh. This way of thinking might naturally lead one to choose W to be the space of quadratic "edge-bumps" which vanish at every vertex in the mesh, in $\mathbb{R}^3$, So that $V \oplus W$ is the space of piecewise quadratic functions on the mesh. However, our analysis proceeds along different lines, which leads us to choose W as we have.*

3. Mesh-Related Notation and Basic Results. In this section, we introduce the mesh-related and other notation which will play a role in the analysis given later. In particular, we lay out the definitions which are needed to define and analyze the quasi-interpolant whose key properties are outlined in Theorem 4.2. Additionally, we collect several basic geometric identities in Lemma 3.2 which will be used in various arguments throughout the manuscript.

Give a mesh $\mathcal{T}$, we distinguish:

Vertices: Dirichlet $\mathcal{V}_D$ ($z \in \partial\Omega_D$), non-Dirichlet $\mathcal{V} = \bar{\mathcal{V}} \setminus \mathcal{V}_D$

Faces: Dirichlet $\mathcal{F}_D$ ($F \subset \partial\Omega_D$), non-Dirichlet $\mathcal{F} = \bar{\mathcal{F}} \setminus \mathcal{F}_D$, Neumann $\mathcal{F}_N$ ($F \subset \partial\Omega_N$), interior $\mathcal{F}_I = \mathcal{F} \setminus \mathcal{F}_N$

Non-Dirichlet faces touching $z \in \bar{\mathcal{V}}$: $\mathcal{F}_z$

Tetrahedra touching $z \in \bar{\mathcal{V}}$ or $F \in \bar{\mathcal{F}}$: $\mathcal{T}_z$, $\mathcal{T}_F$

For $z \in \bar{\mathcal{V}}$, we define the continuous, piecewise linear function ℓ_z by

$$\ell_z(z') = \delta_{zz'} \text{ for all } z \in \mathcal{V}. \tag{3.1}$$

Also, for any $F \in \bar{\mathcal{F}}$, we define $b_F = \ell_z \ell_{z'} \ell_{\hat{z}}$, where z, z' and $\hat{z}$ are the three vertices of F. We these definitions, it is clear that

$$V = \text{span}\{\ell_z : \ z \in \mathcal{V}\} \quad , \quad W = \text{span}\{b_F : \ F \in \mathcal{F}\} \ . \tag{3.2}$$

The following definitions of patches of elements will also be useful in later analysis.

$$\omega_z := \text{supp}(\ell_z) = \bigcup_{T \in \mathcal{T}_z} T \quad , \quad \omega_F := \text{supp}(b_F) = \bigcup_{T \in \mathcal{T}_F} T \ . \tag{3.3}$$

REMARK 3.1. *We implicitly assume that any vertex in $\mathcal{V}_D$ will have at least one adjacent vertex in $\mathcal{V}$ — so $\mathcal{V}_z, \mathcal{F}_z \neq \emptyset$. Such an assumption is very natural, and simple to enforce in practice.*

Let T be a tetrahedron having: vertices $\{z_k\}_1^4$, opposite faces $\{F_k\}_1^4$, and outward unit normals to these faces $\{\mathbf{n}_k\}_1^4$. We denote by θ_{ij} the measure of the dihedral angle at the edge shared by faces F_i and F_j. We also use $\ell_k = \ell_{z_k}$, $b_k = b_{F_k}$ and $d_k = 3|T|/|F_k| = \text{dist}(z_k, F_k)$. In the following lemma, we collect various technical facts which will be used, generally without reference, in the rest of the paper.

LEMMA 3.2. *Let $i, j, k, l \in \{1, 2, 3, 4\}$ be distinct. We have*

$$\ell_i = 1 - \frac{\mathbf{n}_i}{d_i} \cdot (x - z_i) = -\frac{\mathbf{n}_i}{d_i} \cdot (x - z_j) = \nabla \ell_i \cdot (x - z_j) \ \text{ on } T \ , \tag{3.4}$$

$$\int_T \ell_1^p \ell_2^q \ell_3^r \ell_4^s = \frac{|T| 3! p! q! r! s!}{(p + q + r + s + 3)!} \ \text{ for } \ p, q, r, s \in \mathbb{Z}_{\geq 0} \ , \tag{3.5}$$

$$\int_{F_i} \ell_j^p \ell_k^q \ell_l^r = \frac{|F_i| 2! p! q! r!}{(p + q + r + 2)!} \ \text{ for } \ p, q, r \in \mathbb{Z}_{\geq 0} \ , \tag{3.6}$$

$$\int_T \nabla b_i \cdot \nabla b_j = 2 \frac{3! |T|}{7!} \left(\frac{\cos \theta_{kl}}{d_k d_l} - \frac{2 \cos \theta_{ij}}{d_i d_j} \right) \ , \tag{3.7}$$

$$\int_T \nabla b_i \cdot \nabla b_i = 2 \frac{3! |T|}{7!} \sum_{k=1}^4 \frac{1}{d_k^2} \ . \tag{3.8}$$

Finally, for $v \in H^1(T)$,

$$d_k \int_{F_k} v = \int_T 3v + (x - z_k) \cdot \nabla v \ . \tag{3.9}$$

Throughout this manuscript, the notation $|X|$ will be used to denote the length, area, volume, cardinality or Euclidean norm of X, and the appropriate interpretation will be clear from the context.

Proof. Equations (3.4) are well-known, and follow from the facts that such functions are clearly linear and have the correct values at the vertices. Equations (3.5)-(3.6) are also well-known—see [20, pg. 95], or [15, Theorem A.1], for example. The identities (3.7) and (3.8) follow from the definitions of b_i, b_j and the previous results, with simplifications carried out using forms of "generalized" Laws of Cosines for tetrahedra [2]. These are obtained from the fact that $\sum_{m=1}^4 \frac{\mathbf{n}_m}{d_m} = -\sum_{m=1}^4 \nabla \ell_m = \mathbf{0}$, by equating either one term with the other three, or two terms with the other two, and "squaring" both sides. The final result is just a direct application of the Divergence Theorem ($\nabla \cdot (x - z_k) = 3$). ∎

4. The Main Results. Before we state our main results, we take the standard definitions for element and face residuals, R and r, namely

$$R_{|_T} = f - c\hat{u} - \mathbf{b}\cdot\nabla\hat{u} + \nabla\cdot A\nabla\hat{u} \tag{4.1}$$

$$r_{|_F} = \begin{cases} -(A\nabla\hat{u})\cdot\mathbf{n}_T - (A\nabla\hat{u})\cdot\mathbf{n}_{T'} & , \quad F\in\mathcal{F}_I \\ g - (A\nabla\hat{u})\cdot\mathbf{n} & , \quad F\in\mathcal{F}_N \end{cases} . \tag{4.2}$$

Here, T and T' are the two tetrahedra adjacent to $F\in\mathcal{F}_I$, and $\mathbf{n}_T$ and $\mathbf{n}_{T'}$ are their outward unit normals. In order to prove our reliability estimate (2.10) we first establish the following error equation.

LEMMA 4.1. *Suppose that $v_z\in V$ and $w_z\in W$ for all $z\in\bar{\mathcal{V}}$, and let $\hat{v}=\sum_{z\in\bar{\mathcal{V}}} v_z$ and $w=\sum_{z\in\bar{\mathcal{V}}} w_z$. Then for any $v\in\mathcal{H}$,*

$$B(u-\hat{u},v) = B(\varepsilon,w) + \sum_{z\in\bar{\mathcal{V}}}\int_{\omega_z} R(v\ell_z - v_z - w_z)\,dV + \sum_{F\in\mathcal{F}}\int_F r(v-\hat{v}-w)\,dS\ .$$

Proof. We have $\sum_{z\in\bar{\mathcal{V}}} v\ell_z \equiv v$, because the ℓ_z form a partition of unity for Ω. Because of Galerkin orthogonality, $B(u-\hat{u},v_z)=0$, and because of the definition of ε, $B(u-\hat{u},w_z)=B(\varepsilon,w_z)$. Using these facts, we deduce

$$\begin{aligned} B(u-\hat{u},v) &= \sum_{z\in\bar{\mathcal{V}}} B(u-\hat{u},v\ell_z) = \sum_{z\in\bar{\mathcal{V}}} B(u-\hat{u},v\ell_z - v_z - w_z) + \sum_{z\in\bar{\mathcal{V}}} B(\varepsilon,w_z) \\ &= B(\varepsilon,w) + \sum_{z\in\bar{\mathcal{V}}}\int_{\omega_z} R(v\ell_z - v_z - w_z)\,dV + \sum_{F\in\mathcal{F}}\int_F r(v-\hat{v}-w)\,dS\ , \end{aligned}$$

which completes the proof. □

Prudent choices for w and $\hat{v}$ will allow us to bound $B(\varepsilon,w)$ by $K(B,\mathcal{T})\|\varepsilon\|_1|v|_1$, and yield the oscillation term, osc. These choices are reflected in the following theorem, whose technical proof is postponed until Section 5.

THEOREM 4.2. *Let $v\in\mathcal{H}$. There is a quasi-interpolant $\mathcal{I}v=\hat{v}+w\in V\oplus W$, with*

- *$\hat{v}=\sum_{z\in\bar{\mathcal{V}}} v_z$ and $w=\sum_{z\in\bar{\mathcal{V}}} w_z$,*
- *$\mathrm{supp}(v_z),\mathrm{supp}(w_z)\subset\omega_z$ for $z\in\mathcal{V}$,*
- *$\mathrm{supp}(v_z),\mathrm{supp}(w_z)\subset\omega_{z'}$ for $z\in\mathcal{V}_D$, where $z'\in\mathcal{V}$ is adjacent to z,*

which satisfies the zero-mean properties:

$$\int_{\omega_z}(v\ell_z - v_z - w_z) = 0 \quad \textit{and} \quad \int_F (v-\hat{v}-w) = 0 \textit{ for each } F\in\mathcal{F}\ .$$

Furthermore, there are scale-invariant constants C_1, c_{1z}, c_{2z}, c_{3z} and c_{4F} such that

1. *$|w_z|_1 \le c_{1z}|v|_{1,\omega_z}$, $\|w\|_1 \le C_1|v|_1$*
2. *$|v\ell_z - v_z - w_z|_1 \le c_{2z}|v|_{1,\omega_z}$*
3. *$\|v\ell_z - v_z - w_z\|_0 \le c_{3z}D_z|v|_{1,\omega_z}$*
4. *$\|v-\hat{v}-w\|_{0,F} \le c_{4F}|F|^{1/4}|v|_{1,\Omega_F}$ for each $F\in\tilde{\mathcal{F}}$, where $\Omega_F=\omega_z\cup\omega_{z'}\cup\omega_{z''}$, and z,z',z'' are the vertices of F.*

Here, and following, $D_z=\mathrm{diam}(\omega_z)$.

REMARK 4.3. *The use of quasi-interpolants of various sorts (e.g.. [8, 19, 21]) in both a priori and a posteriori error estimates has a long history, though generally not in the context of hierarchical error estimates. Both the choice and the role in analysis*

of the quasi-interpolant used here share some similarities with that in [14], and it is no coincidence that similar notions of data or residual oscillation appear here, as they do in [14] and related works.

LEMMA 4.4. *Let $v \in \mathcal{H}$. There are scale-invariant constants $K_1 = K_1(\mathcal{T}, B)$ and $K_2 = K_2(\mathcal{T})$ such that*

$$|B(u - \hat{u}, v)| \leq K_1 \|\varepsilon\|_1 |v|_1 + K_2 \operatorname{osc}(R, r, \mathcal{T}) |v|_1 \ ,$$

where

$$[\operatorname{osc}(R, r, \mathcal{T})]^2 = \sum_{z \in \bar{\mathcal{V}}} D_z^2 \inf_{R_z \in \mathbb{R}} \|R - R_z\|_{0,\omega_z}^2 + \sum_{F \in \mathcal{F}} |F|^{1/2} \inf_{r_F \in \mathbb{R}} \|r - r_F\|_{0,F}^2 \ .$$

Proof. Using Lemma 4.1 and Theorem 4.2, we see that

$$\begin{aligned}
|B(u - \hat{u}, v)| &\leq M \|\varepsilon\|_1 \|w\|_1 + \sum_{z \in \bar{\mathcal{V}}} \|v\ell_z - v_z - w_z\|_{0,\omega_z} \inf_{R_z \in \mathbb{R}} \|R - R_z\|_{0,\omega_z} \\
&\qquad + \sum_{F \in \mathcal{F}} \|v - \hat{v} - w\|_{0,F} \inf_{r_F \in \mathbb{R}} \|r - r_F\|_{0,F} \\
&\leq M C_1 \|\varepsilon\|_1 |v|_1 + \sum_{z \in \bar{\mathcal{V}}} c_{3z} D_z |v|_{1,\omega_z} \inf_{R_z \in \mathbb{R}} \|R - R_z\|_{0,\omega_z} \\
&\qquad + \sum_{F \in \mathcal{F}} c_{4F} |F|^{1/4} |v|_{1,\Omega_F} \inf_{r_F \in \mathbb{R}} \|r - r_F\|_{0,F} \\
&\leq K_1 \|\varepsilon\|_1 |v|_1 + K_2 \operatorname{osc}(R, r, \mathcal{T}) |v|_1 .
\end{aligned}$$

The last inequality results from discrete Cauchy-Schwarz inequalities, and the (small) finite overlap of the patches ω_z and Ω_F. ∎

REMARK 4.5. *It is not uncommon in applications that the diffusion coefficient $A \in \mathbb{R}^{3\times 3}$ and Neumann data g are piecewise-constant. If we assume this, then the oscillation term simplifies considerably,*

$$[\operatorname{osc}(R, r, \mathcal{T})]^2 = \sum_{z \in \bar{\mathcal{V}}} D_z^2 \inf_{R_z \in \mathbb{R}} \|(f - c\hat{u} - \mathbf{b} \cdot \nabla \hat{u}) - R_z\|_{0,\omega_z}^2 \ .$$

In this case, we see that all of the information about the face-residual is "encoded" in ε. If we further have that $\mathbf{b} = \mathbf{0}$, and $c, f \in H^1(\Omega)$, then $[\operatorname{osc}(R, r, \mathcal{T})]^2 \lesssim \sum_{z \in \bar{\mathcal{V}}} D_z^4$. If c and/or f are merely piecewise smooth, then we will not get these kinds of gains from $\inf_{R_z \in \mathbb{R}} \|(f - c\hat{u}) - R_z\|_{0,\omega_z}^2$ when z is on an interface across which f and/or c is discontinuous. In such cases, one could safeguard the error estimation procedure by including the oscillation terms only in such patches ω_z, although we generally expect that this will be unnecessary.

Based on Lemma 4.4, the following result is immediate.

THEOREM 4.6. *There are scale-invariant constants $K_1 = K_1(B, \mathcal{T})$ and $K_2 = K_2(B, \mathcal{T})$ such that,*

$$\|u - \hat{u}\|_1 \leq K_1 \|\varepsilon\|_1 + K_2 \operatorname{osc}(R, r, \mathcal{T}) \ .$$

REMARK 4.7. *The constants appearing in Theorem 4.6 are obtained from those in Lemma 4.4 through division by the coercivity constant m. In fact, it is clear from*

the argument that coercivity is just a convenient sufficient condition for Theorem 4.6, and that an inf-sup condition (which would also have to be assumed for the discrete problems) is what is necessary for this theorem to hold. In this case, the constants in the bound would have to be modified accordingly.

In practice, the oscillation term is generally ignored both for global error estimates and for local error indicators—this is precisely what we do in the numerical experiments of Section 7. Of course, it is quite possible to construct examples for which the oscillation dominates, or is at least comparable to, the error estimated by ε, but this is something that can often be reliably anticipated from the data itself. At any rate, global estimates and local indicators can be safeguarded by adding on global and local oscillation terms—if only in those areas in which the oscillation is expected to be of similar order (in D_z) to the local interpolation error (cf. Remark 4.5).

5. Proof of Theorem 4.2. In this technical section we prove the existence and key properties of a quasi-interpolant claimed in Theorem 4.2. The following definition will be convenient for some of the Lemmas and Theorems below. For $F \in \mathcal{F}$, ω_F consists of one (in the case $F \in \mathcal{F}_N$) or two (in the case $F \in \mathcal{F}_I$) tetrahedra adjacent to F. In the first case, calling the adjacent tetrahedron T, we define $\mathbf{d}_F = x - z_{FT}$ on ω_F, where z_{FT} is the vertex of T opposite F. In the second case, calling the adjacent tetrahedra T and $\hat{T}$, we define

$$\mathbf{d}_F = \begin{cases} x - z_{FT} & , \quad x \in T \\ x - z_{F\hat{T}} & , \quad x \in \hat{T} \end{cases} . \tag{5.1}$$

LEMMA 5.1. *For $z \in \mathcal{V}$ there are unique $v_z \in V$ and $w_z \in W$, which are supported in ω_z, such that*

$$\int_F v\ell_z - v_z - w_z = 0 \ \text{ for every } F \in \mathcal{F}_z \ \text{ and } \ \int_{\omega_z} v\ell_z - v_z - w_z = 0 \ .$$

For $z \in \mathcal{V}_D$ there are $v_z \in V$ and $w_z \in W$, which are supported in $\omega_{z'}$ for some $z' \in \mathcal{V}$ which is adjacent to z, such that

$$\int_F v\ell_z - v_z - w_z = 0 \ \text{ for every } F \in \mathcal{F}_{z'} \ \text{ and } \ \int_{\omega_z} v\ell_z - v_z - w_z = 0 \ .$$

Proof. For $z \in \mathcal{V}$, we have $v_z = \alpha_z \ell_z$ and $w_z = \sum_{F \in \mathcal{F}_z} \beta_{zF} b_F$, so the vanishing patch-mean and face-mean conditions are equivalent to

$$\frac{|\omega_z|}{4}\,\alpha_z + \sum_{F \in \mathcal{F}_z} \frac{|\omega_F|}{120}\,\beta_{zF} = \int_{\omega_z} v\ell_z \quad , \quad \frac{|F|}{3}\,\alpha_z + \frac{|F|}{60}\,\beta_{zF} = \int_F v\ell_z \,,\ F \in \mathcal{F}_z \ .$$

The latter of these can be converted to

$$|\omega_F|\,\alpha_z + \frac{|\omega_F|}{20}\,\beta_{zF} = \int_{\omega_F} 4v\ell_z + \ell_z \mathbf{d}_F \cdot \nabla v \,,\ F \in \mathcal{F}_z \tag{5.2}$$

via Lemma 3.2. Solving the system yields,

$$\begin{aligned}\alpha_z &= \frac{4}{|\omega_z|}\int_{\omega_z} v\ell_z + \frac{2}{3|\omega_z|}\sum_{F\in\mathcal{F}_z}\int_{\omega_F}\ell_z\mathbf{d}_F\cdot\nabla v \\ &= \kappa_z + \frac{2}{3|\omega_z|}\sum_{F\in\mathcal{F}_z}\int_{\omega_F}\ell_z\mathbf{d}_F\cdot\nabla v \qquad (5.3)\\ \beta_{zF} &= \frac{20}{|\omega_F|}\int_{\omega_F} 4(v-\alpha_z)\ell_z + \ell_z\mathbf{d}_F\cdot\nabla v \\ &= \frac{80}{|\omega_F|}\int_{\omega_F}(v-\kappa_z)\ell_z + \frac{20}{|\omega_F|}\int_{\omega_F}\ell_z\mathbf{d}_F\cdot\nabla v - \frac{40}{3|\omega_z|}\sum_{\hat{F}\in\mathcal{F}_z}\int_{\omega_{\hat{F}}}\ell_z\mathbf{d}_{\hat{F}}\cdot\nabla v \qquad (5.4)\end{aligned}$$

For $z\in\mathcal{V}_D$ we select one $z'\in\mathcal{V}$ which is adjacent to z. We take $v_z=\alpha_z\ell_{z'}$ and $w_z=\sum_{F\in\mathcal{F}_{z'}}\beta_{zF}\,b_F$. Inserting these into the vanishing mean conditions, we obtain

$$\frac{|\omega_{z'}|}{4}\,\alpha_z + \sum_{F\in\mathcal{F}_{z'}}\frac{|\omega_F|}{120}\,\beta_{zF} = \int_{\omega_z} v\ell_z \quad , \quad \frac{|F|}{3}\,\alpha_z + \frac{|F|}{60}\,\beta_{zF} = \int_F v\ell_z\,,\ F\in\mathcal{F}_{z'}\ .$$

As with (5.2), the latter of these can be converted to

$$|\omega_F|\,\alpha_z + \frac{|\omega_F|}{20}\,\beta_{zF} = \int_{\omega_F} 3v\ell_z + \mathbf{d}_F\cdot\nabla(v\ell_z)\,,\ F\in\mathcal{F}_{z'}\ , \qquad (5.5)$$

Solving the system yields

$$\alpha_z = \frac{1}{|\omega_{z'}|}\left(6\int_{\omega_{z'}} v\ell_z - 4\int_{\omega_z} v\ell_z + \sum_{F\in\mathcal{F}_{z'}}\int_{\omega_F}\mathbf{d}_F\cdot\nabla(v\ell_z)\right) \qquad (5.6)$$

$$\beta_{zF} = -20\alpha_z + \frac{20}{|\omega_F|}\int_{\omega_F} 3v\ell_z + \mathbf{d}_F\cdot\nabla(v\ell_z) \qquad (5.7)$$

which completes the proof. □

REMARK 5.2. *Because $\int_F v\ell_z - v_z - w_z = 0$ for each all z, summing over z gives $\int_F v - v - w = 0$.*

LEMMA 5.3. *There exist scale-invariant constants c_{1z} such that $|w_z|_1 \le c_{1z}|v|_{1,\omega_z}$. Furthermore, there is a scale-invariant constant C_1 such that $\|w\|_1 \le C_1|v|_1$.*

Proof. We first consider $z\in\mathcal{V}$, and note that $|w_z|_{1,\omega_z}$ can be bounded by $(\lambda_{\max}\sum_{F\in\mathcal{F}_z}\beta_{zF}^2)^{1/2}$, where $\lambda_{\max}$ is the largest eigenvalue of the $|\mathcal{F}_z|\times|\mathcal{F}_z|$ matrix whose $F\hat{F}$ entry is $\int_{\omega_z}\nabla b_F\cdot\nabla b_{\hat{F}}$. Let $D_z=\mathrm{diam}(\omega_z)$. Recognizing that $\lambda_{\max}\sim D_z$ will allow us to get a bound on $|w_z|_{1,\omega_z}$ via bounds on the β_{zF}. Recall from the proof of Lemma 5.1 that

$$\beta_{zF} = \frac{80}{|\omega_F|}\int_{\omega_F}(v-\kappa_z)\ell_z + \frac{20}{|\omega_F|}\int_{\omega_F}\ell_z\mathbf{d}_F\cdot\nabla v - \frac{30}{|\omega_z|}\sum_{\hat{F}\in\mathcal{F}_z}\int_{\omega_{\hat{F}}}\ell_z\mathbf{d}_{\hat{F}}\cdot\nabla v\ .$$

We have the bound

$$\begin{aligned}\beta_{zF} \le{}& \frac{80\|\ell_z^{1/2}\|_{0,\omega_F}}{|\omega_F|}\,\|(v-\kappa_z)\ell_z^{1/2}\|_{0,\omega_F} + \left|\frac{20}{|\omega_F|}-\frac{30}{|\omega_z|}\right|\|\ell_z\mathbf{d}_F\|_{0,\omega_F}|v|_{1,\omega_F} \\ &+ \frac{30}{|\omega_z|}\sum_{\hat{F}\in[\mathcal{F}_z\setminus\{F\}]}\|\ell_z\mathbf{d}_{\hat{F}}\|_{0,\omega_{\hat{F}}}|v|_{1,\omega_{\hat{F}}}\ .\end{aligned}$$

A Poincaré inequality,

$$\|(v-\kappa_z)\ell_z^{1/2}\|_{0,\omega_F} \le \|(v-\kappa_z)\ell_z^{1/2}\|_{0,\omega_z} = \inf_{a\in\mathbb{R}} \|(v-a)\ell_z^{1/2}\|_{0,\omega_z}$$
$$\le \inf_{a\in\mathbb{R}} \|v-a\|_{0,\omega_z} \lesssim D_z|v|_{1,\omega_z} \ ,$$

allows us to deduce that $\beta_{zF} \lesssim D_z^{-1/2}|v|_{1,\omega_z}$. Therefore, $|w_z|_{1,\omega_z} \le c_{1z}|v|_{1,\omega_z}$ for some scale-invariant constant c_{1z}.

The argument for $z \in \mathcal{V}_D$ follows the same general pattern of establishing that $\beta_{zF} \lesssim D_z^{-1/2}|v|_{1,\omega_z}$. In this case Poincaré-Friedrichs inequalities are used for the α_z and β_{zF}, because v vanishes on a least one face of $\partial\omega_z$. In this way we again deduce that $\beta_{z'F} \lesssim D_z^{-1/2}|v|_{1,\omega_z}$, and therefore that $|w_z|_{1,\Omega_z} \le c_{1z}|v|_{1,\omega_z}$ for some scale-invariant constant c_{1z}.

Standard inverse estimates guarantee the existence of a scale-invariant constant k_{1z} such that $\|w_z\|_{1,\tilde{\omega}_z} \le k_{1z}|v|_{1,\omega_z}$. Using the discrete Cauchy-Schwarz inequality, we can take $C_1^2 = 4\max_{z\in\mathcal{V}} k_{1z}^2$, which completes the proof of the second claim. □

LEMMA 5.4. *There are scale-invariant constants c_{2z} such that $|v\ell_z - v_z - w_z|_1 \le c_{2z}|v|_{1,\omega_z}$.*

Proof. Let $\mathbb{P}_0^3$ consist of functions which are piecewise constant (on the mesh) in each of their three components, and let $\tilde{\omega}_z$ denote the support of $v\ell_z - v_z - w_z$—this will generally be ω_z, but for $z \in \mathcal{V}_D$, it will not. We first note that, for any $\mathbf{F} \in \mathbb{P}_0^3$,

$$\int \mathbf{F}\cdot\nabla(v\ell_z - v_z - w_z) = \sum_{T\subset\tilde{\omega}_z} \int_{\partial T} \mathbf{F}\cdot\mathbf{n}_T(v\ell_z - v_z - w_z) = 0 \ .$$

Therefore, we have

$$|v\ell_z - v_z - w_z|_{1,\tilde{\omega}_z} \le \inf_{\mathbf{F}\in\mathbb{P}_0^3} \|\nabla(v\ell_z) - \mathbf{F} - \nabla w_z\|_{0,\tilde{\omega}_z} \le |w_z|_{1,\tilde{\omega}_z} + \inf_{\mathbf{F}\in\mathbb{P}_0^3} \|\nabla(v\ell_z) - \mathbf{F}\|_{0,\omega_z}$$
$$\le |w_z|_{1,\tilde{\omega}_z} + \|\ell_z\nabla v\|_{0,\omega_z} + \left(\sum_{T\subset\omega_z} \|(v-v_T)\nabla\ell_z\|_{0,T}^2\right)^{1/2} ,$$

where v_T is the average value of v on T. By Lemma 5.3, we can bound $|w_z|_{1,\tilde{\omega}_z}$ in terms of $|v|_{1,\omega_z}$, and it is clear that $\|\ell_z\nabla v\|_{0,\omega_z}$ is bounded by $|v|_{1,\omega_z}$, so we need only consider $\|(v-v_T)\nabla\ell_z\|_{0,T}$. We have

$$\|(v-v_T)\nabla\ell_z\|_{0,T} = |(\nabla\ell_z)_{|_T}|^2 \, \|v-v_T\|_{0,T}^2 \le \frac{h_T^2}{d_T^2\pi^2} \, |v|_{1,T}^2 \ ,$$

where h_T is the longest edge of T and d_T is the distance from z to the opposite face in T. Here we have used [16, 6] for $\|v-v_T\|_{0,T}^2 \le \frac{h_T^2}{\pi^2}\,|v|_{1,T}^2$. □

This leads us to our key Poincaré-type estimate of the quasi-interpolant for this section, namely

LEMMA 5.5. *There are scale-invariant constants c_{3z} such that $\|v\ell_z - v_z - w_z\|_0 \le c_{3z}D_z|v|_{1,\omega_z}$.*

Proof. Noting that $v\ell_z - v_z - w_z$ vanishes on a subset of $\partial\tilde{\omega}_z$ having nonzero measure (generally all of $\partial\tilde{\omega}_z$), we use a Poincaré-Friedrichs inequality to establish that $\|v\ell_z - v_z - w_z\|_{0,\tilde{\omega}_z} \lesssim D_z|v\ell_z - v_z - w_z|_{1,\tilde{\omega}_z}$. Combining this with Lemma 5.4 yields the result. □

We finally give a trace-type estimate for the quasi-interpolant.

LEMMA 5.6. *There are scale-invariant constants c_{4F} such that $\|v-\hat{v}-w\|_{0,F} \leq c_{4F}|F|^{1/4}|v|_{1,\Omega_F}$, where $\mathcal{V}_F$ are the vertices of F and $\Omega_F = \cup\{\tilde{\omega}_z : z \in \mathcal{V}_F\}$.*

Proof. Since $\sum_{z\in\mathcal{V}_F} \ell_z = 1$ on F, we have

$$\|v-\hat{v}-w\|_{0,F} \leq \sum_{z\in\mathcal{V}_F} \|v\ell_z - v_z - w_z\|_{0,F} \ ,$$

so the problem is reduced to that of bounding a single term of this sum. Using Lemma 3.2 we have

$$\begin{aligned}\frac{3|\omega_F|}{|F|}\|v\ell_z - v_z - w_z\|_{0,F}^2 &= \int_{\omega_F} 3(v\ell_z - v_z - w_z)^2 + \mathbf{d}_F \cdot \nabla[(v\ell_z - v_z - w_z)^2] \\ &\leq \int_{\omega_F} 4(v\ell_z - v_z - w_z)^2 + [\mathbf{d}_F \cdot \nabla(v\ell_z - v_z - w_z)]^2 \\ &\leq [4(c_{3z}D_z)^2 + (c_{2z}D_z)^2]|v|_{1,\omega_z}^2 \ .\end{aligned}$$

Using the shape-regularity assumption, we have

$$\|v\ell_z - v_z - w_z\|_{0,F}^2 \lesssim \frac{D_z^2}{|\omega_z|}\,|F|\,|v|_{1,\omega_z}^2 \lesssim |F|^{1/2}\,|v|_{1,\omega_z}^2 \ ,$$

which completes the proof. □

6. Computational Cost. We argue that the matrix associated with the computation of ε, though larger than that associated with the computation of $\hat{u}$, is much better conditioned, and the corresponding linear system can be solved to sufficient accuracy cheaply via a Krylov iteration (CG, GMRES, Bi-CGstab, etc.) with either no preconditioning or very simple preconditioning such as Jacobi or Gauss-Seidel. More particularly, we will argue that this matrix is spectrally equivalent to its diagonal, which is certainly not the case for the matrix associated with computing $\hat{u}$.

Let B, and $\hat{B}$ be the matrices defined by $B_{ij} = B(b_j, b_i)$ and $\hat{B}_{ij} = (b_j, b_i)_{H^1(\Omega)}$, and let D and $\hat{D}$ be their diagonals, respectively. We will argue that:

1. B and $\hat{B}$ are spectrally equivalent.
2. D and $\hat{D}$ are spectrally equivalent.
3. $\hat{B}$ and $\hat{D}$ are spectrally equivalent—this statement is not true for the analogously-defined matrices for piecewise linear elements.
4. B and D are spectrally equivalent—this will follow immediately from the previous three assertions.

LEMMA 6.1. *B and $\hat{B}$ are spectrally equivalent.*

Proof. Let M and m be the optimal boundedness and coercivity constants, $|B(v,w)| \leq M\|v\|_1\|w\|_1$, $B(v,v) \geq m\|v\|_1^2$. Let $\mu = \mu_1 + i\mu_2$ be an eigenvalue of B, having corresponding eigenvector $\mathbf{v} = \mathbf{v}_1 + i\mathbf{v}_2$, with $|\mathbf{v}| = 1$ and $\mu_1, \mu_2, \mathbf{v}_1, \mathbf{v}_2$ real. It is readily deduced that $\mu_1 = \mathbf{v}_1^t B\mathbf{v}_1 + \mathbf{v}_2^t B\mathbf{v}_2$ and $\mu_2 = \mathbf{v}_1^t B\mathbf{v}_2 - \mathbf{v}_2^t B\mathbf{v}_1$. Let $\mathbf{w} \in \mathbb{R}^{|\mathcal{F}|}$; it is the coefficient vector of some $w \in W$. Recognizing that $\mathbf{w}^t\hat{B}\mathbf{w} = \|w\|_1^2$ and $\mathbf{w}^t B\mathbf{w} = B(w,w)$, and using the boundedness and coercivity properties, we have

$$m\lambda_{\min}(\hat{B})|\mathbf{w}|^2 \leq \mathbf{w}^t B\mathbf{w} \leq M\lambda_{\max}(\hat{B})|\mathbf{w}|^2 \ .$$

Applying these inequalities appropriately to the expressions for μ_1 and μ_2 yields

$$m\lambda_{\min}(\hat{B}) \leq \mu_1 \leq M\lambda_{\max}(\hat{B}) \text{ and } |\mu_2| \leq M\lambda_{\max}(\hat{B}) \ .$$

In other words, B and $\hat{B}$ are spectrally equivalent □

REMARK 6.2. *Although the argument given above is more than is really necessary for the assertion, we wished to give more detail about* **how** *the spectrum of A controls the spectrum of B.*

LEMMA 6.3. *D and $\hat{D}$ are spectrally equivalent.*

Proof. The diagonal entries of D and $\hat{D}$ are $D_{ii} = B(b_i, b_i)$ and $\hat{D}_{ii} = (b_i, b_i)_{H^1(\Omega)}$, so $m\hat{D}_{ii} \leq |D_{ii}| \leq M\hat{D}_{ii}$. □

In order to show that $\hat{B}$ and $\hat{D}$ are spectrally equivalent, we consider the element-matrices $\hat{B}_T$ and $\hat{D}_T$ for each T, which we now define. For a given $T \in \mathcal{T}$, let $\hat{B}_T$ be the element-matrix defined by $(\hat{B}_T)_{ij} = (b_{Tj}, b_{Ti})_{H^1(T)}$, where $b_{Tk} \in W$ are the face-bubble functions associated with the non-Dirichlet faces of T, and let $\hat{D}_T$ be its diagonal. By construction, we have, for any $\mathbf{v} \in \mathbb{R}^{|\mathcal{F}|}$,

$$\mathbf{v}^t \hat{B} \mathbf{v} = \sum_{T \in \mathcal{T}} \mathbf{v}_T^t \hat{B}_T \mathbf{v}_T \quad \text{and} \quad \mathbf{v}^t \hat{D} \mathbf{v} = \sum_{T \in \mathcal{T}} \mathbf{v}_T^t \hat{D}_T \mathbf{v}_T \ , \tag{6.1}$$

where $\mathbf{v}_T$ is the sub-vector of $\mathbf{v}$ consisting only of those components associated with (non-Dirichlet) faces of T. Most of these matrices are 4×4, and have the form $\hat{B}_T = G_T + M_T$, where

$$(G_T)_{ij} = 2\,\frac{3!|T|}{7!} \left\{ \begin{array}{lcl} \sum_{m=1}^{4} \frac{1}{d_m^2} & , & i = j \\ \frac{\cos\theta_{kl}}{d_k d_l} - \frac{2\cos\theta_{ij}}{d_i d_j} & , & i \neq j \end{array} \right. \tag{6.2}$$

$$(M_T)_{ij} = \frac{3!|T|}{9!} \left\{ \begin{array}{lcl} 8 & , & i = j \\ 4 & , & i \neq j \end{array} \right. , \tag{6.3}$$

where we have used the notation and results of Lemma 3.2. If some (no more than three) of the faces of T are on the Dirichlet portion of the boundary, then the corresponding element-matrix matrix is a principle submatrix of the 4×4 version, so we lose no generality in our argument that $\hat{B}_T$ and $\hat{D}_T$ are spectrally equivalent by always treating them as 4×4 matrices.

LEMMA 6.4. *Let $\mathcal{F}$ be a shape-regular family of meshes. Then $\hat{B}_T$ and $\hat{D}_T$ are spectrally-equivalent, with constants-of-equivalence independent of the meshes.*

Proof. All four eigenvalues of $\hat{D}_T$ are identical, and equal to

$$\frac{3!|T|}{7!} \sum_{m=1}^{4} \frac{2}{d_m^2} + 8\,\frac{3!|T|}{9!} \ ,$$

so the shape-regularity assumption implies that there are scale-invariant constants $k_0, k_1 > 0$ such that, for any $\mathcal{T} \in \mathcal{F}$ and any $T \in \mathcal{T}$, $\sigma(\hat{D}_T) \subset [k_0 h_T, k_1 h_T]$.

Since $\mathbf{v}^t G_T \mathbf{v}^t = |v|_{1,T}^2$ for some cubic function v on T which vanishes on all of the edges of T, G_T is non-singular. It is clear that the entries of G_T are all on the order of h_T, so its non-singularity, together with the shape-regularity assumption, implies that there are scale-invariant constants $k_2, k_3 > 0$ such that, for any $\mathcal{T} \in \mathcal{F}$ and any $T \in \mathcal{T}$, $\sigma(G_T) \subset [k_2 h_T, k_3 h_T]$. But it also holds that the eigenvalues of M_T are $\frac{3!|T|}{9!}\{4, 4, 4, 20\}$, so there are scale-invariant constants $k_4, k_5 > 0$ such that, for any $\mathcal{T} \in \mathcal{F}$ and any $T \in \mathcal{T}$, $\sigma(\hat{B}_T) \subset [k_4 h_T, k_5 h_T]$. This completes the proof. □

REMARK 6.5. *If this argument was applied to the element matrices corresponding to piecewise-linear finite elements, it would break down at the assertion that the analogous G_T was non-singular. The vector of ones, which corresponds to a constant function in this case, is in the nullspace.*

LEMMA 6.6. *Let $\mathcal{F}$ be a shape-regular family of meshes. Then $\hat{B}$ and $\hat{D}$ are spectrally-equivalent, with constants-of-equivalence independent of the meshes.*

Proof. This is an immediate consequence of Lemma 6.4 and (6.1). □

We finally arrive at our main result for this section, which follows directly from Lemmas 6.1, 6.3 and 6.6:

THEOREM 6.7. *Let $\mathcal{F}$ be a shape-regular family of meshes. Then B and D are spectrally-equivalent, with constants-of-equivalence independent of the meshes.*

To provide some sense of what we might expect from constants of equivalence for the spectra of $\hat{B}_T$ and $\hat{D}_T$, we consider two "reference" tetrahedra for which the spectra are known explicitly. If T is congruent to the tetrahedron having vertices $(0,0,0),(h,0,0),(0,h,0),(0,0,h)$, then the eigenvalues of $\hat{D}_T^{-1}\hat{B}_T$ are

$$\frac{180+h^2}{2(108+h^2)}\ ,\ \frac{27216+576h^2+3h^4 \pm 2\sqrt{49128768-139968h^2-3564h^4+126h^6+h^8}}{2(11664+216h^2+h^4)}\ ,$$

where the first of these is a double eigenvalue. For $h \in (0,1]$, these eigenvalues deviate very little from their values at $h=0$,

$$0.56 < \frac{5}{6}\ ,\ \frac{7\pm\sqrt{13}}{6} < 1.77\ .$$

If T is a regular tetrahedron with side length h, then the eigenvalues of $\hat{D}_T^{-1}\hat{B}_T$ are

$$\frac{234+h^2}{2(108+h^2)}\ ,\ \frac{162+5h^2}{2(108+h^2)}\ ,$$

where the first of these is a triple eigenvalue. For $h \in (0,1]$, the smaller of these eigenvalues decreases from $167/218 \approx 0.766$ to $3/4 = 0.75$, and the larger increases from $235/218 \approx 1.078$ to $13/12 = 1.08\bar{3}$, as $h \to 0$.

REMARK 6.8. **Edge-bumps versus face-bumps.** *Referring back to Remark 2.1, essentially the same analysis could be done for the edge-bump stiffness matrix to argue that it, too, is spectrally equivalent to its diagonal. So solving such a system should be inexpensive. However, a result like* (2.10) *remains elusive for the edge-bump case. Although there tends to be more faces than edges in a tetrahedral mesh, which implies that the face-bump matrix has more rows and columns than its edge-bump counterpart, the face-bump matrix tends to have fewer non-zero entries per row and in total. The number of non-zeros in any row for the face-bump matrix is bounded by seven, regardless of the topology of the mesh. In contrast, the number of non-zeros in a row for the edge-bump matrix is strongly influenced by mesh topology—if a non-boundary edge E is surrounded by a ring of m elements, then there should be $3m+1$ non-zero entries for the corresponding row. To give a concrete example, the highly adapted meshes for the Jump Coefficient problem in Section 7 yield the following comparisons: the number of faces is roughly 1.6 times the number of edges, the average number of non-zeros in a given row of the edge-bump matrix is 16, and the edge-bump matrix has roughly 1.4 times the number total non-zeros than its face-bump counterpart.*

7. Experiments. In this section we provide a variety of examples which illustrate the robust effectivity of our proposed estimator. In most of our examples the bilinear form $B : \mathcal{H} \times \mathcal{H} \to \mathbb{R}$ is an inner-product ($\mathbf{b} = \mathbf{0}$), with induced "energy"-norm defined by $|\!|\!| v |\!|\!|^2 = B(v,v)$. In these cases, it is most natural to assess the energy of the error $|\!|\!| u - \hat{u} |\!|\!|$, and we do so using $\eta = |\!|\!| \varepsilon |\!|\!|$, and take $\eta_T^2 = |B_T(\varepsilon,\varepsilon)|$ as local local error indicators for adaptive refinement. Here $B_T(\cdot,\cdot)$ is the restriction to the

tetrahedron T of the integral defining $B(\cdot,\cdot)$. In our examples, $B_T(\cdot,\cdot)$ is an inner-product as well, and we define the induced norm $\|\!|\cdot|\!\|_T$ accordingly. When B is not an inner-product, we assess the H^1-error $\|u-\hat{u}\|_1$ by $\|\varepsilon\|_1$, and take as local indicators $\eta_T = \|\varepsilon\|_{1,T}$. Among the data reported are

1. Global and local effectivities

$$\frac{|\!|\!|\varepsilon|\!|\!|}{|\!|\!|u-\hat{u}|\!|\!|}\ ,\ \frac{|\!|\!|\varepsilon|\!|\!|_T}{|\!|\!|u-\hat{u}|\!|\!|_T} \quad \text{or} \quad \frac{\|\varepsilon\|_1}{\|u-\hat{u}\|_1}\ ,\ \frac{\|\varepsilon\|_{1,T}}{\|u-\hat{u}\|_{1,T}}$$

   Local effectivities are only reported in cases in which the exact solution is known explicitly, and maximal, minimal, and average local effectivities are provided. To obtain global effectivities in the energy-norm when the exact solution is not known, we use the fact that $|\!|\!|u-\hat{u}|\!|\!|^2 = |\!|\!|u|\!|\!|^2 - |\!|\!|\hat{u}|\!|\!|^2$, and estimate $|\!|\!|u|\!|\!|^2$ from the energy norm of the finite element solution on an extremely fine mesh.
2. Condition numbers for the matrices used to compute ε, both with and without (symmetric) diagonal rescaling.
3. Convergence history of the adaptive method, together with the theoretically optimal rate, and that obtained by uniform refinement.

Comparisons are also made with a standard residual-based error indicator [4, 17] in terms of effectivities and convergence histories. The local residual indicators are given by

$$\eta_{r,T}^2 = \frac{1}{2}\sum_{F\in\mathcal{F}_I} h_T\|r_F\|_{0,F}^2 + \sum_{F\in\partial T\cap\partial\Omega_N} h_T\|r_F\|_{0,F}^2\ ,$$

where r_F is the face residual, and the global indicator is defined by $\eta_r^2 = \sum_{T\in\mathcal{T}}\eta_{r,T}^2$. For all of these problems, a typical adaptive approach based on Dörfler marking [9] and longest-edge bisection [18] is employed, all of which are available in the package FETK [12].

7.1. Poisson Problem. Here, we solve

$$-\triangle u = f$$

in the domain $\Omega = \mathrm{int}([-1,0]^3\cup[0,1]\times[-1,0]^2\cup[-1,0]\times[0,1]\times[-1,0]\cup[-1,0]^2\times[0,1])$ subject to homogeneous Dirichlet conditions. Here, f is chosen so that the exact solution is given by

$$u(x,y,z) = \frac{\sin(\pi x)\sin(\pi y)\sin(\pi z)}{(0.001+x^2+y^2+z^2)^{1.5}}\ .$$

For this problem, due to the behavior of the first derivative of u at the origin, we expect to require most refinement near the origin. The initial coarse mesh is plotted in Figure 7.1. An adapted mesh with around 30 thousand vertices, computed using the face-bump error indicator, is given in Figure 7.2.

We begin by exploring the effectivity of the face-bump error indicator as a global indicator. Figure 7.3 shows the effectivities on the adapted meshes computed by the adaptive finite element algorithm using the given indicator. We see that the effectivity of the face-bump indicator is insensitive to changing mesh size.

In order to study the face-bump indicator as a local error indicator, we can use the exact solution to compute not only the global effectivity, but also the local effectivity

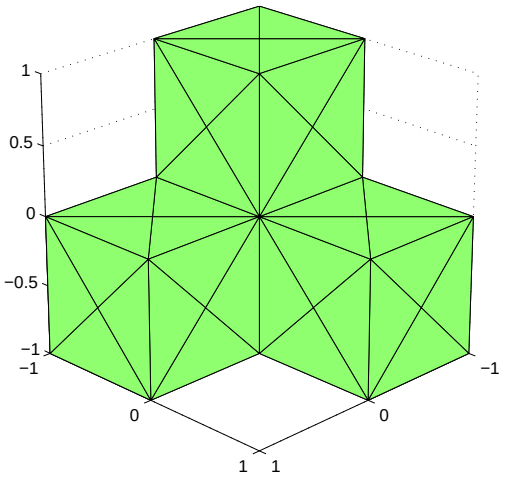

FIG. 7.1. *Initial mesh for the Poisson problem.*

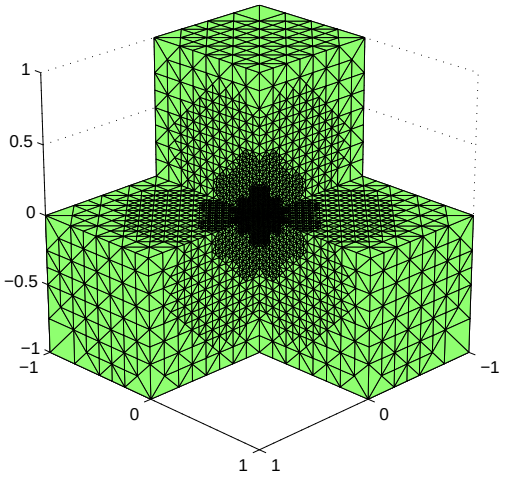

FIG. 7.2. *Adapted mesh with 29549 vertices for the Poisson problem.*

of the error indicator. Figure 7.4 shows a plot of the minimum, maximum, and average element effectivities of both the face-bump and residual error indicators over the series of meshes computed by the adaptive algorithm. From this figure, we see that the average element effectivity of the face-bump error indicator is very stable at around 0.4 and although the maximum and minimum effectivities are not quite as stable, they are much more stable than the effectivities for the residual error indicator.

Another, and perhaps the most important, method of evaluating the face-bump

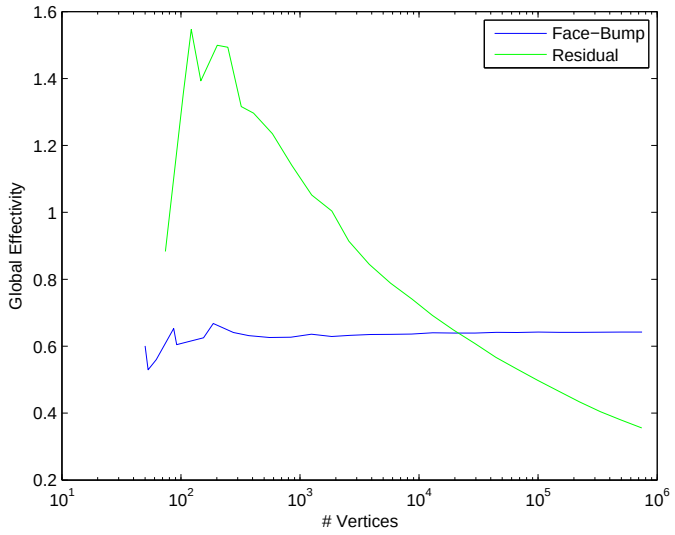

FIG. 7.3. *Global effectivities of the face-bump and residual error indicators for the Poisson problem.*

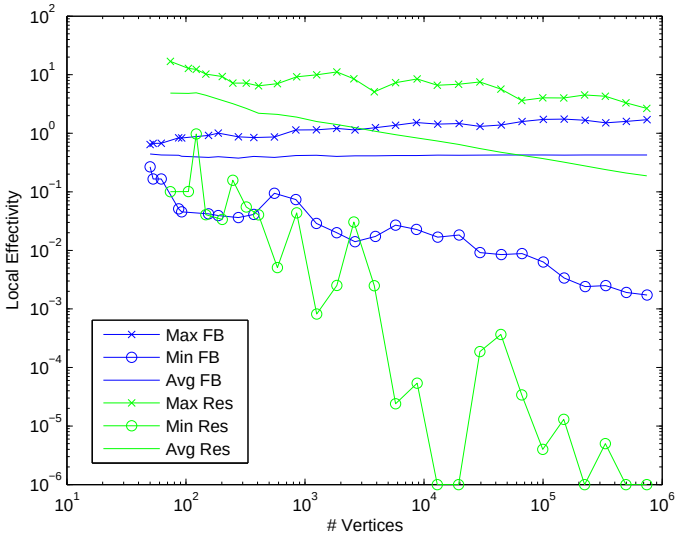

FIG. 7.4. *Maximum, minimum, and average local effectivities of the face-bump and residual error indicators for the Poisson problem.*

error indicator as a local indicator is to see how the error decreases with mesh refinement when using the indicator to control an adaptive algorithm. Figure 7.5 shows the error as a function of the number of vertices in the meshes. It is compared to the error from using the residual indicator, simple uniform refinement, as well as a reference line of optimal order $N^{-\frac{1}{3}}$. We see that both of the adaptive meshes manage

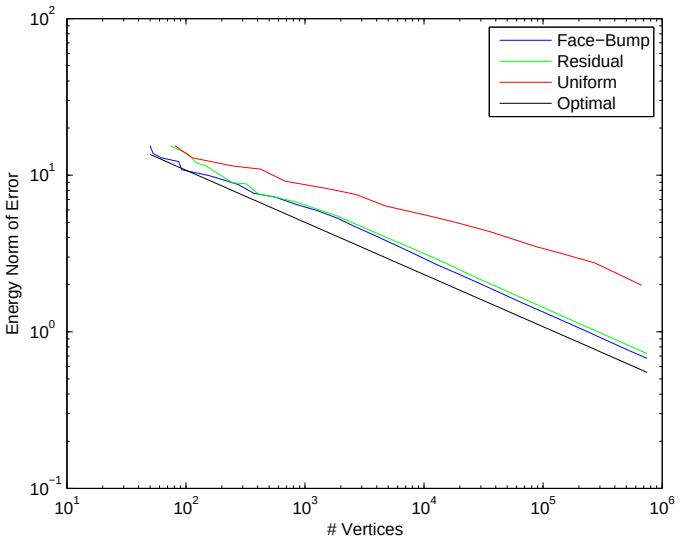

FIG. 7.5. *Energy norm of the error of a solution computed on a mesh computed using face-bump and residual error indicators, as well as using uniform refinement for the Poisson problem.*

to attain optimal convergence rate, while the uniform refinement scheme lags behind.

The final consideration is the cost of the face-bump indicator for this problem. In order to justify the cost of solving the face-bump system, we must demonstrate that the face-bump stiffness matrix is easily preconditioned and that the conjugate gradient method converges quickly with this preconditioner. Figure 7.6 shows that although the condition number of the unscaled matrix grows as the mesh is adaptively refined, the condition number is extremely stable and small after Jacobi preconditioning. Similarly, with a stopping criteria for the conjugate gradient method of reducing the relative residual by a factor of 10^8 over the initial one, the CG method with Jacobi preconditioning converged in under 16 iterations for all mesh sizes.

7.2. Jump Coefficient Problem. Here we turn our attention to the problem

$$-\nabla \cdot (a(x)\nabla u) = 0$$

on the domain $\Omega = (-1,1)^3$ where $a(x) = 10000$ for $x \in (0,1)^3$ and $a(x) = 1$ otherwise, and subject to the boundary conditions $u(1,y,z) = 1$, $u(-1,y,z) = 0$, and homogeneous Neumann boundary condition elsewhere. This test problem can be found in [17]. For this problem, singularities in the solution arise along the interior edges and the interior vertex of the interface. We therefore expect a high amount of refinement along these edges. We use an initial course mesh, shown in Figure 7.7, which is comprised of 8 sub-cubes so as to resolve the different subdomains. Figures 7.8 and 7.9 show an adapted mesh and a cut-away of this mesh, respectively, computed using the face-bump indicator.

Again, we begin by studying the performance of the face-bump error indicator as a global indicator. Figure 7.10 shows the global effectivity of the face-bump and residual error indicators for varying adaptively refined meshes. Recall that for this problem, without an exact solution, the exact error is approximated using the solution

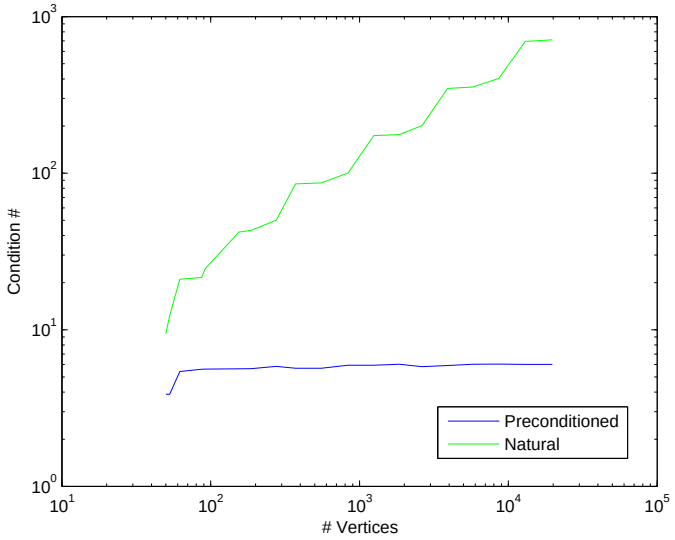

FIG. 7.6. *Condition number of the unpreconditioned and Jacobi preconditioned face-bump stiffness matrix for the Poisson problem.*

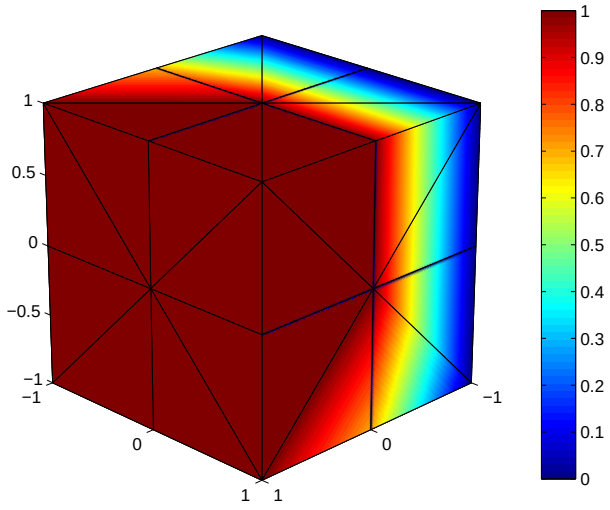

FIG. 7.7. *Initial mesh (colored by solution value) for the jump coefficient problem.*

on a mesh with over 3 million vertices. As we see, the face-bump error indicator has an extremely stable effectivity for this problem.

Without an exact solution, computing local effectivities is more difficult, and so to study the performance of the face-bump indicator as a local error indicator, we simply consider its error reduction properties in an adaptive algorithm. Figure 7.11 shows the error in an approximate solution computed on meshes computed using the

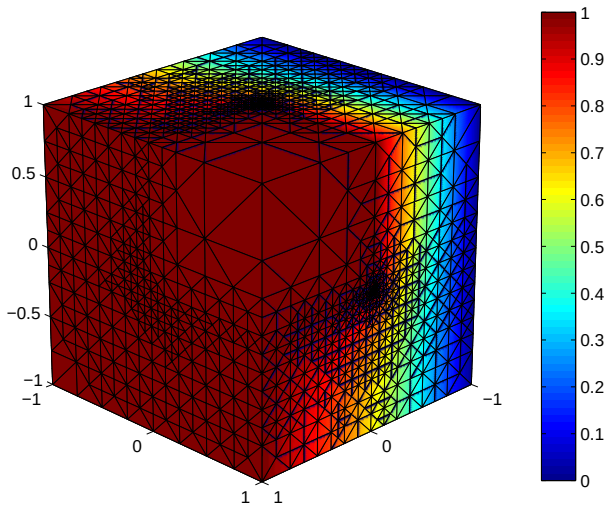

FIG. 7.8. *Adapted mesh (colored by solution value) with 37458 vertices for the jump coefficient problem.*

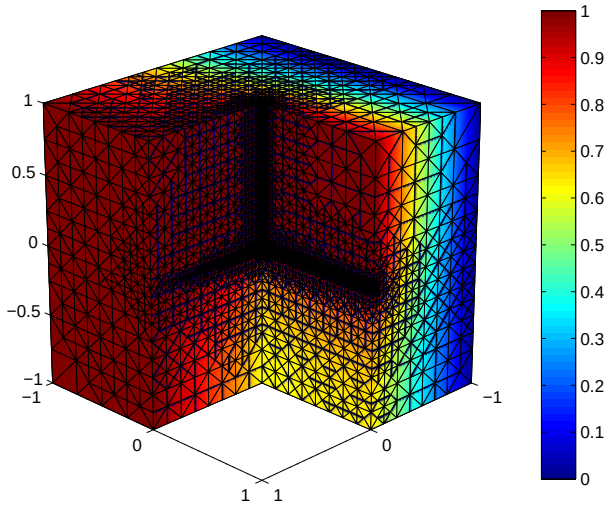

FIG. 7.9. *Adapted mesh (colored by solution value) with 37458 vertices for the jump coefficient problem.*

face-bump indicator, residual indicator, and uniform refinement, as well as a reference line of optimal convergence rate $N^{-\frac{1}{3}}$. We see that the face-bump indicator attains optimal rate and outperforms the residual error indicator.

We again turn our attention to the cost of the face-bump indicator. Figure 7.12 shows the condition number of the unscaled and diagonally scaled face-bump stiffness

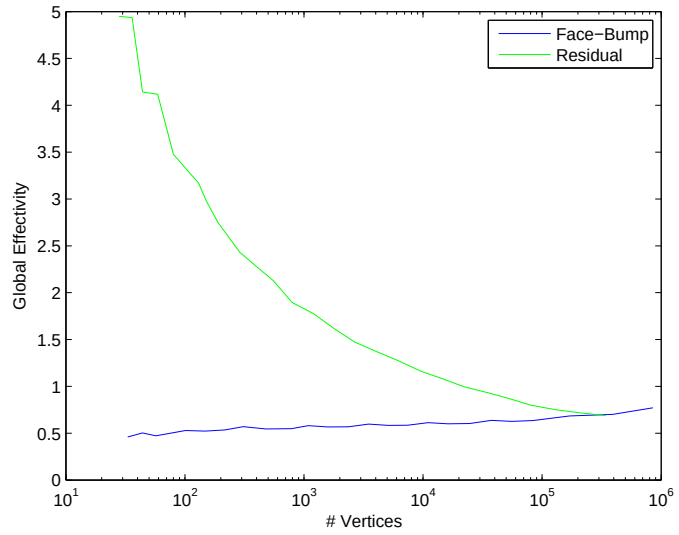

FIG. 7.10. *Global effectivities of the face-bump and residual error indicators for the jump coefficient problem.*

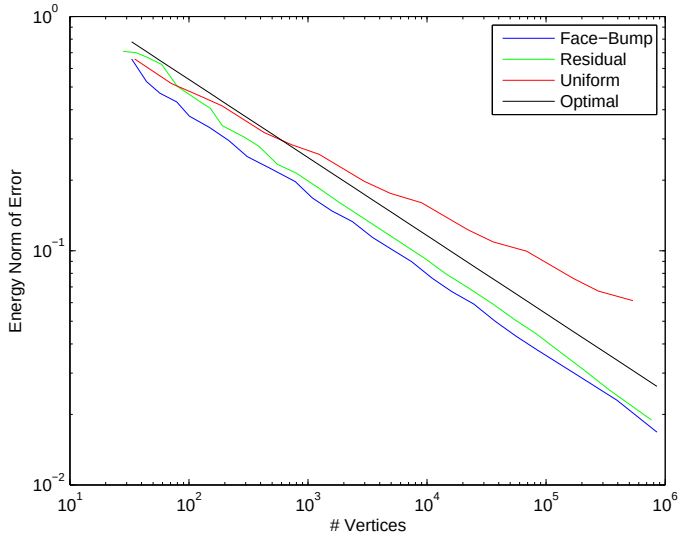

FIG. 7.11. *Energy norm of the error of a solution computed on a mesh computed using face-bump and residual error indicators, as well as using uniform refinement for the jump coefficient problem.*

matrix. Again with simple Jacobi preconditioning, the condition number remains small and stable. With this preconditioner, the number of conjugate gradient iterations required to solve the face-bump problem to high accuracy is less than 17 for all iterations of the adaptive algorithm.

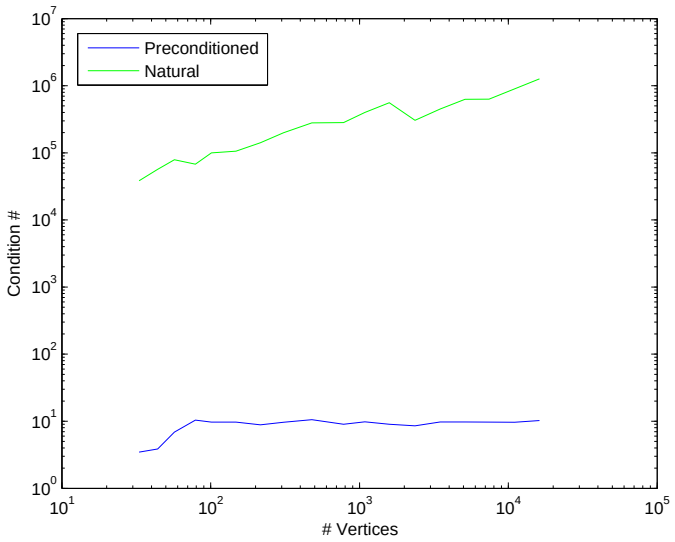

FIG. 7.12. *Condition number of the unpreconditioned and Jacobi preconditioned face-bump stiffness matrix for the jump coefficient problem.*

7.3. Convection-Diffusion Problem. Here we consider a problem with convection in order to test the indicator on a problem which gives a non-symmetric system matrix. Specifically, we solve

$$-\epsilon \triangle u + u_x = 1$$

on the domain $\Omega = (0,1)^3$ subject to $u(x=0) = u(x=1) = 0$ and homogeneous Neumann condition on the other four faces. This problem has exact solution independent of y and z given by

$$u(x,y,z) = x - \frac{e^{\frac{x-1}{\epsilon}} - e^{-\frac{1}{\epsilon}}}{1 - e^{-\frac{1}{\epsilon}}} \ .$$

For these experiments, we used $\epsilon = 0.1$. The initial course mesh is shown in Figure 7.13 while Figure 7.14 shows an adapted solution computed using the face-bump indicator.

The global and local effectivities are shown in Figures 7.15 and 7.16, respectively. As we see, the face bump indicator is remarkably robust both locally and globally even in the presence of the convection term. Figure 7.17 shows the convergence profile (in the H^1 norm) of the face-bump and residual indicator-based methods as well as uniform mesh refinement. For this problem, we see that even uniform mesh refinement is able to attain optimal convergence rate, but the adapted meshes still maintain better placement of the unknowns.

Finally, Figure 7.18 shows that the conditioning of the matrix remains well behaved for this problem. The number of BiCG iterations with diagonal preconditioning remains bounded by 17 for all steps of the adaptive process.

7.4. Anisotropic Diffusion Problem. Finally, we consider a problem with an anisotropic diffusion coefficient. Specifically, let the matrix A be the 3×3 diagonal

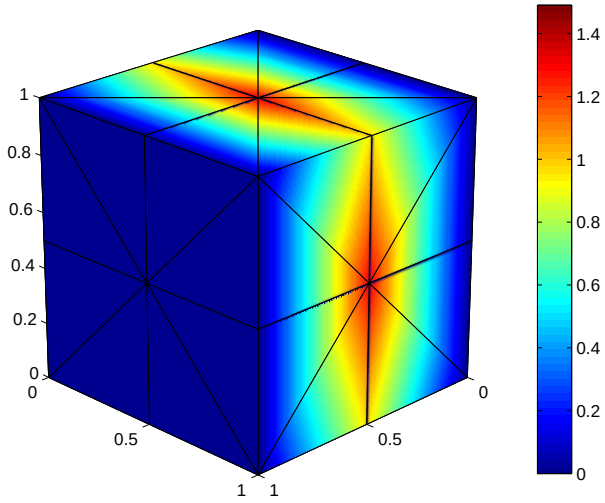

FIG. 7.13. *Initial mesh (colored by solution value) for the convection-diffusion problem.*

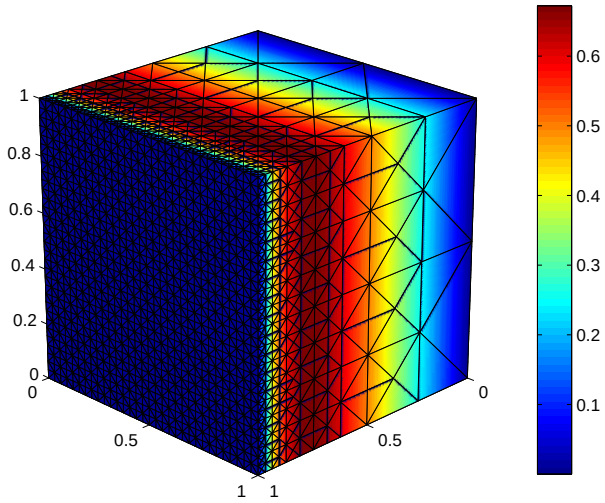

FIG. 7.14. *Adapted mesh (colored by solution value) with 10588 vertices for the convection-diffusion problem.*

matrix with diagonal entries $1, \epsilon, \epsilon^{-1}$, where $\epsilon = 10^{-3}$. Then we consider the problem of solving

$$-\nabla \cdot (A \nabla u) = 1$$

on the domain $\Omega = (-1, 1)^3$ subject to homogeneous Dirichlet boundary conditions. This high anisotropy can cause the residual error indicator to have poor global ef-

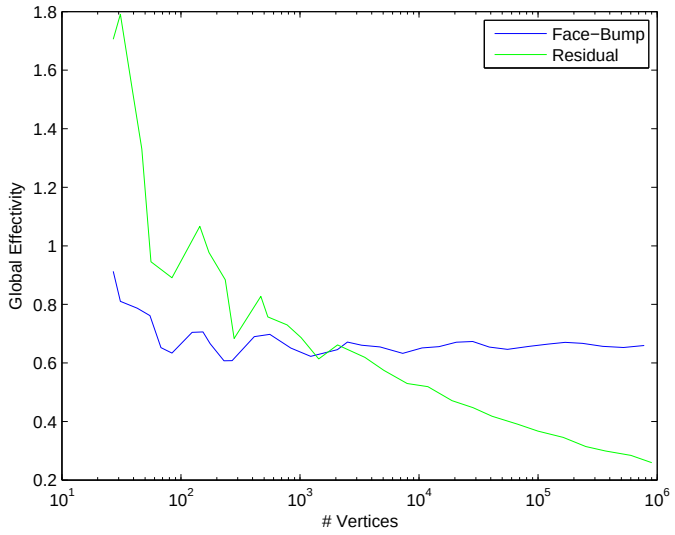

FIG. 7.15. *Global effectivities of the face-bump and residual error indicators for the convection-diffusion problem.*

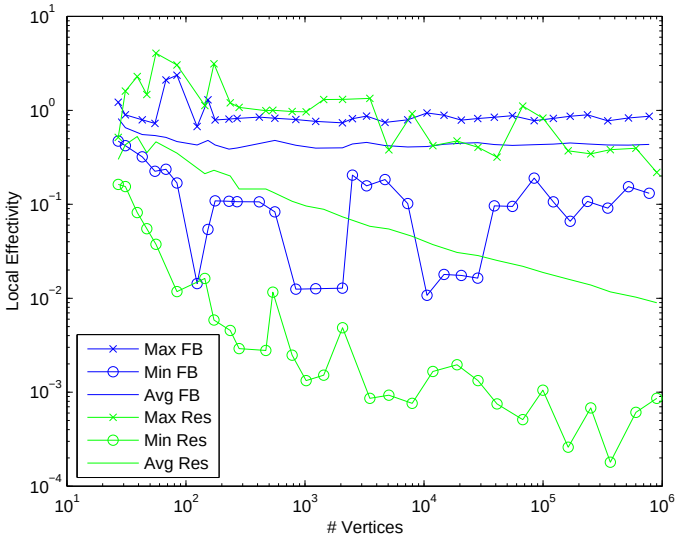

FIG. 7.16. *Maximum, minimum, and average local effectivities of the face-bump and residual error indicators for the convection-diffusion problem.*

fectivity [11], and we use this example only to study the effectivity of the face-bump indicator.

The global effectivities of the two indicators are shown in Figure 7.19. The energy norm of the error for both adaptive algorithms as well as uniform refinement are shown in Figure 7.20. It is apparent that the face-bump indicator retains its usual

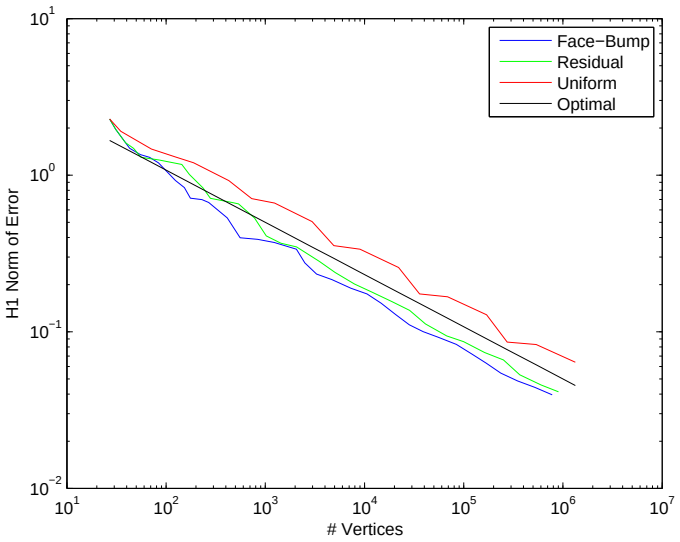

FIG. 7.17. *H^1 norm of the error of a solution computed on a mesh computed using face-bump and residual error indicators, as well as using uniform refinement for the convection-diffusion problem.*

stable global effectivity. However the adaptive algorithms are unable to attain optimal convergence rates—likely due to the longest-edge bisection, which does not allow for anisotropic elements. In this case, the matrix for computing ε is ill-conditioned (even after rescaling) due to the anisotropy in A, which results in more conjugate gradient iterations for the computation of ε. A refinement scheme which allowed for elements which were shape regular with respect to the anisotropic metric induced by A would likely improve not only the conditioning, but also the convergence of the method; but our primary interest here was merely to demonstrate that our hierarchical estimator does not suffer the same deficiencies as residual estimators in terms of effectivity.

8. Final Remarks. For hierarchical estimators, such as the one here proposed, establishing an efficiency estimate (2.9) is a simple exercise—it is reliability estimates which require deeper analysis. In this article, we have proven such an estimate in Theorem 4.6, thereby establishing the equivalence of the error $\|u - \hat{u}\|_1$ and error estimate $\|\varepsilon\|_1$ up to an oscillation term which is readily computable or estimable, if desired. This result is obtained under minimal practical assumptions: the problem data are piecewise smooth, the family of meshes is shape-regular and aligns with discontinuities in the data, and the bilinear form satisfies an inf-sup condition for the continuous and discrete problems. We emphasize that our results are not restricted to the symmetric, energy norm, setting; and the familiar strong Cauchy inequality and saturation assumption are replaced in the analysis and assertions by constants related to quasi-interpolation and residual oscillation. A benefit of such an analysis is that it is completely transparent upon which aspects of the data the constants and oscillation term do (and do not!) depend. In particular, both constants K_1 and K_2 depend only on the shape-regularity of the mesh and continuity and coercivity (or inf-sup) constants of the bilinear form, and it clear precisely how the oscillation term depends on G. In addition to the error equivalence results, we have argued that

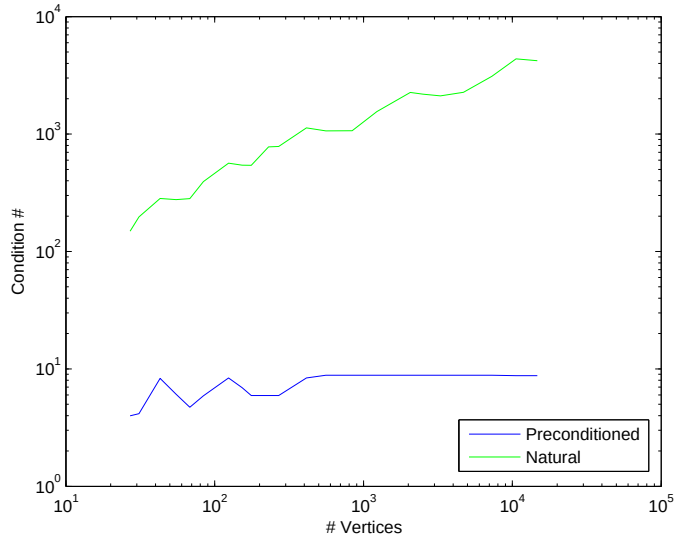

FIG. 7.18. *Condition number of the unpreconditioned and Jacobi preconditioned face-bump stiffness matrix for the convection-diffusion problem.*

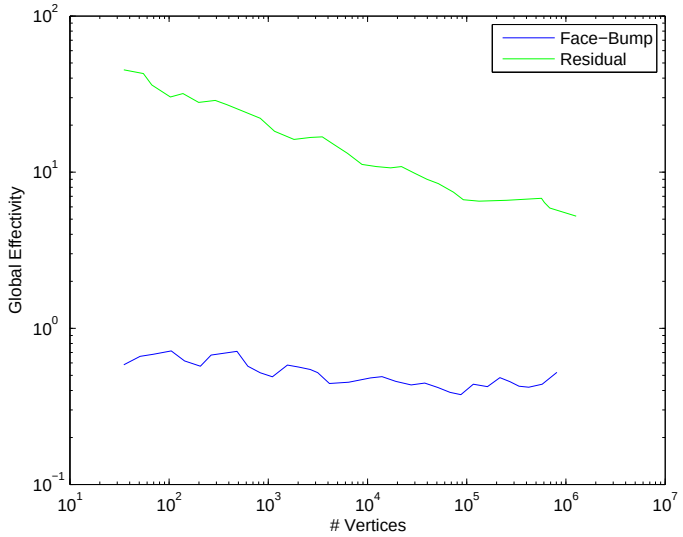

FIG. 7.19. *Global effectivities of the face-bump and residual error indicators for the anisotropic diffusion problem.*

the matrix associated with computing the approximate error function ε is spectrally equivalent to its diagonal, thereby showing that computing ε, which involves the solution of a global system, is not unreasonably expensive, as is sometimes thought. These theoretical results are clearly demonstrated in the numerical experiments, where the effectivity of the estimator and the conditioning of the associated system (after

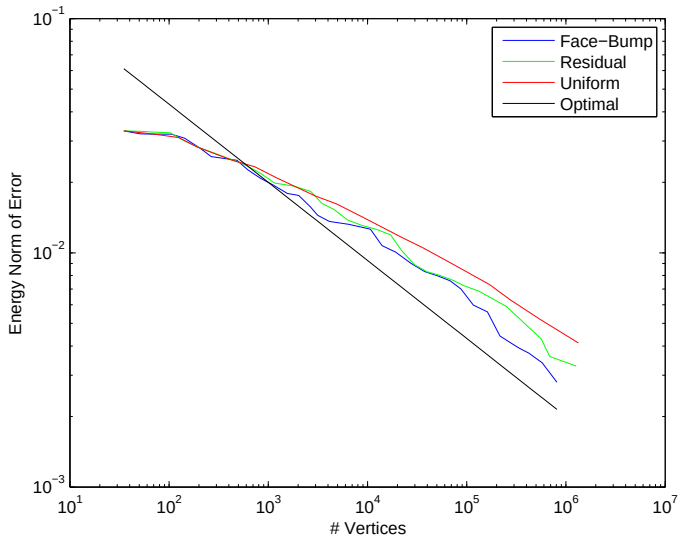

FIG. 7.20. *Energy norm of the error of a solution computed on a mesh computed using face-bump and residual error indicators, as well as using uniform refinement for the anisotropic diffusion problem.*

symmetric diagonal rescaling) are shown to be remarkably consistent—consistently good—for a variety of problems.

The general principle of our reliability analysis may be summarized briefly as follows. Split the variational error into two pieces

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

and choose $\hat{v}+w \in V \oplus W$ in such a way that a useful residual oscillation expression can be obtained from the second piece, while maintaining norm-comparability of v and w. We emphasize that one is free to choose W (and then $\hat{v}+w$) after examining the form of the variational residual $B(u-\hat{u},v-\hat{v}-w)$ to see how additional degrees of freedom from W might best be exploited. One caveat in the selection of W should be noted—the cost of computing of $\varepsilon \in W$ should be provably acceptable within the AFEM (solve-*estimate*-mark-refine) framework. In what may be considered a continuation of the present work, we plan to prove similar effectivity results for non-linear problems, and to prove convergence of the corresponding adaptive method in that context.

REFERENCES

[1] 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.

[2] C. B. Allendoerfer. Generalizations of theorems about triangles. *Math. Mag.*, 38:253–259, 1965.

[3] R. Araya, A. H. Poza, and E. P. Stephan. A hierarchical a posteriori error estimate for an advection-diffusion-reaction problem. *Math. Models Methods Appl. Sci.*, 15(7):1119–1139, 2005.

[4] I. Babuska and W. C. Rheinboldt. Error estimates for adaptive finite element computations. *SIAM Journal on Numerical Analysis*, 15(4):736–754, 1978.

[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] M. Bebendorf. A note on the Poincaré inequality for convex domains. *Z. Anal. Anwendungen*, 22(4):751–756, 2003.

[7] 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.

[8] P. Clément. Approximation by finite element functions using local regularization. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér. RAIRO Analyse Numérique*, 9(R-2):77–84, 1975.

[9] W. Dorfler. A convergent adaptive algorithm for poisson's equation. *SIAM Journal on Numerical Analysis*, 33(3):1106–1124, 1996.

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

[11] F. Fierro and A. Veeser. A posteriori error estimators, gradient recovery by averaging, and superconvergence. *Numer. Math.*, 103(2):267–298, 2006.

[12] M. Holst. Adaptive numerical treatment of elliptic systems on manifolds. *Advances in Computational Mathematics*, 15(1):139 – 191, 2001.

[13] P. Morin, R. H. Nochetto, and K. G. Siebert. Convergence of adaptive finite element methods. *SIAM Rev.*, 44(4):631–658 (electronic) (2003), 2002. Revised reprint of "Data oscillation and convergence of adaptive FEM" [SIAM J. Numer. Anal. **38** (2000), no. 2, 466–488 (electronic); MR1770058 (2001g:65157)].

[14] P. Morin, R. H. Nochetto, and K. G. Siebert. Local problems on stars: a posteriori error estimators, convergence, and performance. *Math. Comp.*, 72(243):1067–1097 (electronic), 2003.

[15] J. S. Ovall. *Duality-Based Adaptive Refinement for Elliptic PDEs*. PhD thesis, University of California, San Diego, La Jolla, CA 92093, USA, 2004.

[16] L. E. Payne and H. F. Weinberger. An optimal Poincaré inequality for convex domains. *Arch. Rational Mech. Anal.*, 5:286–292 (1960), 1960.

[17] M. Petzoldt. A posteriori error estimators for elliptic equations with discontinuous coefficients. *Advances in Computational Mathematics*, 16(1):47 – 75, 2002.

[18] M.-C. Rivara. Local modification of meshes for adaptive and/or multigrid finite-element methods. *Journal of Computational and Applied Mathematics*, 36(1):79 – 89, 1991.

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

[20] G. Strang and G. J. Fix. *An analysis of the finite element method*. Prentice-Hall Inc., Englewood Cliffs, N. J., 1973. Prentice-Hall Series in Automatic Computation.

[21] R. Verfürth. Error estimates for some quasi-interpolation operators. *M2AN Math. Model. Numer. Anal.*, 33(4):695–713, 1999.