MATHEMATICS OF COMPUTATION
Volume 81, Number 278, April 2012, Pages 743–764
S 0025-5718(2011)02546-9
Article electronically published on September 29, 2011

BEST APPROXIMATION PROPERTY IN THE W_∞^1 NORM FOR FINITE ELEMENT METHODS ON GRADED MESHES

A. DEMLOW, D. LEYKEKHMAN, A. H. SCHATZ, AND L. B. WAHLBIN

ABSTRACT. We consider finite element methods for a model second-order elliptic equation on a general bounded convex polygonal or polyhedral domain. Our first main goal is to extend the best approximation property of the error in the W_∞^1 norm, which is known to hold on quasi-uniform meshes, to more general graded meshes. We accomplish it by a novel proof technique. This result holds under a condition on the grid which is mildly more restrictive than the shape regularity condition typically enforced in adaptive codes. The second main contribution of this work is a discussion of the properties of and relationships between similar mesh restrictions that have appeared in the literature.

1. INTRODUCTION

In this paper we consider the model second-order elliptic boundary value problem

$$\begin{aligned} -\Delta u &= f \quad \text{in } \Omega, \\ u &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{1.1}$$

Here Ω is a bounded convex polygonal or polyhedral domain in $\mathbb{R}^n$, $n = 2, 3$, and $f \in L_\infty(\Omega)$. Let S_h be a finite dimensional subspace of $H_0^1(\Omega)$ composed of piecewise polynomials of arbitrary but fixed degree k on a simplicial mesh that can be highly graded (see below for precise assumptions on the mesh). Let $u_h \in S_h$ be the finite element approximation to u given by

$$(\nabla u_h, \nabla \chi)_\Omega = (\nabla u, \nabla \chi)_\Omega, \quad \forall \chi \in S_h, \tag{1.2}$$

where $(\nabla u, \nabla v)_\Omega = \int_\Omega \nabla u \cdot \nabla v$.

Our goal is to establish the best approximation property

$$\|\nabla(u - u_h)\|_{L_\infty(\Omega)} \le C \min_{\chi \in S_h} \|\nabla(u - \chi)\|_{L_\infty(\Omega)}, \tag{1.3}$$

with constant C independent of the mesh size. Such a result has many applications. For example, (1.3) is needed in order to establish the numerically observed L_2 error estimate for bi-harmonic problems (cf. [36]) and to analyze convergence of finite

Received by the editor March 2, 2010 and, in revised form, October 27, 2010.
2010 *Mathematics Subject Classification.* Primary 65N30, 65N15.
Key words and phrases. Maximum norm, finite element, optimal error estimates.
The first author was partially supported by NSF grant DMS-0713770.
The second author was partially supported by NSF grant DMS-0811167.
The third author was partially supported by NSF grant DMS-0612599.

element methods for state-constrained optimal control problems (cf. [11]). By taking $\chi = 0$ in (1.3), we obtain the stability result

$$\|\nabla u_h\|_{L_\infty(\Omega)} \leq C\|\nabla u\|_{L_\infty(\Omega)}, \tag{1.4}$$

which is essential, for example, in analyzing the finite element solution of nonlinear problems (cf. [12, 16, 29]). It is also important to note that ∇u is Hölder continuous with modulus of continuity depending on the geometry of Ω for convex polyhedral domains in two and three space dimensions; cf. the remarks in the introductory paragraphs of [23]. Thus the estimates (1.3) and (1.4) are meaningful under the assumptions made here.

In the case of quasi-uniform grids, this estimate was proved for $n = 2$ in [28] and for $n = 3$ without unnatural restrictions on dihedral angles of the domain Ω in the recent work [23]. The quasi-uniformity assumption in these works rules out finite element grids which possess substantial mesh grading. Mesh grading is needed in order to optimally approximate u in many situations and arises naturally in adaptive codes, however. The main contribution of this work is to extend (1.3) to a class of grids which is sufficiently large to resolve solutions u to (1.1) in quasi-optimal fashion.

Let $\mathcal{T}$ be a decomposition of Ω into simplices. For any $\tau \in \mathcal{T}$, also let $h_\tau = (\text{meas}(\tau))^{1/n}$ and $\underline{h} = \min_{\tau\in\mathcal{T}} h_\tau$ and $\overline{h} = \max_{\tau\in\mathcal{T}} h_\tau$. We make the standard assumption that $\mathcal{T}$ is shape regular, that is, each element $\tau \in \mathcal{T}$ contains and is contained in balls having diameter uniformly equivalent to h_τ. Further restrictions on $\mathcal{T}$ will be described below.

Céa's Lemma trivially yields stability and almost-best-approximation properties for the finite element method in the energy norm independent of the properties of the mesh $\mathcal{T}$, but proofs of other important properties of finite element spaces and solutions typically require some further restriction on the mesh geometry. The shape regularity property described above is viewed as acceptable in most contexts because it is enforced in typical adaptive codes and allows for meshes which are sufficiently graded to optimally resolve many types of singularities, especially in two space dimensions. (Anisotropic meshes possessing "thin" or "flat" elements not satisfying shape regularity properties are also important in many contexts, especially in three space dimensions, but we do not consider such meshes here.) Proofs of some properties of the finite element method have proven elusive assuming only shape regularity, however. The most important examples of such properties are optimal or quasi-optimal error bounds in nonenergy norms such as L_2, L_∞, and W^1_∞ and stability of L_2 projections onto finite element spaces in norms other than L_2. Proofs of these properties appearing in the literature assume technical mesh restrictions that typically have unclear theoretical and practical consequences, so a secondary goal of this paper is to clarify the impact of and relationships between several mesh conditions that have appeared in the literature in connection with proofs of error estimates in nonenergy norms.

As noted above, obtaining optimal error bounds in W^1_∞, L_2, and L_∞ has proven to be difficult when only shape-regularity of the mesh is assumed. In [2], mesh-dependent norms were used to obtain optimal-order error bounds in L_∞ and L_2 for two-point boundary value problems in one space dimension with no restrictions on mesh grading. Optimal local H^1 estimates assuming only shape-regularity of the grid were also recently obtained in [14]. Other than these two papers, proofs of optimal a priori error estimates for norms other than global energy norms have generally required mesh restrictions. In two space dimensions, Eriksson in [18] obtained quasi-optimal L_∞ bounds under the assumption that a regularized mesh function $h(x)$ possessing sufficiently small gradient exists. L_2 estimates in one dimension are proved under a similar restriction in Chapter 0 of [6], and a modified version of these results for arbitrary space dimension is contained in [15]. The latter work also uses a similar mesh restriction to construct an adaptive finite element method for controlling the error in L_2 for which optimal-order complexity can be proved. In [31], L_∞ results assuming local quasi-uniformity of the grid on large element patches were announced. Finally, in [27], local H^1 and L_∞ estimates were obtained under the restriction $\underline{h} \geq \overline{h}^\gamma$ for some $\gamma \geq 1$. Local H^1 estimates were obtained under a similar restriction in [39]. In what follows, we define these three main mesh conditions more precisely and discuss the relationships between them. We also briefly discuss several mesh conditions under which stability of the L_2 projection in various norms has been proved (cf. [5, 7, 9]).

The paper is laid out as follows. In Section 2 we define and discuss mesh restrictions. In Section 3 we give preliminaries, while in Section 4 we prove our main result. Finally, in Section 5 we give a concluding discussion about possible extensions and alternate proof techniques.

2. Mesh conditions

In this section we provide a discussion of several mesh conditions which have appeared in the literature.

2.1. Statements of the conditions.

When proving our results we will assume the following condition, which previously appeared in [31].

Mesh Condition 1 (Local quasi-uniformity). *For $x \in \tau \in \mathcal{T}$, let $h(x) = h_\tau$. There exists a constant $q > 1$ and a sufficiently large constant p, such that for each point $x \in \overline{\Omega}$,*

$$\frac{h(x)}{q} \leq h(y) \leq qh(x),$$

for all $y \in \overline{\Omega}$ satisfying

$$|y - x| \leq ph(x)\ell_h,$$

where $\ell_h = \ln\left(1 + \overline{h}/\underline{h}\right)$.

Any shape-regular mesh is locally quasi-uniform, but only on small element patches. "Large-patch" local quasi-uniformity is thus a natural assumption, but it also places a nontrivial restriction on the class of allowed meshes beyond shape regularity.

The next mesh condition which we consider appeared in slightly different form in [18]. As we outline below, it is essentially equivalent to Mesh Condition 1.

Mesh Condition 2 (Eriksson). *There exists a mesh function $\tilde{h}(x) \in W^1_\infty(\Omega)$ which is uniformly equivalent to $h(x)$ and which satisfies*

$$|\nabla \tilde{h}| \leq \frac{\mu}{\ell_h},$$

for μ sufficiently small.

This condition is a natural but nontrivial extension of the restriction imposed by shape-regularity, since for merely shape-regular meshes $\tilde{h}(x)$ can be constructed so that $\|\nabla \tilde{h}\|_{L_\infty(\Omega)} \le C$ (cf. [27] for a proof when $n = 2$).

It is easy to see that Mesh Condition 2 implies Mesh Condition 1 with $p \sim \frac{1}{\mu}$. The converse is harder to prove, but is also true. That is, given a mesh which is quasi-uniform on patches of size $B_{ph(x_0)}(x_0)$, it is possible to construct the mesh function $\tilde{h}$ so that $\|\nabla \tilde{h}\|_{L_\infty(\Omega)} \sim \frac{1}{p}$ (cf. [15]). The construction of $\tilde{h}(x)$ involves a careful average of the actual mesh sizes h_τ over sufficiently many element rings about x_0. Note that the construction and proofs in [15] assume certain modifications of a standard adaptive bisection routine (newest-node bisection or its generalization to higher space dimensions; cf. [37]) rather than assuming local quasi-uniformity on large patches. Local quasi-uniformity of the mesh is in fact the essential property enforced by this modified bisection routine, however.

A further characterization of meshes satisfying Mesh Conditions 1 and 2 is that for each $x, y \in \Omega$, $h(y) \le \max\{qh(x), \frac{\mu}{\ell_h}|x - y|\}$. This condition essentially places a restriction on the growth of the mesh size as one moves away from each fixed point in the domain. It is equivalent to Conditions 1 and 2 with appropriate small adjustments to μ and q; we do not discuss it further here.

The last main condition, which we consider, first appeared in [27] in connection with proofs of a priori estimates in local energy and L_∞ norms. A priori and a posteriori error estimates in various norms have since been proved under this assumption (cf. [10, 24, 26, 39]).

Mesh Condition 3 (Global mesh restriction)**.** *There exists* $\gamma \ge 1$ *such that*

$$\underline{h} \ge \overline{h}^{\gamma}. \tag{2.1}$$

Mesh Condition 3 is fundamentally different from Mesh Conditions 1 and 2. Conditions 1 and 2 place strong *local or semi-local* restrictions on the mesh structure, prohibiting, in particular, very fast mesh change over large element patches. However, they place only a very weak restriction on the *global* relationship between $\overline{h}$ and $\underline{h}$. From the Fundamental Theorem of Calculus we have under Mesh Condition 2 that $\overline{h} \le \frac{\mu}{\ell_h}$. Thus $\overline{h}$ may decrease very slowly as $\underline{h} \to 0$, and if logarithmic factors are ignored, then in fact $\overline{h}$ may remain bounded away from 0 as $\underline{h} \to 0$. In contrast, Mesh Condition 3 seems to place no restriction on local mesh change outside of that enforced by shape regularity, but it does enforce a much stronger relationship between $\overline{h}$ and $\underline{h}$ than do Mesh Conditions 1 and 2.

We finally remark briefly on mesh restrictions imposed in order to prove stability of L_2 projections onto finite element spaces in various norms. Finite element L_2 projections are of interest in various contexts, especially in the construction and analysis of finite element methods for parabolic problems [3, 4]. Proving stability of L_2 projections in norms such as L_∞ and H^1 has thus far required restrictions on the mesh beyond shape regularity. Such mesh restrictions appear in [5, 7, 9]. We refer to the latter paper for an overview of the restrictions. The conditions appearing in these papers, which we do not discuss in detail, concern properties of mass and related matrices naturally related to the construction of L_2 projections. These matrix conditions are closely related to element geometry and can be seen to hold if the volumes of adjacent elements are not too different (cf. (6.6) in [5] and the following discussion). Thus these restrictions are in a sense similar in spirit to Mesh Condition 2 above. We do not further consider these conditions here.

2.2. **Practicality of the mesh conditions.** On a practical level, Mesh Condition 1 or 2 can be enforced in adaptive codes if the parameter p or μ, respectively, is known (cf. [15]). In particular, a standard newest-node bisection algorithm such as that described in [37] can be modified in order to enforce quasi-uniformity of the grid on large patches, or put in other terms, a sufficiently mild grading of the grid. However, p and μ arise in a rather complicated fashion in our proofs, and must, respectively, be large or small enough to reabsorb terms multiplied by constants depending on various approximation and regularity constants. If one chooses to check rather than enforce Mesh Condition 1, one must then also know the required values of p. Checking Condition 2 would additionally require construction of a smoothed mesh function $\tilde{h}$ which yields a small gradient, which is a nontrivial task. Mesh Conditions 1 and 2 thus are not very practical either to check or to enforce. We finally note that in numerical experiments involving some typical adaptive examples in which a standard adaptive code is run without attempting to enforce any additional mesh smoothness, we have been able to construct a smoothed mesh function $\tilde{h}$ with $\|\nabla\tilde{h}\|_{L_\infty(\Omega)} \approx 0.05$. This value is moderately small but seems unlikely to be small enough to reabsorb the necessary constants in the current context.

Mesh Condition 3 requires only that *some* value of γ (independent of u) exists so that (2.1) holds. This condition is likely to be met in many generic situations without any additional enforcement, since standard gradings that are necessary to resolve typical (corner) singularities indeed satisfy (2.1). The condition is also easily enforceable once a value of γ has been chosen. One potential pitfall associated with enforcing this condition is that choosing γ too small will result in a pessimistically mild grading of the mesh. That is, if a graded mesh satisfying (2.1) for some $\gamma_0 > 1$ is sufficent to optimally resolve u, but (2.1) is instead enforced for a given $1 \le \gamma < \gamma_0$, then elements will unnecessarily be added to the mesh and optimal complexity potentially compromised. Finally, checking (2.1) over several meshes arising, for example, from an adaptive computation is also not difficult. Thus Mesh Condition 3 appears to be easier to check or enforce and also more likely to hold in practical situations without being enforced than are Mesh Conditions 1 and 2.

2.3. **Optimality properties of the mesh conditions.** We finally investigate optimality properties of the mesh conditions. As the discussion below indicates, Mesh Conditions 1 and 2 are natural restrictions of shape regularity and families of meshes satisfying them *essentially preserve the approximation properties of shape-regular meshes*, at least up to logarithmic factors. Families of meshes satisfying Mesh Condition 3, on the other hand, *possess strictly worse approximation properties than corresponding families of merely shape-regular meshes.*

We first investigate all three mesh conditions in the context of standard mesh gradings for optimally resolving corner singularities.

Example 1 (Corner singularity in 2D)**.** Let Ω be a convex polygon in $\mathbb{R}^2$, and assume that the origin O is a vertex of Ω with opening angle ω. Letting $\theta = \frac{\pi}{\omega}$ and $r(x) = \operatorname{dist}(x, O)$, we then have that $u(x) \sim r^\theta$, $\nabla u(x) \sim r^{\theta-1}$, and $D^2u(x) \sim r^{\theta-2}$. Since for $\omega \le \pi/2$ no grading is needed (cf. [1]), we assume that $\omega > \pi/2$ and as a result $1 < \theta < 2$, $0 < \theta - 1 < 1$, and $-1 < \theta - 2 < 0$. Next we fix a small parameter $\delta > 0$ such that $0 < \theta - 1 - \delta < 1$. Given a maximum mesh size $\bar{h}$ and a mesh grading parameter μ, define $\tilde{r} = [\ell_h \bar{h}(2-\theta+\delta)\mu^{-1}]^{1/(\theta-1-\delta)}$, and let $h(x)$ satisfy

$$h(x) = \begin{cases} \bar{h} r^{2-\theta+\delta}, r \geq \tilde{r}, \\ \left[\frac{\ell_h(2-\theta+\delta)}{\mu}\right]^{\frac{2-\theta+\delta}{\theta-1-\delta}} \bar{h}^{\frac{1}{\theta-1-\delta}}, r < \tilde{r}. \end{cases} \tag{2.2}$$

It is easy to check that $h(x) \in W_\infty^1(\Omega)$. We may then construct $\mathcal{T}$ so that $h(x) \sim h_\tau$ uniformly in Ω; (cf. [34], Example 0.1). It is also not difficult to check that $|\nabla h(x)| \leq \frac{\mu}{\ell_h}$. Similar mesh gradings may be carried out at the other vertices of Ω.

Now, let $S_h \subset H_0^1(\Omega)$ be the continuous piecewise linear functions on $\mathcal{T}$, and let $I_h : C(\Omega) \to S_h$ be the standard Lagrange interpolant. For $\tau \in \mathcal{T}$ with $r \sim \mathrm{dist}(\tau, O) \geq \tilde{r}$, we have by standard approximation theory that $\|\nabla(u - I_h u)\|_{L_\infty(\tau)} \leq Ch(\tau) r^{\theta-2} \leq C\bar{h} r^\delta \leq C\bar{h}\,\mathrm{diam}(\Omega)^\delta \leq C\bar{h}$. If $r \leq \tilde{r}$, then by stability of the Lagrange interpolant we have

$$\begin{aligned}\|\nabla(u - I_h u)\|_{L_\infty(\tau)} \leq & C\|\nabla u\|_{L_\infty(\tau)} \leq C\|r^{\theta-1}\|_{L_\infty(\tau)} \leq C\tilde{r}^{\theta-1} \\ \leq & C\bar{h}\bar{h}^{\frac{\delta}{\theta-1-\delta}} \left[\frac{\ell_h(2-\theta+\delta)}{\mu}\right]^{\frac{\theta-1}{\theta-1-\delta}} \\ \leq & C_\mu \bar{h}.\end{aligned}$$

The last inequality follows because here $\ell_h = \ln(1 + \bar{h}/\underline{h}) \leq C \ln \frac{1}{h}$, so that $\bar{h}^\delta \ell_h^\beta$ is uniformly bounded for $0 < \bar{h} \leq 1$ and any fixed β. We thus have shown that

$$\min_{\chi \in S_h} \|\nabla(u-\chi)\|_{L_\infty(\Omega)} \leq C\bar{h} \tag{2.3}$$

for a mesh satisfying Eriksson's mesh smoothness condition, and also Mesh Condition 1 with $p \geq \frac{c}{\mu}$. Note that the mesh grading given by (2.2) appears slightly pessimistic because an extra factor of r^δ is included. If this extra factor is excluded, a similar construction yields (2.3), but with extra logarithmic factors in the upper bound. Observe also that if the grading parameter μ is sufficiently small to ensure that (1.3) holds, then we may combine (1.3) with (2.3) to yield the estimate $\|\nabla(u - u_h)\|_{L_\infty(\Omega)} \leq C\bar{h}$ for the finite element error $u - u_h$. We do not, however, trace the correct value of μ in our proof below and anticipate that it would be very difficult to do so.

We now show that the meshes produced by (2.2) contain $O(\overline{h}^{-2})$ degrees of freedom, which is optimal for the finite element spaces under consideration since u is resolved to a tolerance $O(\overline{h})$ by these meshes. First, by shape regularity the asymptotic number of elements contained in $B_{\tilde{r}}(O)$ is

$$\frac{\tilde{r}^2}{\left(\left[\frac{\ell_h(2-\theta+\delta)}{\mu}\right]^{\frac{2-\theta+\delta}{\theta-1-\delta}} \bar{h}^{\frac{1}{\theta-1-\delta}}\right)^2} \leq \frac{[\ell_h \bar{h}(2-\theta+\delta)\mu^{-1}]^{2/(\theta-1-\delta)}}{\left(\left[\frac{\ell_h(2-\theta+\delta)}{\mu}\right]^{\frac{2-\theta+\delta}{\theta-1-\delta}} \bar{h}^{\frac{1}{\theta-1-\delta}}\right)^2} \leq C(\theta,\mu)\ell_h^2.$$

Let $d_j = 2^j \tilde{r}$ for $j = 0, 1, ...$, let $\Omega_j = \{x \in \Omega : d_{j-1} < \mathrm{dist}(x, O) < d_j\}$ for $j = 1, 2, ...$, and let J be the largest value for which $\Omega_j \neq \emptyset$. Noting that $h|_{\Omega_j}$ is uniformly equivalent to $h(d_j)$, we calculate that the number of elements in $\Omega \setminus B_{\tilde{r}}(O)$ is equivalent to

$$\sum_{j=1}^J \frac{|\Omega_j|}{h(d_j)^2} \leq C \sum_{j=1}^J \frac{d_j^2}{\overline{h}^2 d_j^{4-2\theta+2\delta}} \leq C\overline{h}^{-2} \sum_{j=1}^J d_j^{2\theta-2-2\delta}.$$

Since $2\theta - 2 - \delta > 0$, the latter sum is a geometric sum with the largest summand bounded by $C\mathrm{diam}(\Omega)^{2\theta-2-2\delta}$, so this sum is bounded by a constant that is independent of $\overline{h}$. This completes the proof that the mesh contains $O(\overline{h}^{-2})$ elements.

Finally, we remark that the grading given in (2.2) satisfies Condition 3 with $\gamma = \frac{1}{\theta-1-\delta}$.

We now adopt a viewpoint which is more consistent with the mathematical framework that has been used to assess optimality of adaptive methods. Let $\mathcal{T}_0$ be a shape-regular and quasi-uniform mesh. Then let $\mathbb{T}$ be the family of all conforming meshes that can be derived from $\mathcal{T}_0$ by a standard newest-node bisection algorithm or its generalization to $\mathbb{R}^3$. All meshes in $\mathbb{T}$ are automatically uniformly shape-regular because of the properties of newest-node bisection. Given fixed $\gamma, \mu > 0$, let $\mathbb{T}_\gamma$ and $\mathbb{T}_\mu$ be the (strict) subsets of $\mathbb{T}$ consisting of daughter meshes of $\mathcal{T}_0$ which additionally satisfy Mesh Condition 3 and Mesh Condition 2 (here with ℓ_h taken to be 1 for the time being), respectively. Given $\mathcal{T}$ lying in $\mathbb{T}$, $\mathbb{T}_\gamma$, or $\mathbb{T}_\mu$, we denote by $S_\mathcal{T} \subset H_0^1(\Omega)$ the corresponding conforming piecewise linear finite element space. Whereas typical a priori error estimates for the finite element method seek to show that approximation to u is optimized over a single finite element space corresponding to a single mesh, an adaptive finite element method is considered to be optimal if it optimizes approximation of u over *all* finite element spaces $S_\mathcal{T}$ $(T \in \mathbb{T})$ having a given number of degrees of freedom; cf. [8].

It is shown in Corollary A.6 of [15] that for every mesh $\mathcal{T} \in \mathbb{T}$, there exists a mesh $\mathcal{T}_\mu \in \mathbb{T}_\mu$ which is a daughter mesh of $\mathcal{T}$ such that $\#\mathcal{T}_\mu - \#\mathcal{T}_0 \leq C(\#\mathcal{T} - \#\mathcal{T}_0)$. Here $\#\mathcal{T}$ is the cardinality of $\mathcal{T}$ and C depends only on μ and other nonessential quantities. Put in different terms, when viewed from the standpoint of its ability to approximate functions, so long as logarithmic factors are ignored the class of meshes $\mathbb{T}_\mu$ is essentially as "rich" as $\mathbb{T}$ since $S_\mathcal{T} \subset S_{\mathcal{T}_\mu}$ when $\mathcal{T}$ and $\mathcal{T}_\mu$ are as above. In the present context, this means that if $\min_{\chi \in S_\mathcal{T}} \|\nabla(u-\chi)\|_{L_\infty(\Omega)} \leq \epsilon$, then there exists $\mathcal{T}_\mu \in \mathbb{T}_\mu$ such that $\min_{\chi \in S_{\mathcal{T}_\mu}} \|\nabla(u-\chi)\|_{L_\infty(\Omega)} \leq \epsilon$ and $\#\mathcal{T}_\mu - \#\mathcal{T}_0$ is not more than a fixed constant multiple of $\#\mathcal{T} - \#\mathcal{T}_0$.

We now give two brief examples which allow us to compare the degree to which enforcing Mesh Condition 1 (with logarithmic factors included) or Mesh Condition 3 might inflate the number of degrees in a given mesh.

Example 2 (Corner singularity in 2D, part 2)**.** The mesh grading (2.2) in Example 1 above was hand-constructed in order to both optimally resolve a typical corner singularity and to satisfy Mesh Condition 1. We now consider a simpler and more natural mesh grading which likewise optimally resolves the corner singularity in Example 1, but which does not satisfy Mesh Condition 1. Given a maximum mesh size $\overline{h}$, let $\underline{h} = \overline{h}^{\frac{1}{\theta-1}}$ and define

$$h(x) = \begin{cases} \bar{h} r^{2-\theta}, r \geq \underline{h}, \\ \underline{h},\ r < \underline{h}. \end{cases}$$

The mesh $\mathcal{T}$ thus generated optimally resolves the corner singularity in Example 1 in the sense that it contains $O(\overline{h}^{-2})$ degrees of freedom and that (2.3) is satisfied for the piecewise linear Lagrange finite element space S_h generated by this mesh. We omit the calculations as they are similar to but simpler than those given in Example 1.

A crude calculation using Corollary A.6 of [15] shows that refining additional elements in $\mathcal{T}$ in order to enforce Mesh Condition 1 inflates the number of degrees of freedom to no more than $O(\ell_h^{4-2\theta}\overline{h}^{-2})$ (though by Example 1 this bound may be pessimistic). For the mesh $\mathcal{T}$, Mesh Condition 3 is satisfied with $\gamma_0 = \frac{1}{\theta-1}$. Suppose, however, that we wish to enforce Condition 3, but do not know a priori the correct value of γ and thus choose some $1 < \gamma < \gamma_0$. Without making any assumption about precisely how Mesh Condition 3 is enforced, the outcome will be a new mesh $\tilde{\mathcal{T}}$ with maximum mesh size $\tilde{h}$ and minimum mesh size $\underset{\sim}{h}$, where $\underset{\sim}{h} = \tilde{h}^\gamma$. Note that there must be at least $O(\tilde{h}^{-2})$ elements in $\tilde{\mathcal{T}}$. Now, let $T_O \in \tilde{\mathcal{T}}$ be an element touching the origin O. Recalling that $\nabla u \sim r^{\theta-1}$ near O, we find that there is a constant c such that $\|\nabla(u-\chi)\|_{L_\infty(T_O)} \geq c\|r^{\theta-1}\|_{L_\infty(T_O)} \geq c\underset{\sim}{h}^{\theta-1}$ for any $\chi \in S_h$. But $\underset{\sim}{h}^{\theta-1} = \tilde{h}^{\gamma(\theta-1)} = \tilde{h}^{\frac{\gamma}{\gamma_0}}$. Because $\frac{\gamma}{\gamma_0} < 1$ and $\#\tilde{\mathcal{T}} \geq C\tilde{h}^{-2}$, using $O(\tilde{h}^{-2})$ elements produce a convergence rate that is suboptimal by $1 - \frac{\gamma}{\gamma_0}$, or in other terms, producing an error $O(\epsilon)$ in W^1_∞ requires $O(\overline{\epsilon}^{-\frac{2\gamma_0}{\gamma}})$ degrees of freedoms instead of the optimal number $O(\epsilon^{-2})$. Because no assumption at all concerning the nature of the mesh $\tilde{\mathcal{T}}$ has been made here beyond requiring that it satisfy Mesh Condition 3 with $\gamma < \frac{1}{\theta-1}$, *any* sequence of meshes in $\mathbb{T}_\gamma$ with $\gamma < \gamma_0$ will thus produce suboptimal convergence.

Example 3 (Extreme mesh grading)**.** Let Ω be the unit square, and let $\mathcal{T}_0$ be a uniform coarse mesh consisting of four triangles with vertices given by the four corners and center of Ω. Finally, let $\mathcal{T}_i$, $i > 0$, be derived by iteratively bisecting at each refinement step only those elements touching the origin, plus additional elements to maintain mesh conformity. The number of elements $\#\mathcal{T}_i$ in $\mathcal{T}_i$ is easily seen to be bounded by Ci for some fixed constant C, and $\underline{h} = 2^{-i}$. The resulting meshes are strongly graded toward the origin; here the natural mesh function $h(x,y) \approx \sqrt{x^2+y^2}$ so that $|\nabla h(x,y)| \approx 1$ for all (x,y). Shape regular meshes cannot in general be more strongly graded, since as noted above the gradients of their mesh functions are uniformly bounded.

$\mathcal{T}_i$ always contains an element of diameter (and area) 1, so enforcing Mesh Condition 3 will inflate the number of degrees of freedom from Ci to at least $\overline{h}^{-2} = \underline{h}^{-\frac{2}{\gamma}} \approx (2^i)^{\frac{2}{\gamma}} = 2^{\frac{2i}{\gamma}}$. Enforcing Condition 1, on the other hand, will inflate the number of degrees of freedom from roughly i to roughly $i\ell_h \approx i \ln 2^i \approx i^2$. Thus, in this extreme example, enforcing Mesh Condition 1 with the logarithmic factor taken into account greatly inflates the number of elements in $\mathcal{T}_i$ (from i to i^2). Enforcing Condition 3 will, however, have a much more extreme effect, creating meshes with numbers of degrees of freedom that grow exponentially in the number of refinement steps.

We summarize as follows. If we fix γ, then producing a mesh in $\mathbb{T}_\gamma$ from an adaptively produced mesh lying in $\mathbb{T}$ and thereby enforcing Mesh Condition 3 will in many cases have no effect at all, but in extreme cases it will wildly inflate the number of elements in the mesh even if γ is chosen to be relatively large. Enforcing Condition 1, on the other hand, is likely to lead to some inflation of the number of elements in the mesh in many situations, but the added number of elements is in all cases relatively moderate even if logarithmic factors are taken into account. Thus while Mesh Condition 3 may be expected to hold naturally in many practical situations and can be checked or enforced with relative ease, it has the large disadvantage of producing suboptimal classes of meshes and finite element spaces if enforced. On the other hand, it is difficult to judge whether Mesh Conditions 1 and 2 hold in any given practical situation because the parameters p and μ are difficult to determine. However, these conditions represent a natural "tightening" of restrictions already imposed by shape regularity and at least up to logarithmic factors preserve optimality properties of finite element spaces generated from shape-regular meshes. We shall prove our results under Mesh Condition 1.

3. Preliminaries

In this section we list a number of properties and assumptions that we will need in our proofs. The assumptions we make below concerning properties of finite element spaces are fulfilled, for example, by standard Lagrange finite element spaces of arbitrary polynomial degree k.

3.1. **Local approximation.** There exists a linear operator $I_h : H_0^1(\Omega) \cap H^2(\Omega) \to S_h(\Omega)$ such that for any element $\tau \in \mathcal{T}$,

$$\|I_h v - v\|_{H^s(\tau)} \leq C h_\tau^{2-s} |v|_{H^2(\tau)} \quad \text{for } 0 \leq s \leq 1.$$

Since our presentation is restricted to two and three space dimensions we may simply take I_h to be the Lagrange interpolant.

3.2. **Inverse properties.** For any $\chi \in S_h$ and $\tau \in \mathcal{T}$,

$$\|\chi\|_{W_p^s(\tau)} \leq C h_\tau^{t-s-n(1/q-1/p)} \|\chi\|_{W_q^t(\tau)}, \quad 0 \leq t \leq s \leq 2,\ 1 \leq q \leq p \leq \infty.$$

3.3. **Superapproximation.** Given $D \subset \Omega$ with $\mathcal{T}$ quasi-uniform on D of size h, $d \geq \kappa h > 0$ for sufficiently large κ, and $D_1 \subset D$ with $\operatorname{dist}(D_1, \partial D \setminus \partial\Omega) \geq d$, let ω be a smooth function which is 0 on $\Omega \setminus D_1$ and satisfying $\|D^s \omega\|_{L_\infty(\Omega)} \leq C d^{-s}$, $s \geq 0$. For $D \subset \Omega$, let $S_h^0(D) = S_h \cap H_0^1(D)$. Then for each $\chi \in S_h(D)$ there exists $\eta \in S_h^0(D)$ satisfying

$$\|\nabla(\omega\chi - \eta)\|_D \leq Ch(d^{-1}\|\nabla\chi\|_D + d^{-2}\|\chi\|_D).$$

Furthermore, let $D_4 \subset D_3 \subset D_2 \subset D_1$ with $\operatorname{dist}(D_4, \partial D_3 \setminus \partial\Omega) \geq d \geq \kappa h$ and $\operatorname{dist}(D_3, \partial D_2 \setminus \partial\Omega) \geq d \geq \kappa h$. Then if $\omega \equiv 1$ on D_2, we have $\eta \equiv \chi$ on D_3 and

$$\|\nabla(\omega\chi - \eta)\|_D \leq Ch(d^{-1}\|\nabla\chi\|_{D\setminus D_4} + d^{-2}\|\chi\|_{D\setminus D_4}). \tag{3.1}$$

Superapproximation properties are standard in the finite element literature and are valid for many finite element spaces. For more detailed discussions see [14, 25, 32].

We also need the following result which is very similar to Proposition 2.2 in [32].

Proposition 3.1. *Let the superapproximation property* 3.3 *hold and let* $D_4 \subset D_3 \subset D_2 \subset D_1 \subset D \subset \Omega$ *with* $\operatorname{dist}(D_i, \partial D_{i-1} \setminus \partial\Omega) \geq d$ *as above, and similarly for* D_1 *and* D *and* D *and* Ω. *There is a constant* C *such that for each* $\chi \in S_h(D)$ *there exists an* $\eta \in S_h^0(D_1)$ *with* $\eta \equiv \chi$ *on* D_2 *and*

$$\|\nabla(\chi - \eta)\|_D \leq C(\|\nabla\chi\|_{D\setminus D_4} + d^{-1}\|\chi\|_{D\setminus D_4}).$$

Proof. Let $\omega \in C_0^\infty(D_1)$ be as in Superapproximation assumption 3.3. By the triangle inequality,

$$\|\nabla(\chi - \eta)\|_D \leq \|\nabla(\omega\chi - \eta)\|_D + \|\nabla(\chi - \omega\chi)\|_D.$$

By (3.1)

$$\|\nabla(\omega\chi - \eta)\|_D \leq Ch(d^{-1}\|\nabla\chi\|_{D\setminus D_4} + d^{-2}\|\chi\|_{D\setminus D_4}).$$

If we take $\omega \equiv 1$ on D_2, then since $|\nabla \omega| \leq C d^{-1}$,

$$\begin{aligned}\|\nabla(\chi - \omega\chi)\|_D \leq & \|\chi \nabla(1-\omega)\|_{D\setminus D_4} + \|(1-\omega)\nabla\chi\|_{D\setminus D_4} \\ \leq & C(\|\nabla \chi\|_{D\setminus D_4} + d^{-1}\|\chi\|_{D\setminus D_4}).\end{aligned}$$

Since $h \leq d$ the above two estimates conclude the proof of the proposition. □

3.4. **Scaling.** Let $x_0 \in \overline{\Omega}$ and $R \geq kh$. The linear transformation $y = (x - x_0)/R$ takes $\Omega_R(x_0) = B_R(x_0) \cap \Omega$ into a new domain $\hat{\Omega}_1$ and $S_h(\Omega_R(x_0))$ into a new function space $\hat{S}_{h/R}(\hat{\Omega}_1)$. Then $\hat{S}_{h/R}(\hat{\Omega}_1)$ satisfies assumptions 3.1 through 3.3 with h replaced by h/R. The constants occurring remain unchanged.

3.5. H^2 **regularity.** The following H^2 regularity result is known to hold for convex domains (cf. [21]).

Lemma 3.1. *For any convex domain Ω there exists a constant C depending only on Ω such that*

$$\|u\|_{H^2(\Omega)} \leq C\|f\|_{L_2(\Omega)}.$$

3.6. **Pointwise estimates for the Green's function.** In our proof we will make heavy use of pointwise estimates for the Green's function and its derivatives. The proof for general second order elliptic equation for $n \geq 3$ can be found in [22]. In two dimensions for the Laplace equation a simplified proof can be found in [19].

Lemma 3.2. *Let $G(x,y)$ denote the Green's function for* (1.1) *and let Ω be a bounded convex subset of $\mathbb{R}^n$. Then the following estimates hold:*

$$|G(x,y)| \leq \begin{cases} C(1 + \ln|x-y|), & n = 2, \\ C|x-y|^{2-n}, & n \geq 3, \end{cases} \tag{3.2a}$$

$$|\nabla_x G(x,y)| \leq C|x-y|^{1-n}, \quad n \geq 2, \tag{3.2b}$$

$$|\nabla_x \nabla_y G(x,y)| \leq C|x-y|^{-n}, \quad n \geq 2. \tag{3.2c}$$

Sharper Green's function estimates are derived for two- and three-dimensional convex polygonal and polyhedral domains in [23].

3.7. **Inverse type inequalities for harmonic and discrete harmonic functions.** In our argument we will constantly deal with functions which are harmonic on some parts of the domain. The following inverse type inequalities significantly simplify many arguments. The result is essentially the same as Lemma 8.3 of [33], so we do not provide a proof.

Lemma 3.3. *Let $D \subset D_d \subset \Omega$, and for $d > 0$ let $D_d = \{x \in \Omega : \operatorname{dist}(x, D) \leq d\}$. Assume v vanishes on $\partial D_d \cap \partial\Omega$ and that v is harmonic on D_d, i.e.,*

$$(\nabla v, \nabla w) = 0, \quad \forall w \in H_0^1(D_d).$$

Then,

$$|v|_{H^2(D)} \leq C d^{-1}\|v\|_{H^1(D_d)}, \tag{3.3a}$$

$$\|v\|_{H^1(D)} \leq C d^{-1}\|v\|_{L_2(D_d)}. \tag{3.3b}$$

We will also require similar estimates for *discrete harmonic* functions. We say that a function $v_h \in S_h$ is discrete harmonic over $D \subset \Omega$ if $(\nabla v_h, \nabla \chi) = 0$ for all $\chi \in S_h$ having support in D.

Lemma 3.4. *Let $D \subset \Omega$ have diameter d, and let $D_d = \{x \in \Omega : \operatorname{dist}(x, D) < d\}$. Assume also that $\mathcal{T}_h$ is quasi-uniform with mesh diameter h on D_d for $d \geq kh$ with k sufficiently large. If $v_h \in S_h$ is discrete harmonic on D_d, then*

$$\|\nabla v_h\|_{L_2(D)} \leq Cd^{-1}\|v_h\|_{L_2(D_d)} \leq Cd^{-2}\|v_h\|_{H^{-1}_{<}(D_d)}. \tag{3.4}$$

Here,

$$\|v_h\|_{H^{-1}_{<}(D_d)} = \sup_{\substack{z \in H^1(\Omega), \\ z=0 \text{ on } \Omega \setminus D_d}} \frac{(v_h, z)}{\|z\|_{H^1(D_d)}}.$$

Proof. See [38], Lemma 9.1 for a proof of the first inequality in (3.4); the appropriate power of d is easy to trace. The second estimate is essentially contained in Lemma 9.2 of [38], where again the appropriate powers of d may be traced with slightly more effort. Following the proof of Lemma 9.2 in [38], we see that the only detail missing in the proof of the second inequality is a uniform bound for an H^2 regularity constant. In particular, we must show that D_d is contained in a subdomain $\tilde{D}$ of Ω having diameter equivalent to d and for which $|v|_{H^2(\tilde{D})} \leq C_2\|\Delta v\|_{L_2(\tilde{D})}$ for all $v \in H^2(D_d)$, where the constant C_2 is independent of d and $\tilde{D}$. Independence of C_2 from d follows by a simple scaling argument. Because Ω is convex and polyhedral, we may always take $\tilde{D}$ to be a copy of Ω which is scaled by some factor uniformly equivalent to d and properly translated. In the more general case where Ω is a nonconvex polyhedral domain (which we do not consider here), a similar result can be proved under the additional assumption that D lies a distance d from any re-entrant vertices or edges of $\partial\Omega$. In this case it is possible to define a finite set of convex reference domains $\tilde{D}_1, ..., \tilde{D}_N$ such that at least one of the $\tilde{D}_i$'s will have the desired properties; cf. Section 2.2 of [13]. Thus C_2 is also independent of the shape of D_d, as desired. □

4. Proof of the main result

4.1. **Idea of the proof.** The proof technique is rather novel and roughly can be broken into three parts. Let $e = u - u_h$. In the following we fix a point x_0 such that $\|\nabla e\|_{L^\infty(\Omega)} = |\nabla e(x_0)|$ and let $\Omega_0 = B_d(x_0) \cap \Omega$, where $d \geq kh(x_0)$ for some k sufficiently large.

Part 1: We establish local error estimates. For example in [35], Theorem 1.2, it has been shown for $\Omega_0 \subset\subset \Omega$, that

$$\begin{aligned} |\nabla e(x_0)| \leq & C \min_{\chi \in S_h} \left(\|\nabla(u-\chi)\|_{L_\infty(\Omega_0)} + d^{-1}\|u-\chi\|_{L_\infty(\Omega_0)}\right) \\ & + Cd^{-\frac{n}{2}-1}\|e\|_{L_2(\Omega_0)}. \end{aligned} \tag{4.1}$$

Here the term $d^{-1}\|u-\chi\|_{L_\infty(\Omega_0)}$ can be removed by choosing $\chi = \chi_* + \frac{1}{|\Omega_0|}\int_{\Omega_0}(u - \chi_*)$ with $\chi_* = \operatorname{argmin}_{\chi \in S_h}\|\nabla(u-\chi)\|_{L_\infty(\Omega_0)}$ and applying the Poincaré inequality. The goal of Theorem 1 is to extend the above result all the way to the boundary.

Part 2: Once (4.1) is established the next step is to transform the "slush" term $\|e\|_{L_2(\Omega_0)}$ into a more convenient form. For example, if $\Omega_0 \subset\subset \Omega$, we can replace e by $e - c$ in (4.1), where c is a constant. Only the slush term is affected by this change. Then by choosing $c = (e, 1)_{\Omega_0}$, we can establish (cf. Lemma 4.1 below),

$$\|e\|_{L_2(\Omega_0)} \leq C\|\nabla e\|_{H^{-1}(\Omega_0)}. \tag{4.2}$$

In the case $\Omega_0 \cap \partial\Omega \neq \emptyset$ this argument does not work, but nevertheless we are able to establish a somewhat similar result, where we bound the "slush" term by some combination of derivatives of e in a weaker norm (cf. Lemma 4.2).

Part 3: The final step is to kick back $\|\nabla e\|_{H^{-1}(\Omega_0)}$ to $\|\nabla e\|_{L_\infty(\Omega)}$. Once we transform this slush term, the next step in essence is to show that $d^{-\frac{n}{2}-1}\|e\|_{L_2(\Omega_0)}$ is smaller than $\|\nabla e\|_{L_\infty(\Omega)}$. In order to establish this fact we use a decomposition of Ω on special annuli that give us a limit on the mesh growth on each of them (cf. Proposition 4.1). After that applying approximation theory together with careful bookkeeping finishes the proof.

These steps work almost trivially in the case of $\Omega_0 \subset\subset \Omega$. Therefore, the main challenge is to extend all of these steps up to the boundary, i.e., for the case $\Omega_0 \cap \partial\Omega \neq \emptyset$.

4.2. **Part 1. Local error estimates.** The goal of this part is to establish local error estimates.

Theorem 1. *Let u and u_h satisfy* (1.1) *and* (1.2), *and let the assumptions of Section* 3 *hold. Let $x_0 \in \Omega$ be such that $\|\nabla e\|_{L_\infty(\Omega)} = |\nabla e(x_0)|$. Define $D = \Omega \cap B_d(x_0)$ for $d \geq kh(x_0)$, where $k > 0$ is a sufficiently large number. Assume that the mesh is quasi-uniform on D and can be extended quasi-uniformly to the whole domain Ω with all elements having diameter uniformly equivalent to $h = h(x_0)$. Then there exists a constant C independent of h and u such that for any $\chi \in S_h$,*

$$|\nabla e(x_0)| \leq C\left(\|\nabla(u-\chi)\|_{L_\infty(D)} + d^{-1}\|u-\chi\|_{L_\infty(D)} + d^{-\frac{n}{2}-1}\|e\|_{L_2(D)}\right),\ n = 2, 3.$$

Remark 1. The assumption that the mesh may be extended quasi-uniformly to the whole domain Ω is not essential, but simplifies the proof a lot. This assumption is in any case not very restrictive. It holds, in particular, if the original mesh came from some coarse mesh by successive refinement and coarsening, which holds for many adaptive codes.

Proof. In what follows we will use the abbreviation $mD = B_{md}(x_0) \cap \Omega$. Let ω be a cut-off function with the properties $\omega \equiv 1$ on D, $\mathrm{supp}(\omega) \subset 2D$, and $|\nabla \omega| \leq Cd^{-1}$. Let $\tilde{u} = \omega u$. By the assumption of the theorem, we can extend the quasi-uniform mesh on D to the whole domain Ω quasi-uniformly. Call the corresponding finite element space $\tilde{S}_h$. Define $\tilde{u}_h$ to be the Ritz projection of $\tilde{u}$ onto $\tilde{S}_h$:

$$(\nabla \tilde{u}_h, \nabla \tilde{\chi})_\Omega = (\nabla \tilde{u}, \nabla \tilde{\chi})_\Omega, \quad \forall \tilde{\chi} \in \tilde{S}_h.$$

Then,

$$|\nabla e(x_0)| \leq |\nabla(\tilde{u} - \tilde{u}_h)(x_0)| + |\nabla(\tilde{u}_h - u_h)(x_0)|. \tag{4.3}$$

By Theorem 2 in [23], which gives global best approximation property on quasi-uniform meshes, the first term on the right-hand side of (4.3) can be estimated as

$$\|\nabla(\tilde{u} - \tilde{u}_h)\|_{L_\infty(\Omega)} \leq C\|\nabla \tilde{u}\|_{L_\infty(\Omega)} \leq C\left(\|\nabla u\|_{L_\infty(2D)} + d^{-1}\|u\|_{L_\infty(2D\setminus D)}\right).$$

The term $\tilde{u}_h - u_h$ in (4.3) is discrete harmonic on D; we do not consider the properties of this function outside of D. The rest of the proof is devoted to establishing that

$$|\nabla(\tilde{u}_h - u_h)(x_0)| \leq Cd^{-\frac{n}{2}-1}\|u_h\|_{L_2(D)}.$$

Let $\psi_h = \tilde{u}_h - u_h$. By Proposition 3.1, there exists $\eta_h \in \tilde{S}_h^0(\frac{3}{4}D)$ such that $\eta_h \equiv \psi_h$ on $\frac{1}{2}D$ and

$$\|\nabla \eta_h\|_{L_2(\frac{3}{4}D)} \le C(\|\nabla \psi_h\|_{L_2(\frac{7}{8}D)} + d^{-1}\|\psi_h\|_{L_2(\frac{7}{8}D)}). \tag{4.4}$$

Next we define an approximate derivative Green's function g by

$$\begin{aligned} -\Delta g &= \partial\delta, \quad \text{in } \Omega,\\ g &= 0, \qquad \text{on } \partial\Omega. \end{aligned} \tag{4.5}$$

Here δ is a smooth, discrete δ-function supported in an element containing x_0 and satisfying $\|\delta\|_{L_q} \le Ch(x_0)^{-n(1-\frac{1}{q})}$ (cf. [35], Appendix A, for details). Also, ∂ is a directional derivative. Let $\tilde{g}_h$ be the finite element projection of g onto $\tilde{S}_h$. Then,

$$-\partial\psi_h(x_0) = -\partial\eta_h(x_0) = (\nabla g, \nabla\eta_h) = (\nabla\tilde{g}_h, \nabla\eta_h).$$

Also, by Proposition 3.1, there exists $\zeta_h \in \tilde{S}_h^0(\frac{1}{2}D)$ such that $\zeta_h \equiv \tilde{g}_h$ on $\frac{1}{4}D$ and

$$\|\nabla(\tilde{g}_h - \zeta_h)\|_{L_2(\frac{3}{4}D)} \le C(\|\nabla\tilde{g}_h\|_{L_2(\frac{7}{8}D\setminus\frac{1}{4}D)} + d^{-1}\|\tilde{g}_h\|_{L_2(\frac{7}{8}D\setminus\frac{1}{4}D)}).$$

Recalling that η_h is supported on $\frac{3}{4}D$ and discrete harmonic in $\frac{1}{2}D$ and using (4.4), we have

$$\begin{aligned} -\partial\psi_h(x_0) &= (\nabla(\tilde{g}_h - \zeta_h), \nabla\eta_h) + (\nabla\zeta_h, \nabla\eta_h)\\ &= (\nabla(\tilde{g}_h - \zeta_h), \nabla\eta_h)\\ &\le \|\nabla(\tilde{g}_h - \zeta_h)\|_{L_2(\frac{3}{4}D)}\|\nabla\eta_h\|_{L_2(\frac{3}{4}D)}\\ &\le C(\|\nabla\tilde{g}_h\|_{L_2(\frac{7}{8}D\setminus\frac{1}{4}D)} + d^{-1}\|\tilde{g}_h\|_{L_2(\frac{7}{8}D\setminus\frac{1}{4}D)})\\ &\quad\times(\|\nabla\psi_h\|_{L_2(\frac{7}{8}D)} + d^{-1}\|\psi_h\|_{L_2(\frac{7}{8}D)}). \end{aligned} \tag{4.6}$$

Using that ψ_h is discrete harmonic, the triangle inequality, the fact that $u = \tilde{u}$ on D, and global a priori estimates in the L_2 norm, we have

$$\begin{aligned} \|\nabla\psi_h\|_{L_2(\frac{7}{8}D)} + d^{-1}\|\psi_h\|_{L_2(\frac{7}{8}D)} &\le Cd^{-1}\|\psi_h\|_{L_2(D)}\\ &\le Cd^{-1}\left(\|u - u_h\|_{L_2(D)} + \|\tilde{u} - \tilde{u}_h\|_{L_2(D)}\right)\\ &\le Cd^{-1}\left(\|e\|_{L_2(D)} + h\|\nabla\tilde{u}\|_{L_2(2D)}\right)\\ &\le Cd^{-1}\left(\|e\|_{L_2(D)} + hd^{\frac{n}{2}-1}\|u\|_{L_\infty(2D)} + hd^{\frac{n}{2}}\|\nabla u\|_{L_\infty(2D)}\right). \end{aligned} \tag{4.7}$$

Now we turn to $\|\nabla\tilde{g}_h\|_{L_2(\frac{7}{8}D\setminus\frac{1}{4}D)} + d^{-1}\|\tilde{g}_h\|_{L_2(\frac{7}{8}D\setminus\frac{1}{4}D)}$. Using that $\tilde{g}_h$ is discrete harmonic we have

$$\|\nabla\tilde{g}_h\|_{L_2(\frac{7}{8}D\setminus\frac{1}{4}D)} + d^{-1}\|\tilde{g}_h\|_{L_2(\frac{7}{8}D\setminus\frac{1}{4}D)} \le Cd^{-1-s}\|\tilde{g}_h\|_{H^{-s}(D\setminus\frac{1}{8}D)}, \ s = 0, 1. \tag{4.8}$$

For $n = 2$ we apply (4.8) with $s = 0$. By the Sobolev embedding theorem ($W_1^1 \hookrightarrow L_2$),

$$\|\tilde{g}_h\|_{L_2(D\setminus\frac{1}{8}D)} \le C\|\tilde{g}_h\|_{W_1^1(D\setminus\frac{1}{8}D)}.$$

Note that the Sobolev embedding constant appearing in the inequality above is domain independent. To check that, we can scale the domain D to a unit size domain $\tilde{D}$ by introducing a new variable $y = x/d$. Then it is easy to show that for any general function $V(y) = v(yd)$ we have

$$\|D^sV\|_{L_q(\tilde{D})} = d^{s-n/q}\|D^sv\|_{L_q(D)}, \quad s = 0, 1. \tag{4.9}$$

Thus, if $D\backslash\frac{1}{8}D$ is scaled to a subset of a fixed unit-sized annulus, $\tilde{g}_h$ is extended to zero in this annulus if D abuts $\partial\Omega$, and by using (4.9), we can see that this constant is indeed independent of d.

For $n = 3$ we use (4.8) with $s = 1$. We then use Hölder's inequality and the Sobolev embeddings $W_1^1 \hookrightarrow L_{3/2}$ and $H^1 \hookrightarrow L_6$ to find

$$\begin{aligned}\|\tilde{g}_h\|_{H^{-1}(D\backslash\frac{1}{8}D)} \leq C &\sup_{\|v\|_{H^1_<(D\backslash\frac{1}{8}D)}=1} (\tilde{g}_h, v)\\ &\leq C\|\tilde{g}_h\|_{L_{3/2}(D\backslash\frac{1}{8}D)} \sup_{\|v\|_{H^1_<(D\backslash\frac{1}{8}D)}=1} \|v\|_{L_3(D\backslash\frac{1}{8}D)}\\ &\leq Cd^{1/2}\|\tilde{g}_h\|_{L_{3/2}(D\backslash\frac{1}{8}D)} \sup_{\|v\|_{H^1_<(D\backslash\frac{1}{8}D)}=1} \|v\|_{L_6(D\backslash\frac{1}{8}D)}\\ &\leq Cd^{1/2}\|\tilde{g}_h\|_{W_1^1(D\backslash\frac{1}{8}D)} \sup_{\|v\|_{H^1_<(D\backslash\frac{1}{8}D)}=1} \|v\|_{H^1(D\backslash\frac{1}{8}D)}\\ &\leq Cd^{1/2}\|\tilde{g}_h\|_{W_1^1(D\backslash\frac{1}{8}D)}.\end{aligned}$$

Again the constant C in the above inequality is independent of d.

By the triangle inequality and Theorem 2 in [23],

$$\|\tilde{g}_h\|_{W_1^1(D\backslash\frac{1}{8}D)} \leq \|\tilde{g}_h - g\|_{W_1^1(\Omega)} + \|g\|_{W_1^1(D\backslash\frac{1}{8}D)} \leq C + \|g\|_{W_1^1(D\backslash\frac{1}{8}D)}.$$

Since for some fixed $c > 0$ $\operatorname{dist}(x, \operatorname{supp}(\delta)) \geq cd$ for all $x \in D\backslash\frac{1}{8}D$, we have from Lemma 3.2 that for any such x,

$$\nabla g(x) = \int_{\tau_0} \nabla_x G(x,y)\partial\delta(y)dy = -\int_{\tau_0} \nabla_x \partial_y G(x,y)\delta(y)dy \leq Cd^{-n}.$$

As a result,

$$\|g\|_{W_1^1(D\backslash\frac{1}{8}D)} \leq C.$$

Collecting the above estimates, we thus have that

$$\|\nabla\tilde{g}_h\|_{L_2(\frac{7}{8}D\backslash\frac{1}{4}D)} + d^{-1}\|\tilde{g}_h\|_{L_2(\frac{7}{8}D\backslash\frac{1}{4}D)} \leq Cd^{-\frac{n}{2}}. \tag{4.10}$$

Collecting (4.10) and (4.7) into (4.6) yields

$$\|\nabla e\|_{L_\infty(\Omega)} = |\nabla e(x_0)| \leq C\left(d^{-1}\|u\|_{L_\infty(2D)} + \|\nabla u\|_{L_\infty(2D)} + d^{-\frac{n}{2}-1}\|e\|_{L_2(2D)}\right).$$

We complete the proof of Theorem 1 by inserting $u - \chi$ and $u_h - \chi$ for u and u_h and writing D instead of $2D$. □

Corollary 1. *Assume in addition to the assumptions of Theorem* 1 *that either* $D \subset\subset \Omega$ *or* $\text{meas}^{n-1}(\overline{D} \cap \partial\Omega)$ *is sufficiently large. Then, for any* $\chi \in S_h$

$$|\nabla e(x_0)| \leq C\left(\|\nabla(u-\chi)\|_{L_\infty(D)} + d^{-\frac{n}{2}-1}\|e\|_{L_2(D)}\right),\ n = 2, 3.$$

Proof. If $D \subset\subset \Omega$, then by taking $\chi_* = \chi + \frac{1}{|D|}\int_D(u-\chi)$ in Theorem 1 and applying the Poincaré inequality we have

$$\|u - \chi_*\|_{L_\infty(D)} \leq Cd\|\nabla(u-\chi_*)\|_{L_\infty(D)} = Cd\|\nabla(u-\chi)\|_{L_\infty(D)}.$$

Similarly, in the second case we can apply the Poincaré-Friedrichs inequality since $u - \chi$ vanishes on $\partial\Omega$. □

Remark 2. If Ω is a nonconvex polyhedral domain, our proof technique can also be modified in order to prove similar local estimates near convex portions of $\partial\Omega$ under reasonable assumptions. Let $B_d(x_0) \cap \Omega$ be convex with $\partial B_d(x_0) \cap \partial\Omega \neq \emptyset$. Let $\widetilde{\Omega}$ be a convex polyhedral subset of Ω with $B_d(x_0) \cap \Omega \subset \widetilde{\Omega}$. Assume then that the quasi-uniform mesh on $B_d(x_0) \cap \Omega$ can be extended to a quasi-uniform mesh covering $\widetilde{\Omega}$. In this case essentially the same proof as before works. The only place where modifications are needed is in the definition of g in (4.5), where we need to use $\tilde{\Omega}$ instead of Ω. So long as a finite number of extended subdomains $\widetilde{\Omega}$ are used to cover convex portions of Ω, which is always possible, the corresponding Green's functions estimates of Lemma 3.2 will hold with uniform constants over the extended domains $\tilde{\Omega}$.

4.3. **Part 2. Transforming the "slush" term.** The next step toward establishing (1.3) is to treat $\|e\|_{L_2(\Omega_0)}$, where Ω_0 is a subdomain containing x_0. We will employ the following lemma.

Lemma 4.1. *Let D be a bounded domain with Lipschitz boundary and having diameter d, and assume that v has mean value zero over D. Then*

$$\|v\|_{L_2(D)} \leq C\|\nabla v\|_{H^{-1}(D)},$$

where C depends only on the space dimension n and the ratio of d and the largest ball that can be inscribed in D.

Proof. We will prove this result by a duality argument. Let $\vec{w} \in H_0^1(D)^n$ be a solution to the following problem:

$$\begin{aligned} \nabla \cdot \vec{w} &= v \quad \text{in } D, \\ \vec{w} &= 0 \quad \text{on } \partial D. \end{aligned}$$

By Lemma 3.1 of Chapter III.3 in [20], there is a constant C independent of v and depending only on the ratio of d and the radius of the largest ball that can be inscribed into D such that

$$\|\vec{w}\|_{H_0^1(D)} \leq C\|v\|_{L_2(D)}. \tag{4.11}$$

Thus, integrating by parts and using the estimate above we have,

$$\begin{aligned} \|v\|_{L_2(D)}^2 &= (v, v)_D = (\nabla \cdot \vec{w}, v)_D = -(\vec{w}, \nabla v)_D \\ &\leq \|\vec{w}\|_{H_0^1(D)}\|\nabla v\|_{H^{-1}(D)} \leq C\|v\|_{L_2(D)}\|\nabla v\|_{H^{-1}(D)}. \end{aligned}$$

Canceling $\|v\|_{L_2(D)}$ on both sides we obtain the lemma. □

Remark 3. Let ρ denote the radius of the largest ball B that can be inscribed in D. Duran in [17] constructed a simple example on a rectangular domain showing that the dependence of the constant C on $(d/\rho)^s$ in (4.11) is necessary. However, the power s given in [20] is not optimal. The best constant so far is obtained in [17] and has the form

$$C = C_n \frac{d}{\rho} \left(\frac{|D|}{|B|}\right)^{\frac{n-2}{2(n-1)}} \left(\log \frac{|D|}{|B|}\right)^{\frac{n}{2(n-1)}},$$

where the constant C_n depends on n only.

When $\Omega_0 \subset\subset \Omega$ we may simply replace e by $e-\overline{e}$, where $\overline{e} = (e, 1)_{\Omega_0}$, in Corollary 1 and apply the above lemma. In the case where $\partial\Omega_0 \cap \partial\Omega \neq \emptyset$ we can prove a slightly different result, which again expresses $\|e\|_{L_2(\Omega_0)}$ in terms of derivatives in a weaker norm.

If $\Omega_0 \cap \partial\Omega \neq \emptyset$, we first select a straight $(n-1)$-dimensional hyperplane H containing the portion of $\partial\Omega$ lying closest to x_0. Without loss of generality, we can assume that H is given by $x_n = 0$. Then using odd reflection of the function u across H with respect to Ω_0 we can show the following.

Lemma 4.2. *Assume that* $e(x_1, \dots, x_{n-1}, 0) = 0$. *Then, there exists a constant* C *independent of* u *and* x_0 *such that*

$$\|e\|_{L_2(\Omega_0)} \leq C \sup_{\substack{\vec{\phi} \in H^1_{>}(\Omega_0)^n \\ \|\vec{\phi}\|_{H^1(\Omega_0)}=1}} \big(e, \nabla \cdot \vec{\phi}\big),$$

where $H^1_{>}(\Omega_0)$ *denotes the functions that vanish on all parts of* $\partial\Omega_0$ *except* $x_n = 0$.

Proof. Define

$$B^- = \{(x_1, \dots, x_{n-1}, x_n) \in \Omega_0\}, \quad B^+ = \{(x_1, \dots, x_{n-1}, -x_n) \in \Omega_0\}.$$

Put $\widehat{x} = (x_1, \dots, x_{n-1})$. Extend the function $e(x_1, \dots, x_n)$ from B^- to $B = B^- \cup B^+$ by the odd reflection

$$\widetilde{e}(\widehat{x}, x_n) = \begin{cases} e(\widehat{x}, x_n), & x \in B^- \\ -e(\widehat{x}, -x_n), & x \in B^+. \end{cases}$$

Since e has mean zero over B, by Lemma 4.1,

$$\|e\|_{L_2(B^-)} \leq \|\widetilde{e}\|_{L_2(B)} \leq C\|\nabla\widetilde{e}\|_{H^{-1}(B)}.$$

Here C does not depend on $x_0 \in \Omega$ since $\Omega_0 = B_d(x_0) \cap \Omega$ has uniformly bounded ratio of diameter to radius of largest inscribable ball. Now our goal is to express the last term in terms of e over the original domain B^-. Since

$$\|\nabla\widetilde{e}\|_{H^{-1}(B)} = \sup_{\substack{\vec{\phi} \in H^1_0(B)^n \\ \|\vec{\phi}\|_{H^1(B)}=1}} (\nabla\widetilde{e}, \vec{\phi})_B,$$

we consider separately

$$\left(\frac{\partial\widetilde{e}}{\partial x_i}, \phi_i\right)_B, \quad i = 1, \dots, n-1 \quad \text{and} \quad \left(\frac{\partial\widetilde{e}}{\partial x_n}, \phi_n\right)_B,$$

where $\vec{\phi} = (\phi_1, \dots, \phi_n)^T$. Put

$$\phi_i^{\text{odd}}(\widehat{x}, x_n) = \frac{\phi_i(\widehat{x}, x_n) - \phi_i(\widehat{x}, -x_n)}{2}$$

and

$$\phi_i^{\text{even}}(\widehat{x}, x_n) = \frac{\phi_i(\widehat{x}, x_n) + \phi_i(\widehat{x}, -x_n)}{2}.$$

Using that $\phi_i = \phi_i^{\text{odd}} + \phi_i^{\text{even}}$ and noting that $\frac{\partial\widetilde{e}}{\partial x_i}$ is an odd function with respect to the last variable x_n, for each $1 \leq i \leq n-1$ we have,

$$\left(\frac{\partial\widetilde{e}}{\partial x_i}, \phi_i\right)_B = \left(\frac{\partial\widetilde{e}}{\partial x_i}, \phi_i^{\text{odd}}\right)_B = 2\left(\frac{\partial e}{\partial x_i}, \phi_i^{\text{odd}}\right)_{B^-} = -2\left(e, \frac{\partial\phi_i^{\text{odd}}}{\partial x_i}\right)_{B^-}.$$

There are no boundary terms since $\phi_i^{\text{odd}} = 0$ on ∂B^-.

Similarly, since $\frac{\partial \widetilde{e}}{\partial x_n}$ is now an even function with respect to the last variable x_n, we have

$$\left(\frac{\partial \widetilde{e}}{\partial x_n}, \phi_n\right)_B = \left(\frac{\partial \widetilde{e}}{\partial x_n}, \phi_n^{\text{even}}\right)_B = 2\left(\frac{\partial e}{\partial x_n}, \phi_n^{\text{even}}\right)_{B^-} = -2\left(e, \frac{\partial \phi_n^{\text{even}}}{\partial x_n}\right)_{B^-}.$$

Again, there are no boundary terms since $e\phi_n^{\text{even}} = 0$ on ∂B^-. Thus we have shown that

$$\|e\|_{L_2(B^-)} \leq \sup_{\substack{\vec{\phi} \in H_0^1(B)^n \\ \|\vec{\phi}\|_{H^1(B)}=1}} -2\left(e, \sum_{i=1}^{n-1} \frac{\partial \phi_i^{\text{odd}}}{\partial x_i} + \frac{\partial \phi_n^{\text{even}}}{\partial x_n}\right)_{B^-}$$

That concludes the proof of Lemma 4.2. □

4.4. **Part 3. Partition of the domain and kickback argument.** We use the following decomposition of Ω. Let $d_0 = ph_0\ell_h$ and $d_j = q^{j-1}d_0$, where p and q are as in Mesh Condition 1. Then we have

$$\Omega = \Omega_0 \cup \bigcup_{j=0}^{J-1} \Omega_j \cup \bigcup_{k=0}^{K-1} A_k,$$

where

$$\begin{aligned}
\Omega_0 &= \{x \in \Omega : |x - x_0| \leq d_0\}, \\
\Omega_j &= \{x \in \Omega : d_j \leq |x - x_0| \leq d_{j+1}\}, \quad j = 1, 2, \ldots, J, \\
A_k &= \{x \in \Omega : 2^k d_J \leq |x - x_0| \leq 2^{k+1} d_J\}, \quad k = 0, 1, \ldots, K,
\end{aligned}$$

where J is the smallest integer such that $q^J h_0 \geq \overline{h}$ and K is the smallest integer such that $2^K d_J \geq \operatorname{diam}(\Omega)$. Notice that if we have a quasi-uniform mesh or $h_0 = \overline{h}$, then we can take $J = 0$ and start the dyadic decomposition right away. The next proposition establishes a bound for the mesh function on Ω_j.

Proposition 4.1. *In each Ω_j the mesh size cannot be larger than $q^j h_0$.*

Proof. We will prove this result by induction. Obviously, the statement holds on Ω_0. Assume it holds for Ω_i, $i = 1, 2, \ldots, j$ but not for Ω_{j+1}. Then there exists a point $\overline{x} \in \Omega_{j+1}$ such that $h(\overline{x}) > q^{j+1}h_0$. Take a ball of radius d_{j+1} centered at $\overline{x}$. Call it $B(\overline{x}, d_{j+1})$. This ball will intersect Ω_j. Choose a point $\overline{\overline{x}} \in B(\overline{x}, d_{j+1}) \cap \Omega_j$. By the induction assumption $h(\overline{\overline{x}}) \leq q^j h_0$. On the other hand, by Mesh Condition 1 we have that for all y satisfying $|\overline{x} - y| \leq q^{j+1}d_0 < p\ell_h h(\overline{x})$, $h(y) \geq h(\overline{x})/q > q^j h_0$. For $y = \overline{\overline{x}}$ we get a contradiction. □

We are finally ready to establish our main result.

Theorem 2. *Let u and u_h satisfy* (1.1) *and* (1.2) *with u_h lying in a finite element space satisfying the assumptions of Section* 3 *and defined on a mesh satisfying Mesh Condition* 1*, and assume that Ω is a convex polygonal or polyhedral domain in $\mathbb{R}^n$, $n = 2, 3$. Then there exists a constant C independent of h and u such that*

$$\|\nabla e\|_{L_\infty(\Omega)} \leq C \min_{\chi \in S_h} \|\nabla(u - \chi)\|_{L_\infty(\Omega)}.$$

Proof. **Step 1: Initial decomposition of the error** Let $\|\nabla e\|_{L_\infty(\Omega)} = |e(x_0)|$. We assume that $\operatorname{meas}^{n-1}(\overline{\Omega}_0 \cap \partial\Omega)$ is sufficiently large. The case where $\Omega_0 \subset\subset \Omega$ is slightly easier, and we omit the details. The intermediate case where x_0 is close to $\partial\Omega$ but $\text{meas}^{n-1}(\overline{\Omega}_0 \cap \partial\Omega)$ is small can be avoided by doubling d_0 and reverting to the boundary case; this can always be done with appropriate adjustment of constants.

From Lemma 4.2, we have

$$\|e\|_{L_2(\Omega_0)} \le C \sup_{\substack{\vec{\phi}\in H^1_{>}(\Omega_0)^n \\ \|\vec{\phi}\|_{H^1(\Omega_0)}=1}} \big(e, \nabla \cdot \vec{\phi}\big).$$

For each such fixed ϕ_i, $1 \le i \le n$, we define v by

$$\begin{aligned} -\Delta v &= \frac{\partial \phi_i}{\partial x_i}, \text{ in } \Omega, \\ v &= 0, \text{ on } \partial\Omega. \end{aligned} \tag{4.12}$$

Using the above decomposition of Ω, (4.12), and the Galerkin orthogonality we have

$$\begin{aligned} -(e, \Delta v) &= (\nabla e, \nabla v) = (\nabla e, \nabla(v - v_I)) \\ &\le \|\nabla e\|_{L_\infty(\Omega)} \Bigg(\|\nabla(v - v_I)\|_{L_1(\Omega_0 \cup \Omega_1)} + \sum_{j=2}^{J} \|\nabla(v - v_I)\|_{L_1(\Omega_j)} \\ &\quad + \sum_{k=0}^{K} \|\nabla(v - v_I)\|_{L_1(A_k)} \Bigg) = I_1 + I_2 + I_3. \end{aligned}$$

Step 2: Estimate for I_1. Using approximation theory, the Cauchy-Schwarz inequality, and Lemma 3.1, we have

$$\begin{aligned} I_1 &\le \|\nabla e\|_{L_\infty(\Omega)} \|\nabla(v - v_I)\|_{L_1(\Omega_0 \cup \Omega_1)} \\ &\le C\|\nabla e\|_{L_\infty(\Omega)} h_0 (1+q) d_0^{\frac{n}{2}} (1+q) \|D^2 v\|_{L_2(\Omega)} \\ &\le C\|\nabla e\|_{L_\infty(\Omega)} h_0 (1+q)^2 d_0^{\frac{n}{2}} \|\frac{\partial \phi_i}{\partial x_i}\|_{L_2(\Omega)} \\ &\le C\|\nabla e\|_{L_\infty(\Omega)} h_0 (1+q)^2 d_0^{\frac{n}{2}}. \end{aligned}$$

Step 3: Estimate for I_2. To estimate I_2 we will use the following result.

Proposition 4.2. *Let v be the solution of* (4.12)*, then*

$$\|\nabla v\|_{L_2(\Omega_j)} + \frac{1}{d_j} \|v\|_{L_2(\Omega_j)} \le C d_j^{-\frac{n}{2}} d_0^{\frac{n}{2}+1}, \quad j \ge 2.$$

Proof. Using the Green's function representation, we have

$$v(x) = \int_{\Omega_0} G(x,y) \frac{\partial \phi_i}{\partial y_i} dy = -\int_{\Omega_0} \frac{\partial G(x,y)}{\partial y_i} \phi_i dy.$$

There are no boundary terms since either ϕ_i or $G(x,y)$ vanish on $\partial\Omega_0$.

By the Green's function estimates from Lemma 3.2 and using that $\text{dist}(\Omega_0, \Omega_j) \approx d_j$ for $j \ge 2$, we obtain

$$|\nabla v(x)| \le C d_j^{-n} \|\phi_i\|_{L_1(\Omega_0)}.$$

Using the Poincaré inequality and $\|\phi_i\|_{H^1} \le C$, we have

$$\|\phi_i\|_{L_1(\Omega_0)} \le C d_0 \|\nabla \phi_i\|_{L_1(\Omega_0)} \le C d_0^{\frac{n}{2}+1} \|\nabla \phi_i\|_{L_2(\Omega)} \le C d_0^{\frac{n}{2}+1}.$$

Thus,

$$|\nabla v(x)| \le C d_j^{-n} d_0^{\frac{n}{2}+1}$$

and

$$\|\nabla v\|_{L_2(\Omega_j)} \le C d_j^{-\frac{n}{2}} d_0^{\frac{n}{2}+1}.$$

Very similarly we can obtain

$$\|v\|_{L_2(\Omega_j)} \le C d_j^{-\frac{n}{2}+1} d_0^{\frac{n}{2}+1}. \qquad \square$$

Approximation theory then yields

$$\|\nabla(v - v_I)\|_{L_1(\Omega_j)} \le C h_j q \|D^2 v\|_{L_1(\Omega_j)} \le C h_0 q^{j+1} d_j^{\frac{n}{2}} \|D^2 v\|_{L_2(\Omega_j)}.$$

Using that v is harmonic on Ω_j', by Lemma 3.3, we obtain

$$\|D^2 v\|_{L_2(\Omega_j)} \le C d_j^{-2} \|v\|_{L_2(\Omega_j')}.$$

Using Proposition 4.2 we have

$$\|\nabla(v - v_I)\|_{L_1(\Omega_j)} \le C h_0 q^{j+1} d_j^{-1} d_0^{\frac{n}{2}+1}.$$

Thus, we can estimate I_2 as

$$I_2 \le C h_0 d_0^{\frac{n}{2}+1} \|\nabla e\|_{L_\infty(\Omega)} \sum_{j=1}^{J} \frac{q^j}{d_j} \le C h_0 d_0^{\frac{n}{2}} \ell_h \|\nabla e\|_{L_\infty(\Omega)},$$

where in the last step we used that $d_j = q^{j-1} d_0$ and $J \le C\ell_h$.

Step 4: Estimate for I_3. Using that the maximum mesh size on A_k is at most $\overline{h}$ by the approximation theory, Cauchy-Schwarz inequality, and Lemma 3.3 we have

$$\|\nabla(v - v_I)\|_{L_1(A_k)} \le C\overline{h} \|D^2 v\|_{L_1(A_k)} \le C\overline{h} (2^k d_J)^{\frac{n}{2}-2} \|v\|_{L_2(A_k)}.$$

By Proposition 4.2, with A_k in place of Ω_j we have

$$\|v\|_{L_2(A_k)} \le C d_0^{\frac{n}{2}+1} (2^k d_J)^{1-\frac{n}{2}}.$$

Hence using that $q^J \ge \overline{h}/h_0$,

$$I_3 \le C\overline{h} d_0^{\frac{n}{2}+1} d_J^{-1} \|\nabla e\|_{L_\infty(\Omega)} \sum_{k=0}^{K} 2^{-k} \le C h_0 d_0^{\frac{n}{2}} \|\nabla e\|_{L_\infty(\Omega)}.$$

Step 5: Finish of the proof. Kickback. Using Corollary 1 with $D = \Omega_0$, we have

$$\|\nabla e\|_{L_\infty(\Omega)} \le C \|\nabla(u - \chi)\|_{L_\infty(\Omega_0)} + C d_0^{-\frac{n}{2}-1} \|e\|_{L_2(\Omega_0)}. \tag{4.13}$$

Combining estimates for I_1, I_2, and I_3 we have

$$\|e\|_{L_2(\Omega_0)} \le C h_0 d_0^{\frac{n}{2}} \ell_h \|\nabla e\|_{L_\infty(\Omega)}. \tag{4.14}$$

Thus combining (4.13) and (4.14)

$$\|\nabla e\|_{L_\infty(\Omega)} \le C \|\nabla(u - \chi)\|_{L_\infty(\Omega_0)} + C \frac{h_0}{d_0} \ell_h \|\nabla e\|_{L_\infty(\Omega)}.$$

Recalling that

$$h_0/d_0 \le C/p\ell_h$$

and observing that if p is large enough we can kick back $C\frac{h_0}{d_0}\ell_h \|\nabla e\|_{L_\infty(\Omega)}$, we obtain the desired best approximation result. $\square$

5. Possible extensions

Here, we comment briefly on possible extensions of our results to include "localized" pointwise error estimates and alternative proof techniques.

In [30], pointwise error estimates having a sharply local character were proved. Assume that $\mathcal{T}$ is quasi-uniform of size h. Given a fixed $x_0 \in \Omega$, define $\sigma_{x_0}(y) = \frac{h}{h+|x_0-y|}$. The main result in [30] essentially says that when $\partial\Omega$ is smooth and certain further assumptions are satisfied,

$$|\nabla(u-u_h)(x_0)| \leq C\ell_{h,s} \min_{\chi \in S_h} \|\sigma_{x_0}^s \nabla(u-\chi)\|_{L_\infty(\Omega)}. \tag{5.1}$$

Here, $0 \leq s \leq k$ and $\ell_{h,s}$ is a logarithmic factor which disappears except when $s = k$; recall that k is the polynomial degree. In [23] it was remarked that similar localized estimates hold for convex polyhedral domains as well, except that the allowed range of s above is restricted by the maximum interior angle of $\partial\Omega$ as well as by the polynomial degree k. It is possible to prove a similar result here. In particular, let $\sigma_{x_0}(y) = \frac{\overline{h}}{\overline{h}+|y-x_0|}$. Then if the grid is locally quasi-uniform on balls of size $B_{p(\ln \underline{h})^\alpha h(x_0)}(x_0)$ with $\alpha > 0$ sufficiently large (depending on domain geometry), we have (5.1) with s restricted by k and the domain geometry as before.

The proof of (5.1) may be accomplished by mimicking the proof technique of [30], in particular, by employing local energy estimates over a dyadic decomposition of Ω about x_0 and then using a double kickback argument. The necessary local energy estimates which hold on merely shape regular grids are found in [14]. The proof thus obtained closely follows that of [30] in outline, but is significantly more technical because it must account for changes in mesh size. The proof we use here does not yield such sharply local estimates. It is, however, significantly shorter and also places nominally less restriction on the mesh grading than does the proof outlined above in many cases.

Acknowledgments

The authors would like to acknowledge the helpful comments of two anonymous referees.

References

[1] T. Apel, A. Rösch, and D. Sirch, *L^∞-error estimates on graded meshes with application to optimal control*, SIAM J. Control Optim., 48 (2009), pp. 1771–1796. MR2516188 (2010f:49064)

[2] I. Babuška and J. Osborn, *Analysis of finite element methods for second order boundary value problems using mesh dependent norms*, Numer. Math., 34 (1980), pp. 41–62. MR560793 (81g:65143)

[3] N. Y. Bakaev, M. Crouzeix, and V. Thomée, *Maximum-norm resolvent estimates for elliptic finite element operators on nonquasiuniform triangulations*, M2AN Math. Model. Numer. Anal., 40 (2006), pp. 923–937 (2007). MR2293252 (2007k:35194)

[4] N. Y. Bakaev, V. Thomée, and L. B. Wahlbin, *Maximum-norm estimates for resolvents of elliptic finite element operators*, Math. Comp., 72 (2003), pp. 1597–1610 (electronic). MR1986795 (2004g:65074)

[5] J. H. Bramble, J. E. Pasciak, and O. Steinbach, *On the stability of the L^2 projection in $H^1(\Omega)$*, Math. Comp., 71 (2002), pp. 147–156 (electronic). MR1862992 (2002h:65175)

[6] S. C. Brenner and L. R. Scott, *The Mathematical Theory of Finite Element Methods*, Springer-Verlag, Berlin, Heidelberg, New York, third ed., 2008. MR2373954 (2008m:65001)

[7] C. Carstensen, *Merging the Bramble-Pasciak-Steinbach and the Crouzeix-Thomée criterion for H^1-stability of the L^2-projection onto finite element spaces*, Math. Comp., 71 (2002), pp. 157–163 (electronic). MR1862993 (2002i:65125)

[8] J. M. Cascon, C. Kreuzer, R. H. Nochetto, and K. G. Siebert, *Quasi-optimal convergence rate for an adaptive finite element method*, SIAM J. Numer. Anal., 46 (2008), pp. 2524–2550. MR2421046 (2009h:65174)

[9] M. Crouzeix and V. Thomée, *The stability in L_p and W_p^1 of the L_2-projection onto finite element function spaces*, Math. Comp., 48 (1987), pp. 521–532. MR878688 (88f:41016)

[10] E. Dari, R. G. Durán, and C. Padra, *Maximum norm error estimators for three-dimensional elliptic problems*, SIAM J. Numer. Anal., 37 (2000), pp. 683–700 (electronic). MR1740762 (2001b:65120)

[11] K. Deckelnick and M. Hinze, *Convergence of a finite element approximation to a state-constrained elliptic control problem*, SIAM J. Numer. Anal., 45 (2007), pp. 1937–1953 (electronic). MR2346365 (2008k:49005)

[12] A. Demlow, *Localized pointwise a posteriori error estimates for gradients of piecewise linear finite element approximations to second-order quadilinear elliptic problems*, SIAM J. Numer. Anal., 44 (2006), pp. 494–514 (electronic). MR2218957 (2007c:65105)

[13] ———, *Sharply localized pointwise and W_∞^{-1} estimates for finite element methods for quasilinear problems*, Math. Comp., 76 (2007), pp. 1725–1741 (electronic).

[14] A. Demlow, J. Guzmán, and A. Schatz, *Local energy estimates for the finite element method on sharply varying grids*, Math. Comp., 80 (2011), 1–9. MR2728969

[15] A. Demlow and R. Stevenson, *Convergence and quasi-optimality of an adaptive finite element method for controlling L_2 errors*, Numer. Math., 117 (2011), no. 2, 185–218. MR2754849

[16] J. Douglas, Jr. and T. Dupont, *A Galerkin method for a nonlinear Dirichlet problem*, Math. Comp., 29 (1975), pp. 689–696. MR0431747 (55:4742)

[17] R. G. Durán, *An elementary proof of the continuity from $L_0^2(\Omega)$ to $H_0^1(\Omega)^n$ of Bogovskii's right inverse of the divergence*, preprint.

[18] K. Eriksson, *An adaptive finite element method with efficient maximum norm error control for elliptic problems*, Math. Models Methods Appl. Sci., 4 (1994), pp. 313–329. MR1282238 (95c:65180)

[19] S. J. Fromm, *Potential space estimates for Green potentials in convex domains*, Proc. Amer. Math. Soc., 119 (1993), pp. 225–233. MR1156467 (93k:35076)

[20] G. P. Galdi, *An Introduction to the Mathematical Theory of the Navier–Stokes Equations. Volume I. Linearized Steady Problems*, Springer Tracts in Natural Philosophy, Vol. 38, Springer-Verlag, Berlin, Heidelberg, New York, 1994. MR1284205 (95i:35216a)

[21] P. Grisvard, *Elliptic Problems in Nonsmooth Domains*, Pitman, Boston, London, Melbourne, 1985. MR775683 (86m:35044)

[22] M. Grüter and K.-O. Widman, *The Green function for uniformly elliptic equations*, Manuscripta Math., 37 (1982), pp. 303–342. MR657523 (83h:35033)

[23] J. Guzmán, D. Leykekhman, J. Rossmann, and A. H. Schatz, *Hölder estimates for Green's functions on convex polyhedral domains and their applications to finite element methods*, Numer. Math., 112 (2009), pp. 221–243. MR2495783 (2010a:65237)

[24] Y. He, J. Xu, A. Zhou, and J. Li, *Local and parallel finite element algorithms for the Stokes problem*, Numer. Math., 109 (2008), pp. 415–434. MR2399151 (2009j:76157)

[25] J. A. Nitsche and A. H. Schatz, *Interior estimates for Ritz-Galerkin methods*, Math. Comp., 28 (1974), pp. 937–958. MR0373325 (51:9525)

[26] R. H. Nochetto, *Pointwise a posteriori error estimates for elliptic problems on highly graded meshes*, Math. Comp., 64 (1995), pp. 1–22. MR1270622 (95c:65172)

[27] R. H. Nochetto, M. Paolini, and C. Verdi, *An adaptive finite element method for two-phase Stefan problems in two space dimensions. I. Stability and error estimates*, Math. Comp., 57 (1991), pp. 73–108, S1–S11. MR1079028 (92a:65322)

[28] R. Rannacher and R. Scott, *Some optimal error estimates for piecewise linear finite element approximations*, Math. Comp., 38 (1982), pp. 437–445. MR645661 (83e:65180)

[29] P. Saavedra and L. R. Scott, *Variational formulation of a model free-boundary problem*, Math. Comp., 57 (1991), pp. 451–475. MR1094958 (92a:35166)

[30] A. H. Schatz, *Pointwise error estimates and asymptotic error expansion inequalities for the finite element method on irregular grids. I. Global estimates*, Math. Comp., 67 (1998), pp. 877–899. MR1464148 (98j:65082)

[31] ———, *Maximum norm error estimates for the finite element method allowing highly refined grids*, in Recent advances in adaptive computation, vol. 383 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2005, pp. 133–139. MR2195798

[32] A. H. Schatz and L. B. Wahlbin, *Interior maximum norm estimates for finite element methods*, Math. Comp., 31 (1977), pp. 414–442. MR0431753 (55:4748)
[33] ———, *Maximum norm estimates in the finite element method on plane polygonal domains. I*, Math. Comp., 32 (1978), pp. 73–109. MR0502065 (58:19233a)
[34] ———, *Maximum norm estimates in the finite element method on plane polygonal domains. II. Refinements*, Math. Comp., 33 (1979), pp. 465–492. MR0502067 (58:19233b)
[35] ———, *Interior maximum-norm estimates for finite element methods. II*, Math. Comp., 64 (1995), pp. 907–928. MR1297478 (95j:65143)
[36] R. Scholz, *A mixed method for 4th order problems using linear finite elements*, RAIRO Anal. Numér., 12 (1978), pp. 85–90, iii. MR0483557 (58:3549)
[37] R. Stevenson, *The completion of locally refined simplicial partitions created by bisection*, Math. Comp., 77 (2008), pp. 227–241 (electronic). MR2353951 (2008j:65219)
[38] L. B. Wahlbin, *Local behavior in finite element methods*, in Handbook of numerical analysis, Vol. II, Handb. Numer. Anal., II, North-Holland, Amsterdam, 1991, pp. 353–522. MR1115238
[39] J. Xu and A. Zhou, *Local and parallel finite element algorithms based on two-grid discretizations*, Math. Comp., 69 (2000), pp. 881–909. MR1654026 (2000j:65102)

Department of Mathematics, University of Kentucky, Lexington, Kentucky 40506
E-mail address: alan.demlow@uky.edu.

Department of Mathematics, University of Connecticut, Storrs, Connecticut 06269
E-mail address: leykekhman@math.uconn.edu

Department of Mathematics, Cornell University, Ithaca, New York 14853.
E-mail address: schatz@math.cornell.edu

Department of Mathematics, Cornell University, Ithaca, New York 14853.
E-mail address: wahlbin@math.cornell.edu