Foundations of Computational Mathematics manuscript No.
(will be inserted by the editor)

Approximating gradients with continuous piecewise polynomial functions

Andreas Veeser

Received: date / Accepted: date

Abstract Motivated by conforming finite element methods for elliptic problems of second order, we analyze the approximation of the gradient of a target function by continuous piecewise polynomial functions over a simplicial mesh. The main result is that the global best approximation error is equivalent to an appropriate sum in terms of the local best approximation errors on elements. Thus, requiring continuity does not downgrade local approximation capability and discontinuous piecewise polynomials essentially do not offer additional approximation power, even for a fixed mesh. This result implies error bounds in terms of piecewise regularity over the whole admissible smoothness range. Moreover, it allows for simple local error functionals in adaptive tree approximation of gradients.

Keywords Approximation of gradients · continuous piecewise polynomials · finite elements · Lagrange elements · discontinuous elements · a priori error estimates · adaptive tree approximation

Mathematics Subject Classification (2010) 41A15 · 41A63 · 41A05 · 65N30 · 65N15

1 Introduction

Finite element methods are one of the most successful tools for the numerical solution of partial differential equations. In their simplest form they are Galerkin methods where the discrete space is given by elements that are appropriately coupled. This piecewise structure allows constructing bases that are, on the one hand, relatively easy to implement and, on the other hand, are locally supported. The latter leads

Communicated by Ronald DeVore

Andreas Veeser
Dipartimento di Matematica, Università degli Studi di Milano, Via Saldini 50, 20131 Milano, Italia
E-mail: andreas.veeser@unimi.it

to linear systems with sparse matrices, which can be stored and often solved with optimal linear complexity.

This article concerns the approximation properties of continuous piecewise polynomial functions over a simplicial mesh, which build a prototype finite element space. It analyzes the interplay of global and local best errors when approximating the gradient of a target function.

Continuous piecewise polynomial functions arise when solving elliptic boundary value problems of second order with Lagrange elements (see §2 for a definition). To be more specific, if the associated bilinear form is $H^1_0(\Omega)$-coercive, a typical choice is

$$S := \Big\{ v : \overline{\Omega} \to \mathbb{R} \mid \forall K \in \mathscr{M}\ v_{|K} \in \mathscr{P}_\ell(K), v \in C^0(\overline{\Omega}), v_{|\partial\Omega} = 0 \Big\}, \tag{1}$$

where $\mathscr{M}$ is a conforming simplicial mesh of a domain Ω and $\mathscr{P}_\ell(K)$ denotes the set of polynomials with degree $\leq \ell$ over an element $K \in \mathscr{M}$. Requiring continuity and incorporating the boundary condition in (1) ensure $S \subset H^1_0(\Omega)$ and thus the conformity of finite element method. Céa's lemma therefore implies that the error of the Galerkin solution in S is dictated by the best global approximation error for the exact solution $u \in H^1_0(\Omega)$:

$$E(u) := E(u,S) := \inf \big\{ \|\nabla(u-v)\|_\Omega \mid v \in S \big\},$$

where $\|\cdot\|_\Omega$ stands for the norm of $L^2(\Omega)$. In view of the piecewise structure of space S, the approximation of u on each element $K \in \mathscr{M}$ is limited by the local shape functions $\mathscr{P}_\ell(K)$. This suggests introducing the local best approximation errors

$$e_K(u) := \inf \big\{ \|\nabla(u-p)\|_K \mid p \in \mathscr{P}_\ell(K) \big\}, \qquad \forall K \in \mathscr{M}.$$

The question arises how the global and local best errors are related and how their relationship is affected by requiring conformity.

In the described context the main result Theorem 2 reads as follows. For any conforming mesh, the global best error is equivalent to the appropriately collected local best errors. More precisely, there holds

$$\left(\sum_{K\in\mathscr{M}} e_K(u)^2 \right)^{\frac{1}{2}} \leq E(u) \leq C \left(\sum_{K\in\mathscr{M}} e_K(u)^2 \right)^{\frac{1}{2}}, \tag{2}$$

where C can be bounded in terms of the shape regularity of $\mathscr{M}$. The first inequality in (2) is straight-forward and just a quantitative version of the motivation that suggests introducing the local best errors. The second inequality, which is not obvious and the proper concern of this paper, means that the above requirements for conformity essentially do not downgrade the approximation capability given by the local discrete spaces. It thus confirms in particular the coupling via continuity of the elements. Adopting the broken H^1-seminorm as error notion, there is a second interpretation of (2): discontinuous and continuous piecewise polynomial functions have essentially the same approximation power. We also derive variants of (2) addressing the coupling of partial derivatives of the approximants and mesh conformity; see Theorem 3 and Theorem 6, respectively.

The second inequality in (2) is proved in §3 by means of suitable local error bounds for a continuous interpolant. As interpolant, one can use a variant of the Scott-Zhang interpolant [25] or averages like in S. Brenner [11] of local best approximations. The key ingredients for these local error bounds are the Trace and Poincaré inequalities.

Equivalence (2) reduces the quantification of the global best error to the quantification of decoupled, local best approximation errors. We illustrate the usefulness of this aspect with two applications in §4.

First, the second inequality in (2) may be used in the a priori analysis of finite element solutions. For example, inserting the Bramble-Hilbert lemma in the right-hand side of (2), one readily obtains the upper bound

$$E(u) \leq C \left(\sum_{K \in \mathscr{M}} h_K^{2\ell} \left\| D^{\ell+1} u \right\|_K^2 \right)^{\frac{1}{2}}, \tag{3}$$

where h_K denotes the diameter of an element $K \in \mathscr{M}$. Notice that the right-hand side involves only piecewise regularity and so vanishes whenever $u \in S$. This is also true for Lagrange interpolation error estimates, but not for the available error bounds [15, 25] for interpolation of non-smooth functions. Here it is obtained without invoking the embedding $H^{\ell+1} \subset C^0$ and so also under weaker regularity assumptions on the target function u.

Second, (2) can be applied in constructive nonlinear approximation. When using the squared local best errors $e_K(u)^2$ as local error functionals in the adaptive tree approximation of P. Binev and R. DeVore [6], then equivalence (2) ensures that the approximations, which are constructed with linear complexity, are near best with respect to the H^1-seminorm.

2 Continuous piecewise polynomial functions and gradients

In this section we define the approximants, fix associated notation, and review their relationship with gradients. We also provide a basis for them and, to prepare interpolation, link the coefficients of that basis to the space of target functions.

Let Ω be a non-empty open set of $\mathbb{R}^d$, $d \in \mathbb{N}$. We do not assume that Ω is on one side of its boundary $\partial\Omega$. As usual, $L^2(\Omega)$ denotes the Hilbert space of real-valued functions on Ω that are measurable and square-integrable with respect to the Lebesgue measure of $\mathbb{R}^d$ and $H^1(\Omega)$ is the Hilbert space of all functions that, together with their distributional gradients, are in $L^2(\Omega)$.

Given $k \in \mathbb{N}$ with $k \leq d$, a set K is a k-simplex in $\mathbb{R}^d$ if it is the convex hull of $k+1$ points $a_0, \ldots, a_k \in \mathbb{R}^d$ that do not lie on a plane of dimension $k-1$. The set of extreme points of a convex set C is denoted by $\operatorname{Vert} C$. For example, there holds $\operatorname{Vert} K = \{a_0, \ldots, a_k\}$ in the definition of k-simplex. A m-simplex F with $m \in \{1, \ldots, k\}$ is a m-face of K if $\operatorname{Vert} F \subset \operatorname{Vert} K$. By convention, a vertex is a 0-face. As usual, h_K denotes the diameter of K, while ρ_K stands for the diameter of the largest

ball in K. The boundary of any d-simplex K in $\mathbb{R}^d$ can be represented locally by a Lipschitz function. Hence, the trace operator

$$(\cdot)_{|\partial K} : H^1(K) \to L^2(\partial K) \tag{4}$$

is well-defined. Hereafter $L^2(\partial K)$ stands for the Hilbert space of all real-valued functions on ∂K that are measurable and square-integrable with respect to the $(d-1)$-dimensional Hausdorff measure.

Assume that $\mathscr{M}$ is a conforming simplicial mesh of Ω in the following sense: $\mathscr{M}$ is a finite sequence of d-simplices in $\mathbb{R}^d$ and such that

$$\overline{\Omega} = \bigcup_{K\in\mathscr{M}} K, \qquad \forall K,K' \in \mathscr{M} \quad \operatorname{Vert}(K\cap K') \subset \operatorname{Vert} K \cap \operatorname{Vert} K', \tag{5a}$$

$$\partial\Omega = \bigcup_{F\in\mathscr{F}_{\partial\Omega}} F, \tag{5b}$$

where $\mathscr{F}_{\partial\Omega}$ is a suitable subset of $\mathscr{F}_{d-1}(\Omega)$, the set of all $(d-1)$-dimensional faces of $\mathscr{M}$. In §3 below we make an additional assumption on the mesh $\mathscr{M}$.

The second condition in (5a) has two implications. First, it ensures that $\mathscr{M}$ is a non-overlapping covering in that there holds

$$|\Omega| = \sum_{K\in\mathscr{M}} |K|$$

for the Lebesgue measure $|\cdot|$ in $\mathbb{R}^d$. Second, it entails that $\mathscr{M}$ is conforming or face-to-face, i.e. for any two 'elements' $K,K' \in \mathscr{M}$, the intersection $K\cap K'$ is a k-face of both d-simplices K and K' for some $k \in \{0,\dots,d\}$.

Assumption (5b) allows for domains like the slit domain $\{x=(x_1,x_2)\in\mathbb{R}^2 \mid \max\{|x_1|,|x_2|\}<1, x_2\neq 0 \text{ or } x_1<0\}$. For the latter example, the usual definition of the trace operator $H^1(\Omega)\to L^2(\partial\Omega)$ does not apply as the boundary is not locally a graph of a function. Nevertheless, exploiting (4) and (5b), we can define that a function $v\in H^1(\Omega)$ equals a function $g\in L^2(\partial\Omega)$ on the boundary by

$$v_{|\partial\Omega} = g \quad :\Longleftrightarrow \quad \forall K\in\mathscr{M}\ v_{|\partial\Omega\cap\partial K} = g_{|\partial\Omega\cap\partial K}.$$

As mentioned in the introduction §1, the space

$$S_0^{\ell,0}(\mathscr{M}) := \{V : \overline{\Omega}\to\mathbb{R} \mid \forall K\in\mathscr{M}\ V_{|K}\in\mathscr{P}_\ell(K),\ V\in C^0(\overline{\Omega}),\ V_{|\partial\Omega}=0\}$$

with $\ell\in\mathbb{N}$ may be used when approximating functions in

$$H_0^1(\Omega) := \{v\in L^2(\Omega) \mid \nabla v\in L^2(\Omega),\ v_{|\partial\Omega}=0\}. \tag{6}$$

The first property in the definition of $S_0^{\ell,0}(\mathscr{M})$ determines the basic nature of the approximants: their piecewise structure and that each one can be identified with a finite number of parameters. The role of the other two properties, which constrain this basic nature, is clarified by the following proposition in terms of the space

$$S^{\ell,-1}(\mathscr{M}) := \{V : \overline{\Omega}\to\mathbb{R} \mid \forall K\in\mathscr{M}\ V_{|K}\in\mathscr{P}_\ell(K)\}$$

of all functions that are piecewise polynomial over $\mathscr{M}$ and the space

$$S^{\ell,0}(\mathscr{M}) := \{V \in S^{\ell,-1}(\mathscr{M}) \mid V \in C^0(\overline{\Omega})\}$$

of all functions that are in addition continuous.

Proposition 1 (Characterization of H^1- and H^1_0-conformity) *A piecewise polynomial function $V \in S^{\ell,-1}(\mathscr{M})$ is in $H^1(\Omega)$ if and only if it is continuous in $\overline{\Omega}$. Moreover, a continuous piecewise polynomial function $V \in S^{\ell,0}(\mathscr{M})$ is in $H^1_0(\Omega)$ if and only if $V_{|\partial\Omega} = 0$.*

Proof The first equivalence is a consequence of [10, Chapter II, Theorem 5.1], while then the second one immediately follows from definition (6). □

The requirements $V \in C^0(\overline{\Omega})$ and $V_{|\partial\Omega} = 0$ are therefore sufficient and necessary for conformity of the approximants. Clearly, the first requirement $V \in C^0(\overline{\Omega})$ is independent from the considered boundary condition $v_{|\partial\Omega} = 0$; see also Corollaries 1 and 2 below.

Next, we recall the Lagrange basis of $S^{\ell,0}(\mathscr{M})$, $\ell \in \mathbb{N}$. The principal Lagrange lattice of order ℓ of a k-simplex $K = \mathrm{Conv}\{a_0, \dots, a_k\}$ in $\mathbb{R}^d$ is given by

$$L_\ell(K) := \left\{ \frac{1}{\ell} \sum_{i=0}^{k} \alpha_i a_i \mid \alpha = (\alpha_0, \dots, \alpha_k) \in \mathbb{N}_0^{k+1},\ |\alpha| = \ell \right\},$$

where $|\alpha| := \sum_{i=0}^{d} \alpha_i$ denotes the length of the multi-index α. Fixed a d-simplex K, these lattices have the following property: if F is a k-face of K, then

$$L_\ell(K) \cap F = L_\ell(F). \tag{7}$$

In order to exploit the affine equivalence of simplices and corresponding lattices, we fix $K_d = \mathrm{Conv}\{0, e_1, \dots, e_d\}$ as d-dimensional reference simplex and associate nodes of a given element $K = \mathrm{Conv}\{a_0, \dots, a_d\} \in \mathscr{M}$ to the ones of K_d in the following unique manner. Given $z \in L_\ell(K)$, write $z = \sum_{i=0}^{d} \lambda_i a_i$ as a convex combination of the vertices of K, rearrange the coefficients in decreasing order such that $\lambda_0^* \geq \cdots \geq \lambda_d^*$ and set

$$\hat{z} := \sum_{i=1}^{d} \lambda_i^* e_i \in L_\ell(K_d). \tag{8}$$

Although the rearrangement may be not unique, $\hat{z}$ is well-defined and does not depend on the enumeration of the vertices of K. Moreover, if we identify $\mathbb{R}^k$ with $\mathbb{R}^k \times \{0\} \times \cdots \times \{0\} \subset \mathbb{R}^d$, the reference node $\hat{z}$ does not depend on the vertices with non-zero coefficients in the representation of z, i.e. it does not matter if z is viewed as a node of the simplex K or, if possible, of some of its k-faces.

Lemma 1 (Lagrange basis) *Setting*

$$L_\ell(\mathscr{M}) := \bigcup_{K \in \mathscr{M}} L_\ell(K),$$

there are functions $\{\Phi_z\}_{z\in L_\ell(\mathscr{M})}$ *such that*

$$\Phi_z \in S^{\ell,0}(\mathscr{M}) \qquad \text{and} \qquad \forall y \in L_\ell(\mathscr{M}) \quad \Phi_z(y) = \delta_{yz}.$$

These functions are the Lagrange basis of $S^{\ell,0}(\mathscr{M})$ *and satisfy in particular:*

(i) Any $V \in S^{\ell,0}(\mathscr{M})$ *has the representation*

$$V = \sum_{z\in L_\ell(\mathscr{M})} V(z)\Phi_z.$$

(ii) Each function Φ_z *is locally supported:*

$$\operatorname{supp}\Phi_z = \overline{\{x\in\Omega \mid \Phi_z \neq 0\}} = \omega_z := \bigcup_{K\in\mathscr{M}:K\ni z} K.$$

(iii) It holds

$$\|\Phi_z\|_{0,2;K} = \sqrt{d!}\,|K|^{\frac{1}{2}} \left\|\hat{\Phi}_{\hat{z}}\right\|_{0,2;K_d}$$

where $|K|$ *is the* d*-dimensional Lebesgue measure of* K*, the reference node* $\hat{z}$ *is given by* (8) *and* $\hat{\Phi}_{\hat{z}} \in \mathscr{P}_\ell(K_d)$ *is the polynomial that is 1 at* $\hat{z}$ *and vanishes at the other Lagrange nodes of* K_d*.*

Proof The proofs of the existence of the Lagrange basis, of (i) and (ii) can be found in, e.g., [10] or [13]. In view of Proposition 1, it is worth recalling that the requirement $\Phi_z(y) = \delta_{yz}$ for all $y \in L_\ell(\mathscr{M})$ entails the continuity of Φ_z by an interplay of (7), the implication

$$\forall P,Q \in \mathscr{P}_\ell(K) \qquad P_{|L_\ell(K)} = Q_{|L_\ell(K)} \implies P = Q \text{ on } K$$

for any simplex K of dimension $k \in \{0,\dots,d\}$ and the conformity in (5a) of the mesh $\mathscr{M}$. To show (iii), recall $|K_d| = 1/d!$ and apply the transformation rule with an affine mapping $A : \mathbb{R}^d \to \mathbb{R}^d$ such that $A(K_d) = K$ and $A(\hat{z}) = z$. □

It is useful to extend the so-called global nodal variables

$$S^{\ell,0}(\mathscr{M}) \ni V \mapsto V(z) \in \mathbb{R} \tag{9}$$

to functions in $H^1(\Omega)$. To this end, we invoke the construction of L. R. Scott and S. Zhang [25], which exploits the trace theorem (4) and involves the polynomials

$$\Psi_z^F \in \mathscr{P}_\ell(F) \quad \text{and} \quad \forall y \in L_\ell(F) \int_F \Phi_y \Psi_z^F = \delta_{yz} \tag{10}$$

where $F \in \mathscr{F}_{d-1}(\mathscr{M})$ and $\mathscr{F}_{d-1}(\mathscr{M})$ denotes the set of all $(d-1)$-dimensional faces of $\mathscr{M}$.

Lemma 2 (H^1-extension of nodal variables) *Let* $z \in L_\ell(\mathscr{M})$ *and* $F \in \mathscr{F}_{d-1}(\mathscr{M})$ *be such that* $z \in F$*. Then the functional*

$$N_{z,F}(v) := \int_F v\Psi_z^F, \qquad v \in L^2(F),$$

has the following properties:

(i) *If an element $K \in \mathscr{M}$ contains F and $v \in H^1(K)$, then $N_{z,F}(v)$ is defined. Furthermore, if $v \in \mathscr{P}_\ell(K)$, then $N_{z,F}(v) = v(z)$.*

(ii) *If $v \in H^1_0(\Omega)$, then $N_{z,F}(v) = 0$ whenever $F \subset \partial\Omega$.*

(iii) *It holds*

$$|N_{z,F}(v)| \leq \frac{1}{\sqrt{(d-1)!}} |F|^{-\frac{1}{2}} \left\|\hat{\Psi}_{\hat{z}}\right\|_{0,2;K_{d-1}} \|v\|_{0,2;F},$$

where $|F|$ is the $(d-1)$-dimensional Hausdorff measure of F, the reference node $\hat{z} \in K_{d-1}$ is given by the counterpart of (8) *and $\hat{\Psi}_{\hat{z}}$ is the $L^2(K_{d-1})$-dual basis function corresponding to $\hat{\Phi}_{\hat{z}}$.*

Proof Item (i) readily follows from the trace theorem (4) and from the following fact: the orthogonality condition in (10) extends to

$$\forall y \in L_\ell(\mathscr{M}) \quad \int_F \Phi_y \Psi_z^F = \delta_{yz},$$

because it holds $\Phi_{y|F} = 0$ for every $y \in L_\ell(\mathscr{M}) \setminus L_\ell(F)$. Item (ii) is immediate since we have $v_{|F} = 0$ in such cases. In order to show the remaining item (iii), we take a bi-affine transformation $A : K_{d-1} \to F$ with $A\hat{z} = z$. Applying the transformation rule to the right-hand side of the identity in (10), we obtain the relationship

$$\Psi_z^F = \frac{1}{(d-1)!} |F|^{-1} \hat{\Psi}_{\hat{z}} \circ A^{-1}.$$

Then, a second, direct application of the transformation rule yields

$$\left\|\Psi_z^F\right\|_{0,2;F} = \frac{1}{\sqrt{(d-1)!}} |F|^{-\frac{1}{2}} \left\|\hat{\Psi}_{\hat{z}}\right\|_{0,2;K_{d-1}}$$

and the claimed bound follows from the Cauchy-Schwarz inequality. □

3 Conformity and approximation error

Proposition 1 determines conditions that characterize when piecewise polynomial functions are conforming. The main goal of this section is to analyze the impact of these conditions on the error when approximating the gradient of a function.

3.1 Global and local best errors

We first provide a notion that measures the possible downgrading resulting from conformity. In order to take into account boundary conditions, we consider the following setting for the space X of target functions and the approximants $S(\mathscr{M})$ over a mesh $\mathscr{M}$. Assume that X and $S(\mathscr{M})$ are, respectively, closed affine subspaces of $H^1(\Omega)$ and $S^{\ell,0}(\mathscr{M})$ with $\ell \geq 1$ and that the H^1-seminorm

$$|w|_\Omega := \||\nabla w|\|_{0,2;\Omega} = \left(\int_\Omega |\nabla w|^2\right)^{\frac{1}{2}} \tag{11}$$

is definite on $X - S(\mathscr{M})$. The setting $X = H^1_0(\Omega)$ and $S(\mathscr{M}) = S^{\ell,0}_0(\mathscr{M})$ in the introduction §1 is an example.

The best (possible) error of approximating $v \in X$ is then given by

$$E(v,\mathscr{M}) := E\big(v, S(\mathscr{M})\big) := \inf_{V \in S(\mathscr{M})} |v - V|_\Omega \tag{12}$$

and $E(v,\mathscr{M}) = 0$ implies $v \in S(\mathscr{M})$. Thanks to, e.g., the Projection Theorem in Hilbert spaces, there exists a unique best approximation $V_\mathscr{M}$ such that

$$|v - V_\mathscr{M}|_\Omega = E(v,\mathscr{M}).$$

If $S(\mathscr{M})$ is a linear space, this is equivalent to

$$\forall W \in S(\mathscr{M}) \quad \int_\Omega \nabla V_\mathscr{M} \cdot \nabla W = \int_\Omega \nabla v \cdot \nabla W.$$

and $V_\mathscr{M}$ is called the Ritz projection of v onto $S(\mathscr{M})$.

Similarly, on each single element $K \in \mathscr{M}$, the best error is given by

$$e(v,K) := e\big(v, \mathscr{P}_\ell(K)\big) := \inf_{P \in \mathscr{P}_\ell(K)} |v - P|_K,$$

which depends only on the local gradient $(\nabla v)_{|K}$ of the target function and the shape functions associated with K. Applying the Projection Theorem in the Hilbert space $H^1(K)/\mathbb{R}$, we see that here best approximations also exist but are only unique up to a constant. Let P_K be the best approximation that has the same mean value on K as the target function v. In other words, P_K is characterized by

$$P_K \in \mathscr{P}_\ell(K), \quad \int_K P_K = \int_K v \quad \text{and} \quad |v - P_K|_K = e(v,K), \tag{13}$$

the latter being equivalent to

$$\forall Q \in \mathscr{P}_\ell(K) \quad \int_K \nabla P_K \cdot \nabla Q = \int_K \nabla v \cdot \nabla Q. \tag{14}$$

Since the local best errors cannot be overtaken by any global approximation, one may expect that an appropriate 'sum' of them provides a lower bound for corresponding global best errors. In fact, if S' is any subspace of $S^{\ell,-1}(\mathscr{M})$, one readily verifies

$$\left[\sum_{K \in \mathscr{M}} e(v,K)^2\right]^{\frac{1}{2}} \le E(v,S'), \tag{15}$$

interpreting the right-hand side in a broken manner if necessary. Notice that there holds equality for $S' = S^{\ell,-1}(\mathscr{M})$, while for $S' = S(\mathscr{M}) \subset S^{\ell,0}(\mathscr{M})$ the question arises if the requirement of continuity entails some downgrading of the approximation quality: the local approximants P_K, $K \in \mathscr{M}$, on the left-hand side are decoupled, while their counterparts $V_{\mathscr{M}|K}$, $K \in \mathscr{M}$, on the right-hand side are coupled and so constrained.

In order to measure the possible downgrading, consider the inequality opposite to (15) and denote by $\delta(\mathscr{M})$ the smallest constant C such that

$$\forall v \in X \quad E\big(v, S(\mathscr{M})\big) \leq C \left[\sum_{K\in\mathscr{M}} e(v,K)^2\right]^{\frac{1}{2}} \tag{16}$$

is valid; if there is no such constant C, set $\delta(\mathscr{M}) = \infty$. We refer to $\delta(\mathscr{M})$ as the decoupling coefficient of $S(\mathscr{M})$. If the decoupling coefficient is big, or even ∞, requiring continuity entails that the local approximation potential is not well exploited, at least for some target functions. If it is moderate, or even 1, dispensing with continuity does not improve the approximation quality substantially.

It is instructive to consider, for a moment, (16) with $X = L^2(\Omega)$ and to replace the H^1-seminorm by the L^2-norm. Then a function from $S^{\ell,-1}(\mathscr{M}) \setminus S^{\ell,0}(\mathscr{M})$ is an admissible target function and we immediately obtain that there holds $\delta(\mathscr{M}) = \infty$ in this case.

The following lemma introduces the key property of the H^1-seminorm that ensures a finite decoupling coefficient.

Lemma 3 (Trace and error norm) *Let F be a $(d-1)$-face of a d-simplex K. For any $w \in H^1(K)$ with $\int_K w = 0$, there holds*

$$\|w\|_{0,2;F} \leq C_{\mathrm{Tr}} \left(\frac{h_K\,|F|}{|K|}\right)^{\frac{1}{2}} h_K^{\frac{1}{2}} \|\nabla w\|_{0,2;K},$$

where $C_{\mathrm{Tr}} := \sqrt{C_{\mathrm{P}}(C_{\mathrm{P}} + 2/d)}$ and C_{P} denotes the optimal Poincaré constant for all d-simplices. The ratio between the parentheses is bounded in terms of the shape coefficient h_K/ρ_K of K.

The classical result of L. E. Payne and H. F. Weinberger [23], see also M. Bebendorf [3], ensures $C_{\mathrm{P}} \leq 1/\pi$. In the case $d = 2$, R. S. Laugesen and B. A. Siudeja [21] show $C_{\mathrm{P}} = 1/j_{1,1}$ where $j_{1,1} \approx 3.8317$ denotes the first positive root of the Bessel function J_1.

Proof Corollary 4.5 and Remark 4.6 of [27] imply the trace inequality

$$\frac{1}{|F|}\|w\|_{0,2;F}^2 \leq \frac{1}{|K|}\|w\|_{0,2;K}^2 + \frac{2h_K}{d\,|K|}\|w\|_{0,2;K}\|\nabla w\|_{0,2;K}.$$

We thus obtain the claimed inequality for w by inserting the Poincaré inequality

$$\|w\|_{0,2;K} \leq C_{\mathrm{P}} h_K \|\nabla w\|_{0,2;K},$$

which applies thanks to $\int_K w = 0$.

In order to bound the quotient between the parentheses, observe that Cavallieri's principle yields $|K| = (h_F^{\perp}\,|F|)/d$, where $h_F^{\perp}$ denotes the height of K over F. Hence, we obtain

$$\frac{h_K\,|F|}{|K|} = d\frac{h_K}{h_F^{\perp}} \leq d\frac{h_K}{\rho_K}$$

with the help of the inequality $\rho_K \leq h_F^{\perp}$. □

The significance of Lemma 3 lies in the following observations. If the intersection F of two elements $K_1, K_2 \in \mathscr{M}$ is a common $(d-1)$-face, then the condition $V \in C^0(\overline{\Omega})$ requires that the traces $V_{|\partial K_1}$ and $V_{|\partial K_2}$ coincide on F. The local best approximations P_{K_1} and P_{K_2} from (13) are close to this property in that, thanks to Lemma 3, their properly measured difference is bounded in terms of the local best errors:

$$\begin{aligned} h_F^{-\frac{1}{2}} \|P_{K_1} - P_{K_2}\|_{0,2;F} &\leq h_F^{-\frac{1}{2}} \|P_{K_1} - v\|_{0,2;F} + h_F^{-\frac{1}{2}} \|v - P_{K_2}\|_{0,2;F} \\ &\leq C\big[e(v,K_1) + e(v,K_2)\big], \end{aligned}$$

where $h_F := \operatorname{diam} F$ and C depends on d and on the shape coefficients of K_1 and K_2. Consequently, the trace $V_{|F}$ can be defined by P_{K_1}, P_{K_2} or a mixture of both without substantially downgrading the approximation capability of the shape functions of the two elements K_1 and K_2. The same remark applies to near best approximations in place of P_{K_1} and P_{K_2}.

Similarly, Lemma 3 implies that, thanks to $v_{|\partial\Omega} = 0$, properly measured traces on $\partial\Omega$ of local best approximations are bounded again in terms of local best errors or, in other words, almost vanish. Consequently, enforcing vanishing boundary values will not downgrade the approximation capability.

3.2 Interpolation

In order to show that the decoupling coefficient $\delta(\mathscr{M})$ is finite, we define

$$\Pi : H^1(\Omega) \to S^{\ell,0}(\mathscr{M})$$

with the goal that it satisfies

$$v \in X \implies \Pi v \in S(\mathscr{M}) \tag{17}$$

and

$$|v - \Pi v|_\Omega \leq C \left[\sum_{K \in \mathscr{M}} e(v,K)^2\right]^{\frac{1}{2}} \tag{18}$$

for some constant C, independent of v. Notice that the latter property requires that Π is a projection whenever $S(\mathscr{M}) \subset X$ and stable with respect to (11).

Using the Lagrange basis of $S^{\ell,0}(\mathscr{M})$ from Lemma 1, we can write

$$\Pi v = \sum_{z \in L_\ell(\mathscr{M})} \Pi_z v \, \Phi_z \tag{19a}$$

and defining Π amounts to choosing suitable linear functionals $\Pi_z \in H^{-1}(\Omega)$, $z \in L_\ell(\mathscr{M})$, for the nodal values. For this purpose, it is convenient to introduce the following notion. A node $z \in L_\ell(\mathscr{M})$ is called unconstrained in $S(\mathscr{M})$ if and only if $\operatorname{supp} \Phi_z$ is contained in one element of $\mathscr{M}$ and there holds $\operatorname{span} \Phi_z \subset S(\mathscr{M})$. Extending functions on elements by zero, this is equivalent to requiring that there holds $\operatorname{span} \Phi_{z|K} \subset S(\mathscr{M})$ for any element $K \in \mathscr{M}$. If $\operatorname{span} \Phi_{z|K} \not\subset S(\mathscr{M})$, then z is called constrained and we write $z \in \mathscr{C}$.

Fix an arbitrary node $z \in L_\ell(\mathscr{M})$. If $z \notin \mathscr{C}$, then $\Pi_z v$ affects only the local error $|v - \Pi v|_K$ on that element. In this case, set

$$\Pi_z v := P_K(z), \tag{19b}$$

where P_K is the local best approximation given by (13).

Whereas, if $z \in \mathscr{C}$, then $\Pi_z v$ has to deviate from (19b) for at least one other element or has to assume a prescribed value. In order to meet both issues, we employ the functionals from Lemma 2. To this end, fix some face $F_z \in \mathscr{F}_{d-1}(\mathscr{M})$ containing z; this choice may be subject to further conditions for certain examples of $S(\mathscr{M})$. We then set

$$\Pi_z v := N_{z,F_z}(v). \tag{19c}$$

We illustrate this definition in the setting of the introduction §1 where $X = H^1_0(\Omega)$ and $S(\mathscr{M}) = S^{\ell,0}_0(\mathscr{M})$. Here there holds

$$\mathscr{C} = L_\ell(\mathscr{M}) \cap \Sigma \quad \text{with} \quad \Sigma := \bigcup_{F \in \mathscr{F}_{d-1}(\mathscr{M})} F$$

and, if $z \in \mathscr{C}$, we additionally require that $F_z \subset \partial\Omega$ whenever $z \in \partial\Omega$ lies on the boundary. This readily ensures (17) thanks to Lemma 2 (ii). In Remark 1 below we further discuss the construction of Π, comparing it with existing interpolation operators and indicating alternatives. In particular, we shall see that, irrespective of the choice of F_z, the definition (19c) is 'near to the best (19b)' in a suitable sense.

Notice that, on the one hand, the face in (19c) is linked to an arbitrary element of $\operatorname{supp}\Phi_z$ only through the node z and, on the other hand, traces of H^1-functions are well-defined only on at least $(d-1)$-dimensional faces. The following property of the mesh therefore appears to be essential for local near best approximation properties of Π. A star

$$\omega_z = \bigcup \{K \in \mathscr{M} : K \ni z\}, \qquad z \in L_\ell(\mathscr{M}),$$

is called $(d-1)$-face-connected if for any element K and $(d-1)$-face F containing z there exists a sequence $(K_i)_{i=1}^n$ such that

- any K_i, $i = 0, \dots, n$, is an element of the star,
- any intersection $K_i \cap K_{i+1}$, $i = 0, \dots, n-1$, is a $(d-1)$-face of the star,
- K_0 contains F_z and $K_n = K$.

A star ω_z is $(d-1)$-face-connected if the open set $\omega_z \cap \Omega$ is connected. Consequently, stars of interior nodes $z \in \Omega$ are $(d-1)$-face-connected, as well as stars of boundary nodes $z \in \partial\Omega$ where the boundary is a Lipschitz graph in a sufficiently large neighborhood of z. If $\Omega \cap \omega_z$ is disconnected, the star may or may not be $(d-1)$-face-connected. Figure 1 illustrates this with two planar stars for which $\Omega \cap \omega_z$ consists of two connected components. The left one, which may arise for the slit domain, is edge-connected, while the right one is not. The edge-connectedness of the left one is a consequence of the following observation: a star for which $\Omega \cap \omega_z$ consists of two connected components is $(d-1)$-face-connected if the intersection of their closures in $\mathbb{R}^d$ contains a $(d-1)$-face which in turn contains z. It is worth noting that, if a star is not $(d-1)$-face-connected and $\Omega \cap \omega_z$ consists of two connected components, the

Fig. 1 Edge-connectedness: Planar stars (gray areas) at the boundary (thick lines). The left one is edge-connected, the right one is not.

H^1-norm is not strong enough to couple the components, suggesting that elements belonging to different components should not be coupled in a H^1-conforming finite element space.

3.3 Local decoupling

The definition of Π and the identity

$$|w|_\Omega^2 = \sum_{K\in\mathscr{M}} |w|_K^2,$$

suggest to show (18) by establishing local counterparts, involving the patches

$$\omega_K := \bigcup\{K' \in \mathscr{M} : K' \cap K \neq \emptyset\}$$

in $\mathscr{M}$.

Theorem 1 (Local decoupling) *Given an element $K \in \mathscr{M}$, assume that its stars ω_z, $z \in L_\ell(K)$, are $(d-1)$-face-connected. Then there exists a constant δ_K such that, for all $v \in H^1(\omega_K)$,*

$$|v-\Pi v|_K^2 \le e(v,K)^2 + \delta_K^2 \sum_{z\in L_\ell(K)} \sum_{K'\ni z} e(v,K')^2, \tag{20}$$

where the second sum is over all $K' \in \mathscr{M}$ containing z. The constant δ_K can be bounded as follows:

$$\delta_K^2 \le 4dC_{\mathrm{Tr}}C(\sigma_K) \sum_{z\in L_\ell(K)\cap\mathscr{C}} \left\|\hat\Psi_{\hat z}\right\|_{0,K_{d-1};2}^2 \left\|\nabla\hat\Phi_{\hat z}\right\|_{0,K_d;2}^2, \tag{21}$$

where $\sigma_K := \max_{K'\subset\omega_K} h_{K'}/\rho_{K'}$ stands for the shape coefficient of the patch ω_K.

Proof We start by recalling the definition (13) of the best approximation P_K and exploit the orthogonality (14) to write

$$|v-\Pi v|_K^2 = e(v,K)^2 + |P_K - \Pi v|_K^2,$$

where the second square on the right-hand side measures the deviation of Πv of being locally optimal. Thanks to (19b), there holds

$$(P_K-\Pi v)_{|K} = \sum_{z\in L_\ell(K)} \left[P_K(z) - \Pi_z v\right]\Phi_{z|K} = \sum_{z\in L_\ell(K)\cap\mathscr{C}} \left[P_K(z) - N_{z,F_z}(v)\right]\Phi_{z|K},$$

which reveals that this deviation is entirely related to the requirements for conformity. To bound it, we first apply the triangle inequality to obtain

$$|P_K - \Pi v|_K \leq \sum_{z \in L_\ell(K) \cap \mathscr{C}} \left| P_K(z) - N_{z,F_z}(v) \right| \|\nabla \Phi_z\|_{0,2;K}. \tag{22}$$

We proceed by bounding each term of the sum separately. Fix any $z \in L_\ell(K) \cap \mathscr{C}$ and, using that ω_z is $(d-1)$-face-connected, choose a corresponding sequence $(K_i)_{i=0}^n$ of elements connecting F_z with K. Set $F_0 := F_z$ and $F_i := K_{i-1} \cap K_i$ for $i = 1, \ldots, n$. For the sake of readability, we sometimes replace K_i and F_i by i. Recalling Lemma 2 (i), we note

$$P_K(z) - N_{z,F_z}(v) = N_{z,n}(P_n) - N_{z,0}(v)$$

and

$$\forall i = 0, \ldots, n-1 \quad N_{z,i}(P_i) = P_i(z) = N_{z,i+1}(P_i),$$

which, by telescopic expansion, leads to

$$P_K(z) - N_{z,F_z}(v) = N_{z,0}(P_0 - v) + \sum_{i=1}^{n} N_{z,i}(P_i - P_{i-1}). \tag{23}$$

Since $F_0 \subset K_0$ and $\int_{K_0}(v - P_0) = 0$, one can combine Lemma 2 (iii) and Lemma 3 to derive

$$|N_{z,0}(v - P_0)| \leq \frac{C_{\mathrm{Tr}}}{\sqrt{(d-1)!}} \left\|\hat{\Psi}_{\hat{z}}\right\|_{0,2;K_{d-1}} \frac{h_0}{|K_0|^{\frac{1}{2}}} e(v, K_0),$$

where $|F_0|$ cancels out. Similarly, for $i = 1, \ldots, n$, one obtains

$$\begin{aligned} N_{z,i}(P_i - P_{i-1}) &\leq |N_{z,i}(P_i - v)| + |N_{z,i}(v - P_{i-1})| \\ &\leq \frac{C_{\mathrm{Tr}}}{\sqrt{(d-1)!}} \left\|\hat{\Psi}_{\hat{z}}\right\|_{0,2;K_{d-1}} \left[\frac{h_i}{|K_i|^{\frac{1}{2}}} e(v, K_i) + \frac{h_{i-1}}{|K_{i-1}|^{\frac{1}{2}}} e(v, K_{i-1}) \right]. \end{aligned}$$

Using these bounds in the telescopic expansion (23) gives

$$|P_K(z) - N_{z,F_z}(v)| \leq \frac{2C_{\mathrm{Tr}}}{\sqrt{(d-1)!}} \left\|\hat{\Psi}_{\hat{z}}\right\|_{0,2;K_{d-1}} \sum_{i=0}^{n} \frac{h_i}{|K_i|^{\frac{1}{2}}} e(v, K_i).$$

In order to get independent of the specific choice of F_z, we observe that the sequence $(K_i)_{i=0}^n$ does not allow a double occurrence of some element and that each element is a subset of $\omega_z = \operatorname{supp} \Phi_z$. We therefore replace the preceding inequality by

$$|P_K(z) - N_{z,F_z}(v)| \leq \frac{2C_{\mathrm{Tr}}}{\sqrt{(d-1)!}} \left\|\hat{\Psi}_{\hat{z}}\right\|_{0,2;K_{d-1}} \sum_{K' \subset \omega_z} \frac{h_{K'}}{|K'|^{\frac{1}{2}}} e(v, K'), \tag{24}$$

where the sum is over all elements $K' \in \mathscr{M}$ with $K' \subset \omega_z$. Inserting this inequality and Lemma 1 (iii) into (22), one arrives at

$$|P_K - \Pi v|_K \leq 2\sqrt{d} C_{\mathrm{Tr}} \sum_{z \in L_\ell(K) \cap \mathscr{C}} \sum_{K' \subset \omega_z} b_{\hat{z}} \frac{h_{K'} \|\nabla \Phi_z\|_{0,2;K} |K|^{\frac{1}{2}}}{\|\Phi_z\|_{0,2;K} |K'|^{\frac{1}{2}}} e(v, K')$$

with

$$b_{\hat{z}} := \left\|\hat{\Psi}_{\hat{z}}\right\|_{0,2;K_{d-1}} \left\|\hat{\Phi}_{\hat{z}}\right\|_{0,2;K_d}.$$

Applying a Cauchy-Schwarz inequality, one obtains a sum of the squares of the local best errors and the claimed inequality with

$$\delta_K^2 = 4dC_{\mathrm{Tr}}^2 \sum_{z\in L_\ell(K)\cap\mathscr{C}} b_{\hat{z}}^2 \mu_z \frac{h_K^2 \|\nabla\Phi_z\|_{0,K;2}^2}{\|\Phi_z\|_{0,K;2}^2} \tag{25}$$

and

$$\mu_z := \frac{|K|}{h_K^2} \sum_{K'\subset\omega_z} \frac{h_{K'}^2}{|K'|}.$$

To conclude the proof, we still have to verify (21). Note that only the quantities μ_z and $h_K \|\nabla\Phi_z\|_{0,K;2} / \|\Phi_z\|_{0,K;2}$ in (25) depend on geometrical properties of the patch ω_K. A standard scaling argument, see e.g. [13, (4.5.3)] shows that

$$\frac{h_K \|\nabla\Phi_z\|_{0,K;2}}{\|\Phi_z\|_{0,K;2}} \le \sigma_K \frac{\left\|\nabla\hat{\Phi}_{\hat{z}}\right\|_{0,K_d;2}}{\left\|\hat{\Phi}_{\hat{z}}\right\|_{0,K_d;2}}.$$

Moreover, since the (solid) angles of the elements in ω_K are bounded away from 0 in terms of σ_K, the number of elements in each star of ω_K is bounded in terms of σ_K. Comparing elements having common faces, we thus obtain that the diameters and the volumes of the elements in a star of ω_K are comparable up to σ_K. Consequently, there holds $\mu_z \le C(\sigma_K)$ for each $z \in L_\ell(K)$ and the proof is finished. □

3.4 Global decoupling and boundary conditions

Summing up the inequalities of Theorem 1, we obtain our main result.

Theorem 2 (Decoupling of elements) *Assume that all stars of the mesh $\mathscr{M}$ are $(d-1)$-face-connected. Then there exists a constant C such that*

$$|v - \Pi v|_\Omega \le C \left[\sum_{K\in\mathscr{M}} e(v,K)^2\right]^{\frac{1}{2}}$$

for all $v \in H^1(\Omega)$. The constant C can be bounded in terms of the dimension d, the polynomial degree ℓ and the shape coefficient $\sigma_{\mathscr{M}} := \max_{K\in\mathscr{M}} h_K/\rho_K$ of $\mathscr{M}$.

Proof We sum (20) over all $K \in \mathscr{M}$. On the right-hand side, we hit a given element $K' \in \mathscr{M}$ at most $1 + n_\ell N_{\mathscr{M}}$ times, where $n_\ell = \#\{z \in L_\ell(K) \mid z \in \partial K\}$ indicates the number of boundary Lagrange nodes and

$$N_{\mathscr{M}} := \max_{z\in L_\ell(\mathscr{M})} \#\{K' \in \mathscr{M} \mid K' \ni z\}$$

stands for the maximum number of elements in a star. Consequently, the claimed bounds holds with

$$C = (1 + n_\ell N_{\mathscr{M}}) \max_{K \in \mathscr{M}} \delta_K$$

with δ_K from Theorem 1. Note that $N_{\mathscr{M}}$ can be bounded in terms of the shape coefficient of $\mathscr{M}$, see the end of the proof of Theorem 1. Taking also (21) into account, the claim thus follows. □

Theorem 2 covers various boundary conditions associated with an 'H^1-setting'. We illustrate this by discussing Dirichlet and Neumann boundary conditions for Poisson's equation.

For Dirichlet boundary conditions, we follow the approach of L. R. Scott and S. Zhang in [25, §5]. Denote by $\Pi^{\mathrm{SZ}} : H^1(\Omega) \to S^{\ell,0}(\mathscr{M})$ the interpolation operator therein and recall that the restriction $\Pi^{\mathrm{SZ}} v_{|\partial\Omega}$ depends only on $v_{|\partial\Omega}$. Given boundary values $g \in H^{\frac{1}{2}}(\partial\Omega)$, the weak solution of a Dirichlet problem is from the trial space

$$X_g := \{v \in H^1(\Omega) \mid v_{|\partial\Omega} = g\}$$

and the finite element solution is sought in the space

$$S_g(\mathscr{M}) := \{V \in S^{\ell,-1}(\mathscr{M}) \mid V \in C^0(\overline{\Omega}), V_{|\partial\Omega} = \Pi^{\mathrm{SZ}} g\}$$

which is not necessarily a subspace of X_g. In view of [24, Lemma 2.1], (11) is a definite error notion on $X_g - S_g(\mathscr{M})$. Since however $\Pi^{\mathrm{SZ}} v_{|\partial\Omega} = v_{|\partial\Omega}$ for all $v \in S^{\ell,0}(\mathscr{M})$, there holds $S_g(\mathscr{M}) \subset X_g$, i.e. the ensuing finite element method is conforming, whenever possible. It is not difficult to show that the finite element solution is a near best approximation from $S_g(\mathscr{M})$.

Corollary 1 (Dirichlet boundary values) *For any $v \in X_g$, there holds*

$$E\big(v, S_g(\mathscr{M})\big) \leq C \left[\sum_{K \in \mathscr{M}} e(v,K)^2\right]^{\frac{1}{2}},$$

where C is the constant from Theorem 1. Consequently, the decoupling coefficient of $S_g(\mathscr{M})$ is bounded in terms of d, ℓ and $\sigma_{\mathscr{M}}$.

Proof As for the case corresponding to the introduction, there holds

$$\mathscr{C} = L_\ell(\mathscr{M}) \cap \Sigma \quad \text{with} \quad \Sigma = \bigcup_{F \in \mathscr{F}_{d-1}(\mathscr{M})} F$$

and we require that $F_z \subset \partial\Omega$ whenever $z \in \mathscr{C} \cap \partial\Omega$. Moreover, we require that the choices of F_z for all $z \in L_\ell(\mathscr{M}) \cap \partial\Omega$ in definitions of Π^{SZ} and Π coincide. Consequently,

$$\Pi v_{|\partial\Omega} = \Pi^{\mathrm{SZ}} v_{|\partial\Omega}$$

and (17) holds with $X = X_g$ and $S(\mathscr{M}) = S_g(\mathscr{M})$ and Theorem 2 yields the claim. □

For homogeneous Dirichlet boundary values $g = 0$, Corollary 1 implies the non-obvious part of (2). A further immediate consequence is that 'near best in $S_g(\mathscr{M})$' entails 'near best in $S^{\ell,0}(\mathscr{M})$'. In particular, the aforementioned finite element solution is thus near best in $S^{\ell,0}(\mathscr{M})$.

For Neumann boundary conditions, the trial space is

$$\tilde{X} := H^1(\Omega)/\mathbb{R},$$

which can be approximated by

$$\tilde{S}(\mathscr{M}) := S^{\ell,0}(\mathscr{M})/\mathbb{R}.$$

Again, (11) is a definite error notion and the finite element solution of this space is near best.

Corollary 2 (Neumann boundary values) *For any $v \in \tilde{X}$, there holds*

$$E\big(v, \tilde{S}(\mathscr{M})\big) \leq C \left[\sum_{K \in \mathscr{M}} e(v,K)^2 \right]^{\frac{1}{2}},$$

where C is the constant from Theorem 1. Consequently, the decoupling coefficient of $\tilde{S}(\mathscr{M})$ is bounded in terms of d, ℓ and $\sigma_{\mathscr{M}}$.

Proof Identifying $H^1(\Omega)/\mathbb{R}$ and $\{v \in H^1(\Omega) \mid \int_\Omega v = 0\}$, the implication (17) holds with $X = \tilde{X}$ and Theorem (2) again yields the claim. □

Similarly, one can consider Robin boundary conditions or mixed ones and obtain corresponding statements.

Remark 1 (Construction of interpolation operator) The definition of the nodal values $\Pi_z v$ for constrained nodes $z \in L_\ell(\mathscr{M}) \cap \mathscr{C}$ is the critical part. For constrained nodes on the boundary, we follow the approach of [25]. The role of (24) in proving Theorem 2 reveals that this is a near best choice with respect to the involved local errors and can be adopted also for the other constrained nodes. Inequality (24) shows also that the particular admissible choice of F_z does not matter. Moreover, its proof reveals that, for interior constrained nodes, also $P_K(z)$ where the element K contains z, or some average of these values may be used. Interpolation operators of this type may be viewed as a composition of taking the best approximation in $S^{\ell,-1}(\mathscr{M})$ and a so-called enriching operator. The latter have been used to connect non-conforming finite element methods to conforming ones in various contexts; see, e.g., S. Brenner [11, 12], O. S. Karakashian and F. Pascal [20], T. Gudi [19] and A. Bonito and R. H. Nochetto [8].

3.5 Local gradient conformity

The local best errors $e(v,K)$, $K \in \mathscr{M}$, are related to approximation problems of the following type. Approximate a vector function, which is the gradient of a scalar function, with gradients of polynomial functions. The components of the approximants are coupled whenever $\ell \geq 2$: indeed, the value of higher order partial derivatives then does not depend on their order of application. Hence, the following question arises: Does this coupling lead to some downgrading with respect to approximants, the components of which are independent polynomial functions?

A similar question arises for $E(v,\mathscr{M})$. However, in view of Theorem 2, it suffices for both questions to compare the local best errors $e(v,K)$, $K \in \mathscr{M}$, with the following ones:

$$\underline{e}(v,K) := \inf_{Q \in \mathscr{P}_{\ell-1}(K)^d} \|\nabla v - Q\|_K = \left(\sum_{i=1}^{d} \inf_{R \in \mathscr{P}_{\ell-1}(K)} \|\partial_i v - R\|_K^2 \right)^{\frac{1}{2}},$$

with

$$\|f\|_\Omega := \||f|\|_{0,2;\Omega} = \left(\int_\Omega |f|^2 \right)^{\frac{1}{2}}$$

for $f \in L^2(\Omega)^d$. Also here the inequality

$$\underline{e}(v,K) \leq e(v,K)$$

is straight-forward, while the opposite one is more involved and its proof relies on the construction of a suitable (quasi-)interpolant. We shall use the averaged Taylor polynomial of [18] by T. Dupont and L. R. Scott, which is a variant of the one of S. L. Sobolev. In order to avoid a dependence on the element shape, we follow an idea of S. Dekel and D. Leviatan in [16] and average in a reference configuration.

Theorem 3 (Decoupling of partial derivatives) *There is a constant C depending only on d, $\ell \in \mathbb{N}$ such that, for any element $K \in \mathscr{M}$ and any function $v \in H^1(K)$, there holds*

$$e(v,K) \leq C\underline{e}(v,K).$$

Proof Given $\ell \in \mathbb{N}_0$, denote by I_ℓ the operator corresponding to the averaged Taylor polynomial of order $\ell+1$ (i.e. degree $\leq \ell$) over the largest inscribed ball in K_d. Remarkably, the averaged Taylor polynomial commutes with differentiation in that $\partial_i(I_\ell w) = I_{\ell-1}(\partial_i w)$ for all $w \in H^1(K_d)$ and $i \in \{1,\ldots,d\}$; see, e.g., [13, (4.1.17)]. Corollary 3.4 in [16] generalizes this to

$$\partial_i\Big(\big[I_\ell(v \circ A)\big] \circ A^{-1} \Big) = \Big(I_{\ell-1}\big[(\partial_i v) \circ A\big] \Big) \circ A^{-1}, \tag{26}$$

where $A : K_d \to K$ is an affine bijection, $v \in H^1(K)$ and $\ell \geq 1$. Moreover, I_ℓ is a L^2-stable projection on $\mathscr{P}_\ell(K_d)$: for any $w \in L^2(K_d)$, there hold

$$\begin{aligned} w \in \mathscr{P}_\ell(K_d) &\implies I_\ell w = w, \\ \|I_\ell w\|_{K_d} &\leq C_{d,\ell} \|w\|_{K_d}; \end{aligned}$$

see, e.g., [13, (4.1.15), (4.2.8)]. Lemma 5 of J. Xu and L. Zikatanov [29] therefore implies that I_ℓ is near best with the constant $C_{d,\ell}$:

$$\|w - I_\ell w\|_{K_d} \le C_{d,\ell} \inf_{S \in \mathscr{P}_\ell(K_d)} \|w - S\|_{K_d} . \tag{27}$$

Motivated by (26), we choose $P = \big[I_\ell(v \circ A)\big] \circ A^{-1} \in \mathscr{P}_\ell(K)$ and, also using the transformation rule and (27) with $\ell - 1$ in place of ℓ, we obtain

$$\begin{aligned}
e(v,K)^2 &\le \|\nabla(v-P)\|_K^2 = \sum_{i=1}^d \|\partial_i v - \partial_i P\|_K^2 \\
&= \sum_{i=1}^d \Big\| \partial_i v - \Big(I_{\ell-1}\big[(\partial_i v)\circ A\big] \Big) \circ A^{-1} \Big\|_K^2 \\
&= \frac{|K|}{|K_d|} \sum_{i=1}^d \big\| (\partial_i v)\circ A - I_{\ell-1}\big[(\partial_i v)\circ A\big] \big\|_{K_d}^2 \\
&\le C_{d,\ell-1}^2 \frac{|K|}{|K_d|} \sum_{i=1}^d \inf_{S \in \mathscr{P}_{\ell-1}(K_d)} \|\partial_i v \circ A - S\|_{K_d}^2 \\
&\le C_{d,\ell-1}^2 \sum_{i=1}^d \inf_{R \in \mathscr{P}_{\ell-1}(K)} \|\partial_i v - R\|_K^2 = C_{d,\ell-1}^2 \underline{e}(v,K)^2
\end{aligned}$$

Consequently, the claimed inequality holds with $C = C_{d,\ell-1}$. □

The combination of Theorems 2 and 3 yields the following statement.

Corollary 3 (Decoupling of elements and partial derivatives) *There is a constant C such that for any $v \in X$ there holds*

$$E\big(v, S(\mathscr{M})\big) \le C \left[\sum_{K \in \mathscr{M}} \underline{e}(v,K)^2 \right]^{\frac{1}{2}},$$

with $C \le C_{d,\ell-1}\delta(\mathscr{M})$, which can be bounded in terms of d, ℓ and $\sigma_{\mathscr{M}}$.

It is worth noticing that the right-hand side in Corollary 3 involves best errors of approximation problems that may be considered the simplest ones involving polynomials and measuring the error in some L^2-sense.

4 Applications to error bounds and tree approximation

The main novelty of the preceding section lies in the type of statements that are proven. The goal of this section is to advocate its usefulness by showing that it allows for simplifications and improvements in theory and algorithms. Doing so, we focus on applications of the decoupling of elements and we adopt the setting of §3.1, which allows for essential boundary conditions via the affine spaces $X \subset H^1(\Omega)$ and $S(\mathscr{M}) \subset S^{\ell,0}(\mathscr{M})$.

4.1 Convergence and error bounds

We start by reviewing some approximation results that play an important role in the a priori error analysis of finite element methods.

A minimum requirement for a numerical method for a boundary value problem is that the approximate solution converges to the exact one as the meshsize tends to 0. In case of a finite element method, a necessary (and also sufficient, if for example the Céa Lemma holds) condition for this is that the best error of the corresponding finite element space tends to 0. It is instructive to prove this well-known fact for continuous piecewise polynomial functions with the help of Theorem 2.

Theorem 4 (Convergence) *For any $v \in X \subset H^1(\Omega)$, the global best error* (12) *satisfies*

$$E(v,\mathscr{M}) \to 0 \quad as \quad h := \max_{K\in\mathscr{M}} h_K \to 0$$

within a shape-regular family of meshes with $(d-1)$-face connected stars.

Proof Given a vector function $f \in L^2(\Omega)^d$, let $\overline{f} \in L^\infty(\Omega)^d$ be the piecewise constant function given by

$$\forall K \in \mathscr{M} \quad \overline{f}_{|K} = \frac{1}{|K|}\int_K f.$$

Theorem 2 and $\ell \geq 1$ imply

$$\begin{aligned} E(v,\mathscr{M}) &\leq C\left(\sum_{K\in\mathscr{M}} e(v,K)^2\right)^{\frac{1}{2}} \leq C\left(\sum_{K\in\mathscr{M}} \inf_{P\in\mathscr{P}_1(K)} |v-P|_K^2\right)^{\frac{1}{2}} \\ &= C\left(\sum_{K\in\mathscr{M}} \inf_{c\in\mathbb{R}} \|\nabla v - c\|_K^2\right)^{\frac{1}{2}} = C\left\|\nabla v - \overline{\nabla v}\right\|_\Omega, \end{aligned} \tag{28}$$

which allows concluding with a standard argument: Let $\varepsilon > 0$ be arbitrary. Since $C^0(\overline{\Omega})$ is dense in $L^2(\Omega)$, there exists $g \in C^0(\overline{\Omega})^d$ such that $\|\nabla v - g\|_\Omega \leq \varepsilon/3$. Thanks to $\|\overline{f}\|_\Omega \leq \|f\|_\Omega$, we derive

$$\left\|\nabla v - \overline{\nabla v}\right\|_\Omega \leq \|\nabla v - g\|_\Omega + \|g-\overline{g}\|_\Omega + \left\|\overline{g-\nabla v}\right\|_\Omega \leq (2\varepsilon)/3 + \|g-\overline{g}\|_\Omega$$

and the last term can be made smaller $\varepsilon/3$ for sufficiently small h, because g is uniformly continuous in view of the compactness of $\overline{\Omega}$. □

Notice that the first equality in (28) corresponds to a special case of Theorem 3 and its combination with Theorem 2 simplifies the following density argument in that it does not involve derivatives.

Usually, the quality of a finite element method is theoretically investigated by deriving a priori error estimates, quantifying the convergence speed in terms of powers of h. Accordingly, such estimates for the best error of the corresponding finite element space are of interest. These are usually obtained by directly bounding the error

of some interpolation operator in terms of higher order Sobolev seminorms. Here we use Theorem 2 and

$$|v|_{s,2;K} := \begin{cases} \left(\sum_{|\alpha|=s} \|\partial^\alpha v\|_{0,2;K}^2 \right)^{\frac{1}{2}} & \text{if } s \in \mathbb{N} \\ (1-\theta) \left(\sum_{|\alpha|=\lfloor s \rfloor} \int_K \int_K \frac{|\partial^\alpha v(x) - \partial^\alpha v(y)|^2}{|x-y|^{2\theta+d}} \right)^{\frac{1}{2}} & \text{otherwise,} \end{cases}$$

where $s > 0$ indicates the smoothness, $\theta := s - \lfloor s \rfloor$ its fractional part and the sums are over all multi-indexes of length $\lfloor s \rfloor$; the factor $(1-\theta)$ is motivated by J. Bourgain et al. [9].

Theorem 5 (Error bounds) *Let $v \in X \subset H^1(\Omega)$ a target function, $\mathscr{M}$ be a mesh with $(d-1)$-face connected stars and $1 \le s \le \ell+1$. If $v_{|K} \in H^s(K)$ for all $K \in \mathscr{M}$, then the global best error* (12) *is bounded by*

$$E(v, \mathscr{M}) \le C \left(\sum_{K \in \mathscr{M}} h_K^{2(s-1)} |v|_{s,2;K}^2 \right)^{\frac{1}{2}},$$

where C depends only on d, ℓ, $\sigma_{\mathscr{M}}$ and s. In particular, if $s \in \mathbb{N}$, then

$$C \le \frac{s!}{(\lceil \frac{s}{d} \rceil !)^d} C_{\mathrm{P}}^{s-1} \delta(\mathscr{M})$$

where C_{P} is the optimal Poincaré constant for d-simplices in $\mathbb{R}^d$.

Proof Combine Theorem 2 and the Bramble-Hilbert inequality

$$e(v,K) \le C h_K^{s-1} |v|_{s,2;K} \tag{29}$$

where C depends on d, ℓ, s and the shape coefficient of the d-simplex K. The latter follows, e.g., from [18, Theorems 3.2 and 6.1] and a standard scaling argument. The explicit constant for $s \in \mathbb{N}$ is ensured by choosing a polynomial that allows an iterative application of the Poincaré inequality; see R. Verfürth [28, §3]. □

Theorem 5 provides error bounds in terms of piecewise regularity over the entire admissible smoothness range $[1, \ell+1]$. As is illustrated by the following two remarks, this combines advantages of the error bounds that are available via Lagrange, Clément [15] and Scott-Zhang [25, §4] interpolation. Since the bounds via Clément interpolation [15] are covered by those via Scott-Zhang interpolation, we omit the former in the following discussion. In this discussion, we fix a mesh and vary through functions – a viewpoint differing from the usual one where a function is fixed and meshes vary.

Remark 2 (Entire smoothness range) For a given target function and a given mesh, the most convenient choice of s is not necessarily the maximal one; indeed, the product $h_K^s |v|_{s,2;K}$ may not be monotone decreasing in s for certain functions. This observation is closely related to the following drawback of the bounds via Lagrange interpolation for $d \ge 2$. Consider a sequence of functions $(v_n)_n$ in $H^{\lfloor d/2 \rfloor + 1}(\Omega)$ converging to a function not better than $H^{d/2}(\Omega)$. Then any bound via Lagrange interpolation blows up, while those of Theorem 5 and via Scott-Zhang interpolation remain bounded for suitable $s \in [1, d/2]$.

Remark 3 (Piecewise regularity) The bounds in Theorem 5 are in terms of broken Sobolev seminorms on elements. This can be readily achieved also via Lagrange interpolation; see, e.g., [13, (4.4.20)] and modify the last three steps in its proof. The available bounds via Scott-Zhang interpolation [25] however involve regularity across element boundaries; bounds in terms of broken Sobolev seminorms could be derived with the help of a 'broken' Bramble-Hilbert lemma for element patches, as in F. Camacho and A. Demlow [14].

The broken Sobolev norms on elements have the following advantage: If the global error $E(v,\mathscr{M})$ vanishes, then the bounds of Theorem 5 and Lagrange interpolation also vanish, while those via Scott-Zhang interpolation [25] do not vanish for $s\in]1,3/2[$ and are not applicable, i.e. are ∞, for $s\geq 3/2$ whenever v has non-constant gradient. Similarly as before, this has its counterparts for smooth functions. To illustrate this, let $(v_n)_n$ be a sequence of functions in $H^{\ell+1}(\Omega)$ and consider various conditions ensuring convergence to a function in $S(\mathscr{M})$ with non-constant gradient. If $E(v_n,\mathscr{M})\to 0$, then also the right-hand side of Theorem 2 tends to 0. If additionally the restrictions $v_{n|K}$ converge in $H^s(K)$ for some $1<s\leq \ell+1$, then the corresponding bound of Theorem 5 tends to 0. The same holds for bounds of Lagrange interpolation, but only for $s\in]d/2,\ell+1]$. The situation for bounds via Scott-Zhang interpolation is different: if $1<s<3/2$, then the corresponding bound does not tend to 0 and if $3/2\leq s\leq \ell+1$, it even blows up.

Bounds in terms of broken regularity are useful also in the context of surface finite element methods (SFEM); see [14].

The combination of Remark 2 and 3 entails that the error bounds in Theorem 5 for $1<s\leq d/2$ with $d\geq 3$ are not covered via Lagrange interpolation and those in [15,25].

4.2 Adaptive tree approximation of gradients

The upper bound of the global best error in Theorem 5 locally combines meshsize and higher order derivatives. This suggests that, for a given target function, certain meshes a more convenient than others. I. Babuška and W. C. Rheinboldt [2] formally derive the following criteria, also called equidistribution principle: a mesh minimizing the aforementioned upper bound subject to a fixed number of elements equidistributes the element contributions, e.g., $h_K\,|v|_{2,2;K}$ does not depend on $K\in K$. Obviously this requires in general graded meshes.

An algorithmically simple way of constructing graded meshes arises from a prescribed rule for subdividing elements, which induces a tree structure. An example of this was already in 1967 studied by S. Birman and M. Solomyak [7]. For continuous piecewise polynomial functions over conforming simplicial meshes, one can use bisection with recursive completion; for an overview of this technique, see e.g. Nochetto et al. [22, §4]. Although recursive bisection limits mesh flexibility, in particular mesh grading, the discussion in R. DeVore [17, §6] and the results of P. Binev et al. [5] reveal that the regularity dictating the asymptotic balance of global best error and number of degree of freedoms is close to the best possible one. In particular, the

global best H^1-error of suitably graded two-dimensional meshes decays like $\#\mathscr{M}^{-\frac{1}{2}}$ if $u \in W^{2,p}(\Omega)$ with $1 < p \leq 2$; notice that the latter is weaker than the requirement $u \in H^2(\Omega)$ corresponding to the decay rate with quasi-uniform meshes and that $p = 1$ corresponds to an optimal Sobolev embedding.

The goal of this section is to derive and analyze an instance of the tree algorithm by P. Binev and R. DeVore [6] that constructs near best bisection meshes for the approximation of gradients with piecewise polynomial functions. It may be used for coarsening in adaptive algorithms iterating the main steps

$$\texttt{error reduction} \to \texttt{sparsity adjustment}.$$

Interestingly, this scheme can be applied also if a good a posteriori error estimator is not available. It includes algorithms like in [4, §8] and algorithms that are based upon discretizing the steps of an infinite-dimensional solver. Moreover, the following algorithm can be used to compute an approximation of the best error of bisection meshes with a given number of elements. Such approximations are of interest as a benchmark for corresponding adaptive finite element methods.

In order to introduce the algorithm and to state its main property, we need the following notation. Let $\mathscr{M}_0$ be an initial mesh of Ω that is admissible for bisection with recursive completion; see, e.g., [22, Assumption 11.1 on p. 453]. Denote by $\mathbb{M}'$ the set of all meshes that can be generated by bisections without completion from $\mathscr{M}_0$; these meshes are not necessarily conforming and each one corresponds to a subtree in the master tree given by $\mathscr{M}_0$ and the bisection rule for single d-simplices. Moreover, denote by $\mathbb{M}$ the subset of $\mathbb{M}'$ of all meshes that are conforming. If $\mathscr{M}' \in \mathbb{M}'$ is a possibly non-conforming mesh, denote by $\texttt{complete}(\mathscr{M}')$ the smallest refinement of $\mathscr{M}'$ in $\mathbb{M}$. Since $\mathscr{M}_0$ is admissible, Binev et al. [4, Theorem 2.4] if $d = 2$ and R. Stevenson [26, Theorem 6.1] otherwise ensure the non-obvious relationship

$$\#\texttt{complete}(\mathscr{M}') - \#\mathscr{M}_0 \leq C_{\mathrm{cmpl}}(\#\mathscr{M}' - \#\mathscr{M}_0) \tag{30}$$

with C_{cmpl} depending only on $\mathscr{M}_0$. For any $N \geq \#\mathscr{M}_0$, we associate the best errors related to (12) with the two mesh families $\mathbb{M}$ and $\mathbb{M}'$. Namely, the best approximation error

$$\sigma(v,N) := \min\left\{E\big(v,S(\mathscr{M})\big) \mid \mathscr{M} \in \mathbb{M},\ \#\mathscr{M} \leq N\right\}$$

with continuous piecewise polynomial functions over conforming meshes with less than N elements, which is greater than the corresponding best error

$$\sigma'(v,N) := \min\{E\big(v,S^{\ell,-1}(\mathscr{M}')\big) \mid \mathscr{M}' \in \mathbb{M}',\ \#\mathscr{M}' \leq N\} \tag{31}$$

with possibly discontinuous piecewise polynomial functions over possibly non-conforming meshes. There also holds an inequality in the opposite direction, which may be seen as a generalization of Theorem 2.

Theorem 6 (Non-conforming element decoupling) *Assume that the initial mesh $\mathscr{M}_0$ is conforming, admissible, and that all its stars are $(d-1)$-face-connected. Then there exist constants C_1 and C_2 depending only on $\mathscr{M}_0$ such that, for any $v \in X \subset H^1(\Omega)$ and $N \geq N_0 := \#\mathscr{M}_0$, there holds*

$$\sigma(v,N) \leq C_1\sigma'\left(v, \left\lfloor \frac{N-N_0}{C_2} \right\rfloor + N_0\right).$$

Proof Set $N' := \lfloor (N-N_0)/C_{\mathrm{cmpl}} \rfloor + N_0$ and choose an optimal possibly non-conforming mesh $\mathscr{M} \in \mathbb{M}'$ such that $E\big(v, S^{\ell,-1}(\mathscr{M}')\big) = \sigma'(v,N')$. Since $\mathscr{M}_0$ is admissible, (30) yields $\#\mathscr{M} \leq N$ for $\mathscr{M} := \mathtt{complete}(\mathscr{M}')$. Hence

$$\begin{aligned}\sigma(v,N) &\leq E\big(v,S(\mathscr{M})\big) \leq \delta(\mathscr{M}) E\big(v,S^{\ell,-1}(\mathscr{M})\big) \leq \delta(\mathscr{M}) E\big(v,S^{\ell,-1}(\mathscr{M}')\big) \\ &= \delta(\mathscr{M})\sigma'(v,N')\end{aligned}$$

due to $S^{\ell,-1}(\mathscr{M}) \subset S^{\ell,-1}(\mathscr{M}')$. The shape coefficient for any mesh in $\mathbb{M}$ is bounded in terms of the one of $\mathscr{M}_0$; see, e.g., [22, Lemma 4.1]. Moreover, any mesh in $\mathbb{M}$ inherits from $\mathscr{M}_0$ that all its stars are $(d-1)$-face-connected. Theorem 2 therefore implies $\delta(\mathscr{M}) \leq \delta$, were δ depends only on $\mathscr{M}_0$. The claimed inequality thus holds with $C_1 = \delta$ and $C_2 = C_{\mathrm{cmpl}}$. □

The number of competing meshes for the best errors grows exponentially with $N - \#\mathscr{M}_0$. Nevertheless, the following variant of adaptive tree approximation by P. Binev and R. DeVore [6] constructs near best meshes with $O(N)$ operations and computations of the local error functional

$$\varepsilon(K) := e(v,K)^2,$$

where $K \subset \overline{\Omega}$ is any d-simplex and $v \in H^1(\Omega)$ the target function. Given a threshold $t > 0$, we proceed as follows:

$\mathscr{M}_t' := \emptyset$;
`for all` $K \in \mathscr{M}_0$
  $\eta(K) := \varepsilon(K)$;
  `if` $\eta(K) > t$ `then grow`(K);
`end for`
$\mathscr{M}_t :=$ `complete`$(\mathscr{M}_t')$;

where `grow`(K) grows the subtree generating $\mathscr{M}_t'$ and collects its leafs by

$(K_1,K_2) =$ `bisect`(K);
`for` $i = 1,2$
  $\eta(K_i) := \big[\varepsilon(K_i)^{-1} + \eta(K)^{-1}\big]^{-1}$
  `if` $\eta(K_i) > t$ `then`
    `grow`(K_i);
  `else`
    $\mathscr{M}_t' := \mathscr{M}_t' \cap \{K_i\}$;

and `bisect`(K) implements the bisection of a single simplex; see, e.g., [22, §4.1].

The core of this algorithm is the thresholding algorithm in [7] with the following important difference: the local functional $\eta(K)$ depends not only on the local error functionals but also on their history within in the subdivision hierarchy.

There are noteworthy variants of this algorithm. In particular, the threshold t can be avoided by successively bisecting the elements maximizing the indicators $\eta(K)$ of the current mesh; see [6]. In this case one may also ensure the conformity of the mesh at any intermediate step. For these variants, the following theorem presents only non-essential changes.

Theorem 7 (Tree approximation) *Assume that the initial mesh $\mathscr{M}_0$ is conforming, admissible, and that all its stars are $(d-1)$-face-connected. Then there exist constants C_1 and C_2 depending only on $\mathscr{M}_0$ such that, for any $v \in X$ and any threshold $t > 0$, the output mesh $\mathscr{M}$ of the tree algorithm verifies*

$$E\big(v, S(\mathscr{M})\big) \le C_1 \sigma'\left(v, \left\lceil \frac{\#\mathscr{M}}{C_2} \right\rceil\right)$$

whenever $\#\mathscr{M} \ge C_2(2\#\mathscr{M}_0 + 1)$.

Proof The local error functional $\varepsilon(K) = e(v,K)^2 = \inf_{P\in\mathscr{P}_\ell(K)} |v-P|_K^2$ obviously depends only on the target function v and the simplex K. Moreover, it is subadditive: if $(K_1,K_2) = \texttt{bisect}(K)$, then $\varepsilon(K_1) + \varepsilon(K_2) \le |v-P|_{K_1}^2 + |v-P|_{K_2}^2$ for any $P \in \mathscr{P}_\ell(K)$ and thus

$$\varepsilon(K_1) + \varepsilon(K_2) \le \varepsilon(K).$$

Hence, Theorem 4 of P. Binev's contribution in [1] applies to the above for-loop that constructs $\mathscr{M}_t'$. Writing

$$N_0 := \#\mathscr{M}_0, \quad N' := \#\mathscr{M}_t', \quad L' := N' - N_0$$

and observing $E\big(v, S^{\ell,-1}(\mathscr{M})\big)^2 = \sum_{K\in\mathscr{M}} \varepsilon(K)$ for any $\mathscr{M}' \in \mathbb{M}'$, we therefore have

$$E\big(v, S^{\ell,-1}(\mathscr{M}_t')\big) \le \min_{0\le l\le L'} \left(1 + \frac{l + \min\{l, N_0\}}{L' + 1 - l}\right) \sigma'(v, N_0 + l).$$

Under the assumption $N' \ge 2N_0$ this simplifies to

$$E\big(v, S^{\ell,-1}(\mathscr{M}_t')\big) \le \min_{2N_0\le n\le N'} \frac{N'+1}{N'+1-n} \sigma'(v,n).$$

Since $\mathscr{M}_0$ is admissible, (30) ensures

$$N - N_0 \le C_{\mathrm{cmpl}}(N' - N_0)$$

for $N := \#\mathscr{M}_t$. In view of $N \ge N' \ge 2N_0$, we may use the simpler inequality $N \le \tilde{C}_2 N'$ with $\tilde{C}_2 = 2C_{\mathrm{cmpl}}$. Since $x \mapsto x/(x-c)$ is decreasing and E monotone, one thus can derive

$$E\big(v, S^{\ell,-1}(\mathscr{M}_t)\big) \le \min_{2N_0\le n\le N/\tilde{C}_2} \frac{N}{N - \tilde{C}_2 n} \sigma'(v,n)$$

whenever $N \ge 2\tilde{C}_2 N_0$. Similarly as in the proof of Theorem 6, Theorem 5 therefore implies

$$E\big(v, S(\mathscr{M}_t)\big) \le \delta \min_{2N_0\le n\le N/\tilde{C}_2} \frac{N}{N - \tilde{C}_2 n} \sigma'(v,n)$$

where δ depends only on $\mathscr{M}_0$. Choosing $n \in \mathbb{N}$ such that $N/(2\tilde{C}_2) - 1 < n \le N/(2\tilde{C}_2)$, the claimed inequality follows with $C_1 = 2\delta$ and $C_2 = 2\tilde{C}_2$ as above. □

Theorem 7 is of non-asymptotic nature; as can be seen from the proof, the condition $\#\mathscr{M} \geq C_2(2\#\mathscr{M}_0+1)$ arising from $N' \geq 2N_0$ is of simplifying nature. In particular, it does not suppose any regularity beyond $H^1(\Omega)$ of the target function.

In [6, §7] a similar algorithm is proposed. It relies on the local error functionals

$$\tilde{\varepsilon}(K) := \inf\left\{ |v-V|_{\omega_K} \mid V \in S(\mathscr{M}_K)\right\}$$

where $\mathscr{M}_K$ is the set of elements of the so-called minimal ring $R^-(K)$ around K given by

$$R^-(K) := \bigcap_{\mathscr{M}\in\mathbb{M}:\mathscr{M}\ni K} R(K,\mathscr{M}) \quad \text{with} \quad R(K,\mathscr{M}) := \bigcup_{K'\in\mathscr{M}:K'\cap K\neq\emptyset} K'.$$

In view of the minimality of the ring, these error functionals do not depend on the surrounding mesh. They are not subadditive, but weakly subadditive with respect to repeated bisections and so another variant of the tree algorithm with a similar statement to Theorem 7 still applies. The local error functionals $e(K)$ are however simpler to implement and can be combined with single bisections. Thus, the use of Theorem 2 here permits an algorithmic simplification.

Acknowledgements The author wants to thank Francesco Mora for useful discussions regarding Theorem 2 and the referees for suggesting improvements in the presentation.

References

1. *Adaptive numerical methods for PDEs*, Oberwolfach Rep., 4 (2007), pp. 1663–1739. Abstracts from the workshop held June 10–16, 2007, Organized by Rolf Rannacher, Endre Süli and Rüdiger Verfürth, Oberwolfach Reports. Vol. 4, no. 3.
2. I. Babuška and W. C. Rheinboldt, *Error estimates for adaptive finite element computations*, SIAM J. Numer. Anal., 15 (1978), pp. 736–754.
3. M. Bebendorf, *A note on the Poincaré inequality for convex domains*, Z. Anal. Anwendungen, 22 (2003), pp. 751–756.
4. P. Binev, W. Dahmen, and R. DeVore, *Adaptive finite element methods with convergence rates*, Numer. Math., 97 (2004), pp. 219–268.
5. P. Binev, W. Dahmen, R. DeVore, and P. Petrushev, *Approximation classes for adaptive methods*, Serdica Math. J., 28 (2002), pp. 391–416. Dedicated to the memory of Vassil Popov on the occasion of his 60th birthday.
6. P. Binev and R. DeVore, *Fast computation in adaptive tree approximation*, Numer. Math., 97 (2004), pp. 193–217.
7. M. Š. Birman and M. Z. Solomjak, *Piecewise polynomial approximations of functions of classes W_p^α*, Mat. Sb. (N.S.), 73 (115) (1967), pp. 331–355.
8. A. Bonito and R. H. Nochetto, *Quasi-optimal convergence rate of an adaptive discontinuous Galerkin method*, SIAM J. Numer. Anal., 48 (2010), pp. 734–771.
9. J. Bourgain, H. Brezis, and P. Mirnoescu, *Another look at Sobolev spaces*, in Optimal control and partial differential equations, E. R. J. L. Mendali and A. Sulem, eds., IOS Press, 2001, pp. 439–455.
10. D. Braess, *Finite elements*, Cambridge University Press, Cambridge, second ed., 2001. Theory, fast solvers, and applications in solid mechanics, Translated from the 1992 German edition by Larry L. Schumaker.
11. S. C. Brenner, *Two-level additive Schwarz preconditioners for nonconforming finite element methods*, Math. Comp., 65 (1996), pp. 897–921.
12. ———, *Convergence of nonconforming multigrid methods without full elliptic regularity*, Math. Comp., 68 (1999), pp. 25–53.

13. S. C. BRENNER AND L. R. SCOTT, *The mathematical theory of finite element methods*, vol. 15 of Texts in Applied Mathematics, Springer, New York, third ed., 2008.
14. F. CAMACHO AND A. DEMLOW, *L_2 and pointwise a posteriori error estimates for FEM for elliptic PDE on surfaces*, IMA J. Numer. Anal., published online: July 31, 2014.
15. P. CLÉMENT, *Approximation by finite element functions using local regularization*, Revue Francaise Automat. Informat. Recherche. Operationelle Ser. Rouge Anal. Numér., 9 (1975), pp. 77–84.
16. S. DEKEL AND D. LEVIATAN, *The Bramble-Hilbert lemma for convex domains*, SIAM J. Math. Anal., 35 (2004), pp. 1203–1212.
17. R. A. DEVORE, *Nonlinear approximation*, in Acta numerica, 1998, vol. 7 of Acta Numer., Cambridge Univ. Press, Cambridge, 1998, pp. 51–150.
18. T. DUPONT AND R. SCOTT, *Polynomial approximation of functions in Sobolev spaces*, Math. Comp., 34 (1980), pp. 441–463.
19. T. GUDI, *A new error analysis for discontinuous finite element methods for linear elliptic problems*, Math. Comp., 79 (2010), pp. 2169–2189.
20. O. A. KARAKASHIAN AND F. PASCAL, *A posteriori error estimates for a discontinuous Galerkin approximation of second-order elliptic problems*, SIAM J. Numer. Anal., 41 (2003), pp. 2374–2399 (electronic).
21. R. S. LAUGESEN AND B. A. SIUDEJA, *Minimizing Neumann fundamental tones of triangles: an optimal Poincaré inequality*, J. Differential Equations, 249 (2010), pp. 118–135.
22. R. H. NOCHETTO, K. G. SIEBERT, AND A. VEESER, *Theory of adaptive finite element methods: an introduction*, in Multiscale, nonlinear and adaptive approximation, Springer, Berlin, 2009, pp. 409–542.
23. L. E. PAYNE AND H. F. WEINBERGER, *An optimal Poincaré-inequality for convex domains*, Archive Rat. Mech. Anal., 5 (1960), pp. 286–292.
24. R. SACCHI AND A. VEESER, *Locally efficient and reliable a posteriori error estimators for Dirichlet problems*, Math. Models Methods Appl. Sci., 16 (2006), pp. 319–346.
25. L. R. SCOTT AND S. ZHANG, *Finite element interpolation of nonsmooth functions satisfying boundary conditions*, Math. Comp., 54 (1990), pp. 483–493.
26. R. STEVENSON, *The completion of locally refined simplicial partitions created by bisection*, Math. Comp., 77 (2008), pp. 227–241 (electronic).
27. A. VEESER AND R. VERFÜRTH, *Explicit upper bounds for dual norms of residuals*, SIAM J. Numer. Anal., 47 (2009), pp. 2387–2405.
28. R. VERFÜRTH, *A note on polynomial approximation in Sobolev spaces*, M2AN Math. Model. Numer. Anal., 33 (1999), pp. 715–719.
29. J. XU AND L. ZIKATANOV, *Some observations on Babuška and Brezzi theories*, Numer. Math., 94 (2003), pp. 195–202.