Convergence and quasi-optimality of an adaptive finite element method for controlling L_2 errors

Alan Demlow · Rob Stevenson

Abstract　In this paper, a contraction property is proved for an adaptive finite element method for controlling the global L_2 error on convex polyhedral domains. Furthermore, it is shown that the method converges in L_2 with the best possible rate. The method that is analyzed is the standard adaptive method except that, if necessary, additional refinements are made to keep the meshes sufficiently mildly graded. This modification does not compromise the quasi-optimality of the resulting algorithm.

Mathematics Subject Classification (2000)　65N15 · 65N30

1 Introduction and results

Consider the elliptic model problem

$$-\Delta u = f \ \text{ in } \ \Omega,$$
$$u = 0 \ \text{ on } \ \partial\Omega, \tag{1.1}$$

A. Demlow (✉)
Department of Mathematics, University of Kentucky,
715 Patterson Office Tower, Lexington, KY 40506-0027, USA
e-mail: demlow@ms.uky.edu

R. Stevenson
Korteweg-de Vries (KdV) Institute for Mathematics, University of Amsterdam,
P.O. Box 94248, 1090 GE Amsterdam, The Netherlands
e-mail: R.P.Stevenson@uva.nl

where $\Omega \subset \mathbb{R}^n$, $n \geq 2$, is a convex polyhedral domain and $f \in L_2(\Omega)$. Convexity is imposed so that the (dual) problem is $H^2(\Omega)$ regular, which is used to show reliability of the residual based a posteriori $L_2(\Omega)$ error estimator. The results presented here can be generalized to other linear second order elliptic operators having sufficiently smooth coefficients, but we present the simplest case in order to focus on essential ideas.

In this work we prove contraction and quasi-optimality properties for an adaptive finite element method (AFEM) for controlling errors in the $L_2(\Omega)$-norm. An AFEM is an iterative feedback procedure of the form

$$\mathsf{solve} \to \mathsf{estimate} \to \mathsf{mark} \to \mathsf{refine}. \tag{1.2}$$

Such adaptive algorithms have for many years been a standard tool for efficiently approximating solutions to partial differential equations such as (1.1). The convergence properties of AFEM have become the subject of intense theoretical study only in the past few years, however. We refer to [6,11–14] for an overview of progress in basic convergence theory for AFEM for linear elliptic problems. Optimal convergence rates were demonstrated in [4,19]. A common feature of most theoretical results published to date concerning convergence and quasi-optimality of AFEM is that the error estimated in the "$\mathsf{estimate}$" step in (1.2) is the global energy error. In particular, AFEM optimality results for error notions whose analysis require non-trivial duality arguments have not to our knowledge appeared in the literature. AFEM for controlling other norms are sometimes of practical interest, however; in the current context we refer for example to [5,9,24] where "pollution effects" of global solution properties on the local energy error are measured and controlled adaptively in L_2.

In order to describe our results, let $(\mathcal{T}_i)_{i \geq 0}$ be a nested sequence of conforming, uniformly shape regular partitions of Ω produced by the AFEM. In this work, we consider simplicial partitions as they are produced by the *newest vertex bisection* algorithm or by its generalization to more than two dimensions. Let $S_i \subset H_0^1(\Omega)$ be the standard Lagrange finite element space of some fixed degree k on $\mathcal{T}_i$, and let $u_i \in S_i$ satisfy

$$A(u_i, v_i) := \int_\Omega \nabla u_i \cdot \nabla v_i \, \mathrm{d}x = \int_\Omega f v_i \, \mathrm{d}x, \quad (v_i \in S_i). \tag{1.3}$$

We also let $|||v||| := \sqrt{A(v,v)}$ and $|||v|||_D := \left(\int_D |\nabla v|^2 \, \mathrm{d}x \right)^{1/2}$ denote the global and local energy (semi)norms over D, respectively.

Our first goal in this work is to prove that the AFEM for controlling the L_2-error $\|u - u_i\|_{L_2(\Omega)}$ is majorized linearly convergent. In order to obtain this result, we require that the sequence $(\mathcal{T}_i)_i$ of partitions is sufficiently mildly graded. We will modify the "refine" routine so that, if necessary, it bisects additional simplices in order to preserve the mild grading of the partition. With this modification, and under the assumption that Ω is convex, in a sequence of estimates we prove that the quantities $\|u - u_i\|_{L_2(\Omega)}^2 + \mathrm{osc}_i^2$, $|||h_i(u - u_i)|||^2 + \mathrm{osc}_i^2$ and η_i^2 are all uniformly equivalent. Here osc_i is the L_2-oscillation term, $h_i \in W_\infty^1(\Omega)$ is a regularized local mesh size function, and η_i is the residual based L_2 a posteriori error estimator. A key ingredient

of many of our proofs is that thanks to the mild grading of the partitions, the mesh size function can be designed so that $\|\nabla h_i\|_{L_\infty(\Omega)}$ is sufficiently small. Following an idea from [4], we then show that a suitably weighted sum of $|||h_i(u - u_i)|||^2$ and η_i^2 is contracted by the AFEM, which implies the convergence result.

The approach of treating the L_2-error as a *weighted* energy error and then employing techniques developed for proving convergence of AFEM for global energy errors was used earlier in [5] to prove convergence of an AFEM for controlling local energy errors. We note that the restriction that $\|\nabla h_i\|_{L_\infty(\Omega)}$ must be sufficiently small has previously appeared in the literature in connection with a priori and a posteriori estimates in L_∞ (cf. [7,8]) and a priori estimates in L_2 (cf. [3]).

Convergence of AFEM for controlling L_2-errors was also addressed in [13]. There are several substantial differences between our approach and the one taken in the latter work. Morin et al. [13] obtains convergence of AFEM for controlling weak norms under quite general assumptions on the marking strategy and norm of interest. While the assumptions are general, the convergence result obtained is correspondingly weak in that no estimate of the rate of convergence is obtained. In contrast, we require stronger assumptions; aside from the above-mentioned mesh restriction we also require a Dörfler-type marking strategy. The result is a much stronger convergence theory.

The second main theorem of this work states that for the sequence of partitions produced by the AFEM, $\|u - u_i\|_{L_2(\Omega)} + \text{osc}_i$ can be bounded by some absolute multiple of $(\#\mathcal{T}_i)^{-s}$ for the best possible value of s. To arrive at this quasi-optimality result, we prove a localized a posteriori upper bound for the L_2-difference between Galerkin solutions on nested partitions. Furthermore, we show that, up to an oscillation term, the Galerkin solution is a near best approximation to u from the finite element space in the L_2-norm. The latter result is of some interest independent of the current context, since it is known that the Ritz projection is *not* stable in the L_2 norm (cf. [1]). Both of these results rely on the condition that $\|\nabla h_i\|_{L_\infty(\Omega)}$ is sufficiently small. Finally, we show that, even if additional bisections are made to preserve the mild grading of the partitions, at any stage in the AFEM the cardinality of the current partition can be bounded by the cardinality of the initial partition plus some absolute multiple of the number of all elements that so far were marked for refinement by the routine "mark".

We finally comment on the practicality of the AFEM for which we prove optimality. Existing AFEM optimality results for global energy norms require that an essential user-supplied parameter in the "mark" step be sufficiently small. The L_2 AFEM for which we prove optimality requires *two* user-supplied parameters to be sufficiently small, one in the "mark" and the other in the "refine" step. Whereas in the energy case an upper bound for the parameter in terms of interpolation (Poincaré) constants can in principal be derived theoretically, this will be harder in the L_2-case as the corresponding parameters additionally depend on H^2 regularity constants.

It should also be noted that enforcing mild mesh grading may exacerbate the asymptotic nature of our results, since meshes satisfying a mild grading assumption with small μ will essentially remain quasi-uniform over the first iterations of the adaptive algorithm. On the other hand, practical experience seems to indicate that optimality in L_2 is obtained without taking precautions to keep the mesh sufficiently mildly graded, although there is no proof of that. This might mean either that that mildly gradedness is not really needed, or that that the typical meshes resulting from the application of an

adaptive routine are automatically sufficiently mildly graded. Concerning the latter, note that a mesh that is optimal for controlling L_2 errors is more mildly graded than that for H^1 errors.

The paper is organized as follows. Sections 2 and 3 contain a number of preliminaries and definitions. In Sect. 4 we establish relationships between the error notions used in the paper. Our L_2 AFEM is defined precisely in Sect. 5. Sections 6 and 7 contain proofs of convergence and optimality of our AFEM, respectively. Section 8 contains discussion of the relationship between optimality of AFEM for controlling energy and L_2 norms and a numerical example. Finally, a bisection routine which preserves a mild mesh grading $\mathbb{T}$ given in Appendix A along with results concerning existence and properties of mesh functions.

2 Preliminaries

2.1 Partitions and mesh functions

Let $\mathcal{T}_0$ be a conforming partition of $\bar{\Omega}$ into (essentially) disjoint closed n-simplices. By fixing a local numbering of all vertices of all $T \in \mathcal{T}_0$, all possible descendants $\mathcal{T}$ of $\mathcal{T}_0$ that can be created by newest vertex bisection are uniquely determined. Here, with newest vertex bisection, we mean either the refinement procedure as it was developed in two space dimensions, or its generalization to any space dimension. Details can be found in Appendix A. The simplices in any of those partitions are *uniformly shape regular*, dependent only on the shape regularity parameters of $\mathcal{T}_0$ and the dimension n.

Generally, a descendant of $\mathcal{T}_0$ is non-conforming. Yet, with a suitable numbering of the vertices in the initial partition, any descendant can be refined to a conforming partition by inflating the total number of simplices by not more than some absolute multiple. Possibly after an initial refinement of the original initial partition, such a numbering always exists. Assuming such a numbering, we denote the set of all *conforming* descendants $\mathcal{T}$ of $\mathcal{T}_0$ by $\mathbb{T}$. For $\mathcal{T}, \tilde{\mathcal{T}} \in \mathbb{T}$, we will write $\mathcal{T} \subset \tilde{\mathcal{T}}$ when $\tilde{\mathcal{T}}$ is a refinement of $\mathcal{T}$ (or is equal to $\mathcal{T}$).

For $T \in \mathcal{T} \in \mathbb{T}$, let $h_T = |T|^{1/n}$. Furthermore, let ω_T, $\tilde{\omega}_T$ be the patches of elements (in $\mathcal{T}$) sharing a vertex or a facet (edge, face, etc.) with T. For $\mathcal{T} \in \mathbb{T}$, we construct the continuous piecewise linear "mesh-function" $h_{\mathcal{T}}$ by defining, for any vertex z of $\mathcal{T}$, $h_{\mathcal{T}}(z)$ as the average of the $h_{T'}$ over all $T' \in \mathcal{T}$ with $T' \ni z$. For some constants $c_{\mathbb{T}}$ and $C_{\mathbb{T}}$, it satisfies

$$c_{\mathbb{T}} h_T \leq h_{\mathcal{T}}|_T \leq C_{\mathbb{T}} h_T \quad (T \in \mathcal{T}, \mathcal{T} \in \mathbb{T}). \tag{2.1}$$

By the uniform shape regularity, for another constant $\hat{C}_{\mathbb{T}}$ we have

$$\|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \leq \hat{C}_{\mathbb{T}} \quad (\mathcal{T} \in \mathbb{T}).$$

For a number of our results, we will need a mesh function $h_\mathcal{T}$ that for some *sufficiently small, but fixed constant* $\mu > 0$ satisfies both

$$\|\nabla h_\mathcal{T}\|_{L_\infty(\Omega)} \leq \mu, \tag{2.2}$$

and (2.1) for some absolute constants $c_\mathbb{T}$ and $C_\mathbb{T}$ that are *independent of* μ. By an application of the mean value theorem, the existence of such a mesh function $h_\mathcal{T}$ implies that $h_T \leq c_\mathbb{T}^{-1}(C_\mathbb{T} h_{T'} + \mu \operatorname{dist}(T, T'))$ $(T, T' \in \mathcal{T})$, and so in particular that $\max_{T \in \mathcal{T}} h_T \leq c_\mathbb{T}^{-1}(C_\mathbb{T} \min_{T \in \mathcal{T}} h_T + \mu \operatorname{diam} \Omega)$. We conclude that the existence of such an $h_\mathcal{T}$ imposes a restriction on the grading of the partition $\mathcal{T}$ beyond that imposed by shape regularity alone.

In Appendix A, we show that given a parameter μ, there is a class of sufficiently mildly graded partitions $\mathcal{T} \in \mathbb{T}$ such that mesh functions $h_\mathcal{T}$ satisfying (2.1) independent of μ and (2.2) can be constructed. Given a $\mu > 0$, we denote the class of such partitions by $\mathbb{T}_\mu$. The initial partition $\mathcal{T}_0$, as any of its uniform refinements, has no grading and is therefore contained in $\mathbb{T}_\mu$ for any μ. Each time that we apply a mesh function $h_\mathcal{T}$ for a $\mathcal{T} \in \mathbb{T}_\mu$, obviously we mean the mesh function that satisfies (2.1) and (2.2), independent of μ. These mesh functions are constructed such that, additionally, they are pointwise non-increasing under refinements inside the class $\mathbb{T}_\mu$. Finally, it is shown that any $\mathcal{T} \in \mathbb{T}$ can be refined to a partition in $\mathbb{T}_\mu$ by inflating the number of simplices by not more than some fixed multiple, dependent on μ.

The restriction that $\|\nabla h_\mathcal{T}\|_{L_\infty(\Omega)}$ is sufficiently small appeared previously in [7], where a priori L_∞-estimates were proved under this assumption. There μ was allowed to depend logarithmically on $\min_{T \in \mathcal{T}_i} h_T$. In the introductory Chapter 0 of the standard text [3], a priori L_2-estimates are also proved under the assumption (2.2) in the one-dimensional case; higher-dimensional versions of this result require a slightly more involved technical development but may be obtained with a similar proof. To our knowledge, (2.2) is the least restrictive assumption under which optimal a priori bounds in the L_2 norm have appeared in the literature for general space dimension n.

Finally, it is pointed out in [7] that the graded meshes typically needed to resolve corner singularities occurring in elliptic problems on corner domains satisfy (2.2) with μ small. Thus although we modify the standard "refine" routine so that it bisects additional simplices in order to preserve the mild grading of the partition if necessary, it may be that as in our numerical example in Sect. 8 additional bisections never have to be made.

2.2 Finite element spaces and interpolants

Given $\mathcal{T} \in \mathbb{T}$, let $S_\mathcal{T} \subset H_0^1(\Omega)$ be a space of continuous Lagrange finite element functions which are piecewise polynomials of some fixed degree k on $\mathcal{T}$. As u always denotes the solution of the continuous problem (1.1), the notation $u_\mathcal{T}$ will be reserved for the solution of its Galerkin discretization $A(u_\mathcal{T}, v_\mathcal{T}) = \int_\Omega f v_\mathcal{T} dx$ $(v_\mathcal{T} \in S_\mathcal{T})$. Note that $S_\mathcal{T} \subset S_{\tilde{\mathcal{T}}}$ whenever $\mathcal{T} \subset \tilde{\mathcal{T}}$.

We will employ two standard finite element (quasi-)interpolators onto $S_\mathcal{T}$. The first is the Lagrange interpolator, which we denote by $I_{L,\mathcal{T}}$. We will also employ

the Scott–Zhang interpolator $I_{SZ,\mathcal{T}}$, which is stable in $H^1(\Omega)$, uniformly in $\mathcal{T} \in \mathbb{T}$ (cf. [18]). We do not list further properties of $I_{SZ,\mathcal{T}}$ here, as its application in the establishment of residual-type a posteriori upper bounds is now rather standard.

2.3 Regularity

Using that $f \in L_2(\Omega)$ and Ω is convex, the following H^2-regularity result is well-known.

Theorem 1 *The solution u of* (1.1) *satisfies* $u \in H^2(\Omega)$ *with*

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

2.4 Constants

In what follows, we shall denote by C and C_i ($i = 1, 2, 3 \ldots$) generic constants that may depend on the shape regularity parameters of $\mathcal{T}_0$, the space dimension n, the polynomial degree k, the constants $c_{\mathbb{T}}$ and $C_{\mathbb{T}}$ in (2.2), and the H^2-regularity constant C_{reg} from the theorem above. Other constants will be defined as necessary. In addition, we will often write $a \lesssim b$ instead of $a \leq Cb$ with C as above. Obviously, with $a \gtrsim b$ it is meant that $b \lesssim a$, and with $a \eqsim b$ that both $a \lesssim b$ and $a \gtrsim b$.

3 Residual based L_2 a posteriori error estimator

Given $\mathcal{T} \in \mathbb{T}, v_{\mathcal{T}} \in S_{\mathcal{T}}$, and $T \in \mathcal{T}$, we define the L_2-type error indicator $\eta(v_{\mathcal{T}}, T)$ by

$$\eta(v_{\mathcal{T}}, T)^2 = h_T^4 \|f + \Delta v_{\mathcal{T}}\|_{L_2(T)}^2 + h_T^3 \|[\![\nabla v_{\mathcal{T}}]\!]\|_{L_2(\partial T \setminus \partial \Omega)}^2. \tag{3.1}$$

We omit $v_{\mathcal{T}}$ in the above notation when $v_{\mathcal{T}} = u_{\mathcal{T}}$, that is, $\eta(T) = \eta(u_{\mathcal{T}}, T)$. Next we define error estimators. If $\mathcal{M} \subset \mathcal{T}$, we define

$$\eta(v_{\mathcal{T}}, \mathcal{M})^2 = \sum_{T \in \mathcal{M}} \eta(v_{\mathcal{T}}, T)^2,$$

If $v_{\mathcal{T}} = u_{\mathcal{T}}$, we shall omit the reference to $v_{\mathcal{T}}$ in our notation as above.

The following two results dealing with the reliability and efficiency of the estimator are well-known. Proofs can be found in [22]. The first result makes use of the fact that Ω is convex.

Proposition 1 *It holds that*

$$\|u - u_{\mathcal{T}}\|_{L_2(\Omega)} \lesssim \eta(\mathcal{T}) \quad (\mathcal{T} \in \mathbb{T}).$$

Proposition 2 *There exists a constant* $C_1 > 0$ *such that*

$$\eta(v_{\mathcal{T}}, \mathcal{T}) \leq C_1 [\|u - v_{\mathcal{T}}\|_{L_2(\Omega)} + \mathrm{osc}(\mathcal{T})] \quad (\mathcal{T} \in \mathbb{T}, \; v_{\mathcal{T}} \in S_{\mathcal{T}}).$$

Here osc($\mathcal{M}$) ($\mathcal{M} \subset \mathcal{T}$), *known as the data oscillation term, is defined by*

$$\text{osc}(\mathcal{M})^2 = \sum_{T \in \mathcal{M}} h_T^4 \|f - f_T\|_{L_2(T)}^2,$$

where f_T is the $L_2(T)$-projection of $f|_T$ onto $\mathbb{P}_{k-1}(T)$.

Remark 1 Note that osc($\mathcal{M}$) $\leq \eta(v_{\mathcal{T}}, \mathcal{M})$ ($\mathcal{M} \subset \mathcal{T} \in \mathbb{T}$, $v_{\mathcal{T}} \in S_{\mathcal{T}}$).

Remark 2 Efficient a posteriori control of the L_2 error when H^2 regularity is not present is significantly more difficult than in the present case. When $n = 2$, reliable estimators which control $\|u - u_i\|_{L_2(\Omega)}$ on non-convex polygonal domains by explicitly taking corner singularities into account can be found in [9]; these estimators do not however satisfy a corresponding a posteriori lower bound. Reliable and efficient estimators for bounding a *weighted L_2* norm on non-convex polygonal domains can be found in [23]. Extension of both of these results to $n = 3$ is technically difficult because they rely on knowledge of the exact nature of corner singularities. Analysis of AFEM based on these estimators also would present technical difficulties beyond those encountered in our analysis here for the most basic case, so we restrict our attention to situations with H^2 regularity.

4 Relations between several error notions

In this section we shall establish several relations between different error notions that will be used in the succeeding sections to prove convergence and quasi-optimality of AFEM in the $L_2(\Omega)$-norm.

Instead of directly proving that an AFEM for controlling $\|u - u_i\|_{L_2(\Omega)}$ is contractive, we shall show that a total error notion based on $|||h_i(u - u_i)|||$ contracts. The following proposition establishes that this error notion is meaningful so long as the mesh grading is sufficiently mild. It will be used that Ω is convex.

Proposition 3 *For sufficiently small μ, let $\mathcal{T} \in \mathbb{T}_\mu$. Then*

$$\|u - u_{\mathcal{T}}\|_{L_2(\Omega)} \lesssim |||h_{\mathcal{T}}(u - u_{\mathcal{T}})|||.$$

Proof We first employ a duality argument. Let $e = u - u_{\mathcal{T}}$, and let $z \in H^2(\Omega) \cap H_0^1(\Omega)$ satisfy $A(v, z) = (v, e)$ for all $v \in H_0^1(\Omega)$. Employing Galerkin orthogonality, approximation properties, Theorem 1, $h_{\mathcal{T}}|_T \approx h_T$, and $\|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \leq \mu$, we have that

$$
\begin{aligned}
(e, e) &= A(e, z) \\
&= A(e, z - I_{SZ,\mathcal{T}} z) \lesssim \|h_{\mathcal{T}} \nabla e\|_{L_2(\Omega)} \|h_{\mathcal{T}}^{-1} \nabla(z - I_{SZ,\mathcal{T}} z)\|_{L_2(\Omega)} \\
&\lesssim (|||h_{\mathcal{T}} e||| + \|e \nabla h_{\mathcal{T}}\|_{L_2(\Omega)}) \|z\|_{H^2(\Omega)} \\
&\lesssim |||h_{\mathcal{T}} e||| \|e\|_{L_2(\Omega)} + \mu \|e\|_{L_2(\Omega)}^2 .
\end{aligned}
\tag{4.1}
$$

Taking μ sufficiently small to kick back the last term completes the proof. $\qquad\square$

Next we prove that the a posteriori error estimator provides also an upper bound for $|||h_{\mathcal{T}}(u - u_{\mathcal{T}})|||$. Again, it is used that Ω is convex.

Proposition 4 *For any* $\mathcal{T} \in \mathbb{T}$,

$$
|||h_{\mathcal{T}}(u - u_{\mathcal{T}})||| \lesssim \eta(\mathcal{T}).
$$

Proof Letting $e = u - u_{\mathcal{T}}$, after a short computation and inserting Proposition 1 while recalling that $\|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \lesssim 1$, we find that

$$
\begin{aligned}
A(h_{\mathcal{T}} e, h_{\mathcal{T}} e) &= A(e, h_{\mathcal{T}}^2 e) + \|e \nabla h_{\mathcal{T}}\|_{L_2(\Omega)}^2 \\
&\leq |A(e, h_{\mathcal{T}}^2 e)| + \|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)}^2 \|e\|_{L_2(\Omega)}^2 \\
&\lesssim |A(e, h_{\mathcal{T}}^2 e)| + \eta(\mathcal{T})^2 .
\end{aligned}
\tag{4.2}
$$

Using standard techniques for proving residual-type energy estimates along while recalling that $h_{\mathcal{T}}|_T \eqsim h_T$ and $\|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \lesssim 1$ yields

$$
\begin{aligned}
A(e, h_{\mathcal{T}}^2 e) &= A(e, h_{\mathcal{T}}^2 e - I_{SZ,\mathcal{T}}(h_{\mathcal{T}}^2 e)) \\
&\lesssim \sum_{T \in \mathcal{T}} (h_T \|\Delta u_{\mathcal{T}} + f\|_{L_2(T)} + h_T^{1/2} \|[\![\nabla u_{\mathcal{T}}]\!]\|_{L_2(\partial T)}) \\
&\qquad \times \|\nabla(h_{\mathcal{T}}^2 e)\|_{L_2(\omega_T)} \\
&\lesssim \sum_{T \in \mathcal{T}} \eta(T)(\|\nabla(h_{\mathcal{T}} e)\|_{L_2(\omega_T)} + \|\nabla h_{\mathcal{T}}\|_{L_\infty(\omega_T)} \|e\|_{L_2(\omega_T)}) \\
&\lesssim \eta(\mathcal{T})(|||h_{\mathcal{T}} e||| + \|e\|_{L_2(\Omega)}) .
\end{aligned}
\tag{4.3}
$$

Inserting the result of Proposition 1 into (4.3), inserting the result into (4.2), and kicking back the term $|||h_{\mathcal{T}} e|||$ above completes the proof. $\qquad\square$

Next we prove a stability result for the error indicator, cf. Proposition 3.3 of [4].

Lemma 1 *It holds that*

$$
|\eta(v, T) - \eta(w, T)| \lesssim |||h_{\mathcal{T}}(v - w)|||_{\tilde{\omega}_T} + \|\nabla h_{\mathcal{T}}\|_{L_\infty(\tilde{\omega}_T)} \|v - w\|_{L_2(\tilde{\omega}_T)}
$$

$(T \in \mathcal{T} \in \mathbb{T}, \ v, w \in S_{\mathcal{T}})$.

Proof Recalling the definition (3.1), we use the triangle inequality to compute for $T \in \mathcal{T}_i$ that

$$\eta(v, T) \leq \eta(w, T) + (h_T^4 \|\Delta(v - w)\|_{L_2(T)}^2 + h_T^3 \|[\![\nabla(v - w)]\!]\|_{L_2(\partial T)}^2)^{1/2}. \quad (4.4)$$

Employing an inverse inequality and $h_{\mathcal{T}}|_T \eqsim h_T$, we compute

$$\begin{aligned}
h_T^2 \|\Delta(v - w)\|_{L_2(T)} &\lesssim h_T \|\nabla(v - w)\|_{L_2(T)} \\
&\lesssim \|\nabla[h_{\mathcal{T}}(v - w)]\|_{L_2(T)} + \|(v - w)\nabla h_{\mathcal{T}}\|_{L_2(T)}. \quad (4.5)
\end{aligned}$$

We next compute as above while employing the trace inequality $\|\nabla v\|_{L_2(\partial T)} \lesssim h_T^{-1/2}\|\nabla v\|_{L_2(T)} + h_T^{1/2}|v|_{H^2(T)}$ to obtain for the edge $e = T \cap T'$

$$\begin{aligned}
h_T^{3/2} &\|[\![\nabla(v - w)]\!]\|_{L_2(e)} \\
&\leq h_T^{3/2}(\|\nabla(v - w)_T\|_{L_2(e)} + \|\nabla(v - w)_{T'}\|_{L_2(e)}) \\
&\lesssim h_T^{3/2}\left(h_T^{-1/2}\|\nabla(v - w)\|_{L_2(T)} + h_T^{1/2}|v - w|_{H^2(T)}\right. \\
&\quad\left. + h_T^{-1/2}\|\nabla(v - w)\|_{L_2(T')} + h_T^{1/2}|v - w|_{H^2(T')}\right) \\
&\lesssim \||h_{\mathcal{T}}(v - w)\||_{T \cup T'} + \|(v - w)\nabla h_{\mathcal{T}}\|_{L_2(T \cup T')}. \quad (4.6)
\end{aligned}$$

Above we have employed the convention that for x lying in the edge $e = T \cap T'$, $\nabla(v - w)_T(x) = \lim_{y \in T, y \to x} \nabla(v - w)(y)$. Noting that $\|(v - w)\nabla h_{\mathcal{T}}\|_{L_2(T \cup T')} \leq \|\nabla h_{\mathcal{T}}\|_{L_\infty(\tilde{\omega}_T)}\|v - w\|_{L_2(\tilde{\omega}_T)}$, summing over the edges of T, and collecting (4.5) and (4.6) into (4.4) completes the proof. $\qquad\square$

The following localized upper bound will be used for proving quasi-optimality of the AFEM. It will be used that Ω is convex.

Lemma 2 *There exists a constant $C_2 > 0$ such that for sufficiently small μ, $\mathcal{T} \in \mathbb{T}_\mu$, and $\mathcal{T} \subset \tilde{\mathcal{T}} \in \mathbb{T}$,*

$$\|u_{\tilde{\mathcal{T}}} - u_{\mathcal{T}}\|_{L_2(\Omega)} \leq C_2 \eta(\mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}}).$$

Here $\mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}} \subset \mathcal{T}$ is the subset of elements that are refined in passing from $\mathcal{T}$ to $\tilde{\mathcal{T}}$.

Proof We set $E = u_{\tilde{\mathcal{T}}} - u_{\mathcal{T}}$, so that $A(E, \chi) = 0$ for all $\chi \in S_{\mathcal{T}}$. Arguing as in Proposition 3, it is easy to show that

$$\|E\|_{L_2(\Omega)} \lesssim \|h_{\mathcal{T}} \nabla E\|_{L_2(\Omega)}. \quad (4.7)$$

An elementary computation yields for $\epsilon > 0$

$$\begin{aligned}
\|h_{\mathcal{T}} \nabla E\|_{L_2(\Omega)}^2 &= (h_{\mathcal{T}} \nabla E, h_{\mathcal{T}} \nabla E) = A(E, h_{\mathcal{T}}^2 E) - 2(h_{\mathcal{T}} \nabla E, E \nabla h_{\mathcal{T}}) \\
&\leq |A(E, h_{\mathcal{T}}^2 E)| + \epsilon \|h_{\mathcal{T}} \nabla E\|_{L_2(\Omega)}^2 + \frac{\mu^2}{\epsilon}\|E\|_{L_2(\Omega)}^2. \quad (4.8)
\end{aligned}$$

Taking ϵ small enough to kick back the second term above and inserting the result into (4.7) yields

$$\|h_{\mathcal{T}}\nabla E\|_{L_2(\Omega)}^2 \lesssim |A(E, h_{\mathcal{T}}^2 E)| + \mu^2 \|E\|_{L_2(\Omega)}, \tag{4.9}$$

so that by (4.7) for μ sufficiently small we have

$$\|h_{\mathcal{T}}\nabla E\|_{L_2(\Omega)}^2 \lesssim |A(E, h_{\mathcal{T}}^2 E)|. \tag{4.10}$$

Let $I_{L,\tilde{\mathcal{T}}} : C_0(\Omega) \to S_{\tilde{\mathcal{T}}}$ and $I_{SZ,\mathcal{T}}$ be as in Sect. 2.2. It is shown in [4, Lemma 3.6] that $I_{SZ,\mathcal{T}}$ may be defined so that for $T \in \mathcal{T} \cap \tilde{\mathcal{T}}$ (i.e., for $T \in \mathcal{T} \setminus \mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}}$), $(\chi - I_{SZ,\mathcal{T}}\chi)_T = 0$ for $\chi \in S_{\tilde{\mathcal{T}}}$. We assume that $I_{SZ,\mathcal{T}}$ is so defined. Then

$$\begin{aligned}
&A(E, h_{\mathcal{T}}^2 E) \\
&= \underbrace{A(E, h_{\mathcal{T}}^2 E - I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E))}_{\text{I}} \\
&+ \underbrace{A(E, I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E) - I_{SZ,\mathcal{T}}(I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E)))}_{\text{II}}.
\end{aligned} \tag{4.11}$$

In order to bound the term "I" in (4.11), in essence we employ the classical super approximation tool introduced in [15]. Restricted to $T \in \tilde{\mathcal{T}}$, we have $h_{\mathcal{T}} \in \mathbb{P}_1$ and $E \in \mathbb{P}_k$, and so $|D^2 h_{\mathcal{T}}^2| \lesssim |\nabla h_{\mathcal{T}}|^2$, $D^j h_{\mathcal{T}} = 0$ for $j \geq 3$, and $D^{k+1}E = 0$. Employing inverse inequalities and using that $h_{\mathcal{T}}|_{T'} \eqsim h_{T'}$ ($T' \in \mathcal{T}$) and $\|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \lesssim 1$, we thus compute

$$\begin{aligned}
\|\nabla(h_{\mathcal{T}}^2 E - I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E))\|_{L_2(T)} &\lesssim h_T^k \|D^{k+1}(h_{\mathcal{T}}^2 E)\|_{L_2(T)} \\
&\lesssim h_T^k [\|\nabla h_{\mathcal{T}}^2\|_{L_\infty(T)} \|D^k E\|_{L_2(T)} \\
&\quad + \|D^2 h_{\mathcal{T}}^2\|_{L_\infty(T)} \|D^{k-1}E\|_{L_2(T)}] \\
&\lesssim h_T^k [\|h_{\mathcal{T}}\|_{L_\infty(T)} \|\nabla h_{\mathcal{T}}\|_{L_\infty(T)} \|D^k E\|_{L_2(T)} \\
&\quad + \|\nabla h_{\mathcal{T}}\|_{L_\infty(T)}^2 \|D^{k-1}E\|_{L_2(T)}] \\
&\lesssim [\|h_{\mathcal{T}}\|_{L_\infty(T)} \|\nabla h_{\mathcal{T}}\|_{L_\infty(T)} \\
&\quad + h_T \|\nabla h_{\mathcal{T}}\|_{L_\infty(T)}^2] \|E\|_{L_2(T)} \\
&\lesssim \|h_{\mathcal{T}}\|_{L_\infty(T)} \|\nabla h_{\mathcal{T}}\|_{L_\infty(T)} \|E\|_{L_2(T)}. \tag{4.12}
\end{aligned}$$

Using that $\|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \leq \mu$, for any $\epsilon > 0$ it thus holds that

$$\begin{aligned}
|\text{"I"}| &\lesssim \sum_{T \in \tilde{\mathcal{T}}} \|h_{\mathcal{T}}\nabla E\|_{L_2(T)} \mu \|E\|_{L_2(T)} \\
&\lesssim \|h_{\mathcal{T}}\nabla E\|_{L_2(\Omega)} \mu \|E\|_{L_2(\Omega)} \\
&\leq \epsilon \|h_{\mathcal{T}}\nabla E\|_{L_2(\Omega)}^2 + \frac{\mu^2}{4\epsilon} \|E\|_{L_2(\Omega)}^2 \tag{4.13}
\end{aligned}$$

To bound the term "II", we first note that since $h_{\mathcal{T}}^2 E$ is piecewise polynomial on $\tilde{\mathcal{T}}$, the stability bound $\|\nabla I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E)\|_{L_2(T)} \lesssim \|\nabla(h_{\mathcal{T}}^2 E)\|_{L_2(T)}$ holds. Recall that $I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E) - I_{SZ,\mathcal{T}}(I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E))$ is in $S_{\tilde{\mathcal{T}}}$ and vanishes on all $T \in \mathcal{T} \setminus \mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}}$. We may then compute using Galerkin orthogonality and standard residual techniques that for $\epsilon > 0$,

$$
\begin{aligned}
|\text{"II"}| &= |A(u - u_{\mathcal{T}}, I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E) - I_{SZ,\mathcal{T}}(I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E)))| \\
&\lesssim \sum_{T \in \mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}}} (h_T \|f + \Delta u_{\mathcal{T}}\|_{L_2(T)} + h_T^{1/2} \|[\![\nabla u_{\mathcal{T}}]\!]\|_{L_2(\partial T)}) \\
&\quad \times \|\nabla I_{L,\tilde{\mathcal{T}}}(h_{\mathcal{T}}^2 E)\|_{L_2(\omega_T)} \\
&\lesssim \sum_{T \in \mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}}} h_T^{-1} \eta(T) \|\nabla(h_{\mathcal{T}}^2 E)\|_{L_2(\omega_T)} \\
&\lesssim \sum_{T \in \mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}}} h_T^{-1} \eta(T) \\
&\quad \times (h_T \|h_{\mathcal{T}} \nabla E\|_{L_2(\omega_T)} + h_T \mu \|E\|_{L_2(\omega_T)}) \\
&\lesssim \sum_{T \in \mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}}} \eta(T)(\|h_{\mathcal{T}} \nabla E\|_{L_2(\omega_T)} + \mu \|E\|_{L_2(\omega_T)}) \\
&\lesssim \left(1 + \frac{1}{\epsilon}\right) \eta(\mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}})^2 + \epsilon \|h_{\mathcal{T}} \nabla E\|_{L_2(\Omega)}^2 + \mu^2 \|E\|_{L_2(\Omega)}^2. \quad (4.14)
\end{aligned}
$$

Inserting (4.13) and (4.14) into (4.11) yields for $\epsilon > 0$,

$$
\begin{aligned}
\|h_{\mathcal{T}} \nabla E\|_{L_2(\Omega)}^2 &\lesssim \left(1 + \frac{1}{\epsilon}\right) \eta(\mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}})^2 + \epsilon \|h_{\mathcal{T}} \nabla E\|_{L_2(\Omega)}^2 \\
&\quad + \frac{\mu^2}{\epsilon} \|E\|_{L_2(\Omega)}^2.
\end{aligned} \quad (4.15)
$$

Taking ϵ small enough to kick back the second term above and then inserting the result into (4.10) yields

$$
\|h_{\mathcal{T}} \nabla E\|_{L_2(\Omega)}^2 \lesssim \eta(\mathcal{R}_{\mathcal{T} \to \tilde{\mathcal{T}}})^2 + \mu^2 \|h_{\mathcal{T}} \nabla E\|_{L_2(\Omega)}^2 \quad (4.16)
$$

by (4.7). Taking μ small enough to kick back the last term above and again applying (4.7) completes the proof of the lemma. $\qquad \square$

We next prove a *quasi-orthogonality* result (cf. [11] for a similar estimate in the context of convergence of AFEM in the global energy norm for general second-order linear elliptic problems).

Lemma 3 *For any $\epsilon > 0$, $\mathcal{T} \in \mathbb{T}$, and $v_{\mathcal{T}} \in S_{\mathcal{T}}$, it holds that*

$$
\begin{aligned}
&|||h_{\mathcal{T}}(u - u_{\mathcal{T}})|||^2 + |||h_{\mathcal{T}}(u_{\mathcal{T}} - v_{\mathcal{T}})|||^2 - (1 + \epsilon)|||h_{\mathcal{T}}(u - v_{\mathcal{T}})|||^2 \\
&\lesssim \epsilon^{-1} \|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)}^2 (\|u - u_{\mathcal{T}}\|_{L_2(\Omega)}^2 + \|u - v_{\mathcal{T}}\|_{L_2(\Omega)}^2).
\end{aligned} \quad (4.17)
$$

Proof Writing $\tilde{e}_{\mathcal{T}} = u - v_{\mathcal{T}}$ and $e_{\mathcal{T}} = u - u_{\mathcal{T}}$, we calculate

$$|||h_{\mathcal{T}} e_{\mathcal{T}}|||^2 = |||h_{\mathcal{T}} \tilde{e}_{\mathcal{T}}|||^2 - |||h_{\mathcal{T}}(u_{\mathcal{T}} - v_{\mathcal{T}})|||^2$$
$$-2A(h_{\mathcal{T}} e_{\mathcal{T}}, h_{\mathcal{T}}(u_{\mathcal{T}} - v_{\mathcal{T}})). \tag{4.18}$$

An elementary calculation yields

$$|A(h_{\mathcal{T}} e_{\mathcal{T}}, h_{\mathcal{T}}(u_{\mathcal{T}} - v_{\mathcal{T}}))| = |A(e_{\mathcal{T}}, h_{\mathcal{T}}^2(u_{\mathcal{T}} - v_{\mathcal{T}}))$$
$$+(|\nabla h_{\mathcal{T}}|^2 e_{\mathcal{T}}, u_{\mathcal{T}} - v_{\mathcal{T}}) - (\tilde{e}_{\mathcal{T}} \nabla h_{\mathcal{T}}, \nabla(h_{\mathcal{T}} e_{\mathcal{T}}))$$
$$+(e_{\mathcal{T}} \nabla h_{\mathcal{T}}, \nabla(h_{\mathcal{T}} \tilde{e}_{\mathcal{T}}))|$$
$$\leq \|\tilde{e}_{\mathcal{T}} \nabla h_{\mathcal{T}}\|_{L_2(\Omega)} |||h_{\mathcal{T}} e_{\mathcal{T}}||| + \|e_{\mathcal{T}} \nabla h_{\mathcal{T}}\|_{L_2(\Omega)} |||h_{\mathcal{T}} \tilde{e}_{\mathcal{T}}|||$$
$$+\|e_{\mathcal{T}} \nabla h_{\mathcal{T}}\|_{L_2(\Omega)} \|(u_{\mathcal{T}} - v_{\mathcal{T}}) \nabla h_{\mathcal{T}}\|_{L_2(\Omega)}$$
$$+|A(e_{\mathcal{T}}, h_{\mathcal{T}}^2(u_{\mathcal{T}} - v_{\mathcal{T}}))|. \tag{4.19}$$

Arguments analogous to those in (4.12) and (4.13) yield

$$|A(e_{\mathcal{T}}, h_{\mathcal{T}}^2(u_{\mathcal{T}} - v_{\mathcal{T}}))| = |A(e_{\mathcal{T}}, (I - I_{L,\mathcal{T}})h_{\mathcal{T}}^2(u_{\mathcal{T}} - v_{\mathcal{T}}))|$$
$$\lesssim \|h_{\mathcal{T}} \nabla e_{\mathcal{T}}\|_{L_2(\Omega)} \|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \|u_{\mathcal{T}} - v_{\mathcal{T}}\|_{L_2(\Omega)}$$
$$\leq (|||h_{\mathcal{T}} e_{\mathcal{T}}||| + \|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \|e_{\mathcal{T}}\|_{L_2(\Omega)})$$
$$\times \|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \|u_{\mathcal{T}} - v_{\mathcal{T}}\|_{L_2(\Omega)}.$$

Inserting this bound into (4.19) and applying Young's inequality a few times yields for some constant $C > 0$ and any $\delta > 0$ that

$$|A(h_{\mathcal{T}} e_{\mathcal{T}}, h_{\mathcal{T}}(u_{\mathcal{T}} - v_{\mathcal{T}}))| \leq \frac{\delta}{2}(|||h_{\mathcal{T}} e_{\mathcal{T}}|||^2 + |||h_{\mathcal{T}} \tilde{e}_{\mathcal{T}}|||^2)$$
$$+ C\left(1 + \frac{1}{\delta}\right) \|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)}^2 \left(\|e_{\mathcal{T}}\|_{L_2(\Omega)}^2 + \|\tilde{e}_{\mathcal{T}}\|_{L_2(\Omega)}^2\right). \tag{4.20}$$

Inserting (4.20) into (4.18) then yields

$$(1 - \delta)|||h_{\mathcal{T}} e_{\mathcal{T}}|||^2 \leq (1 + \delta)|||h_{\mathcal{T}} \tilde{e}_{\mathcal{T}}|||^2 - |||h_{\mathcal{T}}(u_{\mathcal{T}} - v_{\mathcal{T}})|||^2$$
$$+ 2C\left(1 + \frac{1}{\delta}\right) \|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)}^2 \left(\|e_{\mathcal{T}}\|_{L_2(\Omega)}^2 + \|\tilde{e}_{\mathcal{T}}\|_{L_2(\Omega)}^2\right). \tag{4.21}$$

Dividing through by $1 - \delta$ and selecting $\frac{1+\delta}{1-\delta}$ as $1 + \epsilon$ completes the proof of Lemma 3.
□

A combination of the Propositions 4 and 2 shows that

$$|||h_{\mathcal{T}}(u - u_{\mathcal{T}})||| \lesssim \|u - u_{\mathcal{T}}\|_{L_2(\Omega)} + \text{osc}(\mathcal{T}).$$

In the following lemma it is shown that this inequality is even valid with $u_{\mathcal{T}}$ reading as any $v_{\mathcal{T}} \in S_{\mathcal{T}}$.

Lemma 4 *For $\mathcal{T} \in \mathbb{T}$, it holds that*

$$|||h_{\mathcal{T}}(u - v_{\mathcal{T}})||| \lesssim \|u - v_{\mathcal{T}}\|_{L_2(\Omega)} + \mathrm{osc}(\mathcal{T}) \quad (v_{\mathcal{T}} \in S_{\mathcal{T}}).$$

Proof Writing $z_{\mathcal{T}} = u - v_{\mathcal{T}}$, from $\|h_{\mathcal{T}}\|_{L_\infty(\Omega)} \lesssim 1$ we have for $\epsilon > 0$

$$
\begin{aligned}
|||h_{\mathcal{T}}z_{\mathcal{T}}|||^2 &= A(h_{\mathcal{T}}z_{\mathcal{T}}, h_{\mathcal{T}}z_{\mathcal{T}}) \\
&= (h_{\mathcal{T}}\nabla z_{\mathcal{T}}, \nabla(h_{\mathcal{T}}z_{\mathcal{T}})) + (z_{\mathcal{T}}\nabla h_{\mathcal{T}}, \nabla(h_{\mathcal{T}}z_{\mathcal{T}})) \\
&\leq \frac{1}{2\epsilon}\left(\|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(\Omega)}^2 + \|z_{\mathcal{T}}\|_{L_2(\Omega)}^2\right) + \epsilon|||h_{\mathcal{T}}z_{\mathcal{T}}|||^2 \quad (4.22)
\end{aligned}
$$

or

$$|||h_{\mathcal{T}}z_{\mathcal{T}}|||^2 \lesssim \|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(\Omega)}^2 + \|z_{\mathcal{T}}\|_{L_2(\Omega)}^2. \tag{4.23}$$

Integrating by parts, we next compute

$$
\begin{aligned}
\|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(\Omega)}^2 &= (\nabla z_{\mathcal{T}}, h_{\mathcal{T}}^2\nabla z_{\mathcal{T}}) \\
&= \sum_{T \in \mathcal{T}} \int_T -z_{\mathcal{T}}\nabla \cdot (h_{\mathcal{T}}^2\nabla z_{\mathcal{T}})dx + \int_{\partial T} z_{\mathcal{T}}h_{\mathcal{T}}^2\nabla z_{\mathcal{T}} \cdot \mathbf{n}\,d\sigma \\
&\leq \sum_{T \in \mathcal{T}} \|z_{\mathcal{T}}\|_{L_2(T)}(\|h_{\mathcal{T}}\|_{L_\infty(T)}^2\|f + \Delta v_{\mathcal{T}}\|_{L_2(T)} \\
&\quad + 2\|\nabla h_{\mathcal{T}}\|_{L_\infty(T)}\|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(T)}) \\
&\quad + \|h_{\mathcal{T}}^2 z_{\mathcal{T}}\|_{L_2(\partial T \setminus \partial \Omega)}\|[\![\nabla v_{\mathcal{T}}]\!]\|_{L_2(\partial T \setminus \partial \Omega)}. \tag{4.24}
\end{aligned}
$$

Inserting the scaled trace inequality $\|v\|_{L_2(\partial T)} \lesssim h_T^{-1/2}\|v\|_{L_2(T)} + h_T^{1/2}|||v|||_T$ into (4.24) yields for $\epsilon > 0$

$$
\begin{aligned}
&\|h_{\mathcal{T}}^2 z_{\mathcal{T}}\|_{L_2(\partial T)}\|[\![\nabla v_{\mathcal{T}}]\!]\|_{L_2(\partial T)} \\
&\lesssim \left(h_T^{1/2}\|h_{\mathcal{T}}z_{\mathcal{T}}\|_{L_2(T)} + h_T^{3/2}|||h_{\mathcal{T}}z_{\mathcal{T}}|||_T\right)\|[\![\nabla v_{\mathcal{T}}]\!]\|_{L_2(\partial T)} \\
&\leq (\|z\|_{L_2(T)} + \|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(T)} + \|\nabla h_{\mathcal{T}}\|_{L_\infty(T)}\|z_{\mathcal{T}}\|_{L_2(T)}) \\
&\quad \times h_T^{3/2}\|[\![\nabla v_{\mathcal{T}}]\!]\|_{L_2(\partial T)} \\
&\lesssim \|z\|_{L_2(T)}^2 + \left(1 + \frac{1}{\epsilon}\right)h_T^3\|[\![\nabla v_{\mathcal{T}}]\!]\|_{L_2(\partial T)}^2 + \epsilon\|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(T)}^2. \tag{4.25}
\end{aligned}
$$

Inserting (4.25) into (4.24) and then employing Proposition 2 and $h_{\mathcal{T}}|_T \approx h_T$ yields for $\epsilon > 0$

$$\|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(\Omega)}^2$$
$$\lesssim \sum_{T\in\mathcal{T}}\left(1+\frac{1}{\epsilon}\right)\left(\|z_{\mathcal{T}}\|_{L_2(T)}^2 + h_T^3\|[\![\nabla v_{\mathcal{T}}]\!]\|_{L_2(\partial T\setminus\partial\Omega)}^2\right)$$
$$+h_T^4\|f+\Delta v_{\mathcal{T}}\|_{L_2(T)}^2 + \epsilon\|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(T)}^2$$
$$\lesssim \left(1+\frac{1}{\epsilon}\right)\left(\|z_{\mathcal{T}}\|_{L_2(\Omega)}^2 + \eta(v_{\mathcal{T}},\mathcal{T})^2\right) + \epsilon\|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(T)}^2$$
$$\lesssim \left(1+\frac{1}{\epsilon}\right)\left(\|z_{\mathcal{T}}\|_{L_2(\Omega)}^2 + \mathrm{osc}(\mathcal{T})^2\right) + \epsilon\|h_{\mathcal{T}}\nabla z_{\mathcal{T}}\|_{L_2(T)}^2. \qquad (4.26)$$

Taking ϵ small enough to kick back the last term above and inserting the result into (4.23) completes the proof.   □

We next prove that if $\mathcal{T}$ is sufficiently mildly graded, then $\|u-u_{\mathcal{T}}\|_{L_2(\Omega)}$ is bounded up to a constant by the best approximation to u lying in $S_{\mathcal{T}}$, as measured in L_2, plus a data oscillation term.

Corollary 1 *There exists a constant $C_3 > 0$ such that for sufficiently small μ and $\mathcal{T}\in\mathbb{T}_\mu$, it holds that*

$$\|u-u_{\mathcal{T}}\|_{L_2(\Omega)} \leq C_3[\inf_{v_{\mathcal{T}}\in S_{\mathcal{T}}}\|u-v_{\mathcal{T}}\|_{L_2(\Omega)} + \mathrm{osc}(\mathcal{T})]. \qquad (4.27)$$

Proof By using that $\|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \leq \mu$, the application of Proposition 3, Lemma 3 for some fixed ϵ, and Lemma 4 shows that for any $v_{\mathcal{T}}\in S_{\mathcal{T}}$

$$\|u-u_{\mathcal{T}}\|_{L_2(\Omega)}^2 \lesssim |||h_{\mathcal{T}}(u-u_{\mathcal{T}})|||^2$$
$$\lesssim |||h_{\mathcal{T}}(u-v_{\mathcal{T}})|||^2 + \mu^2\left(\|u-u_{\mathcal{T}}\|_{L_2(\Omega)}^2 + \|u-v_{\mathcal{T}}\|_{L_2(\Omega)}^2\right)$$
$$\lesssim (1+\mu^2)\|u-v_{\mathcal{T}}\|_{L_2(\Omega)}^2 + \mathrm{osc}(\mathcal{T})^2 + \mu^2\|u-u_{\mathcal{T}}\|_{L_2(\Omega)}^2.$$

By taking μ sufficiently small, the proof is completed.   □

Remark 3 Corollary 1 is of some interest independent of the current context because the Galerkin approximation is *not* stable in L_2 even on quasi-uniform meshes, that is, (4.27) does not hold if the data oscillation term is removed. A simple one-dimensional counterexample can be found in [1]. The counterexample to L_2-stability given in [1] also shows that $\|u-u_{\mathcal{T}}\|_{L_2(\Omega)}$ does *not* bound the right hand side of (4.27) up to a constant. It does however trivially follow from Corollary 1 that $\|u-u_{\mathcal{T}}\|_{L_2(\Omega)}+\mathrm{osc}(\mathcal{T}) \eqsim \inf_{v_{\mathcal{T}}\in S_{\mathcal{T}}}\|u-v_{\mathcal{T}}\|_{L_2(\Omega)}+\mathrm{osc}(\mathcal{T})$ uniformly in all $\mathcal{T}\in\mathbb{T}_\mu$ with μ sufficiently small. That is, on sufficiently mildly graded meshes an analog to Céa's Lemma holds for the *total L_2 error* $\|u-u_{\mathcal{T}}\|_{L_2(\Omega)} + \mathrm{osc}(\mathcal{T})$.

5 Adaptive FEM

In this section we give details of our adaptive FEM. In particular, we give precise definitions of each module solve, estimate, mark, and refine of the generic

adaptive iteration (1.2), with the goal of constructing an AFEM that produces by refinement a sequence of partitions $(\mathcal{T}_i)_{i \geq 0} \subset \mathbb{T}_\mu$ for reducing the weighted energy error $|||h_i(u - u_i)|||$, and thus also the L_2-error $\|u - u_i\|_{L_2(\Omega)}$. Here, and in the following, h_i, u_i, S_i, osc$_i$, η_i, and, for $\mathcal{M} \subset \mathcal{T}_i$, $\eta_i(\mathcal{M})$ denote $h_{\mathcal{T}_i}$, $u_{\mathcal{T}_i}$, $S_{\mathcal{T}_i}$, osc$(\mathcal{T}_i)$, $\eta(u_{\mathcal{T}_i}, \mathcal{T}_i)$, and $\eta(u_{\mathcal{T}_i}, \mathcal{M})$, respectively. The constant μ will be assumed to be sufficiently small.

1. Module **solve**. Given the current partition $\mathcal{T}_i$, solve (1.3) for $u_i \in S_i$. We assume that the finite element system is assembled and solved exactly.
2. Module **estimate**. In principle, the adaptive algorithm is terminated when $C\eta(\mathcal{T}_i) \leq tol$ for some prescribed tolerance tol and user-defined constant C.
3. Module **mark**. We employ a Dörfler marking (cf. [6]). More precisely, we fix a parameter $0 < \theta < 1$, and at each step of the algorithm choose the smallest set $\mathcal{M}_i \subset \mathcal{T}_i$ so that

$$\eta(\mathcal{M}_i) \geq \theta \eta_i. \tag{5.1}$$

4. Module **refine**. Our results below assume that each marked element $T \in \mathcal{M}_i$ is bisected $b \geq 1$ times in passing from $\mathcal{T}_i$ to $\mathcal{T}_{i+1}$ and that generally additional elements are refined in the process in order to ensure that $\mathcal{T}_{i+1}$ is conforming and sufficiently mildly graded in the sense that it is in $\mathbb{T}_\mu$. In Appendix A, we show that, assuming the value of μ is known, the standard newest vertex bisection algorithm or its generalization to more than two dimensions, as it has been implemented in the finite element toolbox ALBERTA [17], can be modified so that the sequence of adaptive meshes is indeed in $\mathbb{T}_\mu$. As we will see in Sect. 7, this modification does not compromise quasi-optimality of the resulting AFEM.

6 Convergence of the AFEM in L_2

In this section we prove a quasi-orthogonality property, an estimator reduction inequality, and finally an error contraction property. We employ the techniques of [4], which do not rely upon local a posteriori lower bounds as do previous proofs of convergence of adaptive FEM.

Proposition 5 *For any $\epsilon > 0$, we have that*

$$|||h_{i+1}(u - u_{i+1})|||^2 + |||h_{i+1}(u_{i+1} - u_i)|||^2 - (1 + \epsilon)|||h_i(u - u_i)|||^2$$
$$\lesssim \epsilon^{-1}\mu^2 \left(\|u - u_{i+1}\|^2_{L_2(\Omega)} + \|u - u_i\|^2_{L_2(\Omega)} \right). \tag{6.1}$$

Proof Writing $e_i = u - u_i$, and using that $\mathcal{T}_i, \mathcal{T}_{i+1} \in \mathbb{T}_\mu$ and $h_{i+1} \leq h_i$, we have

$$|||h_{i+1}e_i||| \leq \|h_{i+1}\nabla e_i\|_{L_2(\Omega)} + \|\nabla h_{i+1}\|_{L_\infty(\Omega)}\|e_i\|_{L_2(\Omega)}$$
$$\leq \|h_i \nabla e_i\|_{L_2(\Omega)} + \mu\|e_i\|_{L_2(\Omega)} \leq |||h_i e_i||| + 2\mu\|e_i\|_{L_2(\Omega)}.$$

and so, for any $\epsilon > 0$,

$$|||h_{i+1}e_i|||^2 \leq (1+\epsilon)|||h_i e_i|||^2 + \left(1 + \frac{1}{\epsilon}\right)4\mu^2\|e_i\|_{L_2(\Omega)}^2.$$

By substituting this result into the estimate of Lemma 3 with $\mathcal{T} = \mathcal{T}_{i+1}$, and thus $u_{\mathcal{T}} = u_{i+1}$, and $v_{\mathcal{T}} = u_i$, we find (6.1) with $1 + \epsilon$ reading as $(1+\epsilon)^2$, which is an equivalent statement.   □

Next we establish an estimator reduction result. Our proof closely follows Corollary 3.4 of [4].

Proposition 6 *With* $\lambda := 1 - 2^{-\frac{3b}{n}}$, *for any* $\delta \in (0, 1]$ *we have*

$$\eta_{i+1}^2 - (1+\delta)(1 - \lambda\theta^2)\eta_i^2$$

$$\lesssim \delta^{-1}\left[|||h_{i+1}(u_{i+1} - u_i)|||^2 + \mu^2\|u_{i+1} - u_i\|_{L_2(\Omega)}^2\right].$$

Proof We first apply Lemma 1 to $u_i, u_{i+1} \in S_{i+1}$, square the result, and apply Young's inequality to the resulting mixed terms to obtain for any $\delta \in (0, 1]$ and $T \in \mathcal{T}_{i+1}$

$$\eta(u_{i+1}, T)^2 - (1+\delta)\eta(u_i, T)^2$$

$$\lesssim \delta^{-1}\left[|||h_{i+1}(u_{i+1} - u_i)|||_{\tilde\omega_T}^2 + \|\nabla h_{\mathcal{T}}\|_{L_\infty(\tilde\omega)}^2\|u_{i+1} - u_i\|_{L_2(\tilde\omega_T)}^2\right].$$

Summing over $T \in \mathcal{T}_{i+1}$ and using the fact that no element is contained in more than $n + 1$ patches $\tilde\omega_T$, we have

$$\eta_{i+1}^2 - (1+\delta)\eta(u_i, \mathcal{T}_{i+1})^2$$

$$\lesssim \delta^{-1}\left[|||h_{i+1}(u_{i+1} - u_i)|||^2 + \mu^2\|u_{i+1} - u_i\|_{L_2(\Omega)}^2\right]. \tag{6.2}$$

For $T' \in \mathcal{T}_i$, let $\mathcal{T}_{T'} = \{T \in \mathcal{T}_{i+1} : T \subset T'\}$. Note that for a marked element $T' \in \mathcal{M}_i$ and $T \in \mathcal{T}_{T'}$, $h_T \leq 2^{-\frac{b}{n}}h_{T'}$ and $[\![\nabla u_i]\!] = 0$ across interfaces of T lying in the interior of T', and so

$$\eta(u_i, \mathcal{T}_{T'})^2 \leq 2^{-\frac{3b}{n}}\eta(u_i, T')^2. \tag{6.3}$$

For $T' \in \mathcal{T}_i \setminus \mathcal{M}_i$, we combine (6.3) with the trivially proved monotonicity property $\eta(u_i, \mathcal{T}_{T'}) \leq \eta(u_i, T')$ and sum over $T \in \mathcal{T}_{i+1}$ to obtain

$$\eta(u_i, \mathcal{T}_{i+1})^2 \leq \eta(u_i, \mathcal{T}_i \setminus \mathcal{M}_i)^2 + 2^{-\frac{3b}{n}}\eta(u_i, \mathcal{M}_i)^2$$
$$= \eta(u_i, \mathcal{T}_i)^2 - \lambda\eta(u_i, \mathcal{M}_i)^2 \leq (1 - \lambda\theta^2)\eta_i^2 \tag{6.4}$$

by (5.1). Combining (6.2) and (6.4) completes the proof.   □

We finally establish a contraction property for $|||h_i(u - u_i)|||^2 + \gamma \eta(\mathcal{T}_i)^2$ for some properly chosen constant $\gamma > 0$.

Theorem 2 *There exist constants $\gamma > 0$ and $\alpha \in (0, 1)$ depending on a generic constant of the type defined in Sect. 2.4, the parameter θ in (5.1), and the number of times b that each element in $\mathcal{M}_i$ is bisected such that for sufficiently small μ,*

$$|||h_{i+1}(u - u_{i+1})|||^2 + \gamma \eta_{i+1}^2 \leq \alpha^2 \left(|||h_i(u - u_i)|||^2 + \gamma \eta_i^2 \right). \qquad (6.5)$$

Proof We will use the abbreviations $e_j = |||h_j(u - u_j)|||$ $(j \in \{i, i + 1\})$, and $E_i = |||h_{i+1}(u_i - u_{i+1})|||$.

From Propositions 1 and 4–6, we know that there exists constants $C_4, C_5, C_6 > 0$ such that for all $\epsilon, \delta > 0$,

$$e_i \leq C_4 \eta_i, \qquad (6.6)$$

$$e_{i+1}^2 \leq (1 + \epsilon)e_i^2 - E_i^2 + C_5\epsilon^{-1}\mu^2 \left(\eta_i^2 + \eta_{i+1}^2 \right), \qquad (6.7)$$

$$\eta_{i+1}^2 \leq (1 + \delta)(1 - \lambda\theta^2)\eta_i^2 + C_6\delta^{-1} \left[E_i + \mu^2 \left(\eta_i^2 + \eta_{i+1}^2 \right) \right]$$

$$\leq (1 + \delta) \left[\left(1 - \frac{1}{2}\lambda\theta^2 \right) \eta_i^2 - \frac{1}{2C_4^2}\lambda\theta^2 e_i^2 \right] \qquad (6.8)$$

$$+ C_6\delta^{-1} \left[E_i + \mu^2 \left(\eta_i^2 + \eta_{i+1}^2 \right) \right] \qquad (6.9)$$

where to arrive at (6.8) we used already (6.6).

Multiplying (6.8) with $\tilde{\gamma} = \tilde{\gamma}(\delta) = \delta C_6^{-1}$ and adding the result to (6.7) yields

$$e_{i+1}^2 + \tilde{\gamma}\eta_{i+1}^2 \leq (1 + \delta) \left(1 - \frac{1}{2}\lambda\theta^2 \right) \tilde{\gamma}\eta_i^2$$

$$+ \left[(1 + \epsilon) - (1 + \delta)\frac{\tilde{\gamma}}{2C_4^2}\lambda\theta^2 \right] e_i^2 + q\mu^2 \left(\eta_i^2 + \eta_{i+1}^2 \right), \qquad (6.10)$$

where $q = q(\epsilon) = C_5\epsilon^{-1} + 1$. Now by fixing a sufficiently small δ and, subsequently, a sufficiently small ϵ such that

$$\tilde{\alpha}^2 := \max \left\{ (1 + \delta) \left(1 - \frac{1}{2}\lambda\theta^2 \right), (1 + \epsilon) - (1 + \delta)\frac{\tilde{\gamma}}{2C_4^2}\lambda\theta^2 \right\} < 1,$$

(6.10) implies that

$$e_{i+1}^2 + (\tilde{\gamma} - q\mu^2)\eta_{i+1}^2 \leq \tilde{\alpha}^2 e_i^2 + (\tilde{\alpha}^2\tilde{\gamma} + q\mu^2)\eta_i^2.$$

Now by choosing μ sufficiently small such that $0 < \frac{\tilde{\alpha}^2 \tilde{\gamma} + q\mu^2}{\tilde{\gamma} - q\mu^2} \leq \alpha^2 := \frac{1+\tilde{\alpha}^2}{2}$, the proof is completed for that α and $\gamma := \tilde{\gamma} - q\mu^2$.   □

We finally show that $\|u - u_i\|_{L_2(\Omega)} \to 0$ with linear rate as $i \to \infty$.

Corollary 2 *Assume that μ is sufficiently small as in Theorem 2. Then with α from that theorem, it holds that for $i \geq j$*

$$\|u - u_i\|_{L_2(\Omega)} + \mathrm{osc}_i \lesssim \alpha^{i-j} [\|u - u_j\|_{L_2(\Omega)} + \mathrm{osc}_j].$$

Proof The proof follows from Theorem 2 and

$$(\|u - u_i\|_{L_2(\Omega)} + \mathrm{osc}_i)^2 \eqsim |||h_i(u - u)|||^2 + \gamma \eta_i^2,$$

the latter being a consequence of Propositions 1, 2 and 4, and Remark 1.   □

7 Quasi-optimality of AFEM in L_2

For $s > 0$, we define the approximation class

$$\mathcal{A}^s = \left\{ v \in H_0^1(\Omega) : -\Delta v \in L_2(\Omega), |v|_{\mathcal{A}^s} \right.$$

$$\left. := \sup_{N \in \mathbb{N}} N^s \inf_{\{\mathcal{T} \in \mathbb{T} : \#\mathcal{T} - \#\mathcal{T}_0 \leq N\}} [\inf_{v_{\mathcal{T}} \in S_{\mathcal{T}}} \|u - v_{\mathcal{T}}\|_{L_2(\Omega)} + \mathrm{osc}(\mathcal{T})] < \infty \right\}.$$

Thus $\mathcal{A}^s$ contains all $v \in H_0^1(\Omega)$ with $-\Delta v \in L_2(\Omega)$, that, for some sequence $(\mathcal{T}_i)_i$, can be approximated at rate s in the L_2-norm by a sequence of functions from $(S_{\mathcal{T}_i})_i$, and for which $-\Delta v$ can be approximated by a sequence of functions from $(\prod_{T \in \mathcal{T}_i} \mathbb{P}_{k-1}(T))_i$ in the $\mathcal{T}_i$-dependent weighted L_2-norm $\sqrt{\sum_{T \in \mathcal{T}_i} h_T^4 \|g\|_{L_2(T)}^2}$.

Note that in our case of Ω being convex, $v \in H_0^1(\Omega)$ with $-\Delta v \in L_2(\Omega)$ is equivalent to $v \in H_0^1(\Omega) \cap H^2(\Omega)$. Standard estimates show that for $s \in \left[\frac{2}{n}, \frac{k+1}{n}\right]$, we have $H_0^1 \cap H^{sn}(\Omega) \subset \mathcal{A}^s$, where for $v \in H_0^1 \cap H^{sn}(\Omega)$ the rate s is already realized with uniform refinements. The class $\mathcal{A}^s$, however, is much larger than $H_0^1 \cap H^{sn}(\Omega)$, which is the reason an AFEM is employed in the first place. What is in essence needed for $v \in H_0^1 \cap H^2(\Omega)$ to be in $\mathcal{A}^s$ is that its snth order partial derivatives are bounded in $L_p(\Omega)$ for some $p > \left(\frac{1}{2} + s\right)^{-1}$.

In this section, we are going to show that if, for whatever $s > 0$, the solution u of (1.1) is in $\mathcal{A}^s$, then for the sequence $(\mathcal{T}_i)_i \subset \mathbb{T}_\mu$ and corresponding Galerkin solutions $(u_i)_i$ produced by our AFEM, it holds that $\|u - u_i\|_{L_2(\Omega)} + \mathrm{osc}_i \lesssim (\#\mathcal{T}_i - \#\mathcal{T}_0)^{-s} |u|_{\mathcal{A}^s}$. Thus our AFEM realizes the best possible convergence rate.

For earlier results on quasi-optimality of AFEM with respect to the energy-norm and more details on approximation classes and references, we refer to [4,16,19].

The following lemma will be used to bound the number of marked elements in the Dörfler marking.

Lemma 5 *With C_1 and C_2 being the constants from Proposition 2 and Lemma 2, let $\theta < \frac{1}{C_1(C_2+1)}$. Let μ be sufficiently small such that Lemma 2 is valid. Then for $(\mathbb{T}_\mu \ni) \mathcal{T}_i \subset \mathcal{T} \in \mathbb{T}$ with*

$$\|u - u_\mathcal{T}\|_{L_2(\Omega)} + \operatorname{osc}(\mathcal{T}) \le [1 - \theta C_1(C_2 + 1)](\|u - u_i\|_{L_2(\Omega)} + \operatorname{osc}_i), \quad (7.1)$$

it holds that

$$\eta(\mathcal{R}_{\mathcal{T}_i \to \mathcal{T}}) \ge \theta \eta_i.$$

Proof By adding the inequalities

$$\|u - u_i\|_{L_2(\Omega)} \le \|u_i - u_\mathcal{T}\|_{L_2(\Omega)} + \|u - u_\mathcal{T}\|_{L_2(\Omega)}$$
$$\operatorname{osc}_i \le \operatorname{osc}(\mathcal{R}_{\mathcal{T}_i \to \mathcal{T}}) + \operatorname{osc}(\mathcal{T})$$

and employing (7.1), we infer that

$$\begin{aligned}
\theta(C_2 + 1)\eta_i &\le \theta C_1(C_2 + 1)(\|u - u_i\|_{L_2(\Omega)} + \operatorname{osc}_i) \\
&\le \|u - u_i\|_{L_2(\Omega)} + \operatorname{osc}_i - \|u - u_\mathcal{T}\|_{L_2(\Omega)} - \operatorname{osc}(\mathcal{T}) \\
&\le \|u_i - u_\mathcal{T}\|_{L_2(\Omega)} + \operatorname{osc}(\mathcal{R}_{\mathcal{T}_i \to \mathcal{T}}) \\
&\le (C_2 + 1)\eta(\mathcal{R}_{\mathcal{T}_i \to \mathcal{T}}),
\end{aligned}$$

where the first and last inequality follow from applications of Proposition 2 and Lemma 2, respectively. $\qquad\square$

Corollary 3 *For some $s > 0$, let $u \in \mathcal{A}^s$. Assume also that μ is sufficiently small that Corollary 1 is valid. Then under the assumptions of Lemma 5, the collection of marked elements $\mathcal{M}_i \subset \mathcal{T}_i$ defined by (5.1) satisfies*

$$\#\mathcal{M}_i \lesssim |u|_{\mathcal{A}^s}^{1/s}(\|u - u_i\|_{L_2(\Omega)} + \operatorname{osc}_i)^{-1/s}.$$

Proof With C_3 the constant from Corollary 1, by definition of $\mathcal{A}^s$ there exists a partition $\mathcal{T}' \in \mathbb{T}$ with

$$\#\mathcal{T}' - \#\mathcal{T}_0 \lesssim |u|_{\mathcal{A}^s}^{1/s}\left(\frac{1 - \theta C_1(C_2 + 1)}{1 + C_3}(\|u - u_i\|_{L_2(\Omega)} + \operatorname{osc}_i)\right)^{-1/s}, \quad (7.2)$$

and a $v_{\mathcal{T}'} \in S_{\mathcal{T}'}$ with

$$\|u - v_{\mathcal{T}'}\|_{L_2(\Omega)} + \operatorname{osc}(\mathcal{T}') \le \frac{1 - \theta C_1(C_2 + 1)}{1 + C_3}(\|u - u_i\|_{L_2(\Omega)} + \operatorname{osc}_i).$$

As is shown in Appendix A, $\mathcal{T}'$ can be refined to a partition $\mathcal{T}'' \in \mathbb{T}_\mu$ with $\#\mathcal{T}'' - \#\mathcal{T}_0 \lesssim \#\mathcal{T}' - \#\mathcal{T}_0$, dependent on μ. The smallest common refinement $\mathcal{T}$ of $\mathcal{T}_i$ and $\mathcal{T}''$ is in

$\mathbb{T}_\mu$ with $\#\mathcal{T} - \#\mathcal{T}_i \le \#\mathcal{T}'' - \#\mathcal{T}_0$ (cf. [19] last lines of the proof of Lemma 5.2). Since $S_\mathcal{T} \subset S_{\mathcal{T}'}$, Corollary 1 shows that

$$\|u - u_\mathcal{T}\|_{L_2(\Omega)} + \mathrm{osc}(\mathcal{T}) \le (C_3 + 1) \left(\inf_{v_{\mathcal{T}'} \in S_{\mathcal{T}'}} \|u - v_{\mathcal{T}'}\|_{L_2(\Omega)} + \mathrm{osc}(\mathcal{T}') \right)$$
$$\le (1 - \theta C_1(C_2 + 1))(\|u - u_i\|_{L_2(\Omega)} + \mathrm{osc}_i),$$

and so $\eta(\mathcal{R}_{\mathcal{T}_i \to \mathcal{T}}) \ge \theta \eta_i$ by Lemma 5. Since $\mathcal{M}_i$ is the *smallest* subset of $\mathcal{T}_i$ with $\eta(\mathcal{M}_i) \ge \theta \eta_i$, we conclude that

$$\#\mathcal{M}_i \le \#\mathcal{R}_{\mathcal{T}_i \to \mathcal{T}} \le \#\mathcal{T} - \#\mathcal{T}_i \le \#\mathcal{T}'' - \#\mathcal{T}_0 \lesssim \#\mathcal{T}' - \#\mathcal{T}_0,$$

so that the proof follows from (7.2).   □

Finally, in the next theorem the quasi-optimality result is stated.

Theorem 3 *For some $s > 0$, let $u \in \mathcal{A}^s$. Then, under the assumptions from Corollary 3, it holds that*

$$\#\mathcal{T}_i - \#\mathcal{T}_0 \lesssim (\|u - u_{i-1}\|_{L_2(\Omega)} + \mathrm{osc}_{i-1})^{-1/s} |u|_{\mathcal{A}^s}^{1/s}.$$

Proof In Theorem 4 it is shown that $\#\mathcal{T}_i - \#\mathcal{T}_0 \lesssim \sum_{j=0}^{i-1} \mathcal{M}_j$. The proof follows by combining this result with Corollaries 2 and 3.   □

8 Numerical example

In this section we discuss the relationship between AFEM for controlling L_2 and energy norms and illustrate our results via a computational example. This example highlights two interesting aspects of AFEM convergence theory. First, one generally cannot expect simultaneous optimal control of the error in two different norms such as the L_2 and H^1 norms. Secondly, in contrast to the usual intuition about the relationship between L_2 and energy convergence rates, an L_2 AFEM can produce a convergence rate which is more than $\frac{1}{n}$ greater than the best possible convergence rate for an energy AFEM for the same problem.

As an example problem, we let Ω be a convex polyhedron in $\mathbb{R}^3$ having maximum edge opening angle $\frac{7\pi}{8}$ on a given edge e_{max}. In our computational example we took Ω to be the union of two tetrahedron, though the precise nature of Ω is not essential so long as the maximum edge opening angle is controlled. We then solve Poisson's problem (1.1) on Ω with right hand side data $f = 1$. By standard theory, we expect the solution u to have an edge singularity of the form $r^{\frac{8}{7}}$, where r is the distance to the edge e_{max}. In our example we also employed polynomial degree $k = 4$. Finally, the Dörfler marking parameter θ was taken to be 0.4. Because AFEM optimality requires that θ be small enough, we also tried lower values of θ but observed no appreciable difference in convergence rates when doing so. Calculations were carried out using the finite element toolbox ALBERTA (cf. [17]).

We now describe the implications of these choices. Following the discussion in Sect. 7, we expect our L_2 AFEM to converge with the rate $s = \frac{k+1}{n}$ so long as the $k + 1$-st (and thus sn-th) order partial derivatives are p-integrable for some $p > \left(\frac{1}{2} + s\right)^{-1}$. Note that because $f = 1$, no data oscillation is present in this problem, so the L_2 error is in fact equivalent to the error estimator and the approximation class $\mathcal{A}^s$ can be defined only in terms of the error. A precise characterization of $\mathcal{A}^s$ in terms of Besov spaces is also possible in this case, though heuristics phrased in terms of L_p spaces are sufficient for our purposes here. In our example we have $k + 1 = 5$ and $n = 3$, so for a convergence rate of $\frac{5}{3}$ we roughly require that $D^5 u \in L_p(\Omega)$ with $p > \frac{6}{13}$. Our example problem has been constructed so that the singularity $r^{\frac{8}{7}}$ is the strongest occurring in the solution, and it is easy to calculate that $D^5 r^{\frac{8}{7}} \approx r^{\frac{8}{7}-5}$ indeed is p-integrable for some values of $p > \frac{6}{13}$.

We now compare the expected L_2 convergence rate with the expected H^1 convergence rate for our example problem. A standard H^1 AFEM will converge with rate $s \in \left[\frac{1}{n}, \frac{k}{n}\right]$ when roughly speaking $u \in H_0^1(\Omega)$ with its $sn + 1$-th partial derivatives lying in L_p for $p > \left(\frac{1}{2} + s\right)^{-1}$. Thus for polynomial degree $k = 4$, the generally best possible convergence rate of $\frac{4}{3}$ can be obtained if $D^5 u \in L_p(\Omega)$ for some $p > \frac{6}{11}$. Note that this condition is more stringent than that required above for obtaining the generally best possible convergence rate of $\frac{5}{3}$ in L_2, and in fact for our example problem it is easy to calculate that $D^5 r^{\frac{8}{7}}$ does *not* lie in L_p for any $p > \frac{6}{11}$. Solving the conditions $D^{\alpha+1} r^{\frac{8}{7}} \in L_p(\Omega)$ with $\alpha = sn$ and $p > \left(\frac{1}{2} + s\right)^{-1}$ yields $s < \frac{8}{7}$, so we predict that a standard H^1 AFEM will converge with rate $s < \frac{8}{7} \approx 1.14$.

In our tests we ran separate computations using L_2 error indicators in the `mark` step and using H^1 error indicators in the `mark` step. In both cases, we recorded both the L_2 and H^1 error estimators, which are uniformly equivalent to the actual respective errors because no data oscillation is present. The results of these tests can be seen in Fig. 1. Approximate convergence rates were calculated by a least squares fit of logarithmic data over the range from 6×10^6 to 1.5×10^8 degrees of freedom, and corresponding trend lines are displayed on the plot along with computed data.

We note several interesting features of this plot. First, the computed convergence rate of $s = 1.65$ in the L_2 norm when using the L_2 AFEM is very close to the predicted optimal rate of $\frac{5}{3}$, while the computed convergence rate in the energy norm of 1.12 when using an energy AFEM is similarly close to the maximum predicted rate of $\frac{8}{7}$ for this particular problem. Thus as predicted by AFEM convergence theory, our L_2 AFEM produces a convergence rate which is more than $\frac{1}{n}$ greater than the best possible convergence rate for an energy AFEM for the same problem.

We next discuss convergence rates in the L_2 norm when energy refinement is used, and vice versa. When energy refinement is employed, we obtain a convergence rate $s = 1.47$ for the L_2 error, which is roughly $\frac{1}{3} = \frac{1}{n}$ more than the observed energy convergence rate of 1.12. Thus refining using energy indicators will *not* in all cases lead to optimal convergence in the L_2 norm. This difference is due to the fact that in
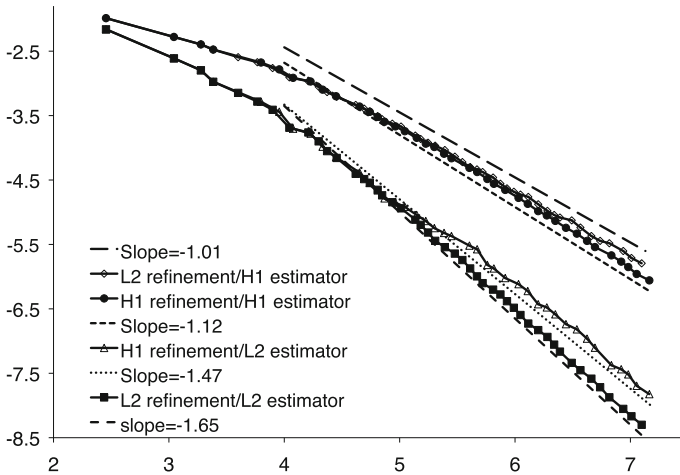
Fig. 1 Error reduction in the L_2 and energy norms when refining using H^1 error indicator ("H1 refinement") and L_2 error indicators "L2 refinement"); *plot* shows $log(estimator)$ versus $log(DOF)$

general more severe local refinement is necessary to obtain the best possible energy convergence rate than is necessary to obtain the best possible L_2 convergence rate. One would correspondingly expect that L_2 refinement will sometimes lead to suboptimal energy error decrease, and in our example we in fact observe a slight decrease in energy convergence rate (from 1.12 to 1.01) when using L_2 instead of energy refinement. Thus we conclude that simultaneous optimal error control in multiple norms is not generally possible.

We close this section with two concluding remarks about our experiments.

Remark 4 In our experiments we employed the standard refinement procedure implemented in ALBERTA, and in particular did not take extra steps to ensure that the meshes remained sufficiently mildly graded as is required in our theoretical results. We are not aware of any experimental evidence that sufficiently mild grading is necessary in practice.

Remark 5 Two-dimensional examples can be constructed for which convergence order $\frac{k+1}{2}$ is expected in the L_2 norm, but where the optimal H^1 AFEM produces a convergence rate of less than $\frac{k}{2}$. However, in contrast to the three-dimensional case such examples require a singular right-hand-side f since corner singularities have infinite smoothness in the relevant scales of Besov spaces. As an example, take Ω to be the unit square, and let $f(x, y) = x^{-0.4}$. Theoretical considerations similar to those above lead us to expect that for polynomial degree $k = 3$, the L_2 AFEM will lead to convergence rate 2 in the L_2 norm, while convergence in the H^1 norm when using the energy AFEM is expected to be less than $\frac{3}{2}$. Coding such examples in a meaningful way is not entirely straightforward, however, because accurate quadrature for singular functions must be constructed and data oscillation is also present and plays an important role in the convergence theory.

Appendix A: Partitions and mesh functions

We specify the type of partitions that we consider, and derive some of their properties. When doing so we recall some results from [20], which generalize upon known results for *newest vertex bisection* in two dimensions. In addition, we introduce the concept of sufficiently mildly graded partitions and define suitable mesh-functions.

A.1 Bisection and uniform shape regularity

For $0 \leq p \leq n - 1$, a (closed) simplex spanned by $p + 1$ vertices of an n-simplex T is called a hyperface of T. For $p = n - 1$, it will be called a *facet*. A partition $\mathcal{T}$ of a domain $\Omega \subset \mathbb{R}^n$, i.e., a subdivision of Ω into (essentially) disjoint closed n-simplices, is called *conforming* when the intersection of any two different $T, T' \in \mathcal{T}$ is either empty, or a hyperface of both simplices. Different simplices T, T' that share a facet will be called *neighbors*.

Simplices are refined by means of bisection. In order to guarantee uniform shape regularity of all descendants, a cyclic choice of the refinement edges has to be made. To that end, given $\{x_0, \ldots x_n\} \subset \mathbb{R}^n$, not on a joint $(n - 1)$-dimensional hyperplane, we distinguish between $n(n + 1)!$ *tagged* simplices given by all possible *ordered* sequences $(x_0, x_1, \ldots, x_n)_\gamma$ and *types* $\gamma \in \{0, \ldots, n - 1\}$. Given a tagged simplex $T = (x_0, x_1, \ldots, x_n)_\gamma$, its children are the tagged simplices

$$\left(x_0, \tfrac{x_0 + x_n}{2}, x_1, \ldots, x_\gamma, x_{\gamma+1}, \ldots, x_{n-1}\right)_{(\gamma+1)\mathrm{mod}n}$$

and

$$\left(x_n, \tfrac{x_0 + x_n}{2}, x_1, \ldots, x_\gamma, x_{n-1}, \ldots, x_{\gamma+1}\right)_{(\gamma+1)\mathrm{mod}n},$$

where the sequences $(x_{\gamma+1}, \ldots, x_{n-1})$ and $(x_1, \ldots, x_\gamma)$ should be read as being void for $\gamma = n - 1$ and $\gamma = 0$, respectively. So these children are defined by bisecting the edge $\overline{x_0 x_n}$ of T, i.e., by connecting its midpoint with the other vertices $x_1, \ldots, x_{n-1}$, and by an appropriate ordering of their vertices, and by having type $(\gamma + 1)\mathrm{mod}n$ (see Fig. 2 for an illustration). This bisection process was introduced in [21], and in different notations, in [10]. The edge $\overline{x_0 x_n}$ is called the *refinement edge* of T. In the $n = 2$ case, the vertex opposite to this edge is known as the *newest vertex*.

Corresponding to a tagged simplex $T = (x_0, \ldots, x_n)_\gamma$, we set

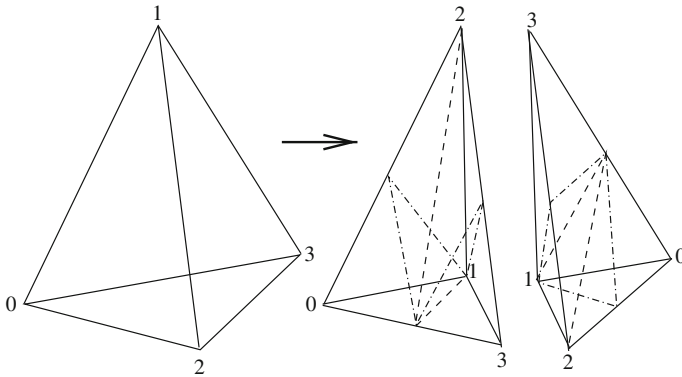$$T_{\mathrm{R}} = (x_n, x_1, \ldots, x_\gamma, x_{n-1}, \ldots, x_{\gamma+1}, x_0)_\gamma,$$

Fig. 2 Bisection of a tagged tetrahedron of type 0 with the next two level cuts indicated

which is the tagged simplex that has the same set of children as T, and in this sense is equal to T. So actually we distinguish between $\frac{1}{2}n(n+1)!$ tagged simplices.

Given a *fixed, conforming initial partition $\mathcal{T}_0$ of $\bar{\Omega}$ into tagged simplices of some fixed type γ*, we exclusively consider partitions that can be created from $\mathcal{T}_0$ by recurrent bisections of tagged simplices, for short, *descendants of $\mathcal{T}_0$*. Simplices from descendants of $\mathcal{T}_0$ are *uniformly shape regular*, dependent only on $\mathcal{T}_0$ and n. In view of the refinement by bisection, this means that there exist constants $d, D > 0$, dependent only on $\mathcal{T}_0$, such that

$$d2^{-\ell(T)} \leq \text{vol}(T) \,(= h_T^n), \quad \text{diam}(T) \leq D2^{-\ell(T)/n},$$

Here $\ell(T)$ denotes the *level* of a T, being the number of bisections needed to create T from a simplex from $\mathcal{T}_0$.

A.2 Conforming partitions

For the application of an a posteriori error estimator, we need partitions that are *conforming*. The set of conforming descendants of $\mathcal{T}_0$ is denoted as $\mathbb{T}$. A partition generated from a conforming one by bisecting some marked simplices is generally non-conforming, so that additional bisections have to be made to restore conformity.

To bound the cardinality of the output partition of a call of AFEM, we would like that the total number of additional bisections needed to restore conformity after bisecting some marked simplices can be bounded by some absolute multiple of the total number of marked simplices inside such a call. To guarantee this property, we assume that the simplices from $\mathcal{T}_0$ are tagged in such a way that any two neighbors $T = (x_0, \ldots, x_n)_\gamma$, $T' = (x_0', \ldots, x_n')_\gamma$ from $\mathcal{T}_0$ *match* in the sense that if $\overline{x_0 x_n}$ or $\overline{x_0' x_n'}$ is on $T \cap T'$, then *either T and T' are reflected neighbors*—meaning that the ordered sequence of vertices of either T or T_R coincides with that of T' on all but one position—*or* the pair of neighboring children of T and T' are reflected neighbors (see Fig. 3 for an illustration).
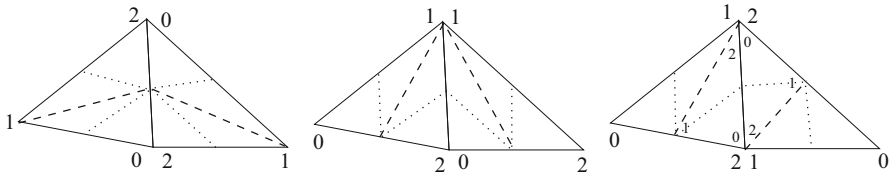
Fig. 3 Matching neighbors for $n = 2$, and their level 1 and 2 descendants. The neighbors in the rightmost picture are not reflected neighbors, but the pair of their neighboring children are

It is known (see [2] and the references therein) that for any conforming partition into triangles, i.e., for $n = 2$, there exists a local numbering of the vertices so that the matching condition is satisfied. For $n > 2$, any conforming partition of n-simplices can be refined, by inflating the number of simplices by not more than an absolute constant factor, into a conforming partition $\mathcal{T}_0$ that allows a local numbering of the vertices so that the matching condition is satisfied.

The matching condition on $\mathcal{T}_0$ is a necessary and sufficient condition so that any uniform refinement of $\mathcal{T}_0$—i.e., a refinement of $\mathcal{T}_0$ in which all simplices have the same level—is conforming. For this reason, *we always assume that $\mathcal{T}_0$ satisfies the matching condition.*

Tagged neighbors are called *compatibly divisible* when they have the same refinement edge. For a descendant $\mathcal{T}$ of $\mathcal{T}_0$, and $T \in \mathcal{T}$, we set

$$N(\mathcal{T}, T) :=$$
$$\{\text{neighbors } T' \text{ of } T \text{ in } \mathcal{T} \text{ that contain the refinement edge of } T\}.$$

As a consequence of the matching condition, we have the following result.

Corollary 4 *For any conforming descendant $\mathcal{T}$ of $\mathcal{T}_0$, and $T \in \mathcal{T}$, it holds that*

1. $|\ell(T) - \ell(T')| \leq 1$ *for any neighbor $T' \in \mathcal{T}$ of T,*
2. *and for $T' \in N(\mathcal{T}, T)$, either*
   – $\ell(T') = \ell(T)$ *and T, T' are compatibly divisible, or*
   – $\ell(T') = \ell(T) - 1$ *and T is compatibly divisible with one of both children of T'.*

In view of the second part of this corollary, an algorithm $\mathsf{bisect}[\mathcal{T}, T]$ for finding the smallest $\mathcal{T} \subset \mathcal{T}' \in \mathbb{T}$ in which some selected simplex $T \in \mathcal{T}$ is bisected may consist of the following two steps: firstly, by recursive calls of $\mathsf{bisect}[\mathcal{T}, T'']$ for some T'' with $\ell(T'') = \ell(T) - 1$, construct a possibly refined partition $\mathcal{T} \subseteq \mathcal{T}' \in \mathbb{T}$ that contains a subset $T \in K \subset \mathcal{T}'$ so that all $T' \in K$ share their refinement edge with T. Secondly, bisect all these $T' \in K$ simultaneously, with which thus no "hanging nodes" are created.

A.3 Graded partitions and the routine "bisect"

For purposes of proving convergence and optimality of L_2 AFEM, we need partitions $\mathcal{T} \in \mathbb{T}$ whose (possible) grading is sufficiently mild. Given a constant $G > 0$, we call

a descendant $\mathcal{T}$ of $\mathcal{T}_0$ *G-graded* when for all $T, T' \in \mathcal{T}$,

$$\text{dist}(T, T') \leq G2^{-\ell(T)/n} \implies \ell(T') \geq \ell(T) - 1.$$

Thus *G*-gradedness locally restricts the *growth* of the size of mesh elements as one moves away from a given element. If we want to bisect a simplex in such a partition, then generally additional bisections have to be made not only to guarantee conformity of the resulting partition, but also its *G*-gradedness. The following routine bisect$[\mathcal{T}, T]$ ensures both properties. The first part of this routine, whose steps were already described at the end of the previous section, can be found in [20], whereas its last, additional loop before the actual bisection of T in the last line ensures the *G*-gradedness of the output partition.

```
bisect[𝒯, T] → 𝒯′:
% 𝒯 ∈ 𝕋 is G-graded, and T ∈ 𝒯.
K := ∅; F = {T}
do Fnew := ∅
   forall T′ ∈ F do
      forall T″ ∈ N(𝒯, T′) with T″ ∉ F ∪ K do
         if T″ compatibly divisible with T′
         then Fnew := Fnew ∪ {T″}
         else 𝒯 := bisect[𝒯, T″]
              add to Fnew the child of T″ that is a
              neighbor of T′
         endif
      endfor
   endfor
   K := K ∪ F
   F := Fnew
until F = ∅
forall T″ ∈ 𝒯 with dist(K, T″) ≤ G2^(-(ℓ(T)+1)/n) and
   ℓ(T″) = ℓ(T) − 1 do
   𝒯 := bisect[𝒯, T″]
endfor
Create 𝒯′ from 𝒯 by simultaneously bisecting all T′ ∈ K
```

Properties of bisect are stated in the following two lemmas. Similar to [20, Theorem 5.1], the first lemma can be proven by induction to $\ell(T)$.

Proposition 7 $\mathcal{T}' := \text{bisect}[\mathcal{T}, T]$ *terminates, and* $\mathcal{T}'$ *is the smallest G-graded refinement of* $\mathcal{T}$ *in* $\mathbb{T}$ *in which* T *has been bisected.*

If $T' \in \mathcal{T}'$ *is newly created by the call, then* $\ell(T') \leq \ell(T) + 1$.

Remark 6 Assuming that the data structure allows us to find all neighbors of a simplex in $\mathcal{T} \in \mathbb{T}$ in $\mathcal{O}(1)$ operations, the number of operations needed for $\mathcal{T}' := \text{bisect}[\mathcal{T}, T]$ is $\mathcal{O}(\#\mathcal{T}' - \#\mathcal{T})$.

Proposition 8 *Any newly created T' by the call* **bisect**$[\mathcal{T}, T]$ *satisfies*

$$\text{dist}(T', T) \leq (D(1 + 2^{1/n}) + 2^{-1/n}G) \sum_{i=\ell(T')}^{\ell(T)} 2^{-i/n} \quad (\approx \text{diam}(T')).$$

Proof For $\ell(T) = 0$, any newly created T' is a child of a $\tilde{T}$ that has its refinement edge on ∂T, so that $\text{dist}(T', T) = 0$. Note that in this case the sum over i is empty since $\ell(T') = \ell(T) + 1$.

Assuming that the theorem holds for $\ell(T) = \ell - 1 \geq 0$, let us consider T with $\ell(T) = \ell$. If T' is created by bisection of any simplex from the set K, then $\text{dist}(T', T) = 0$ as in the $\ell(T) = 0$ case. If T' is created by a recursive call **bisect**$[\mathcal{T}, T'']$, then either $\text{dist}(T'', T) = 0$ or

$$\text{dist}(T'', T) \leq \text{dist}(T'', K) + D2^{-\ell(T)/n} \leq (D + 2^{-1/n}G)2^{-\ell(T)/n},$$

where we used that for any $\tilde{T} \in K$, $\ell(\tilde{T}) = \ell(T)$ and thus $\text{diam}(\tilde{T}) \leq D2^{-\ell(T)/n}$. Now by $\ell(T'') = \ell(T) - 1$, the induction hypothesis shows that

$$\text{dist}(T', T) \leq \text{dist}(T', T'') + \text{diam}(T'') + \text{dist}(T'', T)$$
$$\leq (D(1 + 2^{1/n}) + 2^{-1/n}G) \sum_{i=\ell(T')}^{\ell(T'')} 2^{-i/n}$$
$$+ D2^{1/n}2^{-\ell(T)/n} + (D + 2^{-1/n}G)2^{-\ell(T)/n}$$
$$= (D(1 + 2^{1/n}) + 2^{-1/n}G) \sum_{i=\ell(T')}^{\ell(T)} 2^{-i/n}.$$

$\square$

Our AFEM is a loop of the following form:

```
𝒯 := 𝒯₀
do mark some set ℳ ⊂ 𝒯 for bisection
   for T ∈ ℳ do
      if T ∈ 𝒯   % i.e., if it has not already been bisected as a
                  % byproduct of a previous call of bisect in this
                  % for-loop
      then 𝒯 := bisect[𝒯, T]
      endif
   endfor
until satisfied
```

Remark 7 Actually, we mark simplices possibly for some fixed number of multiple bisections. By scheduling such multiple bisections as a sequence of single ones, the algorithm can still be written in the above form.

The following theorem shows the important result that the difference between the cardinalities of the output and initial partition can be bounded by some absolute multiple of the total number of marked simplices. Besides the uniform shape regularity, the proof given in [20] relies only on the result of Proposition 8 and the second statement from Proposition 7. As we have shown, these results are still valid when considering conforming partitions that are additionally G-graded, and so is the theorem.

We emphasize that the proof from [20] is a harmless modification of the original, ingenious proof given in [2].

Theorem 4 *Let K be the total number of calls of bisect in the above AFEM loop, so that K is no larger than the sum of the cardinalities of all sets of marked simplices. Then for the output partition $\mathcal{T}$, it holds that $\#\mathcal{T} - \#\mathcal{T}_0 \lesssim K$, dependent only on the constants d, D, G, and n.*

Any descendant $\tilde{\mathcal{T}}$ of $\mathcal{T}_0$, not necessarily conforming or G-graded, is constructed from $\mathcal{T}_0$ by a sequence of bisections, where $\#(\tilde{\mathcal{T}} \setminus \mathcal{T}_0)$ is equal to the number of bisections. By replacing these bisections by calls of bisect, we infer the following consequence of Theorem 4. It shows that there is no essential restriction in considering only G-graded partitions from $\mathbb{T}$.

Corollary 5 *Any descendant $\tilde{\mathcal{T}}$ of $\mathcal{T}_0$ can be refined to a G-graded partition $\mathcal{T} \in \mathbb{T}$ with $\#\mathcal{T} - \#\mathcal{T}_0 \lesssim \#\tilde{\mathcal{T}} - \#\mathcal{T}_0$, dependent only on d, D, G, and n.*

Remark 8 For any descendant $\mathcal{T}$ of $\mathcal{T}_0$, it holds that $\#\mathcal{T} - \#\mathcal{T}_0 \leq \#(\mathcal{T} \setminus \mathcal{T}_0) \leq 2(\#\mathcal{T} - \#\mathcal{T}_0)$.

A.4 Mesh functions

In this section we construct piecewise linear mesh functions $h_{\mathcal{T}} \in C(\bar{\Omega}) \cap \prod_{T \in \mathcal{T}} P_1(T)$ such that for any $\mu > 0$, $\|\nabla h_{\mathcal{T}}\|_{L_\infty(\Omega)} \leq \mu$ for all G-graded $\mathcal{T} \in \mathbb{T}$ with $G \cdot \mu$ sufficiently large, and such that $h_{\mathcal{T}}|_T \eqsim h_T$ uniformly in μ, $T \in \mathcal{T}$ and in all those G-graded $\mathcal{T} \in \mathbb{T}$. Moreover, these mesh functions will be pointwise non-increasing under conforming and G-graded refinements. The class of G-graded $\mathcal{T} \in \mathbb{T}$ for this value of G can thus be taken as the class $\mathbb{T}_\mu$ as introduced in Sect. 2. In view of Corollary 5, this class is sufficiently large for our purposes.

For a $\mathcal{T} \in \mathbb{T}$, z a vertex of $\mathcal{T}$, and $p \in \mathbb{N}_0$, we define the patches ("rings")

$$R_p(z, \mathcal{T}) = \begin{cases} \cup_{\{T \in \mathcal{T} : T \ni z\}} T & \text{when } p = 0, \\ \cup_{\{z' : z' \text{ is a vertex of } T \in R_0(z, \mathcal{T})\}} R_{p-1}(z', \mathcal{T}) & \text{when } p > 0. \end{cases}$$

Our approach to constructing a "flat" mesh function $h_{\mathcal{T}}$ is to define $h_{\mathcal{T}}(z)$ as some average of h_T over $T \in R_p(z, \mathcal{T})$ with p being sufficiently large. To ensure that at the same time it is equivalent to the mesh size near z, it will be needed that G is sufficiently large.

In the next proposition, we essentially establish that properly sized element rings in G-graded meshes are in fact quasi-uniform, i.e., all elements in such rings are at almost the same refinement level.

Proposition 9 *For $G \geq (2p+1)D2^{1/n}$, a G-graded $\mathcal{T} \in \mathbb{T}$, and z a vertex of $\mathcal{T}$, the levels of any two simplices in $R_p(z, \mathcal{T})$ differ at most one.*

Proof Assume that $T, T'' \in R_p(z, T)$, and in order to reach a contradiction assume in addition that $\ell(T'') < \ell(T) - 1$. We take T'' to be the closest element to T lying in $R_p(z, T)$ and satisfying $\ell(T'') < \ell(T) - 1$ in the sense that there is a path P along element edges from T to T'' which lies in $R_p(z, T)$, which has the smallest possible number of edges over all such paths from T to T'', and for which each edge in the path P corresponds to an element T' satisfying $\ell(T') \geq \ell(T) - 1$. Since $T, T'' \in R_p(z, T)$, the path P contains at most $2p+1$ element edges, each of length at most $D2^{\frac{-\ell(T)+1}{n}}$. Thus $\mathrm{dist}(T, T'') \leq (2p+1)D2^{-\frac{\ell(T)+1}{n}} \leq G2^{-\frac{\ell(T)}{n}}$, which in contradiction to our initial assumption implies that $\ell(T'') \geq \ell(T) - 1$ since $\mathcal{T}$ is G-graded with $G \geq (2p+1)D2^{1/n}$. Thus $\ell(T'') \geq \ell(T) - 1$ and similarly $\ell(T) \geq \ell(T'') - 1$. That is, $\ell(T'') + 1 \geq \ell(T) \geq \ell(T'') - 1$, which is the desired conclusion. $\square$

Remark 9 Note that a patch $R_p(z, \mathcal{T})$ is not necessarily simply connected. An example is given in Fig. 4.

For z a vertex of a conforming partition $\mathcal{T}$, let

$$\ell(z) := \min_{\{T \in \mathcal{T}: T \ni z\}} \ell(T)$$

be the level of z (in $\mathcal{T}$). Next it is shown that $|\ell(T) - \ell(z)| \leq 1$ for any $T \in \mathcal{T}$ that intersects an Euclidean ball around z with radius being a fixed multiple of the mesh size near z, assuming G be sufficiently large, and thus that the number of such T is uniformly bounded.

Lemma 6 *For a G-graded $\mathcal{T} \in \mathbb{T}$, $\rho \in (0, G2^{-2/n}]$, and z a vertex of $\mathcal{T}$, any two vertices in $B(z, \rho 2^{-\ell(z)/n}) \cap \mathcal{T}$ are connected by a path along the edges of $\mathcal{T}$ of length at most L edges, where L depends only on ρ, n and the shape regularity parameters d and D.*

Furthermore, for any $T \in \mathcal{T}$ with $T \cap B(z, \rho 2^{-\ell(z)/n}) \neq \emptyset$, $|\ell(T) - \ell(z)| \leq 1$.

Proof For all $T \in \mathcal{T}$ with $T \cap B(z, \rho 2^{-\ell(z)/n}) \neq \emptyset$, it holds that $\ell(T) \leq \ell(z) + 1$. Indeed, assume that such a T exists with $\ell(T) \geq \ell(z) + 2$. Then, using that the levels of neighboring simplices in $\mathcal{T}$ differ at most one (Corollary 4(1)), there exists

Fig. 4 An example of a non-simply connected $R_1(z, \mathcal{T})$

z
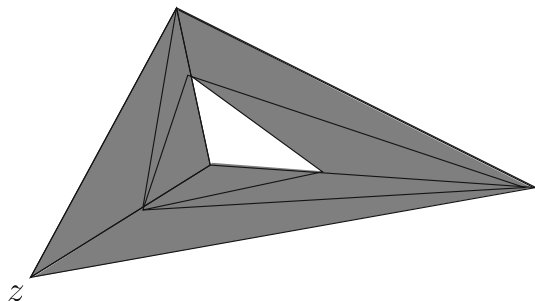
such a T with $\ell(T) = \ell(z) + 2$. However, the existence of such a T is in conflict with the G-gradedness of the partition since $\mathrm{dist}(T, z) \leq G2^{-(\ell(z)+2)/n}$. Finally, the G-gradedness of the partition also implies that $\ell(T) \geq \ell(z) - 1$ for any $T \in \mathcal{T}$ with $T \cap B(z, \rho2^{-\ell(z)/n}) \neq \emptyset$.

We infer that the number of $T \in \mathcal{T}$ with $T \cap B(z, \rho2^{-\ell(z)/n}) \neq \emptyset$ is uniformly bounded, dependent only on ρ, n, d and D, and, consequently, that any two vertices in $B(z, \rho2^{-\ell(z)/n}) \cap \mathcal{T}$ are connected by a path along the edges of $\mathcal{T}$ with a length that is uniformly bounded. $\square$

In the next proposition it is shown that $\#R_p(z, \mathcal{T}) \eqsim (p + 1)^n$ when $G \gtrsim p$ (and when $\ell(z)$ is sufficiently large in relation to p since otherwise the ring "overflows" the domain Ω is all directions). To prove that $\#R_p(z, \mathcal{T}) \gtrsim (p + 1)^n$, it will be demonstrated that if the Euclidean distance between vertices $z, z' \in \mathcal{T}$ is $M \lesssim G$ times the local mesh size at z (or equivalently at z' since G be sufficiently large), then there exists a path over the edges of $\mathcal{T}$ of length $\mathcal{O}(M)$ that connects z and z'. This improves upon Lemma 6 where no linear dependence of the path length L on ρ was demonstrated.

Proposition 10 *It holds that*

$$\#R_p(z, \mathcal{T}) \lesssim (p + 1)^n$$

uniformly in $p \in \mathbb{N}_0$, all G-graded $\mathcal{T} \in \mathbb{T}$ with $G \geq (2p + 1)D2^{1/n}$ and vertices z of $\mathcal{T}$, as well as

$$\#R_p(z, \mathcal{T}) \gtrsim (p + 1)^n$$

when additionally $p2^{-\ell(z)/n} \lesssim 1$, dependent only on the domain Ω.

Finally, it holds that

$$\frac{\#R_p(z, \mathcal{T})}{\sum_{q=0}^{p} \#R_q(z, \mathcal{T})} \lesssim \frac{1}{p + 1},$$

uniformly in $p \in \mathbb{N}_0$, all G-graded partitions $\mathcal{T} \in \mathbb{T}$ with $G \geq (2p + 1)D2^{1/n}$ and vertices z of $\mathcal{T}$.

Proof From Proposition 9, we know that $\mathrm{diam}(T) \eqsim 2^{-\ell(z)/n}$ uniformly in $T \in R_p(z, \mathcal{T})$, and thus that $R_p(z, \mathcal{T})$ is contained in a ball having radius not larger than some absolute multiple of $(p+1)2^{-\ell(z)/n}$. Because $\mathrm{vol}(T) \gtrsim 2^{-\ell(z)}$ for $T \in R_p(z, \mathcal{T})$, we conclude that $\#R_p(z, \mathcal{T}) \lesssim (p + 1)^n$.

For any $z \in \mathcal{T}$, there exists a cone C with vertex z and *fixed* height and opening, dependent only on Ω, such that $C \subset \bar{\Omega}$. We are going to show that there exists a constant $c_{\mathrm{st}} > 0$, such that for all $M \geq 1$,

$$C \cap B(z, M2^{-\ell(z)/n}) \cap \mathcal{T} \subset R_p(z, \mathcal{T}) \quad (G \geq (2p + 1)D2^{1/n}, \ p \geq c_{\mathrm{st}}M). \quad (8.1)$$

Since $\mathrm{vol}(B(z, M2^{-\ell(z)/n}) \cap C) \gtrsim M^n 2^{-\ell(z)}$ when $M2^{-\ell(z)/n} \lesssim 1$, dependent only on the sizes of C, and $\mathrm{vol}(T) \lesssim 2^{-\ell(z)}$ for $T \in R_p(z, \mathcal{T})$, (8.1) implies that $\#R_p(z, \mathcal{T}) \gtrsim (p + 1)^n$ when $p2^{-\ell(z)/n} \lesssim 1$.

Let ρ be a fixed, sufficiently large constant to be determined below. Since $G \geq (2p+1)D2^{1/n}$ and $p \geq c_{st}M \geq c_{st}$, by taking c_{st} sufficiently large, we have $\rho < G2^{-2/n}$.

Given $M \geq 1$, let $z' \in C \cap B(z, M2^{-\ell(z)/n})$ be a vertex of $\mathcal{T}$ and $M' := |z - z'|2^{\ell(z)/n}$. If $M' \leq \rho$, i.e., $z' \in B(z, \rho2^{-\ell(z)/n})$, then, by taking $c_{st} \geq L - 1$, where $L = L(\rho, n, d, D) \in \mathbb{N}$ is the constant from Lemma 6, we have $p \geq c_{st}M \geq L - 1$, and so z' is a vertex in $R_p(z, \mathcal{T})$ thanks to this lemma.

Now let $M' > \rho$. Since for any $T \in \mathcal{T}$ with $T \cap B(z, \rho2^{-\ell(z)/n}) \neq \emptyset$, $|\ell(T) - \ell(z)| \leq 1$ by Lemma 6, for ρ being sufficiently large in relation to the opening angle α of the cone C and the shape regularity parameter D, there exists a vertex z_1 of $\mathcal{T}$ in $C \cap B(z, \rho2^{-\ell(z)/n})$ with

$$|z_1 - z'| \leq \left(M' - \tfrac{1}{2}\rho\right) 2^{-\ell(z)/n}.$$

See Fig. 5 for an illustration.

With $z_0 := z$, a repeated application of this argument shows that there exists a sequence of vertices $z_1, z_2, \ldots$ of $\mathcal{T}$ with z_i in $C \cap B(z_{i-1}, \rho2^{-\ell(z_{i-1})/n})$ and

$$|z_i - z'| \leq |z_{i-1} - z'| - \tfrac{1}{2}\rho2^{-\ell(z_{i-1})/n} \tag{8.2}$$

whenever $|z_{i-1} - z'| > \rho2^{-\ell(z_{i-1})/n}$, and with $z_i = z'$ otherwise.

Let $m \geq 2$ be the smallest integer with $M - \tfrac{1}{2}\rho - (m-1)\tfrac{1}{2}2^{-\frac{1}{n}}\rho \leq 0$. When $p \geq (m-1)L$, for $1 \leq i \leq m - 1$ we have that any $T \in \mathcal{T}$ with $T \ni z_i$ is in $R_p(z, \mathcal{T})$ by Lemma 6. Taking $G \geq (2p+1)D2^{1/n}$ then ensures that $|\ell(z_i) - \ell(z)| \leq 1$ by Proposition 9. By substituting $\ell(z_{i-1}) \leq \ell(z) + 1$ in (8.2), we conclude that $z_m = z'$. In addition requiring that $p \geq mL - 1$ implies that $z' = z_m \in R_p(z, \mathcal{T})$.

Since $m \lesssim M$, our assumption that $p \geq mL - 1$ is satisfied when c_{st} is sufficiently large, with which we have completed the proof of (8.1), and thus that of $\#R_p(z, \mathcal{T}) \gtrsim (p+1)^n$ whenever $p2^{-\ell(z)/n} \lesssim 1$.

Combining the upper and lower bounds for $\#R_p(z, \mathcal{T})$ yields the last statement of this Proposition in the case when $p2^{-\ell(z)/n} \lesssim 1$. The above arguments show that the lower bound $\#R_p(z, \mathcal{T}) \gtrsim (p+1)^n$ will only be violated when p is large enough that
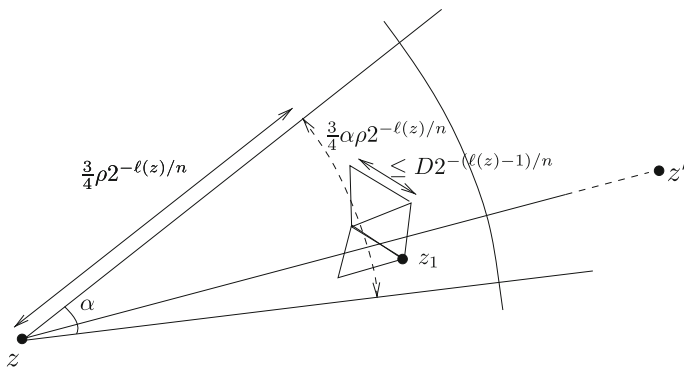

Fig. 5 Illustration with the proof of Proposition 10

$R_p(z, \mathcal{T})$ (nearly) fills the domain, i.e., that it is equal to $\mathcal{T}$. From that value of p on, $\#R_p(z, \mathcal{T})$ will not grow anymore, from which we infer the last statement also in that case. $\square$

For any $\mathcal{T} \in \mathbb{T}$, we define $\bar{h}_{\mathcal{T},p} \in C(\bar{\Omega}) \cap \prod_{T \in \mathcal{T}} P_1(T)$ by

$$\bar{h}_{\mathcal{T},p}(z) = \frac{\sum_{q=0}^{p} \sum_{T \in R_q(z,\mathcal{T})} h_T}{\sum_{q=0}^{p} \#R_q(z, \mathcal{T})}$$

for any vertex z of $\mathcal{T}$. The next proposition shows that $\bar{h}_{\mathcal{T},p}$ is appropriately called a *mesh function*. The proof follows from Proposition 9 and the uniform shape regularity.

Proposition 11 *For G-graded $\mathcal{T} \in \mathbb{T}$, it holds that*

$$\bar{h}_{\mathcal{T},p}|_T \eqsim h_T$$

uniformly in $p \in \mathbb{N}_0$, $G \geq (2p + 1)D2^{1/n}$, *and* $T \in \mathcal{T}$.

Next we show that $\bar{h}_{\mathcal{T},p}$ can be made arbitrarily flat by decreasing the grading of the partition, i.e., by increasing G, and by increasing p proportionally.

Proposition 12 *For G-graded $\mathcal{T} \in \mathbb{T}$, it holds that*

$$|\bar{h}_{\mathcal{T},p}(z) - \bar{h}_{\mathcal{T},p}(z')| \lesssim \frac{1}{p+1} h_T,$$

uniformly in $p \in \mathbb{N}_0$, $G \geq (2p + 1)D2^{1/n}$, $T \in \mathcal{T}$, *and vertices* $z, z' \in T$, *and so* $\|\nabla \bar{h}_{\mathcal{T},p}\|_{L_\infty(\Omega)} \lesssim \frac{1}{p+1}$.

Proof For $z'' \in \{z, z'\}$, let

$$n(z'') = \sum_{q=0}^{p} \sum_{T' \in R_q(z'',\mathcal{T})} h_{T'}, \quad d(z'') = \sum_{q=0}^{p} \#R_q(z'', \mathcal{T}).$$

Letting $R_{-1}(\cdot, \cdot) := \emptyset$, we have $R_q(z'', \mathcal{T}) \setminus (R_q(z, \mathcal{T}) \cap R_q(z', \mathcal{T})) \subset R_q(z'', \mathcal{T}) \setminus R_{q-1}(z'', \mathcal{T})$. Employing Proposition 10, we compute

$$\sum_{q=0}^{p} \#(R_q(z'', \mathcal{T}) \setminus R_{q-1}(z'', \mathcal{T})) = \#R_p(z'', \mathcal{T}) \lesssim \frac{1}{p+1} d(z'').$$

Applying Proposition 9 then yields $|n(z) - n(z')| \lesssim \frac{h_T}{p+1}(d(z) + d(z'))$, $|d(z) - d(z')| \lesssim \frac{1}{p+1}(d(z) + d(z'))$, and $\left|\frac{n(z'')}{d(z'')}\right| \lesssim h_T$. Assuming $d(z') \geq d(z)$, we may complete the proof by writing

$$\frac{n(z)}{d(z)} - \frac{n(z')}{d(z')} = \frac{n(z)}{d(z)}\frac{d(z') - d(z)}{d(z')} + \frac{n(z) - n(z')}{d(z')}.$$

$\square$

Remark 10 Proposition 11 is still valid if the more obvious definition $\bar{h}_{\mathcal{T},p} = \sum_{T \in R_p(z,\mathcal{T})} h_T / \#R_p(z,\mathcal{T})$ of the mesh function is used. In order to prove a result such as Proposition 12 for this mesh function, however, it is necessary to show that, for $G \geq (2p+1)D2^{1/n}$, $\#(R_p(z'',\mathcal{T}) \setminus R_{p-1}(z'',\mathcal{T}))/\#R_p(z'',\mathcal{T}) \lesssim \frac{1}{p+1}$, or at least that it tends to zero for $p \to \infty$. It turns out that such an estimate is difficult to establish, cf. also Remark 9.

For some applications it is important to have a family of mesh functions that is non-increasing under (conforming, G-graded) refinements. For G-graded $\mathcal{T} \in \mathbb{T}$, we define $h_{\mathcal{T},p} \in C(\bar{\Omega}) \cap \prod_{T \in \mathcal{T}} P_1(T)$ by

$$h_{\mathcal{T},p}(z) = \min_{\{\mathcal{T}' \in \mathbb{T} \text{ is } G\text{-graded}: \mathcal{T}' \subset \mathcal{T}\}} \bar{h}_{\mathcal{T},p}(z),$$

where z is any vertex of $\mathcal{T}$.

Proposition 13 *The mapping* $\mathbb{T} \ni \mathcal{T} \mapsto h_{\mathcal{T},p}$ *is pointwise* non-increasing *under G-graded refinements, and for some constants $c_{\mathbb{T}}$, $C_{\mathbb{T}}$, $\tilde{C}_{\mathbb{T}}$, which are independent of p, it holds that*

$$c_{\mathbb{T}} h_T \leq h_{\mathcal{T},p}|_T \leq C_{\mathbb{T}} h_T, \quad \|\nabla h_{\mathcal{T},p}\|_{L_\infty(\Omega)} \leq \tilde{C}_{\mathbb{T}}\frac{1}{p+1},$$

for all G-graded $\mathcal{T} \in \mathbb{T}$ with $G \geq (2p+1)D2^{1/n}$, and $T \in \mathcal{T}$.

Proof The first two statements follow by definition of $h_{\mathcal{T},p}$ and Proposition 11.

For vertices z_1, z_2 of $\mathcal{T}$, let $\bar{h}_{\mathcal{T}_1,p}(z_1) = h_{\mathcal{T},p}(z_1) \leq h_{\mathcal{T},p}(z_2) = \bar{h}_{\mathcal{T}_2,p}(z_2)$. Then we have $|h_{\mathcal{T},p}(z_1) - h_{\mathcal{T},p}(z_2)| \leq |\bar{h}_{\mathcal{T}_1,p}(z_1) - \bar{h}_{\mathcal{T}_1,p}(z_2)| \leq \|\nabla \bar{h}_{\mathcal{T}_1}\|_{L_\infty(\Omega)}\|z_1 - z_2\|$. Now an application of Proposition 12 shows the second statement. $\square$

We conclude that given a $\mu > 0$, for $p+1 \geq \tilde{C}_{\mathbb{T}}\mu^{-1}$ and $G \geq (2p+1)D2^{1/n}$, the collection of all G-graded $\mathcal{T} \in \mathbb{T}$ can be used as the class $\mathbb{T}_\mu$.

References

1. Babuška, I., Osborn, J.: Analysis of finite element methods for second order boundary value problems using mesh dependent norms. Numer. Math. **34**, 41–62 (1980)
2. Binev, P., Dahmen, W., DeVore, R.: Adaptive finite element methods with convergence rates. Numer. Math. **97**, 219–268 (2004)
3. Brenner, S.C., Scott, L.R.: The Mathematical Theory of Finite Element Methods. Texts in Applied Mathematics, vol. 15, 3rd edn. Springer, New York (2008)
4. Cascon, J., Kreuzer, C., Nochetto, R.H., Siebert, K.G.: Quasi-optimal convergence rate for an adaptive finite element method. SIAM J. Numer. Anal. **46**, 2524–2550 (2008)
5. Demlow, A.: Convergence of an adaptive finite element method for controlling local energy errors. SIAM J. Numer. Anal. **48**(2), 470–497 (2010)

6. Dörfler, W.: A convergent adaptive algorithm for Poisson's equation. SIAM J. Numer. Anal. **33**, 1106–1124 (1996)
7. Eriksson, K.: An adaptive finite element method with efficient maximum norm error control for elliptic problems. Math. Models Methods Appl. Sci. **4**, 313–329 (1994)
8. Eriksson, K., Johnson, C.: Adaptive finite element methods for parabolic problems. II. Optimal error estimates in $L_\infty L_2$ and $L_\infty L_\infty$. SIAM J. Numer. Anal. **32**, 706–740 (1995)
9. Liao, X., Nochetto, R.H.: Local a posteriori error estimates and adaptive control of pollution effects. Numer. Methods Partial Differ. Equ. **19**, 421–442 (2003)
10. Maubach, J.: Local bisection refinement for n-simplicial grids generated by reflection. SIAM J. Sci. Comput. **16**, 210–227 (1995)
11. Mekchay, K., Nochetto, R.H.: Convergence of adaptive finite element methods for general second order linear elliptic PDEs. SIAM J. Numer. Anal. **43**, 1803–1827 (2005) (electronic)
12. Morin, P., Nochetto, R.H., Siebert, K.G.: Convergence of adaptive finite element methods. SIAM Rev. **44**, 631–658 (2002) (electronic) (2003) Revised reprint of "Data oscillation and convergence of adaptive FEM" [SIAM J. Numer. Anal. **38**(2), 466–488 (2000) (electronic); MR1770058 (2001g:65157)]
13. Morin, P., Siebert, K., Veeser, A.: Convergence of Finite Elements Adapted for Weaker Norms. Applied and Industrial Mathematics in Italy II. Ser. Adv. Math. Appl. Sci., vol. 75, pp. 468–479. World Sci. Publ., Hackensack, NJ (2007)
14. Morin, P., Siebert, K., Veeser, A.: A basic convergence result for conforming adaptive finite elements. Math. Models Methods Appl. Sci. **18**, 707–737 (2008)
15. Nitsche, J.A., Schatz, A.H.: Interior estimates for Ritz-Galerkin methods. Math. Comput. **28**, 937–958 (1974)
16. Nochetto, R., Siebert, K., Veeser, A.: Theory of adaptive finite element methods: an introduction. In: DeVore, R., Kunoth, A. (eds.) Multiscale, Nonlinear and Adaptive Approximation: Dedicated to Wolfgang Dahmen on the Occasion of his 60th Birthday, pp. 409–542. Springer, Berlin (2009)
17. Schmidt, A., Siebert, K.G.: Design of adaptive finite element software. The finite element toolbox ALBERTA, with 1 CD-ROM (Unix/Linux). Lecture Notes in Computational Science and Engineering, vol. 42, pp. xii+315. Springer, Berlin (2005)
18. Scott, L.R., Zhang, S.: Finite element interpolation of nonsmooth functions satisfying boundary conditions. Math. Comput. **54**, 483–493 (1990)
19. Stevenson, R.: Optimality of a standard adaptive finite element method. Found. Comput. Math. **7**, 245–269 (2007)
20. Stevenson, R.: The completion of locally refined simplicial partitions created by bisection. Math. Comput. **77**, 227–241 (2008)
21. Traxler, C.T.: An algorithm for adaptive mesh refinement in n dimensions. Computing **59**, 115–137 (1997)
22. Verfürth, R.: A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques. Wiley-Teubner, Chichester (1996)
23. Wihler, T.P.: Weighted L^2-norm a posteriori error estimation of FEM in polygons. Int. J. Numer. Anal. Model. **4**, 100–115 (2007)
24. Xu, J., Zhou, A.: Local and parallel finite element algorithms based on two-grid discretizations. Math. Comput. **69**, 881–909 (2000)