



**HAL**  
open science

# Finite element method with local damage on the mesh

Michel Duprez, Vanessa Lleras, Alexei Lozinski

► **To cite this version:**

Michel Duprez, Vanessa Lleras, Alexei Lozinski. Finite element method with local damage on the mesh. 2018. hal-01858033v1

**HAL Id: hal-01858033**

**<https://hal.science/hal-01858033v1>**

Preprint submitted on 18 Aug 2018 (v1), last revised 14 Mar 2019 (v2)

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

# Finite element method with local damage on the mesh <sup>\*</sup>

Michel Duprez<sup>†</sup>, Vanessa Lleras<sup>‡</sup> and Alexei Lozinski<sup>§</sup>

August 18, 2018

## Abstract

We consider the finite element method on locally damaged meshes allowing for some distorted cells which are isolated from one another. In the case of the Poisson equation and piecewise linear Lagrange finite elements, we show that the usual *a priori* error estimates remain valid on such meshes. We also propose an alternative finite element scheme which is optimally convergent and, moreover, well conditioned, *i.e.* its conditioning is of the same order as that of a standard finite element method on a regular mesh of comparable size.

## 1 Introduction

We are interested in the finite element method on meshes containing some isolated degenerated cells. The meshes of this type can be encountered in bio-mechanical applications where the objects with very complicated geometry (as a human face) should be meshed, and the mesh generators or mesh morphing techniques are not always able to satisfy the usual regularity constraints (see *e.g.* [8, p.3]). Our work is a preliminary study in which we propose a suitable finite element approximation in such situations without requiring to reconstruct a high quality mesh everywhere. We restrict ourselves to the simplest model: the Poisson equation with Dirichlet boundary conditions

$$\begin{cases} -\Delta u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega \end{cases} \quad (1)$$

where  $\Omega$  is a bounded polygonal (resp. polyhedral) domain in  $\mathbb{R}^n$ ,  $n = 2$  (resp.  $n = 3$ ),  $\partial\Omega$  is its boundary, and  $f \in L^2(\Omega)$  is a given function. We only consider the standard piecewise linear continuous finite elements on a simplicial mesh without hanging nodes. The formal (quite usual) definitions of the exact and approximated solutions to (1) in the appropriate functional spaces are given in the beginning of Section 2.

The first goal of the present work is to highlight that we can recover the optimal convergence of the finite element method even if the mesh contains several isolated almost degenerated simplexes. More precisely, we shall assume that the majority of the simplexes in the mesh are regular in the usual Ciarlet sense [11] but there are some distorted simplexes that are typically adjacent to regular mesh cells and well separated from one another by layers of regular cells. The formal assumptions will be given in the beginning of Section 2. To prove the optimal convergence of the standard finite element method, we shall construct a modification of the nodal interpolation operator replacing the standard interpolating

---

<sup>\*</sup>This work has been carried out in the framework of Archimède Labex (ANR-11-LABX-0033) and of the A\*MIDEX project (ANR-11-IDEX-0001-02), funded by the “Investissements d’Avenir” French Government programme managed by the French National Research Agency (ANR). The authors acknowledge the support of the ANR project CroCo ANR-16-CE33-0008 and that of Région Bourgogne Franche-Comté “Convention Région 2015C-4991. Modèles mathématiques et méthodes numériques pour l’élasticité non-linéaire”.

<sup>†</sup>Aix Marseille Université, CNRS, Centrale Marseille, I2M, UMR 7373, 13453 Marseille, France. e-mail: [mduprez@math.cnrs.fr](mailto:mduprez@math.cnrs.fr)

<sup>‡</sup>IMAG, Univ Montpellier, CNRS, Montpellier, France. [vanessa.llerass@umontpellier.fr](mailto:vanessa.llerass@umontpellier.fr)

<sup>§</sup>Laboratoire de Mathématiques de Besançon, UMR CNRS 6623, Université Bourgogne Franche-Comté, 16, route de Gray, 25030 Besançon Cedex, France. e-mail: [alexei.lozinski@univ-fcomte.fr](mailto:alexei.lozinski@univ-fcomte.fr)

polynomial on a degenerated cell by another one obtained by averaging the interpolated function on a patch of cells surrounding the degenerated one.

Although the standard finite element method turns out optimally convergent on the locally damaged meshes, as outlined above, it can suffer from bad conditioning of the stiffness matrix. Indeed, the gradient operator can have an arbitrary large norm on the space of piecewise polynomial functions on a mesh containing very elongated cells even if all the cells are of approximately the same diameter  $h$ . The same issue of bad conditioning can be found in recent finite element methods on geometrically unfitted meshes such as CutFEM, cf. the review in [10]. The mesh is allowed to be cut by the domain boundary in this approach giving rise to eventually very narrow computational cells, and consequently to very ill-conditioned stiffness matrices. The workaround consists in introducing some stabilization terms which come in (at least) two forms: (i) an augmented Lagrangian type approach involving the polynomial extension from “good” (uncut) to “bad” (cut) cells [13]; (ii) *ghost penalty* terms on the facets [9] of the cut cells which reduce the jumps between the gradients of the finite element solution between the “bad” and “good” cells. In both cases, the goal is to make the finite element solution on a “bad” cell to be aligned with its counterpart on a neighboring “good” cell. In the present paper, we inspire ourselves from both approaches described above and propose an alternative finite element discretization in which the approximated solution on the degenerated cells is made to be aligned with that on neighboring regular cells. We are able to prove that such a scheme is optimally convergent and well conditioned, *i.e.* its conditioning is of the same order as that of a standard finite element method on a usual regular mesh of comparable size, provided the number of degenerated cells remains uniformly bounded.

The present article is a contribution to the already rich literature studying the influence of the the mesh cell geometry on the convergence of finite element approximations. The optimal  $H^1$ -convergence has been proved in [21] for second order elliptic equation and in [20] for linear elasticity equations under the *minimum angle condition* in 2D: there exists  $\alpha_0 \in (0, \pi)$  such that for any mesh cell  $K$ ,

$$0 < \alpha_0 \leq \alpha_K, \tag{2}$$

where  $\alpha_K$  is the minimum angle of  $K$ . In [6, 5], this condition was generalized to the higher dimensions. If we denote by  $h_K$  the diameter of  $K$  and  $\rho_K$  the diameter of the largest ball contained in  $K$ , then (2) is equivalent to already mentioned *Ciarlet condition* [11]: there exists  $c_0$  such that for all mesh cells  $K$

$$h_K/\rho_K \leq c_0. \tag{3}$$

The conditions above were further relaxed in several ways. Three groups (see [3, 4, 14]) have proposed independently in 1976 a weaker assumption called the *maximum angle condition*: there exists  $\beta_0 \in (0, \pi)$  such that for any mesh cell  $K$

$$\beta_K \leq \beta_0 < \pi, \tag{4}$$

where  $\beta_K$  is the maximum angle of  $K$ . The first condition (2) implies the second (4). The second condition was generalized for higher dimensions in [15, 18].

Furthermore, it is shown in [12] that even the maximum angle condition may be not necessary. More precisely, if a degenerated triangulation is included in a non-degenerated one, then optimal convergence rates. The convergence on appropriate anisotropic meshes is studied in [2]. A sufficient condition for convergence (not necessarily of optimal order) was derived in [16] under the name of the *circumradius condition*:  $\max_K R_K \rightarrow 0$  as  $h \rightarrow 0$  where  $R_K$  is the circumradius of the mesh cell  $K$ . Both the maximum angle and circumradius conditions for  $O(h^\alpha)$  convergence are generalized in [17]. It is proved that the triangulations can contain many elements violating these conditions as long as their maximum angle vertexes are sufficiently small size. However, one cannot hope for an optimal convergence on completely arbitrary meshes: an example of a heavily distorted mesh family stemming from [3] has been recently analyzed in [19] showing rigorously that the finite element method may fail to converge at all. The present paper propose yet another choice of assumptions on the mesh in the spirit of, but different from [12], guaranteeing the optimal convergence.

The rest of the paper is organized as follows: in Section 2, we prove that one can allow degenerated cells, which violate Condition (2) or (4), if they are isolated in some sense. We shall establish in Subsection 2.1 the optimal  $L^2$ - and  $H^1$ -convergence of the standard finite elements method on such meshes. We also recall, in Subsection 2.2, the well known fact that the presence of degenerated cells may induce a large conditioning number of the stiffness matrix. We then propose in Section 3 a modified finite element method that preserves the optimal convergence while ensuring a good conditioning. We conclude with some numerical illustrations in Section 4.

## 2 Approximation by linear finite elements under local mesh damage assumption

Let us first recall the notions of the weak and approximated solutions to System (1). We call a *weak solution* in  $V := H_0^1(\Omega)$  to System (1) a function  $u \in V$  such that

$$a(u, v) = l(v) \text{ for all } v \in V, \quad (5)$$

where the bilinear form  $a$  and the linear form  $l$  are defined for all  $u, v \in V$  by

$$a(u, v) := \int_{\Omega} \nabla u \cdot \nabla v \, dx \quad \text{and} \quad l(v) := \int_{\Omega} f v \, dx.$$

It is well known that System (1) admits a unique weak solution thanks to Lax-Milgram lemma.

Consider now a simplicial mesh  $\mathcal{T}_h$  on  $\Omega$  without hanging nodes. This means that  $\bar{\Omega} = \cup_{K \in \mathcal{T}_h} K$  with each mesh cell  $K \in \mathcal{T}_h$  being a simplex (triangle in 2D, tetrahedron in 3D) and every two mesh cells  $K_1, K_2 \in \mathcal{T}_h$  being either disjoint or sharing a vertex, an edge, or a face (in 3D). We recall that  $\rho_K$  denotes the diameter of the largest ball contained in a mesh cell  $K$ . Moreover,  $h_{\omega}$  will denote the diameter of any bounded domain  $\omega$  and we set  $h = \max_{K \in \mathcal{T}_h} h_K$ . As mentioned in the Introduction, we will assume that the cells of mesh  $\mathcal{T}_h$  satisfy Ciarlet Condition (3) up to some isolated cells.

**Assumption 1.** *Let  $c_0 > 0$  and  $K_1^{deg}, \dots, K_I^{deg}$  be the degenerated cells violating Ciarlet Condition (3), i.e. for  $i \in \{1, \dots, I\}$*

$$K_i^{deg} \in \mathcal{T}_h \text{ and } h_{K_i^{deg}} / \rho_{K_i^{deg}} > c_0.$$

*Each  $K_i^{deg}$  is included in a patch  $\mathcal{P}_i$ , which is a union of mesh cells, star-shaped with respect to a ball of diameter  $\rho_{\mathcal{P}_i}$  such that*

$$h_{\mathcal{P}_i} / \rho_{\mathcal{P}_i} \leq c_1.$$

*We denote by  $\tilde{\mathcal{P}}_i \supset \mathcal{P}_i$  the larger patch composed of mesh cells sharing at least a vertex with  $\mathcal{P}_i$ . Then*

- *The patches  $\tilde{\mathcal{P}}_i$  are mutually disjoint, i.e.  $\tilde{\mathcal{P}}_i$  and  $\tilde{\mathcal{P}}_j$  have no common cells for  $i \neq j$ .*
- *The number of cells in each  $\tilde{\mathcal{P}}_i$  is bounded by a constant  $M$ .*

*The intersection of boundaries  $\partial\mathcal{P}_i$  and  $\partial\Omega$  is either empty, or is reduced to a point, or is a subset of one side of the polygon/polyhedron  $\Omega$  containing an  $(n-1)$ -dimensional ball of radius  $\geq c_2 h_{\mathcal{P}_i}$ .*

**Notational warning.** In what follows, the letter  $C$  will stand for constants which depend only on the generalized mesh regularity in the sense of Assumption 1 (unless stated otherwise). This means that  $C$  can depend on  $c_0$ ,  $c_1$ ,  $c_2$ , and  $M$ , but otherwise independent from the choice of mesh  $\mathcal{T}_h$ . As usual, the value of  $C$  can change from one line to another.

An example of patches  $\mathcal{P}_i$  and  $\tilde{\mathcal{P}}_i$  is given in Fig. 1. We illustrate there a typical situation of a degenerated triangle  $K_i^{deg}$  (dashed in red) adjacent to a regular triangle  $K_i^{nd}$  (dashed in grey). The patch  $\mathcal{P}_i$  is then formed of these two triangles  $K_i^{deg}$  and  $K_i^{nd}$ . It is obviously star-shaped with respect to a ball (for example, the largest ball inscribed in  $K_i^{nd}$ ). Its chunky parameter  $h_{\mathcal{P}_i} / \rho_{\mathcal{P}_i}$  is close to that of

surrounding regular triangles. Note, however, that Assumption 1 allows for more general configurations, for example, a patch can contain several degenerated cells.

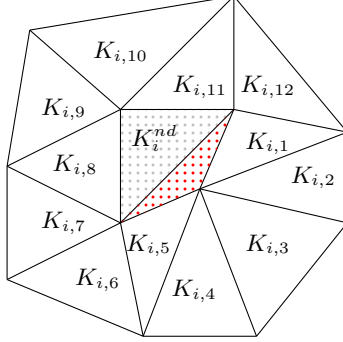


Figure 1: Example of configuration: patch  $\mathcal{P}_i$  (dashed), patch  $\tilde{\mathcal{P}}_i$  (all the cells), non-degenerated cell  $K_i^{nd}$  (gray) of the patch  $\mathcal{P}_i$ .

We now set the finite element space on mesh  $\mathcal{T}_h$  and the finite element approximation to System (1). Let

$$V_h := \{v_h \in V : v_h|_K \in \mathbb{P}_1(K) \forall K \in \mathcal{T}_h\},$$

where  $\mathbb{P}_1(K)$  is the space of polynomials of degree  $\leq 1$  on cell  $K$ . Consider the following finite element approximation to System (5): find  $u_h \in V_h$  such that:

$$a(u_h, v_h) = l(v_h) \text{ for all } v_h \in V_h. \quad (6)$$

## 2.1 A priori error estimate

In what follows,  $|\cdot|_{i,A}$  and  $\|\cdot\|_{i,A}$  denote the semi-norm and the norm associated to  $H^i(A)$ .

**Theorem 2.** *Let  $u \in V$  and  $u_h \in V_h$  be the solutions to System (5) and System (6), respectively. Then, under Assumption 1,*

$$|u - u_h|_{1,\Omega} \leq Ch|u|_{2,\Omega}. \quad (7)$$

Moreover, if  $\Omega$  is convex,

$$\|u - u_h\|_{0,\Omega} \leq Ch^2|u|_{2,\Omega}. \quad (8)$$

The proof of this theorem is completely standard (*cf.* [11, 7]) provided one has constructed an interpolant to  $V_h$  satisfying the optimal error estimates. We thus go directly to the construction of such an interpolation operator which we shall call  $\tilde{\mathcal{I}}_h$  and properly introduce in Definition 1. The necessary properties of this operator will be established in the Proposition 1. We start with some technical lemmas.

**Lemma 1.** *Under Assumption 1, for any  $v \in H^2(\Omega) \cap H_0^1(\Omega)$  on any patch  $\mathcal{P}_i$  there exists a polynomial  $Q_h^i(v)$  on  $\mathcal{P}_i$  of degree  $\leq 1$  vanishing on  $\partial\mathcal{P}_i \cap \partial\Omega$  such that*

$$|v - Q_h^i(v)|_{1,\mathcal{P}_i} \leq Ch_{\mathcal{P}_i}|v|_{2,\mathcal{P}_i}, \quad \|v - Q_h^i(v)\|_{0,\mathcal{P}_i} \leq Ch_{\mathcal{P}_i}^2|v|_{2,\mathcal{P}_i}, \quad \|v - Q_h^i(v)\|_{L^\infty(\mathcal{P}_i)} \leq Ch_{\mathcal{P}_i}^{2-n/2}|v|_{2,\mathcal{P}_i}. \quad (9)$$

*Proof.* We consider first the case of the patch  $\mathcal{P}_i$  lying completely inside  $\Omega$ . We take then  $Q_h^i(v)$  on  $\mathcal{P}_i$  as the Taylor polynomial  $Q^2v$ , *cf.* Definition (4.1.3) from [7], averaged over the ball of diameter  $\rho_{\mathcal{P}_i}$  mentioned in Assumption 1. The estimates (9) for  $Q_h^i(v) = Q^2v$  are thus given by Proposition (4.3.2) and Bramble-Hilbert Lemma (4.3.8) from [7].

We now turn to the case when the boundary  $\partial\mathcal{P}_i$  intersects  $\partial\Omega$  in only one point, say  $x$ . The polynomial  $Q_h^i(v)$  should vanish at  $x$  so that we correct  $Q^2v$  by subtracting from it its value at point. We set thus  $Q_h^i(v) = Q^2v - c_h$  where  $c_h = Q^2v(x)$ . Since  $v(x) = 0$ , we have by the above mentioned properties of  $Q^2v$

$$|c_h| = |Q^2v(x) - v(x)| \leq \|Q^2v - v\|_{L^\infty(\mathcal{P}_i)} \leq Ch^{2-n/2}|v|_{2,\mathcal{P}_i}$$

which entails

$$\|Q_h^i v - v\|_{0,\mathcal{P}_i} \leq \|Q^2v - v\|_{0,\mathcal{P}_i} + |c_h| |\mathcal{P}_i|^{1/2} \leq Ch^2|v|_{2,\mathcal{P}_i}$$

and

$$\|Q_h^i v - v\|_{L^\infty(\mathcal{P}_i)} \leq \|Q^2v - v\|_{L^\infty(\mathcal{P}_i)} + |c_h| \leq Ch|v|_{2,\mathcal{P}_i}.$$

The  $H^1$  semi-norm of the error is not affected by the constant  $c_h$ , so that the announced estimate for  $|Q_h^i v - v|_{1,\mathcal{P}_i}$  is also valid.

The last case to consider is when  $\partial\mathcal{P}_i$  has a non-empty intersection with a side, say  $\Gamma$ , of  $\partial\Omega$ , which is not reduced to one point. We recall that  $\partial\mathcal{P}_i \cap \Gamma$  is assumed then to contain a ball of radius of order  $h_{\mathcal{P}_i}$ . We introduce the polynomial  $c_h$  of degree  $\leq 1$  that coincides with  $Q^2v$  on  $\partial\mathcal{P}_i \cap \Gamma$  and does not vary in the direction perpendicular to  $\Gamma$ . Setting  $Q_h^i(v) = Q^2v - c_h$  we see immediately that  $Q_h^i(v)$  vanishes on  $\Gamma$ . Moreover, thanks to our geometrical assumptions and the fact that  $v$  vanishes on  $\partial\mathcal{P}_i \cap \Gamma$ ,

$$\|c_h\|_{L^\infty(\mathcal{P}_i)} \leq C \|c_h\|_{L^\infty(\partial\mathcal{P}_i \cap \Gamma)} = C \|Q^2v - v\|_{L^\infty(\partial\mathcal{P}_i \cap \Gamma)} \leq Ch_{\mathcal{P}_i}^{2-n/2}|v|_{2,\mathcal{P}_i}.$$

We can thus prove the desired estimates for  $\|Q_h^i v - v\|_{0,\mathcal{P}_i}$  and  $\|Q_h^i v - v\|_{L^\infty(\mathcal{P}_i)}$  as in the previous case. Finally, by an inverse inequality,

$$\|\nabla c_h\|_{L^\infty(\mathcal{P}_i)} \leq \frac{C}{h_{\mathcal{P}_i}} \|c_h\|_{L^\infty(\partial\mathcal{P}_i \cap \Gamma)} \leq Ch_{\mathcal{P}_i}^{1-n/2}|v|_{2,\mathcal{P}_i}$$

so that

$$|Q_h^i v - v|_{1,\mathcal{P}_i} \leq |Q^2v - v|_{1,\mathcal{P}_i} + \|\nabla c_h\|_{L^\infty(\mathcal{P}_i)} |\mathcal{P}_i|^{1/2} \leq Ch|v|_{2,\mathcal{P}_i}.$$

□

We also recall the usual interpolation error estimates on regular cells for the standard Lagrange interpolation operator  $\mathcal{I}_h$  to the space of piecewise linear functions, cf. [11, 7].

**Lemma 2.** *Under Assumption 1, we have on each mesh cell  $K \in \mathcal{T}_h$  outside of patches  $\mathcal{P}_i$*

$$|v - \mathcal{I}_h(v)|_{1,K} \leq Ch|v|_{2,K}, \quad \|v - \mathcal{I}_h(v)\|_{0,K} \leq Ch^2|v|_{2,K}, \quad \|v - \mathcal{I}_h(v)\|_{L^\infty(K)} \leq Ch^{2-n/2}|v|_{2,K}, \quad (10)$$

for any  $v \in H^2(K)$ .

**Definition 1.** For all  $v \in H^2(\Omega) \cap H_0^1(\Omega)$ , let  $\tilde{\mathcal{I}}_h(v)$  be the function in  $V_h$  that coincides with  $Q_h^i(v)$  from Lemma 1 on each patch  $\mathcal{P}_i$ , and with the standard Lagrange interpolation  $\mathcal{I}_h v$  on all the cells  $K \in \mathcal{T}_h$  out of the extended patches  $\tilde{\mathcal{P}}_i$ , i.e.  $\tilde{\mathcal{I}}_h(v)(x) = v(x)$  at all the mesh nodes  $x \in \bar{\Omega} \setminus \cup_{i \in \{1, \dots, I\}} \tilde{\mathcal{P}}_i$ .

Note that  $\tilde{\mathcal{I}}_h(v)$  is uniquely defined also on the mesh cells from  $\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i$ ,  $i = 1, \dots, I$  although they are not explicitly mentioned above. Indeed, all the vertices of such cells are shared either with a patch  $\mathcal{P}_i$  or with a regular cell from  $\bar{\Omega} \setminus \cup_{i \in \{1, \dots, I\}} \tilde{\mathcal{P}}_i$ . Since the values of  $\tilde{\mathcal{I}}_h(v)$  are given at all these nodes by the definition above, the piecewise linear function  $\tilde{\mathcal{I}}_h(v)$  is well defined everywhere.

We now prove the global interpolation estimates for the interpolation operator  $\tilde{\mathcal{I}}_h$ .

**Proposition 1.** *Under Assumption 1, we have for all  $v \in H^2(\Omega) \cup H_0^1(\Omega)$*

$$|v - \tilde{\mathcal{I}}_h(v)|_{1,\Omega} \leq Ch|v|_{2,\Omega}, \quad \|v - \tilde{\mathcal{I}}_h(v)\|_{0,\Omega} \leq Ch^2|v|_{2,\Omega}.$$

*Proof.* The contributions to the interpolation errors on the patches  $\mathcal{P}_i$  and on the mesh cells outside the patches  $\tilde{\mathcal{P}}_i$  (where the interpolators  $\tilde{\mathcal{I}}_h$  and  $\mathcal{I}_h$  coincide) are already covered by Lemmas 1 and 2. It remains to bound the error on mesh cells in  $\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i$ .

Let  $K \in \mathcal{T}_h$  and  $K \subset \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i$ . By the triangle inequality and Lemma 2,

$$|v - \tilde{\mathcal{I}}_h(v)|_{1,K} \leq |v - \mathcal{I}_h(v)|_{1,K} + |\mathcal{I}_h(v) - \tilde{\mathcal{I}}_h(v)|_{1,K} \leq Ch|v|_{2,K} + |r_h|_{1,K},$$

where we have denoted  $r_h := \mathcal{I}_h(v) - \tilde{\mathcal{I}}_h(v)$ . By a homogeneity argument and the equivalence of norms on finite dimensional space, we see easily

$$|r_h|_{1,K} \leq Ch^{n/2-1} \|r_h\|_{L^\infty(K)}.$$

Recalling that  $r_h$  is a polynomial of degree  $\leq 1$  vanishing at the vertices of  $K$  on  $\partial\tilde{\mathcal{P}}_i$ , the other vertices belonging to  $\partial\mathcal{P}_i$ , we conclude

$$\|r_h\|_{L^\infty(K)} \leq \|r_h\|_{L^\infty(\partial K \cap \partial\mathcal{P}_i)} \leq \|v - \mathcal{I}_h(v)\|_{L^\infty(K)} + \|v - \tilde{\mathcal{I}}_h(v)\|_{L^\infty(\mathcal{P}_i)} \leq Ch^{2-n/2} (|v|_{2,K} + |v|_{2,\mathcal{P}_i}).$$

Putting the estimates above together yields

$$|v - \tilde{\mathcal{I}}_h(v)|_{1,K} \leq Ch(|v|_{2,K} + |v|_{2,\mathcal{P}_i}). \quad (11)$$

Similarly,

$$\|v - \tilde{\mathcal{I}}_h(v)\|_{0,K} \leq Ch^2(|v|_{2,K} + |v|_{2,\mathcal{P}_i}). \quad (12)$$

Taking the square on both sides of (11) and (12), summing them over all the mesh cells  $K \subset \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i$ ,  $i = 1, \dots, I$  (recall that the number of such cells on each patch is bounded by a predefined constant  $M$ ), adding the estimates from lemma 1 on the patches  $\mathcal{P}_i$  and those of Lemma 2 on the mesh cells outside the patches  $\tilde{\mathcal{P}}_i$  gives the desired result.  $\square$

## 2.2 Poor conditioning of the system matrix

In this section, we shall recall the well known fact that the presence of degenerated cells can induce an arbitrary large conditioning number of the associated finite element matrix. In the following proposition, we consider a particular example of a mesh satisfying Assumption 1 and give an estimator for the conditioning number. This result should be contrasted with the “normal” conditioning number of order  $1/h^2$  on a quasi-uniform mesh.

**Proposition 2.** *Suppose that the mesh  $\mathcal{T}_h$  satisfies Assumption 1 and contains a degenerate cell  $K^{deg}$  such that*

$$\rho_{K^{deg}} = \varepsilon, \quad h_{K^{deg}} \geq C_1 h. \quad (13)$$

*Then the conditioning number  $\kappa(\mathbf{A}) := \|\mathbf{A}\|_2 \|\mathbf{A}^{-1}\|_2$  of the matrix  $\mathbf{A}$  associated to the bilinear form  $a$  in  $V_h$  satisfies*

$$\kappa(\mathbf{A}) \geq \frac{C}{h\varepsilon}$$

*for sufficiently small  $h$ , with  $C$  depending only on  $C_1$  and  $\Omega$ . Here,  $\|\cdot\|_2$  stands for the matrix norm associated to the vector 2-norm.*

*Proof.* Denote by  $N$  the dimension of  $V_h$ . Consider  $\phi_h$  the basis function of  $V_h$  equal to 1 at the node of  $K^{deg}$  opposite to the largest edge (face) of  $K^{deg}$ , vanishing at all the other nodes, and  $\phi \in \mathbb{R}^N$  the vector representing  $\phi_h$  in the basis of hat functions. Then, denoting by  $|\cdot|_2$  the vector 2-norm on  $\mathbb{R}^N$  and by  $(\cdot, \cdot)$  the associated inner product,

$$\|\mathbf{A}\|_2 = \sup_{\mathbf{u} \in \mathbb{R}^N} \frac{(\mathbf{A}\mathbf{u}, \mathbf{u})}{|\mathbf{u}|_2^2} \geq (\mathbf{A}\phi, \phi) = a(\phi_h, \phi_h) = |\phi_h|_{1,\Omega}^2 \geq |\phi_h|_{1,K^{deg}}^2.$$

By (13), the gradient of  $\phi_h$  is of order  $1/\varepsilon$  on  $K^{deg}$ , and the area of  $K^{deg}$  is of order  $\varepsilon h^{n-1}$ . Thus,

$$\|\mathbf{A}\|_2 \geq |\phi_h|_{1,K^{deg}}^2 \geq C \frac{h^{n-1}}{\varepsilon}.$$

Now take any  $\psi \in H^2(\Omega) \cap H_0^1(\Omega)$ ,  $\psi \neq 0$  and let  $\boldsymbol{\psi} \in \mathbb{R}^N$  be the vector associated to  $\tilde{\mathcal{I}}_h \psi$ . Then

$$\|\mathbf{A}^{-1}\|_2 = \sup_{\mathbf{u} \in \mathbb{R}^N} \frac{|\mathbf{u}|_2^2}{(\mathbf{A}\mathbf{u}, \mathbf{u})} \geq \frac{|\boldsymbol{\psi}|_2^2}{(\mathbf{A}\boldsymbol{\psi}, \boldsymbol{\psi})} = \frac{|\boldsymbol{\psi}|_2^2}{a(\tilde{\mathcal{I}}_h \psi, \tilde{\mathcal{I}}_h \psi)} \geq \frac{C}{h^n} \frac{\|\tilde{\mathcal{I}}_h \psi\|_{0,\Omega}^2}{|\tilde{\mathcal{I}}_h \psi|_{1,\Omega}^2}.$$

We have used here the bound

$$\|v_h\|_{0,\Omega}^2 \leq Ch^n |\mathbf{v}|^2$$

valid for any  $v_h \in V_h$  and the corresponding vector  $\mathbf{v}$  since all the mesh cells are of diameter  $\leq h$ . Proposition 1 implies

$$\begin{cases} \|\tilde{\mathcal{I}}_h \psi\|_{0,\Omega} \geq \|\psi\|_{0,\Omega} - Ch^2 |\psi|_{2,\Omega} \geq C \|\psi\|_{0,\Omega}, \\ |\tilde{\mathcal{I}}_h \psi|_{1,\Omega} \leq |\psi|_{1,\Omega} + Ch |\psi|_{2,\Omega} \leq C |\psi|_{1,\Omega}. \end{cases}$$

So that

$$\|\mathbf{A}^{-1}\|_2 \geq \frac{C}{h^n}.$$

This gives the desired result.  $\square$

### 3 A well conditioned alternative finite element scheme

In this section, we build an alternative finite element method for which the optimal convergence rates (7) and (8) hold true and the conditioning number of the finite element matrix is of order  $C/h^2$  if all the mesh cells are of diameter  $\sim h$ . We start by the observation that such a method could be based on a subspace  $\tilde{V}_h \subset V_h$  which is the image of interpolation operator  $\tilde{\mathcal{I}}_h$ , *i.e.*

$$\tilde{V}_h := \{v_h \in V_h : [\nabla v_h]_F = 0 \text{ for all } F \in \mathcal{F}_i, i \in \{1, \dots, I\}\},$$

where  $\mathcal{F}_i$  is the set of interior edges (faces) of the patch  $\mathcal{P}_i$  and  $[\cdot]_F$  represents the jump on  $F$ . In view of our interpolation estimates, the problem of finding  $\tilde{u}_h \in \tilde{V}_h$  such that

$$a(\tilde{u}_h, \tilde{v}_h) = l(\tilde{v}_h) \text{ for all } \tilde{v}_h \in \tilde{V}_h$$

would produce an approximate solution with optimal error. Moreover, it is easy to see that the matrix would be well-conditioned since the space  $\tilde{V}_h$  ignores the degenerated cells. Such a method is only of theoretical interest because one cannot easily construct a basis for  $\tilde{V}_h$  using available finite element libraries. In what follows, we use this problem rather as an inspiration in constructing an implementable finite element scheme.

In doing so, we shall impose further restrictions on the mesh:

**Assumption 3.** *The mesh satisfies Assumption 1. Moreover,*

- *The number of patches  $I$  is bounded by some  $I_{\max}$ .*
- *Each patch  $\mathcal{P}_i$  contains a non-degenerated cell  $K_i^{nd}$ , *i.e.* such that  $h_{\mathcal{P}_i}/\rho_{K_i^{nd}} \leq c_0$ .*

In what follows, the constants  $C$  will be allowed to depend on the additional parameter in Assumption 3, *i.e.*  $I_{\max}$ .

We shall need the following modification of the previously defined interpolation operator  $\tilde{\mathcal{I}}_h$ , which makes sense under Assumption 3, *cf.* also Fig. 1, and will be incorporated explicitly into our modified finite element scheme.



**Definition 2.** For all  $v \in H^2(\Omega) \cap H_0^1(\Omega)$ , let  $\widehat{\mathcal{I}}_h(v)$  be the function in  $V_h$  that coincides with the standard Lagrange interpolation  $\mathcal{I}_h v$  on all the cells  $K \in \mathcal{T}_h$  out of the extended patches  $\widetilde{\mathcal{P}}_i$ , and is given on each patch  $\mathcal{P}_i$ , not touching the boundary  $\partial\Omega$ , by

$$\widehat{\mathcal{I}}_h(v)|_{\mathcal{P}_i} := \text{Ext} \left( \mathcal{I}_h(v)|_{K_i^{nd}} \right), \quad (14)$$

where  $\mathcal{I}_h$  stands again for the standard Lagrange interpolation operator on  $K_i^{nd}$ , and Ext stands for the extension of a polynomial from  $K_i^{nd} \subset \mathcal{P}_i$  to the whole  $\mathcal{P}_i$  without changing the coefficients of the polynomial. If the patch  $\mathcal{P}_i$  touches  $\partial\Omega$ , then  $\widehat{\mathcal{I}}_h(v)$  is also based there on formula (14), corrected as in Lemma 1.

REMARK 1. The new interpolation operator  $\widehat{\mathcal{I}}_h(v)$  satisfies the same optimal estimates as that for the old operator  $\widetilde{\mathcal{I}}_h(v)$  which are given in Proposition 1, the proof of which is based on Lemma 1. To prove that Lemma 1 remains valid for  $\widehat{\mathcal{I}}_h$ , i.e. redefining in (9) the original  $Q_h^i$  by  $Q_h^i := \widehat{\mathcal{I}}_h(v)|_{\mathcal{P}_i}$  as in (14), we refer to Theorem (4.4.4) and Corollary (4.4.7) from [7]. Following their proofs, one can see that the only thing to check is the boundedness of operator Ext in (14) as a linear map on the space of polynomials of degree  $\leq 1$  equipped with the norm of  $L^\infty(K_i^{nd})$  to  $L^\infty(\mathcal{P}_i)$ . This, in turn, follows easily from our geometrical Assumptions 1, 3.

### 3.1 An alternative scheme

We denote by  $a_\omega$  the restriction of  $a$  on a subset  $\omega$  of  $\Omega$ , and by  $(\cdot, \cdot)_\omega$  the inner product in  $L^2(\omega)$ . Consider the bilinear form  $a_h$  defined for all  $u_h, v_h \in V_h$  by

$$a_h(u_h, v_h) := a_{\Omega_h^{nd}}(u_h, v_h) + \sum_i a_{\mathcal{P}_i}(\widehat{\mathcal{I}}_h u_h, \widehat{\mathcal{I}}_h v_h) + \sum_i \frac{1}{h_{\mathcal{P}_i}^2} ((\text{Id} - \widehat{\mathcal{I}}_h)u_h, (\text{Id} - \widehat{\mathcal{I}}_h)v_h)_{\mathcal{P}_i}, \quad (15)$$

where  $\Omega_h^{nd} := \Omega \setminus (\cup_i \overline{\mathcal{P}}_i)$  and the interpolation operator  $\widehat{\mathcal{I}}_h$  is defined by (14), i.e.  $u_h$  is not used directly inside the patches in the second term of  $a_h$ , but rather it is extended from a non-degenerate cell inside each patch. The third term in  $a_h$  will serve, loosely speaking, to penalize the eventual non-alignment of the approximate solution  $u_h$  with the optimal subspace  $\widetilde{V}_h$ .

We now introduce the following method approximating System (5): find  $u_h \in V_h$  such that

$$a_h(u_h, v_h) = l(v_h) \text{ for all } v_h \in V_h. \quad (16)$$

The idea of using the polynomial extension from “good” to “bad” mesh cells in the scheme (16) is borrowed from [13]. We shall also see that the scheme can be recast in a form using the interior penalization on the mesh facets between “good” and “bad” cells, as in the ghost penalty method [9].

### 3.2 A priori estimate

The approximation of System (5) by (16) induces a quasi-optimal convergence rate:

**Theorem 4** (*A priori estimate*). *Let  $u \in V$  and  $u_h \in V_h$  be the solutions to System (5) and System (16), respectively. Then, under Assumption 3, we have for any  $\varepsilon > 0$  if  $n = 2$ , and for  $\varepsilon = 0$  if  $n = 3$ ,*

$$|u - \Pi_h u_h|_{1,\Omega} := |u - u_h|_{1,\Omega_h^{nd}} + \sum_i |u - \widehat{\mathcal{I}}_h u_h|_{1,\mathcal{P}_i} \leq Ch^{1-\varepsilon} |u|_{2,\Omega}, \quad (17)$$

where  $\Pi_h u_h$  is equal to  $u_h$  on  $\Omega_h^{nd}$  and  $\widehat{\mathcal{I}}_h u_h$  on  $\mathcal{P}_i$ . Moreover, if  $\Omega$  is convex,

$$\|u - u_h\|_{0,\Omega} \leq Ch^{2-\varepsilon} |u|_{2,\Omega}.$$

In the case  $n = 2$ ,  $C$  depends on  $\varepsilon$  (in addition to its dependence on the mesh regularity).

Before proving Theorem 4, we first give some auxiliary results.

**Lemma 3** (Galerkin orthogonality). *Consider  $u$  and  $u_h$  the solution to Systems (5) and (16). Then*

$$a_{\Omega_h^d}(u_h - u, v_h) - \sum_i a_{\mathcal{P}_i}(u, v_h) + \sum_i a_{\mathcal{P}_i}(\widehat{\mathcal{I}}_h u_h, \widehat{\mathcal{I}}_h v_h) + \sum_i \frac{1}{h_{\mathcal{P}_i}^2} ((\text{Id} - \widehat{\mathcal{I}}_h)u_h, (\text{Id} - \widehat{\mathcal{I}}_h)v_h)_{0, \mathcal{P}_i} = 0,$$

for all  $v_h \in V_h$ .

The proof of Lemma 3 is immediate.

We shall need the norm  $\|\cdot\|$  defined for all  $v_h \in V_h$  by

$$\|v_h\| := a_h(v_h, v_h)^{1/2} = \left( |v_h|_{1, \Omega_h^d}^2 + \sum_i |\widehat{\mathcal{I}}_h v_h|_{1, \mathcal{P}_i}^2 + \sum_i \frac{1}{h_{\mathcal{P}_i}^2} \|v_h - \widehat{\mathcal{I}}_h v_h\|_{0, \mathcal{P}_i}^2 \right)^{\frac{1}{2}}.$$

Note for the future use that this norm is also well defined on  $V \cap H^2(\Omega)$ .

**Lemma 4.** *Under Assumption 3, for all  $v_h \in V_h$ , it holds*

$$\sum_i |v_h - \widehat{\mathcal{I}}_h v_h|_{1, \widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \leq C \|v_h\|^2.$$

*Proof.* It suffices to prove for each patch

$$|v_h - \widehat{\mathcal{I}}_h v_h|_{1, \widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \leq C \left( |v_h|_{1, \widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 + |\widehat{\mathcal{I}}_h v_h|_{1, \mathcal{P}_i}^2 + \frac{1}{h_{\mathcal{P}_i}^2} \|(\text{Id} - \widehat{\mathcal{I}}_h)v_h\|_{0, \mathcal{P}_i}^2 \right). \quad (18)$$

Using the fact that the number of cells in each  $\widetilde{\mathcal{P}}_i$  is bounded by a constant  $M$  (see Assumption 1), the functional space involved in (18) is of finite dimension and the geometry of the patch  $\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i$  is governed by a finite number of parameters, the existence of the constant  $C$  will follow from maximization of the ratio of the left-hand side to the right-hand side over all polynomials and all the acceptable geometries and from a homogeneity argument. We need only to verify that if the right-hand side vanishes, *i.e.*

$$|v_h|_{1, \widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 + |\widehat{\mathcal{I}}_h v_h|_{1, \mathcal{P}_i}^2 + \frac{1}{h_{\mathcal{P}_i}^2} \|(\text{Id} - \widehat{\mathcal{I}}_h)v_h\|_{0, \mathcal{P}_i}^2 = 0$$

on a patch  $\mathcal{P}_i$  and for some function  $v_h \in V_h$ , then the left-hand side vanishes as well, *i.e.*  $v_h - \widehat{\mathcal{I}}_h v_h$  is constant on  $\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i$ . Since  $|v_h|_{1, \widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i} = 0$  and  $|\widehat{\mathcal{I}}_h v_h|_{1, \mathcal{P}_i} = 0$

$$v_h = C_i \text{ in } \widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i, \quad \widehat{\mathcal{I}}_h v_h = D_i \text{ in } \mathcal{P}_i$$

with some constants  $C_i$  and  $D_i$ . Since  $\|(\text{Id} - \widehat{\mathcal{I}}_h)v_h\|_{0, \mathcal{P}_i} = 0$ , we deduce that

$$v_h = \widehat{\mathcal{I}}_h v_h = C_i = D_i \text{ in } \mathcal{P}_i$$

so that  $(\text{Id} - \widehat{\mathcal{I}}_h)v_h = 0$  on  $\partial\mathcal{P}_i$ . This entails  $(\text{Id} - \widehat{\mathcal{I}}_h)v_h = 0$  in  $\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i$  by construction of  $\widehat{\mathcal{I}}_h$ .  $\square$

**Lemma 5.** *Under Assumption 3, we have for any  $\varepsilon > 0$  if  $n = 2$ , and for  $\varepsilon = 0$  if  $n = 3$ ,*

$$\left( \sum_i |u|_{1, \widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq C_\varepsilon h^{1-\varepsilon} \|u\|_{2, \Omega}$$

for all  $u \in H^2(\Omega)$ .

*Proof.* Let  $\overline{\nabla u}^i$  denote the average of  $\nabla u$  on  $\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i$ , *i.e.*

$$\overline{\nabla u}^i := \frac{1}{|\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i|} \int_{\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i} \nabla u.$$

Then, by Poincaré inequality, it holds

$$|u|_{1, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i} \leq \|\nabla u - \overline{\nabla u}^i\|_{0, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i} + \|\overline{\nabla u}^i\|_{0, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i} \leq Ch|u|_{2, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i} + \|\overline{\nabla u}^i\|_{0, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i}.$$

We now consider separately the cases  $n = 2$  and  $n = 3$ . If  $n = 2$ , using Hölder inequality with exponents  $q > 2$  and  $\frac{q}{q-1}$  and the assumption that  $|\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i|$  is of order  $h^2$ , we have

$$\|\overline{\nabla u}^i\|_{0, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i} = \frac{1}{|\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i|^{\frac{1}{2}}} \left| \int_{\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i} \nabla u \right| \leq \|\nabla u\|_{L^q(\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i)} |\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i|^{\frac{q-1}{q} - \frac{1}{2}} \leq Ch^{1-\varepsilon} \|\nabla u\|_{L^q(\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i)}$$

with  $\varepsilon = \frac{2}{q}$ . Summing over all the patches, we have by the discrete Hölder inequality with exponents  $\frac{q}{2}$  and  $\frac{q/2}{q/2-1}$  (recall that the number of patches  $I$  is assumed uniformly bounded) and by the Sobolev embedding  $H^1(\Omega) \rightarrow L^q(\Omega)$

$$\left( \sum_i \|\overline{\nabla u}^i\|_{0, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq Ch^{1-\varepsilon} \left( \sum_i \|\nabla u\|_{L^q(\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i)}^2 \right)^{\frac{1}{2}} \leq Ch^{1-\varepsilon} \|\nabla u\|_{L^q(\Omega)} I^{\frac{1}{2} - \frac{1}{q}} \leq C_\varepsilon h^{1-\varepsilon} \|\nabla u\|_{1, \Omega}$$

with  $C_\varepsilon$  depending both on  $q$  (thus on  $\varepsilon$ ) and on  $\Omega$ .

Similarly, if  $n = 3$ , using Hölder inequality with exponents 6 and  $\frac{6}{5}$  and the assumption that  $|\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i|$  is of order  $h^3$ , we have

$$\|\overline{\nabla u}^i\|_{0, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i} \leq \|\nabla u\|_{L^6(\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i)} |\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i|^{\frac{5}{6} - \frac{1}{2}} \leq Ch \|\nabla u\|_{L^6(\tilde{\mathcal{P}}_i \setminus \mathcal{P}_i)}.$$

Summing over all the patches, we have by the discrete Hölder inequality with exponents 3 and  $\frac{3}{2}$  and by the Sobolev embedding  $H^1(\Omega) \rightarrow L^6(\Omega)$

$$\left( \sum_i \|\overline{\nabla u}^i\|_{0, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq Ch \|\nabla u\|_{L^6(\Omega)} I^{\frac{1}{3}} \leq Ch \|\nabla u\|_{1, \Omega}$$

with a constant depending on  $\Omega$  and without introducing an additional parameter  $\varepsilon$ , *i.e.* setting  $\varepsilon = 0$ . We conclude for both  $n = 2$  and 3

$$\left( \sum_i |u|_{1, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq Ch \left( \sum_i |u|_{2, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} + C_\varepsilon h^{1-\varepsilon} \left( \sum_i \|\nabla u\|_{1, \tilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq C_\varepsilon h^{1-\varepsilon} \|u\|_{2, \Omega}.$$

□

*Proof of Theorem 4.* Let  $e_h := \widehat{\mathcal{I}}_h u - u_h$ . We remark that

$$\begin{aligned} \|e_h\|^2 &= a_h(e_h, e_h) \\ &= a_{\Omega_h^{nd}}(\widehat{\mathcal{I}}_h u - u_h, e_h) + \sum_i a_{\mathcal{P}_i}(\widehat{\mathcal{I}}_h u - \widehat{\mathcal{I}}_h u_h, \widehat{\mathcal{I}}_h e_h) \\ &\quad + \sum_i \frac{1}{h_{\mathcal{P}_i}^2} ((\widehat{\mathcal{I}}_h - \text{Id})u_h, (\text{Id} - \widehat{\mathcal{I}}_h)e_h)_{\mathcal{P}_i}. \end{aligned}$$

Lemma 3 leads to

$$\begin{aligned} \|e_h\|^2 &= a_{\Omega_h^{nd}}(\widehat{\mathcal{I}}_h u - u, e_h) + \sum_i a_{\mathcal{P}_i}(\widehat{\mathcal{I}}_h u, \widehat{\mathcal{I}}_h e_h) - \sum_i a_{\mathcal{P}_i}(u, e_h) \\ &= a_{\Omega_h^{nd}}(\widehat{\mathcal{I}}_h u - u, e_h) + \sum_i a_{\mathcal{P}_i}(\widehat{\mathcal{I}}_h u - u, \widehat{\mathcal{I}}_h e_h) + \sum_i a_{\mathcal{P}_i}(u, \widehat{\mathcal{I}}_h e_h - e_h). \end{aligned} \tag{19}$$

We now estimate each term in the right-hand side. Using Proposition 1 for the interpolation operator  $\widehat{I}_h$ , cf. Remark 1, it holds

$$\begin{aligned} a_{\Omega_h^{nd}}(\widehat{I}_h u - u, e_h) &\leq |\widehat{I}_h u - u|_{1,\Omega} |e_h|_{1,\Omega_h^{nd}} \\ &\leq Ch|u|_{2,\Omega} \| \| e_h \| \| \end{aligned} \quad (20)$$

and

$$\begin{aligned} \sum_i a_{\mathcal{P}_i}(\widehat{I}_h u - u, \widehat{I}_h e_h) &\leq \sum_i |\widehat{I}_h u - u|_{1,\mathcal{P}_i} |\widehat{I}_h e_h|_{1,\mathcal{P}_i} \\ &\leq Ch|u|_{2,\Omega} \| \| e_h \| \| . \end{aligned} \quad (21)$$

Concerning the third term, it holds

$$\begin{aligned} \sum_i a_{\mathcal{P}_i}(u, \widehat{I}_h e_h - e_h) &= \sum_i (\partial_n u, \widehat{I}_h e_h - e_h)_{0,\partial\mathcal{P}_i} - \sum_i (\Delta u, \widehat{I}_h e_h - e_h)_{0,\mathcal{P}_i} \\ &= \sum_i (\Delta u, \widehat{I}_h e_h - e_h)_{0,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i} - \sum_i (\Delta u, \widehat{I}_h e_h - e_h)_{0,\mathcal{P}_i} + \sum_i a_{\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}(u, \widehat{I}_h e_h - e_h). \end{aligned}$$

Since  $h_{\widetilde{\mathcal{P}}_i} \leq Ch$ , we obtain the Poincaré type inequality

$$\| \widehat{I}_h e_h - e_h \|_{0,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i} \leq Ch \| \widehat{I}_h e_h - e_h \|_{1,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}. \quad (22)$$

By Cauchy-Schwarz inequality, inequality (22), Lemma 4 it holds

$$\sum_i (\Delta u, \widehat{I}_h e_h - e_h)_{0,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i} \leq \left( \sum_i |u|_{2,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{1/2} \left( \sum_i \| \widehat{I}_h e_h - e_h \|_{0,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{1/2} \leq Ch|u|_{2,\Omega} \| \| e_h \| \| . \quad (23)$$

Again, using Cauchy-Schwarz inequality, we obtain

$$\sum_i (\Delta u, \widehat{I}_h e_h - e_h)_{0,\mathcal{P}_i} \leq \left( \sum_i |u|_{2,\mathcal{P}_i}^2 \right)^{1/2} \left( \sum_i \| \widehat{I}_h e_h - e_h \|_{0,\mathcal{P}_i}^2 \right)^{1/2} \leq Ch|u|_{2,\Omega} \| \| e_h \| \| . \quad (24)$$

By Lemmas 4 and 5

$$\sum_i a_{\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}(u, \widehat{I}_h e_h - e_h) \leq \left( \sum_i |u|_{1,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \left( \sum_i \| \widehat{I}_h e_h - e_h \|_{1,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq C_\varepsilon h^{1-\varepsilon} \| \| u \| \| \| \| e_h \| \| . \quad (25)$$

Thus, Proposition 1 for the interpolation operator  $\widehat{I}_h$  and (19), (20), (21), (23), (24), (25) lead to

$$\| \| u - u_h \| \| \leq \| \| u - \widehat{I}_h u \| \| + \| \| \widehat{I}_h u - u_h \| \| \leq C_\varepsilon h^{1-\varepsilon} |u|_{2,\Omega}.$$

Consider now the solution  $w \in V$  to

$$a(w, v) = (u - \widehat{I}_h u_h, v), \quad \forall v \in V.$$

Observe

$$a(\widehat{I}_h u_h, \widehat{I}_h w) = a_h(u_h, \widehat{I}_h w) - \sum_i a_{\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}(u_h - \widehat{I}_h u_h, \widehat{I}_h w) = (f, \widehat{I}_h w) - \sum_i a_{\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}(u_h - \widehat{I}_h u_h, \widehat{I}_h w)$$

so that

$$a(u - \widehat{I}_h u_h, \widehat{I}_h w) = \sum_i a_{\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}(u_h - \widehat{I}_h u_h, \widehat{I}_h w) = \sum_i a_{\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}(e_h - \widehat{I}_h e_h, \widehat{I}_h w)$$

with  $e_h = u_h - \widehat{\mathcal{I}}_h u$ . Thus,

$$\begin{aligned} \|u - \widehat{\mathcal{I}}_h u_h\|_{0,\Omega}^2 &= a(u - \widehat{\mathcal{I}}_h u_h, w - \widehat{\mathcal{I}}_h w) + \sum_i a_{\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}(e_h - \widehat{\mathcal{I}}_h e_h, \widehat{\mathcal{I}}_h w) \\ &\leq Ch|u - \widehat{\mathcal{I}}_h u_h|_{1,\Omega}|w|_{2,\Omega} + \left( \sum_i |e_h - \widehat{\mathcal{I}}_h e_h|_{1,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \left( \sum_i |\widehat{\mathcal{I}}_h w|_{1,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}}, \end{aligned} \quad (26)$$

where we have used the interpolation estimate. Using Lemma 3 and (above), we obtain

$$\left( \sum_i |e_h - \widehat{\mathcal{I}}_h e_h|_{1,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq C \| \| e_h \| \| \leq C_\varepsilon h^{1-\varepsilon} |u|_{2,\Omega}$$

and

$$\begin{aligned} |u - \widehat{\mathcal{I}}_h u_h|_{1,\Omega} &\leq C \left( |u - u_h|_{1,\Omega^{nd}}^2 + |u_h - \widehat{\mathcal{I}}_h u_h|_{1,\Omega^{nd}}^2 + \sum_i |u - \widehat{\mathcal{I}}_h u_h|_{1,\mathcal{P}_i}^2 \right)^{\frac{1}{2}} \\ &\leq C_\varepsilon h^{1-\varepsilon} |u|_{2,\Omega} + \left( \sum_i |e_h - \widehat{\mathcal{I}}_h e_h|_{1,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq C_\varepsilon h^{1-\varepsilon} |u|_{2,\Omega}. \end{aligned}$$

We also have by regularity of elliptic problem in a convex polygon (polyhedron)

$$|w|_{2,\Omega} \leq C \| \| u - \widehat{\mathcal{I}}_h u_h \| \|_{0,\Omega}$$

and by Lemma 5

$$\left( \sum_i |\widehat{\mathcal{I}}_h w|_{1,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq |w - \widehat{\mathcal{I}}_h w|_{1,\Omega} + \left( \sum_i |w|_{1,\widetilde{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \right)^{\frac{1}{2}} \leq Ch|w|_{2,\Omega} + C_\varepsilon h^{1-\varepsilon} \| \| w \| \|_{2,\Omega} \leq C_\varepsilon h^{1-\varepsilon} \| \| u - \widehat{\mathcal{I}}_h u_h \| \|_{0,\Omega}.$$

Substituting into (26) gives

$$\| \| u - \widehat{\mathcal{I}}_h u_h \| \|_{0,\Omega}^2 \leq C_\varepsilon h^{2-2\varepsilon} |u|_{2,\Omega} \| \| u - \widehat{\mathcal{I}}_h u_h \| \|_{0,\Omega}$$

which in combination with the triangle inequality and the estimate for  $\| \| u_h - \widehat{\mathcal{I}}_h u_h \| \|_{0,\mathcal{P}_i}$  contained in the estimate for  $\| \| u - u_h \| \|$  gives the announced  $L^2$ -error estimate.  $\square$

### 3.3 Conditioning of the system matrix

We are now going to prove that the conditioning number of the finite element matrix associated to the bilinear form  $a_h$  of the alternative scheme does not deteriorate in the presence of degenerated cells: it is of order  $1/h^2$  if the mesh is quasi-uniform in a sense specified below.

**Proposition 3** (Conditioning). *Suppose that Assumption 3 holds and the union of mesh cells  $\omega_x$  attached to each node  $x$  of  $\mathcal{T}_h$  satisfies*

$$c_1 h^n \leq |\omega_x| \leq c_2 h^n \quad (27)$$

*with some constants  $c_1, c_2$ . Then, the conditioning number  $\kappa(\mathbf{A})$  of the matrix  $\mathbf{A}$  associated to the bilinear form  $a_h$  in  $V_h$  satisfies*

$$\kappa(\mathbf{A}) \leq Ch^{-2}.$$

REMARK 2. Condition (27) is satisfied for instance if the mesh is quasi-uniform in the sense  $h_K \geq c_3 h$  for all  $K \in \mathcal{T}_h$ , and each patch is constituted of a degenerated cell and a non-degenerated cell. This is the situation considered in our numerical simulations given in Section 4.

Before proving Proposition 3, we first introduce some auxiliary results:

**Lemma 6** (Coercivity of  $a_h$ ). *Under the assumptions of Proposition 3, it holds for all  $v_h \in V_h$*

$$a_h(v_h, v_h) \geq C \|v_h\|_{0,\Omega}^2.$$

*Proof.* Let  $v_h \in V_h$ . Observe, using triangle and Poincaré inequalities,

$$\begin{aligned} \|v_h\|_{0,\Omega} &\leq \|\widehat{\mathcal{I}}_h v_h\|_{0,\Omega} + \|v_h - \widehat{\mathcal{I}}_h v_h\|_{0,\Omega} \\ &\leq C |\widehat{\mathcal{I}}_h v_h|_{1,\Omega} + \|v_h - \widehat{\mathcal{I}}_h v_h\|_{0,\Omega} \\ &\leq C \left( |v_h|_{1,\Omega_h^{nd}}^2 + \sum_i |\widehat{\mathcal{I}}_h v_h|_{1,\mathcal{P}_i}^2 + \sum_i |v_h - \widehat{\mathcal{I}}_h v_h|_{1,\widehat{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 + \sum_i \|v_h - \widehat{\mathcal{I}}_h v_h\|_{0,\widehat{\mathcal{P}}_i}^2 \right)^{\frac{1}{2}}. \end{aligned}$$

Following the proof of Lemma 4, we see easily

$$\sum_i |v_h - \widehat{\mathcal{I}}_h v_h|_{1,\widehat{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 + \sum_i \|v_h - \widehat{\mathcal{I}}_h v_h\|_{0,\widehat{\mathcal{P}}_i \setminus \mathcal{P}_i}^2 \leq C \|v_h\|^2.$$

This implies  $\|v_h\|_{0,\Omega} \leq C \|v_h\|$  which is equivalent to the desired result.  $\square$

**Lemma 7** (Continuity of  $a_h$ ). *Under the assumptions of Proposition 3, it holds for all  $u_h, v_h \in V_h$*

$$a_h(u_h, v_h) \leq \frac{C}{h^2} \|u_h\|_{0,\Omega} \|v_h\|_{0,\Omega}.$$

*Proof.* Let  $u_h, v_h \in V_h$ . Since the cells of  $\Omega_h^{nd}$  and the patches  $\mathcal{P}_i$  are regular, we obtain using the inverse inequality

$$a_{\Omega_h^{nd}}(u_h, v_h) \leq \frac{C}{h^2} \|u_h\|_{0,\Omega_h^{nd}} \|v_h\|_{0,\Omega_h^{nd}}$$

and

$$a_{\mathcal{P}_i}(\widehat{\mathcal{I}}_h u_h, \widehat{\mathcal{I}}_h v_h) \leq \frac{C}{h^2} \|\widehat{\mathcal{I}}_h u_h\|_{0,\mathcal{P}_i} \|\widehat{\mathcal{I}}_h v_h\|_{0,\mathcal{P}_i}.$$

Using the equivalence of the norm in finite dimensional spaces and the fact that  $\mathcal{P}_i$  and  $K_i^{nd}$  are regular, for all  $w_h \in V_h$ , it holds

$$\|\widehat{\mathcal{I}}_h w_h\|_{0,\mathcal{P}_i} \leq C \|w_h\|_{0,K_i^{nd}}.$$

We deduce that

$$\|(\text{Id} - \widehat{\mathcal{I}}_h)w_h\|_{0,\mathcal{P}_i} + \|\widehat{\mathcal{I}}_h w_h\|_{0,\mathcal{P}_i} \leq C (\|w_h\|_{0,\mathcal{P}_i} + \|w_h\|_{0,K_i^{nd}}) \leq C \|w_h\|_{0,\mathcal{P}_i}$$

which leads to the conclusion.  $\square$

*Proof of Proposition 3.* We first remark using (27), that there exists  $C_1, C_2 > 0$  such that for all  $w_h \in V_h$  and  $\mathbf{w}$  its associated vector in  $\mathbb{R}^N$

$$C_1 h^{n/2} |\mathbf{w}|_2 \leq \|w_h\|_0 \leq C_2 h^{n/2} |\mathbf{w}|_2. \quad (28)$$

Indeed, denoting by  $\mathcal{N}_h$  the set of nodes of  $\mathcal{T}_h$ , by  $\mathcal{N}_h(K)$  the set of nodes of a simplex  $K \in \mathcal{T}_h$ , and using  $\sim$  to denote the equivalence with universal constant, as in (28), we can conclude

$$\|w_h\|_0^2 \sim \sum_{K \in \mathcal{T}_h} |K| \sum_{x \in \mathcal{N}_h(K)} |w_h(x)|^2 = \sum_{x \in \mathcal{N}_h} |w_h(x)|^2 |\omega_x| \sim h^n |\mathbf{w}|_2^2.$$

In what follows,  $\mathbf{v} \in \mathbb{R}^N$  denotes the vector associated to  $v_h \in V_h$ . Inequality (28) with Lemma 7 imply

$$\|\mathbf{A}\|_2 = \sup_{\mathbf{v} \in \mathbb{R}^N} \frac{(\mathbf{A}\mathbf{v}, \mathbf{v})}{|\mathbf{v}|_2^2} = \sup_{\mathbf{v} \in \mathbb{R}^N} \frac{a(v_h, v_h)}{|\mathbf{v}|_2^2} \leq C h^n \sup_{v_h \in V_h} \frac{a(v_h, v_h)}{\|v_h\|_0^2} \leq C h^{n-2}.$$

Similarly, (28) with Lemma 6 imply

$$\|\mathbf{A}^{-1}\|_2 = \sup_{\mathbf{v} \in \mathbb{R}^N} \frac{|\mathbf{v}|_2^2}{(\mathbf{A}\mathbf{v}, \mathbf{v})} = \sup_{\mathbf{v} \in \mathbb{R}^N} \frac{|\mathbf{v}|_2^2}{a(v_h, v_h)} \leq Ch^{-n} \sup_{v_h \in V_h} \frac{\|v_h\|_0^2}{a(v_h, v_h)} \leq Ch^{-n}.$$

These estimates lead to the desired result.  $\square$

### 3.4 An equivalent, easily implementable variational formulation with interior penalty

Since implementing the interpolation operator  $\widehat{\mathcal{I}}_h$  is not necessary trivial, we rewrite in this section the bilinear form  $a_h$  given in (15) in an equivalent form, which introduces the jumps of the gradients over the interior facets. The resulting method is similar to the ghost penalty from [9].

**Lemma 8.** *Under Assumption 3, suppose moreover that each patch  $\mathcal{P}_i$  is composed of a non-degenerated cell  $K_i^{nd}$  and a degenerated cell  $K_i^{deg}$ . Denote by  $F_i$  the facet between  $K_i^{nd}$  and  $K_i^{deg}$ , as illustrated in Fig. 2.*

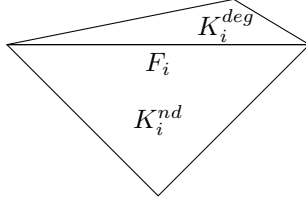


Figure 2: Example of patch  $\mathcal{P}_i = K_i^{nd} \cup K_i^{deg}$ .

Then, for all  $u_h, v_h \in V_h$ , it holds

$$a_h(u_h, v_h) = a_{\Omega_h^{nd}}(u_h, v_h) + \sum_i \frac{|\mathcal{P}_i|}{|K_i^{nd}|} a_{K_i^{nd}}(u_h, v_h) + \kappa_n \sum_i \frac{|K_i^{deg}|^3}{h_{\mathcal{P}_i}^2 |F_i|^2} [\nabla u_h]_{F_i} \cdot [\nabla v_h]_{F_i} \quad (29)$$

with  $\kappa_n := \frac{2n^2}{(n+1)(n+2)}$ .

*Proof.* Let us assume, without loss of generality, that the coordinate axes are chosen so that the  $y$  axis is orthogonal to  $F_i$ , as in Fig. 3. We also denote by  $h_i$  the height of the simplex  $K_i^{deg}$  drawn to the base  $F_i$ .

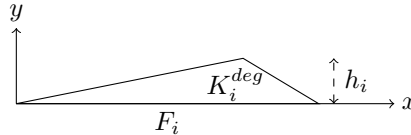


Figure 3: Degenerated cell  $K_i^{deg}$ .

We first remark that, for all  $u_h \in V_h$ ,

$$(\text{Id} - \widehat{\mathcal{I}}_h)u_h = \begin{cases} [\nabla u_h]_{F_i} y & \text{on } K_i^{deg}, \\ 0 & \text{on } K_i^{nd}. \end{cases}$$

Hence, we deduce that

$$((\text{Id} - \widehat{\mathcal{I}}_h)u_h, (\text{Id} - \widehat{\mathcal{I}}_h)v_h)_{\mathcal{P}_i} = [\nabla u_h]_{F_i} \cdot [\nabla v_h]_{F_i} \int_{K_i^{deg}} y^2.$$

Moreover

$$\int_{K_i^{deg}} y^2 = \int_0^{h_i} y^2 |F_i| \left(1 - \frac{y}{h_i}\right)^{n-1} dy = |F_i| h_i^3 \frac{2}{n(n+1)(n+2)} = \frac{|K_i^{deg}|^3}{|F_i|^2} \frac{2n^2}{(n+1)(n+2)},$$

since  $|K_i^{deg}| = \frac{1}{n}|F_i|h_i$ . This leads to the conclusion.  $\square$

## 4 Numerical simulations

In this section, we will illustrate with some numerical examples the sharpness of the *a priori* estimates of Theorem 2 and the efficiency of the method proposed in Section 3.4 to ensure the good conditioning of the matrix. The simulations of this section have been implemented using the finite element library FEniCS [1].

We consider problem (1) on the domain  $\Omega := (0,1) \times (0,1)$  with the right hand side  $f(x,y) = 2\pi^2 \sin(\pi x) \sin(\pi y)$  so that the exact solution is given by  $u(x,y) = \sin(\pi x) \sin(\pi y)$  for  $(x,y) \in \Omega$ . To construct the meshes  $\mathcal{T}_h$  in all our numerical experiments presented below, we start from a uniform Cartesian mesh of step  $h$  and degenerate certain cells so that for each degenerated cell  $K^{deg}$ ,  $h_{K^{deg}} = h$ ,  $\rho_{K^{deg}} \sim h^2$  (more precisely, the distance between the longer side and the opposite node will be equal to  $h^2$ ). In doing so, we take care that each degenerated cell be included in a patch of surrounding cells, and the patches corresponding to distinct degenerated cells do not intersect each other, cf. Fig. 4. Assumptions 1 and 3 are thus satisfied.

We report in Fig. 5 the numerical results obtained on a series of meshes with decreasing  $h$ , taking 10 degenerated cells (as described above) for every  $h$ . We use here the standard scheme (6) to produce the approximated solution  $u_h$ . The  $L^2$  and  $H^1$  absolute errors between  $u$  and  $u_h$  are given on the left in Fig. 5. The optimal convergence rates are indeed observed, as predicted by Theorem 2. However, the conditioning number of the associated finite element matrix is much bigger than  $1/h^2$ , which would be expected on a quasi-uniform mesh with step  $h$ . This is illustrated by Fig. 5, right. The estimate on the conditioning number from Proposition 2 is recovered, *i.e.*  $\kappa(A) \sim 1/(h\varepsilon) \sim 1/h^3$ , since  $\varepsilon = \rho_{K^{deg}} \sim h^2$ .

We now turn to the alternative scheme (16). We have implemented it using the reformulation (29). The results are reported in Fig. 6 using the same meshes containing 10 degenerated cells as above. The errors are reported on the left. We recall that Theorem 4 predicts the optimal convergence in the  $H^1$  norm only if the approximate solution  $u_h$  is post-processed on the degenerated cells, by replacing the actual polynomial giving  $u_h$  on such a cell by the extension  $\Pi_h u_h$  of  $u_h$  from the attached regular cell, cf. the definition of  $|u - \Pi_h u_h|_{1,\Omega}$  in (17). Numerical experiments confirm the optimal  $H^1$  convergence of the post-processed solution and also the necessity of such a post-processing. Indeed, the error with respect to the non-processed approximate solution  $|u - u_h|_{1,\Omega}$  is not of optimal order  $h$ . It is also much bigger than  $|u - \Pi_h u_h|_{1,\Omega}$ . We also note that the optimal  $L^2$  convergence is recovered without any post-processing, as predicted by Theorem 4. We recall that the introduction of the alternative scheme (16) was motivated by the desire to obtain less ill-conditioned matrices. The results in Fig. 6 (right) confirm that conditioning number for this scheme is indeed no longer affected by the presence of degenerated cells, in accordance with Proposition 3.

We recall that the theory of Section 3 concerning the alternative scheme (16) is developed under Assumption 3 supposing, in particular, that the number of degenerate cells is uniformly bounded. In the numerical experiments reported in Figs. 7 and 8, we wish to check if such an assumption is indeed necessary. We consider to this end a sequence of meshes constructed as above, but containing an increasing number of degenerate cells, cf. Fig. 7. We consider namely the densest packing of the degenerated cells allowed by Assumption 1 (the non-intersection of the surrounding patches), which gives approximately 5.5% of degenerated cells. Otherwise, the procedure for degenerating the cells is as above, in particular,  $\rho_{K_i^{deg}} \approx h^2$ . The results are presented in Fig. 8 both for the standard scheme on the left, and the alternative scheme (16) on the right. We first remark that the standard scheme remains optimally convergence in  $L^2$  and  $H^1$ , in accordance with Theorem 2. On the contrary, the alternative scheme (16) does not converge. This observation highlights the sharpness of the results given in Theorem 4.



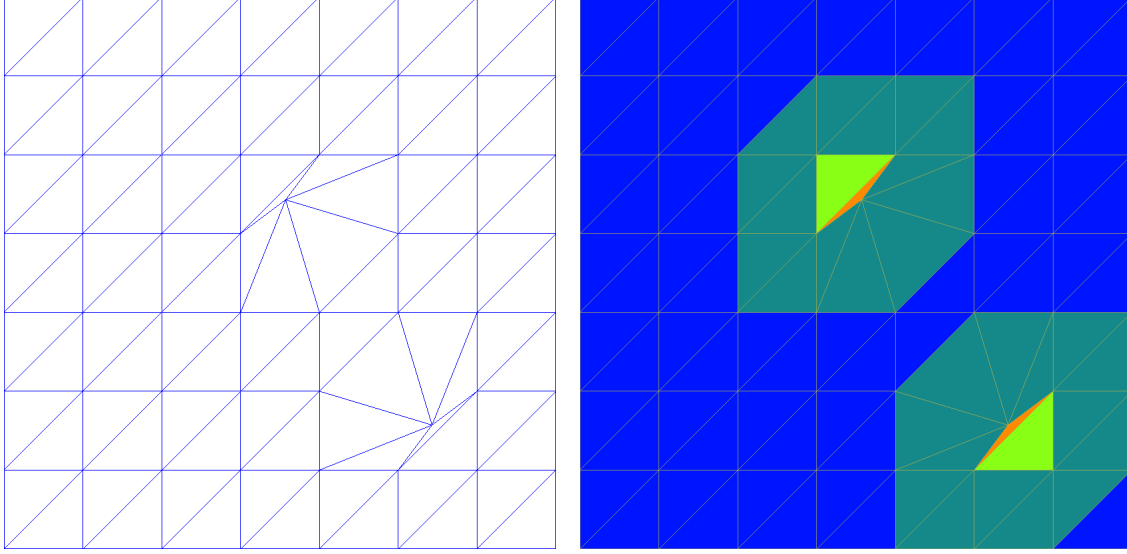


Figure 4: Example of a uniform mesh with 2 arbitrarily chosen degenerated cells  $K_i^{deg}$ ,  $\rho_{K_i^{deg}} \sim h^2$  (left). On the right, the degenerated cells are painted in red, the adjacent regular cells in light green, and the surrounding patches in dark green.

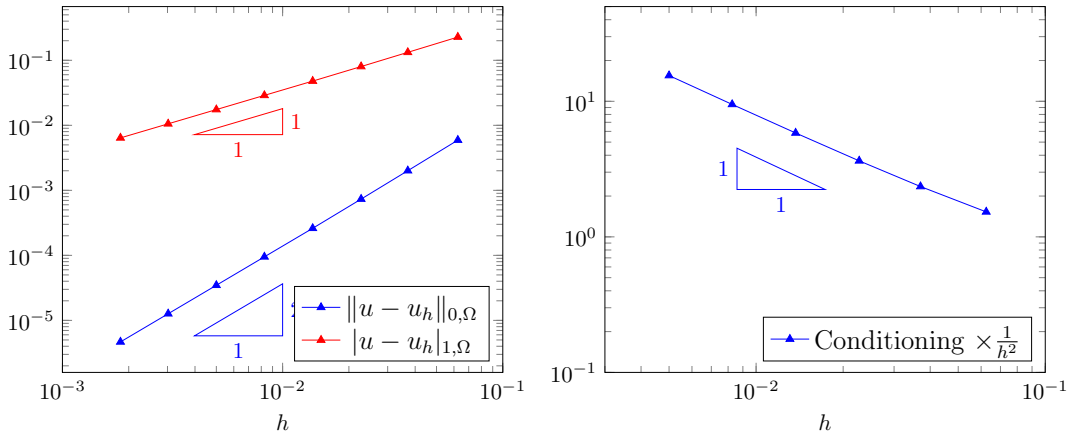


Figure 5: Errors (left) and conditioning (right) for the standard finite element scheme (6) on a sequence of meshes containing 10 degenerated cells with  $\rho_{K_i^{deg}} \sim h^2$ .

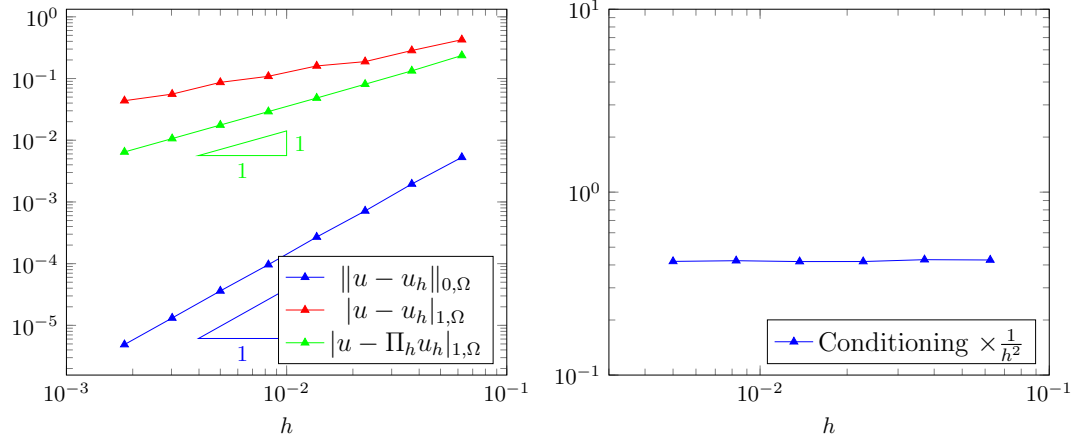


Figure 6: Errors (left) and conditioning (right) for the alternative finite element scheme (16) on a sequence of meshes containing 10 degenerated cells with  $\rho_{K_i^{deg}} \sim h^2$ . The  $H^1$  norm is calculated both using the approximate solution  $u_h$  directly and extending it to the degenerated cells from the adjacent regular cells, as in (17).

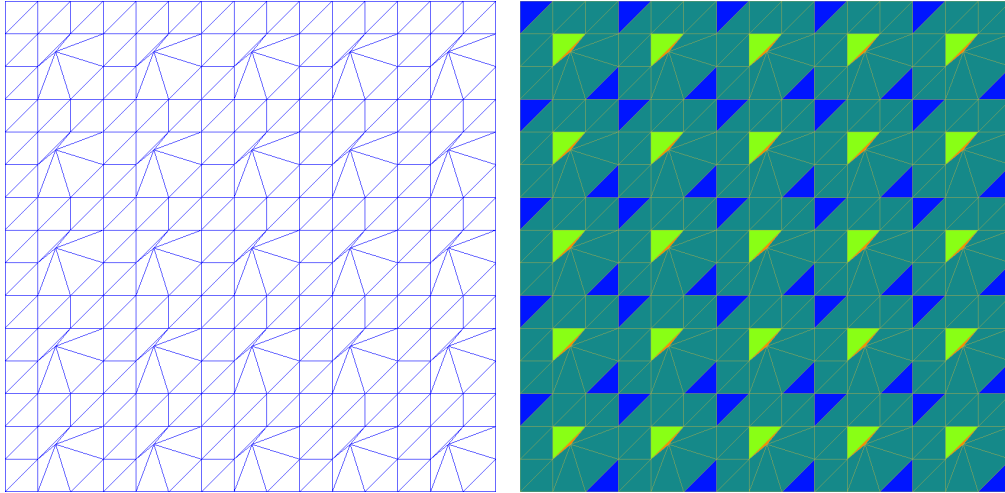


Figure 7: Example of a mesh with densely packed degenerated cells ( $\approx 5.5\%$  of degenerated cells). Left: the mesh. Right: the disjoint patches surrounding the degenerated cells.

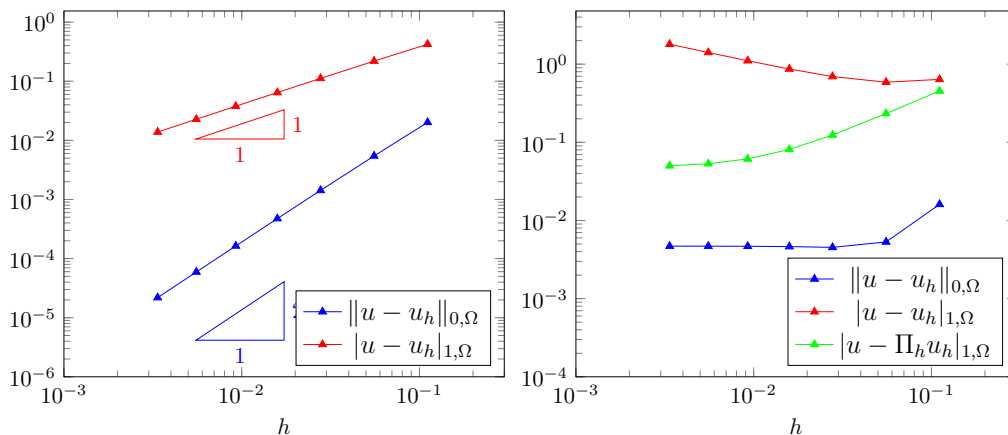


Figure 8: Errors on the meshes containing  $\approx 5.5\%$  of degenerated cells. Left: the standard scheme (6). Right: alternative scheme (16).

## Acknowledgements

The authors are thankful to Marek Bucki (TexiSense) and Franz Chouly (Université de Bourgogne Franche-Comté) for inspiring discussions which were at the origin of this project and for constant support during its realization.

## References

- [1] L. Anders, M. Kent-Andre, G. N. Wells, and al. *Automated Solution of Differential Equations by the Finite Element Method*. Springer, 2012.
- [2] T. Apel. *Anisotropic finite elements: local estimates and applications*. Advances in Numerical Mathematics. B. G. Teubner, Stuttgart, 1999.
- [3] I. Babuška and A. K. Aziz. On the angle condition in the finite element method. *SIAM J. Numer. Anal.*, 13(2):214–226, 1976.
- [4] R. E. Barnhill and J. A. Gregory. Sard kernel theorems on triangular domains with application to finite element error bounds. *Numer. Math.*, 25(3):215–229, 1975/76.
- [5] J. Brandts, S. Korotov, and M. Křížek. On the equivalence of regularity criteria for triangular and tetrahedral finite element partitions. *Comput. Math. Appl.*, 55(10):2227–2233, 2008.
- [6] J. Brandts, S. Korotov, and M. Křížek. Generalization of the Zlámal condition for simplicial finite elements in  $\mathbb{R}^d$ . *Appl. Math.*, 56(4):417–424, 2011.
- [7] S. Brenner and R. Scott. *The mathematical theory of finite element methods*, volume 15. Springer Science & Business Media, 2007.
- [8] M. Bucki, C. Lobos, and Y. Payan. A fast and robust patient specific finite element mesh registration technique: application to 60 clinical cases. *Medical image analysis*, 14(3):303–317, 2010.
- [9] E. Burman. Ghost penalty. *C. R. Acad. Sci. Paris, (Ser. I 348)*:1217–1220, 2010.
- [10] E. Burman, S. Claus, P. Hansbo, M. G. Larson, and A. Massing. CutFEM: discretizing geometry and partial differential equations. *Internat. J. Numer. Methods Engrg.*, 104(7):472–501, 2015.

- [11] P. G. Ciarlet. *The finite element method for elliptic problems*. North-Holland Publishing Co., Amsterdam-New York-Oxford, 1978. Studies in Mathematics and its Applications, Vol. 4.
- [12] A. Hannukainen, S. Korotov, and M. Křížek. The maximum angle condition is not necessary for convergence of the finite element method. *Numer. Math.*, 120(1):79–88, 2012.
- [13] J. Haslinger and Y. Renard. A new fictitious domain approach inspired by the extended finite element method. *SIAM J. Numer. Anal.*, 47(2):1474–1499, 2009.
- [14] P. Jamet. Estimations d’erreur pour des éléments finis droits presque dégénérés. *Rev. Française Automat. Informat. Recherche Opérationnelle Sér.*, 10(R-1):43–60, 1976.
- [15] P. Jamet. Estimation of the interpolation error for quadrilateral finite elements which can degenerate into triangles. *SIAM J. Numer. Anal.*, 14(5):925–930, 1977.
- [16] K. Kobayashi and T. Tsuchiya. On the circumradius condition for piecewise linear triangular elements. *Jpn. J. Ind. Appl. Math.*, 32(1):65–76, 2015.
- [17] V. Kučera. A note on necessary and sufficient conditions for convergence of the finite element method. In *Applications of mathematics 2015*, pages 132–139. Czech. Acad. Sci., Prague, 2015.
- [18] M. Křížek. On the maximum angle condition for linear tetrahedral elements. *SIAM J. Numer. Anal.*, 29(2):513–520, 1992.
- [19] P. Oswald. Divergence of FEM: Babuška-Aziz triangulations revisited. *Appl. Math.*, 60(5):473–484, 2015.
- [20] A. Zeníšek. Convergence of the finite element method for boundary value problems of a system of elliptic equations. *Apl. Mat.*, 14:355–377, 1969.
- [21] M. Zlámal. On the finite element method. *Numer. Math.*, 12:394–409, 1968.