AN ANALYSIS OF THE SATURATION ASSUMPTION

RANDOLPH E. BANK*, JINCHAO XU†, AND HARRY YSERENTANT‡

Abstract. The saturation assumption plays a central role in much of the analysis of a posteriori error estimates and refinement algorithms for adaptive finite element methods. In this work we provide an analysis of this assumption in the simple setting of interpolation.

Key words. Saturation Assumption, Adaptive Feedback Loop, A Posteriori Error Estimates

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

1. Introduction.

Consider a single shape regular simplicial finite element t in $\mathcal{R}^d$. Let h denote the diameter of t. On element t, we define a family of polynomial spaces $\mathcal{S}_p(t)$, consisting of polynomials of degree less than or equal to p. Let $H^k(t)$ denote the Sobolev space for appropriately chosen k and let $\mathcal{I}_p : H^k(t) \rightarrow \mathcal{S}_p(t)$ denote the usual Lagrange interpolation operator associated with the space $\mathcal{S}_p(t)$. We assume that k is sufficiently large that the usual a priori estimates of the form

$$\|(1-\mathcal{I}_p)u\|_{H^r(t)} \le C(q,r)h^{q-r}|u|_{H^q(t)} \tag{1}$$

for $0 \le r < q \le p+1$ hold. In this work, we focus mainly on the $H^1(t)$ seminorm. $|u|_{H^1(t)} \equiv \|\nabla u\|_{L_2(t)}$, although much of the analysis can be generalized to other Sobolev norms. For simplicity in notation, we will write $|\cdot|_1$ or $|\cdot|_{1,t}$ where appropriate.

Next we assume some refinement of element t. This could be h-refinement (division of t into several smaller elements with the same polynomial degree as t) or p-refinement (increasing the degree of the polynomial space by one.) Let $\hat{\mathcal{I}}_p$ denote the canonical interpolation operator applied to the refinement of t. Thus in the case of p-refinement we have $\hat{\mathcal{I}}_p \equiv \mathcal{I}_{p+1}$, or in the case of h-refinement, $\hat{\mathcal{I}}_p$ represents $\mathcal{I}_p$ applied to each of the newly refined child elements of t.

The local *saturation assumption* reads

$$|u-\hat{\mathcal{I}}_p u|_{1,t} \le \beta |u-\mathcal{I}_p u|_{1,t} \tag{2}$$

for some $\beta = \beta(u) < 1$. Informally, the saturation assumption asserts that the refinement of element t will result in a reduction of the interpolation error.

Our main interest in (2) is its use in the study of adaptive finite element methods for solving partial differential equations. Here one makes an a posteriori error estimate using the finite element solution u_h, and then employs this estimate to guide refinement of the finite element subspace in an adaptive feedback loop. Success of such an

---

[*] Department of Mathematics, University of California, San Diego, La Jolla, California 92093-0112.Email: rbank@ucsd.edu. The work of this author was supported by the National Science Foundation under contract DMS-2012857, and the Alexander von Humboldt Foundation through a Humboldt Research Award.

[†] Applied Mathematics and Computational Science Program, Electrical and Mathematical Science and Engineering Division, King Abdullah University of Science and Technology, Thuwal 23955 Saudi Arabia. Department of Mathematics, Penn State University, State College, Pennsylvania 16802.Email: xu@multigrid.org. The work of this author was supported by the National Science Foundation under contract DMS-1819157, the Alexander von Humboldt Foundation through a Humboldt Research Award, and KAUST Baseline Research Fund BAS16990101.

[‡] Institut für Mathematik, Technische Universität Berlin, 10623 Berlin, Germany.Email: yserentant@math.tu-berlin.de.

adaptive procedure depends on both the quality of the a posteriori error estimates, and that the adaptive feedback loop results in a significant reduction of the error. Saturation assumptions can play an important role in analyzing both of these aspects of adaptive procedures. The books of Babuška and Strouboulis [1], Babuška, Whiteman, and Strouboulis [2], Deuflhard and Weiser [9], Verfürth [14], and the references cited therein, together provide a rather complete overview of these topics and the role played by the saturation assumption. Several authors have studied the validation of this assumption, as well as possible alternative approaches. See for example, Nochetto [12], Dörfler and Nochetto [10], Carstensen, Gallisti and Gedicke [8], Bulle, Chouly, Hale, and Lozinski [7], Praetorius, Ruggeri, and Stephan [13], and Bank, Parsania and Sauter [3].

What these and other works have in common is that they consider the application of a global saturation assumption to the finite element solution u_h of the partial differential equation, often restricted to the case of p-refinement. In (2), we consider its application to interpolation, and allow both h- and p-refinement. Since interpolation is also quite local, this setting is simple in comparison. On the other hand, (2) still has important implications for adaptive finite element methods. In [4], Bank and Yserentant show that interpolation error is a local lower bound on the error for any finite element approximation of a given function u, in particular the finite element solution u_h. Estimate (2) is the key assumption in that analysis. Coupled with standard a priori estimates, this shows that interpolation error is both reliable and efficient as an a posteriori error estimator. As a practical matter, of course it could not be used as such, since the interpolant is generally not available in such situations.

However, if we have a practical a posteriori error estimate e_h that is both reliable and efficient, we have the estimates

$$c_1|u-\mathcal{I}_p u|_{1,\Omega} \le |u-u_h|_{1,\Omega} \le c_2|u-\mathcal{I}_p u|_{1,\Omega}$$
$$C_1|e_h|_{1,\Omega} \le |u-u_h|_{1,\Omega} \le C_2|e_h|_{1,\Omega}$$

where Ω is the domain of the given partial differential equation. It follows that

$$\frac{c_1}{C_2}|u-\mathcal{I}_p u|_{1,\Omega} \le |e_h|_{1,\Omega} \le \frac{C_1}{c_2}|u-\mathcal{I}_p u|_{1,\Omega}$$

that suggests the true finite element error, the interpolation error, and the computed a posteriori error estimate all behave in the same way. See also [11].

This provides an important tool for analyzing the convergence properties of adaptive feedback loops. One can choose known functions of the same class as the true solution of the partial differential equation, and then apply the adaptive feedback loop to these functions, using interpolation error in place of the a posteriori error estimate. This removes noise that might arise from errors in assembly and approximate solution of the finite element system, as well as the details of the practical a posteriori error estimate, and allows one to focus mainly on the properties of the adaptive feedback loop itself. In [5], Bank and Yserentant show that this indirect approach provides a framework to prove (optimal) convergence for adaptive feedback loops. We note that (2) is also a critical assumption in that analysis.

The remainder of this paper is organized as follows. In Section 2 we provide a simple analysis that can be used to prove (2). In particular, we show

$$\beta \le \beta_0 + C(u)h \tag{3}$$

where $\beta_0 \in [0,1)$ is independent of u. $\beta_0 = 0$ in the case of p-refinement, while in in Section 3, we show $\beta_0 = 2^{-p}$ for several common h-refinement schemes in

$1 \le d \le 3$ dimensions. In Section 4 we consider variable coefficients as might arise in the numerical solution of elliptic pdes. In Section 5 we consider the effect of low regularity and other exceptional behavior that could be associated with such pdes. In Section 6, we provide some numerical examples that illustrate some of our results. In Section 7 we show estimates analogous to (2)–(3) also hold for the L_2 norm.

Finally, we note that (3) depends on an asymptotic a priori estimate and therefore is itself an asymptotic estimate. Thus we consider the issue of its practical value. By analogy, global a priori error estimates are asymptotic but have proven critical in the development and theoretical error analysis of finite element discretizations. In a similar fashion, better understanding the behavior of local h-refinement and p-refinement is important to the development of adaptive finite element methods, for example deciding between h or p refinement in a hp algorithm. Also, more precise characterization of the saturation assumption adds to our understanding when used in the theoretical analysis of a posteriori error estimates, adaptive feedback loops, and other aspects of adaptive finite elements.

2. Analysis of the Saturation Assumption. We begin with the assumption $\beta_0 < 1$ where

$$\beta_0 \equiv \max_{\nu \in \mathcal{S}_{p+1}} \frac{|\nu - \hat{\mathcal{I}}_p \nu|_{1,t}}{|\nu - \mathcal{I}_p \nu|_{1,t}}. \tag{4}$$

β_0 is similar to β but restricted to polynomials of degree $p+1$. The optimization problem (4) is equivalent to finding the maximum eigenvalue of a generalized eigenvalue problem involving two symmetric, positive semidefinite matrices. Thus it is easy to see that while β generally depends on the function u, β_0 is independent of the function. It is also independent of the size h of the element t but does depend on its shape of t and its orientation due to $|\cdot|_{1,t}$[1].

The operator $1 - \mathcal{I}_p$ has a large nullspace in $\mathcal{S}_{p+1}$, namely $\mathcal{S}_p$, that is also a subspace of the nullspace of $1 - \hat{\mathcal{I}}_p$. When $\hat{\mathcal{I}}_p = \mathcal{I}_{p+1}$, its nullspace is $\mathcal{S}_{p+1}$, so $\beta_0 = 0$ for the case of p-refinement. Hence β_0 is mainly relevant to the study of h-refinement schemes. In this case it is important to require that the number of interpolation nodes of $\hat{\mathcal{I}}_p$ be at least as large as the dimension of $\mathcal{S}_{p+1}$ (here we assume the nodes of $\mathcal{I}_p$ are a subset of those of $\hat{\mathcal{I}}_p$). If not, one can show $\beta_0 = 1$ by choosing a nonzero $\nu \in \mathcal{S}_{p+1}$ such that $\mathcal{I}_p \nu = \hat{\mathcal{I}}_p \nu = 0$. As an example, in the case $p = 1$ for triangular elements in two space dimensions, a simple bisection yields four nodes for the refined piecewise linear space while the dimension of quadratic space on the original element t is six. To overcome this issue, we define $\hat{\mathcal{I}}_p$ to correspond to several levels of refinement such that the h-refined space has sufficiently many nodes.

In Section 3, we show $\beta_0 = 2^{-p}$ for several common h-refinement schemes in $1 \le d \le 3$ space dimensions. The main challenge is that the H^1-norm depends, in contrast to the L_2-norm, on the shape and orientation of the elements and not only on their size.

Lemma 1. *Let $u \in H^{p+2}(t)$ satisfy (1), $u \notin \mathcal{S}_p$, and β_0 be given by (4). Then there is a constant C depending on u, the degree p, the shape of element t, but not on*

---

[1] In this context *orientation* refers to the geometric orientation of an element within a fixed vector space. If the entire vector space is reoriented using an orthogonal transformation the norm remains invariant as usual.

its diameter h, such that

$$\frac{|u - \hat{\mathcal{I}}_p u|_{1,t}}{|u - \mathcal{I}_p u|_{1,t}} \le \beta_0 + Ch. \tag{5}$$

Proof. Using (4) and the triangle inequality, we obtain

$$\begin{aligned} |u - \hat{\mathcal{I}}_p u|_{1,t} &\le |(1 - \hat{\mathcal{I}}_p)\mathcal{I}_{p+1}u|_{1,t} + |(1 - \hat{\mathcal{I}}_p)(u - \mathcal{I}_{p+1}u)|_{1,t} \\ &\le \beta_0|(1 - \mathcal{I}_p)\mathcal{I}_{p+1}u|_{1,t} + |(1 - \hat{\mathcal{I}}_p)(u - \mathcal{I}_{p+1}u)|_{1,t} \\ &\le \beta_0|u - \mathcal{I}_p u|_{1,t} + \beta_0|(1 - \mathcal{I}_p)(u - \mathcal{I}_{p+1}u)|_{1,t} + |(1 - \hat{\mathcal{I}}_p)(u - \mathcal{I}_{p+1}u)|_{1,t} \end{aligned}$$

If $u \in \mathcal{S}_p$, $|u - \mathcal{I}_p u|_{1,t} = 0$, then (2) is trivially satisfied for any choice of β, so we exclude the nullspace of $1 - \mathcal{I}_p$. As shown by Lin, Xie, and Xu in [11], it follows that

$$|u - \mathcal{I}_p u|_{1,t} \ge C_0(u)h^p > 0.$$

Then using this estimate and (1)

$$\begin{aligned} |(1 - \mathcal{I}_p)(u - \mathcal{I}_{p+1}u)|_{1,t} &\le C_1 h|u - \mathcal{I}_{p+1}u|_{2,t} \\ &\le C_2 h^{p+1}|u|_{p+2,t} \\ &\le \left(\frac{C_2|u|_{p+2,t}}{C_0(u)}\right) h|u - \mathcal{I}_p u|_{1,t} \\ &\equiv C_3(u)h|u - \mathcal{I}_p u|_{1,t} \end{aligned}$$

for functions $u \in H^{p+2}$. A similar argument shows

$$|(1 - \hat{\mathcal{I}}_p)(u - \mathcal{I}_{p+1}u)|_{1,t} \le \hat{C}_3(u)h|u - \mathcal{I}_p u|_{1,t}.$$

The lemma now follows with $C(u) = \beta_0 C_3(u) + \hat{C}_3(u)$. □

Combining (2), (4) and (5) we have

$$\beta_0 \le \frac{|u - \hat{\mathcal{I}}_p u|_{1,t}}{|u - \mathcal{I}_p u|_{1,t}} \le \beta \le \beta_0 + Ch$$

for $u \in H^{p+2}(t)$, $u \notin \mathcal{S}_p$, and satisfying (1). If $\beta_0 < 1$, this shows the saturation assumption holds for h sufficiently small, with β approaching β_0 with decreasing h.

3. Some h-Refinement Examples. In this section we estimate β_0 of (4) for some typical h-refinement algorithms. While it is easy to see that β_0 does not depend on the size of the element, as an H^1 seminorm, it does depend on its shape and orientation. This limits the types of h-refinement schemes that can be easily analyzed. The easiest case is one where the refined elements have the same shape and orientation as the original element. Then the impact of shape is the same in both $|(1-\hat{\mathcal{I}}_p)\nu|_1$ and $|(1 - \mathcal{I}_p)\nu|_1$, and for these cases we are able to make an exact calculation for β_0. We are also able to provide some analysis in situations where the refinement process of a given element generates elements that are members of a small number of geometric congruence classes. In these situations, instead of comparing the error in element t to that of the children of t, we compare the error in the child elements to that of the next generation (grandchildren). In particular, $\mathcal{I}_p$ will refer to interpolation on the child elements and $\hat{\mathcal{I}}_p$ to interpolation on the h-refined child elements. Also, we pick just one child element from each congruence class, and the domain of $|\cdot|_1$ is integration over just this set of child elements and not all of t, LEMMA 2. *For any given polynomial $v \in \mathcal{S}_{p+1}$ and all elements t*

$$|v - \mathcal{I}_p v|_{1,t} = c_0\, h^{p+d/2}$$

holds, where the constant c_0 depends on v and on the shape and orientation of the element t, but not on the diameter h of t, and d is the space dimension.

Proof. The H^1-seminorm does not depend on the position of element t in space. Thus we fix a reference element $\widetilde{t}$ of given shape and orientation with diameter $h = 1$ and one vertex at the origin. Without loss, we restrict analysis to the elements t consisting of the points hx, $x \in \widetilde{t}$. Let hx_k, $k = 1, \ldots, n$, be the interpolation nodes for $\mathcal{I}_p$, and let polynomials ϕ_k of order p attain the values $\phi_k(x_\ell) = \delta_{k\ell}$. The interpolant $\mathcal{I}_p v$ of a function v is then

$$(\mathcal{I}_p v)(x) = \sum_{k=1}^{n} v(hx_k)\phi_k\Big(\frac{x}{h}\Big).$$

Let $v \in \mathcal{S}_{p+1}$ with Taylor representation

$$v(x) = \sum_{|\alpha| \le p+1} \frac{(\partial^\alpha v)(0)}{\alpha!}\, x^\alpha.$$

As $\mathcal{I}_p$ reproduces polynomials in $\mathcal{S}_p$, one obtains then the error representation

$$(v - \mathcal{I}_p v)(x) = h^{p+1} \sum_{|\alpha| = p+1} \frac{(\partial^\alpha v)(0)}{\alpha!}\, \psi_\alpha\Big(\frac{x}{h}\Big),$$

with the polynomials

$$\psi_\alpha(x) = x^\alpha - \sum_{k=1}^{n} x_k^\alpha\, \phi_k(x)$$

of order $p + 1$. The squared H^1-error is therefore

$$|v - \mathcal{I}_p v|_{1,t}^2 = h^{2p} \int_t \Big| \sum_{|\alpha| = p+1} \frac{(\partial^\alpha v)(0)}{\alpha!}\, (\nabla \psi_\alpha)\Big(\frac{x}{h}\Big) \Big|^2 dx,$$

or, after transformation of the integral to our reference element $\widetilde{t}$,

$$|v - \mathcal{I}_p v|_{1,t}^2 = c_0^2 h^{2p+d},$$

with the constant c given by

$$c_0^2 = \int_{\widetilde{t}} \Big| \sum_{|\alpha| = p+1} \frac{(\partial^\alpha v)(0)}{\alpha!}\, (\nabla \psi_\alpha)(x) \Big|^2 dx,$$

and d the space dimension. □

Given reference element $\tilde{t}$, we define a second reference element $\bar{t}$ by reflection of $\tilde{t}$ about all coordinate axes. Then without loss, we restrict attention to elements t consisting of points $-hx$, $x \in \bar{t}$. We then reprise the proof of Lemma 2, with a few systematic but minor changes in sign to show

Corollary 3. *Element t and its reflection $\hat{t}$ about all coordinate axes have the same value of c_0 in Lemma 2 for the same polynomial $v \in \mathcal{S}_{p+1}$.*

We now analyze some specific examples of h-refinement. We first, consider Lagrange interpolation in one space dimension. The refinement is simple bisection. Lemma 2 can be applied to both $|(1-\mathcal{I})\nu|_1$ and $|(1-\hat{\mathcal{I}})\nu|_1$ yielding the same value of c_0 for both t of size h and its two child elements of size $h/2$. Thus

$$|(1-\hat{\mathcal{I}}_p)\nu|_1^2 = 4^{-p}|(1-\mathcal{I}_p)\nu|_1^2$$

and

$$\beta_0 = 2^{-p}.$$

Next consider the case of Lagrange polynomials on triangular elements. We first consider the case of regular (red) refinement, illustrated in Figure 1 (left). In this case all child elements are similar to the original triangle t, but with size $h/2$. The three child elements sharing a vertex with the parent element have the same orientation as the parent, while the the center element has the reflective orientation described in Corollary 3 (in two dimensions, equivalent to rotation by π).

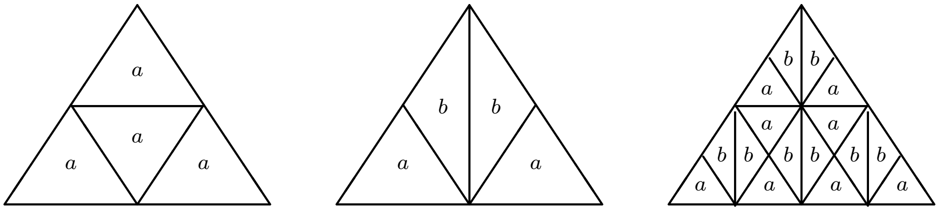

Fig. 1. *Left: regular refinement. Center: two levels of newest node bisection. Right: four levels of newest node bisection. Triangles with similar geometry are labeled a and b.*

Let T_a be the set of four refined elements. Then Lemma 2 and Corollary 3 show that

$$|(1-\hat{\mathcal{I}}_p)\nu|_1^2 \equiv \sum_{t\in T_a} |(1-\hat{\mathcal{I}}_p)\nu|_{1,t}^2 = 4^{-p}|(1-\mathcal{I}_p)\nu|_1^2.$$

Thus

$$\beta_0 = 2^{-p}.$$

We now consider an alternative refinement, illustrated in the center of Figure 1. This refinement pattern could arise from two levels of uniform refinement using newest node bisection. This refinement pattern generates two congruence classes of child elements, labeled a and b in Figure 1. To analyze this situation, we refine the four child elements using the same scheme, resulting in 16 elements, eight in each congruence class. Each class of elements contains elements with two orientations, differing by reflection about both coordinate axes as in Corollary 3.

We now let $|\cdot|_1$ refer to a pair of the child elements in the center image in Figure 1, one of class a (t_a) and one of class b (t_b), that share a common edge. Together, t_a and t_b have eight child elements, four of each class. Let T_a be the set of the four refined elements in congruence class a; T_b is defined analogously. As before, we consider the case of Lagrange elements of degree p. Since both t_a and t_b are children of the same parent element t, they share the same values for the constant partial derivatives $\partial^\alpha \nu$, $|\alpha| = p+1$. Using the same technique as above in the case of regular refinement, we obtain the pair of estimates

$$\begin{aligned} |(1-\hat{\mathcal{I}}_p)\nu|^2_{1,t_a} &= 4^{-p} \sum_{t \in T_a} |(1-\mathcal{I}_p)\nu|^2_{1,t} \\ |(1-\hat{\mathcal{I}}_p)\nu|^2_{1,t_b} &= 4^{-p} \sum_{t \in T_b} |(1-\mathcal{I}_p)\nu|^2_{1,t}. \end{aligned}$$

We can then combine these estimates as

$$|(1-\hat{\mathcal{I}}_p)\nu|^2_1 = 4^{-p} \sum_{t \in T_a \cup T_b} |(1-\mathcal{I}_p)\nu|^2_{1,t} = 4^{-p} |(1-\mathcal{I}_p)\nu|^2_1$$

that immediately implies

$$\beta_0 = 2^{-p}.$$

As our final example, we consider regular refinement of tetrahedral elements. Now there are generally three congruence classes denoted T_i, $1 \le i \le 3$. See Bey [6] for discussion of this point. If $t \in T_i$ is regularly refined, the child elements consist of four elements in T_i and two elements in each of the other two classes. The four elements in class T_i share a vertex with the parent and have the same orientation as the parent. Each pair of elements in the other two classes contain one element that has the reflective orientation of the other. If t is uniformly refined many times, the distribution of child elements grows as powers of the matrix

$$\begin{pmatrix} 4 & 2 & 2 \\ 2 & 4 & 2 \\ 2 & 2 & 4 \end{pmatrix}.$$

This matrix has one eigenvalue $\lambda = 8$ (eigenvector $(1,1,1)^t$) and two eigenvalues $\lambda = 2$. Thus whatever congruence class the original element, the child elements rapidly attain an approximately equal distribution among the three classes. After k levels of uniform refinement, the distribution of the 8^k elements has $(8^k - 2^k)/3$ triplets consisting of one element of each class and 2^k unmatched elements similar to the original. Each of the three classes contain elements with one of two possible orientations, the second the reflection of the first about all three coordinate axes.

We treat this case in a fashion similar to the two dimensional case above. In particular choose three elements resulting from the original refinement, one from each class, each sharing at least one face with another element of the trio. Let $|\cdot|_1$ refer to the three elements. Each element is refined using regular refinement, resulting in 24 child elements, eight of each class. Estimate the decrease in error using the same strategy as above in the two dimensional case, resulting in the estimate

$$\beta_0 = 2^{-p}.$$

4. Variable Coefficients. We consider briefly the case of variable coefficients. Let

$$a(u,v)_t = \int_t a(x) \nabla u \cdot \nabla v \, dx$$

where $a \in W^1_\infty(t)$ and $a(x) > a_0 > 0$ in t. The seminorm $|u|^2_{a,t} = a(u,u)_t$ is comparable to the $H^1(t)$ seminorm

$$\hat{c}_1 |u|_{a,t} \le |u|_{1,t} \le \hat{c}_2 |u|_{a,t} \tag{6}$$

for positive constants $\hat{c}_1$ and $\hat{c}_2$.

The constant β_0 is now defined as

$$\beta_0 \equiv \max_{\nu \in \mathcal{S}_{p+1}} \frac{|\nu - \hat{\mathcal{I}}_p \nu|_{a,t}}{|\nu - \mathcal{I}_p \nu|_{a,t}}. \tag{7}$$

β_0 can now depend on the coefficient a but retains other properties of the constant coefficient case.

Lemma 4. *Let $u \in H^{p+2}(t)$ satisfy (1), $u \notin \mathcal{S}_p$, and β_0 be given by (7). Then there is a constant C depending on u and a, the degree p, the shape of element t, but not on its diameter h, such that*

$$\frac{|u - \hat{\mathcal{I}}_p u|_{a,t}}{|u - \mathcal{I}_p u|_{a,t}} \le \beta_0 + Ch. \tag{8}$$

Proof. The proof follows the pattern of proof of Lemma 1. As before, using (7) and the triangle inequality

$$|u - \hat{\mathcal{I}}_p u|_{a,t} \le \beta_0 |u - \mathcal{I}_p u|_{a,t} + \beta_0 |(1 - \mathcal{I}_p)(u - \mathcal{I}_{p+1} u)|_{a,t} + |(1 - \hat{\mathcal{I}}_p)(u - \mathcal{I}_{p+1} u)|_{a,t}.$$

Using (6) and following the proof of Lemma 1, we have

$$\hat{c}_1 |(1 - \mathcal{I}_p)(u - \mathcal{I}_{p+1} u)|_{a,t} \le |(1 - \mathcal{I}_p)(u - \mathcal{I}_{p+1} u)|_{1,t} \le C_3 h |u - \mathcal{I} u|_{1,t} \le \hat{c}_2 C_3 h |u - \mathcal{I} u|_{a,t}$$

and

$$\hat{c}_1 |(1 - \hat{\mathcal{I}}_p)(u - \mathcal{I}_{p+1} u)|_{a,t} \le \hat{c}_2 \hat{C}_3 h |u - \mathcal{I}_p u|_{a,t}$$

for functions $u \in H^{p+2}$. The lemma follows with $C = \hat{c}_2 (\beta_0 C_3 + \hat{C}_3)/\hat{c}_1$. ☐

As before, $\beta_0 = 0$ for the case of p-refinement. For the case of h-refinement, Let

$$\hat{a} = \sqrt{\max_t a \cdot \min_t a}.$$

Since $a \in W^1_\infty(t)$

$$\|a/\hat{a}\|_{L_\infty(t)} = \|\hat{a}/a\|_{L_\infty(t)} \le 1 + C_4 h.$$

We first consider the one dimensional case. We assume $\beta_0 = 2^{-p}$ for the constant coefficient case $a \equiv 1$. Then

$$\begin{aligned}
|(1 - \hat{\mathcal{I}}_p) v|_{a,t} &\le \|a/\hat{a}\|^{1/2}_{L_\infty(t)} \; |(1 - \hat{\mathcal{I}}_p) v|_{1,t} \\
&\le \sqrt{1 + C_4 h} \; |(1 - \hat{\mathcal{I}}_p) v|_{1,t} \\
&= \frac{\sqrt{1 + C_4 h}}{2^p} \; |(1 - \mathcal{I}_p) v|_{1,t} \\
&\le \frac{1 + C_4 h}{2^p} \; |(1 - \mathcal{I}_p) v|_{a,t}
\end{aligned}$$

that implies

$$\beta_0 \le (1 + C_4 h) 2^{-p}. \tag{9}$$

A similar perturbation argument can be applied to the two and three dimensional examples of Section 3, all yielding a result similar to (9). Thus

$$\beta \le \beta_0 + Ch \le 2^{-p} + \hat{C}(a, u) h$$

an asymptotic result similar to the constant coefficient case.

5. Low Regularity and Other Exceptional Cases. We first consider the issue of regularity. Many pde solutions do not have the very high regularity assumed in Lemmas 1 and 4. It is straightforward to weaken the assumption in these Lemmas to $u \in H^{p+1+\alpha}(t)$ for some $0 < \alpha < 1$, leading to estimates of the form

$$\beta \leq \beta_0 + Ch^{\alpha}.$$

However, such estimates still require better than $H^2(t)$ regularity. Often pdes have point singularities with regularity $H^{1+\alpha}$, $0 < \alpha < 1$, that are not addressed by our Lemmas. In such cases, we still expect most elements in the mesh have higher local regularity and are covered by our estimates. It is mainly those elements that contain the singular point, often as a vertex, where our estimates cannot be applied.

We now analyze the case of h-refinement, $p = 1$ in $d = 1$ dimension, and without loss of generality, we assume the element $t = (0, h)$. This case is unique because the piecewise linear finite element solution of the two-point Dirichlet boundary value problem $-u'' = f$ on $(0, h)$ is exact at the vertices, and thus the finite element solution is the piecewise linear interpolant. Thus the interpolant $\hat{\mathcal{I}}_1 u$ satisfies the usual Galerkin orthogonality and best approximation properties associated with the finite element solution. In particular, Galerkin orthogonality implies

$$\int_0^h (u - \hat{\mathcal{I}}_1 u)'(\mathcal{I}_1 u - \hat{\mathcal{I}}_1 u)' \, dx = 0$$

leading to

$$|u - \mathcal{I}_1 u|_{1,t}^2 = |u - \hat{\mathcal{I}}_1 u|_{1,t}^2 + |\mathcal{I}_1 u - \hat{\mathcal{I}}_1 u|_{1,t}^2.$$

Assuming both $|\mathcal{I}_1 u - \hat{\mathcal{I}}_1 u|_{1,t}$ and $|u - \mathcal{I}_1 u|_{1,t}$ are nonzero, we have

$$\beta^2 = 1 - \frac{|\mathcal{I}_1 u - \hat{\mathcal{I}}_1 u|_{1,t}^2}{|u - \mathcal{I}_1 u|_{1,t}^2} < 1. \tag{10}$$

This suggests that the saturation property holds even in cases of low regularity. Additional assumptions provide more quantitative estimates for β. Thus we examine a few special cases. Let

$$\begin{aligned} D_1 &= \frac{u(h) - u(0)}{h} = (\mathcal{I}_1 u)', \\ D_2 &= \frac{-u(h) + 2u(h/2) - u(0)}{(h/2)^2}. \end{aligned}$$

Then

$$\begin{aligned} |u - \mathcal{I}_1 u|_{1,t}^2 &= \int_0^h (u')^2 \, dx - hD_1^2, \\ |\mathcal{I}_1 u - \hat{\mathcal{I}}_1 u|_{1,t}^2 &= 4h^3 D_2^2. \end{aligned}$$

Combining these results

$$\beta^2 = 1 - \frac{4h^3 D_2^2}{\int_0^h (u')^2 \, dx - hD_1^2}. \tag{11}$$

We make formal series expansions of these terms centered at the point $h/2$.

$$\begin{aligned} 4h^3 D_2^2 &= \frac{h^3}{16}\left(u''(h/2)^2 + O(h^2)\right), \\ \int_0^h (u'(x))^2 - hD_1^2 &= \frac{h^3}{12}\left(u''(h/2)^2 + O(h^2)\right). \end{aligned}$$

Thus

$$\beta = \left(1 - \frac{3}{4} + 0(h^2)\right)^{1/2} \leq \frac{1}{2} + O(h),$$

as predicted by our analysis.

A second special case is the computation of β for a function that does not satisfy the minimal regularity assumption in Lemma 1. Let $u = x^\alpha$ for $1/2 < \alpha < 1$. Then

$$\begin{aligned} |u - \mathcal{I}_1 u|_{1,t}^2 &= \frac{\alpha^2 h^{2\alpha-1}}{2\alpha - 1} - h^{2\alpha-1} = h^{2\alpha-1}(\alpha-1)^2/(2\alpha-1), \\ |\mathcal{I}_1 u - \hat{\mathcal{I}}_1 u|_{1,t}^2 &= h^{2\alpha-1}(1 - 2^{1-\alpha})^2, \end{aligned}$$

and

$$\beta^2 \equiv \beta_\alpha^2 = 1 - \left(\frac{1 - 2^{1-\alpha}}{\alpha - 1}\right)^2 (2\alpha - 1).$$

In this case $\beta_\alpha < 1$ is independent of h but depends strongly on α. In some sense β_α plays a role similar to β_0 when x^α is the dominant part of a more general function. We explore this in more detail below.

We next consider the case of p-refinement for $p = 1$, $d = 1$ on the interval $(0, h)$. We develop the interpolant $\mathcal{I}_2$ in the hierarchical basis. Let

$$\mathcal{I}_2 u = \mathcal{I}_1 u + \psi = \mathcal{I}_1 u + \frac{(h - x)x}{2} D_2.$$

To estimate β, we begin with

$$\begin{aligned} \int_0^h (u' - (\mathcal{I}_2 u)')^2 &= \int_0^h (u' - (\mathcal{I}_1 u)' + \psi')^2 \\ &= \int_0^h (u' - (\mathcal{I}_1 u)')^2 - 2(u' - (\mathcal{I}_1 u)')\psi' + (\psi')^2 \\ &= \int_0^2 (u')^2\, dx - (D_1)^2 h + D_2 \left(2 \int_0^h xu' \ dx - h^2 D_1 + h^3 D_2/12\right). \end{aligned}$$

As before

$$\int_0^h (u' - (\mathcal{I}_1 u)')^2 = \int_0^h (u')^2\, dx - hD_1^2,$$

and

$$\beta^2 = 1 + \frac{D_2 \left(2 \int_0^h xu' \ dx - h^2 D_1 + h^3 D_2/12\right)}{\int_0^2 (u')^2\, dx - (D_1)^2 h}.$$

If we make formal series expansions centered and $h/2$, we see

$$\begin{aligned}
\beta^2 &= 1 + \frac{D_2 \left(2 \int_0^h x u' \, dx - h^2 D_1 + h^3 D_2/12\right)}{\int_0^2 (u')^2 \, dx - (D_1)^2 h} \\
&= 1 - \frac{u''(h/2)^2 h^3/6 - u''(h,2)^2 h^3/12 + O(h^5)}{u''(h/2)^2 h^3/12 + O(h^5)} \\
&= 1 - (1 + O(h^2)) \\
&= O(h^2),
\end{aligned}$$

and $\beta = O(h)$ as expected.

For the singular case, let $u = x^\alpha$ for $1/2 < \alpha < 1$. Then

$$\beta^2 \equiv \beta_\alpha^2 = 1 + \frac{4(2\alpha - 1)(2^{1-\alpha} - 1)}{(\alpha - 1)^2} \left(\frac{\alpha - 1}{\alpha + 1} + \frac{(2^{1-\alpha} - 1)}{3} \right) < 1.$$

We see that β_α is a positive constant as in the h-refinement case. This is consistent with the principle that h-refinement is preferable at singular points. First, in the case of h-refinement, child elements not containing the singular vertex have improved local regularity and become covered by Lemma 1. Second, child elements containing the singular vertex become smaller, and since the singular function is of size $O(h^\alpha)$ its impact on the global solution also becomes smaller.

We expect this behavior to extend to cases $p \geq 1$ and $d \geq 1$ dimensions, although direct calculation becomes much more challenging. However, in the next section we will numerically verify that one does see similar behavior for $1 \leq p \leq 4$ and $1 \leq d \leq 2$. Finally, we remark there are other possible sources for reduced local regularity, for example insufficient local smoothness in the coefficient function $a(x)$.

Since our estimates for β are both asymptotic and local, it is also possible to find functions that have sufficient regularity but may not exhibit the expected behavior. As a first example, it is simple to find functions where h-refinement or p-refinement result in no reduction of the interpolation error. In our example of $p = 1$ and $d = 1$, any smooth function $u \notin \mathcal{S}_1$ with $D_2 = 0$ in element t will have $\beta = 1$ for both h and p refinement. For such functions, one can expect this behavior to be transient and most likely to occur on coarse meshes, although it could occur in other scenarios as well.

A more interesting example is $u = x^\alpha$ with $u \in H^{p+2}(t)$ and $u \notin \mathcal{S}_p$. For our example case of $p = 1$, $d = 1$, and h-refinement,

$$\beta^2 \equiv \beta_\alpha^2 = 1 - \left(\frac{1 - 2^{1-\alpha}}{\alpha - 1} \right)^2 (2\alpha - 1).$$

using the same computations as in the singular case above. In our numerical experiments, we observe similar behavior $\beta \equiv \beta_{\alpha,p} < 1$ for both h and p refinement for $p \geq 1$. Such functions have an interesting impact on the assumption of local a priori error estimates. In particular $|u|_{k,t} = \tilde{c}_k h^{d/2+\alpha-k}$ and $|u - \mathcal{I}_p u|_{k,t} = c_k h^{d/2+\alpha-k}$ so the a priori estimate

$$|u - \mathcal{I}_p u|_{m,t} \leq C h^{k-m} |u|_{k,t}$$

becomes

$$c_m \leq C \tilde{c}_k.$$

While the local a priori estimate is formally true, and shows that the interpolation error approaches zero as $h \to 0$, the convergence is a mixture of the solution and relevant derivatives approaching zero coupled with approximation properties. Indeed, the usual interpretation of (1) concerning rates of convergence is problematic in such cases. Here $\beta < 1$ provides more meaningful information about the local convergence behavior, as it indicates the local relative error reduction for a single refinement step separately from the convergence behavior of the function. Also, note that for the case $p = 1$, $d = 1$, and h-refinement, we have $\beta_\alpha \to 1$ as $\alpha \to \infty$, but this is balanced by $x^\alpha \to 0$ as $\alpha \to \infty$ (provided $h < 1$). We expect this to remain true for $p \geq 1$ and both h-refinement and p-refinement.

Let $d > 1$ and $u = x^\xi = \sum_{k=1}^d x_k^{\xi_k}$ where the multi index $|\xi| = \alpha$ on a shape regular simplicial element of size h with one vertex at the origin. We expect such functions to exhibit similar behavior with respect to a priori estimates as the case $d = 1$. Functions composed as a linear combination of such terms all with multi indices size α should also behave this way, as will functions such as r^α where $r = (\sum_{k=1}^d x_k^2)^{1/2}$. If u contains such a function as the dominant term then $\beta_{\alpha,p}$ will behave in a role similar to β_0 in Lemma 1.

To see this, let $u = v + w \in H^{1+\gamma}$ on a shape regular simplicial element t of size h in $\mathcal{R}^d$ with one vertex at the origin. The function v is the dominant term and has the exceptional property

$$\begin{aligned} |v|_{k,t} &= \tilde{c}_k(v) h^{d/2+\alpha-k}, \\ |v - \mathcal{I}_p v|_{k,t} &= c_k(v) h^{d/2+\alpha-k}, \\ |v - \hat{\mathcal{I}}_p v|_{k,t} &= \hat{c}_k(v) h^{d/2+\alpha-k}, \end{aligned}$$

for $0 \leq k \leq 1 + \gamma$. It is possible $\gamma < 1$ and v is a singular function but it is also possible that γ could be quite large. We assume $v \notin S_p$ and compute

$$\beta_{\alpha,p} = \frac{|v - \hat{\mathcal{I}}_p v|_{1,t}}{|v - \mathcal{I}_p v|_{1,t}} = \frac{\hat{c}_1}{c_1}$$

in analogy to β_0.

The function w is smoother than v, although we can allow it to be a singular function but with a weaker singularity. All we require is

$$|w - \mathcal{I}_p w|_{1,t} \leq C_1(w) h^{d/2+\xi-1}, \tag{12}$$

$$|w - \hat{\mathcal{I}}_p w|_{1,t} \leq \hat{C}_1(w) h^{d/2+\xi-1}, \tag{13}$$

with $\xi > \alpha$.

Lemma 5. *Let $u = v + w \in H^{1+\gamma}$ on a shape regular simplicial element t of size h in $\mathcal{R}^d$ with one vertex at the origin, and let $v \notin S_p$, w, and $\beta_{\alpha,p}$ be defined as above. Then for h sufficiently small*

$$\frac{|u - \hat{\mathcal{I}}_p u|_{1,t}}{|u - \mathcal{I}_p u|_{1,t}} \leq \beta_{\alpha,p} + C h^{\xi-\alpha}.$$

Proof. The proof here follows the pattern of our proof of Lemma 1.

$$\begin{aligned} |u - \hat{\mathcal{I}}_p u|_{1,t} &\leq |(1 - \hat{\mathcal{I}}_p) v|_{1,t} + |(1 - \hat{\mathcal{I}}_p) w|_{1,t} \\ &\leq \beta_{\alpha,p} |(1 - \mathcal{I}_p) v|_{1,t} + |(1 - \hat{\mathcal{I}}_p) w|_{1,t} \\ &\leq \beta_{\alpha,p} |u - \mathcal{I}_p u|_{1,t} + \beta_{\alpha,p} |(1 - \mathcal{I}_p) w|_{1,t} + |(1 - \hat{\mathcal{I}}_p) w|_{1,t}. \end{aligned}$$

Since $v \notin S_p$

$$|v - \mathcal{I}_p v|_{1,t} = c_1(v) h^{d/2+\alpha-1} > 0.$$

Then using this estimate and (12)

$$\begin{aligned} |(1-\mathcal{I}_p)w|_{1,t} &\le C_1(w) h^{d/2+\xi-1} \\ &\le \left(\frac{C_1(w)}{c_1(v)}\right) h^{\xi-\alpha} |(1-\mathcal{I}_p)v|_{1,t} \\ &\le C_2 h^{\xi-\alpha} \left(|(1-\mathcal{I}_p)u|_{1,t} + |(1-\mathcal{I}_p)w|_{1,t}\right). \end{aligned}$$

For h sufficiently small,

$$|(1-\mathcal{I}_p)w|_{1,t} \le \frac{C_2}{1 - C_2 h^{\xi-\alpha}} h^{\xi-\alpha} |(1-\mathcal{I}_p)u|_{1,t} \equiv C_3 h^{\xi-\alpha} |(1-\mathcal{I}_p)u|_{1,t}.$$

A similar argument using (13) shows

$$|(1-\hat{\mathcal{I}}_p)w|_{1,t} \le \hat{C}_3 h^{\xi-\alpha} |(1-\mathcal{I}_p)u|_{1,t}.$$

The lemma now follows with $C(u) = \beta_{\alpha,p} C_3 + \hat{C}_3$. $\square$

While we have been emphasizing the similarities between Lemma 5 and Lemma 1, it is equally important to emphasize the differences. While the calculation of β_0 requires stronger assumptions, it achieves the important advantage of being independent of the function u. Thus Lemma 1 describes a very broad and general case, while Lemma 5 is more narrowly focused on a particular singularity or exceptional case. In a typical pde, one is likely to find only a finite number of singular and exceptional points. A finite number of elements contain these points and the remainder should be covered by Lemma 1. As the (shape regular) mesh is refined, the number of singular and exceptional elements remains bounded, while the number covered by Lemma 1 grows. Thus the exceptional and singular elements should have a decreasing impact on the overall error in the finite element solution.

To look at this from a different point of view, one might wonder if the exceptional cases could adversely effect the convergence of global adaptive refinement algorithms based on local error indicators and feedback loops. To address this point, we begin with a shape regular triangulation consisting of simplicial elements and corresponding finite element space S_h^0 on a domain Ω in $\mathcal{R}^d$ for $1 \le d \le 3$. We assume a local a priori estimate holds for each element in the initial domain. The local smoothness could be different in each element, but we assume at least $u \in H^{1+\alpha}(t)$ for $\alpha > 0$. For convenience we assume $h_t \le h_0 < 1$ for all elements in the initial mesh, and the local degree p in $\mathcal{I}_p u$ could be different in each element. Because interpolation is local, we allow S_h^0 and the adaptive sequence of refined finite element spaces $\{S_h^k\}$, $k \ge 1$ to be nonconforming. The sequence $\{S_h^k\}$ is generated as follows. Given S_h^k, we choose an element with the largest interpolation error and refine it, creating the space S_h^{k+1}. The refinement could be h-refinement or p-refinement. If the initial regularity violates the assumption of Lemma 1 or p-refinement would cause such a violation in the child element, then h-refinement is required. In either case, we require the new elements to be shape regular and satisfy an a priori estimate (1) with equal or improved regularity compared to the parent. In the case that both h- and p-refinement are allowed, we prefer an option providing the largest error reduction. This algorithm is similar to the so-called reference adaptive procedure analyzed in [5].

Let $u_k \in S_h^k$ be the (possibly discontinuous) piecewise polynomial interpolant of u generated in the k-th step of this process.

Lemma 6. *Assume that the sequence $\{u_k \in S_h^k\}$ is generated as described above, and that (1) holds for all elements for all k. Then*

$$\lim_{k\to\infty} |u - u_k|_{1,\Omega} = 0. \tag{14}$$

Proof. The proof is by the method of contradiction. If the adaptive procedure did not converge, there must be one or more elements with maximum positive error that was not reduced even after multiple (infinite) refinement steps. This contradicts the a priori error estimate (1). ∎

Thus if we have only a priori error estimates, not necessarily a local saturation property, that is sufficient to (abstractly) prove convergence of our reference adaptive scheme. More detailed arguments in [5] show similar h and hp reference adaptive refinement schemes exhibit optimal rates of convergence. From a practical point of view, one can construct scenarios where refinement of some element (either h- or p-refinement) does not reduce the error. However, unless the error is already zero, such constructions must be transient (finite) in nature, or else the a priori convergence estimates could not hold. Thus a finite number of refinements of that element must yield error reduction by a fixed factor $\beta < 1$. Overall, such scenarios seem most likely to occur for specially chosen functions on specially chosen initial coarse meshes, although one can imagine such events might occur later in the adaptive procedure. Nonetheless the local saturation assumption (2) should hold asymptotically, even if not demonstrated by Lemmas 1, 4, or 5 and the accumulated costs of all the violations should remain bounded.

We conclude with a remark. Consider the model self-adjoint elliptic boundary value problem: find $u \in H_0^1(\Omega)$ such that

$$a(u,v) \equiv \int_\Omega a\nabla u \cdot \nabla v + buv\, dx = (f,v)$$

for all $v \in H_0^1(\Omega)$, where $a(x) \ge a_0 > 0$, $b(x) \ge 0$ are smooth and $\Omega \subset R^d$. Let $|||u|||^2 = a(u,u)$ denote the (global) energy norm. Let $S_h \subset \hat{S}_h \subset H_0^1(\Omega)$ be two finite element spaces corresponding to shape regular (but not necessarily quasi-uniform) triangulations of Ω. We assume $\hat{S}_h$ was created by refinement (h, p, or hp) of S_h.

Let $u_h \in S_h$ solve

$$a(u_h, v) = (f, v)$$

for all $v \in S_h$ and let $\hat{u}_h \in \hat{S}_h$ solve

$$a(\hat{u}_h, v) = (f, v)$$

for all $v \in \hat{S}_h$. Both of these problems have Galerkin orthogonality and best approximation properties. Since $S_h \subset \hat{S}_h$, we have

$$a(u - \hat{u}_h, \hat{u}_h - u_h) = 0$$

leading to

$$|||u - u_h|||^2 = |||u - \hat{u}_h|||^2 + |||u_h - \hat{u}_h|||^2.$$

If $|||u_h - \hat{u}_h|||$ and $|||u - u_h|||$ are nonzero,

$$\beta^2 = 1 - \frac{|||u_h - \hat{u}_h|||^2}{|||u - u_h|||^2} < 1.$$

This is a global rather than local saturation property, applying to finite element solutions rather than interpolants, but like a previous example above, it does not depend on a strong regularity assumption.

6. Numerical Illustrations. In this section we present simple examples with $d = 1$ and $d = 2$ that illustrate some of our results. The computations for $d = 1$ use the interval $(0, h)$, and those for $d = 2$ use triangle t with vertices $(0, 0)$, $(h, 0)$, and $(0, h)$. In each case, we compute the value of β for six values of h, $h = 2^{-k}$ for $2 \leq k \leq 7$, and four values of p, $1 \leq p \leq 4$. As usual, for $d = 1$ h-refinement consists of bisecting the given interval. For $d = 2$ we use regular (red) refinement. In both cases, p-refinement consists of increasing the polynomial degree by one. All integrals are approximated using numerical quadrature. To control the discretization error, for $d = 1$, the interval $(0, h)$ is partitioned into 50000 subintervals and a 12-point, order 24 Gaussian quadrature rule is applied on each subinterval. For $d = 2$ we partition the triangle t into 2500 triangles similar to t and a 91-point order 22 quadrature rule is applied to each subelement. To minimize the effects of round-off, all calculations are done using quadruple precision arithmetic. However, the node locations and weights of the quadrature formulae are known only to double precision, so we expect those errors will dominate the round-off error behavior.

h	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 1$	$p = 2$	$p = 3$	$p = 4$
	h-refinement, $d = 1$				h-refinement, $d = 2$			
$h = 1/4$	.5190	.2826	.1588	.0910	.5204	.2555	.1509	.0807
$h = 1/8$	.5067	.2617	.1373	.0730	.5111	.2505	.1373	.0702
$h = 1/16$	.5021	.2536	.1288	.0658	.5058	.2494	.1309	.0658
$h = 1/32$	.5006	.2510	.1261	.0634	.5030	.2494	.1279	.0640
$h = 1/64$	.5002	.2503	.1253	.0627	.5015	.2496	.1264	.0632
$h = 1/128$	.5000	.2501	.1251	.0626	.5008	.2498	.1257	.0628
	p-refinement, $d = 1$				p-refinement, $d = 2$			
$h = 1/4$	.1185	.1379	.1549	.1696	.0733	.1380	.1373	.1654
$h = 1/8$	.0696	.0810	.0909	.0996	.0411	.0852	.0797	.1002
$h = 1/16$	.0384	.0446	.0501	.0549	.0218	.0484	.0431	.0562
$h = 1/32$	.0203	.0236	.0264	.0290	.0112	.0259	.0224	.0299
$h = 1/64$	.0104	.0121	.0136	.0149	.0057	.0135	.0114	.0154
$h = 1/128$	.0053	.0062	.0069	.0076	.0029	.0069	.0058	.0079

Table 1

The value of β for several values of h and p for the functions $u = (x + 1/4)^{2/3}$ on the interval $(0, h)$ and $u = ((x + 1/4)^2 + (y + 1/4)^2)^{5/8}$ on the triangle with vertices $(0, 0)$, $(h, 0)$, and $(0, h)$.

In our first experiment, we illustrate the behavior described in Lemma 1. For $d = 1$, we use the function $u = (x + 1/4)^{2/3}$. This function has a singularity at $x = -1/4$ but satisfies the assumptions of Lemma 1 on $(0, h)$. For $d = 2$, we use the function $((x + 1/4)^2 + (y + 1/4)^2)^{5/8}$. This function has a singular point near but not within the triangle of interest. The results are given in Table 1. In the case of h-refinement we observe the convergence $\beta \to 2^{-p}$ predicted by the bounds $2^{-p} \leq \beta \leq 2^{-p} + Ch$. In the case of p-refinement we see $\beta \to 0$ predicted by the bounds $0 \leq \beta \leq Ch$.

For our second experiment, we consider functions with a point singularity at the origin and compute $\beta = \beta_{\alpha,p}$, as u contains only the dominant singular term. For $d = 1$, we chose $u = x^{2/3}$ and for $d = 2$ we chose $u = (x^2 + y^2)^{5/8} \equiv r^{5/4}$. The results are shown in Table 2. For both h-refinement and p-refinement, $\beta_{\alpha,p}$ appears to be a constant for each value of p and is independent of h as predicted in our

h	$p=1$	$p=2$	$p=3$	$p=4$	$p=1$	$p=2$	$p=3$	$p=4$
	h-refinement, $d=1$				h-refinement, $d=2$			
$h=1/4$	.8930	.8909	.8909	.8909	.6106	.4544	.4398	.4254
$h=1/8$	.8930	.8909	.8909	.8909	.6106	.4544	.4398	.4254
$h=1/16$	.8930	.8909	.8909	.8909	.6106	.4544	.4398	.4254
$h=1/32$	.8930	.8909	.8909	.8909	.6106	.4544	.4398	.4254
$h=1/64$	.8930	.8909	.8909	.8909	.6106	.4544	.4398	.4254
$h=1/128$	.8930	.8909	.8909	.8909	.6106	.4544	.4398	.4254
	p-refinement, $d=1$				p-refinement, $d=2$			
$h=1/4$	.8013	.8835	.9222	.9423	.2904	.4687	.5433	.6167
$h=1/8$	.8013	.8835	.9222	.9423	.2904	.4687	.5433	.6167
$h=1/16$	.8013	.8835	.9222	.9423	.2904	.4687	.5433	.6167
$h=1/32$	.8013	.8835	.9222	.9423	.2904	.4687	.5433	.6167
$h=1/64$	.8013	.8835	.9222	.9423	.2904	.4687	.5433	.6167
$h=1/128$	.8013	.8835	.9222	.9423	.2904	.4687	.5433	.6167

Table 2

The value of $\beta_{\alpha,p}$ *for several values of* h *and* p *for the functions* $u = x^{2/3}$ *on the interval* $(0,h)$ *and* $u = (x^2+y^2)^{5/8}$ *on the triangle with vertices* $(0,0)$, $(h,0)$, *and* $(0,h)$.

h	$p=1$	$p=2$	$p=3$	$p=4$	$p=1$	$p=2$	$p=3$	$p=4$
	h-refinement, $d=1$				h-refinement, $d=2$			
$h=1/4$	.8737	.8874	.8903	.8908	.5787	.4306	.4316	.4236
$h=1/8$	.8831	.8899	.8908	.8909	.5831	.4443	.4379	.4252
$h=1/16$	.8884	.8907	.8909	.8909	.5898	.4508	.4394	.4254
$h=1/32$	.8910	.8909	.8909	.8909	.5963	.4533	.4397	.4254
$h=1/64$	.8922	.8909	.8909	.8909	.6014	.4541	.4398	.4254
$h=1/128$	.8926	.8909	.8909	.8909	.6048	.4543	.4398	.4254
	p-refinement, $d=1$				p-refinement, $d=2$			
$h=1/4$	.7706	.8793	.9216	.9422	.2229	.4344	.5321	.6142
$h=1/8$	.7856	.8823	.9220	.9423	.2325	.4533	.5406	.6164
$h=1/16$	.7941	.8832	.9221	.9423	.2472	.4630	.5428	.6167
$h=1/32$	.7982	.8834	.9221	.9423	.2610	.4668	.5432	.6167
$h=1/64$	.8000	.8835	.9222	.9423	.2715	.4681	.5433	.6167
$h=1/128$	.8007	.8835	.9222	.9423	.2786	.4685	.5433	.6167

Table 3

The value of β *for several values of* h *and* p *for the functions* $u = x^{2/3} + (x+1/4)^{2/3}$ *on the interval* $(0,h)$ *and* $u = (x^2+y^2)^{5/8} + ((x+1/4)^2+(y+1/4)^2)^{5/8}$ *on the triangle with vertices* $(0,0)$, $(h,0)$, *and* $(0,h)$.

results. In the case $d = 2$, we tried this experiment (not shown) with different triangle geometries, and observed that β also depends on the shape and orientation of the elements involved, as well as on p and the character of the singularity. In the case of h-refinement, child elements that do not have the origin as a vertex become covered by Lemma 1 and begin the behave as in our first example. Child elements that contain the singular vertex have both the element size and the magnitude of

u reduced as $h \to 0$, so have a diminishing effect on global error estimates. In the case p-refinement, lack of convergence of β to zero provides another reason to prefer h-refinement to p-refinement in the case of point singularities. Note in each quadrature only one subelement contains the origin as a vertex. For such subelements the high order of the quadrature rule is partly offset by the low local regularity of the integrand. This single subelement thus made a large contribution to the overall error in that quadrature, mitigated by our use of many subelements.

In our third experiment we consider the functions $u = x^{2/3} + (x + 1/4)^{2/3}$ for the case $d = 1$ and $u = (x^2+y^2)^{5/8}+((x+1/4)^2+(y+1/4)^2)^{5/8}$ for the case $d = 2$. These consist of the addition of the smooth function used in the first experiment and the singular function of the second. These functions satisfy the assumptions of Lemma 5, and illustrate the convergence $\beta \to \beta_{\alpha,p}$ as $h \to 0$ predicted by that lemma. The convergence becomes more rapid with increasing p due to a larger values of ξ in (12) and (13). The results are shown in Table 3.

h	$p = 1$	$p = 2$	$p = 3$	$p = 4$	$p = 1$	$p = 2$	$p = 3$	$p = 4$
	h-refinement, $d = 1$				h-refinement, $d = 2$			
$h = 1/4$	.7662	.4085	.1768	.0693	.6945	.3963	.1698	.0897
$h = 1/8$	.7662	.4085	.1768	.0693	.6945	.3963	.1698	.0897
$h = 1/16$	.7662	.4085	.1768	.0693	.6945	.3963	.1698	.0897
$h = 1/32$	.7662	.4085	.1768	.0693	.6945	.3963	.1698	.0897
$h = 1/64$	.7662	.4085	.1768	.0693	.6945	.3963	.1698	.0897
$h = 1/128$	.7662	.4085	.1768	.0693	.6945	.3963	.1698	.0897
	p-refinement, $d = 1$				p-refinement, $d = 2$			
$h = 1/4$	.5768	.3816	.2356	.1126	.6108	.4786	.2475	.3630
$h = 1/8$	.5768	.3816	.2356	.1126	.6108	.4786	.2475	.3630
$h = 1/16$	.5768	.3816	.2356	.1126	.6108	.4786	.2475	.3630
$h = 1/32$	.5768	.3816	.2356	.1126	.6108	.4786	.2475	.3630
$h = 1/64$	.5768	.3816	.2356	.1126	.6108	.4786	.2475	.3630
$h = 1/128$	.5768	.3816	.2356	.1126	.6108	.4786	.2475	.3630

TABLE 4

The value of $\beta_{\alpha,p}$ for several values of h and p for the functions $u = x^6$ on the interval $(0, h)$ and $u = (x^2 + y^2)^3$ on the triangle with vertices $(0, 0)$, $(h, 0)$, and $(0, h)$.

In our fourth experiment we compute $\beta = \beta_{\alpha,p}$ for functions that satisfy the regularity assumption of Lemma 1 but exhibit exceptional behavior with respect to the a priori estimates. We chose $u = x^6$ for $d = 1$ and $u = (x^2 + y^2)^3 \equiv r^6$ for $d = 2$. The results are shown in Table 4. As in the previous experiment we observe $\beta_{\alpha,p}$ depending on p and the character of u (and element geometry and orientation in the case $d = 2$) but not on h. We remark that while $\beta_{\alpha,p}$ is independent of h, the function $u \to 0$ as $O(h^6)$.

7. The L_2 Norm. In this section we consider saturation with respect to the L_2 norm. As before for simplicity $\|u\|_{L_2(t)}$ will be expressed as $\|u\|_0$ or $\|u\|_{0,t}$ where appropriate. In this norm, the local saturation assumption becomes

$$\|u - \hat{\mathcal{I}}_p u\|_{0,t} \leq \beta \|u - \mathcal{I}_p u\|_{0,t} \tag{15}$$

for some $\beta = \beta(u) < 1$. As before, we begin by assuming the saturation property holds for all $\nu \in \mathcal{S}_{p+1}$. In particular,

$$\beta_0 \equiv \max_{\nu \in \mathcal{S}_{p+1}} \frac{|\nu - \hat{\mathcal{I}}_p \nu|_{0,t}}{|\nu - \mathcal{I}_p \nu|_{0,t}}. \tag{16}$$

for some $\beta_0 < 1$. The analogue of Lemma 1 is

Lemma 7. *Let $u \in H^{p+2}(t)$ satisfy (1), $u \notin \mathcal{S}_p$, and β_0 be given by (16). Then there is a constant C depending on u, the degree p, but not on its diameter h, such that*

$$\frac{|u - \hat{\mathcal{I}}_p u|_{0,t}}{|u - \mathcal{I}_p u|_{0,t}} \leq \beta_0 + Ch. \tag{17}$$

The proof of Lemma 7 follows exactly the steps of the proof Lemma 1.

Similar to the case of the H^1 seminorm, if $\beta_0 < 1$, the saturation assumption (15) holds for $\beta = \beta_0 + Ch < 1$ for h sufficiently small, with β approaching β_0 with decreasing h. Since $\beta_0 = 0$ when $\hat{\mathcal{I}}_p = \mathcal{I}_{p+1}$, we need only analyze the case of h-refinement in estimating β_0. Because $\|\cdot\|_0$ is a strong norm, some technical details related to the $|\cdot|_1$ seminorm are avoided. Perhaps more important, $\|\cdot\|_0$ is far less dependent on the shape and orientation of element t, simplifying the analysis in many cases. The analogue of Lemma 2 is

Lemma 8. *For any given polynomial $v \in \mathcal{S}_{p+1}$ and all elements t*

$$\|v - \mathcal{I}_p v\|_0 = c_0\, h^{p+1+d/2}$$

holds, where the constant c_0 depends on v, but not on the shape, orientation or the diameter h of t, and d is the space dimension.

The proof is similar to the proof of Lemma 2, but avoids the technical details about element shape and orientation associated with the $H^1(t)$ seminorm. The analogue of Corollary 3 is unnecessary.

Consider a simplex t in $1 \leq d \leq 3$ dimensions, and let $\hat{\mathcal{I}}_p$ represent uniform regular (red) refinement of z into 2^d simplicies of of size $|t|2^{-d}$. Since element shape and orientation do not enter in the computation of any of the relevant mass matrices, the analysis used in the first example in Section 4 for $d = 1$ applies in higher dimensions as well and

$$\beta_0 = 2^{-(p+1)} \tag{18}$$

for $1 \leq d \leq 3$. Similarly, when we consider the case of newest node bisection in Figure 1 center, estimate (18) applies, as the four refined elements have equal areas.

REFERENCES

[1] Babuška, I., Strouboulis, T.: The finite element method and its reliability. Numerical Mathematics and Scientific Computation. The Clarendon Press, Oxford University Press, New York (2001)

[2] Babuška, I., Whiteman, J.R., Strouboulis, T.: Finite elements. Oxford University Press, Oxford (2011). An introduction to the method and error estimation

[3] Bank, R.E., Parsania, A., Sauter, S.: Saturation estimates for *hp*-finite element mathods. Computing and Visualization in Science **16**, 195–218 (2013)

[4] Bank, R.E., Yserentant, H.: A note on interpolation, best approximation, and the saturation property. Numer. Math. **131**(1), 199–203 (2015). DOI 10.1007/s00211-014-0687-0. URL https://doi.org/10.1007/s00211-014-0687-0

[5] Bank, R.E., Yserentant, H.: On the convergence of adaptive feedback loops. Computing and Visualization in Science **20**, 59–70 (2019)

[6] Bey, J.: Simplicial grid refinement: on Freudenthal's algorithm and the optimal number of congruence classes. Numer. Math. **85**(1), 1–29 (2000). DOI 10.1007/s002110050475. URL https://doi.org/10.1007/s002110050475

[7] Bulle, R., Chouly, F., Hale, J.S., Lozinski, A.: Removing the saturation assumption in Bank-Weiser error estimator analysis in dimension three. Appl. Math. Lett. **107**, 106,429, 7 (2020). DOI 10.1016/j.aml.2020.106429. URL https://doi.org/10.1016/j.aml.2020.106429

[8] Carstensen, C., Gallistl, D., Gedicke, J.: Justification of the saturation assumption. Numer. Math. **134**(1), 1–25 (2016). DOI 10.1007/s00211-015-0769-7. URL https://doi.org/10.1007/s00211-015-0769-7

[9] Deuflhard, P., Weiser, M.: Adaptive numerical solution of PDEs. De Gruyter Textbook. Walter de Gruyter & Co., Berlin (2012). DOI 10.1515/9783110283112. URL https://doi.org/10.1515/9783110283112

[10] Dörfler, W., Nochetto, R.H.: Small data oscillation implies the saturation assumption. Numer. Math. **91**(1), 1–12 (2002). DOI 10.1007/s002110100321. URL https://doi.org/10.1007/s002110100321

[11] Lin, Q., Xie, H., Xu, J.: Lower bounds of the discretization error for piecewise polynomials. Math. Comp. **83**(285), 1–13 (2014). DOI 10.1090/S0025-5718-2013-02724-X. URL https://doi.org/10.1090/S0025-5718-2013-02724-X

[12] Nochetto, R.H.: Removing the saturation assumption in a posteriori error analysis. Istit. Lombardo Accad. Sci. Lett. Rend. A **127**(1), 67–82 (1994) (1993)

[13] Praetorius, D., Ruggeri, M., Stephan, E.P.: The saturation assumption yields optimal convergence of two-level adaptive BEM. Appl. Numer. Math. **152**, 105–124 (2020). DOI 10.1016/j.apnum.2020.01.014. URL https://doi.org/10.1016/j.apnum.2020.01.014

[14] Verfürth, R.: A posteriori error estimation techniques for finite element methods. Numerical Mathematics and Scientific Computation. Oxford University Press, Oxford (2013). DOI 10.1093/acprof:oso/9780199679423.001.0001. URL https://doi.org/10.1093/acprof:oso/9780199679423.001.0001