POSITIVITY PRESERVING GRADIENT APPROXIMATION WITH LINEAR FINITE ELEMENTS

ANDREAS VEESER

Abstract. Preserving positivity precludes that linear operators onto continuous piecewise affine functions provide near best approximations of gradients. Linear interpolation thus does not capture the approximation properties of positive continuous piecewise affine functions. To remedy, we assign nodal values in a nonlinear fashion such that their global best error is equivalent to a suitable sum of local best errors with positive affine functions. As one of the applications of this equivalence, we consider the linear finite element solution to the elliptic obstacle problem and derive that its error is bounded in terms of these local best errors.

1. Introduction

Finite element functions are very useful in the numerical solution of partial differential equations. In the context of linear elliptic equations of second order, a basic result about their approximation properties is

$$\forall u \in H^1_0(\Omega) \quad \inf_{s\in S_0} |u-s|_{1;\Omega} \approx \left(\sum_{K\in\mathcal{M}} \inf_{p\in\mathbb{P}_1(K)} |u-p|^2_{1;K} \right)^{\frac{1}{2}}, \tag{1}$$

where $\mathcal{M}$ is a simplicial face-to-face mesh of a domain $\Omega \subseteq \mathbb{R}^d$, $d \geq 2$, and S_0 denotes the space of piecewise affine functions that are continuous and vanish on $\partial\Omega$. Employing the H^1-seminorm as a measure, this equivalence relates the global best error in S_0 to the local best errors with affine functions over elements. Note that the left-hand side involves continuity across interelement faces and a boundary condition, while the right-hand side does not. This simpler nature of the left-hand side prepares the ground for at least the following applications:

- derivation of a priori error bounds in terms of broken extra regularity,
- adaptive tree approximation of Binev and DeVore [1],
- parallel approximate computation of the global best error in S_0.

The proof of the nontrival part '$\lesssim$' of (1) relies on the construction of a suitable element v in S_0. Assuming that this construction is represented by the operator $\Pi : H^1_0(\Omega) \to S_0$, we readily see that the following global invariance and stability conditions are necessary:

$$\forall s \in S_0 \;\; \Pi s = s \quad \text{and} \quad \forall u \in H^1_0(\Omega) \;\; |\Pi u|_{1;\Omega} \lesssim |u|_{1;\Omega}\,.$$

The linear interpolation operator of Scott and Zhang [9] verifies these conditions and even local counterparts thereof. The latter allows to prove '$\lesssim$'; see Veeser [10].

The elliptic obstacle problem motivates to establish a counterpart of the best error decomposition (1) for the following setting. In order to simplify the discussion, let us neglect the boundary condition for the rest of this introduction and replace

- $H^1_0(\Omega)$ by $H^1(\Omega)^+ := \{v \in H^1(\Omega) \mid v \geq 0\}$,
- S_0 by the space S^+ of positive continuous piecewise affine functions and
- $\mathbb{P}_1(K)$ by $\mathbb{P}^+_1(K) := \{p \in \mathbb{P}_1(K) \mid p \geq 0\}$.

Moreover, since $|\cdot|_{1;K}$ is insensitive to additive constants and so

$$\inf_{p\in\mathbb{P}_1(K)} |u-p|_{1;K} = \inf_{p\in\mathbb{P}_1^+(K)} |u-p|_{1;K}\,,$$

we augment the H^1-seminorm with an L^2-contribution that scales in the same way:

$$\|v\|_{1;\omega} := \Big(|v|_{1;\omega}^2 + \operatorname{diam}(\omega)^{-2}\,|v|_{0;\omega}^2\Big)^{\frac{1}{2}}, \quad v\in H^1(\omega), \quad \omega\subseteq\Omega. \tag{2}$$

In other words: we are interested in the best error decomposition

$$\forall u\in H^1(\Omega)^+ \quad \inf_{s\in S^+} \|u-s\|_{1;\Omega} \approx \left(\sum_{K\in\mathcal{M}} \inf_{p\in\mathbb{P}_1^+(K)} \|u-p\|_{1;K}^2\right)^{\frac{1}{2}}, \tag{3}$$

where target function, global and local approximants are positive.

In order to prove it, we may mimic the proof of the best error decomposition (1) without constraints. Then the construction of a suitable $v=\Pi^+u$ element in S^+ has to satisfy

$$\forall s\in S^+ \;\; \Pi^+ s = s \quad \text{and} \quad \forall u\in H^1(\Omega)^+ \;\; \big\|\Pi^+ u\big\|_{1;\Omega} \lesssim \|u\|_{1;\Omega}\,. \tag{4}$$

If Π^+ is a linear operator from $H^1(\Omega)$ onto the space S of continuous piecewise affine functions, then these conditions strengthen to

$$\Pi^+\big(H^1(\Omega)^+\big)\subset S^+, \quad \forall s\in S \;\; \Pi^+ s = s, \quad \forall u\in H^1(\Omega) \;\; \big\|\Pi^+u\big\|_{1;\Omega} \lesssim \|u\|_{1;\Omega}\,. \tag{5}$$

Exploiting the results of Nochetto and Wahlbin [8], we derive that the linearity and the first two conditions force Π to be the Lagrange interpolation operator. Since Lagrange interpolation is not H^1-stable for $d\geq 2$, this is in contradiction with the third condition.

Given the impossibility of an appropriate linear operator, we resort to a nonlinear projection operator in the vein of the Scott-Zhang construction. It satisfies (4) and local counterparts thereof. Using this nonlinear projection in the approach of [10], we then establish the best error decomposition (3), where positivity is preserved.

Turning back to the elliptic obstacle problem, let us consider the case of a lower obstacle χ, continuous and piecewise affine for the sake of simplicity. The preceding results then yield the following error bound for the finite element approximation:

$$|u-U|_{1;\Omega} \lesssim \left[\sum_{K\in\mathcal{M}} e(K)\left(e(K) + \sum_{z\in\mathcal{L}_1\cap\omega_K} \|\mu\|_{-1;\omega_z}\right)\right]^{\frac{1}{2}}, \tag{6}$$

where $e(K)$ is the best $\|\cdot\|_{1;K}$-error with affine functions that are larger than the obstacle, $\mathcal{L}_1\cap\omega_K$ denotes the vertices of the patch ω_K around K, ω_z is the star in $\mathcal{M}$ around z, and μ is the Lagrange multiplier associated with the obstacle. Notice that this bound

- consists of local terms and is free of extra regularity,
- vanishes whenever the error is zero,
- is essentially independent of load perturbations that do not affect the error,
- implies first order convergence under suitable smoothness assumptions.

The rest of this article is organized as follows. Section 2 fixes the notation for piecewise affine functions and characterizes their positivity. Section 3 first shows that there is no linear operator satisfying (5) and then establishes the constrained best error decomposition (3). The concluding Section 4 discusses the application to the elliptic obstacle problem, establishing in particular the error bound (6).

2. Continuous piecewise affine functions and positivity

We introduce our notations around continuous piecewise affine functions over simplicial meshes and discuss their positivity.

Let $d \in \mathbb{N}$ denote the dimension. Given $n \in \{0, \dots, d\}$, an *n-simplex* $C \subseteq \mathbb{R}^d$ is the convex hull of $n+1$ points $z_1, \dots, z_{n+1} \in \mathbb{R}^d$ spanning an n-dimensional affine space. The uniquely determined points $z_1, \dots, z_{n+1}$ are the vertices of C and form the set $\mathcal{L}_1(C)$. If $n \geq 1$, we let $\mathcal{F}_C$ denote the set of the $(n-1)$-dimensional faces of C, which are the $(n-1)$-simplices arising by picking n distinct vertices from $\mathcal{L}_1(C)$. We write $h_C := \operatorname{diam}(C)$ for the diameter of C, ρ_C for the diameter of its largest inscribed n-dimensional ball, and γ_C for its *shape coefficient* $\gamma_C := h_C/\rho_C$.

$\mathbb{P}_1(C)$ indicates the space of affine functions on C. An affine function $p \in \mathbb{P}_1(C)$ is determined by its point values at the vertices $\mathcal{L}_1(C)$. For each vertex $z \in \mathcal{L}_1(C)$, its *barycentric coordinate* λ_z^C is the unique affine function on C such that $\lambda_z^C(y) = \delta_{zy}$ for all $y \in \mathcal{L}_1(C)$. It is well known, see [5], that

$$\{\lambda_z^C\}_{z \in \mathcal{L}_1(C)} \text{ forms a basis of } \mathbb{P}_1(C)$$

with the representation formula

$$\forall p \in \mathbb{P}_1(C) \quad p = \sum_{z \in \mathcal{L}_1(C)} p(z)\lambda_z^C \tag{7}$$

and

$$\sum_{z \in \mathcal{L}_1(C)} \lambda_z^C = 1 \quad \text{where} \quad 0 \leq \lambda_z^C \leq 1 \quad \text{in } C. \tag{8}$$

We thus have the following simple description of

$$\mathbb{P}_1^+(C) := \{p \in \mathbb{P}_1(C) \mid p \geq 0\}.$$

Lemma 1 (Positivity and vertices)**.** *Any affine function $p \in \mathbb{P}_1(C)$ is positive if and only if it is positive at the vertices of C.*

Let $d \geq 2$ and $\Omega \subset \mathbb{R}^d$ be a nonempty, open, connected, bounded and polyhedral set with Lipschitz boundary $\partial\Omega$. Furthermore, let $\mathcal{M}$ be a simplicial, face-to-face *mesh* of Ω. More precisely, $\mathcal{M}$ is a finite collection of d-simplices in $\mathbb{R}^d$ such that

- $\overline{\Omega} = \bigcup_{K \in \mathcal{M}} K$,
- the intersection of two arbitrary elements $K_1, K_2 \in \mathcal{M}$ is either empty or an n-simplex with $n \in \{0 \dots, d\}$,
- $\mathcal{L}_1(K_1 \cap K_2) = \mathcal{L}_1(K_1) \cap \mathcal{L}_1(K_2)$ whenever the intersection of $K_1, K_2 \in \mathcal{M}$ is nonempty.

We let $\mathcal{F} := \bigcup_{K \in \mathcal{M}} \mathcal{F}_K$ denote the $(d-1)$-dimensional faces of $\mathcal{M}$ and distinguish between boundary faces $\mathcal{F}_{\partial\Omega} := \{F \in \mathcal{F} \mid F \subseteq \partial\Omega\}$ and interior faces $\mathcal{F}_\Omega := \mathcal{F} \setminus \mathcal{F}_{\partial\Omega}$. The shape coefficient of $\mathcal{M}$ is

$$\gamma_{\mathcal{M}} := \max_{K \in \mathcal{M}} \gamma_K.$$

The space of continuous functions that are piecewise affine over $\mathcal{M}$ is

$$S = \left\{s \in C^0(\overline{\Omega}) \mid \forall K \in \mathcal{M} \;\; s_{|K} \in \mathbb{P}_1(K)\right\}$$

and a subspace of the Sobolev space $H^1(\Omega)$. The global counterpart of the local bases λ_z^K, $z \in \mathcal{L}_1(K)$, is given as follows. We let $\mathcal{L}_1 := \bigcup_{K \in \mathcal{M}} \mathcal{L}_1(K)$ denote the vertices of $\mathcal{M}$ and, for every vertex $z \in \mathcal{L}_1(K)$, we define a function λ_z by

$$\lambda_{z|K} := \begin{cases} \lambda_z^K, & \text{if } K \ni z, \\ 0, & \text{otherwise,} \end{cases} \qquad K \in \mathcal{M}.$$

Again, it is well known that

$$\{\lambda_z\}_{z\in\mathcal{L}_1} \text{ forms a basis of } S$$

with the representation formula

$$\forall s \in S \quad s = \sum_{z\in\mathcal{L}_1} s(z)\lambda_z. \tag{9}$$

The support of the basis function λ_z is the star $\omega_z := \bigcup_{K'\ni z} K'$ around $z \in \mathcal{L}_1$. Since $\partial\Omega$ is Lipschitz, stars are face-connected in the sense of [10]: given a vertex $z \in \mathcal{L}_1$, an element $K \in \mathcal{M}$, and a face $F \in \mathcal{F}$ with $z \in K \cap F$, there exists a path $(K_i)_{i=1}^n \subset \mathcal{M}$ such that

$$K_1 = K \text{ and } K_n \supset F, \tag{10a}$$

$$\text{each } K_i \text{ contains } z, \tag{10b}$$

$$\text{each } K_i \cap K_{i+1} \in \mathcal{F}_\Omega \text{ is an interior face.} \tag{10c}$$

All supports of the basis functions associated with the element $K \in \mathcal{M}$ are contained in the patch $\omega_K := \bigcup_{K'\cap K\neq\emptyset} K'$.

If V is some linear space of real-valued functions, then

$$V^+ := \{v \in V \mid v \geq 0\}$$

denotes its convex cone of positive functions. If $V = S$, Lemma 1 implies

$$S^+ = \big\{s \in S \mid \forall z \in \mathcal{L}_1 \;\; s(z) \geq 0\big\}. \tag{11}$$

In other words: S^+ is the conical hull of the functions $\{\lambda_z\}_{z\in\mathcal{L}_1}$.

If not specified differently, C_* stands for a function which is not necessarily the same at each occurrence and possibly depending on $* \subseteq \{d, \gamma_\mathcal{M}\}$, increasing in $\gamma_\mathcal{M}$ if present. For instance, we have, for $K, K' \in \mathcal{M}$,

$$K \cap K' \neq \emptyset \quad \Longrightarrow \quad |K| \leq C_{\gamma_\mathcal{M}} |K'| \text{ and } h_K \leq C_{\gamma_\mathcal{M}} \rho_{K'} \tag{12}$$

and, for $K \in \mathcal{M}$, $z \in \mathcal{L}_1(K)$ and the norm from (2),

$$c_d |K|^{\frac{1}{2}} h_K^{-1} \leq \|\lambda_z\|_{1;K} \leq C_d |K|^{\frac{1}{2}} \rho_K^{-1}. \tag{13}$$

If there is no danger of confusion, $A \leq C_* B$ may be abbreviated as $A \lesssim B$.

3. Positivity preserving gradient approximation

We are interested in the following conical approximation problem: given a function $u \in H^1(\Omega)^+$, find a function from S^+ that is close-by with respect to the norm $\|\cdot\|_\Omega$ from (2). An approximation operator $\Pi^+ : H^1(\Omega)^+ \to S^+$ is $\mathbb{P}_1^+$-quasi-optimal whenever there exists a constant $D \geq 1$ such that

$$\forall u \in H^1(\Omega)^+ \quad \left\|u - \Pi^+ u\right\|_{1;\Omega} \leq D \left(\sum_{K\in\mathcal{M}} \inf_{p\in\mathbb{P}_1^+(K)} \|u - p\|_{1;K}^2 \right)^{\frac{1}{2}}.$$

It is the goal of this section to devise an operator that satisfies this property and is local.

3.1. Impossibility of linearity.

Generally speaking, linearity is considered to be a desirable property for approximation operators. For the above approximation problem, it is however precluded by the following observation.

Lemma 2 (No quasi-optimal positive linear interpolation). *There is no linear operator $L : H^1(\Omega) \to S$ such that*

$$\forall u \in H^1(\Omega)^+ \qquad Lu \in S^+ \quad \textit{and} \quad \|u - Lu\|_{1;\Omega} \leq C \inf_{s\in S^+} \|u - s\|_{1;K} \tag{14}$$

with some constant $C \geq 1$.

Proof. Let us assume that $L : H^1(\Omega) \to S$ is linear and satisfies (14) and look for a contradiction. We first show that the quasi-optimality in (14) requires the invariance

$$\forall s \in S \quad Ls = s. \tag{15}$$

To this end, we note that, since $S^+ \subseteq H^1(\Omega)^+$, it readily gives the partial invariance

$$\forall s \in S^+ \quad Ls = s. \tag{16}$$

In order to generalize to the full one, we introduce

$$I^{\pm}s := \sum_{z \in \mathcal{L}_1} \max\{0, \pm s(z)\}\lambda_z, \qquad s \in S,$$

such that $s = I^+s - I^-s$. Combining this identity with the linearity of L and the partial invariance (16), we infer (15): for all $s \in S$,

$$Ls = L(I^+s) + L(-I^-s) = I^+s - I^-s = s.$$

Next, we prove that L is defined on all $C^0(\overline{\Omega})$ and satisfies there

$$\sup_{\Omega} |Lu| \leq \sup_{\Omega} |u| \quad \text{and} \quad u \geq 0 \implies Lu \geq 0. \tag{17}$$

We start by showing that L is defined on $C^0(\overline{\Omega})$ and satisfies there the stability estimate. As $H^1(\Omega) \cap C^0(\overline{\Omega})$ is a dense subspace of $C^0(\overline{\Omega})$, c.f. [7, §7.2] and L linear, it suffices to verify the stability bound for $u \in H^1(\Omega) \cap C^0(\overline{\Omega})$. Writing $M := \sup_{\Omega} |u| \in [0, \infty[$, we have $u - M \leq 0$. Hence, by the first part of (14) and (16),

$$Lu = LM + L(u - M) \leq M.$$

Similarly, $-u - M \leq 0$ implies $Lu \geq -M$ and so the stability bound on $H^1(\Omega)$ in (17) is verified. In order to show the positivity, let $u \in C^0(\overline{\Omega})$ be positive and choose $(u_k)_k \subseteq H^1(\Omega) \cap C^0(\overline{\Omega})$ such that $\sup_{\Omega} |u - u_k| \to 0$ as $k \to \infty$. From the first parts of (14) and (17), we then obtain

$$\begin{aligned} Lu &= L(u - u_k) + Lu_k \geq -\sup_{\Omega} |u - u_k| + \inf_{\Omega} u_k \\ &\geq -\sup_{\Omega} |u - u_k| + \inf_{\Omega}(u_k - u) \geq -2 \sup_{\Omega} |u - u_k| \to 0 \quad \text{as} \quad k \to \infty. \end{aligned}$$

In view of (15) and (17), the operator $L : C^0(\overline{\Omega}) \to S$ is bounded, linear, positive and reproduces all continuous piecewise affine functions. Consequently, Corollary 2 of Nochetto and Wahlbin [8] reveals that L is the Lagrange interpolation operator relying on the point evaluations:

$$\forall z \in \mathcal{L}_1 \quad Lu(z) = u(z). \tag{18}$$

On the other hand, the quasi-optimality of (14) entails also that L is H^1-stable with

$$\|Lu\|_{1;\Omega} \leq \sqrt{2}(1 + C) \|u\|_{1;\Omega} . \tag{19}$$

To see this, we proceed similarly as for the full invariance. We first note that the quasi-optimality of (14) readily implies

$$\forall u \in H^1(\Omega)^+ \quad \|Lu\|_{1;\Omega} \leq \|u\|_{1;\Omega} + \|Lu - u\|_{1;\Omega} \leq (1 + C) \|u\|_{1;\Omega}$$

because of $0 \in S^+$. Thus, given a general $u \in H^1(\Omega)$, we may write $u = u^+ - u^-$, where $u^{\pm} = \max\{\pm u, 0\} \geq 0$ are the positive and negative parts u and obtain

$$\begin{aligned} \|Lu\|_{1;\Omega} &\leq \left\|Lu^+\right\|_{1;\Omega} + \left\|Lu^-\right\|_{1;\Omega} \leq (1 + C) \left(\left\|u^+\right\|_{1;\Omega} + \left\|u^-\right\|_{1;\Omega} \right) \\ &= \sqrt{2}(1 + C) \|u\|_{1;\Omega} . \end{aligned}$$

In the last step, we have used $\nabla u = \nabla u^+ - \nabla u^-$, where the supports of ∇u^+ and ∇u^- are disjoint; see [7, Lemma 7.6].

Since $d \geq 2$, the H^1-stability (19) is in contradiction with the point evaluations in (18). In fact, given any vertex $z \in \mathcal{L}_1$, the functions

$$u_k := \min\{k, u\}, \quad k \in \mathbb{N}, \quad \text{with} \quad u(x) := \ln \left| \ln \left(\frac{|x-z|}{2\operatorname{diam}(\Omega)} \right) \right|, \quad x \in \Omega,$$

satisfy, on the one hand, $\lim_{k\to\infty} u_k(y) = u(y)$ for all $y \in \Omega \setminus \{z\}$ and $u_k(z) = k$ and, on the other hand, $\|u_k\|_{1;\Omega} \leq \|u\|_{1;\Omega} < \infty$, cf. Example (1.4.3) of [2]. Hence, we have

$$k \|\lambda_z\|_{1;\Omega} - \left\| \sum_{y\in\mathcal{L}_1\setminus\{z\}} u_k(y)\lambda_y \right\|_{1;\Omega} \leq \|Lu_k\|_{1;\Omega} \leq \sqrt{2}(1+C) \|u\|_{1;\Omega} < \infty$$

and obtain a contradiction for $k \to \infty$. □

Condition (14) in Lemma 2 does not specify a boundary condition. Incorporating one reduces the contained information. Nevertheless, the conclusion of Lemma 2 persists. Let us illustrate this in the special case of vanishing boundary values.

Lemma 3 (. . . with vanishing boundary values). *Assume that there exists a vertex $z \in \mathcal{L}_1$ such that $\omega_z \cap \partial\Omega \neq \emptyset$. Then there is no linear operator $L : H^1_0(\Omega) \to S_0$ such that*

$$\forall u \in H^1_0(\Omega)^+ \qquad Lu \in S_0^+ \quad \textit{and} \quad \|u - Lu\|_{1;\Omega} \leq C_0 \inf_{s\in S_0^+} \|u - s\|_{1;K} \tag{20}$$

with some constant $C_0 \geq 1$.

Proof. Proceed as in the proof of Lemma 2 with obvious replacements of the spaces $H^1(\Omega)$, S, their cones of positive functions, and $C^0(\overline{\Omega})$. The only important change is in the use of Nochetto and Wahlbin [8]: instead of Corollary 2, apply Corollary 3 therefore and obtain

$$\forall z \in \mathcal{L}_1 \quad \omega_z \cap \partial\Omega \neq \emptyset \implies Lu(z) = u(z)$$

in lieu of (18). □

3.2. **Positive Scott-Zhang-like approximation.** Scott-Zhang interpolation [9] enjoys local stability and invariance properties as well as linearity. While the latter was not actually used in proving the best error localization (1), the other two appear to be crucial. In view of the preceding section, we thus drop linearity but otherwise mimic Scott-Zhang interpolation as closely as possible.

To this end, the key devices are the following local approximation operators. Given any face $F \in \mathcal{F}$, define a map $Q_F^+ : L^2(F) \to \mathbb{P}_1^+(F)$ by

$$Q_F^+ v \in \mathbb{P}_1^+(F) \quad \text{such that} \quad \left\| v - Q_F^+ v \right\|_{0;F} = \inf_{q\in\mathbb{P}_1^+(F)} \|v - q\|_{0;F}\,, \tag{21}$$

where $|\cdot|_{0;F}$ indicates the $L^2(F)$-norm. Since $\mathbb{P}_1^+$ is a nonempty, closed and convex subset of the Hilbert space $L^2(F)$, the projection theorem implies the following lemma.

Lemma 4 (Face projections). *For each $F \in \mathcal{F}$, the map Q_F^+ is well-defined, invariant on $\mathbb{P}_1^+(F)$, and satisfies the stability estimate*

$$\forall v_1, v_2 \in L^2(F) \qquad \left\| Q_F^+ v_1 - Q_F^+ v_1 \right\|_{0;F} \leq \|v_1 - v_2\|_{0;F}\,. \tag{22}$$

The operators Q_F, $F \in \mathcal{F}$ can be applied to any $u \in H^1(\Omega)$ thanks to the trace theorem [2, (1.6.6)]. For this purpose, we pick a face $F_z \in \mathcal{F}$ for any vertex $z \in \mathcal{L}_1$ such that

$$F_z \ni z \quad \text{and} \quad z \in \partial\Omega \implies F_z \subset \partial\Omega. \tag{23}$$

These chosen F_z are intended to be fixed with respect to the mesh $\mathcal{M}$. The role of the implication in (23) will be clarified in §3.4 below. Given $u \in H^1(\Omega)^+$, we then set

$$\Pi^+ u := \sum_{z \in \mathcal{L}_1} \big(Q^+_{F_z} u\big)(z) \lambda_z. \tag{24}$$

We readily see that the face projections impart their positivity and invariance,

$$\Pi^+ u \in S^+ \quad \text{and} \quad \forall s \in S^+ \;\; \Pi^+ s = s \tag{25}$$

so that Π^+ is a projection onto S^+. In light of the second condition in (23), the trace $\Pi^+ u_{|\partial\Omega}$ of the approximant depends only on the trace $u_{|\partial\Omega}$ of target functions, with corresponding invariance.

The construction of Π resembles the one of Scott-Zhang interpolation. The next remark shows that it may be viewed even as a generalization.

Remark 5 (Π^+ and Scott-Zhang interpolation)*.* Given $F \in \mathcal{F}$ and $z \in \mathcal{L}_1(F)$, let $\Psi_{F,z} \in \mathbb{P}_1(F)$ be given by

$$\forall q \in \mathbb{P}_1(F) \quad \int_F q \Psi_{F,z} = q(z).$$

Then the Scott-Zhang interpolation operator from [9] is

$$\Pi u = \sum_{z \in \mathcal{L}_1} \left(\int_{F_z} u \Psi_{F_z,z} \right) \lambda_z, \quad u \in H^1(\Omega).$$

On the other hand, if we replace in (21) the closed convex cone $\mathbb{P}^+_1(F)$ by the linear space $\mathbb{P}_1(F)$ and call the resulting operator Q_F, then Q_F is the $L^2(F)$-orthogonal projection onto $\mathbb{P}_1(F)$ and we have

$$\int_{F_z} u \Psi_{F_z,z} = \int_{F_z} Q_{F_z} u \Psi_{F_z,z} = \big(Q_{F_z} u\big)(z).$$

In other words: Π^+ arises from Scott-Zhang interpolation Π only by the change of the admissible shape functions on faces.

3.3. **$\mathbb{P}^+_1$-quasi-optimality of Π^+.** This section analyzes the approximation error of the projection Π^+. The most local result is formulated with the help of the best approximant on an element $K \in \mathcal{M}$, which is the output of $P^+_K : H^1(K)^+ \to \mathbb{P}^+_1(K)$ given by

$$P^+_K u \in \mathbb{P}^+_1(K) \quad \text{such that} \quad \big\|u - P^+_K u\big\|_{1;K} = \inf_{p \in \mathbb{P}^+_1(K)} \|u - p\|_{1;K} \tag{26}$$

with $\|\cdot\|_{1;K}$ as in (2). Similarly as for Q^+_F, the projection theorem ensures that P^+_K is well-defined, invariant on $\mathbb{P}^+_1(K)$, and $\|\cdot\|_{1;K}$-stable.

Theorem 6 (Nodal quasi-optimality)**.** *For any $u \in H^1(\Omega)^+$ and every vertex $z \in \mathcal{L}_1(K)$ of an element $K \in \mathcal{M}$, we have*

$$|\Pi^+ u(z) - P^+_K u(z)| \le C_d \sum_{K' \in \mathcal{M} : K' \ni z} \frac{h_{K'}}{|K'|^{\frac{1}{2}}} \inf_{p \in \mathbb{P}^+_1(K')} \|u - p\|_{1;K'}$$

Proof. Let $(K_i)_{i=1}^n$ be a minimal path from K to F_z satisfying (10) and, abbreviating the subscript K_i to i, write

$$\big|\Pi^+ u(z) - P^+_K u(z)\big| \le \big|Q^+_{F_z} u(z) - P^+_n u(z)\big| + \sum_{i=1}^{n-1} \big|P^+_{i+1} u(z) - P^+_i u(z)\big|. \tag{27}$$

We derive suitable bounds for the absolute values on the right-hand side and start with the first one. Using an inverse estimate in $\mathbb{P}_1(F)$, Lemma 4, and the trace identity [11, Proposition 4.2], we derive

$$\begin{aligned}\left|Q^+_{F_z}u(z) - P^+_n u(z)\right| &\leq C_d\,|F|^{-\frac{1}{2}}\left\|Q^+_{F_z}u - Q^+_{F_z}P^+_n u\right\|_{0;F_z}\\ &\leq C_d\,|F|^{-\frac{1}{2}}\left\|u - P^+_n u\right\|_{0;F_z} \leq C_d h_n\,|K_n|^{-\frac{1}{2}}\left\|u - P^+_n u\right\|_{1;n}.\end{aligned} \tag{28}$$

In order to bound the other absolute values, let $i \in \{1,\dots,n-1\}$, set $F_i := K_{i+1}\cap K_i$ and abbreviate also the subscript F_i to i. Here we derive

$$\begin{aligned}\left|P^+_{i+1}u(z) - P^+_i u(z)\right| &\leq C_d\,|F|^{-\frac{1}{2}}\left\|P^+_{i+1}u - P^+_i u\right\|_{0;i}\\ &\leq C_d\,|F|^{-\frac{1}{2}}\left(\left\|P^+_{i+1}u - u\right\|_{0;i} + \left\|u - P^+_i u\right\|_{0;i}\right)\\ &\leq C_d\left(\frac{h_{i+1}}{|K_{i+1}|^{\frac{1}{2}}}\left\|u - P^+_{i+1}u\right\|_{1;i+1} + \frac{h_i}{|K_i|^{\frac{1}{2}}}\left\|u - P^+_i u\right\|_{1;i}.\right)\end{aligned} \tag{29}$$

Inserting (28) and (29) into (27), we conclude the claimed bound, which is independent of the chosen path. □

Next, we consider the error of Π^+ within elements. Note that the following results implies a local stability estimate in terms of a broken H^1-norm.

Corollary 7 (Quasi-optimality within elements). *For any $u \in H^1(\Omega)^+$ and every element $K \in \mathcal{M}$, we have*

$$\left\|u - \Pi^+ u\right\|^2_{1;K} \leq C_{d,\gamma_{\mathcal{M}}} \sum_{K'\cap K\neq\emptyset}\ \inf_{p\in\mathbb{P}^+_1(K')} \left\|u - p\right\|^2_{1;K'},$$

where K' varies in $\mathcal{M}$.

Proof. We start with

$$\left\|u - \Pi^+ u\right\|_{1;K} \leq \left\|u - P^+_K u\right\|_{1;K} + \left\|\Pi^+ u - P^+_K u\right\|_{1;K}$$

and it remains to bound the second term suitably. Exploiting $\Pi^+u - P^+_K u \in \mathbb{P}_1(K)$, (7), Theorem 6, (12), and (13), we infer

$$\begin{aligned}\left\|\Pi^+ u - P^+_K u\right\|_{1;K} &\leq \sum_{z\in\mathcal{L}_1(K)} \left|\Pi^+ u(z) - P^+_K u(z)\right|\,\|\lambda_z\|_{1;K}\\ &\lesssim \sum_{K'\cap K\neq\emptyset}\ \inf_{p\in\mathbb{P}^+_1(K')}\|u - p\|_{1;K'}.\end{aligned}$$

We conclude by applying the Cauchy-Schwarz inequality on the sum and noting

$$\#\{K' \in \mathcal{M} \mid K'\cap K \neq \emptyset\} \leq C_{d,\gamma_{\mathcal{M}}}. \tag{30}$$

□

We conclude this section with the resulting bound of the global error, which in particular shows that Π^+ is superior to any linear approximation operator in that it provides near best approximations.

Corollary 8 (Global quasi-optimality). *For any $u \in H^1(\Omega)^+$, we have*

$$\left\|u - \Pi^+ u\right\|_{1;\Omega} \leq C_{d,\gamma_{\mathcal{M}}}\left(\sum_{K\in\mathcal{M}}\ \inf_{p\in\mathbb{P}^+_1(K)}\|u - p\|^2_{1;K}\right)^{\frac{1}{2}}.$$

Proof. We sum the local bound in Corollary 7 over all $K \in \mathcal{M}$ and obtain

$$\left\|u - \Pi^+ u\right\|^2_{1;\Omega} \leq \sum_{K\in\mathcal{M}}\left\|u - \Pi^+ u\right\|^2_{1;K} \lesssim \sum_{K\in\mathcal{M}}\ \inf_{p\in\mathbb{P}^+_1(K)}\|u - p\|^2_{1;K},$$

where we used also (30). □

The operator Π^+ is defined via an implicit local procedure. This procedure can be replaced by an explicit one, preserving positivity and $\mathbb{P}_1^+$-quasi-optimality.

Remark 9 (Alternative construction). Another possibility to modify the Scott-Zhang construction to achieve positivity of the approximation is

$$\widetilde{\Pi}^+ u := \sum_{z \in \mathcal{L}_1} \max\left\{0, \int_{F_z} u \Psi_{F_z,z}\right\} \lambda_z,$$

where we use the same notation as in Remark 5. Then Theorem 6, Corollaries 7 and 8 hold also with $\widetilde{\Pi}^+$ in place of Π^+. This follows just by replacing (28) with

$$\begin{aligned} \left|\max\left\{0, Q_{F_z} u(z)\right\} - P_n^+ u(z)\right| &= \left|\max\left\{0, Q_{F_z} u(z)\right\} - \max\left\{0, P_n^+ u(z)\right\}\right| \\ &\leq \left|Q_{F_z} u(z) - P_n^+ u(z)\right| \leq C_d \left|F\right|^{-\frac{1}{2}} \left\|Q_{F_z} u - Q_{F_z} P_n^+ u\right\|_{0;F_z} \\ &\leq C_d \left|F\right|^{-\frac{1}{2}} \left\|u - P_n^+ u\right\|_{0;F_z} \leq C_d h_n \left|K_n\right|^{-\frac{1}{2}} \left\|u - P_n^+ u\right\|_{1;n}. \end{aligned}$$

Therefore, in what follows, Π^+ can be always replaced by $\widetilde{\Pi}^+$.

3.4. **Best error decompositions.**

Resorting to the approximation properties of Π^+, we show that gluing, or coupling, elements via continuity and prescribing boundary values do not impair the approximation potential provided by the admissible shape functions $\mathbb{P}_1^+(K)$, $K \in \mathcal{M}$.

Let us first verify the best error localization (3) of the introduction.

Theorem 10 (Best error decomposition with positivity)**.** *For any $u \in H^1(\Omega)^+$, we have*

$$\inf_{s \in S^+} \|u - s\|_{1;\Omega} \leq C_{d,\gamma_\mathcal{M}} \left(\sum_{K \in \mathcal{M}} \inf_{p \in \mathbb{P}_1^+(K)} \|u - p\|_{1;K}^2 \right)^{\frac{1}{2}}.$$

Proof. We simply use $\Pi^+ u \in S^+$ and apply Corollary 8:

$$\inf_{s \in S^+} \|u - s\|_{1;\Omega}^2 \leq \left\|u - \Pi^+ u\right\|_{1;\Omega}^2 \leq C_{d,\gamma_\mathcal{M}} \sum_{K \in \mathcal{M}} \inf_{p \in \mathbb{P}_1^+(K)} \|u - p\|_{1;K}^2. \qquad \square$$

Next, we incorporate prescribed boundary values. For this purpose, we note first that the construction of Π^+ yields an approximation operator for boundary values as side-product. Indeed, given $v \in L^2(\partial\Omega)^+$, we set

$$\Pi_{\partial\Omega}^+ v = \sum_{z \in \mathcal{L}_1 \cap \partial\Omega} \big(Q_{F_z} v\big)(z) \lambda_{z|\partial\Omega},$$

which is well-defined thanks to the second part of (23) and satisfies

$$\Pi_{\partial\Omega}^+ v \in S_\partial^+ \quad \text{and} \quad \forall s \in S_\partial^+ \;\; \Pi_{\partial\Omega}^+ s = s.$$

with $S_\partial := \{s \in C^0(\partial\Omega) \mid \forall F \in \mathcal{F}_{\partial\Omega} \;\, s_{|F} \in \mathbb{P}_1(F)\}$. Moreover, we have

$$\forall u \in H^1(\Omega)^+ \quad \Pi^+ u_{|\partial\Omega} = \Pi_{\partial\Omega}^+(u_{|\partial\Omega}).$$

thanks to (7). As restriction of Scott-Zhang interpolation to the boundary, $\Pi_{\partial\Omega}^+$ can be applied to all admissible boundary values, in contrast to Lagrange interpolation, which is not defined for boundary values in $H^{\frac{1}{2}}(\partial\Omega) \setminus C^0(\partial\Omega)$.

Given $g \in H^{\frac{1}{2}}(\partial\Omega)^+$, we set

$$H_g^1(\Omega) := \{u \in H^1(\Omega) \mid v = g \text{ on } \partial\Omega\} \quad \text{and} \quad S_g := \{s \in S \mid s = \Pi^+ g \text{ on } \partial\Omega\}.$$

Although S_g^+ is strictly smaller than S^+, we can still bound the global best error es before if the target function is from $H_g^1(\Omega)^+$.

Theorem 11 (Decoupling with positivity and boundary values)**.** *Let $g \in H^{\frac{1}{2}}(\partial\Omega)^+$. For any $u \in H^1_g(\Omega)^+$, we have*

$$\inf_{s \in S_g^+} \|u - s\|_{1;\Omega} \le C_{d,\gamma_{\mathcal{M}}} \left(\sum_{K \in \mathcal{M}} \inf_{p \in \mathbb{P}_1^+(K)} \|u - p\|^2_{1;K} \right)^{\frac{1}{2}}.$$

Proof. Here we use the fact $\Pi^+ u \in S_g^+$ and conclude again with Corollary 8:

$$\inf_{s \in S_g^+} \|u - s\|^2_{1;\Omega} \le \left\|u - \Pi^+ u\right\|^2_{1;\Omega} \le C_{d,\gamma_{\mathcal{M}}} \sum_{K \in \mathcal{M}} \inf_{p \in \mathbb{P}_1^+(K)} \|u - p\|^2_{1;K}. \qquad \square$$

4. Application to the elliptic obstacle problem

Corollary 8 splits a global approximation problem into many local ones, which are independent of each other. Obviously, this can be used for the parallel approximate computation of the best error $\inf_{s \in S^+} \|u - s\|_{1;\Omega}$ with continuous, piecewise affine and positive functions. The splitting and the reciprocal independence is also useful in adaptive tree approximation. It can be applied exactly as in [10, §4.2]. One only has to employ $\inf_{p \in \mathbb{P}_1^+} \|u - p\|_{1;K}$ instead of $\inf_{p \in \mathbb{P}_1} |u - p|_{1;K}$ as local error functionals.

An application with new aspects concerns the elliptic obstacle problem, which reads as follows; see [4]: given a force $f \in H^{-1}(\Omega) = H^1_0(\Omega)^\star$, a lower obstacle $\chi \in H^1(\Omega)$, and $g \in H^{\frac{1}{2}}(\partial\Omega)$ with $g \ge \chi$ on $\partial\Omega$, set

$$\mathcal{A} := \{v \in H^1(\Omega) \mid v \ge \chi \text{ in } \Omega,\, v = g \text{ on } \partial\Omega\}$$

and

$$\text{find } u \in \mathcal{A} \quad \text{such that} \quad \forall v \in \mathcal{A} \ \int_\Omega \nabla u \cdot \nabla(v - u) \ge \langle f, v - u \rangle. \tag{31}$$

Since $\mathcal{A}$ is nonempty, closed and convex, the solution exists and is unique by [4, Theorems 1.1.1 and 1.1.2]. It is useful to introduce the functional $\mu \in H^{-1}(\Omega)$ given by

$$\langle \mu, \varphi \rangle := \langle f, \varphi \rangle - \int_\Omega \nabla u \cdot \nabla \varphi, \tag{32}$$

which can be viewed as virtual force exerted by the obstacle and Lagrange multiplier associated with the constraint $u \ge \chi$. We have

$$\forall \varphi \in H^1_0(\Omega)^+ \ \langle \mu, \varphi \rangle \le 0 \quad \text{and} \quad \forall \varphi \in H^1_0(\{u > \chi\}) \ \langle \mu, \varphi \rangle = 0,$$

where $\{u > \chi\} := \bigcup_{\varepsilon > 0} \{u - \chi - \varepsilon \ge 0\}$ and $\{u - \chi - \varepsilon \ge 0\}$ is the largest open set U such that $\int_U (u - \chi - \varepsilon)\varphi \ge 0$ for all $\varphi \in C^\infty_0(U)^+$. The set $\{u > \chi\}$ is the non-coincidence set, while the coincidence set $\{u = \chi\}$ is the largest open set U such that $\int_U (u - \chi)\varphi = 0$ for all $\varphi \in C^\infty_0(U)$. Here μ may be strictly negative.

A remarkable feature of the obstacle problem is that the solution may not change under certain perturbations of the force. For example, if u is the solution corresponding to f, it is also the solution for $f + \delta f$ whenever $\delta f \in H^{-1}(\Omega)$ such that $\operatorname{supp} \delta f \subseteq \{u = \chi\}$ and $\delta f \le -\mu$ in $H^{-1}(\Omega)$, i.e. $\langle \delta f + \mu, \varphi \rangle \le 0$ for all $\varphi \in H^1_0(\Omega)$. In other words: the solution operator of the obstacle problem losses information and so the force f cannot be recovered from the solution only.

We discretize problem with linear finite elements, using the notations from §3.4. In order to minimize technicalities in presenting the application of §3, we assume

$$\chi \in S \quad \text{and} \quad g \in S_\partial.$$

Then

$$\mathcal{A}_S := \{s \in S \mid s \ge \chi \text{ in } \Omega,\, s = g \text{ on } \partial\Omega\} \subset \mathcal{A} \tag{33}$$

is conforming as well as nonempty, closed and convex. Hence

$$\text{find } U \in \mathcal{A}_S \quad \text{such that} \quad \forall v \in \mathcal{A}_S \quad \int_\Omega \nabla U \cdot \nabla(v-U) \geq \langle f, v-U \rangle \tag{34}$$

defines a unique conforming approximation to u in (31).

We are interested in the error

$$|u-U|_{1;\Omega} := \left(\int_\Omega |\nabla(u-U)|^2 \right)^{\frac{1}{2}}.$$

The departure point of our analysis is the following relationship between $|u-U|_{1;\Omega}$ and the approximation properties of $\mathcal{A}_S$, which already appears in Falk [6] implicitly. We provide a proof for the sake of completeness.

Proposition 12 (Error of U and approximation with $\mathcal{A}_S$)**.** *We have*

$$|u-U|_{1;\Omega} \leq \inf_{v \in \mathcal{A}_S} \Big(|u-v|^2_{1;\Omega} + 2\langle \mu, u-v \rangle \Big)^{\frac{1}{2}}.$$

Proof. Choosing $v = U \in \mathcal{A}$ in (31) and $v \in \mathcal{A}_S$ arbitrary in (34), we derive

$$\begin{aligned} |u-U|^2_{1;\Omega} &= \int_\Omega \nabla u \cdot \nabla(u-U) + \int_\Omega \nabla U \cdot \nabla U - \int_\Omega \nabla U \cdot \nabla u \\ &\leq \langle f, u-U \rangle + \int_\Omega \nabla U \cdot \nabla v + \langle f, U-v \rangle - \int_\Omega \nabla U \cdot \nabla u \\ &\leq \langle f, u-v \rangle + \int_\Omega \nabla U \cdot \nabla(v-u) \\ &= \int_\Omega \nabla(u-U) \cdot \nabla(u-v) + \langle \mu, u-v \rangle, \end{aligned}$$

where in the last step we have used $u - v \in H^1_0(\Omega)$ and the definition (32) of μ. Applying Cauchy-Schwarz on the integral then gives

$$|u-U|^2_{1;\Omega} - |u-v|_{1;\Omega}\, |u-U|_{1;\Omega} - \langle \mu, u-v \rangle \leq 0.$$

Preferring simplicity to sharpness with respect to constants, we deduce

$$|u-U|^2_{1;\Omega} \leq |u-v|^2_{1;\Omega} + 2\langle \mu, u-v \rangle$$

and the proof is finished. □

It is worth noting that Proposition 12 is not a quasi-optimality result like Céa's lemma. In view of the augmentation $\langle \mu, u-v \rangle \geq 0$, it is not clear the right-hand side is bounded by the left-hand side. We shall therefore assess the sharpness of the given bound by other criteria below.

We next provide bounds for the best error of the approximation problem in Proposition 12 with the help of the approximation operator Π^+ of §3. To this end, we fix the solution u of (31) and introduce the local errors

$$e(K) := \inf \big\{ \|u-p\|_{1;K} \mid p \in \mathbb{P}_1(K),\, p \geq \chi \text{ in } K \big\}, \quad K \in \mathcal{M}, \tag{35}$$

and the local dual norms

$$\|\mu\|_{-1;\omega_z} := \sup \big\{ \langle \mu, \varphi \rangle \mid \varphi \in H^1_0(\omega_z),\, |\varphi|_{1;\omega_z} = 1 \big\}, \quad z \in \mathcal{L}_1. \tag{36}$$

Proposition 13 (Localization of augmented gradient error)**.** *In the above notation, we have*

$$\inf_{v \in \mathcal{A}_S} \Big(|u-v|^2_{1;\Omega} + 2\langle \mu, u-v \rangle \Big) \leq C_{d,\gamma_\mathcal{M}} \sum_{K \in \mathcal{M}} e(K) \left(e(K) + \sum_{z \in \mathcal{L}_1(K)} \|\mu\|_{-1;\omega_z} \right).$$

Proof. Let us start with an observation regarding the local best errors (35). As $\chi_{|K} \in \mathbb{P}_1(K)$ and $u - p = (u - \chi) - (p - \chi)$, we see that $p = \chi + P_K^+(u - \chi)$ and

$$e(K) = \left\| u - \chi - P_K^+(u - \chi) \right\|_{1;K} = \inf_{q \in \mathbb{P}_1^+(K)} \|w - q\|_{1;K} \tag{37}$$

for $w := u - \chi \in H^1(\Omega)^+$. Moreover, we have

$$\chi + \Pi^+(u - \chi) \in \mathcal{A}_S \quad \text{and} \quad w - \Pi^+ w \in H_0^1(\Omega),$$

which yields

$$\inf_{v \in \mathcal{A}_S} \left(|u - v|_{1;\Omega}^2 + 2\langle \mu, u - v \rangle \right) \leq \left| w - \Pi^+ w \right|_{1;\Omega}^2 + 2\langle \mu, w - \Pi^+ w \rangle.$$

For the first term on the right-hand side, Corollary 8 and (37) imply

$$\left| w - \Pi^+ w \right|_{1;\Omega}^2 \leq C_{d,\gamma_\mathcal{M}} \sum_{K \in \mathcal{M}} e(K)^2, \tag{38}$$

while the second term has to be localized 'ad hoc'. For the purpose we may use the partition of unity $\sum_{z \in \mathcal{L}_1} \lambda_z = 1$ in Ω and write

$$\langle \mu, w - \Pi^+ w \rangle = \sum_{z \in \mathcal{L}_1} \langle \mu, (w - \Pi^+ w)\lambda_z \rangle. \tag{39}$$

Let $z \in \mathcal{L}_1$ be arbitrary. In view of $0 \leq \lambda_z \leq 1$ and $|\nabla \lambda_z| \leq \rho_K^{-1}$ for all $K \in \mathcal{M}$ containing z as well as Corollary 7, we obtain

$$\begin{aligned} \left| (w - \Pi^+ w)\lambda_z \right|_{1;\omega_z}^2 &\leq 2 \left| (w - \Pi^+ w) \right|_{1;\omega_z}^2 + 2 \sum_{K \ni z} \rho_K^{-2} \left| w - \Pi^+ w \right|_{0;K}^2 \\ &\leq C_{\gamma_\mathcal{M}} \sum_{K \ni z} \left\| w - \Pi^+ w \right\|_{1;K}^2 \leq C_{\gamma_\mathcal{M}} \sum_{K : \omega_K \ni z} e(K)^2, \end{aligned}$$

where K varies in $\mathcal{M}$. Hence, a norm equivalence in some $\mathbb{R}^n$ yields

$$\begin{aligned} \langle \mu, (w - \Pi^+ w)\lambda_z \rangle &\leq C_{\gamma_\mathcal{M}} \|\mu\|_{-1;\omega_z} \left(\sum_{K : \omega_K \ni z} e(K)^2 \right)^{\frac{1}{2}} \\ &\leq C_{d,\gamma_\mathcal{M}} \sum_{K : \omega_K \ni z} \|\mu\|_{-1;\omega_z} e(K), \end{aligned}$$

Inserting this bound into (39) and rearranging terms, we arrive at

$$\langle \mu, w - \Pi^+ w \rangle \leq \sum_{K \in \mathcal{M}} e(K) \left(\sum_{z \in \mathcal{L}_1 \cap \omega_K} \|\mu\|_{-1;\omega_z} \right). \tag{40}$$

Summing the two localizations (38) and (40), we conclude the claimed bound. □

Combining Propositions 12 and 13, we obtain the main result of this section, an a priori error bound in terms of the local best errors and local dual norms of the Lagrange multiplier given in (35) and (36), respectively.

Theorem 14 (Localized and regularity-free a priori bound)**.** *If* u *is the solution of* (31) *and* U *its approximation from* (34)*, then*

$$|u - U|_{1;\Omega} \leq C_{d,\gamma_\mathcal{M}} \left[\sum_{K \in \mathcal{M}} e(K) \left(e(K) + \sum_{z \in \mathcal{L}_1 \cap \omega_K} \|\mu\|_{-1;\omega_z} \right) \right]^{\frac{1}{2}}.$$

The error bound in Theorem 14 can be used, in the setting at hand, instead of, e.g., of Falk [6, Theorem 1]. Confronting with this and the bound in Brezzi et al. [3, Theorem 2.1], we note:

- The best errors $e(K)$, $K \in \mathcal{M}$, and the dual norms $\|\mu\|_{-1,\omega_z}$, $z \in \mathcal{L}_1$, are local and do not involve regularity beyond $u \in H^1_0(\Omega)$ and $\mu \in H^{-1}(\Omega)$, which are required or follow from the problem formulation (31).
- The discrete solution U does not appear on the right-hand side.
- In the special case $\chi = -\infty$ and thus $\mu = 0$ which corresponds to the Poisson problem, the bound and the local Poincaré inequality $\left\|v - \fint_K v\right\|_{0;K} \leq \pi^{-1}\operatorname{diam}(K)\,|v|_{1;K}$ imply
$$|u-U|_{1;\Omega} \leq C_{d,\gamma_\mathcal{M}} \left[\sum_{K\in\mathcal{M}} \inf_{p\in\mathbb{P}_1(K)} |u-p|^2_{1;K}\right]^{\frac{1}{2}}$$
i.e. that the error is $\mathbb{P}_1$-quasi-optimal. This reproduces the application of (1) after Céa's lemma.
- Introducing the two-layer neighborhood
$$N_{\mathcal{M};+} := \bigcup_{K\in\mathcal{M}: u\neq\chi \text{ on } K} \widetilde{\omega}_K \quad \text{with} \quad \widetilde{\omega}_K := \bigcup_{z\in\mathcal{L}_1\cap\omega_K} \omega_z$$
of the non-coincidence set $\{u > \chi\}$, we have the following implication: if $\delta f \in H^{-1}(\Omega)$ such that $\delta f \leq -\mu$ and $\operatorname{supp}\delta f \cap N_{\mathcal{M};+} = \emptyset$, then the bound yields the same value for the forces f and $f + \delta f$.

The abstract bound in [6, Theorem 1] assumes that the Lagrange multiplier $\mu \in L^2(\Omega)$ is a function in order to derive first order error decay; cf. comment (iv) after Theorem 23.1 in [5]. Let us check that the bound in Theorem 14 also implies this decay rate. To this end, we derive an error bound involving $h_K := \operatorname{diam}(K)$ and
$$\left\|hD^2u\right\|_{0;K} := \left(\sum_{|\alpha|=2} \|\partial^\alpha u\|^2_{0;K}\right)^{\frac{1}{2}},$$
where $K \in \mathcal{M}$ and we used multi-index notation.

Corollary 15 (Localized a priori bounds with regularity)**.** *Let u and U be as in Theorem 14. If $\mu \in L^2(\Omega)$, then*
$$|u-U|_{1;\Omega} \leq C_{d;\gamma_\mathcal{M}} \left[\sum_{K\in\mathcal{M}} e(K)\left(e(K) + h_K\|\mu\|_{0;\widetilde{\omega}_K}\right)\right]^{\frac{1}{2}}.$$
Furthermore, if in addition $d \in \{2,3\}$, $u \in H^2(K)$ for all $K \in \mathcal{M}$, then
$$|u-U|_{1;\Omega} \leq C_{d;\gamma_\mathcal{M}} \left[\sum_{K\in\mathcal{M}} h_K^2 \left\|D^2u\right\|_{0;K}\left(\left\|D^2u\right\|_{0;K} + \|\mu\|_{0;\widetilde{\omega}_K}\right)\right]^{\frac{1}{2}}.$$

Proof. We only have to provide bounds for the local quantities in Theorem 14 and start with the local dual norms. Given $z \in \mathcal{L}_1$ and $\varphi \in H^1_0(\omega_z)$, the Poincaré-Friedrichs inequality
$$\|\varphi\|_{0;\omega_z} \leq \operatorname{diam}(\omega_z)\,|\varphi|_{1;\omega_z}$$
yields
$$\langle\mu,\varphi\rangle = \int_\Omega \mu\varphi = \int_{\omega_z} \mu\varphi \leq \|\mu\|_{0;\omega_z}\|\varphi\|_{0;\omega_z} \leq \operatorname{diam}(\omega_z)\,|\varphi|_{1;\omega_z}.$$
Therefore
$$\|\mu\|_{-1;\omega_z} \leq \operatorname{diam}(\omega_z)\,\|\mu\|_{0;\omega_z}$$
and the first bound is proved since $\operatorname{diam}(\omega_z) \leq C_{d,\gamma_\mathcal{M}} h_K$ whenever $z \in \omega_K$.

Next, we consider a local best error $e(K)$, $K \in \mathcal{M}$. Since $d \in \{2,3\}$, the Lagrange interpolation $I_K u \in \mathbb{P}_1(K)$ of $u \in H^2(K)$ is well-defined and we have $I_K u \geq \chi$ on K, whence

$$e(K) \leq \|u - I_K u\|_{1;K} \leq C_{d,\gamma_{\mathcal{M}}} h_K \left\|D^2 u\right\|_{0;K};$$

cf. [5, Theorem 16.1]. With this, the second bound follows from the first one. □

Comparing the bounds of Corollary 15 with previous asymptotic error bounds, the following comments are in order.

- If the exact solution u happens to be in $\mathcal{A}_S$, both bounds vanish thanks to the fact that the approximation or the regularity of u enters only in a piecewise manner.
- Assume shape regular and uniform refinement such that $\gamma_{\mathcal{M}} \approx 1$ and $h := \max_{K\in\mathcal{M}} h_K \approx \min_{K\in\mathcal{M}} h_K$ tends to 0. Then the second bound and $\sum_{K\in\mathcal{M}} \|\mu\|^2_{0;\widetilde{\omega}_K} \lesssim \|\mu\|^2_{0;\Omega}$ yield the maximal decay rate
$$|u - U|_{1;\Omega} = O(h)$$
when approximating in H^1 with piecewise affine functions.
- Assume shape regular and uniform refinement as in the preceding item and, additionally, that the Lagrange multiplier is bounded, $\mu \geq -\mu_0$ with $\mu_0 \in \mathbb{R}$, and that the free boundary is a hyper surface. Then there are $O(h^{-(d-1)})$ elements $K \in \mathcal{M}$ with $K \subset \{u \neq \chi\}$ and $\widetilde{\omega}_K \not\subseteq \{u > \chi\}$. Since $K \subseteq \{u = \chi\}$ entails $e(K) = 0$ and $\widetilde{\omega}_K \subset \{u > \chi\}$ entails $\mu = 0$ on $\widetilde{\omega}_K$, we infer
$$\begin{aligned}\sum_{K\in\mathcal{M}} & e(K) h_K \|\mu\|_{0;\widetilde{\omega}_K} \\ &\leq \frac{1}{2}\sum_{K\in\mathcal{M}} e(K)^2 + \frac{1}{2}\sum_{K\subset\{u\neq\chi\},\widetilde{\omega}_K\not\subseteq\{u>\chi\}} h_K^2 \|\mu\|^2_{0;\widetilde{\omega}_K} \\ &\lesssim \sum_{K\in\mathcal{M}} e(K)^2 + \mu_0^2 h^3.\end{aligned}$$
Inserting this inequality into the first bound, we arrive at
$$|u - U|_{1;\Omega} \lesssim \left(\sum_{K\in\mathcal{M}} e(K)^2\right)^{\frac{1}{2}} + \mu_0 h^{\frac{3}{2}}.$$
As the first term has at most the rate h, we may say that the error $|u - U|_{1;\Omega}$ is asymptotically $\mathbb{P}_1^+$-quasi-optimal for shape regular and uniform refinement.

In summary, the presented approach relying on Corollary 7 maintains the advantages of Falk's method, offering additional locality and covering any regularity.

Acknowledgment. Support by the GNCS of the INdAM is gratefully acknowledged.

References

[1] P. Binev and R. DeVore, *Fast computation in adaptive tree approximation*, Numer. Math., 97 (2004), pp. 193–217.

[2] 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.

[3] F. Brezzi, W. W. Hager, and P.-A. Raviart, *Error estimates for the finite element solution of variational inequalities*, Numer. Math., 28 (1977), pp. 431–443.

[4] P. G. Ciarlet, *The finite element method for elliptic problems*, vol. 4 of Studies in Mathematics and its Applications, North–Holland, Amsterdam, 1978.

[5] ———, *Basic error estimates for elliptic problems*, in Handbook of Numerical Analysis, Vol. II, P. G. Ciarlet and J.-L. Lions, eds., North-Holland, 1991, pp. 17–352.

[6] R. S. Falk, *Error estimates for the approximation of a class of variational inequalities*, Math. Comput., 28 (1974), pp. 963–971.

[7] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, Classics in Mathematics, Springer-Verlag, Berlin, 2001. Reprint of the 1998 edition.

[8] R. H. Nochetto and L. B. Wahlbin, *Positivity preserving finite element approximation*, Math. Comp., 71 (2002), pp. 1405–1419.

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

[10] A. Veeser, *Approximating gradients with continuous piecewise polynomial functions*, Found. Comput. Math., 16 (2016), pp. 723–750.

[11] A. Veeser and R. Verfürth, *Explicit upper bounds for dual norms of residuals*, SIAM J. Numer. Anal., 47 (2009), pp. 2387–2405.

Andreas Veeser, Dipartimento di Matematica, Università degli Studi di Milano, Via Saldini 50, 20131 Milano, Italia

E-mail address: andreas.veeser@unimi.it

URL: www.mat.unimi.it/users/veeser/