Chapter 6

A hybrid Mortar Virtual Element method for Discrete Fracture Networks simulations

This chapter deals with the theoretical details of the hybrid Mortar Virtual approach introduced in Section 5.4.4 and displays numerical results that show optimal convergence rate for the error and good behaviour on DFN test cases. These results were published in [5].

6.1 Introduction

In [6, 7] the newly developed Virtual Element Method [1] (VEM) was applied in the DFN framework: the methods proposed therein exploit the flexibility of VEM, that allows the treatment of elements with an arbitrary number of edges, even with flat angles. Thanks to this property, a conforming mesh is easily obtained at a moderate computational cost.

In [5], the use of VEM in the DFN framework proposed in [6] is coupled with the well established Mortar Method [8]. A major advantage of this new coupling with respect to previous works that also work with a primal formulation of the problem is that the flux entering/exiting each fracture from its intersections is directly obtained as part of the solution of the discrete problem and not through a post-processing of the results. This approach was already described in Section 5.4.4. In this chapter we give the proofs of well-posedness of the method and show optimal a priori error estimates and numerical tests confirming the theoretical results.

The chapter is organized as follows: in Section 6.2 we state the problem setting, we briefly recall the main features of the VEM needed for the description of our method, and describe the hybrid method obtained from coupling the VEM with the mortar method; Section 6.3 addresses some implementation issues related to the generation of the locally conforming mesh; finally, Section 6.4 reports some numerical results assessing the behaviour of the method.

6.2 The VEM-Mortar method for DFNs

First, let us recall some of the notation established in Section 5.2. A DFN, Ω, is a set of N open planar polygons F_i, $i = 1, \dots, N$, representing the fractures in the medium. In the sequel, we will identify the fractures with the polygons. Fractures intersect each other along segments called *traces*, that we assume to be given by the intersection of exactly two fractures: whenever two traces intersect each other, we split both traces into two sub-traces. The generic trace is indicated by Γ_m, with $m \in \mathcal{M} = \{1, \dots, M\}$. For each $m \in \mathcal{M}$, it is convenient to identify the set $\mathcal{I}_S = \{i, j\}$ of the indices of the two fractures intersecting at Γ_m. For any function or set defined on the whole DFN, its restriction to fracture F_i will be denoted using the subscript i.

As detailed in Section 5.3, we consider the Darcy's law locally on each fracture as a model for the equilibrium of the hydraulic head h. We introduce on each fracture the transmissivity K_i, which is assumed, for the sake of simplicity, to be a scalar function of the local tangential coordinates system on F_i. Let Γ^D be a non-empty portion of $\partial\Omega$ on which the Dirichlet boundary condition h^D is imposed, and let us set $\Gamma_i^D = \Gamma^D \cap F_i$. Note that Γ_i^D is allowed to be empty for some i. Let us assume that $h_i^D \in \mathrm{H}^{\frac{1}{2}}\left(\Gamma_i^D\right)$ for all $i \in \{1, \dots, N\}$. Furthermore, let $\Gamma_i^N = \partial F_i \setminus \Gamma_i^D$ be the local Neumann boundary and let $h_i^N \in \mathrm{H}^{-\frac{1}{2}}\left(\Gamma_i^N\right)$ be the Neumann boundary condition imposed therein. We define the following functional spaces:

$$
\begin{aligned}
V_i &= \left\{ v \in \mathrm{H}^1(F_i) : \gamma_{\Gamma_i^D}(v) = 0 \right\} \quad \forall i = 1, \dots, N\,, \\
V_i^D &= \left\{ v \in \mathrm{H}^1(F_i) : \gamma_{\Gamma_i^D}(v) = h_i^D \right\} \quad \forall i = 1, \dots, N\,, \\
V^D &= \prod_{i=1}^N V_i^D\,, \qquad V = \prod_{i=1}^N V_i\,,
\end{aligned}
$$

where $\gamma_{\Gamma_i^D} : \mathrm{H}^1(F_i) \mapsto \mathrm{H}^{\frac{1}{2}}\left(\Gamma_i^D\right)$ is the trace operator on Γ_i^D. The problem of interest is to find $h \in V^D$ such that $h = h_0 + \mathcal{R}\left(h^D\right)$ where $\mathcal{R}\left(h^D\right)$ is a lifting of h^D on V^D and $h_0 \in V$ satisfies, for any given $v \in V$ and any $i = 1, \dots, N$,

$$
\begin{aligned}
\left(\mathsf{K}_i \nabla h_{0i}, \nabla v_i\right)_{F_i} - \left\langle \left[\!\left[\frac{\partial h_i}{\partial \hat{n}_{\mathcal{M}_i}} \right]\!\right]_{\mathcal{M}_i}, v_i \right\rangle_{\mathcal{M}_i} &= \left(f_i, v_i\right)_{F_i} + \left\langle H_i^N, v_i \right\rangle_{\pm\frac{1}{2}, \Gamma_i^N} \\
&\quad - \left(\mathsf{K}_i \nabla \mathcal{R}_i\left(h_i^D\right), \nabla v_i\right)_{F_i}\,, \quad (6.1)
\end{aligned}
$$

where we refer the reader to Section 5.2 for the details of the notation. The equations on each fracture are coupled by the balance of fluxes on traces:

$$
\forall m \in \mathcal{M}, \text{ if } \mathcal{I}_m = \{i, j\}, \quad \left[\!\left[\mathsf{K}_i \frac{\partial h_i}{\partial \hat{n}_m^i} \right]\!\right]_{\Gamma_m} + \left[\!\left[\mathsf{K}_j \frac{\partial h_j}{\partial \hat{n}_m^j} \right]\!\right]_{\Gamma_m} = 0\,, \tag{6.2}
$$

and by the continuity of the solution across traces, that can be written as:

$$
\forall m \in \mathcal{M}, \forall \psi \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_m), \ \left\langle [\![h]\!]_{\mathcal{M}}, \psi \right\rangle_{\mathcal{M}} = 0\,. \tag{6.3}
$$

As described in Section 5.4.2, problem (6.1)–(6.3) can be written as a minimization problem for the energy functional in (5.9), whose solution is the couple $(h, \lambda) \in V^D \times M = \prod_{m \in \mathcal{M}} \mathrm{H}^{-\frac{1}{2}}(\Gamma_m)$ such that, $h = h^0 + \mathcal{R}\left(h^D\right)$, $h^0 \in V$ and

$$
\begin{cases}
a\left(h^0, v\right) + b^{\mathcal{M}}(v, \Lambda) = (f, v) + \left\langle h^N, v \right\rangle_{\pm\frac{1}{2}, \Gamma^N} - a\left(\mathcal{R}\left(h^D\right), v\right) & \forall v \in V, \\
b^{\mathcal{M}}\left(h^0, \psi\right) = -b^{\mathcal{M}}\left(\mathcal{R}\left(h^D\right), \psi\right) & \forall \psi \in M\,,
\end{cases} \tag{6.4}
$$

being, $\forall w, v \in V$, $\psi \in M$,

$$a(w, v_\delta) = \sum_{i \in \mathcal{I}} (\mathsf{K}_i \nabla w, \nabla v)_{F_i} \ ,$$
$$b^{\mathcal{M}}(v, \psi) = \sum_{m \in \mathcal{M}} \left\langle \psi_m, [\![v]\!]_{\Gamma_m} \right\rangle_{\Gamma_m} \ ,$$

Let us endow V^D and V with the norm

$$\|v\|_V = \left(\sum_{i=1}^{N} \|v_i\|_{F_i}^2 + a_i(v_i, v_i) \right)^{\frac{1}{2}} . \tag{6.5}$$

Well-posedness of problem (6.4) follows observing that, introducing the Hilbert space

$$W = \left\{ v \in V \colon \forall m \in \mathcal{M}, \forall \psi \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_m) \, , \ \left\langle [\![v]\!]_{\Gamma_m}, \psi \right\rangle_{\pm\frac{1}{2}, \Gamma_m} = 0 \right\} = \ker\left(b^{\mathcal{M}}\right) \ ,$$

problem (6.4) is equivalent to: *find* $h^0 \in W$ *such that*

$$a\left(h^0, v\right) = (f, v) + \left\langle h^N, v \right\rangle_{\pm\frac{1}{2}, \Gamma^N} - a\left(\mathcal{R}\left(h^D\right), v\right) \quad \forall v \in W \, .$$

The latter problem is well posed because the bilinear form in the left-hand side of (6.1) is coercive on said space. According to [9, Corollary 1.1], problem (6.4) is well posed since a is coercive on the space

$$W = \ker(b^{\mathcal{M}}) = \left\{ v \in V \colon b^{\mathcal{M}}(v, \psi) = 0 \quad \forall \psi \in M \right\} , \tag{6.6}$$

equipped with the norm

$$|\!|\!| v |\!|\!| \ = a(v, v)^{\frac{1}{2}} \ , \tag{6.7}$$

and thanks to the inf-sup condition

$$\exists \beta > 0 \colon \inf_{\psi \in M} \sup_{v \in V} \frac{b^{\mathcal{M}}(v, \psi)}{\|v\|_V \, \|\psi\|_M} \geq \beta \tag{6.8}$$

which is satisfied by M and V.

Remark 6.1. The proof that $|\!|\!|\cdot|\!|\!|$ is a norm on W follows from the definition of W, see e.g. Proposition 6.1 for a proof in a wider space. The existence of the inf-sup constant β follows from [4, 10], which guarantees the existence of a constant $\beta_m > 0$, $\forall m \in \mathcal{M}$, such that

$$\inf_{\psi \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_m)} \sup_{v \in V} \frac{\left\langle [\![v]\!]_{\Gamma_m}, \psi \right\rangle_{\pm\frac{1}{2}, \Gamma_m}}{\|v\|_V \, \|\psi\|_{-\frac{1}{2}, \Gamma_m}} \geq \beta_m \, .$$

Thus, (6.8) holds with $\beta = \min_{m \in \mathcal{M}} \beta_m$, that is strictly positive because there is a finite number of traces.

As detailed in Section 5.4, problem (6.4) can be discretized by coupling a VEM discretization on each fracture via the Mortar method, thus imposing weak continuity and obtaining a piecewise polynomial approximation of the flux on each trace. Indicating such approximation with λ_δ, we have the following discrete formulation of the problem: find $h_\delta = h_\delta^0 + \mathcal{R}_\delta\left(h^D\right)$, with $h_\delta^0 \in V_\delta$ and $\lambda_\delta \in M_\delta$ such that,

$$\begin{cases} a_\delta\left(h_\delta^0, v_\delta\right) + b^{\mathcal{M}}(v_\delta, \lambda_\delta) = (f, v_\delta)_\delta + \left(h^N, v_\delta\right)_{\Gamma^N} - a_\delta\left(\mathcal{R}_\delta\left(h^D\right), v_\delta\right) & \forall v_\delta \in V_\delta \, , \\ b^{\mathcal{M}}\left(h_\delta^0, \psi_\delta\right) = -b^{\mathcal{M}}\left(\mathcal{R}_\delta\left(h^D\right), \psi_\delta\right) & \forall \psi_\delta \in M_\delta \, , \end{cases} \tag{6.9}$$

where V_δ is the global VEM space defined by (5.4), (5.5) and (5.6), $M_\delta = \Pi_{m\in\mathcal{M}} M_{\delta m}$, being $M_{\delta m} \subset \mathrm{L}^2(\Gamma_m)$ a finite dimensional space (tipically a piecewise polynomial space, see Section 6.3.3) and, $\forall v_\delta, w_\delta \in V_\delta$,

$$a_\delta(w_\delta, v_\delta) = \sum_{\substack{i\in\mathcal{I}\\ E\in\mathcal{T}_{\delta,i}}} \left(\mathsf{K}_i \Pi^0_{k-1}\nabla w_\delta, \Pi^0_{k-1}\nabla v_\delta\right)_E + S^E\left(w_\delta - \Pi^\nabla_k w_\delta, v_\delta - \Pi^\nabla_k v_\delta\right) ,$$
$$(f, v_\delta)_\delta = \sum_{i\in\mathcal{I}} \left(f_i, \Pi^0_{k-1} v_\delta\right)_{F_i} .$$

Finally we notice that, if $\psi_\delta, v_\delta \in \mathrm{L}^2(\Gamma_m)$,

$$b^{\mathcal{M}}(v_\delta, \psi_\delta) = \sum_{m\in\mathcal{M}} \left\langle \psi_{\delta m}, [\![v_\delta]\!]_{\Gamma_m}\right\rangle_{\pm\frac{1}{2},\Gamma_m} = \sum_{m\in\mathcal{M}} \left(\psi_{\delta m}, [\![v_\delta]\!]_{\Gamma_m}\right)_{\Gamma_m} ,$$

6.2.1 Well-posedness of the discrete problem. Following [9, Corollary 2.1], the well-posedness of problem (6.9) is guaranteed if a_δ is coercive on

$$W_\delta = \{v_\delta \in V_\delta : b(v_\delta, \psi_\delta) = 0 \quad \forall \psi_\delta \in M_\delta\} , \tag{6.10}$$

and an *inf-sup* condition holds:

$$\exists \beta > 0 : \inf_{\psi_\delta\in M_\delta} \sup_{v_\delta\in V_\delta} \frac{b^{\mathcal{M}}(v_\delta, \psi_\delta)}{\|v_\delta\|_V \|\psi_\delta\|_M} \geq \beta . \tag{6.11}$$

The existence of a constant β independent of δ satisfying (6.11) was proved in [4] making use of [10, Lemma 10] in the case of a polynomial Finite Element approximation on a regular triangulation. The same proof applies here under the following assumption.

Assumption 6.1. There exists a constant $\sigma > 0$ independent of δ such that, for each $E \in \tau_{\delta,i}$, for $i = 1, \ldots, N$, the distance between any two vertices of E is larger then or equal to σh_E, where h_E is the diameter of E.

Under this assumption, consider a trace Γ_m and a segment e belonging to the discretization of Γ_m. Let E be one of the two polygons sharing e. By Assumption 6.1, we can construct in the interior of E a triangle $T_{e,E}$ having e as one of its edges and having a shape regularity which depends uniquely on σ (for example, for convex elements, by connecting the extrema of e with the barycenter of E). The area of such a triangle scales as the area of E divided by the number of edges of E. We are thus led to make the following assumption.

Assumption 6.2. The number of edges of the elements of τ_δ is limited independently of δ.

With this last assumption, the area of $T_{e,E}$ scales like the area of E and thus, the norm of any function belonging to the finite dimensional space on $T_{e,E}$ is equivalent to the one on E. From [10, Lemma 10], we obtain the existence of an inf-sup constant independent of δ for $T_{e,E}$ and thus prove the existence of such a constant for E by the equivalence of the norms.

To prove the coercivity of a_δ on W_δ, we first prove the coercivity of a on such space and then use the equivalence (5.8). The key result needed is the following.

Proposition 6.1. *Assume that M_δ contains the functions which are constant on each trace. Then, the functional $v_\delta \mapsto |\!|\!| v_\delta |\!|\!|$ is a norm over W_δ.*

Proof. It is enough to verify that $\|\!|v_\delta|\!\| = 0$ only if $v_\delta = 0$. Let $v_\delta \in W_\delta$ be such that $\|\!|v_\delta|\!\| = 0$. Then it must be constant on each fracture, since its gradient on each fracture is null. Furthermore, v_δ clearly vanishes on all fractures such that $\Gamma_i^D \neq \emptyset$. It is now easy to prove that v_δ vanishes on all fractures. Indeed, let Γ_m be a trace shared by fractures F_i and F_j, with $\gamma_{\Gamma_m}(v_{\delta i}) = 0$; thanks to the mortar condition one has

$$\left(\left[\!\left[v_\delta\right]\!\right]_{\Gamma_m}, 1\right)_{\Gamma_m} = |\Gamma_m| \left[\!\left[v_\delta\right]\!\right]_{\Gamma_m} = 0 \Rightarrow \gamma_{\Gamma_m}(v_{\delta j}) = \gamma_{\Gamma_m}(v_{\delta i}) = 0$$

and since $v_{\delta i}$ and $v_{\delta j}$ are constant, it follows that $v_{\delta j} = 0$. Thanks to the network connectivity, this ensures that v_δ vanishes on all the fractures. □

From now on, M_δ is required to satisfy the assumption of Proposition 6.1. It follows that a is coercive with coercivity constant 1 on W_δ. By (5.8), a_δ is coercive with coercivity constant α_*.

6.2.2 A priori error estimates.

We are now able to derive an *a priori* error estimate. To this aim, we introduce the operators $\mathscr{F}, \mathscr{F}_\delta \in V'$ defined such that

$$\langle \mathscr{F}, v \rangle_{\pm 1} = (f, v)\,, \qquad\qquad \langle \mathscr{F}_\delta, v \rangle_{\pm 1} = (f, v)_\delta\,.$$

Furthermore, define

$$W_\delta^D = \left\{ v \in V_\delta^D \colon b^{\mathcal{M}}(v, \psi) = 0, \quad \forall \psi \in M_\delta \right\}, \tag{6.12}$$

$$\mathbb{P}_k^D(\Omega) = \left\{ p \in V_\delta^D \,:\, p \in \mathbb{P}_k(E), \quad \forall E \in \tau_\delta \right\}. \tag{6.13}$$

The main result concerning the a priori error estimate is stated in the following Theorem.

Theorem 6.1. *The solution* $(h_\delta, \lambda_\delta)$ *to problem* (6.9) *and the solution* (h, λ) *to problem* (6.4) *satisfy*

$$\begin{aligned}\|\!|h - h_\delta|\!\| \leq \left(1 + \frac{\alpha^*}{\alpha_*}\right) \inf_{v_\delta \in W_\delta^D} \|\!|h - v_\delta|\!\| + \frac{1+\alpha^*}{\alpha_*} \inf_{p_k \in \mathbb{P}_k^D(\Omega)} \|\!|h - p_k|\!\| \\ + \frac{1}{\alpha_*} \left(\inf_{\psi_\delta \in M_\delta} \sup_{v_\delta \in W_\delta} \frac{b^{\mathcal{M}}(v_\delta, \lambda - \psi_\delta)}{\|\!|v_\delta|\!\|} \right) + \frac{1 + C_\Omega}{\alpha_*} \|\mathscr{F} - \mathscr{F}_\delta\|_{V'}\,.\end{aligned} \tag{6.14}$$

Moreover, assume (6.11) *is satisfied. Then,*

$$\begin{aligned}\|\lambda - \lambda_\sigma\|_M \leq \left(1 + \frac{1}{\beta}\right) \inf_{\psi_\delta \in M_\delta} \|\lambda - \psi_\delta\|_M + \frac{\sqrt{\alpha^*}}{\beta} \|\!|h - h_\delta|\!\| \\ + \frac{1 + \sqrt{\alpha^*}}{\beta} \inf_{p_k \in \mathbb{P}_k^D(\Omega)} \|\!|h - p_k|\!\| + \frac{1}{\beta} \|\mathscr{F} - \mathscr{F}_\delta\|_{V'}\,.\end{aligned} \tag{6.15}$$

The proof follows the lines of proofs of [10, Theorem 3] and [1, Theorem 3.1]. We first prove the following preliminary result, which extends Poincaré's inequality to a DFN.

Lemma 6.1. *Let* $\tilde{W} = \left\{ v \in V \colon \int_S [\![v]\!] = 0 \quad \forall S \in \mathcal{S} \right\}$. *Then*

$$\exists C_\Omega > 0 \colon \forall w \in \tilde{W} \qquad \left(\sum_{i=1}^N \|w\|_{F_i}^2 \right)^{\frac{1}{2}} \leq C_\Omega \|\!|w|\!\| \tag{6.16}$$

Proof. First, notice that $\|\!|\cdot|\!\|$ is a norm on $\tilde{W}$ (see Proposition 6.1), thus the right hand side of (6.16) does not vanish, unless w is identically zero. By contradiction, suppose

$$\forall C > 0,\, \exists w_C \in \tilde{W} \colon \|w_C\| = \left(\sum_{i=1}^{N} \|w_C\|_{F_i}^2 \right)^{\frac{1}{2}} > C \, |\!|\!| w_C |\!|\!| \,,$$

then it is possible to build a sequence $w_k \in \tilde{W}$, $k \in \mathbb{N}$, of functions such that $\|w_k\| > k \, |\!|\!| w_k |\!|\!|$ and, without loss of generality, suppose that $\|w_k\| = 1$ for all k. Then, since $\|w_k\|_{1,F_i}$ is limited for all $i = 1, \dots, N$, w_k converges weakly in V to a function $w^\star$ up to sub-sequences. Clearly, ∇w_k converges to $\nabla w^\star$ weakly. Then, since

$$0 \leq \|\nabla w_k - \nabla w^\star\|_{F_i} = \|\nabla w_k\|_{F_i}^2 - 2\,(\nabla w_k, \nabla w^\star)_{F_i} + \|\nabla w^\star\|_{F_i}^2 \,,$$

and $\|\nabla w_k\|_{F_i} < \frac{1}{k}$, taking the limit for $k \to \infty$, it follows that $\|\nabla w^\star\|_{F_i} = 0$. Then, $w^\star$ is constant on each fracture. By the same arguments used in the proof of Proposition 6.1, it follows that $w^\star$ must be the null function. Moreover, since $\mathrm{H}^1(F_i)$ is compactly embedded in $\mathrm{L}^2(F_i)$, w_k converges strongly to $w^\star$ in $\mathrm{L}^2(F_i)$, for all $i = 1, \dots, N$. Since $\|w_k\|_{F_i} \overset{k\to\infty}{\longrightarrow} \|w^\star\|_{F_i}$ for all $i = 1, \dots, N$, we obtain $\|w^\star\| = 1$, which is a contradiction. □

We can now prove the a priori error estimate.

Proof of Theorem 6.1. Let $h_l \in W_\delta^D$ be the a-orthogonal projection of $h \in V^D$ over W_δ^D, such that

$$\forall v_\delta \in W_\delta^D, \qquad a(h - h_l, v_\delta) \, = 0 \,.$$

Exploiting the properties of the projection, we have

$$|\!|\!| h - h_\delta |\!|\!|^{\,2} = |\!|\!| h - h_l |\!|\!|^{\,2} + |\!|\!| h_l - h_\delta |\!|\!|^{\,2} = \left(\inf_{v_\delta \in W_\delta^D} |\!|\!| h - v_\delta |\!|\!| \right)^2 + |\!|\!| h_l - h_\delta |\!|\!|^{\,2} \,.$$

As far as the second term is concerned, recalling (5.8) we have

$$\alpha_* \, |\!|\!| h_l - h |\!|\!|^{\,2} = \alpha_* a(h_l - h, h_l - h) \leq a_\delta(h_l - h, h_l - h) \,.$$

By using the problem definitions (6.4) and (6.9), and introducing an arbitrary $p \in \mathbb{P}_k^D$, for which the polynomial consistency property $a(v_\delta, p) = a_\delta(v_\delta, p)$ holds for any given $v_\delta \in V_\delta$, we have

$$\begin{aligned}
&a_\delta(h_l - h_\delta, h_l - h_\delta) = a_\delta(h_l - p, h_l - h_\delta) + a_\delta(p, h_l - h_\delta) - a_\delta(h_\delta, h_l - h_\delta) = a_\delta(h_l - p, h_l - h_\delta) \\
&+ a(p, h_l - h_\delta) - (f, h_l - h_\delta)_\delta + b^{\mathcal{M}}(h_l - h_\delta, \lambda) - \left(h^N, h_l - h_\delta\right)_{\Gamma^N} = a_\delta(h_l - p, h_l - h_\delta) \\
&+ a(p - h, h_l - h_\delta) + a(h, h_l - h_\delta) - (f, h_l - h_\delta)_\delta + b^{\mathcal{M}}(h_l - h_\delta, \lambda_\delta) - \left(h^N, h_l - h_\delta\right)_{\Gamma^N} \\
&= a_\delta(h_l - p, h_l - h_\delta) + a(p - h, h_l - h_\delta) - (f, h_l - h_\delta)_\delta + (f, h_l - h_\delta) - b^{\mathcal{M}}(h_l - h_\delta, \lambda) \,,
\end{aligned}$$

where we have used that $b^{\mathcal{M}}(h_l - h_\delta, \lambda_\delta) = 0$ because $h_l - h_\delta \in W_\delta$. Introducing $\mathscr{F}$, $\mathscr{F}_\delta$ and

a generical $\psi_\delta \in M_\delta$, since $b^{\mathcal{M}}(h_l - h_\delta, \psi_\delta) = 0$ we have

$$a_\delta(h_l - h_\delta, h_l - h_\delta) = a_\delta(h_l - p, h_l - h_\delta) + a(p - h, h_l - h_\delta) - b^{\mathcal{M}}(h_l - h_\delta, \lambda)$$
$$+ {}_{V'}\langle \mathscr{F} - \mathscr{F}_\delta, h_l - h_\delta \rangle_V \leq \left(\alpha^* \left|\!\left|\!\left| h_l - p \right|\!\right|\!\right| + \left|\!\left|\!\left| h - p \right|\!\right|\!\right| + \frac{b^{\mathcal{M}}(h_l - h_\delta, \lambda - \psi_\delta)}{\left|\!\left|\!\left| h_l - h_\delta \right|\!\right|\!\right|} \right) \left|\!\left|\!\left| h_l - h_\delta \right|\!\right|\!\right|$$
$$+ \|\mathscr{F} - \mathscr{F}_\delta\|_{V'} \|h_l - h_\delta\|_V \leq \left(\alpha^* \left|\!\left|\!\left| h_l - p \right|\!\right|\!\right| + \left|\!\left|\!\left| h - p \right|\!\right|\!\right| + \frac{b^{\mathcal{M}}(h_l - h_\delta, \lambda - \psi_\delta)}{\left|\!\left|\!\left| h_l - h_\delta \right|\!\right|\!\right|} \right.$$
$$\left. + (1 + C_\Omega) \|\mathscr{F} - \mathscr{F}_\delta\|_{V'} \right) \left|\!\left|\!\left| h_l - h_\delta \right|\!\right|\!\right| ,$$

where in the last step inequality (6.16) has been used (see (6.5) for the definition of the V-norm). The proof of (6.14) is thus completed using the triangle inequality and suitably taking the supremums and infimums.

In order to prove (6.15), let us consider an arbitrary $\psi_\delta \in M_\delta$. By applying (6.11), (6.4) and (6.9) we get:

$$\beta \|\psi_\delta - \lambda_\delta\|_M \leq \sup_{v_\delta \in V_\delta} \frac{b^{\mathcal{M}}(v_\delta, \psi_\delta - \lambda_\delta)}{\|v_\delta\|_V} = \sup_{v_\delta \in V_\delta} \frac{b^{\mathcal{M}}(v_\delta, \lambda - \lambda_\delta) + b^{\mathcal{M}}(v_\delta, \psi_\delta - \lambda)}{\|v_\delta\|_V}$$
$$= \sup_{v_\delta \in V_\delta} \frac{a_\delta(h_\delta, v_\delta) - (f, v_\delta)_\delta - a(h, v_\delta) + (f, v_\delta) + b^{\mathcal{M}}(v_\delta, \psi_\delta - \lambda)}{\|v_\delta\|_V} .$$

Next, introducing an arbitrary $p \in \mathbb{P}_k^D(\Omega)$, by polynomial consistency we get

$$\beta \|\psi_\delta - \lambda_\delta\|_M \leq \sup_{v_\delta \in V_\delta} \|v_\delta\|_V^{-1} \left[a_\delta(h_\delta - p, v_\delta) + a(p - h, v_\delta) + {}_{V'}\langle \mathscr{F} - \mathscr{F}_\delta, v_\delta \rangle_V \right.$$
$$\left. + b^{\mathcal{M}}(v_\delta, \psi_\delta - \lambda) \right] \leq \sup_{v_\delta \in V_\delta} \|v_\delta\|_V^{-1} \left[\sqrt{a_\delta(h_\delta - p, h_\delta - p)} \sqrt{a_\delta(v_\delta, v_\delta)} \right] + \left|\!\left|\!\left| h - p \right|\!\right|\!\right|$$
$$+ \|\mathscr{F} - \mathscr{F}_\delta\|_{V'} + \|\lambda - \psi_\delta\|_M \leq \sqrt{\alpha^*} \left|\!\left|\!\left| h_\delta - p \right|\!\right|\!\right| + \left|\!\left|\!\left| h - p \right|\!\right|\!\right| + \|\mathscr{F} - \mathscr{F}_\delta\|_{V'}$$
$$+ \|\lambda - \psi_\delta\|_M \leq \sqrt{\alpha^*} \left|\!\left|\!\left| h - h_\delta \right|\!\right|\!\right| + (1 + \sqrt{\alpha^*}) \left|\!\left|\!\left| h - p \right|\!\right|\!\right| + \|\mathscr{F} - \mathscr{F}_\delta\|_{V'} + \|\lambda - \psi_\delta\|_M .$$

The proof is concluded by the triangle inequality and taking the infimum over $\mathbb{P}_k^D(\Omega)$. □

6.3 Implementation

We describe in this section some details concerning the practical implementation of the method.

6.3.1 Mesh generation and trace management. Following closely the ideas in [6], we start by independently introducing a good quality triangular mesh on each fracture, disregarding trace positions. Such triangulation will be called *base mesh*. On each fracture, the base mesh is then modified in such a way that a new polygonal mesh is obtained, that is locally conforming to the traces of the fractures. This means that traces will be covered by edges of the new polygonal elements, though we remark that elements on meshes from different fractures induce a different discretization of the same trace. This new mesh will be suitable for the application of the method described in the previous sections and it will be called *VEM mesh*. The procedure for obtaining the VEM mesh is the following. Whenever a trace intersects an edge of the triangulation, a new node is created at the intersection. Each

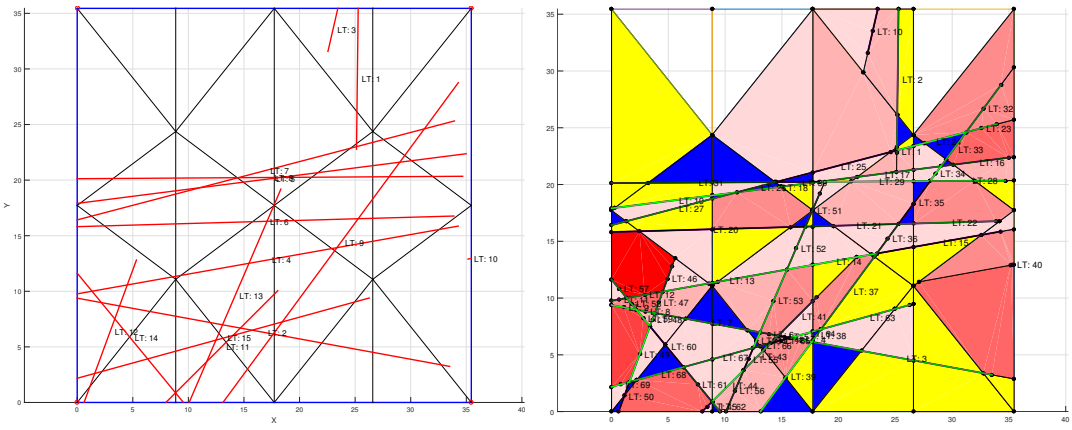


Figure 6.1: Mesh examples. Left: base mesh; right: VEM mesh.

trace tip defines a new node and the trace segment is prolonged up to the nearest edge of the triangulation, thereby creating a new edge and a new node. When two traces intersect each other, they are split into two sub-traces and in their intersection a new node is created. Whenever an element of the mesh is cut by a (possibly prolonged) trace segment, it is split into two parts which become new elements of the polygonal mesh in their own right. Finally, traces without internal nodes receive the addition of a new node in its midpoint, which is necessary to define the discrete Mortar space for the trace. The overall procedure thus results in a polygonal mesh whose elements are convex polygons made of an arbitrary number of edges.

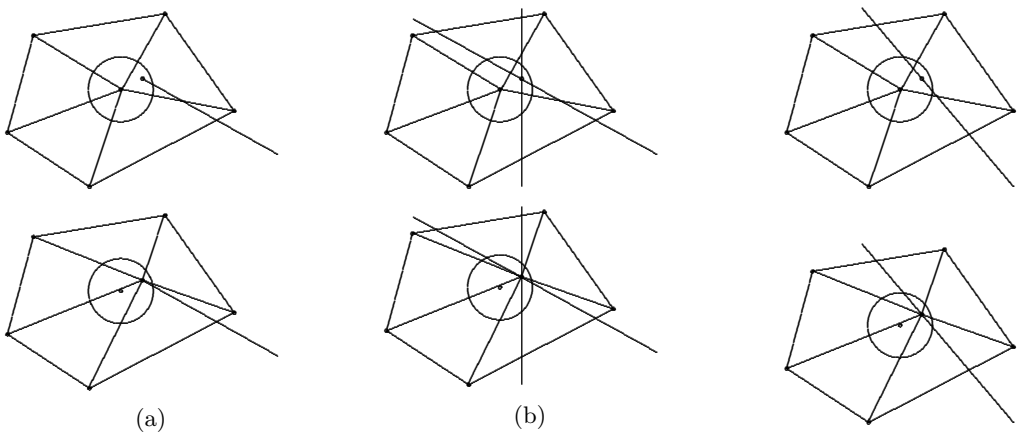
Figure 6.1 is illustrative for such procedure. Focusing on a single fracture, we depict on the left the base mesh introduced, and the local traces present on the fracture, denoted by LT and with a fracture-local numbering from 1 to 15. On the right, the VEM mesh obtained is represented. Note that new traces are introduced by splitting the original traces into sub-traces. Note, as well, the generation of new nodes and elements obtained via trace segment prolongation and the addition of one internal node (see, e.g., the original local trace 3 on the top of the fracture). To better highlight the number of edges in the elements, a different coloring is used for elements with a different number of edges.

Remark 6.2. In order to verify Assumption 6.1, a mesh smoothing process can be designed, in order to improve the quality of the VEM mesh, reduce the number of DOFs and prevent irregular elements in the discretization. Let us introduce for each vertex a quantity r_{m} called *moving radius*, defined as a fixed rate of the smallest edge connected to that vertex. Correspondingly, we define a *moving ball* as a ball with center the vertex and radius r_{m}. Then:

1. if a trace tip lies within a moving ball of a vertex, the vertex is moved on the tip (see Figure 6.2a);

2. if the intersection between two traces is within the moving ball of a vertex not previously moved to a tip, the latter is moved on the intersection (see Figure 6.2b);

3. if a vertex not previously moved is closer to a trace than the moving radius, it is moved orthogonally onto the trace (see Figure 6.2c).



(a) (b) (c)

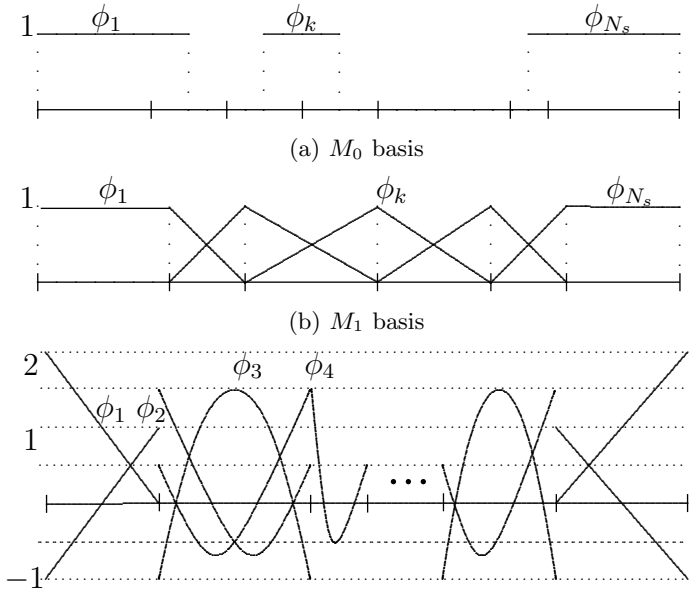
Figure 6.2: Mesh smoothing process. Top: before mesh smoothing; bottom: after mesh smoothing. Cases: (a) trace tip close to a vertex; (b) traces intersecting close to a vertex; (c) trace very close to a vertex.

This procedure does not cover the case in which two traces intersect each other with a very small angle or very small traces, but from the numerical results (see, in particular, Section 6.4.2) we can say that the method is sufficiently robust to deal with this kind of issues.

Remark 6.3. Assumption 6.2 is satisfied by the VEM mesh. Indeed, the triangles of the base mesh are only split when a trace cuts them. Thus, the number of edges of the new polygonal elements is limited by the number of traces cutting the element (that is bounded by the number of traces on the fracture), plus 3.

6.3.2 Matrix Formulation of the problem. On the discretization of Γ_m induced by the triangulation on the non-mortar fracture, we introduce a finite dimensional subspace of dimension N_{Γ_m}, containing the constant functions (this is required for well-posedness, see Proposition 6.1). Let N_h and N_λ be the total number of degrees of freedom for h_δ and λ_δ, respectively, and set $N_{\text{dof}} = N_h + N_\lambda$; let us denote by ϕ_k, $k = 1, \dots, N_h$, and ψ_l, $l = 1, \dots, N_\lambda$, the basis functions for h_δ and λ_δ, respectively. Finally, let N^D be the number of basis functions ϕ_j^D used to define the lifting $\mathcal{R}_\delta\big(h^D\big)$ of the Dirichlet boundary condition. Then, problem (6.9) can be written as

$$\begin{cases} \displaystyle\sum_{j=1}^{N_h} a_\delta(\phi_j, \phi_k)\, h_j + \sum_{l=1}^{N_\lambda} b^{\mathcal{M}}(\phi_k, \psi_l)\, \lambda_l = (f, \phi_k)_\delta + \big(H^N, \phi_k\big)_{\Gamma^N} \\ \qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad - \displaystyle\sum_{j=1}^{N_D} a_\delta\big(\phi_j^D, \phi_k\big)\, h_j^D \\ \sum_{j=1}^{N_h} b^{\mathcal{M}}(\phi_j, \psi_m)\, h_j = -\sum_{j=1}^{N_D} b^{\mathcal{M}}\big(\phi_j^D, \psi_m\big)\, h_j^D \end{cases}$$



(a) M_0 basis

(b) M_1 basis

(c) M_2 basis

Figure 6.3: Lagrange multiplier basis

$\forall k = 1, \dots, N_h$ and $\forall m = 1, \dots, N_\lambda$, where h_j^D is the value of $\Pi^0_{k,\Gamma^D}(H^D)$ at the boundary node corresponding to ϕ_j^D. Summarizing, we have to solve the system

$$\begin{pmatrix} A \in \mathbb{R}^{N_h,N_h} & B \in \mathbb{R}^{N_h,N_\lambda} \\ B^\intercal \in \mathbb{R}^{N_\lambda,N_h} & O \in \mathbb{R}^{N_\lambda,N_\lambda} \end{pmatrix} \begin{pmatrix} \mathbf{h} \\ \boldsymbol{\lambda} \end{pmatrix} = \begin{pmatrix} \mathbf{F} \\ \boldsymbol{\Psi} \end{pmatrix}, \tag{6.17}$$

where

$$A_{kj} = a_\delta(\phi_k, \phi_j)\,, \qquad B_{jl} = b^{\mathcal{M}}(\phi_j, \psi_l)$$

$$F_k = (f, \phi_k)_\delta + \left(H^N, \phi_k\right)_{\Gamma^N} - \sum_{j=1}^{N_D} a_\delta\left(\phi_j^D, \phi_k\right) h_j^D, \qquad \Psi_m = -\sum_{j=1}^{N_D} b^{\mathcal{M}}\left(\phi_j^D, \psi_m\right) h_j^D.$$

For the practical construction of the VEM stiffness matrix and right hand side vector, we refer the reader to [2]. We remark that the construction of the matrix B can be done by standard quadrature formulas, since the analytical expression of the basis functions on the edges of each element is known.

6.3.3 Bases for the discrete Lagrange multipliers. In this subsection we give details about the choice adopted for the space $M_{\delta,S}$, for each $S \in \mathcal{S}$. For a thorough description of the possible choices of Mortar bases, we refer the reader to [11].

In this work we have used three bases: the basis M_0, composed by piecewise constant functions; the basis M_1, given by continuous piecewise linear functions, except for the first and last intervals on which the functions are taken constant; the basis M_2, given by discontinuous piecewise quadratic functions, except for the first and last interval where the functions are linear. These bases are depicted in Figure 6.3.

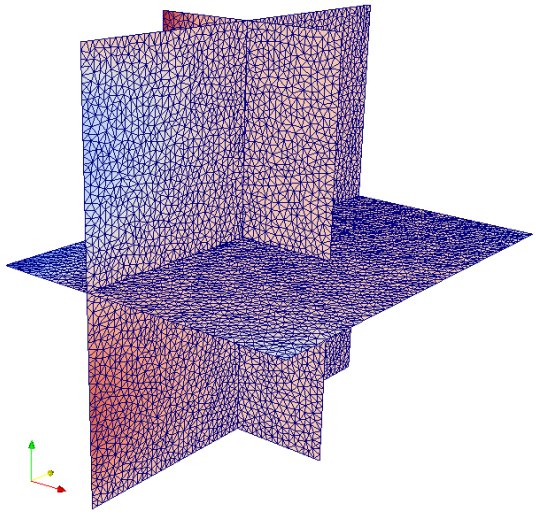


Figure 6.4: **Benchmark problem** Geometry of the network

6.4 Numerical results

We present in this section some numerical results aimed at assessing the practical behavior of the method. The results are obtained on two classes of problems: firstly, we present a benchmark problem for which the exact solution is known, with some convergence results; secondly, we analyse the performance of the method on larger DFNs that introduce several geometrical complexities. All the numerical results here reported are obtained without any kind of mesh smoothing (see Remark 6.2), in order to test the robustness of the method.

6.4.1 Benchmark problem. The benchmark DFN consists of 3 fractures as shown in Figure 6.4. Despite being a simple network, it presents two geometrical features (a trace intersection and a trace tip) which make it worthwhile to analyse the behavior of the method at tackling them. The computational domain $\Omega = F_1 \cup F_2 \cup F_3$ is defined by

$$
\begin{aligned}
F_1 &= \left\{(x,y,z) \in \mathbb{R}^3 : -1 \le x \le 1/2,\ -1 \le y \le 1,\ z = 0\right\},\\
F_2 &= \left\{(x,y,z) \in \mathbb{R}^3 : -1 \le x \le 0,\ y = 0,\ -1 \le z \le 1\right\},\\
F_3 &= \left\{(x,y,z) \in \mathbb{R}^3 : x = -1/2,\ -1 \le y \le 1,\ -1 \le z \le 1\right\},
\end{aligned}
$$

with traces

$$
\begin{aligned}
\Gamma_1 = F_1 \cap F_2 &= \left\{(x,y,z) \in \mathbb{R}^3 : -1 \le x \le 1/2,\ y = 0,\ z = 0\right\},\\
\Gamma_2 = F_1 \cap F_3 &= \left\{(x,y,z) \in \mathbb{R}^3 : x = -1/2,\ -1 \le y \le 1,\ z = 0\right\},\\
\Gamma_3 = F_2 \cap F_3 &= \left\{(x,y,z) \in \mathbb{R}^3 : x = -1/2,\ y = 0,\ -1 \le z \le 1\right\}.
\end{aligned}
$$

The problem is defined setting non-homogeneous Dirichlet boundary conditions on the whole boundary $\partial\Omega$, and a load term on each fracture in such a way that the exact solution

is given by:

$$h_1(x,y) = \frac{1}{10}\left(-x-\frac{1}{2}\right)\left(8xy\left(x^2+y^2\right)\operatorname{arctan2}(y,x)+x^3\right),$$
$$h_2(x,z) = \frac{1}{10}\left(-x-\frac{1}{2}\right)x^3 - \frac{4}{5}\pi\left(-x-\frac{1}{2}\right)x^3\,|z|\,,$$
$$h_3(y,z) = (y-1)y(y+1)(z-1)z,$$

where $\operatorname{arctan2}(y,x)$ is the four quadrant inverse tangent function with 2 arguments, that returns the appropriate quadrant of the computed angle y/x. Note that since $H_1, H_2 \notin \mathrm{C}^1$, a net flux is expected between F_1 and F_2.

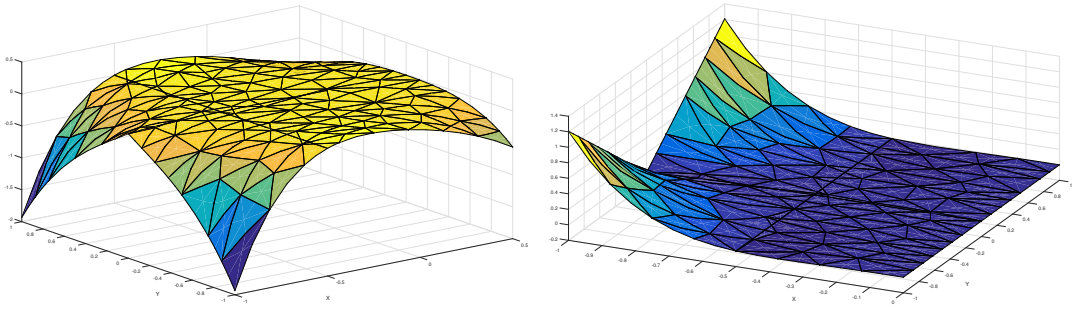


Figure 6.5: **Benchmark problem** Computed hydraulic head on fractures F_1 (left) and F_2 (right).

The computed solutions obtained for the hydraulic head on such fractures are shown in Figure 6.5. Fluxes exchanged between F_1 and F_2, computed with all three considered choices for the mortar bases are shown in Figure 6.6, where they are compared with the exact one.

In order to present convergence results, we remark that since the values of the discrete solution are not explicitly known inside the elements but only on the set of DOFs, the errors are computed by projecting the discrete solution on the space of polynomials of degree k, as is the usual procedure with the VEM [3]:

$$\left(Err^H_{\mathrm{L}^2}\right)^2 = \sum_{E\in\mathcal{T}_\delta} \left\|h - \Pi^\nabla_k h_\delta\right\|^2_E\,,$$
$$\left(Err^H_{\mathrm{H}^1}\right)^2 = \sum_{E\in\mathcal{T}_\delta} \left\|h - \Pi^\nabla_k h_\delta\right\|^2_{1,E}\,,$$

Regarding the errors of approximation of λ, we measure them on each trace both in $\mathrm{L}^2(\Gamma_m)$ and $\mathrm{H}^{-\frac{1}{2}}(\Gamma_m)$ norm; for practical computational issues, we approximate this latter norm with a weighted $\mathrm{L}^2(\Gamma_m)$ norm:

$$\left(Err^\Lambda_{\mathrm{L}^2}\right)^2 = \sum_{m\in\mathcal{M}} \|\lambda-\lambda_\delta\|^2_e\,,$$
$$\left(Err^\Lambda_{\mathrm{H}^{-\frac{1}{2}}}\right)^2 = \sum_{S\in\mathcal{S}}\sum_{e\subset S} |e|\,\|\lambda-\lambda_\delta\|^2_e\,.$$

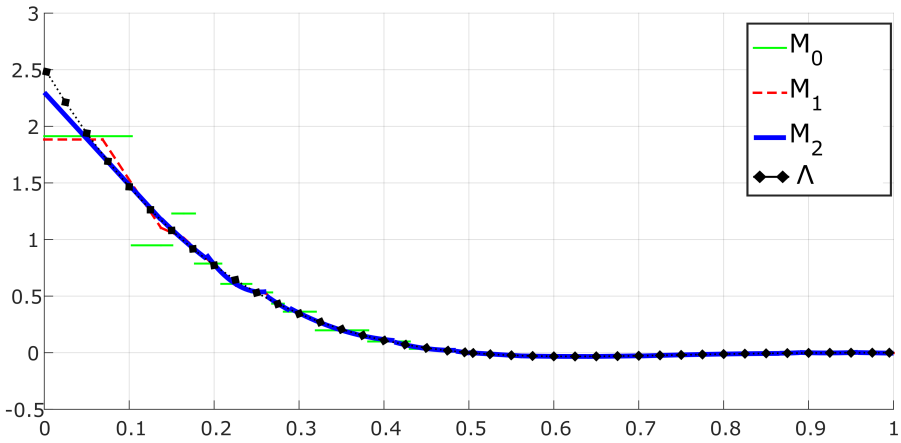


Figure 6.6: **Benchmark problem** Computed and exact fluxes

		h_δ		λ_δ on Γ_1	
VEM order	Mortar basis	L^2 Norm	H^1 Norm	L^2 Norm	$H^{-\frac{1}{2}}$ Norm
1	M_0	1.00 (1)	0.50 (0.5)	1.19	1.79
1	M_1	1.00 (1)	0.50 (0.5)	1.26	1.87
2	M_0	1.38 (1.5)	0.91 (1)	0.98	1.54
2	M_1	1.50 (1.5)	1.01 (1)	1.54	2.05
2	M_2	1.51 (1.5)	1.01 (1)	2.45	3.02

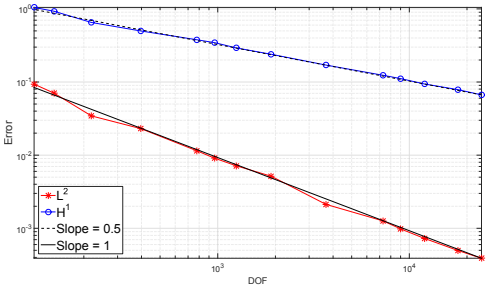
Table 6.1: **Benchmark problem** Convergence rates for several VEM orders and Mortar bases. The numbers in parentheses indicate the expected rates.

In Figure 6.7, focusing on fracture F_1, we present the convergence curves for different combinations of the order k for the VEM space and of the type of Mortar basis. Namely, in the left column we report the behavior of the errors $Err^H_{L^2}$ and $Err^H_{H^1}$ (labeled by L^2 and H^1, respectively); the errors are plot versus the total number of h-DOFs on the fracture. In the right column we report the errors $Err^\Lambda_{L^2}$ and $Err^\Lambda_{H^{-\frac{1}{2}}}$ (labeled by L^2 and $H^{-\frac{1}{2}}$, respectively); here, the errors are plot versus the number of λ-DOFs on the traces of F_1.

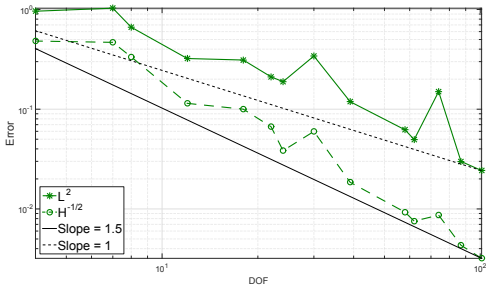
Finally, Table 6.1 reports, for all the analysed cases, the computed convergence rates with respect to the number of DOFs. Namely, we report the computed rates of convergence for h with respect to the h-DOFs (the expected values being reported in parentheses); note the very good agreement between the computed and the expected rates, except for the case $k = 2$ and M_0, in which the low order of the mortar basis slows down the rate of convergence for the hydraulic head. Focusing on trace Γ_1, we also report the computed rates of convergence for λ_δ with respect to the number of λ-DOFs. The rates of convergence for the λ-errors with respect to the number of h-DOFs, not listed here, are approximately one half of the reported values; this is in agreement with the fact that the number of λ-DOFs scales as the square root of the number of h-DOFs.

6.4.2 Complex networks.

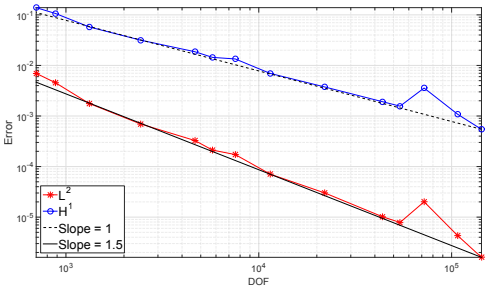
In this section we present results obtained on more complex networks. The first one, DFN36, consists of 36 fractures. The geometry of the DFN is depicted in Figure 6.8, from which the geometrical complexity of the domain can be seen. A



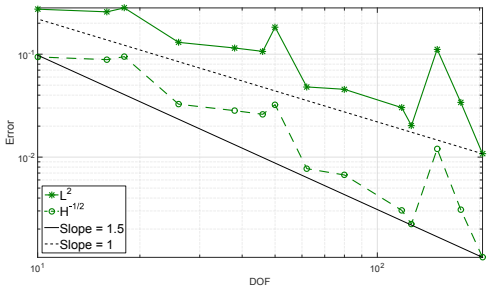
(a) Order 1 - M_0 - Error in h



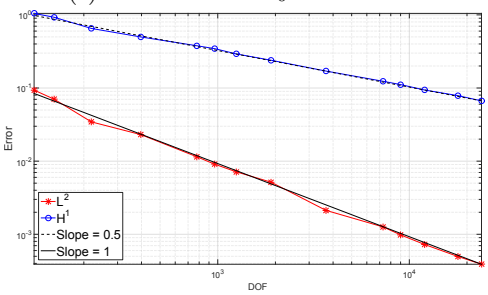
(b) $k = 1$ - Basis M_0 - Error in λ



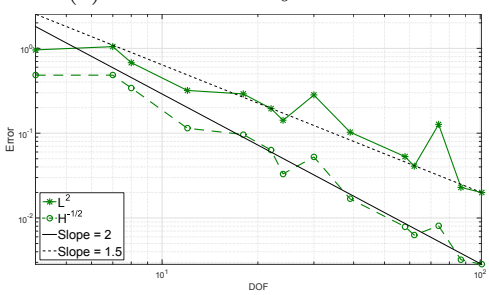
(c) $k = 2$ - Basis M_0 - Error in h



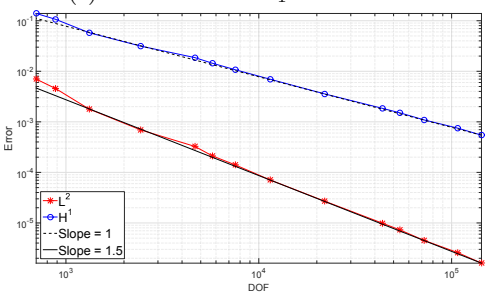
(d) $k = 2$ - Basis M_0 - Error in λ



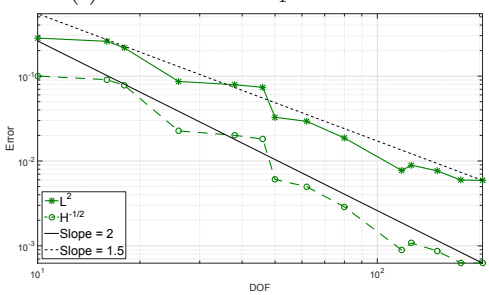
(e) $k = 1$ - Basis M_1 - Error in h



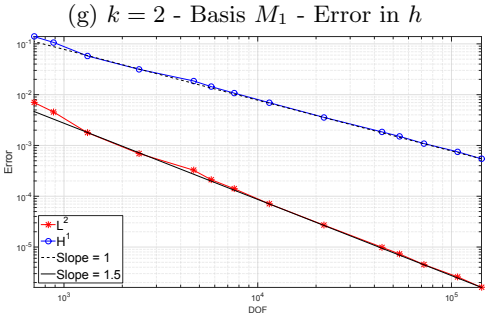
(f) $k = 1$ - Basis M_1 - Error in λ



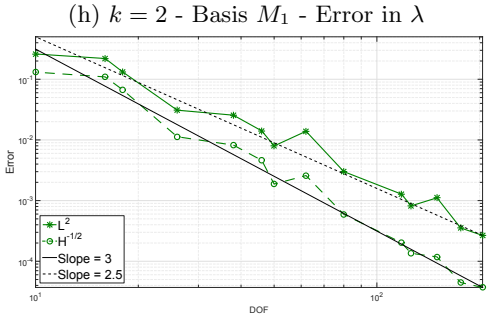
(g) $k = 2$ - Basis M_1 - Error in h



(h) $k = 2$ - Basis M_1 - Error in λ



(i) $k = 2$ - Basis M_2 - Error in h



(j) $k = 2$ - Basis M_2 - Error in λ

Figure 6.7: **Benchmark problem** Convergence curves measured on fracture F_1

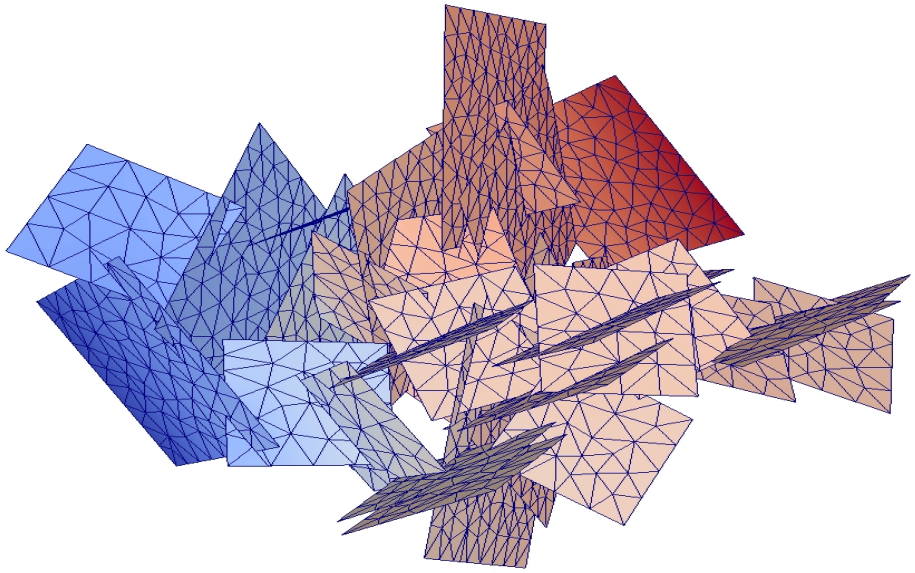


Figure 6.8: **DFN36** Geometry of the network and computed hydraulic head (as a scale of colours)

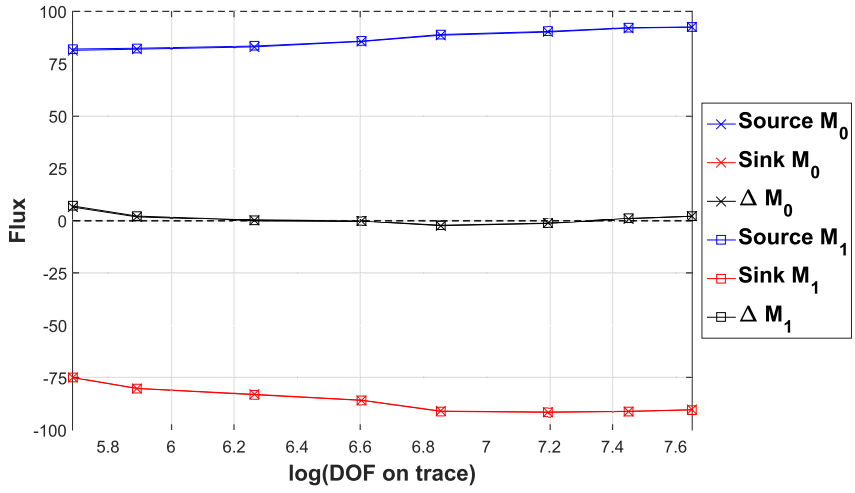
non-homogeneous constant Neumann boundary condition ($h^N = 100$) has been set on one fracture (called source fracture), and a homogeneous Dirichlet boundary condition has been set on another fracture (sink fracture). Homogeneous Neumann boundary conditions on the remaining part of the boundary isolate all the other fractures from the surrounding medium.

The plots in Figure 6.9 report the computed total net fluxes exchanged by the source and sink fracture versus the number of DOFs on traces (logaritmic x-scale), for VEM of order $k = 1, 2$ and 3, and mortar bases M_0, M_1 and M_2. The value Δ reported is the difference between the two curves and is an indication of the global conservation state of the method in the whole DFN. Results show the tendency to approximate the expected values and we note that, interestingly, almost no difference in flux values is appreciated for different choices of mortar bases. As a further quality indicator for the obtained solution, we introduce a measure of the error of the jump of the hydraulic head on traces. Namely, we set

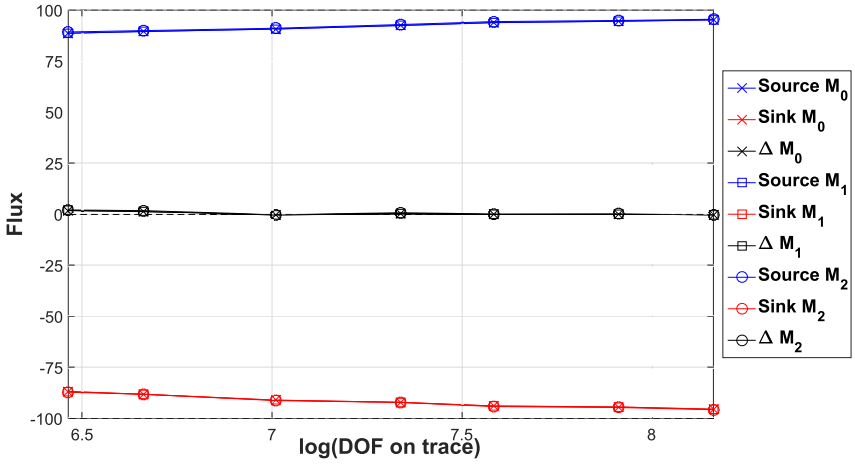
$$E_h = \sum_{m \in \mathcal{M}} \left\| [\![h]\!]_{\Gamma_m} \right\|^2_{\Gamma_m}.$$

The computed values are shown in Figure 6.10 for VEM of order $k = 1, 2$ and 3, using the basis M_1. For all orders, a decrease in this parameter was observed with increasing number of DOFs as expected, but interestingly, with a similar rate. Since the defined quantity does not constitute a norm, no further conclusions about convergence can be drawn.

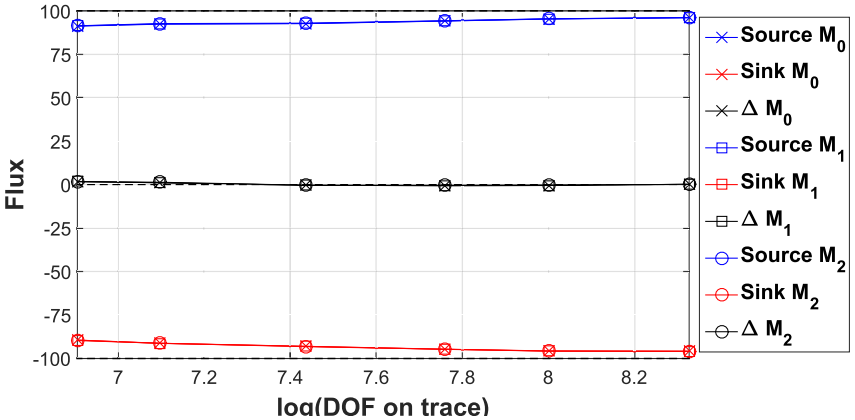
As a second example, a 134 fracture network is proposed (DFN134, Figure 6.11). As far as geometrical complexities are concerned, this DFN is far more challenging than DFN36, as it exhibits several critical features: very small angles at trace intersections (thus challenging the shape regularity of the elements stated by Assumption 6.1 and discussed in Remark 6.2),



(a) $k = 1$



(b) $k = 2$



(c) $k = 3$

Figure 6.9: **DFN36** Flux results

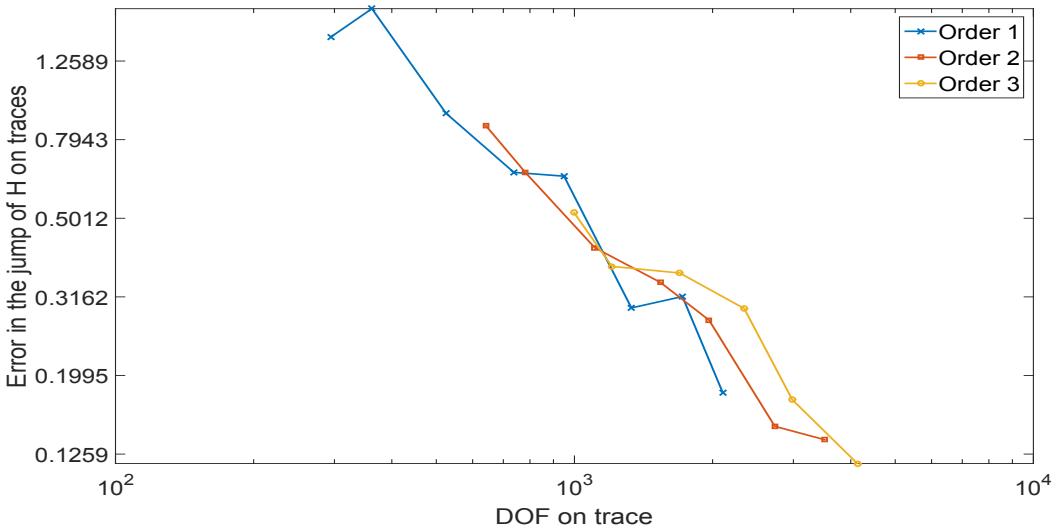


Figure 6.10: **DFN36** Error in the jump of the hydraulic head on traces

almost parallel traces, large variation of trace lengths and fracture sizes. Three fractures were chosen as source fractures by imposing non-homogeneous Neumann boundary conditions. A fourth fracture was set as sink fracture, and on one of its edges a homogeneous Dirichlet boundary conditions was set. Homogeneous Neumann conditions were imposed on all the remaining components of the boundary.

In Figure 6.12 we report some data for a particularly intricate fracture, where the problem has been solved using VEM of order $k = 2$ and the M_1 basis. The VEM mesh is presented (top left figure), as well as the affine interpolation of the computed hydraulic head solution (bottom left) and the corresponding velocity field obtained from the gradients of the computed hydraulic head (top right figure). From the detail reported in the bottom right figure, it can be seen how elements of order 2 allow for a better representation of the change in slope between close traces thanks to the added DOFs in the midpoints of each of the edges.

6.5 Conclusions

We have introduced a new approach for flow simulations in Discrete Fracture Networks. The key feature is given by its capability to work with arbitrary (good quality) meshes generated on the fractures. Taking advantage of the versatility of the Virtual Element Method in handling polygonal meshes, each arbitrary mesh is easily modified in such a way that local conformity of the meshes is obtained for almost any trace disposition. Using the hybrid formulation of the Mortar method, only "weak" continuity is required for the hydraulic head along the intersections between fractures.

The main advantage of the approach presented here, with respect to the method proposed in [7], is that, besides the computation of the hydraulic head, the present approach allows for a direct approximation of the flux on each trace, whereas in [7] the flux exchange is derived from the values of the hydraulic head.

The validity of the approach proposed is supported by numerical experiments, showing

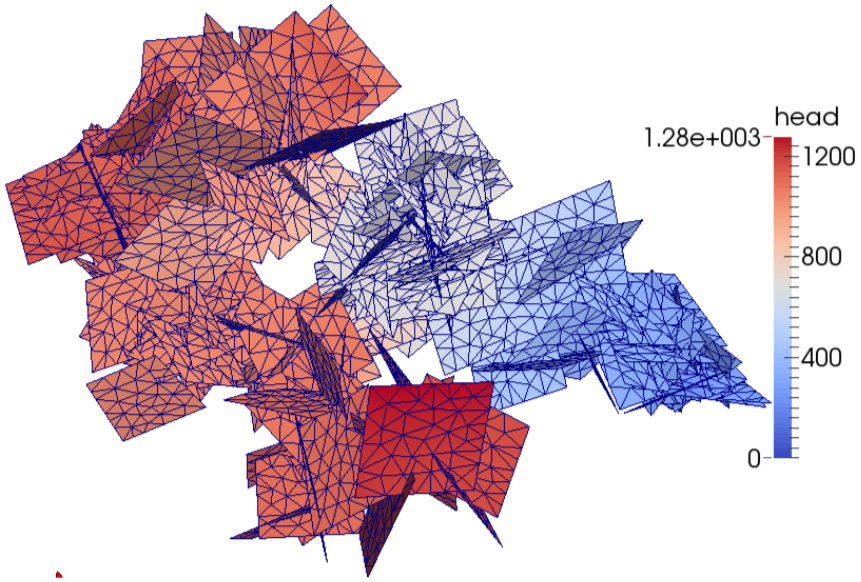


Figure 6.11: **DFN134** Geometry of the network and computed hydraulic head (as a scale of colours)

optimal convergence for the primal variable; furthermore, the behaviour of the method is quite satisfactory also when it is applied to DFNs with complex geometry.

Future developments include the extension to more complex flow models and in particular to the case of non-constant transmissivity values. Furthermore, we aim at investigating a possible parallel implementation, which is recommended for tackling large scale DFNs for realistic underground flow simulations.

References for Chapter 6

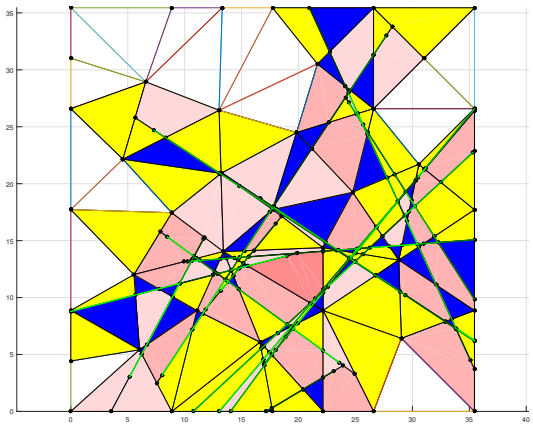
[1] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. D. Marini, and A. Russo. "Basic principles of virtual element methods". In: *Mathematical Models and Methods in Applied Sciences* 23.01 (2013), pp. 199–214.

[2] L. Beirão Da Veiga, F. Brezzi, L. D. Marini, and A. Russo. "The hitchhiker's guide to the Virtual Element method". In: *Math. Models Methods Appl. Sci* 24.8 (2014), pp. 1541–1573.

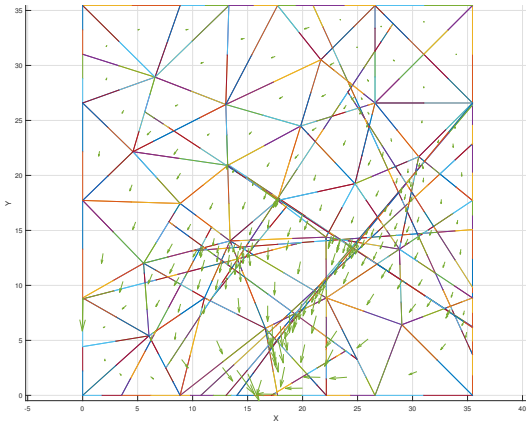
[3] L. Beirão da Veiga, F. Brezzi, L. D. Marini, and A. Russo. "Virtual Element Methods for General Second Order Elliptic Problems on Polygonal Meshes". In: *Mathematical Models and Methods in Applied Sciences* 26.04 (2015), pp. 729–750. DOI: 10.1142/S0218202516500160.

[4] F. B. Belgacem. "The mortar finite element method with Lagrange multipliers". In: *Numerische Mathematik* 84.2 (1999), pp. 173–197.

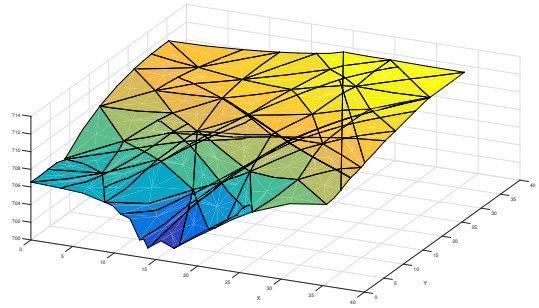
[5] M. Benedetto, S. Berrone, A. Borio, S. Pieraccini, and S. Scialò. "A Hybrid Mortar Virtual Element Method For Discrete Fracture Network Simulations". In: *J. Comput. Phys.* 306 (2016), pp. 148–166. DOI: 10.1016/j.jcp.2015.11.034.



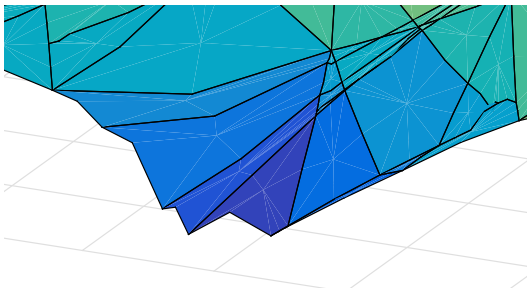
(a) Mesh



(b) Velocity field solution



(c) Hydraulic head solution



(d) Detail of hydraulic head solution

Figure 6.12: **DFN134** A selected fracture

[6] M. Benedetto, S. Berrone, S. Pieraccini, and S. Scialò. "The virtual element method for discrete fracture network simulations". In: *Comput. Methods Appl. Mech. Engrg.* 280.0 (2014), pp. 135–156. ISSN: 0045-7825. DOI: 10.1016/j.cma.2014.07.016.

[7] M. Benedetto, S. Berrone, and S. Scialò. "A Globally Conforming Method For Solving Flow in Discrete Fracture Networks Using the Virtual Element Method". In: *Finite Elem. Anal. Des.* 109 (2016), pp. 23–36. DOI: 10.1016/j.finel.2015.10.003.

[8] C. Bernardi, Y. Maday, and A. T. Patera. "A new nonconforming approach to domain decomposition: the mortar element method". In: *Nonlinear partial differential equations and their applications. Collège de France Seminar, Vol. XI (Paris, 1989–1991).* Vol. 299. Pitman Res. Notes Math. Ser. Longman Sci. Tech., Harlow, 1994, pp. 13–51.

[9] F. Brezzi. "On the existence, uniqueness and approximation of saddle-point problems arising from lagrangian multipliers". In: *Revue française d'automatique, informatique, recherche opérationelle. Analyse numérique* 8.2 (1974), pp. 129–151.

[10] P. Raviart and J. Thomas. "Primal hybrid finite element methods for 2nd order elliptic equations". In: *Mathematics of computation* 31.138 (1977), pp. 391–413.

[11] B. I. Wohlmuth. *Discretization Methods And Iterative Solvers Based On Domain Decomposition.* Berlin: Springer, 2001.

Chapter 7

Orthogonal polynomials in badly shaped polygonal elements for the Virtual Element Method

This chapter addresses the issue of containing the numerical oscillations arising on badly shaped polygons when using a Virtual Element formulation, with particular focus on the applications to DFN simulations described in Chapter 5. This work is published in [12].

7.1 Introduction

The flexibility of the recently developed Virtual Element Method (VEM) has been applied in the field of geological poro-fractured media [7–11]. Geosciences very often produce applications with huge domains and terrific geometrical complexities. Within this context, the Discrete Fracture Network (DFN) model was developed for modeling the flow in the geological fractured media [1, 20, 22, 27] and is object of a very large numerical bibliography [21, 23–26, 28–33]. Due to the huge uncertainty in the definition of the underground fracture distribution, this model instantiates a fracture distribution by a stochastic procedure starting from probabilistic distributions of geometrical parameters: direction, dimension, aspect ratio; and from probabilistic distributions of thickness and other hydro-geological properties. The stochastic procedure that instantiates the fracture distribution can create geometrical complexities arbitrarily demanding for a numerical method; typically these complexities are related, for example, to very small angles between couple of fractures, to a huge variability in the length of fracture-intersections, and to disjoint fractures very close to each other [19]. The VEM applied to this problem has proved a good reliability in dealing with these complexities, but, sometimes, some fracture configurations have lead to unfeasible numerical solutions [11]. A possible solution, sometimes viable, is to relax the mesh conformity requirement, resorting to the Mortar fracture matching method described in Chapters 5 and 6 [8] or applying a preliminary *mesh smoothing process* [10]. Nonetheless, some very badly shaped configurations cannot be avoided, mainly on coarse meshes.

Starting from these observations, in this paper we propose a different basis for assembling the local linear systems within the VEM, that, at a very small additional cost with respect to a classical implementation based on monomials, can largely improve the reliability of the

method by limiting the condition numbers of local matrices in badly shaped elements. We remark that the proposed method aims at improving the reliability of the computations performed in the set up of the consistent part of the VEM formulation of the problem and is completely independent of the VEM stabilization that is added to the consistent part in order to get a well posed problem [6]. Moreover, our description is organized in such a way that it can be easily plugged in a standard VEM code based on scaled monomials.

In Section 7.2 we recall the VEM formulation for a advection-diffusion-reaction problems, as described in Chapter 2. In Section 7.3 we introduce the computation of a quasi-orthogonal polynomial basis for assembling the VEM linear system that is fully compatible with the traditional monomial basis. The two bases can be mixed on elements in the same mesh using the quasi-orthogonal basis on badly shaped elements and the traditional monomial basis on all the other elements. In Section 7.4 we provide a brief validation of the modified VEM construction on a general reaction-convection-diffusion problem with variable coefficients. In Section 7.6 we compare the results provided by the classical monomial basis with the presented quasi-orthogonal basis on two critical Discrete Fracture Networks. In this last section we further discuss some simple criteria useful to determine in which elements it is beneficial to resort to the new basis and in which elements it is safe to use the monomial basis, as well as some limitations of the proposed approach.

7.2 Model problem: VEM for advection-diffusion-reaction equations

Following [5] and Chapter 2, we consider the general second order problem

$$\begin{cases} -\nabla \cdot (\mathsf{K}\nabla u) + \beta \cdot \nabla u + \gamma u = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

whose variational formulation reads

$$(\mathsf{K}\nabla u, \nabla v) + (\beta \cdot \nabla u, v) + (\gamma u, v) = (f, v) \ .$$

The VEM discretization of the problem consists in partitioning the domain using a mesh $\mathcal{T}_\delta$ made up of open star-shaped polygons having an arbitrary number of sides (even different from one polygon to another, but uniformly bounded independently of the mesh size). We make the following regularity assumption: $\exists \gamma > 0$ such that $\forall E \in \mathcal{T}_\delta$, with diameter h_E, E is star-shaped with respect to a ball of radius larger than γh_E. Let us define the functional spaces

$$\begin{aligned} V_\delta^E = \big\{ & v_\delta \in \mathrm{H}^1(E) : \Delta v_\delta \in \mathbb{P}_k(E),\, v_\delta \in \mathbb{P}_k(e)\ \forall e \subset \partial E,\, \gamma_{\partial E}(v_\delta) \in \mathrm{C}^0(\partial E), \\ & (v_\delta, p)_E = \left(\Pi_k^\nabla v_\delta, p\right)_E \ \forall p \in \mathbb{P}_k(E)/\mathbb{P}_{k-2}(E) \big\} , \\ V_\delta = \big\{ & v_h \in \mathrm{C}^0(\Omega) : v_h \in V_\delta^E \quad \forall E \in \mathcal{T}_\delta \big\} , \end{aligned}$$

and the following discrete counterpart of the bilinear form, which is computable from the VEM degrees of freedom (see Definition 2.1). Let

$$\begin{aligned} a_\delta(u_\delta, v_\delta) &= \left(\mathsf{K}\Pi_{k-1}^0 \nabla u_\delta, \Pi_{k-1}^0 \nabla v_\delta\right) + S\left(\left(I - \Pi_k^\nabla\right) u_\delta, \left(I - \Pi_k^\nabla\right) v_\delta\right) , \\ b_\delta(u_\delta, v_\delta) &= \left(\beta \cdot \Pi_{k-1}^0 \nabla u_\delta, \Pi_{k-1}^0 v_\delta\right) , \\ c_\delta(u_\delta, v_\delta) &= \left(\Pi_{k-1}^0 u_\delta, \Pi_{k-1}^0 v_\delta\right) , \\ \mathscr{B}_\delta(u_\delta, v_\delta) &= a_\delta(u_\delta, v_\delta) + b_\delta(u_\delta, v_\delta) + c_\delta(u_\delta, v_\delta) , \end{aligned}$$

where S is the VEM stabilization [3, 6] such that

$$\exists c_*, c^* > 0 \colon \forall v_\delta \in \ker\left(\Pi_k^\nabla\right),\ c_* \left\|\nabla v_\delta\right\|^2 \leq S(v_\delta, v_\delta) \leq c^* \left\|\nabla v_\delta\right\|^2 ,$$

and all the other terms of the operator $\mathscr{B}_\delta$ provide the consistent part of the operator. Within these terms, the operator Π^0_{k-1} is the elementwise $\mathrm{L}^2(E)$ projection on $\mathbb{P}_{k-1}(E)$, for any $E \in \mathcal{T}_\delta$. For the ease of notation, we will use the same symbol also for the application of the projection operator to vectors, such as gradients, meaning a component-wise application.

Using the above definitions, we define the discrete VEM solution as the function $u_\delta \in V_\delta$ satisfying

$$\mathscr{B}_\delta(u_\delta, v_\delta) = \left(f, \Pi^0_{k-1} v_\delta\right) \quad \forall v_\delta \in V_\delta .$$

This problem is well-posed and satisfies optimal a priori error estimates (see Chapters 2 and 3). In the following we focus on the construction of the local projection matrices and the local matrices and vectors required for the set up of the global discrete problem.

In the presentation given here we have considered the minimal requirement in the projections in order to preserve the expected polynomial rate of convergence (k in the energy norm) of the numerical solution. In the first VEM papers the projection used for the right-hand-side computation was ${}^{\mathbf{m}}\mathbf{\Pi}^0_k$.

7.3 Orthogonal polynomials on the generic element

All the computations performed in order to set up the VEM linear system providing the solution are based on operations between polynomial functions representing the projection of functions appearing in the consistent part of the operator and in the right-hand-side. A key issue in performing all the computations is a suitable basis for the polynomial spaces on general polygonal elements. Among the several possible options the classical and more simple choice is the scaled monomial basis [3, 5]. In the following we describe the construction of a suitable different almost orthogonal basis. A key issue to be considered in this construction process is that we need a basis for the space of polynomials of order $k-1$ for the construction of the Π^0_{k-1} projector, largely used in the consistent part of the discretization of the problem. This is the first step of our construction. Moreover, we also need a basis for the full space of polynomials of order k for the computations involved by the Π^∇_k construction required in the VEM stabilization considered in [3, 5]. For this reason we need a basis for the space $\mathbb{P}_k(E)$ obtained by the chosen basis functions for $\mathbb{P}_{k-1}(E)$ and by a set of additional linearly independent basis functions. We remark that the proposed construction of a polynomial basis aims at improving the reliability of the projector operator and is not dependent on the VEM stabilization chosen [6].

7.3.1 Basis construction. In the following we introduce a number of vector of basis functions, mass matrices and projectors; for all of them we adopt the following common notation: we use a right superscript to denote the polynomial order, and we indicate the polynomial basis used for the construction of the mass matrices and the projectors as the left superscript. For the mass matrices we also introduce the right superscript $k/k-1$ to indicate that monomilas of order exactly k are used in the construction.

Let $\mathbf{m}^k$ be the column vector of the n_k scaled monomial basis functions of the space of polynomials up to degree k usually used in the VEM definition, $\boldsymbol{p}^k$ and $\mathbf{p}^k$ are the column vectors of two suitable sets of linearly independent polynomials of degree k, whose construction will be discussed in the following. The construction of the target basis $\mathbf{p}^k$ of $\mathbb{P}_k(E)$ is split in two steps: first we construct the orthonormal basis $\mathbf{p}^{k-1}$ of $\mathbb{P}_{k-1}(E)$ used for the

construction of the projectors ${}^{\mathfrak{p}}\Pi^0_{k-1}$ and then we complete the basis for $\mathbb{P}_k(E)$ adding suitable basis functions, this basis is required for the construction of the projection Π^∇_k needed for the computation of the VEM stabilization. In each of the two steps intermediate bases $\boldsymbol{p}^{k-1}$ and $\boldsymbol{p}^k$ are introduced to explain the construction.

Let $\mathbf{R}^k$ be the matrix whose i-th row represent the coefficients of the i-th polynomial p_i^k of the orthogonal basis in terms of the monomial basis $\mathbf{m}^k$:

$$\mathfrak{p}_i^k = \sum_{j=1,..n_k} r_{i,j} m_j^k = \mathbf{R}^k_{i,:}\mathbf{m}^k\,.$$

In a compact form we can write

$$\mathfrak{p}^k = \mathbf{R}^k\mathbf{m}^k.$$

Let us introduce the mass matrix ${}^{\mathfrak{m}}\mathbf{H}^k \in \mathbb{R}^{n_k\times n_k}$ defined as

$${}^{\mathfrak{m}}\mathbf{H}^k = \int_E \mathbf{m}^k \left(\mathbf{m}^k\right)^\intercal \mathrm{d}\Omega\,,$$

and let us consider the principal sub-matrix of order n_{k-1}, that is the mass matrix of the monomials up to the order $k-1$:

$${}^{\mathfrak{m}}\mathbf{H}^{k-1} = \int_E \mathbf{m}^{k-1} \left(\mathbf{m}^{k-1}\right)^\intercal \mathrm{d}\Omega\,.$$

Moreover, let us denote by ${}^{\mathfrak{m}}\mathbf{H}^{k,k-1}$ the block of the mass matrix ${}^{\mathfrak{m}}\mathbf{H}^k$ with the last $n_k - n_{k-1}$ rows and the first n_{k-1} columns, and by ${}^{\mathfrak{m}}\mathbf{H}^{k/k-1}$ the block matrix with the last $n_k - n_{k-1}$ rows and columns.

Orthonormal basis for $\mathbb{P}_{k-1}(E)$. Let us define the matrix $\mathbf{R}^{k-1}$ such that the mass matrix ${}^{\mathfrak{p}}\mathbf{H}^{k-1}$ with respect to the basis $\mathfrak{p}^{k-1}$ is diagonal:

$$\begin{aligned}{}^{\mathfrak{p}}\mathbf{H}^{k-1} = \int_E \mathfrak{p}^{k-1} \left(\mathfrak{p}^{k-1}\right)^\intercal \mathrm{d}\Omega = \int_E \mathbf{R}^{k-1}\mathbf{m}^{k-1} \left(\mathbf{m}^{k-1}\right)^\intercal \left(\mathbf{R}^{k-1}\right)^\intercal \mathrm{d}\Omega = \\ = \mathbf{R}^{k-1}\,{}^{\mathfrak{m}}\mathbf{H}^{k-1} \left(\mathbf{R}^{k-1}\right)^\intercal = \mathbf{\Lambda}^{k-1}\,.\end{aligned}$$

Namely, the matrix $\left(\mathbf{R}^{k-1}\right)^\intercal$ is the matrix of the column-wise right-eigenvectors of ${}^{\mathfrak{m}}\mathbf{H}^{k-1}$, and the diagonal matrix $\mathbf{\Lambda}^{k-1}$ is the matrix of the eigenvalues of ${}^{\mathfrak{m}}\mathbf{H}^{k-1}$.

We finally introduce the orthogonal matrix

$$\mathbf{Q}^{k-1} = \sqrt{\left(\mathbf{\Lambda}^{k-1}\right)^{-1}}\,\mathbf{R}^{k-1}\,, \tag{7.1}$$

and then define the set of $\mathrm{L}^2(E)$-orthonormal polynomials that is a basis of the space $\mathbb{P}_{k-1}(E)$:

$$\mathbf{p}^{k-1} = \mathbf{Q}^{k-1}\mathbf{m}^{k-1}, \tag{7.2}$$

with an identity mass matrix:

$$\begin{aligned}{}^{\mathfrak{p}}\mathbf{H}^{k-1} = \int_E \mathbf{p}^{k-1} \left(\mathbf{p}^{k-1}\right)^\intercal \mathrm{d}\Omega = \int_E \mathbf{Q}^{k-1}\mathbf{m}^{k-1} \left(\mathbf{m}^{k-1}\right)^\intercal \left(\mathbf{Q}^{k-1}\right)^\intercal \mathrm{d}\Omega \\ = \mathbf{Q}^{k-1}\,{}^{\mathfrak{m}}\mathbf{H}^{k-1} \left(\mathbf{Q}^{k-1}\right)^\intercal = \sqrt{\left(\mathbf{\Lambda}^{k-1}\right)^{-1}}\mathbf{\Lambda}^{k-1} \left(\sqrt{\left(\mathbf{\Lambda}^{k-1}\right)^{-1}}\right)^\intercal = \mathbf{I}^{k-1}.\end{aligned}$$

Improved basis for $\mathbb{P}_k(E)$. In order to build a basis for the full space $\mathbb{P}_k(E)$ we add to the basis functions $\mathbf{p}^{k-1}$ a set of suitable linearly independent basis functions denoted by $\mathfrak{p}^{k/k-1}$, and obtained removing from the monomials $\mathbf{m}^{k/k-1}$ of order (exactly) k their components in the space of polynomials of order up $k-1$. Let us apply a Gram-Schmidt ortogonalization:

$$\begin{aligned}\mathfrak{p}^{k/k-1} &= \mathbf{m}^{k/k-1} - \left(\int_E \mathbf{m}^{k/k-1}\left(\mathbf{p}^{k-1}\right)^{\mathsf{T}} \mathrm{d}\Omega\right)\mathbf{p}^{k-1} = \\ &= \mathbf{m}^{k/k-1} - \left(\int_E \mathbf{m}^{k/k-1}\left(\mathbf{m}^{k-1}\right)^{\mathsf{T}} \mathrm{d}\Omega\right)\mathbf{m}^{k-1} = \\ &= \mathbf{m}^{k/k-1} - {}^{\mathtt{m}}\mathbf{H}^{k,k-1}\,\mathbf{m}^{k-1} = \left[-\left({}^{\mathtt{m}}\mathbf{H}^{k,k-1}\right)^{\mathsf{T}} \quad \mathbf{I}^{k/k-1}\right]\mathbf{m}^{k}.\end{aligned}$$

Let us define the matrix

$$\mathbf{R}_a^{k/k-1} = \left[-\left({}^{\mathtt{m}}\mathbf{H}^{k,k-1}\right)^{\mathsf{T}} \quad \mathbf{I}^{k/k-1}\right] \in \mathbb{R}^{(n_k - n_{k-1})\times n_{k-1}}. \tag{7.3}$$

Note that the set of functions $\mathfrak{p}^{k/k-1}$ is obtained starting from the set of monomials of order k, but they are general polynomials of order k orthogonal to the polynomial basis functions of order $k-1$.

Now, let us extract from these polynomials a set of linearly independent $\mathrm{L}^2(E)$ orthogonal functions $\mathbf{p}^{k/k-1}$. Let us consider the mass matrix relative to the polynomials $\mathfrak{p}^{k/k-1}$

$${}^{\mathfrak{p}^{k/k-1}}\mathbf{H}^{k/k-1} = \int_E \mathfrak{p}^{k/k-1}\left(\mathfrak{p}^{k/k-1}\right)^{\mathsf{T}} \mathrm{d}\Omega = \mathbf{R}_a^{k/k-1}\left(\int_E \mathbf{m}^{k}\left(\mathbf{m}^{k}\right)^{\mathsf{T}} \mathrm{d}\Omega\right)\left(\mathbf{R}_a^{k/k-1}\right)^{\mathsf{T}},$$

and let $\mathbf{R}_b^{k/k-1}$ the orthogonal matrix of change of basis that leads to a diagonal mass matrix starting from ${}^{\mathfrak{p}^{k/k-1}}\mathbf{H}^{k/k-1}$:

$$\begin{aligned}\mathbf{\Lambda}^{k/k-1} &= \left(\mathbf{R}_b^{k/k-1}\right)\left({}^{\mathfrak{p}^{k/k-1}}\mathbf{H}^{k/k-1}\right)\left(\mathbf{R}_b^{k/k-1}\right)^{\mathsf{T}} = \\ &= \left(\mathbf{R}_b^{k/k-1}\right)\left(\mathbf{R}_a^{k/k-1}\right){}^{\mathtt{m}}\mathbf{H}^{k}\left(\mathbf{R}_a^{k/k-1}\right)^{\mathsf{T}}\left(\mathbf{R}_b^{k/k-1}\right)^{\mathsf{T}}.\end{aligned}$$

We, finally, define the basis functions

$$\mathbf{p}^{k/k-1} = \sqrt{\left(\mathbf{\Lambda}^{k/k-1}\right)^{-1}}\mathbf{R}_b^{k/k-1}\mathbf{R}_a^{k/k-1}\mathbf{m}^{k} = \mathbf{Q}^{k/k-1}\mathbf{m}^{k}, \tag{7.4}$$

and the new full "almost $\mathrm{L}^2(E)$-othonormal" basis is

$$\mathbf{p}^{k} = \mathbf{Q}^{k}\mathbf{m}^{k}, \tag{7.5}$$

where, defined the zero-matrix $\mathbf{O}^{k-1,k} \in \mathbb{R}^{n_{k-1}\times n_k - n_{k-1}}$, the matrix $\mathbf{Q}^k$ has the following structure:

$$\mathbf{Q}^{k} = \begin{bmatrix}\mathbf{Q}^{k-1} & \mathbf{O}^{k-1,k} \\ & \mathbf{Q}^{k/k-1}\end{bmatrix}, \tag{7.6}$$

and, in exact arithmetic, the resulting mass matrix is

$${}^{\mathtt{p}}\mathbf{H}^{k} = \int_E \mathbf{p}^{k}\left(\mathbf{p}^{k}\right)^{\mathsf{T}} \mathrm{d}\Omega = \mathbf{Q}^{k}\,{}^{\mathtt{m}}\mathbf{H}^{k}\,\left(\mathbf{Q}^{k}\right)^{\mathsf{T}} = \begin{bmatrix}\mathbf{I}^{k-1} & {}^{\mathtt{p}}\mathbf{H}^{k-1,k} \\ {}^{\mathtt{p}}\mathbf{H}^{k,k-1} & \mathbf{I}^{k/k-1}\end{bmatrix}. \tag{7.7}$$

For badly shaped elements, the computation of the eigenvalues-eigenvectors can be affected by a non negligible numerical error. When this happens, the diagonal blocks of the matrix

${}^{\mathbf{p}}\mathbf{H}^k$ are no longer identity matrices, and, for this reason, in Section 7.6 we consider the following definitions:

$$ {}^{\mathbf{p}}\mathbf{H}^{k-1} = \mathbf{Q}^{k-1}\, {}^{\mathbf{m}}\mathbf{H}^{k-1} \left(\mathbf{Q}^{k-1}\right)^{\mathsf{T}}, \tag{7.8} $$

$$ {}^{\mathbf{p}}\mathbf{H}^{k} = \mathbf{Q}^{k}\, {}^{\mathbf{m}}\mathbf{H}^{k} \left(\mathbf{Q}^{k}\right)^{\mathsf{T}}, \tag{7.9} $$

with the matrices $\mathbf{Q}^{k-1}$ and $\mathbf{Q}^k$ given by (7.1) and (7.6), respectively.

7.3.2 Computation of the projector operator matrices for $\Pi^0_{k-1}\nabla$.

In this section we describe how to obtain the $\mathrm{L}^2(E)$ projection of the gradient components of a VEM basis function following the description provided in [4, 5].

Let $\Pi^0_{k-1}\phi_{i,x}$ be the projection of the derivative with respect to the variable x of the VEM basis function ϕ_i. Let us write this projection with respect to the scaled monomial basis $\mathbf{m}$ and the basis $\mathbf{p}$ built in the previous section, respectively:

$$ {}^{\mathbf{m}}\Pi^0_{k-1}\,\phi_{i,x} = \left(\mathbf{m}^{k-1}\right)^T\, {}^{\mathbf{m}}\mathbf{\Pi}^0_x\,(:,i)\,, \qquad {}^{\mathbf{p}}\Pi^0_{k-1}\,\phi_{i,x} = \left(\mathbf{p}^{k-1}\right)^T\, {}^{\mathbf{p}}\mathbf{\Pi}^0_x\,(:,i)\,, \tag{7.10} $$

and similarly for the derivatives with respect to the variable y.

Let us define the matrix ${}^{\mathbf{m}}\mathbf{E}_x$ of the $\mathrm{L}^2\,(E)$ scalar product of the x derivative of the VEM basis function ϕ_i with respect to the monomial basis $\mathbf{m}^{k-1}$ and the matrix ${}^{\mathbf{p}}\mathbf{E}_x$ with respect to the orthonormal basis $\mathbf{p}^{k-1}$, respectively:

$$ {}^{\mathbf{m}}\mathbf{E}_x\,(l,i) = \int_E m_l \phi_{i,x}, \qquad\qquad {}^{\mathbf{p}}\mathbf{E}_x\,(l,i) = \int_E p_l \phi_{i,x}, $$

the relation between the two matrices is ${}^{\mathbf{p}}\mathbf{E}_x = \mathbf{Q}^{k-1}\, {}^{\mathbf{m}}\mathbf{E}_x$. Moreover, the $\mathrm{L}^2\,(E)$ projections ${}^{\mathbf{m}}\Pi^0_{k-1}\,\phi_{i,x}$ and ${}^{\mathbf{p}}\Pi^0_{k-1}\,\phi_{i,x}$ are defined by the systems of equations

$$ \int_E \mathbf{m}^{k-1} \left({}^{\mathbf{m}}\Pi^0_{k-1}\,\phi_{i,x}\right) \mathrm{d}\Omega = \int_E \mathbf{m}^{k-1}\phi_{i,x}\mathrm{d}\Omega, \tag{7.11} $$

$$ \int_E \mathbf{p}^{k-1} \left({}^{\mathbf{p}}\Pi^0_{k-1}\,\phi_{i,x}\right) \mathrm{d}\Omega = \int_E \mathbf{p}^{k-1}\phi_{i,x}\mathrm{d}\Omega, \tag{7.12} $$

respectively. Let us write the projections in (7.11), (7.12) by (7.10), we have

$$ \int_E \mathbf{m}^{k-1}\phi_{i,x}\mathrm{d}\Omega = \left(\int_E \mathbf{m}^{k-1}\left(\mathbf{m}^{k-1}\right)^T \mathrm{d}\Omega\right)\, {}^{\mathbf{m}}\mathbf{\Pi}^0_x\,(:,i), $$

$$ \int_E \mathbf{p}^{k-1}\phi_{i,x}\mathrm{d}\Omega = \left(\int_E \mathbf{p}^{k-1}\left(\mathbf{p}^{k-1}\right)^T \mathrm{d}\Omega\right)\, {}^{\mathbf{p}}\mathbf{\Pi}^0_x\,(:,i), $$

that is

$$ {}^{\mathbf{m}}\mathbf{E}_x\,(:,i) = {}^{\mathbf{m}}\mathbf{H}^{k-1}\, {}^{\mathbf{m}}\mathbf{\Pi}^0_x\,(:,i), \qquad {}^{\mathbf{p}}\mathbf{E}_x\,(:,i) = {}^{\mathbf{p}}\mathbf{H}^{k-1}\, {}^{\mathbf{p}}\mathbf{\Pi}^0_x\,(:,i), $$

$$ {}^{\mathbf{m}}\mathbf{E}_x = {}^{\mathbf{m}}\mathbf{H}^{k-1}\, {}^{\mathbf{m}}\mathbf{\Pi}^0_x\,, \qquad {}^{\mathbf{p}}\mathbf{E}_x = {}^{\mathbf{p}}\mathbf{H}^{k-1}\, {}^{\mathbf{p}}\mathbf{\Pi}^0_x\,, $$

and

$$ {}^{\mathbf{m}}\mathbf{\Pi}^0_x = \left({}^{\mathbf{m}}\mathbf{H}^{k-1}\right)^{-1}\, {}^{\mathbf{m}}\mathbf{E}_x\,, \qquad {}^{\mathbf{p}}\mathbf{\Pi}^0_x = \left({}^{\mathbf{p}}\mathbf{H}^{k-1}\right)^{-1}\, {}^{\mathbf{p}}\mathbf{E}_x\,. \tag{7.13} $$

In exact arithmetic we have

$$ {}^{\mathbf{p}}\mathbf{\Pi}^0_x = {}^{\mathbf{p}}\mathbf{E}_x = \mathbf{Q}^{k-1}\, {}^{\mathbf{m}}\mathbf{E}_x\,, \tag{7.14} $$

and proceeding in a similar way we get ${}^{\mathtt{p}}\mathbf{\Pi}^0_y = {}^{\mathtt{p}}\mathbf{E}_y$. For the computation of the matrices ${}^{\mathtt{m}}\mathbf{E}_x$ and ${}^{\mathtt{m}}\mathbf{E}_y$ resorting to the VEM-dofs we refer to [3, 5] and remark that, by the Green formula, all these computations can be written in term of integrals on the elements of polynomials of order $k-2$ that are VEM dofs and integrals on the boundary of VEM basis functions and polynomials of order $k-1$. In the computations performed in the following we use the expressions

$$
\begin{aligned}
{}^{\mathtt{p}}\mathbf{\Pi}^0_x &= \left({}^{\mathtt{p}}\mathbf{H}^{k-1}\right)^{-1}\mathbf{Q}^{k-1}\,{}^{\mathtt{m}}\mathbf{E}_x = \mathbf{Q}^{k-1}\,{}^{\mathtt{m}}\mathbf{\Pi}^0_x\,, && (7.15)\\
{}^{\mathtt{p}}\mathbf{\Pi}^0_y &= \left({}^{\mathtt{p}}\mathbf{H}^{k-1}\right)^{-1}\mathbf{Q}^{k-1}\,{}^{\mathtt{m}}\mathbf{E}_y = \mathbf{Q}^{k-1}\,{}^{\mathtt{m}}\mathbf{\Pi}^0_y\,. && (7.16)
\end{aligned}
$$

The matrix $\mathbf{Q}^{k-1}$ works as a preconditioner for the projection matrices ${}^{\mathtt{p}}\mathbf{\Pi}^0_x$ and ${}^{\mathtt{p}}\mathbf{\Pi}^0_y$.

7.3.3 Stiffness matrix computation.

Denoting by Φ the column vector of the VEM basis functions ϕ_i, $i = 1, .., n_k$, and by $\nabla\Phi^T$ the matrix with two rows and n_k columns with the gradient $\nabla\phi_i$ in the column i. Let us assume that K is a positive scalar function. The element stiffness matrix is given by

$$
\begin{aligned}
{}^{\mathtt{p}}\mathbf{K}_{\mathsf{K}} &= \int_E \mathsf{K}\left({}^{\mathtt{p}}\Pi^0_{k-1}\frac{\partial\Phi}{\partial x}\right)\left({}^{\mathtt{p}}\Pi^0_{k-1}\frac{\partial\Phi}{\partial x}\right)^{\mathsf{T}}\mathrm{d}\Omega + \int_E \mathsf{K}\left({}^{\mathtt{p}}\Pi^0_{k-1}\frac{\partial\Phi}{\partial y}\right)\left({}^{\mathtt{p}}\Pi^0_{k-1}\frac{\partial\Phi}{\partial y}\right)^{\mathsf{T}}\mathrm{d}\Omega =\\
&= \int_E \mathsf{K}\left({}^{\mathtt{p}}\mathbf{\Pi}^0_x\right)^{\mathsf{T}}\mathbf{p}^{k-1}\left(\mathbf{p}^{k-1}\right)^{\mathsf{T}}{}^{\mathtt{p}}\mathbf{\Pi}^0_x\,\mathrm{d}\Omega + \int_E \mathsf{K}\left({}^{\mathtt{p}}\mathbf{\Pi}^0_y\right)^{\mathsf{T}}\mathbf{p}^{k-1}\left(\mathbf{p}^{k-1}\right)^{\mathsf{T}}{}^{\mathtt{p}}\mathbf{\Pi}^0_y\,\mathrm{d}\Omega =\\
&= \left({}^{\mathtt{p}}\mathbf{\Pi}^0_x\right)^{\mathsf{T}}{}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}}\,{}^{\mathtt{p}}\mathbf{\Pi}^0_x + \left({}^{\mathtt{p}}\mathbf{\Pi}^0_y\right)^{\mathsf{T}}{}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}}\,{}^{\mathtt{p}}\mathbf{\Pi}^0_y\,,
\end{aligned}
$$

where we have defined

$$
{}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}} = \int_E \mathbf{p}^{k-1}\mathsf{K}\left(\mathbf{p}^{k-1}\right)^{\mathsf{T}}\mathrm{d}\Omega = \mathsf{K}\,\mathbf{I}^{k-1},
$$

and we can write

$$
{}^{\mathtt{p}}\mathbf{K}_{\mathsf{K}} = \begin{bmatrix}\left({}^{\mathtt{p}}\mathbf{\Pi}^0_x\right)^{\mathsf{T}} & \left({}^{\mathtt{p}}\mathbf{\Pi}^0_y\right)^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}{}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}} & 0\\ 0 & {}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}}\end{bmatrix}\begin{bmatrix}{}^{\mathtt{p}}\mathbf{\Pi}^0_x\\ {}^{\mathtt{p}}\mathbf{\Pi}^0_y\end{bmatrix}. \qquad (7.17)
$$

If K is constant in the element E, in exact arithmetic, we have ${}^{\mathtt{p}}\mathbf{K}_{\mathsf{K}} = \mathsf{K}\,\mathbf{I}^{k-1}$.

In case K is a simmetric positive definite tensor whose components are denoted by $\mathsf{K}_{x_ix_j}$ with $i, j = 1, 2$ and the usual convention $x_1 = x$ and $x_2 = y$, we define

$$
\begin{aligned}
{}^{\mathtt{m}}\mathbf{H}^{k-1}_{\mathsf{K}_{x_ix_j}} &= \int_E \mathsf{K}_{x_ix_j}\mathbf{m}^{k-1}\left(\mathbf{m}^{k-1}\right)^{\mathsf{T}}\mathrm{d}\Omega,\\
{}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}_{x_ix_j}} &= \int_E \mathsf{K}_{x_ix_j}\mathbf{p}^{k-1}\left(\mathbf{p}^{k-1}\right)^{\mathsf{T}}\mathrm{d}\Omega = \mathbf{Q}^{k-1}\,{}^{\mathtt{m}}\mathbf{H}^{k-1}_{\mathsf{K}_{x_ix_j}}\left(\mathbf{Q}^{k-1}\right)^{\mathsf{T}},
\end{aligned}
$$

and proceeding in a similar way we finally get

$$
{}^{\mathtt{p}}\mathbf{K}_{\mathsf{K}} = \begin{bmatrix}\left({}^{\mathtt{p}}\mathbf{\Pi}^0_x\right)^{\mathsf{T}} & \left({}^{\mathtt{p}}\mathbf{\Pi}^0_y\right)^{\mathsf{T}}\end{bmatrix}\begin{bmatrix}{}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}_{x_1x_1}} & {}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}_{x_1x_2}}\\ {}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}_{x_2x_1}} & {}^{\mathtt{p}}\mathbf{H}^{k-1}_{\mathsf{K}_{x_2x_2}}\end{bmatrix}\begin{bmatrix}{}^{\mathtt{p}}\mathbf{\Pi}^0_x\\ {}^{\mathtt{p}}\mathbf{\Pi}^0_y\end{bmatrix} \qquad (7.18)
$$

7.3.4 Computation of the projector operator Π^∇_k.

First let us recall the definition of the Π^∇_k operator [2–4]:

$$
\left(\nabla\Pi^\nabla_k v_\delta, \nabla q_k\right) = \left(\nabla v_\delta, \nabla q_k\right), \quad \forall q_k \in \mathbb{P}_E(k)\,. \qquad (7.19)
$$

Equation (7.19) defines the projection $\Pi_k^\nabla v_\delta$ of the VEM function v_δ up to a constant that can be fixed prescribing a projector operator $P_0 : V_k(E) \to \mathbb{P}_E(0)$ such that

$$P_0 \Pi_k^\nabla v_\delta = P_0 v_\delta \,.$$

Several options for this operator are possible. As in [3, 4] we choose

$$\begin{cases} (P_0 v_\delta, 1)_{\partial E} = (v_\delta, 1)_{\partial E}, & \text{for } k = 1, \\ (P_0 v_\delta, 1)_E = (v_\delta, 1)_E, & \text{for } k \geq 2. \end{cases} \tag{7.20}$$

Since $\Pi_k^\nabla \phi_i \in \mathbb{P}_k(E)$ we can represent it with respect to the bases $\mathbf{m}$ and $\mathbf{p}$, with coefficients ${}^{\mathbf{m}}\mathbf{\Pi}_k^\nabla(:,i)$ and ${}^{\mathbf{p}}\mathbf{\Pi}_k^\nabla(:,i)$, respectively

$$\Pi_k^\nabla \phi_i = \left(\mathbf{m}^k\right)^\intercal \, {}^{\mathbf{m}}\mathbf{\Pi}_k^\nabla(:,i) = \left(\mathbf{m}^k\right)^\intercal \left(\mathbf{Q}^k\right)^\intercal \, {}^{\mathbf{p}}\mathbf{\Pi}_k^\nabla(:,i).$$

Let us define the $\mathbb{R}^{n_{k-1}, n_{k-1}}$ matrix

$${}^{\mathbf{m}}\tilde{\mathbf{G}} = \int_E \nabla^T \mathbf{m}^k \nabla \mathbf{m}^k \mathrm{d}\Omega,$$

and

$${}^{\mathbf{m}}\tilde{\mathbf{B}}(:,i) = \int_E \nabla^T \mathbf{m}^k \nabla \Phi \mathrm{d}\Omega.$$

Using the monomial basis we get

$$\int_E \nabla^\intercal \mathbf{m}^k \nabla \Pi_k^\nabla \phi_i \mathrm{d}\Omega = \int_E \nabla^\intercal \mathbf{m}^k \nabla \mathbf{m}^{k\intercal} \mathrm{d}\Omega \, {}^{\mathbf{m}}\mathbf{\Pi}_k^\nabla(:,i) = {}^{\mathbf{m}}\tilde{\mathbf{G}} \, {}^{\mathbf{m}}\mathbf{\Pi}_k^\nabla(:,i) = \\ = \int_E \nabla^\intercal \mathbf{m}^k \nabla \phi_i \mathrm{d}\Omega = {}^{\mathbf{m}}\tilde{\mathbf{B}}(:,i).$$

Whereas, using the basis of polynomials $\mathbf{p}^k$

$${}^{\mathbf{p}}\tilde{\mathbf{G}} \, {}^{\mathbf{p}}\mathbf{\Pi}_k^\nabla(:,i) = \mathbf{Q}^k \, {}^{\mathbf{m}}\tilde{\mathbf{G}} \left(\mathbf{Q}^k\right)^\intercal \, {}^{\mathbf{p}}\mathbf{\Pi}_k^\nabla(:,i) = {}^{\mathbf{p}}\tilde{\mathbf{B}}(:,i) = \mathbf{Q}^k \, {}^{\mathbf{m}}\tilde{\mathbf{B}}(:,i). \tag{7.21}$$

The first row and first column of the matrix ${}^{\mathbf{m}}\tilde{\mathbf{G}}$ is trivially vanishing appearing in the integrals the gradient of constants. We can say that the matrix ${}^{\mathbf{p}}\tilde{\mathbf{G}}$ is singular as well. For this reason we define the matrices ${}^{\mathbf{m}}\mathbf{G}$ and ${}^{\mathbf{p}}\mathbf{G}$ in the following way. As in [4], let us consider the matrix ${}^{\mathbf{m}}\tilde{\mathbf{G}}$ and replace its first row with the vector $P_0\left(\mathbf{m}^k\right)^\intercal$ obtaining the matrix ${}^{\mathbf{m}}\mathbf{G}$, and replace the first row of ${}^{\mathbf{m}}\tilde{\mathbf{B}}$ with $P_0\left(\Phi\right)^\intercal$, obtaining ${}^{\mathbf{m}}\mathbf{B}$. The undetermined linear system ${}^{\mathbf{m}}\tilde{\mathbf{G}} \, {}^{\mathbf{m}}\mathbf{\Pi}_k^\nabla = {}^{\mathbf{m}}\tilde{\mathbf{B}}$ is replaced by

$${}^{\mathbf{m}}\mathbf{G} \, {}^{\mathbf{m}}\mathbf{\Pi}_k^\nabla = {}^{\mathbf{m}}\mathbf{B}. \tag{7.22}$$

Instead of computing ${}^{\mathbf{p}}\mathbf{G}$ by the transformation ${}^{\mathbf{p}}\mathbf{G} = \mathbf{Q}^k \, {}^{\mathbf{m}}\mathbf{G} \left(\mathbf{Q}^k\right)^\intercal$ we could directly compute the matrix ${}^{\mathbf{p}}\mathbf{G}$ by performing a QR-rank-revealing factorization of the matrix ${}^{\mathbf{p}}\tilde{\mathbf{G}} = \mathbf{Q}^k \, {}^{\mathbf{m}}\tilde{\mathbf{G}} \left(\mathbf{Q}^k\right)^\intercal$, and then by replacing the row of the matrix corresponding to the lowest singular value with the the vector $P_0\left(\mathbf{p}^k\right)^\intercal = P_0\left(\mathbf{m}^k\right)^\intercal \left(\mathbf{Q}^k\right)^\intercal$ and the corresponding element of the right hand side ${}^{\mathbf{p}}\tilde{\mathbf{B}} = \mathbf{Q}^k \, {}^{\mathbf{m}}\tilde{\mathbf{B}}$ with $P_0 \Phi^\intercal$, we get

$${}^{\mathbf{p}}\mathbf{G} \, {}^{\mathbf{p}}\mathbf{\Pi}_k^\nabla = {}^{\mathbf{p}}\mathbf{B}. \tag{7.23}$$

7.3.5 Computation of the projector operator matrices Π^0_{k-1}.

In this section we describe how to obtain the $\mathrm{L}^2(E)$ projection of a VEM basis function following the description provided in [2, 5].

Let $\Pi^0_{k-1}\phi_i$ be the projection of the VEM basis function ϕ_i. Let us write this projection with respect to the scaled monomial basis $\mathbf{m}$ and the basis $\mathbf{p}$ built in the previous section, respectively:

$$ {}^{\mathbf{m}}\Pi^0_{k-1}\,\phi_i = \left(\mathbf{m}^{k-1}\right)^{\intercal}\,{}^{\mathbf{m}}\mathbf{\Pi}^0_{k-1}\,(:,i), \qquad {}^{\mathbf{p}}\Pi^0_{k-1}\,\phi_i = \left(\mathbf{p}^{k-1}\right)^{\intercal}\,{}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}\,(:,i). \tag{7.24} $$

Let us define the matrix ${}^{\mathbf{m}}\mathbf{C}$ of the $\mathrm{L}^2(E)$ scalar product of the VEM basis function ϕ_i with respect to the monomial basis $\mathbf{m}^{k-1}$ and the matrix ${}^{\mathbf{p}}\mathbf{C}$ with respect to the basis $\mathbf{p}^{k-1}$, respectively:

$$ {}^{\mathbf{m}}\mathbf{C}\,(l,i) = \int_E m_l\phi_i, \qquad {}^{\mathbf{p}}\mathbf{C}\,(l,i) = \int_E p_l\phi_i, \quad l = 1,\dots n_{k-1} $$

the relation between the two matrices is ${}^{\mathbf{p}}\mathbf{C} = \mathbf{Q}^{k-1}\,{}^{\mathbf{m}}\mathbf{C}$. In the definition of the VEM space we ask that $(q,\phi_i)_E = (q,\Pi^{\nabla}_k\phi_i)_E$, $\forall q \in \mathbb{P}_k(E)/\mathbb{P}_{k-2}(E)$ and this is the way we can compute the last row of the matrix ${}^{\mathbf{m}}\mathbf{C}$ and consequantly the matrix ${}^{\mathbf{p}}\mathbf{C}$ [2, 4]. Moreover, the $\mathrm{L}^2(E)$ projections ${}^{\mathbf{m}}\Pi^0_{k-1}\,\phi_i$ and ${}^{\mathbf{p}}\Pi^0_{k-1}\,\phi_i$ are defined by the systems of equations

$$ \int_E \mathbf{m}^{k-1}\,{}^{\mathbf{m}}\Pi^0_{k-1}\,\phi_i \mathrm{d}\Omega = \int_E \mathbf{m}^{k-1}\phi_i \mathrm{d}\Omega, \tag{7.25} $$

$$ \int_E \mathbf{p}^{k-1}\,{}^{\mathbf{p}}\Pi^0_{k-1}\,\phi_i \mathrm{d}\Omega = \int_E \mathbf{p}^{k-1}\phi_i \mathrm{d}\Omega, \tag{7.26} $$

respectively. Let us write the projections in (7.25), (7.26) by (7.24), we have

$$ {}^{\mathbf{m}}\mathbf{C}\,(:,i) = {}^{\mathbf{m}}\mathbf{H}^{k-1}\,{}^{\mathbf{m}}\mathbf{\Pi}^0_{k-1}\,(:,i), \qquad {}^{\mathbf{p}}\mathbf{C}\,(:,i) = {}^{\mathbf{p}}\mathbf{H}^{k-1}\,{}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}\,(:,i), $$
$$ {}^{\mathbf{m}}\mathbf{C} = {}^{\mathbf{m}}\mathbf{H}^{k-1}\,{}^{\mathbf{m}}\mathbf{\Pi}^0_{k-1}, \qquad {}^{\mathbf{p}}\mathbf{C} = {}^{\mathbf{p}}\mathbf{H}^{k-1}\,{}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}, $$

and

$$ {}^{\mathbf{m}}\mathbf{\Pi}^0_{k-1} = \left({}^{\mathbf{m}}\mathbf{H}^{k-1}\right)^{-1}\,{}^{\mathbf{m}}\mathbf{C}, $$
$$ {}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1} = \left({}^{\mathbf{p}}\mathbf{H}^{k-1}\right)^{-1}\,{}^{\mathbf{p}}\mathbf{C} = {}^{\mathbf{p}}\mathbf{C} = \mathbf{Q}^{k-1}\,{}^{\mathbf{m}}\mathbf{C}. $$

From a numerical point of view, in the following, we prefer to use

$$ {}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1} = \left({}^{\mathbf{p}}\mathbf{H}^{k-1}\right)^{-1}\,{}^{\mathbf{p}}\mathbf{C} = \left({}^{\mathbf{p}}\mathbf{H}^{k-1}\right)^{-1}\,\mathbf{Q}^{k-1}\,{}^{\mathbf{m}}\mathbf{C}. \tag{7.27} $$

7.3.6 Advection matrix computation.

Let us consider the elemental matrix of the advection term

$$ {}^{\mathbf{p}}\mathbf{K}_\beta = \int_E \beta_x \left({}^{\mathbf{p}}\Pi^0_{k-1}\,\Phi\right)\left({}^{\mathbf{p}}\Pi^0_{k-1}\,\frac{\partial\Phi}{\partial x}\right)^{\intercal}\mathrm{d}\Omega + \int_E \beta_y \left({}^{\mathbf{p}}\Pi^0_{k-1}\,\Phi\right)\left({}^{\mathbf{p}}\Pi^0_{k-1}\,\frac{\partial\Phi}{\partial y}\right)^{\intercal}\mathrm{d}\Omega = $$
$$ = \int_E \beta_x \left({}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}\right)^{\intercal}\mathbf{p}^{k-1}\left(\mathbf{p}^{k-1}\right)^{\intercal}\,{}^{\mathbf{p}}\mathbf{\Pi}^0_x\,\mathrm{d}\Omega + \int_E \beta_y \left({}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}\right)^{\intercal}\mathbf{p}^{k-1}\left(\mathbf{p}^{k-1}\right)^{\intercal}\,{}^{\mathbf{p}}\mathbf{\Pi}^0_y\,\mathrm{d}\Omega = $$
$$ = \left({}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}\right)^{\intercal}\,{}^{\mathbf{p}}\mathbf{H}^{k-1}_{\beta_x}\,{}^{\mathbf{p}}\mathbf{\Pi}^0_x + \left({}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}\right)^{\intercal}\,{}^{\mathbf{p}}\mathbf{H}^{k-1}_{\beta_y}\,{}^{\mathbf{p}}\mathbf{\Pi}^0_y $$

where, with $i = 1, 2$, we have defined

$$ {}^{\mathbf{m}}\mathbf{H}^{k-1}_{\beta_{x_i}} = \int_E \beta_{x_i}\mathbf{m}^{k-1}\left(\mathbf{m}^{k-1}\right)^{\intercal}\mathrm{d}\Omega, $$
$$ {}^{\mathbf{p}}\mathbf{H}^{k-1}_{\beta_{x_i}} = \int_E \beta_{x_i}\mathbf{p}^{k-1}\left(\mathbf{p}^{k-1}\right)^{\intercal}\mathrm{d}\Omega = \mathbf{Q}^{k-1}\,{}^{\mathbf{m}}\mathbf{H}^{k-1}_{\beta_{x_i}}\,\left(\mathbf{Q}^{k-1}\right)^{\intercal}. $$

7.3.7 Reaction matrix computation. Let us consider the elemental matrix of the reaction term

$$
\begin{aligned}
{}^{\mathbf{p}}\mathbf{K}_\gamma &= \int_E \gamma \left({}^{\mathbf{p}}\Pi^0_{k-1}\, \Phi \right) \left({}^{\mathbf{p}}\Pi^0_{k-1}\, \Phi \right)^{\mathsf{T}} \mathrm{d}\Omega = \\
&= \int_E \gamma \left({}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1} \right)^{\mathsf{T}} \mathbf{p}^{k-1} \left(\mathbf{p}^{k-1} \right)^{\mathsf{T}} {}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}\, \mathrm{d}\Omega = \left({}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1} \right)^{\mathsf{T}} {}^{\mathbf{p}}\mathbf{H}^{k-1}_\gamma\, {}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}\,,
\end{aligned}
$$

where we have defined

$$
\begin{aligned}
{}^{\mathbf{m}}\mathbf{H}^{k-1}_\gamma &= \int_E \gamma \mathbf{m}^{k-1} \left(\mathbf{m}^{k-1} \right)^{\mathsf{T}} \mathrm{d}\Omega, \\
{}^{\mathbf{p}}\mathbf{H}^{k-1}_\gamma &= \int_E \gamma \mathbf{p}^{k-1} \left(\mathbf{p}^{k-1} \right)^{\mathsf{T}} \mathrm{d}\Omega = \mathbf{Q}^{k-1}\, {}^{\mathbf{m}}\mathbf{H}^{k-1}_\gamma \left(\mathbf{Q}^{k-1} \right)^{\mathsf{T}}.
\end{aligned}
$$

7.4 Validation on a reaction-advection-diffusion problem

Before proceding to a detailed analysis of the effects of the basis $\mathbf{p}$ in preventing instabilities on badly shaped elements, we report some numerical results for a validation of the method. In particular, we aim at showing that the use of the new basis yields a discretization displaying rates of convergence for the error which correspond to the theoretical ones. Let $\Omega = (0,1) \times (0,1)$ and consider the reaction-convection-diffusion problem:

$$
\begin{cases}
-\nabla \cdot (\mathsf{K} \nabla u) + \beta \cdot \nabla u + \gamma u = f & \text{in } \Omega, \\
u = 0 & \text{su } \partial\Omega,
\end{cases}
$$

where $\mathsf{K}(x,y) = \begin{pmatrix} 1+y^2 & 0 \\ 0 & 1+x^2 \end{pmatrix}$ is a non-constant tensor diffusivity parameter, $\beta(x,y) = (x,-y)$ is the convection velocity, $\gamma(x,y) = xy$ is the reaction parameter and f is the right-hand-side chosen such that the solution is

$$
u(x,y) = -200\sqrt{\sin(1-x/\pi)}\cos(\pi x)(1-x)(1-y)xy^2\,.
$$

	k = 1	k = 2	k = 3	k = 4	k=5	k=6
$\mathrm{L}^2(\Omega)$	2.08	3.14	4.29	5.25	6.60	7.53
$\mathrm{H}^1_0(\Omega)$	1.03	2.12	3.20	4.25	5.55	6.40

Table 7.1: Validation test: rates of convergence on triangular mesh

	k = 1	k = 2	k = 3	k = 4	k=5	k=6
$\mathrm{L}^2(\Omega)$	1.98	3.01	3.97	4.95	6.05	6.98
$\mathrm{H}^1_0(\Omega)$	1.00	1.97	2.98	3.96	5.06	6.00

Table 7.2: Validation test: rates of convergence on hexagonal mesh

The computed rates of convergence for the norms $\mathrm{L}^2(\Omega)$ and $\mathrm{H}^1_0(\Omega)$ are reported in Tables 7.1 and 7.2 and are very close to the expected ones. Being the mesh a good quality mesh we have that the errors display the same values both with the basis $\mathbf{m}$ and $\mathbf{p}$. The rates of convergence in Table 7.1 are obtained on a triangular mesh with elements of area

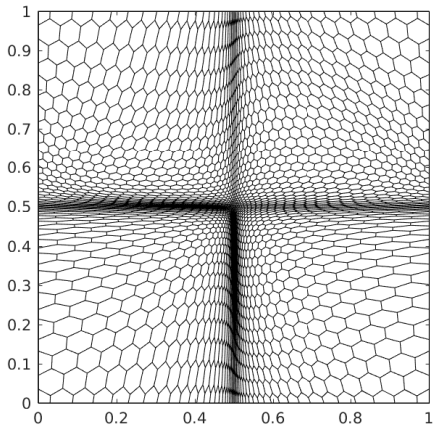


Figure 7.1: **Validation** Highly distorted Voronoi mesh

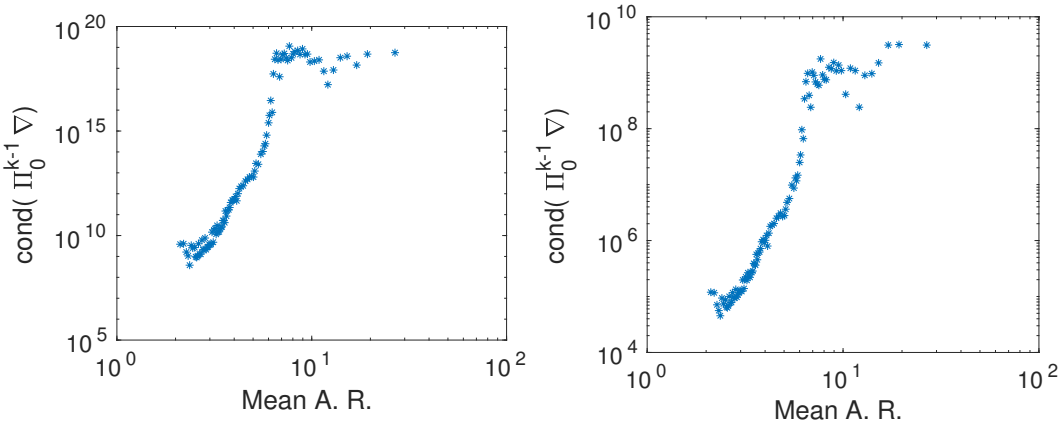
equal to 0.1, 0.01, 0.001 and 0.0001 for $k = 1, \dots, 4$, and with area equal to 0.1, 0.05, 0.01, 0.005 for $k = 5, 6$, while the results in Table 7.2 are obtained on progressively refined meshes of mildly distorted hexagons, with diameters spanning from 0.219 to 0.0266 for orders 1 up to 5, and from 0.219 to 0.071 for order 6.

In order to describe the effect of the use of the basis $\mathbf{p}$ we compare the condition numbers of the projection matrices computed solving the previous problem on an highly distorted Voronoi mesh displayed in Figure 7.1. Figures 7.2 to 7.4 display the condition numbers of the projection matrices ${}^*\mathbf{\Pi}^0_{k-1,x}$ and ${}^*\mathbf{\Pi}^0_{k-1,y}$ (mixed in Figure 7.2), ${}^*\mathbf{\Pi}^\nabla_k$ and ${}^*\mathbf{\Pi}^0_{k-1}$, for the basis $\mathbf{m}$ and the basis $\mathbf{p}$ with respect to the aspect ratio. We can observe a very strong reduction of their condition numbers when $\mathbf{p}$ is used. In order to draw these plots we define the aspect ratio, as the ratio between the largest distance and the smallest distance between couples of vertices of the polygon. For each element in the mesh we compute the aspect ratio and we partition the full range of aspect ratios in 100 uniform intervals. In the plots we report the mean condition numbers computed on all the elements with an aspect ratio in each of these intervals. We remark that the effect of the use of the basis $\mathbf{p}$ is local, and that the global condition number of the final matrix is not significantly reduced by the process.

7.5 Interface problem with highly anisotropic mesh

To show the capability of the described change of basis in providing a more accurate solution also with very bad shaped meshes, we consider here a case where oscillations are observed due to very badly shaped elements. Let

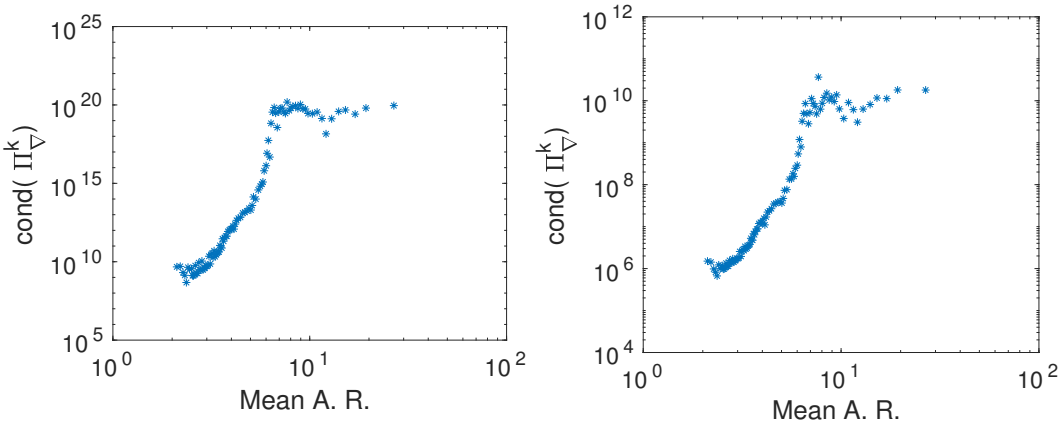
$$\mathsf{K}(x, y) = \begin{cases} 10 & \text{if } (x, y) \in (0, 0.5)^2 \cup (0.5, 1)^2 \,, \\ 1 & \text{otherwise} \,, \end{cases}$$



(a) Standard basis

(b) Orthogonal basis

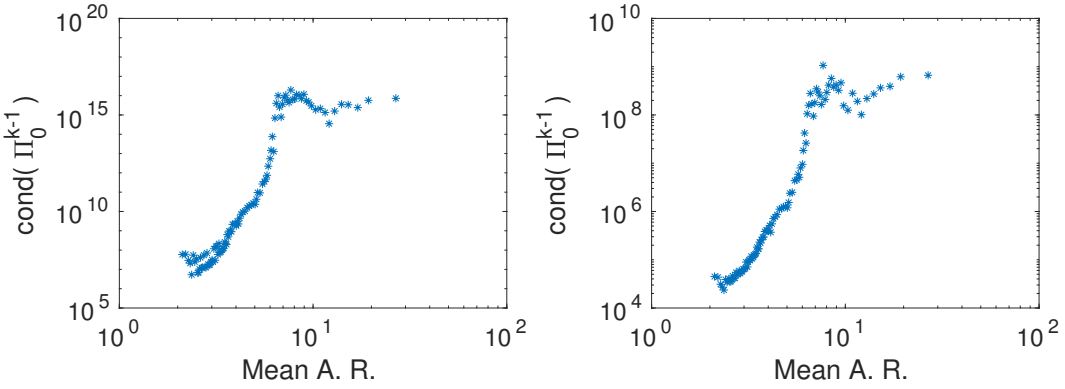
Figure 7.2: **Validation**, order 6 Mean condition number of the matrix representation of $\Pi^0_{k-1}\nabla$.



(a) Standard basis

(b) Orthogonal basis

Figure 7.3: **Validation**, order 6 Mean condition number of the matrix representation of Π^∇_k.



(a) Standard basis

(b) Orthogonal basis

Figure 7.4: **Validation**, order 6 Mean conditioning number of the matrix representation of Π^0_{k-1}.

and let

$$\psi(x) = -\frac{1}{\mu}\begin{cases} \frac{x^2}{2} + cx & \text{if } (x,y) \in (0,0.5)^2, \\ \frac{x^2}{2} + cx - c - \frac{1}{2} & \text{if } (x,y) \in (0.5,1) \times (0,0.5), \\ \frac{(1-x)^2}{2} + c(1-x) & \text{if } (x,y) \in (0.5,1)^2, \\ \frac{(1-x)^2}{2} + c(1-x) - c - \frac{1}{2} & \text{if } (x,y) \in (0,0.5) \times (0.5,1), \end{cases}$$

where $c = -31/44$ is chosen in such a way that the co-normal derivative of ψ is continuous. Furthermore, let $Y(y) = y\,(1-y)\left(y - \frac{1}{2}\right)^2$. We consider the problem

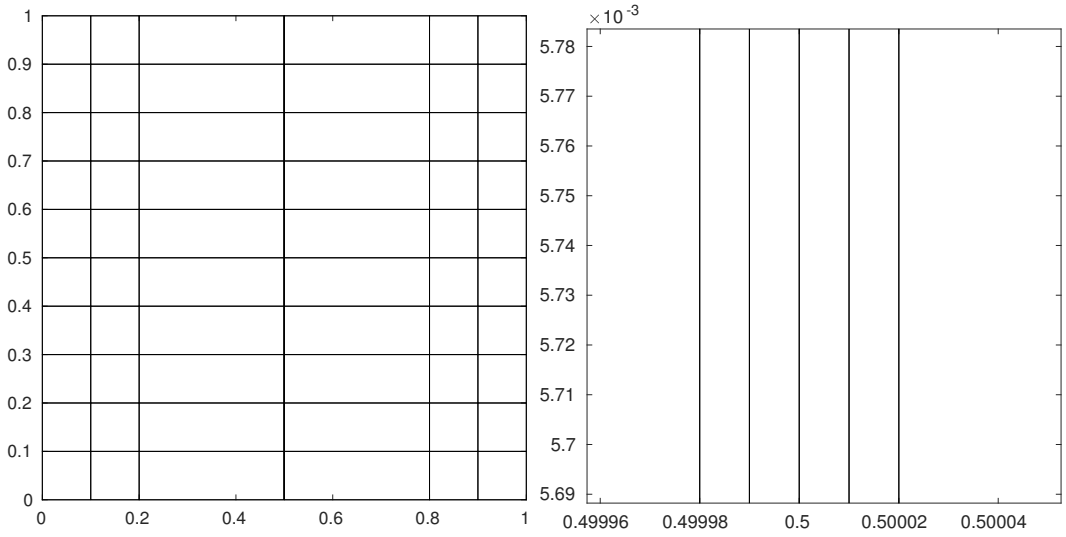
$$\begin{cases} -\nabla \cdot (\mathsf{K}\nabla u) = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases}$$

with f chosen in such a way that the solution is $u(x,y) = \psi(x)Y(y)$. First, we solve the problem using standard Virtual Elements on the mesh in Figure 7.5a, that is obtained from a regular 10×10 square mesh by moving the edges in the region $(0.25, 0.75) \times (0, 1)$ towards the axis $x = 0.5$ in such a way that the resulting aspect ratio of the central polygons is 10^4 (see the detail of the central band in Figure 7.5b).

As we can see from Figure 7.6, the use of badly shaped elements in conjunction with high order VEM ($k = 6$) causes large errors in the discrete solution, on the badly shaped elements, where we witness a wrong behaviour in the discrete solution (note the different behaviours in the region $\{x \sim 0.5,\, 0 \leq y \leq 0.5\}$ in Figure 7.6b compared to Figures 7.6a and 7.6c). These errors are remarkably reduced by the change of basis (Figure 7.6c). In this test, the orthogonal basis was used on all polygons.

7.6 Numerical results on Discrete Fracture Networks

In this section we consider a computational framework where instabilities arise when performing high order simulations, namely the computation of the hydraulic head inside Discrete



(a) Mesh

(b) Detail of the central polygon band of the mesh

Figure 7.5: **Interface** The mesh used for the test on the unity square.

Fracture Networks. These kind of domains are used in geomechanics to model fractured media in those cases where the rock matrix can be considered fully impervious: fractures are seen as planar polygons that intersect in the three-dimensional space, and the intersections are commonly called traces (see Figure 7.7 for a visualization of the DFNs that are considered in the following).

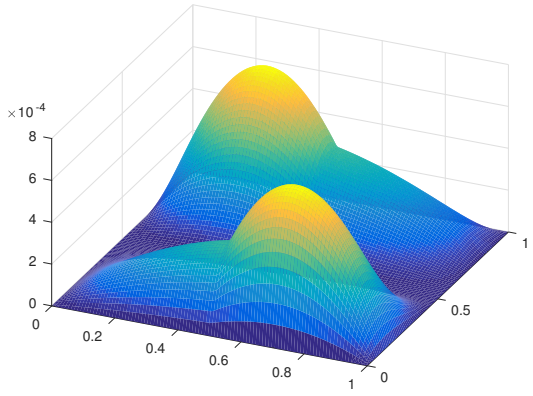
order	minimum aspect ratio	**m** polygons	ill-conditioned polygons	badly shaped polygons	both causes
5	150	4256	124	66	9
5	50	4177	115	145	18
5	10	3775	60	547	73
6	150	3193	1187	43	32
6	50	3143	1149	93	70
6	10	2888	947	348	272

Table 7.3: **DFN 27** Number of polygons where orthogonal polynomials were used and the motivations for their use.

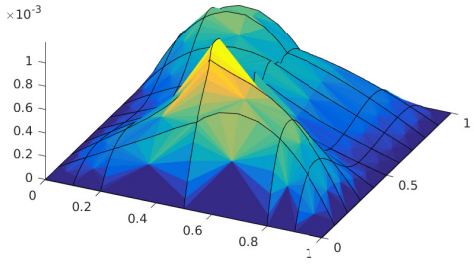
In practical applications, DFNs are generated randomly to respect the properties of the medium, which can be estimated experimentally, and are then used, for example, to determine certain quantities of interest through uncertainty quantification techniques [14].

In [7, 8, 11], the use of polygonal meshes in the VEM framework is exploited to obtain meshes which are conforming to traces, starting from an independent triangulation whose elements are then cut along the traces. Since these cuts are in fact random, the resulting polygons are convex but are likely to be very badly shaped.

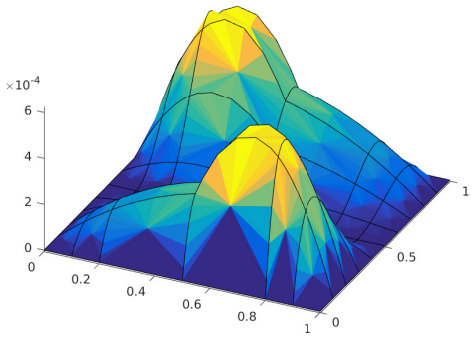
In order to circumvent the mesh generation problem an optimization approach working on totally non conforming meshes was developed [13, 15–19]. In this section we show that the



(a) Exact solution

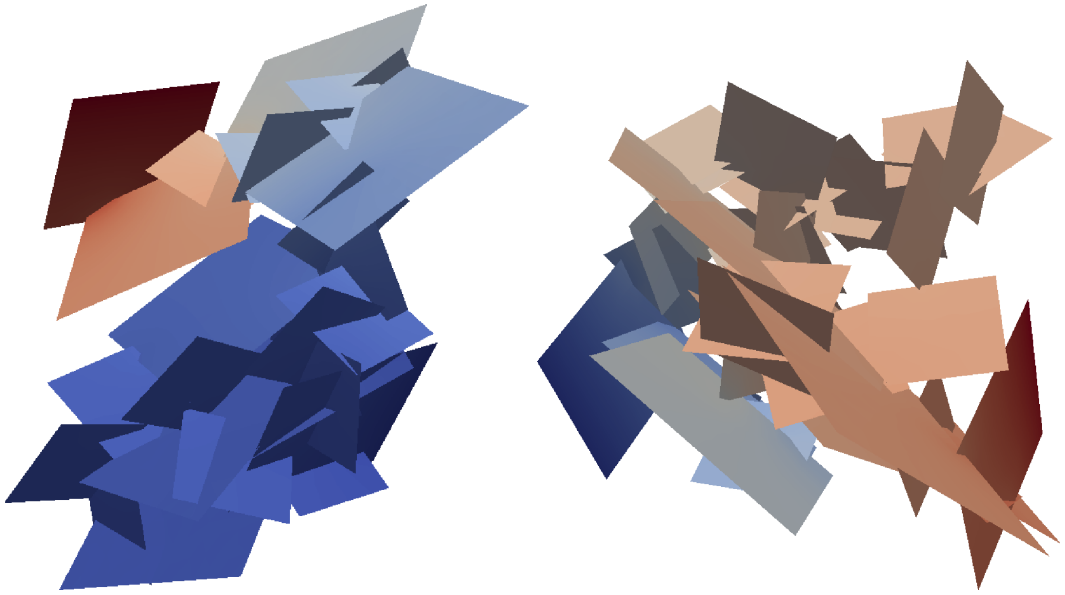


(b) Standard basis



(c) Orthogonal basis

Figure 7.6: **Interface**. Results for standard VEM of order 6 and VEM with orthogonal polynomials.



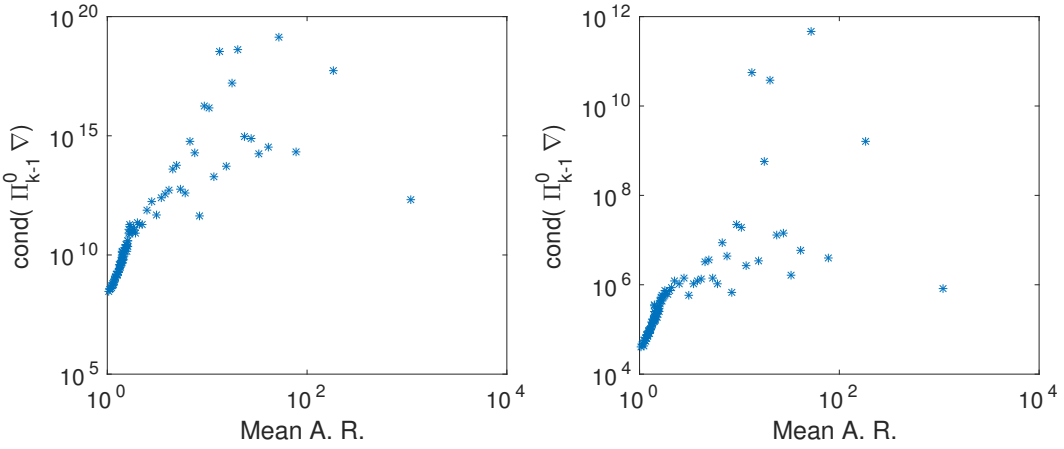
(a) **DFN 27** (b) **DFN 36**

Figure 7.7: The DFNs considered for numerical tests

use of orthogonal polynomials as described in the previous sections can prevent instabilities caused by a very large condition number of the projector matrices arising from the use of high order VEM on badly shaped polygons.

7.6.1 Mesh Generation process on the DFN fractures. In this subsection we briefly recall the process described in [11]: we refer the reader to this reference for a detailed description. A starting triangular mesh is generated on each fracture independently of traces (fracture intersections) position. The next process of polygonal mesh generation consists of the generation of a fracture-local mesh conforming to the traces, obtained splitting the triangles of the baseline mesh into polygons conforming to the traces, iteratively for all the traces. In this step if a trace ends within an original triangle or in one of the children polygons we extend the cut segment of this trace up to the next edge. In this operation the trace is unchanged: only the segment that is cutting the polygons is extended. All the points generated by intersections between cut segments and mesh edges are added to the mesh as new vertices. At the end of this step we have a polygonal mesh on each fracture that is *locally conforming* with the traces. Finally, for each couple of intersecting fractures F_i and F_j, generating the trace T_l, we consider on the trace the union of the mesh points coming from at least one of the two fractures that are on T_l. On each fracture, polygon edges lying on T_l are accordingly split in several aligned edges at the newly added points. In such a process we, first, generate a forest of polygons with root in the original triangles. Then, we modify the leaves polygons with edges on the traces converting the edges on the traces with the aligned edges generated by the mesh points on the trace of the twin fracture.

We remark that applying a preliminary mesh smoothing step as described in [8] the aspect ratio of many elements can be strongly reduced, nevertheless in these kind of applications the geometry can unavoidably produce very badly shaped elements whatever is the conforming



(a) Standard basis

(b) Orthogonal basis

Figure 7.8: **DFN 27**, order 6 Mean condition number of the matrix representation of $\Pi^0_{k-1}\nabla$ and its standard deviation.

mesh generation and smoothing process performed. In order to consider the worst possible cases, in the presented simulations we decide not to apply any mesh smoothing step.

7.6.2 Problem formulation on the DFN.

The computation of the hydraulic head on the DFN is provided by the solution of coupled problems on each fracture. The model we are considering is a simple Darcy model for the flow. Let $\mathcal{I}$ be the set of the indices of all the fractures in the DFN. The hydraulic head is given by the following equations $\forall i \in \mathcal{I}$:

$$\begin{cases} -\nabla \cdot (\mathsf{K}\nabla h) = 0 & \text{in } F_i, \\ h = h_D & \text{on } \partial F_{i,D}, \\ \nabla h \cdot \hat{\mathbf{n}} = 0 & \text{on } \partial F_{i,N}, \end{cases}$$

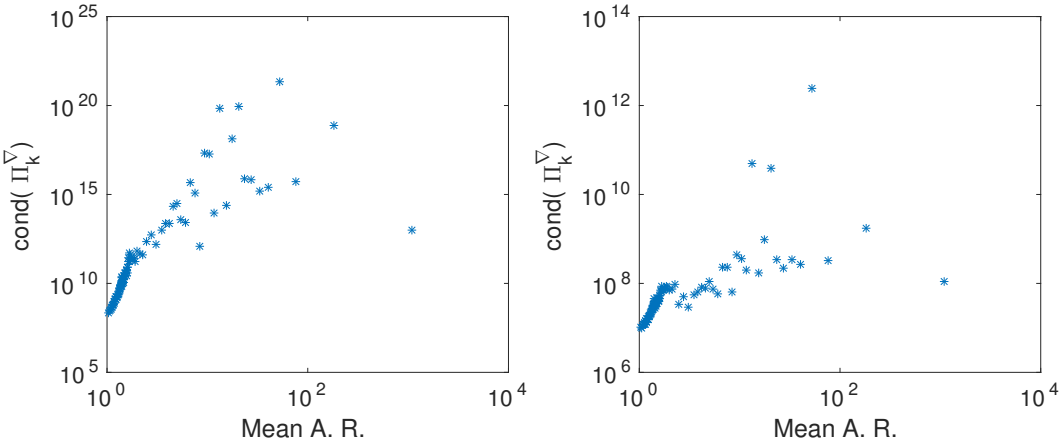
where $\partial F_{i,D}$ is the subset of the boundary of the fracture F_i with Dirichlet boundary conditions and $\partial F_{i,N}$ is the subset of the boundary of the fracture F_i with Neumann boundary conditions.

Continuity matching conditions for the solution h are imposed at the traces as in [11]. We set a non-homogeneous Dirichlet boundary condition on one side of a source fracture and a homogeneous Dirichlet condition on one side of a sink fracture and homogeneous Neumann boundary conditions on all the other fracture-sides of the DFN.

7.6.3 DFN 27.

We first consider a DFN composed by 27 fractures and displaying 57 traces (see Figure 7.7a). Starting from a mesh of triangular elements with area smaller than 60, we have created the globally conforming VEM polygonal mesh and assembled the linear system. We first focus on the condition numbers of the several projection matrices needed for the solution of the problem.

In Figures 7.8 to 7.10 we report the behaviour of the condition numbers of the projectors ${}^{\mathrm{m}}\boldsymbol{\Pi}^0_{k-1}\nabla$, ${}^{\mathrm{m}}\boldsymbol{\Pi}^{\nabla}_{k}$, ${}^{\mathrm{m}}\boldsymbol{\Pi}^0_{k-1}$, ${}^{\mathrm{p}}\boldsymbol{\Pi}^0_{k-1}\nabla$, ${}^{\mathrm{p}}\boldsymbol{\Pi}^{\nabla}_{k}$, ${}^{\mathrm{p}}\boldsymbol{\Pi}^0_{k-1}$, for different aspect ratios of the VEM polygonal elements using VEM of order 6, following the same procedure as in the plots of



(a) Standard basis

(b) Orthogonal basis

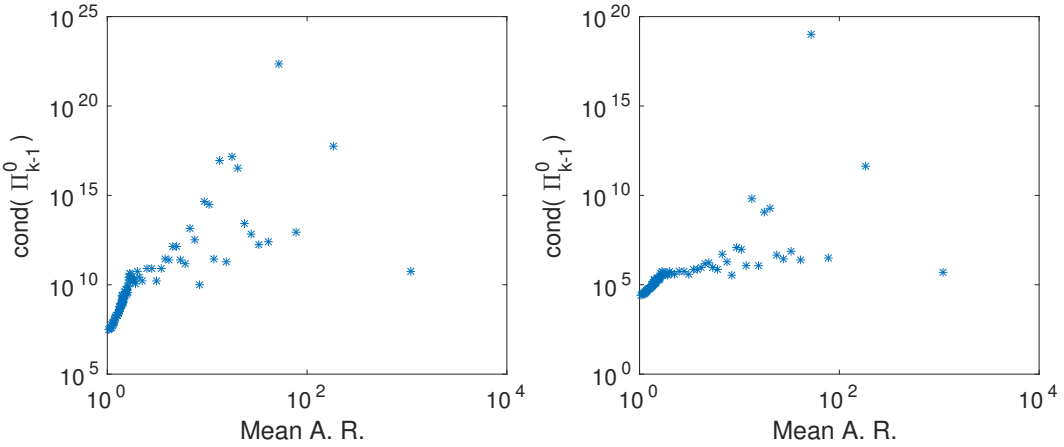
Figure 7.9: **DFN 27**, order 6 Mean condition number of the matrix representation of Π_k^∇ and its standard deviation.

order	minimum aspect ratio	**m** polygons	ill-conditioned polygons	badly shaped polygons	both causes
4	150	4465	22	49	3
4	50	4373	15	141	10
4	10	3874	1	640	24
5	150	4322	165	38	14
5	50	4234	154	126	25
5	10	3795	80	565	99

Table 7.4: **DFN 36** Number of polygons where orthogonal polynomials were used and the motivations for their use.

Figure 7.2. In order to draw these plots we compute the aspect ratio, defined as the ratio of the largest distance over the smallest distance between any couple of vertices of the polygon, of all the elements in the DFN and partition the full range of aspect ratios in 100 intervals uniformly. In the plots we report the mean condition numbers computed on all the elements with an aspect ratio in each of these intervals. In Figure 7.8 we compare the conditioning of ${}^{\mathbf{m}}\mathbf{\Pi}^0_{k-1}\nabla$ (left) and ${}^{\mathbf{p}}\mathbf{\Pi}^0_{k-1}\nabla$ (right), and we can appreciate a strong reduction of the mean condition numbers induced by the use of the basis **p**. We can came to the same conclusion observing Figure 7.9, concerning the projector used in the VEM stabilization, as well as Figure 7.10. Again, we remark that the effect of the change of basis is purely local, and the condition number of the global system is not significantly reduced. However, this process improves the quality of the local projections needed to build the final system, and this results to be sufficient to correct the instabilities.

In the following figures we report some examples of the instabilities due to the ill conditioned projectors obtained using the monomial basis **m** and the improved solution obtained with the new basis. In Figure 7.11 we show the low order solutions on two fractures in the DFN (Fracture 3 and Fracture 4) obtained with $k = 1$ and 4. Comparing these pictures we can appreciate an improvement in the quality of the solution using $k = 4$. In Figure 7.12



(a) Standard basis

(b) Orthogonal basis

Figure 7.10: **DFN 27**, order 6 Mean conditioning number of the matrix representation of Π^0_{k-1} and its standard deviation.

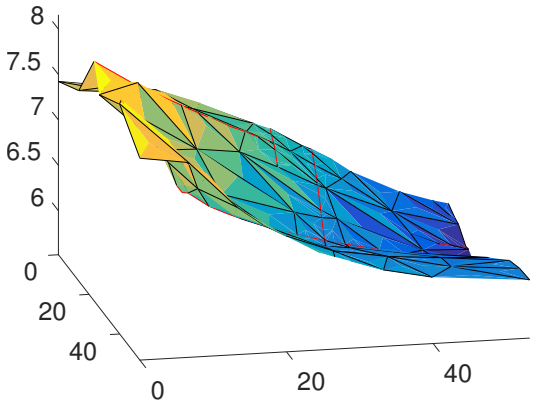
order	2	3	4	5	6
error $> 1e-4$	0	6	98	1105	4124
error > 1	0	0	8	29	352
error > 10	0	0	1	5	48
error > 100	0	0	0	1	6
max. orthog. error	$1.59 \cdot 10^{-10}$	$9.92 \cdot 10^{-01}$	$1.18 \cdot 10$	$1.77 \cdot 10^{3}$	$5.28 \cdot 10^{2}$

Table 7.5: **DFN 36** Counts of the elements with large orthogonalization error and maximum orthogonalization error for different orders.

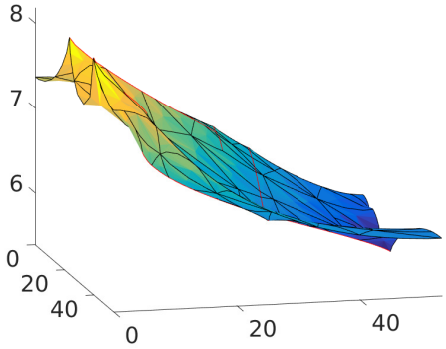
we report the solution obtained on Fracture 3 with $k = 5$ and 6. Observing Figures 7.12a and 7.12c compared with Figures Figures 7.11a and 7.11b, we can appreciate the instabilities arising due to the ill conditioning of the local matrices with respect to the monomial basis, that gets higher as the VEM order increases. We can see that both the shape of the solution and the values are completely wrong. In Figures Figures 7.12b and 7.12d we can see that the use of the basis $\mathbf{p}$ has a clear stabilizing effect. The same conclusion can be driven observing Figure 7.13 compared with Figures 7.11c and 7.11d.

For these results, orthogonal polynomials are used only on those polygons such that the conditioning number of the local ${}^{\mathbf{m}}\mathbf{H}^{k-1}$ is larger than 10^{10} or such that the aspect ratio is larger than 150.

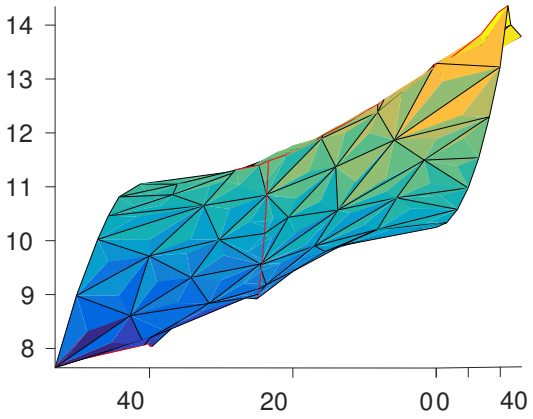
In Table 7.3 we report the number of polygons for which orthogonal polynomials are used for different threshold values on the aspect ratio, ranging from 10 to 150. The third column reports the number of polygons of the mesh where the monomial basis $\mathbf{m}$ is used, the fourth column reports the number of polygons on which the basis $\mathbf{p}$ is introduced only due to the large conditioning of the mass matrix ${}^{\mathbf{m}}\mathbf{H}^{k-1}$, in the fifth column the number of polygons on which $\mathbf{p}$ is used due to the large aspect ratio of the element. In the last column we report the number of polygons that require $\mathbf{p}$ for both the previous reasons.



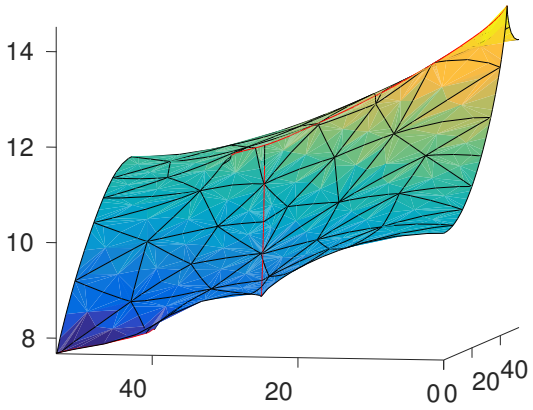
(a) Fracture 3, order 1



(b) Fracture 3, order 4

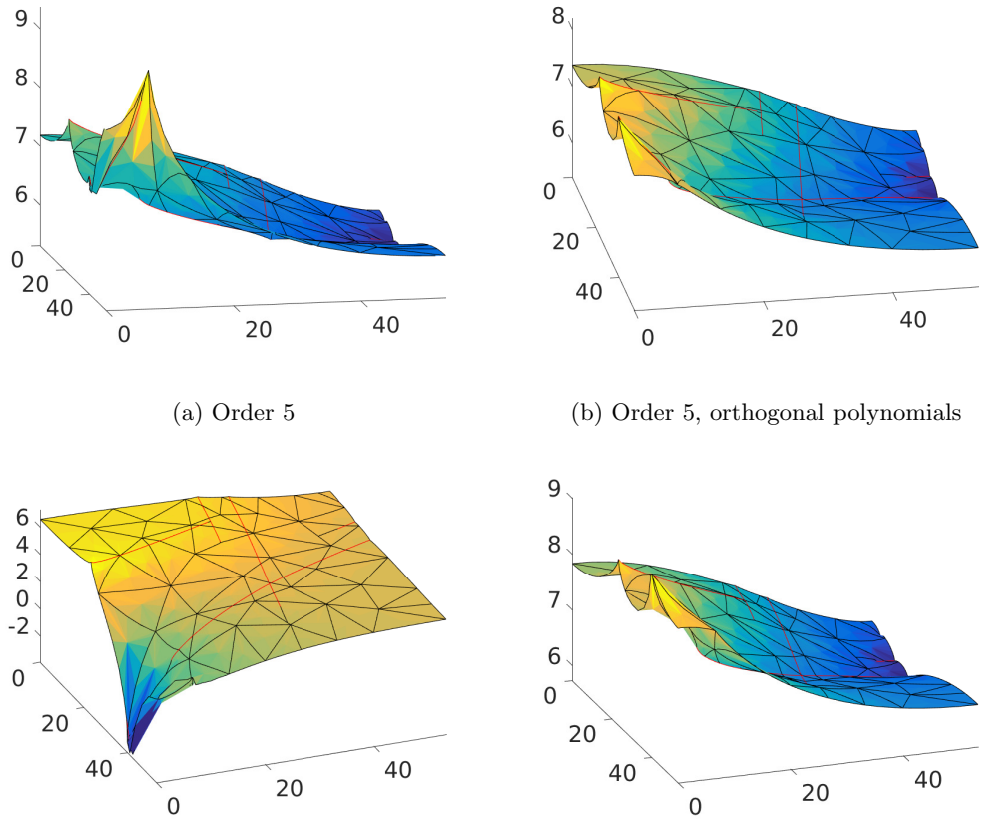


(c) Fracture 4, order 1



(d) Fracture 4, order 4

Figure 7.11: **DFN 27** Reference solutions with low order VEM.



(a) Order 5

(b) Order 5, orthogonal polynomials

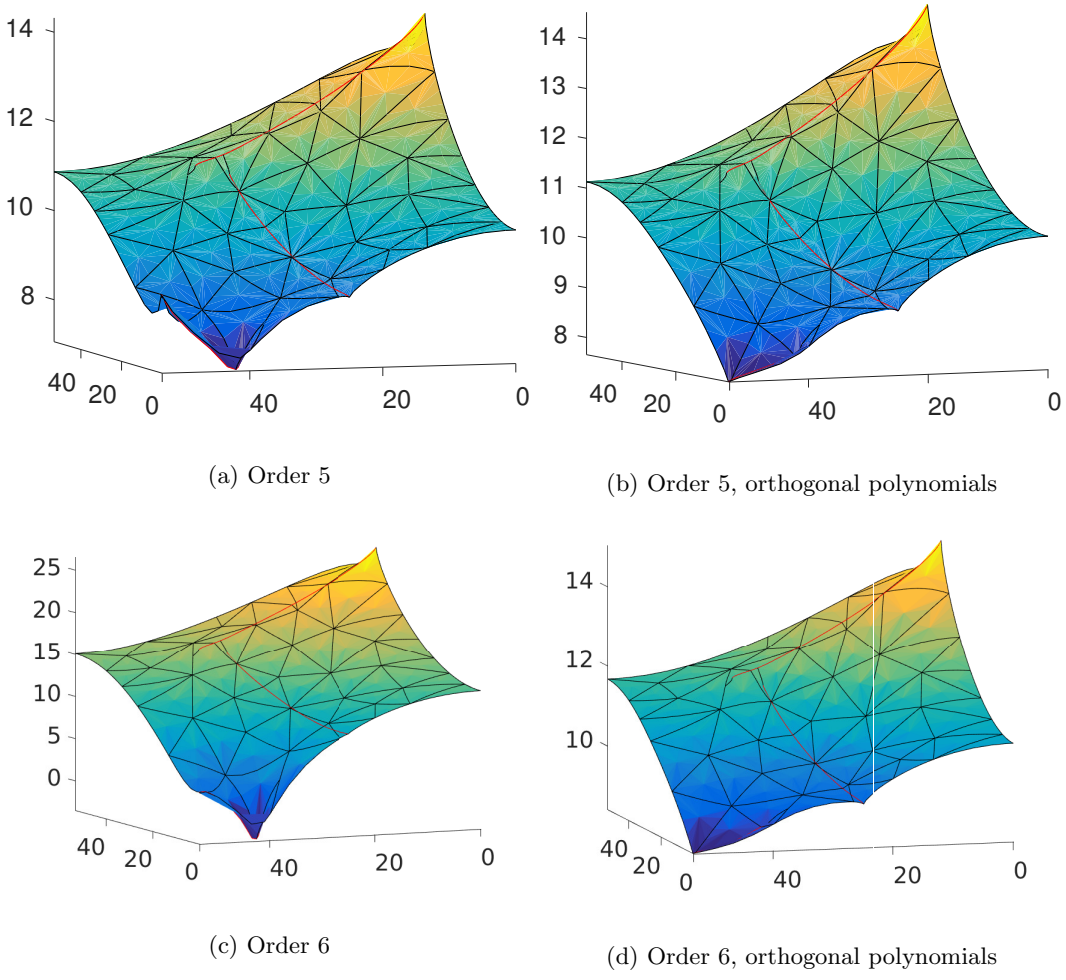
(c) Order 6

(d) Order 6, orthogonal polynomials

Figure 7.12: **DFN 27**, fracture 3 Solutions with increasing VEM order using standard VEM and behaviour of orthogonal polynomials in correcting the instabilities

7.6.4 DFN 36. Our second test considers a 36 fracture network with 65 traces. We focus on two particular fractures, where instabilities arise on high order VEM and observe, in Figures 7.14 and 7.15, how the use of the proposed basis for the space of polynomials in the construction of the projectors prevents the generation of non-physical oscillations. We notice that, although using the monomial basis the shape of the solution seems correct, its values are completely wrong (see Figures 7.14c, 7.14e, 7.15c and 7.15e). Again, the figures refer to the choice of applying the change of basis only on those polygons where ${}^{\mathbf{m}}\mathbf{H}^{k-1}$ displays a condition number larger than 10^{10} or with an aspect ratio greater than or equal to 150. In Table 7.4 we show how the condition number of the matrix ${}^{\mathbf{m}}\mathbf{H}^{k-1}$ is influenced by the shape of the polygons and the VEM order, and the number of elements on which the change of basis is applied. We notice again that it is sufficient to apply the change of basis only locally on certain polygons to cure global instabilities.

The proposed approach is effective for this DFN up to the VEM order 5, but it fails to stabilize the solution for VEM of order 6. Indeed, in Figure 7.16 we see that instabilities are still present even using orthogonal polynomials on all the elements (compare Figures 7.16a



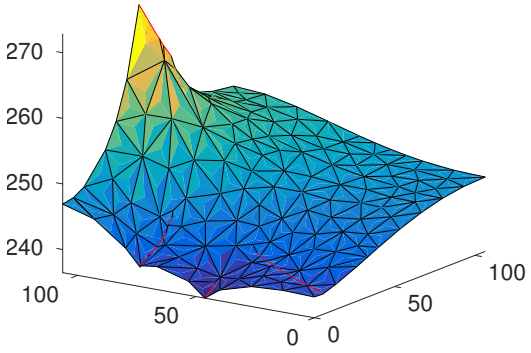
(a) Order 5

(b) Order 5, orthogonal polynomials

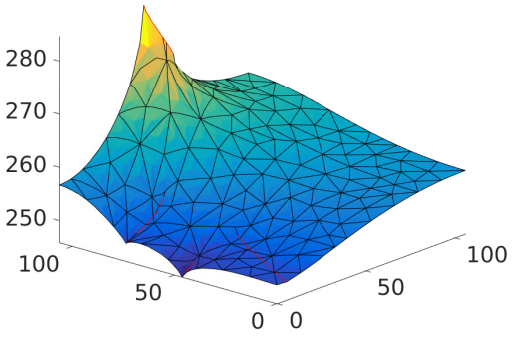
(c) Order 6

(d) Order 6, orthogonal polynomials

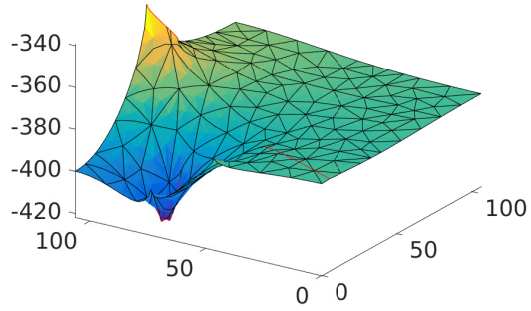
Figure 7.13: **DFN 27**, fracture 4 Solutions with increasing VEM order using standard VEM and behaviour of orthogonal polynomials in correcting the instabilities



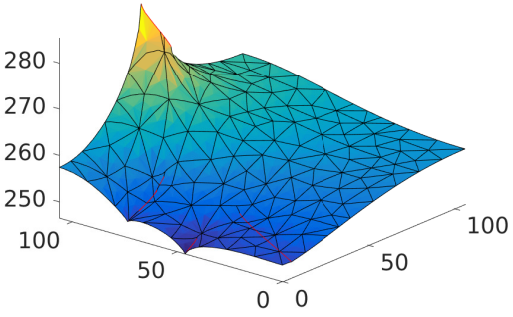
(a) Order 1



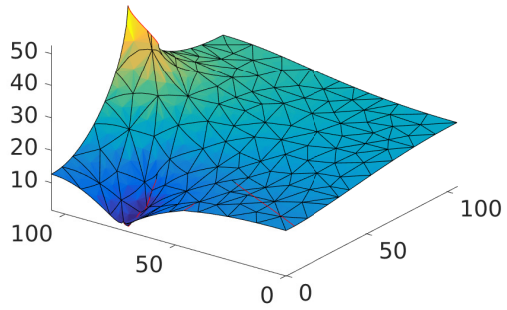
(b) Order 3



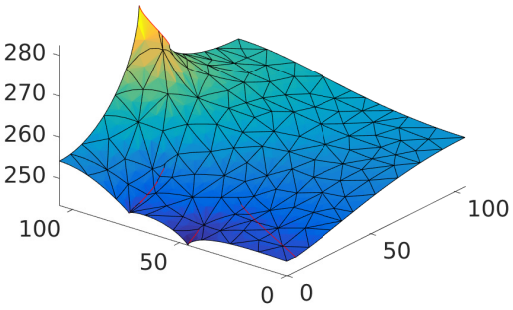
(c) Order 4



(d) Order 4, orthogonal polynomials



(e) Order 5



(f) Order 5, orthogonal polynomials

Figure 7.14: **DFN 36**, fracture 27 Solutions with increasing VEM order using standard VEM and behaviour of orthogonal polynomials in correcting the instabilities

and 7.16b with Figures 7.14a and 7.14b and Figures 7.16c and 7.16d with Figures 7.15a and 7.15b). This behaviour is related to the ill conditioning of some of the mass matrices ${}^{\mathbf{m}}\mathbf{H}^{k-1}$ that induces a large approximation error in the computation of the eigenvectors, that leads to a largely polluted polynomial basis. We remark that these situations can be easily detected by an evaluation of the orthogonalization error on each element:

$$\left\| \mathbf{Q}^{k-1}\, {}^{\mathbf{m}}\mathbf{H}^{k-1}\, \mathbf{Q}^{k-1\mathsf{T}} - \mathbf{I}^{k-1} \right\|_{\infty}. \tag{7.28}$$

In Figure 7.17 we report the orthogonalization error with respect to the aspect ratio of the elements, and in Figure 7.18 the orthogonalization error is plot with respect to the condition number of ${}^{\mathbf{m}}\mathbf{H}^{k-1}$. As expected, we can notice an evident correlation between them. We can remark that when these orthogonalization errors become large the generation of the orthogonal basis is not reliable and the method should be applied prudently. We can notice that for order 5 the orthogonalization error is large, but the method provides a basis for the space of polynomials that is still better than the scaled monomial basis. This is because only few elements are affected by a large error. In Table 7.5, we report the number of elements in the DFN with an orthogonalization error larger than $1.0E-4$, 1, 10, 100 for $k = 1, ..., 6$, and in the last row the largest orthogonalization error. In order to be more accurate also on problematic elements, in the computations we use equation (7.8) for the computation of ${}^{\mathbf{p}}\mathbf{H}^{k-1}$ instead of the identity matrix in order to take advantage from all that situations in which the basis $\mathbf{p}^{k-1}$ is no longer orthogonal, but provides a better conditioned mass matrix. As a rule of thumb we can say that when the largest orthogonalization error is not large or large orthogonalization errors occur on very few elements the method can be used, otherwise the computations cannot be considered reliable.

Finally, to further assess the behaviour of the method, we show in Figure 7.19 the effect of the change of basis on the conditioning of the matrices representing the projectors $\Pi^0_{k-1}\nabla$, Π^{∇}_k and Π^0_{k-1}, respectively. These graphs show the mean condition number with respect to the aspect ratio of all the elements of the DFN. We see how the use of orthogonal polynomials strongly mitigates the dependance of the condition number on the aspect ratio.

7.7 Conclusions

Dealing with problems with very complex geometries can easily lead to very strong mesh generation problems. In these situations the use of more flexible polygonal methods is very helpful. The VEM is a suitable and effective approach for the discretization of Partial Differential Equations. Nevertheless, in some of these applications the polygonal mesh generated for the VEM applications can suffer for very low quality elements. An applicative example in which these situations are likely to happen is in geophysical simulations following the DFN model. For the most badly shaped elements the use of the classical monomial basis for the construction of the local matrices can lead to large problems due to the large condition number of the local matrices.

In this paper, for high order VEM, we have presented the construction of a polynomial basis that leads to better conditioned local matrices and more accurate solutions. The construction is based on a local eigenvalue-eigenvector computation. This approach is very effective for very badly shaped elements, but for some elements with a huge aspect ratio the eigenvalue-eigenvector problem can be inaccurate and also this approach does not provide a reliable solution.

We have reported the success of the method in providing good solutions in some applications and have provided a criterion to evaluate the reliability of the method when the most

problematic elements are met. The method has also the attractive property to be simply added to a standard VEM implementation and can be applied selectively only on the elements that really need an improvement in term of accuracy of the computations, and provides an indicator that alerts the user when the method is no longer reliable.

Acknowledgments

We wish to thank Matías Fernando Benedetto for his crucial help in developing the code that was used for numerical simulations.

References for Chapter 7

[1] P. M. Adler. *Fractures and Fracture Networks*. Kluwer Academic, Dordrecht, 1999.

[2] B. Ahmad, A. Alsaedi, F. Brezzi, L. D. Marini, and A. Russo. "Equivalent projectors for virtual element methods". In: *Computers & Mathematics with Applications* 66 (3 Sept. 2013), pp. 376–391.

[3] L. Beirão da Veiga, F. Brezzi, A. Cangiani, G. Manzini, L. D. Marini, and A. Russo. "Basic principles of virtual element methods". In: *Mathematical Models and Methods in Applied Sciences* 23.01 (2013), pp. 199–214.

[4] L. Beirão Da Veiga, F. Brezzi, L. D. Marini, and A. Russo. "The hitchhiker's guide to the Virtual Element method". In: *Math. Models Methods Appl. Sci* 24.8 (2014), pp. 1541–1573.

[5] L. Beirão da Veiga, F. Brezzi, L. D. Marini, and A. Russo. "Virtual Element Methods for General Second Order Elliptic Problems on Polygonal Meshes". In: *Mathematical Models and Methods in Applied Sciences* 26.04 (2015), pp. 729–750. DOI: 10.1142/S0218202516500160.

[6] L. Beirão da Veiga, C. Lovadina, and A. Russo. *Stability Analysis for the Virtual Element Method*. 2016. arXiv: 1607.05988.

[7] M. Benedetto, S. Berrone, and A. Borio. "The Virtual Element Method for underground flow simulations in fractured media". In: *Advances in Discretization Methods*. Vol. 12. SEMA SIMAI Springer Series. Switzerland: Springer International Publishing, 2016. Chap. 8, pp. 167–186. DOI: 10.1007/978-3-319-41246-7.

[8] M. Benedetto, S. Berrone, A. Borio, S. Pieraccini, and S. Scialò. "A Hybrid Mortar Virtual Element Method For Discrete Fracture Network Simulations". In: *J. Comput. Phys.* 306 (2016), pp. 148–166. DOI: 10.1016/j.jcp.2015.11.034.

[9] M. Benedetto, S. Berrone, A. Borio, S. Pieraccini, and S. Scialò. "Order preserving SUPG stabilization for the virtual element formulation of advection-diffusion problems". In: *Computer Methods in Applied Mechanics and Engineering* 311 (2016), pp. 18–40. ISSN: 0045-7825. DOI: 10.1016/j.cma.2016.07.043.

[10] M. Benedetto, S. Berrone, S. Pieraccini, and S. Scialò. "The virtual element method for discrete fracture network simulations". In: *Comput. Methods Appl. Mech. Engrg.* 280.0 (2014), pp. 135–156. ISSN: 0045-7825. DOI: 10.1016/j.cma.2014.07.016.

[11] M. Benedetto, S. Berrone, and S. Scialò. "A Globally Conforming Method For Solving Flow in Discrete Fracture Networks Using the Virtual Element Method". In: *Finite Elem. Anal. Des.* 109 (2016), pp. 23–36. DOI: 10.1016/j.finel.2015.10.003.

[12] S. Berrone and A. Borio. "Orthogonal polynomials in badly shaped polygonal elements for the Virtual Element Method". In: *Finite Elements in Analysis & Design* 129 (C 2017), pp. 14–31. ISSN: 0168-874X. DOI: 10.1016/j.finel.2017.01.006.

[13] S. Berrone, A. Borio, and S. Scialò. "A posteriori error estimate for a PDE-constrained optimization formulation for the flow in DFNs". In: *SIAM J. Numer. Anal.* 54.1 (2016), pp. 242–261. DOI: 10.1137/15M1014760.

[14] S. Berrone, C. Canuto, S. Pieraccini, and S. Scialò. "Uncertainty quantification in Discrete Fracture Network models: stochastic fracture transmissivity". In: *Comput. Math. Appl.* 70.4 (2015), pp. 603–623. DOI: 10.1016/j.camwa.2015.05.013.

[15] S. Berrone, S. Pieraccini, and S. Scialò. "A PDE-constrained optimization formulation for discrete fracture network flows". In: *SIAM J. Sci. Comput.* 35.2 (2013), B487–B510. ISSN: 1064-8275. DOI: 10.1137/120865884.

[16] S. Berrone, S. Pieraccini, and S. Scialò. "An optimization approach for large scale simulations of discrete fracture network flows". In: *J. Comput. Phys.* 256 (2014), pp. 838–853. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2013.09.028.

[17] S. Berrone, S. Pieraccini, and S. Scialò. "On simulations of discrete fracture network flows with an optimization-based extended finite element method". In: *SIAM J. Sci. Comput.* 35.2 (2013), A908–A935. ISSN: 1064-8275. DOI: 10.1137/120882883.

[18] S. Berrone, S. Pieraccini, and S. Scialò. "Towards effective flow simulations in realistic discrete fracture networks". In: *J. Comput. Phys.* 310 (2016), pp. 181–201. DOI: 10.1016/j.jcp.2016.01.009.

[19] S. Berrone, S. Pieraccini, S. Scialò, and F. Vicini. "A parallel solver for large scale DFN flow simulations". In: *SIAM J. Sci. Comput.* 37.3 (2015), pp. C285–C306. DOI: 10.1137/140984014.

[20] M. C. Cacas, E. Ledoux, G. de Marsily, B. Tillie, A. Barbreau, E. Durand, B. Feuga, and P. Peaudecerf. "Modeling fracture flow with a stochastic discrete fracture network: calibration and validation: 1. The flow model". In: *Water Resour. Res.* 26 (1990), pp. 479–489. DOI: 10.1029/WR026i003p00479.

[21] M. Cravero and C. Fidelibus. "A code for scaled flow simulations on generated fracture networks". In: *Comput. Geosci.* 25.2 (1999), pp. 191–195.

[22] W. S. Dershowitz and C. Fidelibus. "Derivation of equivalent pipe networks analogues for three-dimensional discrete fracture networks by the boundary element method". In: *Water Resource Res.* 35 (1999), pp. 2685–2691. DOI: 10.1029/1999WR900118.

[23] J.-R. de Dreuzy, G. Pichot, B. Poirriez, and J. Erhel. "Synthetic benchmark for modeling flow in 3D fractured media". In: *Computers & Geosciences* 50.0 (2013), pp. 59–71.

[24] C. Fidelibus. "The 2D hydro-mechanically coupled response of a rock mass with fractures via a mixed BEM-FEM technique". In: *International Journal for Numerical and Analytical Methods in Geomechanics* 31.11 (2007), pp. 1329–1348.

[25] C. Fidelibus, G. Cammarata, and M. Cravero. *Hydraulic characterization of fractured rocks. In: Abbie M, Bedford JS (eds) Rock mechanics: new research.* Nova Science Publishers Inc., New York, 2009.

[26] J. Hyman, C. Gable, S. Painter, and N. Makedonska. "Conforming Delaunay Triangulation of Stochastically Generated Three Dimensional Discrete Fracture Networks: A Feature Rejection Algorithm for Meshing Strategy". In: *SIAM Journal on Scientific Computing* 36 (4 2014), A1871–A1894. DOI: 10.1137/130942541.

[27] J. Jaffré and J. E. Roberts. "Modeling flow in porous media with fractures; Discrete fracture models with matrix-fracture exchange". In: *Numerical Analysis and Applications* 5.2 (2012), pp. 162–167.

[28] V. Lenti and C. Fidelibus. "A *BEM* solution of steady-state flow problems in discrete fracture networks with minimization of core storage". In: *Computers & Geosciences* 29.9 (2003), pp. 1183–1190. ISSN: 0098-3004. DOI: 10.1016/S0098-3004(03)00140-7.

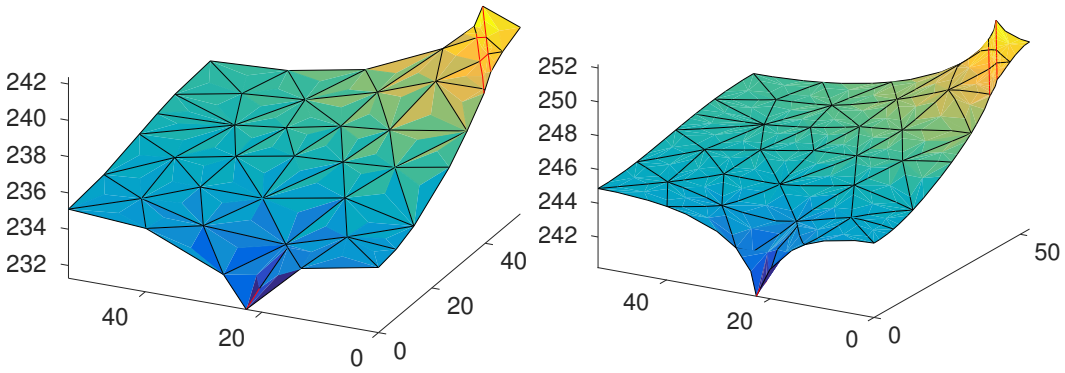
[29] B. Nœtinger. "A quasi steady state method for solving transient Darcy flow in complex 3D fractured networks accounting for matrix to fracture flow". In: *J. Comput. Phys.* 283 (2015), pp. 205–223. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2014.11.038.

[30] B. Nœtinger and N. Jarrige. "A quasi steady state method for solving transient Darcy flow in complex 3D fractured networks". In: *J. Comput. Phys.* 231.1 (2012), pp. 23–38. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2011.08.015.

[31] G. Pichot, J. Erhel, and J. de Dreuzy. "A generalized mixed hybrid mortar method for solving flow in stochastic discrete fracture networks". In: *SIAM Journal on scientific computing* 34 (2012), B86–B105. DOI: 10.1137/100804383.

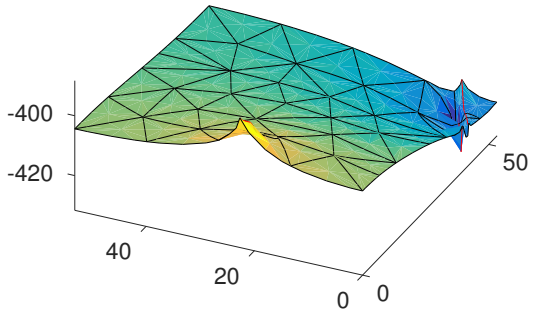
[32] G. Pichot, J. Erhel, and J. de Dreuzy. "A mixed hybrid Mortar method for solving flow in discrete fracture networks". In: *Applicable Analysis* 89 (2010), pp. 1629–643. DOI: 10.1080/00036811.2010.495333.

[33] G. Pichot, B. Poirriez, J. Erhel, and J.-R. de Dreuzy. "A Mortar BDD method for solving flow in stochastic discrete fracture networks". In: *Domain Decomposition Methods in Science and Engineering XXI*. Lecture Notes in Computational Science and Engineering. Springer, 2014, pp. 99–112.

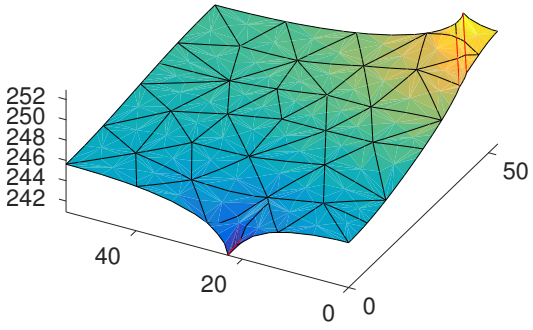


(a) Order 1

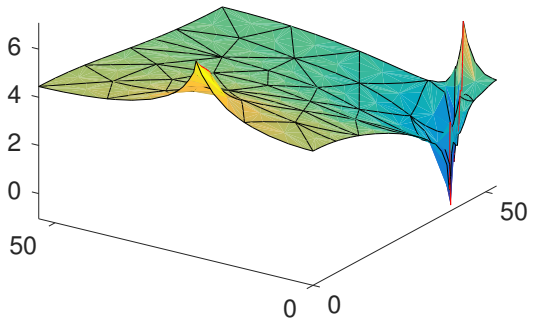
(b) Order 3



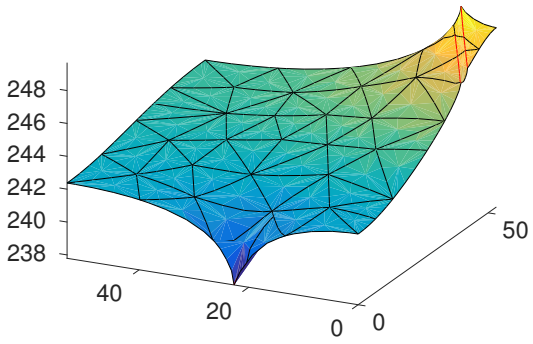
(c) Order 4



(d) Order 4, orthogonal polynomials

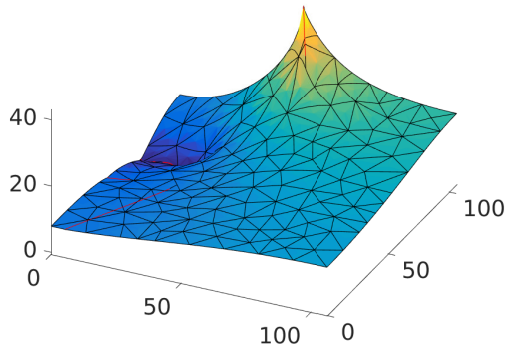


(e) Order 5

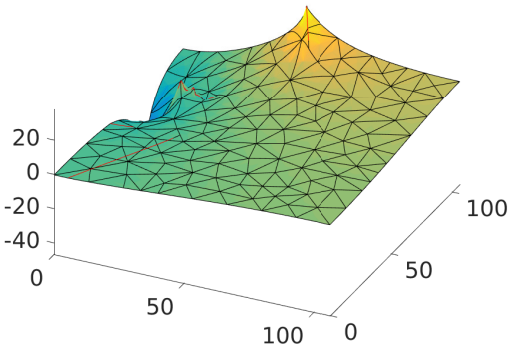


(f) Order 5, orthogonal polynomials

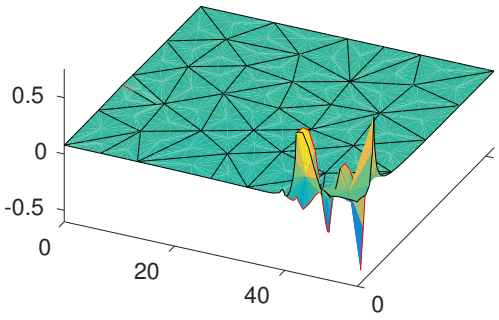
Figure 7.15: **DFN 36**, fracture 29 Solutions with increasing VEM order using standard VEM and behaviour of orthogonal polynomials in correcting the instabilities



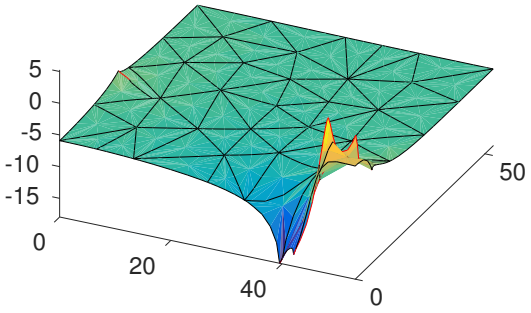
(a) Fracture 27, standard basis



(b) Fracture 27, orthogonal polynomials

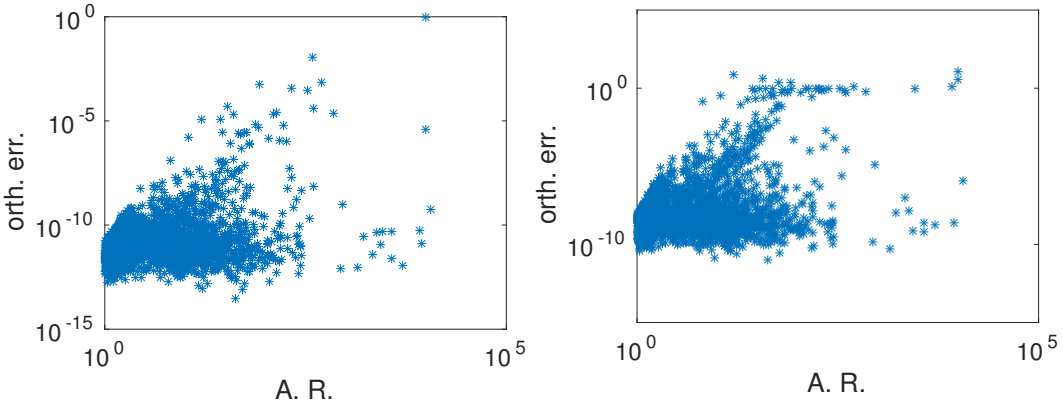


(c) Fracture 29, standard basis



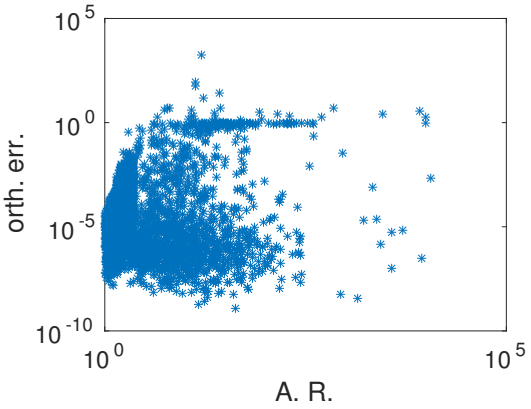
(d) Fracture 29, orthogonal polynomials

Figure 7.16: **DFN 36**, order 6 Solutions using standard VEM polynomial basis and orthogonal polynomials

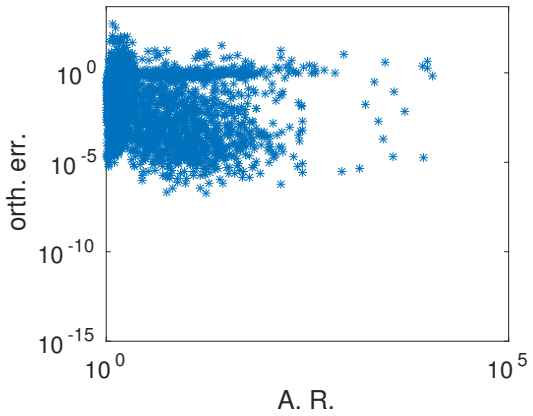


(a) Order 3

(b) Order 4

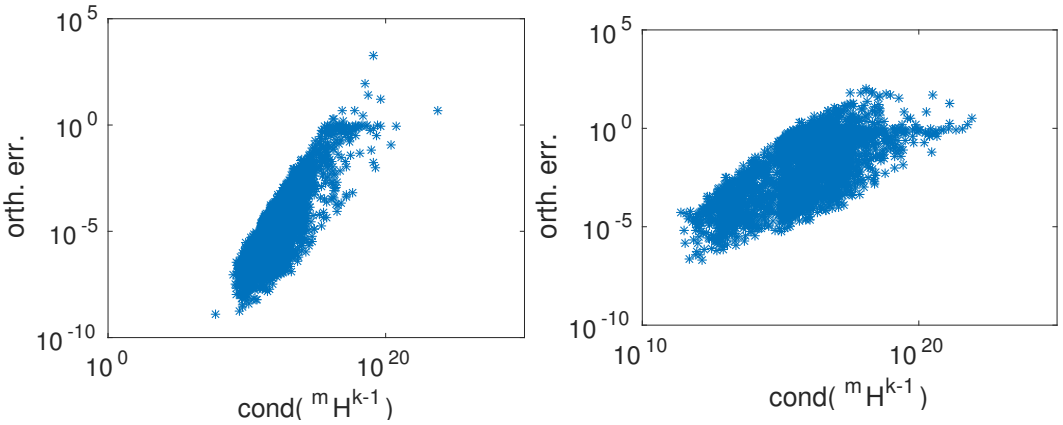


(c) Order 5



(d) Order 6

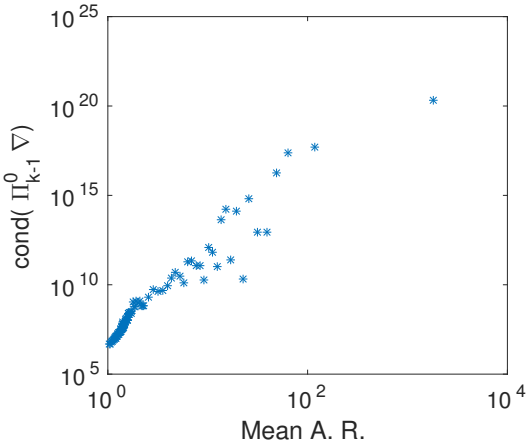
Figure 7.17: **DFN 36** Error of orthogonalization of ${}^{\mathbf{m}}\mathbf{H}^{k-1}$ vs. aspect ratio



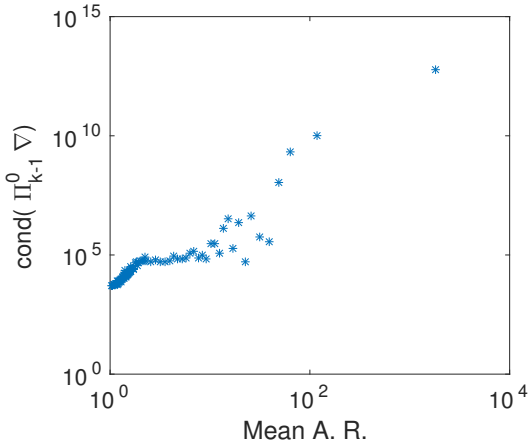
(a) Order 5

(b) Order 6

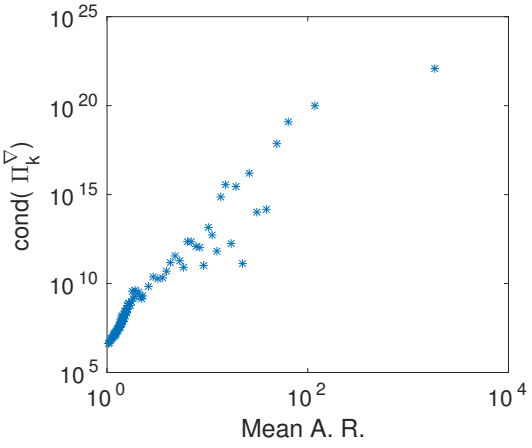
Figure 7.18: **DFN 36** Error of orthogonalization of $^{\mathbf{m}}\mathbf{H}^{k-1}$ vs. its condition number



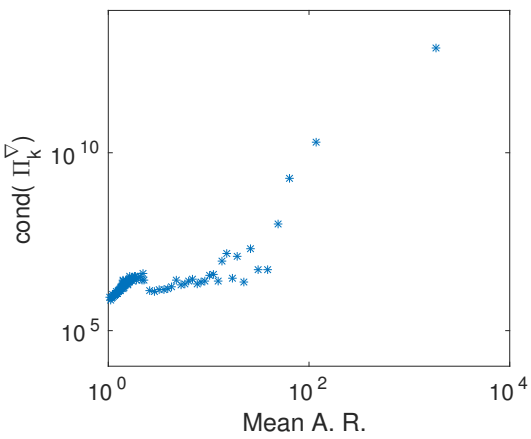
(a) Standard basis, $k = 5$



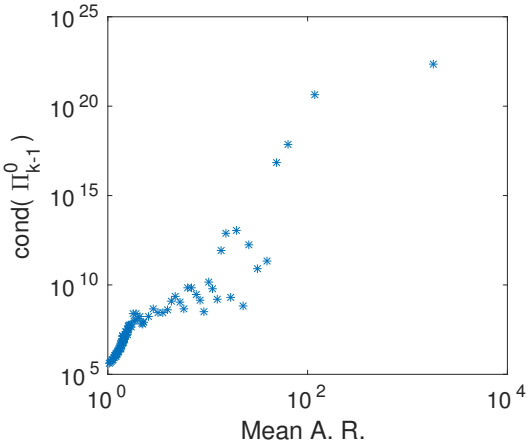
(b) Orthogonal basis, $k = 5$



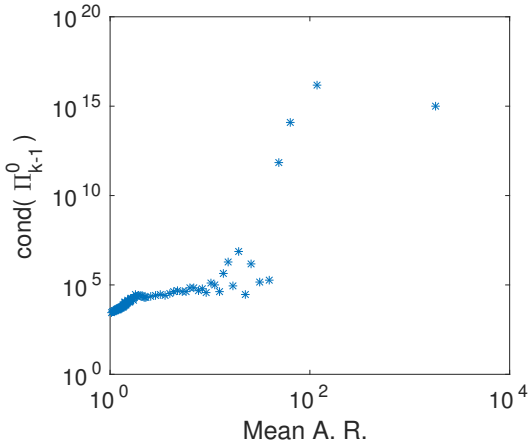
(c) Standard basis, $k = 5$



(d) Orthogonal basis, $k = 5$



(e) Standard basis, $k = 5$



(f) Orthogonal basis, $k = 5$

Figure 7.19: **DFN 36**, order 5 Mean condition number and standard deviation of $\Pi^0_{k-1}\nabla$, $\Pi^0_{k-1}\nabla$ and Π^0_{k-1}.

Chapter 8

Conclusions about VEM in DFN simulations

The use of generally shaped polygons to discretize the domain can become crucial in the context of Discrete Fracture Network flows simulation, expecially as the number of traces per fracture increases and some kind of conformity to intersection is required, as it is for the most common domain decomposition techniques. The recently developed Virtual Element Method allows such flexibility, and thus it is very suitable for DFN simulations.

We have developed a general framework for the use of VEM in the computation of the distribution of hydraulic head in a DFN (see also [1]), with a particular focus on the coupling of the Mortar method with the Virtual Element Method [2], proving optimal rates of convergence and obtaining good numerical results. In view of the solution of advection-diffusion problems on DFNs, with such applications in mind as the simulation of the stationary distribution of a passive pollutant in the underground, we developed and analyzed a Streamline Upwind Petrov Galerkin formulation of the method (see also [3]), preserving optimal rates of convergence and stabilizing advection dominated problems, whose solution can be completely altered when obtained by a pure Galerkin approach. Furthermore, since the method could strongly benefit from an adaptive strategy, we considered the problem of the a posteriori estimation of the error, developing error estimators which are independent of the stabilization terms introduced by the VEM, which are arbitrary and cannot be fully estimated. Finally, since the meshing process we adopted can sometimes generate very badly shaped polygons which yield instabilities on high order VEM, we pointed out a possible way to avoid such instabilities without changing the mesh, but solving local eigenproblems in order to obtain a local quasi orthogonal polynomial basis to be used for computing the projections needed for the method [4].

These results will be the starting point for the study of more complex situations such as the simulation of time evolving transport of pollutants in the DFN and the coupling of the DFN with a non impervious surrounding rock matrix. Future work will also include the development of an adaptive refinement strategy, exploiting the a posteriori error estimates in order to obtain a well balanced distribution of the error in the domain.

References for Chapter 8

[1] M. Benedetto, S. Berrone, and A. Borio. "The Virtual Element Method for underground flow simulations in fractured media". In: *Advances in Discretization Methods*. Vol. 12. SEMA SIMAI Springer Series. Switzerland: Springer International Publishing, 2016. Chap. 8, pp. 167–186. DOI: 10.1007/978-3-319-41246-7.

[2] M. Benedetto, S. Berrone, A. Borio, S. Pieraccini, and S. Scialò. "A Hybrid Mortar Virtual Element Method For Discrete Fracture Network Simulations". In: *J. Comput. Phys.* 306 (2016), pp. 148–166. DOI: 10.1016/j.jcp.2015.11.034.

[3] M. Benedetto, S. Berrone, A. Borio, S. Pieraccini, and S. Scialò. "Order preserving SUPG stabilization for the virtual element formulation of advection-diffusion problems". In: *Computer Methods in Applied Mechanics and Engineering* 311 (2016), pp. 18–40. ISSN: 0045-7825. DOI: 10.1016/j.cma.2016.07.043.

[4] S. Berrone and A. Borio. "Orthogonal polynomials in badly shaped polygonal elements for the Virtual Element Method". In: *Finite Elements in Analysis & Design* 129 (C 2017), pp. 14–31. ISSN: 0168-874X. DOI: 10.1016/j.finel.2017.01.006.

Appendix A

A posteriori error estimate for a PDE constrained optimization formulation for the flow in DFNs

This appendix contains the work published in [5], concerning the development of an a posteriori error estimate for a PDE constrained optimization approach to the computation of the hydraulic head inside a Discrete Fracture Network.

A.1 Introduction

In the approach developed in [6–8] the DFN problem is seen as a PDE constrained optimization problem, in which a cost functional measuring the discontinuity and flux unbalance at fracture intersections is minimized, constrained by the Darcy law on the fractures. In this framework, no mesh conformity is required at fracture intersections and the solution is obtained through the resolution of small weakly dependent sub-problems on the fractures with an iterative solver. Any difficulty related to the generation of the mesh is avoided and the approach has a natural parallel implementation with good scalability performances [9]. Further, no modification of the geometry of the network is required, and this is particularly important for uncertainty quantification procedures, in which a modification of the disposition of fracture would imply a modification of the probabilistic law at the basis of the generation of the network.

In the present chapter, residual based *a posteriori* error estimates [1, 11, 12, 15–17] are derived for the optimization formulation of the DFN problem described above, in view of a possible future use within an adaptive algorithm. In deriving the a posteriori error estimates, particular attention is devoted to highlight the effect of discontinuities of the discrete solution and unbalance of fluxes at fracture intersections that can cross the interior of mesh elements. Indeed, the error estimator proposed herein contains several additional terms with respect to classical residual based a posteriori error estimates; some of these additional terms exploit known properties of the exact solution. Moreover, part of the work is devoted to estimate the errors generated by the non-conformity of triangles to fracture intersections and to track the influence of this non-conformity on the effectivity of the global estimate. In particular, in deriving the lower bounds (Theorem A.3) we explicitly define a non-conformity coefficient

that affects the effectivity index.

The structure of the paper is as follows. In Appendix A.2 some useful notations concerning DFNs are introduced; in Appendix A.3 the problem and its discrete formulation are stated; in Appendix A.4 and Appendix A.5 suitable estimators are defined and an upper bound of the error is provided; in Appendix A.6 the efficiency of these estimators is proved and in Appendix A.7 some numerical results are described.

A.2 Nomenclature and main assumptions

In the present work we consider a network of fractures surrounded by an impervious rock matrix, with flow occurring only through fractures and across fracture intersections in the normal direction. Let us denote by Ω the DFN, composed of N intersecting fractures (see Figure A.1a). Each fracture F_i, $i \in \mathcal{I} = \{1, \dots, N\}$ is a planar open polygon, with boundary ∂F_i and the boundary of Ω is $\partial\Omega = \bigcup_{i\in\mathcal{I}} \partial F_i$. We assume that all the fractures in Ω are connected, i.e. each fracture has at least one intersection with another fracture in the network, and we call *traces* these intersections, each denoted by Γ_m, with $m \in \mathcal{M} = \{1, \dots, M\}$. For the sake of simplicity we assume that there are no intersections between traces and that each trace is shared by exactly two fractures, so, if $\Gamma_m = \bar{F}_i \cap \bar{F}_j$, there is a bijective correspondence between the index m and the couple of indexes (i, j), thus allowing us to define the ordered couple $\mathcal{I}_m = (i, j)$, $i < j$, (see [6] for relaxing these hypotheses). We further introduce, for each fracture F_i, the ordered set $\mathcal{M}_i \subset \mathcal{M}$ (Figure A.1a) collecting indexes of traces belonging to $\bar{F}_i$ in increasing order, with $M_i = \#\mathcal{M}_i$. $\mathcal{M}_i(k)$, for $k = 1, \dots, M_i$ indicates the k-th index of a trace in $\mathcal{M}_i$. For each $i \in \mathcal{I}$ and each $m \in \mathcal{M}_i$, $\hat{n}^i_m$ is a fixed normal unit vector to the trace Γ_m on F_i (Figure A.1b). The reader can refer to Figure A.2 for some simple DFNs and to [2, 4, 9] for more complex ones.

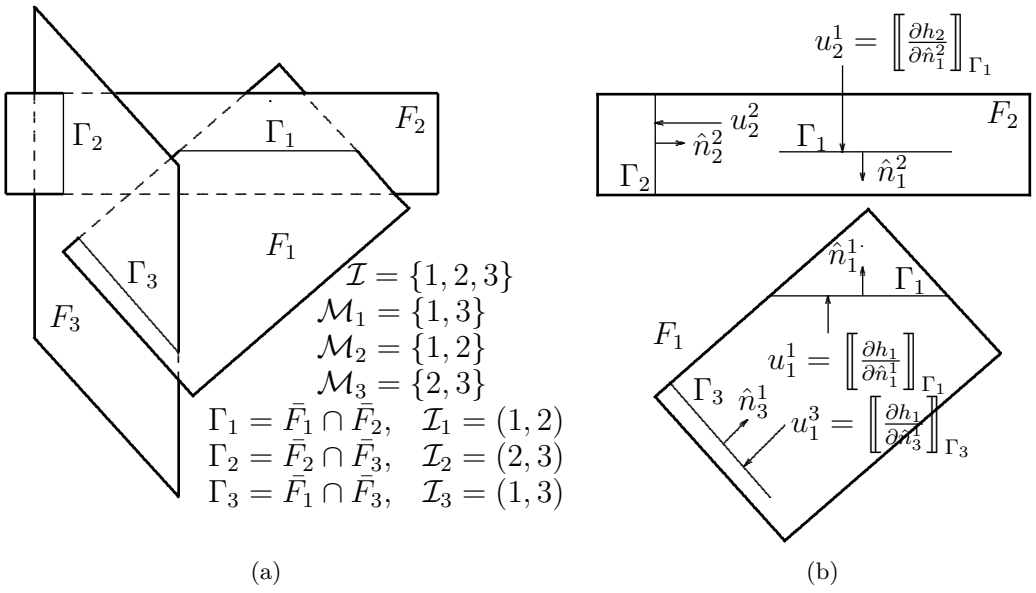


Figure A.1: (a) Simple DFN with three rectangular fractures. For each fracture F_i we list the set of the indices of the traces of that fracture $\mathcal{M}_i$, and for each trace Γ_m we list the ordered set of the intersecting fractures along that trace $\mathcal{I}_m = (i, j)$, with $i < j$. (b) Detail of normals and fluxes on the fractures F_1 and F_2.

For any given segment $\sigma \subset F_i$, $i \in \mathcal{I}$, $\gamma^i_\sigma : \mathrm{H}^1_0(F_i) \to \mathrm{H}^{\frac{1}{2}}(\sigma)$ is the trace operator and

$$\langle \mu, \beta \rangle_\sigma := {}_{\mathrm{H}^{-\frac{1}{2}}(\sigma)} \langle \mu, \beta \rangle_{\mathrm{H}^{\frac{1}{2}}(\sigma)} \quad \forall \mu \in \mathrm{H}^{-\frac{1}{2}}(\sigma), \forall \beta \in \mathrm{H}^{\frac{1}{2}}(\sigma)$$

is the duality between $\mathrm{H}^{\frac{1}{2}}(\sigma)$ and $\mathrm{H}^{-\frac{1}{2}}(\sigma)$. For any given function $v \in \mathrm{H}^1_0(F_i)$, $\gamma_{\mathcal{M}_i}(v) \in \prod_{m \in \mathcal{M}_i} \mathrm{H}^{\frac{1}{2}}(\Gamma_m)$ is the tuple of functions $\gamma^i_{\Gamma_m}(v)$, $m \in \mathcal{M}_i$ ordered by increasing trace index m, and we denote the duality between product spaces on the set of the traces of a fracture as

$$\forall \boldsymbol{\mu} \in \prod_{m \in \mathcal{M}_i} \mathrm{H}^{-\frac{1}{2}}(\Gamma_m), \forall \boldsymbol{\beta} \in \prod_{m \in \mathcal{M}_i} \mathrm{H}^{\frac{1}{2}}(\Gamma_m), \quad \langle \boldsymbol{\mu}, \boldsymbol{\beta} \rangle_{\mathcal{M}_i} := \sum_{m \in \mathcal{M}_i} \langle \mu_m, \beta_m \rangle_{\Gamma_m}.$$

Let us introduce the functional space $V := V_1 \times \ldots \times V_N$, where $V_i := \mathrm{H}^1_{0,F_i}(1)$, $\forall i \in \mathcal{I}$. For any function $\mathbf{g} \in V$, we define the jump operator across a trace Γ_m as $[\![\mathbf{g}]\!]_{\Gamma_m} := \gamma^i_{\Gamma_m}(g_i) - \gamma^j_{\Gamma_m}(g_j)$, $\forall m \in \mathcal{M}$ and $(i,j) = \mathcal{I}_m$. Then $[\![\mathbf{g}]\!]_{\mathcal{M}_i}$ is the vector of jumps of $\mathbf{g}$ across the traces in $\mathcal{M}_i$, ordered according to trace index: $[\![\mathbf{g}]\!]_{\mathcal{M}_i} := \left([\![\mathbf{g}]\!]_{\Gamma_{\mathcal{M}_i(1)}}, \ldots, [\![\mathbf{g}]\!]_{\Gamma_{\mathcal{M}_i(M_i)}} \right)$. Similarly, given a function $g_i \in V_i$, $\left[\!\left[\frac{\partial g_i}{\partial \hat{n}^i_m} \right]\!\right]_{\Gamma_m}$ is the jump of the co-normal derivative across Γ_m on F_i, and we define the tuple

$$\left[\!\left[\frac{\partial g_i}{\partial \hat{n}_{\mathcal{M}_i}} \right]\!\right]_{\mathcal{M}_i} := \left(\left[\!\left[\frac{\partial g_i}{\partial \hat{n}^i_{\mathcal{M}_i(1)}} \right]\!\right]_{\Gamma_{\mathcal{M}_i(1)}}, \ldots, \left[\!\left[\frac{\partial g_i}{\partial \hat{n}^i_{\mathcal{M}_i(M_i)}} \right]\!\right]_{\Gamma_{\mathcal{M}_i(M_i)}} \right).$$

A.3 Problem formulation

Let us denote the unknown hydraulic head in Ω as $\boldsymbol{h} = (h_1, \ldots, h_N) \in V$, where $h_i \in V_i$, for $i = 1, \ldots, N$ is the hydraulic head on F_i. Then, in a simplified setting, using homogeneous Dirichlet boundary conditions, the DFN problem can be stated as: *find* $\boldsymbol{h} \in V$ *such that,* $\forall i \in \mathcal{I}$

$$(\nabla h_i, \nabla v)_{F_i} = (f_i, v)_{F_i} + \left\langle \left[\!\left[\frac{\partial h_i}{\partial \hat{n}_{\mathcal{M}_i}} \right]\!\right]_{\mathcal{M}_i}, \gamma_{\mathcal{M}_i}(v) \right\rangle_{\mathcal{M}_i} \quad \forall v \in V_i, \tag{A.1}$$

where $f_i \in \mathrm{L}^2(F_i)$ $\forall i \in \mathcal{I}$ is a function representing source terms on the fracture. At fracture intersections additional matching conditions are added, enforcing continuity of the hydraulic head and conservation of fluxes: $\forall m \in \mathcal{M}$, $\mathcal{I}_m = (i,j)$

$$\gamma^i_{\Gamma_m}(h_i) - \gamma^j_{\Gamma_m}(h_j) = 0 \tag{A.2}$$

$$\left[\!\left[\frac{\partial h_i}{\partial \hat{n}^i_m} \right]\!\right]_{\Gamma_m} + \left[\!\left[\frac{\partial h_j}{\partial \hat{n}^j_m} \right]\!\right]_{\Gamma_m} = 0. \tag{A.3}$$

A.3.1 Formulation as an optimization problem. We now aim at a different formulation of the above problem as an optimization problem of a suitable functional. First

of all, we define the fluxes (see Figure A.1b)

$$\forall i \in \mathcal{I}, \forall m \in \mathcal{M}_i,\ u_i^m := \left[\!\left[\frac{\partial h_i}{\partial \hat{n}_m^i} \right]\!\right]_{\Gamma_m} \in \mathrm{H}^{-1/2}(\Gamma_m)$$
$$\forall i \in \mathcal{I},\ \boldsymbol{u}_i := \left(u_i^{\mathcal{M}_i(1)}, \dots, u_i^{\mathcal{M}_i(M_i)}\right) \in \prod_{m \in \mathcal{M}_i} \mathrm{H}^{-1/2}(\Gamma_m) := U_i$$
$$\boldsymbol{u} := (\boldsymbol{u}_1, \dots, \boldsymbol{u}_N) \in \prod_{i \in \mathcal{I}} U_i := U.$$

In general, an element $\mathbf{w}$ of U is a $2(\#\mathcal{M})$-tuple of functions each belonging to $\mathrm{H}^{-\frac{1}{2}}(\Gamma_m)$, for some $m \in \mathcal{M}$. For all $\mathbf{w} \in U$, we indicate by $\mathbf{w}_i$ the M_i-tuple of functions in $\mathbf{w}$ which are defined on the traces lying on fracture F_i. The component of $\mathbf{w}$ related to the trace Γ_m and the fracture F_i is denoted by $w_i^m \in \mathrm{H}^{-\frac{1}{2}}(\Gamma_m)$. Moreover, for any $\mathbf{w} \in U$ we set $\{\{\mathbf{w}\}\}_{\Gamma_m} = w_i^m + w_j^m$, $\forall m \in \mathcal{M}$ with $\mathcal{I}_m = (i,j)$ and indicate by $\{\{\mathbf{w}\}\}_{\mathcal{M}_i}$ the vector whose k-th component is $\{\{\mathbf{w}\}\}_{\Gamma_{\mathcal{M}_i(k)}}$. Let us define by U_i^* and U^* the dual spaces of U_i and U, respectively.

Let us define the operator $\mathcal{H} : U \to V$, which associates to each vector $\mathbf{w} \in U$ a vector $\mathcal{H}(\mathbf{w}) = (h_1^w, \dots, h_N^w)$ of solutions to a Darcy's problem on each fracture independently, that is

$$(\nabla h_i^w, \nabla v_i)_{F_i} = (f_i, v_i)_{F_i} + \left\langle \mathbf{w}_i, \gamma_{\mathcal{M}_i}(v_i) \right\rangle_{\mathcal{M}_i} \quad \forall i \in \mathcal{I},\ \forall v_i \in V_i. \tag{A.4}$$

Moreover, for each $m \in \mathcal{M}$, we define the constrained functional $J_m : U \to \mathbb{R}$ such that

$$J_m(\mathbf{w}) = \left\| [\![\boldsymbol{h}^w]\!]_{\Gamma_m} \right\|^2_{\mathrm{H}^{\frac{1}{2}}(\Gamma_m)} + \left\| \{\{\mathbf{w}\}\}_{\Gamma_m} \right\|^2_{\mathrm{H}^{-\frac{1}{2}}(\Gamma_m)} \quad \text{where } \boldsymbol{h}^w = \mathcal{H}(\mathbf{w}). \tag{A.5}$$

The first term of the functional $J_m(\mathbf{w})$ represents the jump of the hydraulic head on the two fractures sharing the trace Γ_m, we call this functional "constrained" because we assume that these hydraulic heads satisfy equations (A.4) on the two fractures sharing Γ_m. The second term of the functional represents the flux conservation at the trace Γ_m.

We can define the global (constrained) functional $J : U \to \mathbb{R}$ such that $J(\mathbf{w}) = \sum_{m \in \mathcal{M}} J_m(\mathbf{w})$. We formulate the problem (A.1)-(A.3) as a constrained optimization problem:

$$\text{find } \boldsymbol{u} \in U \text{ such that } \ \boldsymbol{u} = \arg\min_{\mathbf{w} \in U} J(\mathbf{w}). \tag{A.6}$$

The functional $J(\mathbf{w})$ is positive for all $\mathbf{w} \in U \setminus \{\boldsymbol{u}\}$ and $J(\boldsymbol{u}) = 0$.

A.3.2 Equivalence with an elliptic differential problem.

We recall here the equivalence between problem (A.6) and a system of partial differential equations involving $\boldsymbol{h}$, $\boldsymbol{u}$ and an auxiliary pressure $\boldsymbol{p} \in V$ (see [6, Proposition 2.4]).

Proposition A.1. *The unique minimum of the functional $J(\mathbf{w})$ corresponds to the first order stationary conditions $\forall i \in \mathcal{I}$:*

$$\left\langle \{\{\boldsymbol{u}\}\}_{\mathcal{M}_i}, \mu_i \right\rangle_{\mathcal{M}_i} = -\left\langle \gamma_{\mathcal{M}_i}(p_i), \mu_i \right\rangle_{\mathcal{M}_i} \qquad \forall \mu_i \in U_i^*, \tag{A.7}$$
$$(\nabla p_i, \nabla q_i)_{F_i} = \left\langle [\![\boldsymbol{h}]\!]_{\mathcal{M}_i}, \gamma_{\mathcal{M}_i}(q_i) \right\rangle_{\mathcal{M}_i} \qquad \forall q_i \in V_i, \tag{A.8}$$
$$(\nabla h_i, \nabla v_i)_{F_i} = (f_i, v_i)_{F_i} + \left\langle \boldsymbol{u}_i, \gamma_{\mathcal{M}_i}(v_i) \right\rangle_{\mathcal{M}_i} \qquad \forall v_i \in V_i. \tag{A.9}$$

Remark A.1. From (A.2)-(A.3), we see that $[\![\boldsymbol{h}]\!]_{\mathcal{M}}$ is the null vector as well as $\{\{\boldsymbol{u}\}\}_{\mathcal{M}}$. Therefore, the exact solution of (A.8) corresponds to $p_i \equiv 0\ \forall i \in \mathcal{I}$.

We now want to find a suitable elliptic operator that describes our problem. We define the functional spaces $L := V \times V \times U$ and $L^* := V \times V \times U^*$ whose norms are:

$$\|(\boldsymbol{h}, \boldsymbol{p}, \boldsymbol{u})\|_L := \left[\sum_{i\in\mathcal{I}} \left(\|h_i\|^2_{F_i} + \|p_i\|^2_{F_i} + \sum_{m\in\mathcal{M}_i} \|u_i^m\|^2_{\mathrm{H}^{-\frac{1}{2}}(\Gamma_m)} \right)\right]^{\frac{1}{2}}, \tag{A.10}$$

$$\|(\mathbf{v}, \mathbf{q}, \boldsymbol{\mu})\|_{L^*} := \left[\sum_{i\in\mathcal{I}} \left(\|v_i\|^2_{F_i} + \|q_i\|^2_{F_i} + \sum_{m\in\mathcal{M}_i} \|\mu_i^m\|^2_{\mathrm{H}^{\frac{1}{2}}(\Gamma_m)} \right)\right]^{\frac{1}{2}}. \tag{A.11}$$

We can define the bilinear continuous operator $\mathscr{L}\colon\ L \times L^* \to \mathbb{R}$ such that

$$\begin{aligned}\mathscr{L}((\boldsymbol{h}, \boldsymbol{p}, \boldsymbol{u}), (\mathbf{v}, \mathbf{q}, \boldsymbol{\mu})) := \sum_{i\in\mathcal{I}} \Big\{ &(\nabla h_i, \nabla v_i)_{F_i} - \left\langle \boldsymbol{u}_i, \gamma_{\mathcal{M}_i}(v_i)\right\rangle_{\mathcal{M}_i} \\ &+ (\nabla p_i, \nabla q_i)_{F_i} - \left\langle [\![\boldsymbol{h}]\!]_{\mathcal{M}_i}, \gamma_{\mathcal{M}_i}(q_i)\right\rangle_{\mathcal{M}_i} + \left\langle \gamma_{\mathcal{M}_i}(p_i) + \{\{\boldsymbol{u}\}\}_{\mathcal{M}_i}, \boldsymbol{\mu}_i\right\rangle_{\mathcal{M}_i} \Big\}. \end{aligned} \tag{A.12}$$

Using this definition, the system of equations of Proposition A.1 can be written in compact form as

$$\forall\, (\mathbf{v},\, \mathbf{q},\, \boldsymbol{\mu}) \in L^*, \quad \mathscr{L}((\boldsymbol{h},\, \boldsymbol{p},\, \boldsymbol{u}), (\mathbf{v},\, \mathbf{q},\, \boldsymbol{\mu})) = \sum_{i\in\mathcal{I}} (f_i, v_i)_{F_i}\,. \tag{A.13}$$

This problem has a unique solution being equivalent to (A.1)–(A.3), thus applying Nečas Theorem (see, for example, [13, Theorem 3.3]), we can say that $\mathscr{L}$ satisfies an inf-sup condition:

$$\exists \beta > 0 : \|(\boldsymbol{h}, \boldsymbol{p}, \boldsymbol{u})\|_L \le \beta \sup_{(\mathbf{v},\mathbf{q},\boldsymbol{\mu})\in L^*} \frac{\mathscr{L}((\boldsymbol{h}, \boldsymbol{p}, \boldsymbol{u}), (\mathbf{v}, \mathbf{q}, \boldsymbol{\mu}))}{\|(\mathbf{v}, \mathbf{q}, \boldsymbol{\mu})\|_{L^*}}. \tag{A.14}$$

A.3.3 Problem discretization.

An important advantage of the formulation introduced in the previous sections is that the discretization of each fracture has not to be conforming to the traces, i.e. triangles can freely cross traces. In the following, we assume that each fracture is meshed by a good quality triangulation $\mathcal{T}_{\delta,i}$ [10]. Let $\mathcal{T}_\delta = \bigcup_{i\in\mathcal{I}} \mathcal{T}_{\delta,i}$ be the set of all the triangles on the DFN. Let $\mathcal{V}_\delta$ be the set of the vertices of the triangles in $\mathcal{T}_\delta$, $\mathcal{E}_\delta$ the set of all the edges of the triangles in $\mathcal{T}_\delta$. $\mathcal{V}_{\delta,i}$ and $\mathcal{E}_{\delta,i}$ coherently are the subsets of $\mathcal{V}_\delta$ and $\mathcal{E}_\delta$ containing the objects defined on fracture F_i. For all $T \in \mathcal{T}_\delta$, we indicate by $\mathring{T}$ the interior of T, by $\mathcal{M}_T$ the set of indices of those traces having non empty intersection with $\mathring{T}$, and by ℓ_T^m the segment $\Gamma_m \cap \mathring{T}$, for all $m \in \mathcal{M}_T$. Coherently, for any given $\sigma \in \mathcal{E}_\delta$ we indicate by $\mathcal{M}_\sigma$ the set of those $m \in \mathcal{M}$ such that $|\Gamma_m \cap \sigma| \neq \emptyset$. Moreover, on each trace Γ_m shared by two fractures F_i and F_j, we fix two discretizations $\Lambda_{m,i}$ and $\Lambda_{m,j}$ defined on the two fractures respectively. In the following, the symbol $h_\sharp$ denotes the diameter of an arbitrary geometrical object $\sharp$.

To solve the minimization problem in (A.6) we start by discretizing (A.13). Let us define the following finite dimensional subspaces: $V_{\delta,i} \subset V_i$, $\forall i \in \mathcal{I}$, $U_{\delta i}^m \subset \mathrm{L}^2(\Gamma_m) \subset \mathrm{H}^{-\frac{1}{2}}(\Gamma_m) = U_i^m \quad \forall i \in \mathcal{I}, m \in \mathcal{M}_i$, and let us set $U_{\delta i} := \prod_{m\in\mathcal{M}_i} U_{\delta i}^m$, $\forall i \in \mathcal{I}$, $U_\delta := \prod_{i\in\mathcal{I}} U_{\delta i}$, $V_\delta := \prod_{i\in\mathcal{I}} V_{\delta i}$. Our discrete problem is to find $\boldsymbol{h}_\delta, \boldsymbol{p}_\delta \in V_\delta$ and $\boldsymbol{u}_\delta \in U_\delta$ such that

$$\mathscr{L}((\boldsymbol{h}_\delta,\, \boldsymbol{p}_\delta,\, \boldsymbol{u}_\delta), (\mathbf{v}_\delta,\, \mathbf{q}_\delta,\, \boldsymbol{\mu}_\delta)) = \sum_{i\in\mathcal{I}} (f_i, v_{\delta i})_{F_i} \quad \forall \mathbf{v}_\delta,\, \mathbf{q}_\delta \in V_\delta,\, \boldsymbol{\mu}_\delta \in U_\delta\,, \tag{A.15}$$

that is to solve the following system of equations $\forall i \in \mathcal{I}$:

$$\left(\{\{\boldsymbol{u}_\delta\}\}_{\mathcal{M}_i}, \mu_{\delta i}\right)_{\mathcal{M}_i} = -\left(\gamma_{\mathcal{M}_i}(p_{\delta i}), \mu_{\delta i}\right)_{\mathcal{M}_i} \qquad \forall \mu_{\delta i} \in U_{\delta i}\,, \tag{A.16}$$

$$\left(\nabla p_{\delta i}, \nabla v_{\delta i}\right)_{F_i} = \left(\left[\!\left[\boldsymbol{h}_\delta\right]\!\right]_{\mathcal{M}_i}, \gamma_{\mathcal{M}_i}(v_{\delta i})\right)_{\mathcal{M}_i} \qquad \forall v_{\delta i} \in V_{\delta i}\,, \tag{A.17}$$

$$\left(\nabla h_{\delta i}, \nabla v_{\delta i}\right)_{F_i} = \left(f_i, v_{\delta i}\right)_{F_i} + \left(\boldsymbol{u}_{\delta i}, \gamma_{\mathcal{M}_i}(v_{\delta i})\right)_{\mathcal{M}_i} \qquad \forall v_{\delta i} \in V_{\delta i}\,. \tag{A.18}$$

This is equivalent (see [6]) to minimize a functional with the same structure of J but involving $\mathrm{L}^2(\Gamma_m)$ norms of the discrete functions $\boldsymbol{h}_\delta$ and $\boldsymbol{u}_\delta$. Indeed, if we define $\mathcal{H}_\delta : U_\delta \to V_\delta$ such that $(h_{\delta 1}, \ldots, h_{\delta N}) = \mathcal{H}_\delta(\boldsymbol{u}_\delta)$ is the solution vector of

$$\left(\nabla h_{\delta i}, \nabla v_{\delta i}\right)_{F_i} = \left(f_i, v_{\delta i}\right)_{F_i} + \left(\boldsymbol{u}_{\delta i}, \gamma_{\mathcal{M}_i}(v_{\delta i})\right)_{\mathcal{M}_i} \quad \forall v_{\delta i} \in V_{\delta i},\, \forall i \in \mathcal{I}\,, \tag{A.19}$$

then we can define, for any given $\mathbf{w}_\delta \in U_\delta$ and any $m \in \mathcal{M}$ the functional $J_{m\delta}$ such that

$$J_{m\delta}(\mathbf{w}_\delta) = \left\|\left[\!\left[\boldsymbol{h}_\delta^w\right]\!\right]_{\Gamma_m}\right\|^2_{\Gamma_m} + \left\|\{\{\mathbf{w}_\delta\}\}_{\Gamma_m}\right\|^2_{\Gamma_m} \quad \text{with } \boldsymbol{h}_\delta^w = \mathcal{H}_\delta(\mathbf{w}_\delta)\ . \tag{A.20}$$

The system (A.16)–(A.18) is equivalent to the following minimum problem:

$$\boldsymbol{u}_\delta = \operatorname*{arg\,min}_{\mathbf{w}_\delta \in U_\delta} J_\delta(\mathbf{w}_\delta) = \operatorname*{arg\,min}_{\mathbf{w}_\delta \in U_\delta} \sum_{m \in \mathcal{M}} J_{m\delta}(\mathbf{w}_\delta)\,. \tag{A.21}$$

A.4 Error and error estimators

The following quantities define the error performed approximating (A.13) by (A.15):

$$e_{\boldsymbol{h}} = \boldsymbol{h} - \boldsymbol{h}_\delta, \quad e_{\boldsymbol{p}} = \boldsymbol{p} - \boldsymbol{p}_\delta, \quad e_{\boldsymbol{u}} = \boldsymbol{u} - \boldsymbol{u}_\delta\,. \tag{A.22}$$

Since $(e_{\boldsymbol{h}}, e_{\boldsymbol{p}}, e_{\boldsymbol{u}}) \in L$, we have the following Lemma.

Lemma A.1. *Let $\mathscr{L}$ be defined by* (A.12) *and $e_{\boldsymbol{h}}$, $e_{\boldsymbol{p}}$, $e_{\boldsymbol{u}}$ by* (A.22)*. Then, for any $\mathbf{v}_\delta$, $\mathbf{q}_\delta \in V_\delta$ and $\mu_\delta \in U_\delta$, $\mathscr{L}((e_{\boldsymbol{h}}, e_{\boldsymbol{p}}, e_{\boldsymbol{u}}), (\mathbf{v}_\delta, \mathbf{q}_\delta, \mu_\delta)) = 0$.*

We define the *error measure* as

$$err := \left\|(e_{\boldsymbol{h}}, e_{\boldsymbol{p}}, e_{\boldsymbol{u}})\right\|_L\,. \tag{A.23}$$

The main result of next section is that the error measure (A.23) can be controlled by the following quantities $\forall i \in \mathcal{I}$:

Residual estimator:

$$\eta_{R,T} := h_T \left\|f_i + \Delta h_{\delta i}\right\|_T \quad \forall T \in \mathcal{T}_{\delta,i}\,. \tag{A.24}$$

Estimator for the approximation of the flux through edges: $\forall \sigma \in \mathcal{E}_{\delta,i}$,

$$\xi_{F,\sigma} \;:=\; (h_\sigma)^{\frac{1}{2}} \left\| \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right\|_\sigma \quad \text{where} \quad \tilde{u}_{\delta i,\sigma} \;:=\; \begin{cases} u^m_{\delta i} & \forall m \in \mathcal{M}_\sigma, \\ 0 & \text{elsewhere}\,. \end{cases} \tag{A.25}$$

Estimator for the nonconformity of the discretization:

$$\xi^m_{NC,T} := \left(h_{\ell^m_T}\right)^{\frac{1}{2}} \left\|u^m_{\delta i}\right\|_{\ell^m_T} \quad \forall T \in \mathcal{T}_{\delta,i},\, m \in \mathcal{M}_T\,. \tag{A.26}$$

Local estimator for the pressure induced by discontinuity:

$$\eta_{P,T} := \|\!|p_{\delta i}|\!\|_T \quad \forall T \in \mathcal{T}_{\delta,i}. \tag{A.27}$$

Local estimator for the pressure induced by the unbalancing of fluxes:

$$\xi^i_{P,\lambda} := h_\lambda^{\frac{1}{2}} \left\|\gamma^i_{\Gamma_m}(p_{\delta i})\right\|_\lambda \quad \forall m \in \mathcal{M}_i,\ \lambda \in \Lambda_{m,i}. \tag{A.28}$$

Local estimator of the minimization error:

$$J_{\delta,\lambda}(\boldsymbol{u}_\delta) := h_\lambda^{\frac{1}{2}} \left(\left\|\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}\right\|_\lambda + \left\|[\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}\right\|_\lambda \right) \quad \forall m \in \mathcal{M}_i,\ \lambda \in \Lambda_{m,i}. \tag{A.29}$$

A.5 Reliability

This section is devoted to obtain an a posteriori upper bound for the error norm (A.23) based on condition (A.14). After stating some auxiliary results (Appendix A.5.1), we obtain (Theorem A.2) our estimate.

A.5.1 Auxiliary results. In the following, we apply the well know properties of the Clement's pseudo-interpolation operator on the fracture F_i, $i \in \mathcal{I}$, denoted by $\Pi_{\delta i}$. Given a fracture F_i, $i \in \mathcal{I}$, let us consider a triangle $T \in \mathcal{T}_{\delta,i}$ and an edge $\sigma \in \mathcal{E}_{\delta,i}$. Then, for any $v \in \mathrm{H}^1_0(F_i)$,

$$\|v - \Pi_{\delta i}(v)\|_T \lesssim h_T \|\!|v|\!\|_{\omega_T}\ , \tag{A.30}$$

$$\|\!|v - \Pi_{\delta i}(v)|\!\|_T \lesssim \|\!|v|\!\|_{\omega_T}\ , \tag{A.31}$$

$$\left\|\gamma^i_\sigma(v - \Pi_{\delta i}(v))\right\|_\sigma \lesssim (h_\sigma)^{\frac{1}{2}} \|\!|v|\!\|_{\omega_\sigma}\ , \tag{A.32}$$

where ω_T is the union of all triangles having a side or a vertex in common with T and ω_σ is the union of the two triangles having σ in common.

Concerning trace spaces, given a bounded open set $\Omega \subset \mathbb{R}^2$, a segment $\lambda \subseteq \partial\Omega$ and a function $g \in \mathrm{H}^{\frac{1}{2}}(\lambda)$, one can define the set $\mathrm{H}^1_{g,\lambda}(\Omega) := \left\{v \in \mathrm{H}^1(\Omega) :\ \gamma_\lambda(v) = g\right\} \subseteq \mathrm{H}^1(\Omega)$ and the seminorm

$$|g|_{\frac{1}{2},\lambda} := \inf_{v \in \mathrm{H}^1_{g,\lambda}} \|\nabla v\|_\Omega\ .$$

The following Lemma A.2 defines the function of minimum norm.

Lemma A.2. *Let Ω be a bounded open set, $\lambda \subseteq \partial\Omega$. Let $g \in \mathrm{H}^{\frac{1}{2}}(\lambda)$. Then*

$$\exists!\, u \in \mathrm{H}^1(\Omega) \quad \textit{tale che} \quad |g|_{\frac{1}{2},\lambda} = |u|_{1,\Omega}$$

and u is the unique solution of problem

$$\textit{Find } u \in \mathrm{H}^1_{g,\lambda}(\Omega) \textit{ such that}$$

$$a(u,v) = 0 \quad \forall v \in \mathrm{H}^1_{0,\lambda}(\Omega) \tag{A.33}$$

where

$$a:\ \mathrm{H}^1(\Omega) \times \mathrm{H}^1(\Omega) \to \mathbb{R} \quad \textit{such that} \quad a(w,v) = \int_\Omega \nabla w \cdot \nabla v$$

Proof. Let $u \in \mathrm{H}^1_{g,\lambda}(\Omega)$ be the solution of (A.33), $R_g \in \mathrm{H}^1_{g,\lambda}(\Omega)$ such that $R_g \neq u$. We need to show that $|u|_{1,\Omega} < |R_g|_{1,\Omega}$. If we introduce $u_0 = u - R_g \in \mathrm{H}^1_{0,\lambda}(\Omega)$, we have that $a(u, u_0) = 0$. Since the seminorm $\mathrm{H}^1(\Omega)$ is generated by a,

$$\begin{aligned} |R_g|_{1,\Omega} &= a(R_g, R_g) = a(u - u_0, u - u_0) = \\ &= a(u,u) - 2\cancel{a(u, u_0)} + a(u_0, u_0) = \\ &= |u|_{1,\Omega} + |u_0|_{1,\Omega} > |u|_{1,\Omega} \ , \end{aligned}$$

being $|u_0|_{1,\Omega} > 0$ because $u_0 \in \mathrm{H}^1_{0,\lambda}(\Omega)$ is the unique solution of

$$a(u_0, v) = a(R_g, v) \quad \forall v \in \mathrm{H}^1_{0,\lambda} \, .$$

□

The existence of the function u of Lemma A.2 is exploited in the proof of the following important result.

Theorem A.1. *Let λ be a segment of length h_λ and $P : \mathrm{H}^{\frac{1}{2}}(\lambda) \to \mathrm{L}^2(\lambda)$ a continuous linear operator preserving a.e. constant functions. Then,*

$$\exists C > 0 : \ \forall g \in \mathrm{H}^{\frac{1}{2}}(\lambda) \, , \ \|g - Pg\|_{0,\lambda} \leq C h_\lambda^{\frac{1}{2}} \, |g|_{\frac{1}{2},\lambda} \, . \tag{A.34}$$

Proof. The hypothesis that P is the identity for constants is necessary since

$$g \in P^0(\lambda) \Rightarrow |g|_{\frac{1}{2},\lambda} = 0$$

which invalidates the thesis. To start we consider a segment $\hat{\lambda}$ of length 1. By contraddiction, suppose

$$\forall C > 0, \, \exists \hat{g} \in \mathrm{H}^{\frac{1}{2}}\left(\hat{\lambda}\right) : \ \|\hat{g} - P\hat{g}\| \, 0, \hat{\lambda} > C \, |\hat{g}|_{\frac{1}{2},\hat{\lambda}}$$

Then, $\forall N \in \mathbb{N}$ one can find $\hat{g}_N \in \mathrm{H}^{\frac{1}{2}}\left(\hat{\lambda}\right)$ such that,

$$\|\hat{g}_N - P\hat{g}_N\|_{0,\hat{\lambda}} > N \, |\hat{g}_N|_{\frac{1}{2},\hat{\lambda}} \tag{A.35}$$

and it is clearly always possible to choose $\hat{g}_N$ such that

$$|\hat{g}_N|_{\frac{1}{2},\hat{\lambda}} = 1 \tag{A.36}$$

Since $\mathrm{H}^{\frac{1}{2}}\left(\hat{\lambda}\right)$ is compact in $\mathrm{L}^2\left(\hat{\lambda}\right)$, there exists a subsequence $\{\hat{g}_{N_k}\}$ that converges to some element $\hat{g}_\star \in \mathrm{L}^2\left(\hat{\lambda}\right)$. Then, by continuity of P,

$$\hat{g}_{N_k} - P\hat{g}_{N_k} \to \hat{g}_\star - P\hat{g}_\star \, .$$

From (A.35),

$$\begin{cases} \forall N_k, \ |\hat{g}_{N_k}|_{\frac{1}{2},\hat{\lambda}} < \dfrac{\left\|\hat{g}_{N_k} - P\hat{g}_{N_k}\right\|_{0,\hat{\lambda}}}{N_k} \\ \lim_{N_k \to \infty} \dfrac{\left\|\hat{g}_{N_k} - P\hat{g}_{N_k}\right\|_{0,\hat{\lambda}}}{N_k} = 0 \end{cases} \Rightarrow |\hat{g}_{N_k}|_{\frac{1}{2},\hat{\lambda}} \xrightarrow{N_k \to \infty} 0$$

which contradicts the choice (A.36). Thus, the following inequality must hold:

$$\forall \hat{g} \in \mathrm{H}^{\frac{1}{2}}\left(\hat{\lambda}\right),\, \exists C > 0 :\, \|\hat{g} - P\hat{g}\|\, 0, \hat{\lambda} \leq C\, |\hat{g}|_{\frac{1}{2},\hat{\lambda}} \tag{A.37}$$

Now, consider segment λ and turn our attention to the first member of (A.34). Let $g \in \mathrm{H}^{\frac{1}{2}}(\lambda)$ and $\hat{g}$ be its mapping on $\hat{\lambda}$, that is $\hat{g}(\hat{x}) = g(x/h_\lambda)$, $\hat{x} \in (0,1)$ and $x \in (0, h_\lambda)$. Then,

$$\begin{aligned}\|g - Pg\|^2_{0,\lambda} &= \int_\lambda (g(x) - Pg(x))^2\, \mathrm{d}x = h_\lambda \int_{\hat{\lambda}} (\hat{g}(\hat{x}) - P\hat{g}(\hat{x}))^2\, \mathrm{d}\hat{x} = \\ &= h_\lambda\, \|\hat{g} - P\hat{g}\|^2_{0,\hat{\lambda}} \Rightarrow \exists C > 0 :\, \|g - Pg\|_{0,\lambda} \leq C h_\lambda^{\frac{1}{2}}\, |\hat{g}|_{\frac{1}{2},\hat{\lambda}}\end{aligned}$$

Now, consider a square Ω having λ as one of its sides, mapped on $\hat{\Omega}$, having $\hat{\lambda}$ as one of its sides, by the transformation $(\hat{x}, \hat{y}) = (x/h_\lambda,\, y/h_\lambda)$. Then, if we define $\hat{\nabla} = \left(\frac{\partial}{\partial \hat{x}}, \frac{\partial}{\partial \hat{y}}\right)$, $v_g \in \mathrm{H}^1(\Omega)$ such that $\gamma(v)\,\lambda = g$ and $\|v_g\|_{1,\Omega} = |g|_{\frac{1}{2},\lambda}$ (whose existence is guaranteed by Lemma A.2) and $\hat{v}_g$ its mapping on $\hat{\Omega}$ with just stated transformation. Then it is easy to check that $\hat{v}$ is the solution of (A.33). Thus we have

$$\begin{aligned}|\hat{g}|^2_{\frac{1}{2},\hat{\lambda}} = |\hat{v}_g|^2_{\frac{1}{2},\hat{\lambda}} &= \int_{\hat{\Omega}} \left(\hat{\nabla}\hat{v}_g\right)^2 \mathrm{d}\hat{x}\mathrm{d}\hat{y} = \int_{\hat{\Omega}} \left[\left(\frac{\partial \hat{v}_g}{\partial \hat{x}}\right)^2 + \left(\frac{\partial \hat{v}_g}{\partial \hat{y}}\right)^2\right] \mathrm{d}\hat{x}\mathrm{d}\hat{y} = \\ &= \frac{1}{\cancel{h_\lambda^2}} \int_\Omega \left[\cancel{h_\lambda^2}\left(\frac{\partial v_g}{\partial x}\right)^2 + \cancel{h_\lambda^2}\left(\frac{\partial v_g}{\partial y}\right)^2\right] = |v_g|^2_{1,\Omega} = |g|^2_{\frac{1}{2},\lambda}\end{aligned}$$

and this concludes the proof. □

A.5.2 Upper bound.

In this section we derive an upper bound for the error.

Theorem A.2. *Let $e_{\boldsymbol{h}}$, $e_{\boldsymbol{p}}$, $e_{\boldsymbol{u}}$ be defined by* (A.22) *and let all the quantities defined in* (A.24)–(A.29) *be given. Then,*

$$\begin{aligned}err \lesssim \sum_{i\in\mathcal{I}} \Bigg[\sum_{T\in\mathcal{T}_{\delta,i}} \left(\eta_{R,T} + \eta_{P,T} + \sum_{m\in\mathcal{M}_T} \xi^m_{NC,T}\right) + \sum_{\sigma\in\mathcal{E}_{\delta,i}} \xi^m_{F,\sigma} \\ + \sum_{m\in\mathcal{M}_i} \sum_{\lambda\in\Lambda_{m,i}} \left(\xi^i_{P,\lambda} + J_{\delta,\lambda}(\boldsymbol{u}_\delta)\right) \Bigg] .\end{aligned}$$

Proof. From (A.14) we have

$$err = \|(e_{\boldsymbol{h}}, e_{\mathbf{p}}, e_{\boldsymbol{u}})\|_L \lesssim \sup_{(\mathbf{v},\mathbf{q},\boldsymbol{\mu})\in L^*} \frac{\mathscr{L}((e_{\boldsymbol{h}}, e_{\mathbf{p}}, e_{\boldsymbol{u}}), (\mathbf{v}, \mathbf{q}, \boldsymbol{\mu}))}{\|(\mathbf{v}, \mathbf{q}, \boldsymbol{\mu})\|_{L^*}} .$$

From Lemma A.1 we know that, for any given $\mathbf{v}_\delta$, $\mathbf{q}_\delta \in V_\delta$ and $\boldsymbol{\mu}_\delta \in U_\delta$,

$$\begin{aligned}\mathscr{L}((e_{\boldsymbol{h}}, e_{\boldsymbol{p}}, e_{\boldsymbol{u}}), (\mathbf{v}, \mathbf{q}, \boldsymbol{\mu})) &= \mathscr{L}((e_{\boldsymbol{h}}, e_{\mathbf{p}}, e_{\boldsymbol{u}}), (\mathbf{v} - \mathbf{v}_\delta, \mathbf{q} - \mathbf{q}_\delta, \boldsymbol{\mu} - \boldsymbol{\mu}_\delta)) = \\ &= \sum_{i\in\mathcal{I}} \Big\{ (\nabla(h_i - h_{\delta i}), \nabla(v_i - v_{\delta i}))_{F_i} - \left\langle \boldsymbol{u} - \boldsymbol{u}_\delta, \gamma_{\mathcal{M}_i}(v_i - v_{\delta i})\right\rangle_{\mathcal{M}_i} \\ &\quad + (\nabla(p_i - p_{\delta i}), \nabla(q_i - q_{\delta i}))_{F_i} - \left\langle [\![\boldsymbol{h} - \boldsymbol{h}_\delta]\!]_{\mathcal{M}_i}, \gamma_{\mathcal{M}_i}(q_i - q_{\delta i})\right\rangle_{\mathcal{M}_i} \\ &\quad + \left\langle \gamma_{\mathcal{M}_i}(p_i - p_{\delta i}) + \{\{\boldsymbol{u} - \boldsymbol{u}_\delta\}\}_{\mathcal{M}_i}, \boldsymbol{\mu}_i - \boldsymbol{\mu}_{\delta i}\right\rangle_{\mathcal{M}_i} \Big\} .\end{aligned}$$

We now proceed by estimating separately the terms involving different test functions. Let $i \in \mathcal{I}$ be fixed:

TERMS WITH TEST FUNCTION $v_i - v_{\delta i}$. Since $\boldsymbol{h} = \mathcal{H}(\boldsymbol{u})$ (thus h_i and $\boldsymbol{u}_i$ are linked by (A.9)) and, by Green's formula applied on $\mathcal{T}_{\delta,i}$,

$$\begin{aligned}(\nabla(h_i - h_{\delta i}), \nabla(v_i - v_{\delta i}))_{F_i} - \left\langle (\boldsymbol{u}_i - \boldsymbol{u}_{\delta i}), \gamma_{\mathcal{M}_i}(v_i - v_{\delta i}) \right\rangle_{\mathcal{M}_i} = \\ = \sum_{T \in \mathcal{T}_{\delta,i}} (f_i + \Delta h_{\delta i}, v_i - v_{\delta i})_T - \sum_{\sigma \in \mathcal{E}_{\delta,i}} \left\langle \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma, \gamma^i_\sigma(v_i - v_{\delta i}) \right\rangle_\sigma \\ + \sum_{m \in \mathcal{M}_i} \left\langle \boldsymbol{u}^m_{\delta i}, \gamma^i_{\Gamma_m}(v_i - v_{\delta i}) \right\rangle_{\Gamma_m} .\end{aligned}$$

Then, since $\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma \in \mathrm{L}^2(\sigma)$, $u^m_{\delta i} \in \mathrm{L}^2(\Gamma_m) \subset \mathrm{H}^{-\frac{1}{2}}(\Gamma_m)$, $\gamma^i_\sigma(v_i - v_{\delta i}) \in \mathrm{H}^{\frac{1}{2}}(\sigma) \subset \mathrm{L}^2(\sigma)$, $\gamma^i_{\Gamma_m}(v_i - v_{\delta i}) \in \mathrm{H}^{\frac{1}{2}}(\Gamma_m) \subset \mathrm{L}^2(\Gamma_m)$ it is possible to write the duality product on each trace as a scalar product in L^2:

$$\begin{aligned}- \sum_{\sigma \in \mathcal{E}_{\delta,i}} \left\langle \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma, \gamma^i_\sigma(v_i - v_{\delta i}) \right\rangle_\sigma + \sum_{m \in \mathcal{M}_i} \left\langle u^m_{\delta i}, \gamma^i_{\Gamma_m}(v_i - v_{\delta i}) \right\rangle_{\Gamma_m} = \\ = \sum_{\sigma \in \mathcal{E}_\delta} \left(\tilde{u}_{\delta i,\sigma} - \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma, \gamma^i_\sigma(v_i - v_{\delta i}) \right)_\sigma + \sum_{\substack{T \in \mathcal{T}_{\delta,i} \\ m \in \mathcal{M}_T}} \left(u^m_{\delta i}, \gamma^i_{\Gamma_m}(v_i - v_{\delta i}) \right)_{\ell^m_T} .\end{aligned}$$

Then, taking $v_{\delta i} = \Pi_{\delta i}(v_i)$ and using inequalities (A.30) and (A.32),

$$\begin{aligned}(\nabla(h_i - h_{\delta i}), \nabla(v_i - v_{\delta i}))_{F_i} - \left\langle \boldsymbol{u}_i - \boldsymbol{u}_{\delta i}, \gamma_{\mathcal{M}_i}(v_i - v_{\delta i}) \right\rangle_{\mathcal{M}_i} \leq \\ \leq \left\{ \sum_{T \in \mathcal{T}_{\delta,i}} \left(h_T \left\| f_i + \Delta h_{\delta i} \right\|_T + \sum_{m \in \mathcal{M}_T} h^{\frac{1}{2}}_{\ell^m_T} \left\| u^m_{\delta i} \right\|_{\ell^m_T} \right) \right. \\ \left. + \sum_{\sigma \in \mathcal{E}_\delta} h^{\frac{1}{2}}_\sigma \left\| \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right\|_\sigma \right\} \left\| v_i \right\|_{F_i} .\end{aligned}$$

TERMS WITH TEST FUNCTION $q_i - q_{\delta i}$. Using equation (A.8), Green's formula applied on $\mathcal{T}_{\delta,i}$ and the fact that $[\![\boldsymbol{h}]\!]_{\mathcal{M}_i} = \boldsymbol{0}$ and $p_i = 0$,

$$\begin{aligned}(\nabla(p_i - p_{\delta i}), \nabla(q_i - q_{\delta i}))_{F_i} - \left\langle [\![\boldsymbol{h} - \boldsymbol{h}_\delta]\!]_{\mathcal{M}_i}, \gamma_{\mathcal{M}_i}(q_i - q_{\delta i}) \right\rangle_{\mathcal{M}_i} = \\ = \sum_{T \in \mathcal{T}_{\delta,i}} (-\nabla p_{\delta i}, \nabla(q_i - q_{\delta i}))_T + \sum_{m \in \mathcal{M}_i} \left(- [\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}, \gamma^i_{\Gamma_m}(q_i - q_{\delta i}) \right)_{\Gamma_m} .\end{aligned}$$

For any given $m \in \mathcal{M}_i$, we introduce a discretization $\Lambda_{m,i}$ writing

$$\left(- [\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}, \gamma^i_{\Gamma_m}(q_i - q_{\delta i}) \right)_{\Gamma_m} \leq \sum_{\lambda \in \Lambda_{m,i}} \left\| [\![\boldsymbol{h}_\delta]\!]_{\Gamma_m} \right\|_\lambda \left\| \gamma^i_{\Gamma_m}(q_i - q_{\delta i}) \right\|_\lambda ,$$

then, choosing $q_{\delta i} = \Pi_{\delta i}(q_i)$ and using inequalities (A.30) and (A.32),

$$\begin{aligned}(\nabla(p_i - p_{\delta i}), \nabla(q_i - q_{\delta i}))_{F_i} - \left\langle [\![\boldsymbol{h} - \boldsymbol{h}_\delta]\!]_{\mathcal{M}_i}, \gamma_{\mathcal{M}_i}(q_i - q_{\delta i}) \right\rangle_{\mathcal{M}_i} \leq \\ \leq \left(\sum_{T \in \mathcal{T}_{\delta,i}} \left\| p_{\delta i} \right\|_T + \sum_{m \in \mathcal{M}_i} \sum_{\lambda \in \Lambda_{m,i}} h^{\frac{1}{2}}_\lambda \left\| [\![\boldsymbol{h}_\delta]\!]_{\Gamma_m} \right\|_\lambda \right) \left\| q_i \right\| .\end{aligned}$$

TERM WITH TEST FUNCTION $\boldsymbol{\mu}_i - \boldsymbol{\mu}_{\delta i}$. Using (A.7) and since $\gamma^i_{\Gamma_m}(p_{\delta i}), \mu^m_i \in \mathrm{H}^{\frac{1}{2}}(\Gamma_m) \subset \mathrm{L}^2(\Gamma_m)$, $u^m_{\delta i} \in \mathrm{L}^2(\Gamma_m) \subset \mathrm{H}^{-\frac{1}{2}}(\Gamma_m)$ we obtain, rewriting the duality product as a scalar product in $\mathrm{L}^2(\Gamma_m)$ and using discretization $\Lambda_{m,i}$,

$$\begin{aligned}\left\langle \gamma_{\mathcal{M}_i}(p_i - p_{\delta i}) + \{\{\boldsymbol{u}_i - \boldsymbol{u}_{\delta i}\}\}_{\mathcal{M}_i}, \boldsymbol{\mu}_i - \boldsymbol{\mu}_{\delta i} \right\rangle_{\mathcal{M}_i} = \\ = \sum_{m\in\mathcal{M}_i}\sum_{\lambda\in\Lambda_{m,i}} \left(-\gamma^i_{\Gamma_m}(p_{\delta i}), \mu^m_i - \mu^m_{\delta i}\right)_\lambda + \left(\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}, \mu^m_i - \mu^m_{\delta i}\right)_\lambda \lesssim \\ \lesssim \sum_{m\in\mathcal{M}_i}\sum_{\lambda\in\Lambda_{m,i}} h_\lambda^{\frac{1}{2}} \left(\left\|\gamma^i_{\Gamma_m}(p_{\delta i})\right\|_\lambda + \left\|\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}\right\|_\lambda \right) \|\mu^m_i\|_{\frac{1}{2},\Gamma_m} ,\end{aligned}$$

where last estimate is obtained by supposing $\mu^m_{\delta i}$ is the image of μ^m_i through a linear continuous operator that preserves constants, by applying Theorem A.1 and since

$$|\mu^m_i|_{\frac{1}{2},\Gamma_m} \le \|\mu^m_i\|_{\frac{1}{2},\Gamma_m} = \inf_{\substack{v\in\mathrm{H}^1(\omega_{\Gamma_m,i}) \\ \gamma^i_m(v)=\mu^m_i}} \|v\|_{1,\omega_{\Gamma_m,i}} ,$$

where $\omega_{\Gamma_m,i}$ is a subregion of F_i having Γ_m on its boundary.

The proof is concluded since $\forall i \in \mathcal{I}$ and $\forall m \in \mathcal{M}_i$

$$|||v_i|||_{F_i} \le \|(\mathbf{v},\mathbf{q},\boldsymbol{\mu})\|_{L^*} , \; |||q_i|||_{F_i} \le \|(\mathbf{v},\mathbf{q},\boldsymbol{\mu})\|_{L^*} , \; \|\mu^m_i\|_{\frac{1}{2},\Gamma_m} \le \|(\mathbf{v},\mathbf{q},\boldsymbol{\mu})\|_{L^*} .$$

□

For the sake of notational simplicity, we define the global estimator

$$est_\delta := \sum_{i\in\mathcal{I}} \left[\sum_{T\in\mathcal{T}_{\delta,i}} \left(\eta_{R,T} + \eta_{P,T} + \sum_{m\in\mathcal{M}_T} \xi^m_{NC,T} \right) + \sum_{\sigma\in\mathcal{E}_\delta} \xi^m_{F,\sigma} \right. \\ \left. + \sum_{m\in\mathcal{M}_i}\sum_{\lambda\in\Lambda_{m,i}} \left(\xi^i_{P,\lambda} + J_{\delta,\lambda}(\boldsymbol{u}_\delta)\right) \right] . \quad \text{(A.38)}$$

A.6 Efficiency of the a posteriori error estimate

In this section we prove the efficiency of the estimators presented in Theorem A.2, i.e. we show that for the a posteriori error estimator of Theorem A.2 we can write a lower bound in terms of a multiple of the error norm defined by (A.23).

From now on, $\forall i \in \mathcal{I}$ we assume that the discretization $\Lambda_{m,i}$ which was fixed in Appendix A.3.3 is the one induced on Γ_m by the triangulation $\mathcal{T}_{\delta,i}$, that is $\Lambda_{m,i} = \bigcup_{T\in\mathcal{T}_{\delta,i}} \overline{\ell^m_T}$. For any triangle $T \in \mathcal{T}_\delta$, the following non-conformity measure can be defined:

$$h_{NC,T} := \sum_{m\in\mathcal{M}_T} h_{\ell^m_T} . \quad \text{(A.39)}$$

Such quantity is zero for all triangles having empty intersection with all traces and is less or equal than $\#\mathcal{M}_T\, h_T$ for those intersecting some of them. It is not too restrictive to suppose $h_{NC,T} < 1$, assuming that the problem is written in non-dimensional way. The results in Appendices A.6.1 to A.6.3 together prove the following theorem.

Theorem A.3. *Let est_δ be defined by* (A.38)*, $e_{\boldsymbol{h}}$, $e_{\boldsymbol{u}}$ and $e_{\boldsymbol{p}}$ be defined by* (A.22) *and $h_{NC,T}$ be defined by* (A.39) *for all $T \in \mathcal{T}_\delta$. Then, if $h_{NC,T} < 1$ $\forall T \in \mathcal{T}_\delta$,*

$$est_\delta \lesssim |\!|\!| e_{\boldsymbol{p}} |\!|\!| + C_{NC} \left[\max_{\sigma \in \mathcal{E}_\delta} \{1, h_\sigma\} \left(|\!|\!| e_{\boldsymbol{h}} |\!|\!| + \|e_{\boldsymbol{u}}\|_U \right) + \max_{\substack{i \in \mathcal{I} \\ T \in \mathcal{T}_{\delta,i}}} h_T \|f_i - f_T\| \right],$$

where f_T is the mean of f_i on triangle $T \in \mathcal{T}_{\delta,i}$ and

$$C_{NC} := \max_{T \in \mathcal{T}_\delta} \left(\frac{1 + h_{NC,T}}{1 - h_{NC,T}} \right).$$

Remark A.2. We remark that the efficiency of the a posteriori estimate depends on the non-conformity of the triangulation through C_{NC}, which tends to 1 by refining the mesh. In a non-dimensional formulation of the problem, it is however always possible to ask that the coarsest considered triangulation satisfies $h_{NC,T} \leq \frac{1}{2}$ (thus having $C_{NC} \leq 3$).

A.6.1 Auxiliary results.

In the following we apply classical results about suitable cut-off functions [16], which exploit the properties of special polynomial functions with compact support, called *bubble functions*.

Given $i \in \mathcal{I}$ and any triangle $T \subset \mathcal{T}_{\delta,i}$, let us denote by $\mathfrak{b}_T$ the triangle bubble function, as defined in [16]. It has the following properties:

$$\operatorname{supp} \mathfrak{b}_T = T, \quad 0 \leq \mathfrak{b}_T \leq 1, \quad \max_{\mathbf{x} \in T} \mathfrak{b}_T(\mathbf{x}) = 1, \quad (\mathfrak{b}_T, 1)_T = \frac{9}{20} |T| ,$$

from which, since $\mathfrak{b}_T \in \mathrm{H}^1_0(T)$, the following estimates can be obtained (see [16, Lemma 1.3]):

$$\|\mathfrak{b}_T\|_T = [(\mathfrak{b}_T, \mathfrak{b}_T)_T]^{\frac{1}{2}} \leq [(\mathfrak{b}_T, 1)_T]^{\frac{1}{2}} \Rightarrow \|\mathfrak{b}_T\|_T \lesssim h_T , \tag{A.40}$$

$$\|\nabla \mathfrak{b}_T\|_T \lesssim h_T^{-1} \|\mathfrak{b}_T\|_T \Rightarrow \|\nabla \mathfrak{b}_T\|_T \lesssim 1 . \tag{A.41}$$

Let $l \subset T$ be a segment not necessarily intersecting ∂T, L its prolongation up to ∂T. Then, since $\gamma_L(\mathfrak{b}_T) \in \mathrm{H}^{\frac{1}{2}}_{00}(L)$ and $h_L \leq h_T$, applying the continuity of the trace operator on L, we have

$$\|\gamma_l(\mathfrak{b}_T)\|_{\frac{1}{2},l} \leq \|\gamma_L(\mathfrak{b}_T)\|_{\frac{1}{2},L} \lesssim \|\gamma_L(\mathfrak{b}_T)\|_{\mathrm{H}^{\frac{1}{2}}_{00}(L)} \lesssim \|\nabla \mathfrak{b}_T\|_T \lesssim 1 . \tag{A.42}$$

All the constants depend on the quality of the considered triangle, namely on the minimum of its angles. It is possible to prove the following useful result [14, Lemma 4.1].

Lemma A.3. *Let $i \in \mathcal{I}$, $T \in \mathcal{T}_{\delta,i}$, $\mathfrak{b}_T$ the bubble function on T. Let $\mathcal{P}(T) \subset \mathrm{H}^1(T)$ be a finite dimensional space. Then, for any given $v \in \mathcal{P}(T)$,*

$$\|v\|_T^2 \lesssim \left(\mathfrak{b}_T, v^2\right)_T , \qquad \|v\mathfrak{b}_T\|_T \leq \|v\|_T , \tag{A.43}$$

$$|\!|\!| \mathfrak{b}_T v |\!|\!|_T \lesssim h_T^{-1} \|v\|_T . \tag{A.44}$$

We now consider a side $\sigma \subset F_i$ shared by two triangles R and L belonging to a regular triangulation, that is such that $h_R \sim h_L \sim h_\sigma$. We denote by $\mathfrak{b}_\sigma$ the side bubble function of σ, as defined in [16]. It has the following properties:

$$\operatorname{supp} \mathfrak{b}_\sigma = \omega_\sigma, \quad 0 \leq \mathfrak{b}_\sigma \leq 1, \quad \max_{\mathbf{x} \in \omega_\sigma} \mathfrak{b}_\sigma = 1, \quad (\gamma_\sigma(\mathfrak{b}_\sigma), 1)_\sigma = \frac{2}{3} h_\sigma \sim h_\sigma ,$$

$$\forall T \in \{R, L\}, \quad (\mathfrak{b}_\sigma, 1)_T = \frac{1}{3} |T| ,$$

from which, since $\mathfrak{b}_\sigma \in \mathrm{H}^1_0(\omega_\sigma)$,

$$\|\gamma_\sigma(\mathfrak{b}_\sigma)\|^2_\sigma = (\gamma_\sigma(\mathfrak{b}_\sigma), \gamma_\sigma(\mathfrak{b}_\sigma))_\sigma \leq (\gamma_\sigma(\mathfrak{b}_\sigma), 1)_\sigma \Rightarrow \|\gamma_\sigma(\mathfrak{b}_\sigma)\|_\sigma \lesssim h_\sigma^{\frac{1}{2}},$$
$$\|\mathfrak{b}_\sigma\|^2_{\omega_\sigma} = (\mathfrak{b}_\sigma, \mathfrak{b}_\sigma)_{\omega_\sigma} \leq (\mathfrak{b}_\sigma, 1)_{\omega_\sigma} \Rightarrow \|\mathfrak{b}_\sigma\|_{\omega_\sigma} \lesssim h_\sigma,$$
$$\|\nabla \mathfrak{b}_\sigma\|_{\omega_\sigma} \lesssim h_\sigma^{-1} \|\mathfrak{b}_\sigma\|_{\omega_\sigma} \Rightarrow \|\nabla \mathfrak{b}_\sigma\|_{\omega_\sigma} \lesssim 1.$$

Let $l \subset \omega_\sigma$ be a segment that does not necessarily intersects $\partial\omega_\sigma$, L its straight prolongation whose extrema intersect $\partial\omega_\sigma$. Then, since $\mathfrak{b}_\sigma \in \mathrm{H}^{\frac{1}{2}}_{00}(L)$ and $h_L \leq h_\sigma$, by applying the continuity of the trace operator on L,

$$\|\gamma_l(\mathfrak{b}_\sigma)\|_{\frac{1}{2},l} \leq \|\gamma_L(\mathfrak{b}_\sigma)\|_{\frac{1}{2},L} \lesssim \|\gamma_L(\mathfrak{b}_\sigma)\|_{\mathrm{H}^{\frac{1}{2}}_{00}(L)} \lesssim \|\nabla \mathfrak{b}_\sigma\|_{\omega_\sigma} \lesssim 1. \tag{A.45}$$

Again, constants depend on the minimum angle in ω_σ. The following Lemma, analogous to Lemma A.3, also involves the *continuation operator* defined in [16], which extends a function from a side σ of a triangle T to the whole triangle. We denote this operator by $\mathcal{C}_T : L^\infty(E) \to L^\infty(T)$. The following Lemmas can be proved using [14, Lemma 4.1] and the techniques in [16].

Lemma A.4. *Let $\sigma \subset F_i$ be a segment shared by two triangles R and L, $\mathfrak{b}_\sigma$ its bubble function defined on the union of the two triangles. Let $\mathcal{P}(\sigma) \subset \mathrm{H}^{\frac{1}{2}}(\sigma)$ be a finite dimensional space. Then, for any given $v \in \mathcal{P}(\sigma)$ and any triangle $T \in \{R, L\}$,*

$$\|v\|^2_\sigma \lesssim (v, v\,\gamma_\sigma(\mathfrak{b}_\sigma))_\sigma, \qquad \|v\mathfrak{b}_\sigma\|_\sigma \leq \|v\|_\sigma, \tag{A.46}$$
$$\|\mathcal{C}_T(v)\ \mathfrak{b}_\sigma\|_T \lesssim h_T^{-\frac{1}{2}} \|v\|_\sigma, \qquad \|\mathcal{C}_T(v)\ \mathfrak{b}_\sigma\|_T \lesssim h_T^{\frac{3}{2}} \|v\|_\sigma. \tag{A.47}$$

A.6.2 Efficiency of estimators.

In these subsection we report the results about the efficiency of the estimators. These results, together with the ones in the following subsection, prove Theorem A.3. Lemmas A.5 and A.6 are used in the proofs of the subsequent Propositions.

Lemma A.5. *Let $\boldsymbol{u} \in U$, $\boldsymbol{u}_\delta \in U_\delta$ be the solutions of* (A.6) *and* (A.21)*. Let $h = \mathcal{H}(\boldsymbol{u})$ and $h_\delta = \mathcal{H}_\delta(\boldsymbol{u}_\delta)$. Then, for any given $i \in \mathcal{I}$, $v \in V_i$,*

$$\sum_{\sigma \in \mathcal{E}_{\delta,i}} \left(\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma, \gamma_\sigma(v) \right)_\sigma = \sum_{T \in \mathcal{T}_\delta} (f_i + \Delta h_{\delta i}, v)_T + \left\langle \boldsymbol{u}_i, \gamma_{\mathcal{M}_i}(v) \right\rangle_{\mathcal{M}_i} + (\nabla(h_{\delta i} - h_i), \nabla v)_{F_i}.$$

Proof. Let $i \in \mathcal{I}$, $v \in V_i$. By applying definition (A.4) and using Green's formula on $\mathcal{T}_{\delta,i}$ we obtain

$$(\nabla(h_i - h_{\delta i}), \nabla v)_{F_i} = (f_i, v)_{F_i} + \left\langle \boldsymbol{u}_i, \gamma_{\mathcal{M}_i}(v) \right\rangle_{\mathcal{M}_i} - (\nabla h_{\delta i}, \nabla v)_{F_i} =$$
$$= \sum_{T \in \mathcal{T}_\delta} (f_i + \Delta h_{\delta i}, v)_T + \left\langle \boldsymbol{u}_i, \gamma_{\mathcal{M}_i}(v) \right\rangle_{\mathcal{M}_i} - \sum_{\sigma \in \mathcal{E}_{\delta,i}} \left(\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma, \gamma^i_\sigma(v) \right)_\sigma.$$

□

Lemma A.6. *Let $\mathbf{f}$ be the vector of forcing terms in problem* (A.1) *and $f_T = \frac{1}{|T|}(f_i, 1)_T$ its mean value on each $T \in \mathcal{T}_{\delta,i}$ ($i \in \mathcal{I}$). Then,*

$$\|f_T + \Delta h_{\delta i}\|_T \lesssim \|f_i - f_T\|_T + \frac{1}{h_T} \left[\|h_i - h_{\delta i}\|_T + \sum_{m \in \mathcal{M}_i} \|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2}, \ell_T^m} + \xi^m_{NC,T} \right].$$

Proof. Let $\mathfrak{b}_T$ be the bubble function of T. Then, if $w_T := (f_T + \Delta h_{\delta i})\,\mathfrak{b}_T \in \mathrm{H}^1_0(F_i)$, using Lemmas A.3 and A.5 (where the terms in $\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma}\right]\!\right]_\sigma$ vanish because $\gamma^i_\sigma(w_T) = 0\ \forall \sigma \in \mathcal{E}_{\delta,i}$), using inequalities (A.40) and (A.41) and the definition of w_T, we obtain

$$\begin{aligned}\|f_T + \Delta h_{\delta i}\|_T^2 \lesssim (f_T + \Delta h_{\delta i}, w_T)_T &= (f_i + \Delta h_{\delta i}, w_T)_T + (f_T - f_i, w_T)_T = \\ &= (\nabla(h_i - h_{\delta i}), \nabla w_T)_T - \sum_{m\in\mathcal{M}_i} \left\langle u_i^m, \gamma^i_{\Gamma_m}(w_T)\right\rangle_{\ell_T^m} + (f_T - f_i, w_T)_T \leq \\ &\leq \|f_i - f_T\|_T \|w_T\|_T + |\!|\!| h_i - h_{\delta i} |\!|\!|_T |\!|\!| w_T |\!|\!|_T \\ &+ \sum_{m\in\mathcal{M}_i} \left(\|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\ell_T^m} \left\|\gamma^i_{\Gamma_m}(w_T)\right\|_{\frac{1}{2},\ell_T^m} + \|u_{\delta i}^m\|_{\ell_T^m} \left\|\gamma^i_{\Gamma_m}(w_T)\right\|_{\ell_T^m} \right),\end{aligned}$$

by which the thesis is proved using (A.43), (A.44), the continuity of $\gamma^i_{\Gamma_m}$ from $\mathrm{H}^1_0(T)$ to $\mathrm{H}^{\frac{1}{2}}_{00}(\ell_T^m)$, the fact that $0 \leq \mathfrak{b}_T \leq 1$ and a standard trace inequality:

$$\begin{aligned}\|w_T\|_T = \|(f_T + \Delta h_{\delta i})\,\mathfrak{b}_T\|_T &\leq \|f_T + \Delta h_{\delta i}\|_T\ , \\ |\!|\!| w_T |\!|\!|_T = |\!|\!| (f_T + \Delta h_{\delta i})\,\mathfrak{b}_T |\!|\!|_T &\lesssim h_T^{-1} \|f_T + \nabla h_{\delta i}\|_T\ , \\ \left\|\gamma^i_{\Gamma_m}(w_T)\right\|_{\frac{1}{2},\ell_T^m} \lesssim |\!|\!| w_T |\!|\!|_T &\lesssim h_T^{-1} \|f_T + \Delta h_{\delta i}\|_T\ , \\ \left\|\gamma^i_{\Gamma_m}(w_T)\right\|_{\ell_T^m} \lesssim h_{\ell_T^m}^{\frac{1}{2}} |\!|\!| w_T |\!|\!|_T &\lesssim h_{\ell_T^m}^{\frac{1}{2}} h_T^{-1} \|f_T + \Delta h_{\delta i}\|_T\ .\end{aligned}$$

□

Proposition A.2 (efficiency of $\eta_{R,T}$)**.** *Let $i \in \mathcal{I}$, $T \in \mathcal{T}_{\delta,i}$. Let $\eta_{R,T}$ be the estimator defined by* (A.24)*. Then*

$$\eta_{R,T} \lesssim |\!|\!| h_i - h_{\delta i} |\!|\!|_T + \sum_{m\in\mathcal{M}_i} \left(\|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\ell_T^m} + \xi^m_{NC,T} \right) + h_T \|f_i - f_T\|_T\,.$$

Proof. The thesis immediately follows from Lemma A.6:

$$\begin{aligned}\eta_{R,T} = h_T \|f_i + \Delta h_{\delta i}\|_T \leq h_T \left(\|f_T + \Delta h_{\delta i}\|_T + \|f_i - f_T\|_T\right) &\lesssim \\ \lesssim |\!|\!| h_i - h_{\delta i} |\!|\!|_T + \sum_{m\in\mathcal{M}_i} \left(\|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\ell_T^m} + \xi^m_{NC,T} \right) &+ h_T \|f_i - f_T\|_T\ .\end{aligned}$$

□

Proposition A.3 (efficiency of $\xi_{F,\sigma}$)**.** *Let $i \in \mathcal{I}$, $\sigma \in \mathcal{E}_{\delta,i}$, let $\xi_{F,\sigma}$ be defined by* (A.25)*, let $\xi^m_{NC,T}$ defined by* (A.26) *and $\eta_{R,T}$ by* (A.24)*. Then,*

$$\xi_{F,\sigma} \lesssim \sum_{T\in\omega_\sigma} \left(|\!|\!| h_i - h_{\delta i} |\!|\!|_T + h_\sigma \eta_{R,T} + \sum_{m\in\mathcal{M}_i} \xi^m_{NC,T} \right) + \sum_{m\in\mathcal{M}_i} \|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\Gamma_m\cap\mathring{\omega}_\sigma}\,.$$

Proof. If we define w_σ as the function such that $w_\sigma|_T := \mathcal{C}_T\left(\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma}\right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right) \mathfrak{b}_\sigma\ \forall T \subset \omega_\sigma$ and $w_\sigma \equiv 0$ elsewhere, since $\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma}\right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma}$ belongs to a finite dimensional subspace of

$\mathrm{H}^{\frac{1}{2}}(\sigma)$, it is possible to apply (A.46):

$$\left\| \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right\|_\sigma^2 \lesssim \left(\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma}, \gamma_\sigma^i(w_\sigma) \right)_\sigma = \\ = \left(\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma, w_\sigma \right)_\sigma - \sum_{m \in \mathcal{M}_i} \left(u_{\delta i}^m, \gamma_\sigma^i(w_\sigma) \right)_{\Gamma_m \cap \sigma} .$$

We now use Lemma A.5 on the first term, considering that w_σ vanishes on $\mathcal{E}_{\delta,i} \setminus \{\sigma\}$ and partitioning the traces having $|\Gamma_m \cap \omega_\sigma| \neq 0$ in those for which $|\Gamma_m \cap \sigma| \neq 0$ and those for which $|\Gamma_m \cap \sigma| = 0$:

$$\left\| \left[\!\left[\frac{\partial h_{\delta}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right\|_\sigma^2 \lesssim (f_i + \Delta h_{\delta i}, w_\sigma)_{\omega_\sigma} + (\nabla(h_{\delta i} - h_i), \nabla w_\sigma)_{\omega_\sigma} \\ + \sum_{\substack{m \in \mathcal{M}_i \\ |\Gamma_m \cap \sigma| \neq 0}} \left\langle u_i^m - u_{\delta i}^m, \gamma_{\Gamma_m}^i(w_\sigma) \right\rangle_{\Gamma_m \cap \sigma} + \sum_{\substack{m \in \mathcal{M}_i \\ |\Gamma_m \cap \sigma| = 0}} \left\langle u_i^m, \gamma_{\Gamma_m}^i(w_\sigma) \right\rangle_{\Gamma_m \cap \mathring{\omega}_\sigma} .$$

Now, we add and subtract the quantity $\left(u_{\delta i}^m, \gamma_{\Gamma_m}^i(w_\sigma) \right)_{\Gamma_m \cap \mathring{\omega}_\sigma}$ for all those m such that $|\Gamma_m \cap \sigma| = 0$:

$$\left\| \left[\!\left[\frac{\partial h_{\delta}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right\|_\sigma^2 \lesssim (f_i + \Delta h_{\delta i}, w_\sigma)_{\omega_\sigma} + (\nabla(h_{\delta i} - h_i), \nabla w_\sigma)_{\omega_\sigma} \\ + \sum_{m \in \mathcal{M}_i} \left(\left\langle u_i^m - u_{\delta i}^m, \gamma_{\Gamma_m}^i(w_\sigma) \right\rangle_{\Gamma_m \cap \mathring{\omega}_\sigma} + \left(u_{\delta i}^m, \gamma_{\Gamma_m}^i(w_\sigma) \right)_{\Gamma_m \cap \left(\mathring{\omega}_\sigma \setminus \sigma \right)} \right) = \\ = \sum_{T \subset \omega_\sigma} \Bigg[(f_i + \Delta h_{\delta i}, w_\sigma)_T + (\nabla(h_{\delta i} - h_i), \nabla w_\sigma)_T + \sum_{m \in \mathcal{M}_T} \Big(\left(u_{\delta i}^m, \gamma_{\Gamma_m}^i(w_\sigma) \right)_{\ell_T^m} \\ + \left\langle u_i^m - u_{\delta i}^m, \gamma_{\Gamma_m}^i(w_\sigma) \right\rangle_{\Gamma_m \cap T} \Big) \Bigg] \leq \sum_{T \subset \omega_\sigma} \Bigg[\|f_i + \Delta h_{\delta i}\|_T \|w_\sigma\|_T \\ \times \|w_\sigma\|_T + |\!|\!| h_i - h_{\delta i} |\!|\!|_T |\!|\!| w_\sigma |\!|\!|_T + \sum_{m \in \mathcal{M}_T} \Big(\|u_{\delta i}^m\|_{\ell_T^m} \left\| \gamma_{\Gamma_m}^i(w_\sigma) \right\|_{\ell_T^m} \\ + \|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2}, \Gamma_m \cap T} \left\| \gamma_{\Gamma_m}^i(w_\sigma) \right\|_{\frac{1}{2}, \Gamma_m \cap T} \Big) \Bigg] .$$

It is possible to control norms of w_σ using Appendix A.6.1, (A.46) and (A.47), the continuity of the trace operator and a trace inequality. Indeed, for any given $T \subset \omega_\sigma$

$$\|w_\sigma\|_T = \left\| \mathcal{C}_T \left(\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right) \mathfrak{b}_\sigma \right\|_T \lesssim h_T^{\frac{3}{2}} \left\| \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right\|_\sigma ,$$

$$|\!|\!| w_\sigma |\!|\!|_T = \left\| \mathcal{C}_T \left(\left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right) \mathfrak{b}_\sigma \right\|_T \lesssim h_T^{-\frac{1}{2}} \left\| \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right\|_\sigma ,$$

$$\left\| \gamma_{\Gamma_m}^i(w_\sigma) \right\|_{\ell_T^m} \lesssim h_{\ell_T^m}^{\frac{1}{2}} |\!|\!| w_\sigma |\!|\!|_T \lesssim h_{\ell_T^m}^{\frac{1}{2}} h_T^{-\frac{1}{2}} \left\| \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right\|_\sigma ,$$

$$\left\| \gamma_{\Gamma_m}^i(w_\sigma) \right\|_{\frac{1}{2}, \Gamma_m \cap T} \lesssim |\!|\!| w_\sigma |\!|\!|_T \lesssim h_T^{-\frac{1}{2}} \left\| \left[\!\left[\frac{\partial h_{\delta i}}{\partial \hat{n}_\sigma} \right]\!\right]_\sigma - \tilde{u}_{\delta i,\sigma} \right\|_\sigma ,$$

from which the thesis is obtained by definition of $\xi_{F,\sigma}$, $\eta_{R,T}$, and $\xi_{NC,T}^m$. □

Proposition A.4 (efficiency of $\xi^m_{NC,T}$)**.** *Let $i \in \mathcal{I}$, $T \in \mathcal{T}_{\delta,i}$, $m \in \mathcal{M}_T$ and let $\xi^m_{NC,T}$ be the estimator defined by* (A.26)*. Then, assuming that $u^m_{\delta i}$ has a finite number of jumps in ℓ^m_T,*

$$\xi^m_{NC,T} \lesssim \|u^m_i - u^m_{\delta i}\|_{-\frac{1}{2},\ell^m_T} + \|\!|h_i - h_{\delta i}|\!\|_T + h_{\ell^m_T}\eta_{R,T}\,.$$

Proof. First, suppose $u^m_{\delta i}$ is continuous on ℓ^m_T. Let R, $L \subset T$ be two triangles lying in the interior of T and sharing ℓ^m_T as a side. Let $\mathfrak{b}_{\ell^m_T}$ be the bubble function of ℓ^m_T, having support on $\omega_{\ell^m_T} = R \cup L \subset T$, $\mathcal{C}_R$ and $\mathcal{C}_L$ the continuation operators of R and L respectively. Let $w_{\ell^m_T}$ be the function such that $w_{\ell^m_T}\big|_E := \mathcal{C}_E(u^m_{\delta i})\,\mathfrak{b}_{\ell^m_T}\ \forall E \in \{R, L\}$. Since $\mathcal{C}_R(u^m_{\delta i})$ and $\mathcal{C}_L(u^m_{\delta i})$ belong to a finite dimensional subspace, we can apply (A.46) on the two triangles R and L, obtaining

$$\|u^m_{\delta i}\|^2_{\ell^m_T} \lesssim \Big(u^m_{\delta i}, \gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big)\Big)_{\ell^m_T}\,.$$

Since $u^m_{\delta i} \in \mathrm{L}^2\left(\ell^m_T\right) \subset \mathrm{H}^{-\frac{1}{2}}\left(\ell^m_T\right)$ and $\gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big) \in \mathrm{H}^{\frac{1}{2}}\left(\ell^m_T\right)$, we can rewrite the scalar product above as a duality product. Then, adding and subtracting u^m_i,

$$\begin{aligned}
\|u^m_{\delta i}\|^2_{\ell^m_T} \lesssim{}& \Big\langle u^m_{\delta i}, \gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big)\Big\rangle_{\ell^m_T} = \Big\langle u^m_{\delta i} - u^m_i, \gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big)\Big\rangle_{\ell^m_T} \\
&+ \Big\langle u^m_i, \gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big)\Big\rangle_{\ell^m_T} = \Big\langle u^m_{\delta i} - u^m_i, \gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big)\Big\rangle_{\ell^m_T} + \big(\nabla h_i, \nabla w_{\ell^m_T}\big)_{\omega_{\ell^m_T}} \\
&- \big(f_i, w_{\ell^m_T}\big)_{\omega_{\ell^m_T}} = \Big\langle u^m_{\delta i} - u^m_i, \gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big)\Big\rangle_{\ell^m_T} + \big(\nabla(h_i - h_{\delta i}), \nabla w_{\ell^m_T}\big)_{\omega_{\ell^m_T}} \\
&+ \big(\nabla h_{\delta i}, \nabla w_{\ell^m_T}\big)_{\omega_{\ell^m_T}} - \big(f_i, w_{\ell^m_T}\big)_{\omega_{\ell^m_T}} = \Big\langle u^m_{\delta i} - u^m_i, \gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big)\Big\rangle_{\ell^m_T} \\
&+ \big(\nabla(h_i - h_{\delta i}), \nabla w_{\ell^m_T}\big)_{\omega_{\ell^m_T}} + \big(-f_i - \Delta h_{\delta i}, w_{\ell^m_T}\big)_{\omega_{\ell^m_T}} \le \|u^m_i - u^m_{\delta i}\|_{-\frac{1}{2},\ell^m_T} \\
&\times \Big\|\gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big)\Big\|_{\frac{1}{2},\ell^m_T} + \|\!|h_i - h_{\delta i}|\!\|_{\omega_{\ell^m_T}} \big\|\!\big|w_{\ell^m_T}\big|\!\big\|_{\omega_{\ell^m_T}} + \|f_i + \Delta h_{\delta i}\|_{\omega_{\ell^m_T}} \big\|w_{\ell^m_T}\big\|_{\omega_{\ell^m_T}}\,,
\end{aligned}$$

where Green's formula has been applied, using the fact that there are no jumps of $\nabla h_{\delta i}$ inside $\omega_{\ell^m_T}$. Using the continuity of the trace operator and (A.47), we obtain, for all $E \in \{R, L\}$,

$$\Big\|\gamma^i_{\ell^m_T}\big(w_{\ell^m_T}\big)\Big\|_{\frac{1}{2},\ell^m_T} \lesssim \big\|\!\big|w_{\ell^m_T}\big|\!\big\|_E = \big\|\!\big|\mathcal{C}_E(u^m_{\delta i})\,\mathfrak{b}_{\ell^m_T}\big|\!\big\|_E \lesssim h^{-\frac{1}{2}}_{\ell^m_T}\,\|u^m_{\delta i}\|_{\ell^m_T}\,,$$

and, since $h_{\ell^m_T} \le h_T$,

$$\big\|\!\big|w_{\ell^m_T}\big|\!\big\|_E = \big\|\!\big|\mathcal{C}_E(u^m_{\delta i})\,\mathfrak{b}_{\ell^m_T}\big|\!\big\|_E \lesssim h^{\frac{3}{2}}_{\ell^m_T}\,\|u^m_{\delta i}\|_{\ell^m_T} \le h_T\,h^{\frac{1}{2}}_{\ell^m_T}\,\|u^m_{\delta i}\|_{\ell^m_T}\,.$$

The thesis comes from the definitions of $\eta_{R,T}$, $\xi^m_{NC,T}$ and from $\|\!|\cdot|\!\|_{\omega_{\ell^m_T}} \le \|\!|\cdot|\!\|_T$. If $u^m_{\delta i}$ has some jumps, it is sufficient to apply this procedure on each of the sub-segments of ℓ^m_T upon which it is continuous. □

Proposition A.5 (efficiency of $\eta_{P,T}$ and ξ^i_{P,Γ_m})**.** *Let $i \in \mathcal{I}$, $T \in \mathcal{T}_{\delta,i}$. Then*

$$\eta_{P,T} = \|\!|p_i - p_{\delta i}|\!\|_T\,.$$

Moreover, let $m \in \mathcal{M}_i$, $\lambda \in \Lambda_{m,i}$. Then

$$\xi^i_{P,\lambda} \lesssim \|\!|p_i - p_{\delta i}|\!\|_{\omega_\lambda}\,,$$

where ω_λ is the union of two triangles having λ as one of their sides.

Proof. The first estimate derives directly from the fact that $p_i = 0$. For what concerns the estimate of $\xi^i_{P,\lambda}$, let R and L be two triangles sharing only λ as a side and such that $h_\lambda = h_R = h_L$. Let $\mathfrak{b}_\lambda$ be the bubble function of λ defined on $\omega_\lambda = R \cup L$ and let $w_\lambda = p_{\delta i}\, \mathfrak{b}_\lambda$. We can apply Lemma A.4 because w_λ belongs to a finite dimensional space. Moreover, we can add $\gamma^i_\lambda(p_i) \equiv 0$:

$$\begin{aligned}\left\|\gamma^i_\lambda(p_{\delta i})\right\|^2_\lambda \lesssim (p_{\delta i}, w_\lambda)_\lambda = \left(\gamma^i_\lambda(p_{\delta i} - p_i)\,, \gamma^i_\lambda(w_\lambda)\right)_\lambda \leq \\ \leq \left\|\gamma^i_\lambda(p_i - p_{\delta i})\right\|_\lambda \left\|\gamma^i_\lambda(w_\lambda)\right\|_\lambda \leq \left\|\gamma^i_\lambda(p_i - p_{\delta i})\right\|_{\frac{1}{2},\lambda} \left\|\gamma^i_\lambda(w_\lambda)\right\|_{\frac{1}{2},\lambda}.\end{aligned}$$

Thus, for the continuity of the trace operator, (A.47) and the fact that $h_\lambda = h_R = h_L$,

$$\begin{aligned}\left\|\gamma^i_\lambda(p_{\delta i})\right\|^2_\lambda \lesssim \left\|\gamma^i_\lambda(p_i - p_{\delta i})\right\|_{\frac{1}{2},\lambda} \left\|\gamma^i_\lambda(p_{\delta i}\, \mathfrak{b}_\lambda)\right\|_{\frac{1}{2},\lambda} \lesssim \left|\!\left|\!\left|p_i - p_{\delta i}\right|\!\right|\!\right|_{\omega_\lambda} \left|\!\left|\!\left|p_{\delta i}\, \mathfrak{b}_\lambda\right|\!\right|\!\right|_{\omega_\lambda} \lesssim \\ \lesssim \left|\!\left|\!\left|p_i - p_{\delta i}\right|\!\right|\!\right|_{\omega_\lambda} \sum_{T \in \omega_\lambda} h_T^{-\frac{1}{2}} \left\|\gamma^i_\lambda(p_{\delta i})\right\|_\lambda \lesssim h_\lambda^{-\frac{1}{2}} \left|\!\left|\!\left|p_i - p_{\delta i}\right|\!\right|\!\right|_{\omega_\lambda} \left\|\gamma^i_\lambda(p_{\delta i})\right\|_\lambda.\end{aligned}$$

The thesis is proved since $\xi^i_{P,\lambda} = h_\lambda^{\frac{1}{2}} \left\|\gamma^i_\lambda(p_{\delta i})\right\|_\lambda$. □

Proposition A.6 (efficiency of $J^i_{\delta,\lambda}$)**.** *Let $\boldsymbol{u}_\delta$ be the solution of* (A.21)*, $\boldsymbol{h}_\delta = \mathcal{H}_\delta(\boldsymbol{u}_\delta)$. Let $i \in \mathcal{I}$, $m \in \mathcal{M}_i$, $\lambda \in \Lambda_{m,i}$, $J^i_{\delta,\lambda}(\boldsymbol{u}_\delta)$ be the quantity defined by* (A.29)*. Then, if $\Gamma_m = F_i \cap F_j$,*

$$J^i_{\delta,\lambda} \lesssim \left|\!\left|\!\left|h_i - h_{\delta i}\right|\!\right|\!\right|_{F_i} + \left|\!\left|\!\left|h_j - h_{\delta j}\right|\!\right|\!\right|_{F_j} + \left\|u^m_i - u^m_{\delta i}\right\|_{-\frac{1}{2},\lambda} + \left\|u^m_j - u^m_{\delta j}\right\|_{-\frac{1}{2},\lambda}.$$

Proof. We recall that

$$J^i_{\delta,\lambda}(\boldsymbol{u}_\delta) = h_\lambda^{\frac{1}{2}} \left(\left\|\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}\right\|_\lambda + \left\|[\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}\right\|_\lambda\right).$$

The proof follows the same pattern as the one done for Proposition A.5. Consider two triangles R, $L \subset F_i$ (possibly not in $\mathcal{T}_{\delta,i}$) having λ as common side and such that $h_\lambda = h_R = h_L$. Define $\omega_\lambda = R \cup L$ and let $\mathfrak{b}_\lambda$ be the bubble function of λ defined on ω_λ. We estimate the two terms of the sum separately.

First, let $w_{\lambda,u}|_E := \mathcal{C}_E\big(\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}\big)\, \mathfrak{b}_\lambda \ \forall E \in \{R, L\}$. It is possible to apply (A.46) and consider $u^m_i + u^m_j = 0$:

$$\begin{aligned}\left\|\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}\right\|^2_\lambda \lesssim \left(\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}, \gamma_{\Gamma_m}(w_{\lambda,u})\right)_\lambda = \left(u^m_{\delta i} + u^m_{\delta j}, \gamma_{\Gamma_m}(w_{\lambda,u})\right)_\lambda \\ = \left\langle u^m_{\delta i} + u^m_{\delta j} + u^m_i - u^m_j, w_{\lambda,u}\right\rangle_\lambda \leq \left[\left\|u^m_i - u^m_{\delta i}\right\|_{-\frac{1}{2},\lambda} + \left\|u^m_j - u^m_{\delta j}\right\|_{-\frac{1}{2},\lambda}\right] \\ \times \left\|w_{\lambda,u}\right\|_{\frac{1}{2},\lambda}.\end{aligned}$$

Since by continuity of the trace operator and applying (A.47) we have

$$\left\|w_{\lambda,u}\right\|_{\frac{1}{2},\lambda} \lesssim h_\lambda^{-\frac{1}{2}} \left\|\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}\right\|_\lambda,$$

and since $h_\lambda = h_R = h_L$,

$$\left\|\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}\right\|^2_\lambda \lesssim h_\lambda^{-\frac{1}{2}} \left(\left\|u^m_i - u^m_{\delta i}\right\|_{-\frac{1}{2},\lambda} + \left\|u^m_j - u^m_{\delta j}\right\|_{-\frac{1}{2},\lambda}\right) \left\|\{\{\boldsymbol{u}_\delta\}\}_{\Gamma_m}\right\|_\lambda,$$

and this estimates the first term.

Similarly, let $w_{\lambda,h}|_E := \mathcal{C}_E\big([\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}\big)\,\mathfrak{b}_\lambda\ \forall E \in \{R, L\}$. It is possible to apply (A.46). Then, since $\gamma^i_{\Gamma_m}(h_i) \equiv \gamma_j(h_j)\,\Gamma_m$ and $\gamma^i_{\Gamma_m}(h_i)\,,\ \gamma^j_{\Gamma_m}(h_j) \in \mathrm{H}^{\frac{1}{2}}(\lambda) \subset \mathrm{L}^2(\lambda)$, we have

$$\begin{aligned}\big\|[\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}\big\|_\lambda \lesssim \big([\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}\,,\gamma^i_{\Gamma_m}(w_{\lambda,h})\big)_\lambda &= \Big(\gamma^i_{\Gamma_m}(h_{\delta i}) - \gamma^j_{\Gamma_m}(h_{\delta j})\,,\gamma^i_{\Gamma_m}(w_{\lambda,h})\Big)_\lambda = \\ &= \Big(\gamma^i_{\Gamma_m}(h_{\delta i}) - \gamma^j_{\Gamma_m}(h_{\delta j}) - \gamma^i_{\Gamma_m}(h_i) + \gamma^j_{\Gamma_m}(h_j)\,,\gamma^i_{\Gamma_m}(w_{\lambda,h})\Big)_\lambda \le \\ &\le \left(\big\|\gamma^i_{\Gamma_m}(h_i - h_{\delta i})\big\|_{\frac{1}{2},\lambda} + \Big\|\gamma^j_{\Gamma_m}(h_j - h_{\delta j})\Big\|_{\frac{1}{2},\lambda}\right)\big\|[\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}\,\gamma^i_{\Gamma_m}(\mathfrak{b}_\lambda)\big\|_{\frac{1}{2},\lambda}\,,\end{aligned}$$

if we define $\omega^j_{\lambda,h} \subset F_j$ such that $\lambda \subset \partial\omega^j_{\lambda,h}$. Applying the continuity of the trace operator on F_i and F_j, (A.47) and since $h_\lambda = h_R = h_L$, we have estimated the second term:

$$\begin{aligned}\big\|[\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}\big\|_\lambda \lesssim \Big(|\!|\!|h_i - h_{\delta i}|\!|\!|_{\omega_\lambda} + |\!|\!|h_j - h_{\delta j}|\!|\!|_{\omega^j_\lambda}\Big)\,|\!|\!|w_{\lambda,h}|\!|\!|_{\omega_\lambda} &\lesssim \\ \lesssim h_\lambda^{-\frac{1}{2}}\Big(|\!|\!|h_i - h_{\delta i}|\!|\!|_{F_i} + |\!|\!|h_j - h_{\delta j}|\!|\!|_{F_j}\Big)\,\big\|[\![\boldsymbol{h}_\delta]\!]_{\Gamma_m}\big\|_\lambda\,.\end{aligned}$$

□

A.6.3 Final lower bounds. In this subsection we collect the previous efficiency results to complete the proof of Theorem A.3.

Assuming $h_{NC,T} < 1\ \forall T \in \mathcal{T}_\delta$, we first look at the result of Proposition A.2, together with the result of Proposition A.4. From these we can obtain an efficiency estimate for $\eta_{R,T}$ involving only the exact errors and higher order terms (this is standard: see for example [16]). For any given $i \in \mathcal{I}$ and a triangle $T \in \mathcal{T}_{\delta,i}$,

$$\begin{aligned}\eta_{R,T} \lesssim |\!|\!|h_i - h_{\delta i}|\!|\!|_T + \sum_{m\in\mathcal{M}_T}\|u_i - u_{\delta i}\|_{-\frac{1}{2},\ell^m_T} + h_T\,\|f_i - f_T\|_T \\ + \sum_{m\in\mathcal{M}_T}\Big(\|u_i - u_{\delta i}\|_{-\frac{1}{2},\ell^m_T} + |\!|\!|h_i - h_{\delta i}|\!|\!|_T + h_{\ell^m_T}\eta_{R,T}\Big)\,.\end{aligned}$$

Then, since $\#\mathcal{M}_T \le \#\mathcal{M}_i$ and $\#\mathcal{M}_i$ is fixed, thanks to the assumption $h_{NC,T} < 1$ we have

$$\begin{aligned}\eta_{R,T} \lesssim \frac{1}{1-h_{NC,T}}\left[|\!|\!|h_i - h_{\delta i}|\!|\!|_T + \sum_{m\in\mathcal{M}_T}\|u^m_i - u^m_{\delta i}\|_{-\frac{1}{2},\ell^m_T}\right] \\ + \frac{h_T}{1-h_{NC,T}}\,\|f_i - f_T\|_T\,. \quad \text{(A.48)}\end{aligned}$$

Now we consider Proposition A.4. Since $\#\mathcal{M}_T$ is bounded independently on the discretization, summing on all $m \in \mathcal{M}_T$ both members we obtain

$$\begin{aligned}\sum_{m\in\mathcal{M}_T}\xi_{NC,T} \lesssim \#\mathcal{M}_T\,|\!|\!|h_i - h_{\delta i}|\!|\!|_T + \sum_{m\in\mathcal{M}_T}\|u^m_i - u^m_{\delta i}\|_{-\frac{1}{2},\ell^m_T} + \underbrace{\left(\sum_{m\in\mathcal{M}_T}h_{\ell^m_T}\right)}_{h_{NC,T}}\eta_{R,T} \lesssim \\ \lesssim |\!|\!|h_i - h_{\delta i}|\!|\!|_T + \sum_{m\in\mathcal{M}_T}\|u^m_i - u^m_{\delta i}\|_{-\frac{1}{2},\ell^m_T} + h_{NC,T}\eta_{R,T}.\end{aligned}$$

We remark that the constants may depend on $\#\mathcal{M}_T \leq \#\mathcal{M}_i$. We now make use of Proposition A.2 for bounding $\eta_{R,T}$ with the exact error and $\xi_{NC,T}$, obtaining

$$\sum_{m\in\mathcal{M}_T} \xi_{NC,T} \lesssim |\!|\!| h_i - h_{\delta i} |\!|\!|_T + \sum_{m\in\mathcal{M}_T} \|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\ell_T^m} + h_{NC,T} \Bigg[|\!|\!| h_i - h_{\delta i} |\!|\!|_T + \sum_{m\in\mathcal{M}_T} \|u_i - u_{\delta i}\|_{-\frac{1}{2},\ell_T^m} + h_T \|f_i - f_T\|_T \Bigg] + h_{NC,T} \sum_{m\in\mathcal{M}_T} \xi_{NC,T} \,.$$

Then,

$$\sum_{m\in\mathcal{M}_T} \xi_{NC,T}^m \lesssim \frac{1+h_{NC,T}}{1-h_{NC,T}} \left[|\!|\!| h_i - h_{\delta i} |\!|\!|_T \sum_{m\in\mathcal{M}_T} \|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\ell_T^m} \right] + \frac{h_{NC,T} h_T}{1-h_{NC,T}} \|f_i - f_T\|_T \,. \quad \text{(A.49)}$$

The influence of non-conformity on the efficiency of our estimate is clear.

Finally, let's turn to the result of Proposition A.3. To have an explicit estimate for $\xi_{F,\sigma}$ we use equations (A.48), (A.49) and the fact that

$$\|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\Gamma_m\cap\mathring{\omega}_\sigma} \leq \sum_{T\in\omega_\sigma} \|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\ell_T^m} \,.$$

For any given $i \in \mathcal{I}$, $\sigma \in \mathcal{E}_{\delta,i}$ and indicating by ω_σ the set of triangles in $\mathcal{T}_{\delta,i}$ having σ as one of their sides, algebraic calculations yield

$$\xi_{F,\sigma} \lesssim \max\{1, h_\sigma\} \sum_{T\in\omega_\sigma} \frac{1}{1-h_{NC,T}} \left[|\!|\!| h_i - h_{\delta i} |\!|\!|_T + + \sum_{m\in\mathcal{M}_T} \|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\ell_T^m} \right] + \frac{h_T(1+h_{NC,T})}{1-h_{NC,T}} \|f_i - f_T\|_T \,, \quad \text{(A.50)}$$

where we see the same kind of dependence. Using the results from Propositions A.5 and A.6 and (A.48)–(A.50), we can prove Theorem A.3.

A.7 Numerical Results

We show here the results of numerical experiments mainly performed in order to evaluate the *effectivity index*, defined as the ratio between the *true error* err and the *estimated error* est_δ (see Table A.1):

$$\varepsilon := \frac{err}{est_\delta} \,.$$

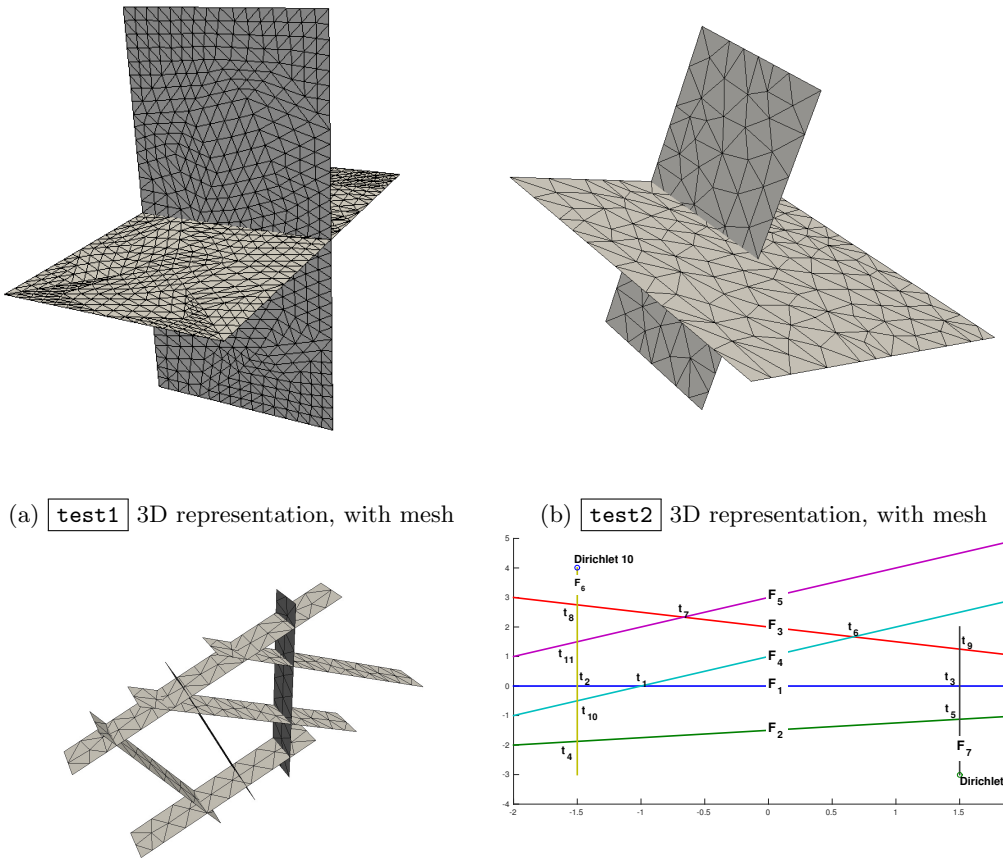
In order to approximate the norm of the error $\boldsymbol{u} - \boldsymbol{u}_\delta$ in the dual space $\mathrm{H}^{-\frac{1}{2}}(\Gamma_m)$, for all $m \in \mathcal{M}$, we have used the following weighted $\mathrm{L}^2(\Gamma_m)$ norm:

$$\forall i \in \mathcal{I},\ \forall m \in \mathcal{M}_i,\ \|u_i^m - u_{\delta i}^m\|_{-\frac{1}{2},\Gamma_m} \sim \left(\sum_{\lambda\in\Lambda_{m,i}} h_\lambda \|u_i^m - u_{\delta i}^m\|_{\Gamma_m}^2 \right)^{\frac{1}{2}} \,,$$

where $\Lambda_{m,i}$ is defined as the discretization of the trace Γ_m induced by the mesh $\mathcal{T}_{\delta,i}$.

We consider three DFNs, as shown in Figure A.2. All simulations are performed using linear finite elements on a sequence of refined grids, and using, for each trace Γ_m, a continuous piecewise linear approximation for u_i^m and u_j^m on the induced meshes $\Lambda_{m,i}$ and $\Lambda_{m,j}$, respectively. In all the considered meshes, traces are arbitrarily placed with respect to the mesh-edges (full non-conformity between the meshes).

All the results are collected in Table A.1, where the effectivity index ε for the different cases is reported. Further, Figure A.3 shows the behaviour of the error estimator est_δ and of the error err with respect to the meshsize for the three DFNs.



(a) `test1` 3D representation, with mesh

(b) `test2` 3D representation, with mesh

(c) `7_fract` 3D representation, with mesh

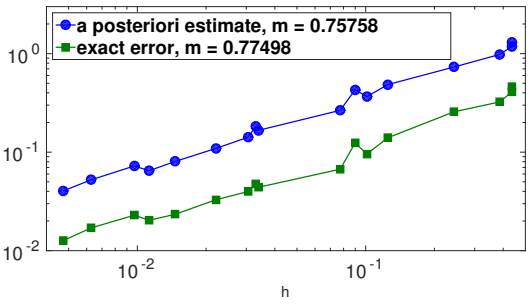
(d) `7_fract` 2D representation, with boundary conditions

Figure A.2: Views of the considered DFNs

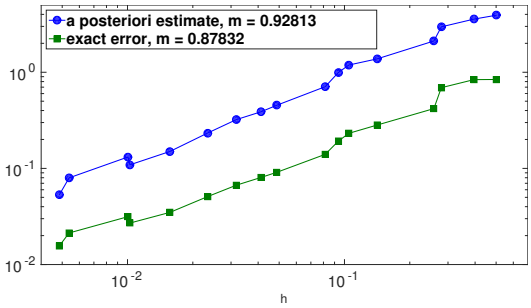
A.7.1 Problem `test1`.

The first test problem deals with two identical fractures intersecting each other orthogonally (see Figure A.2a):

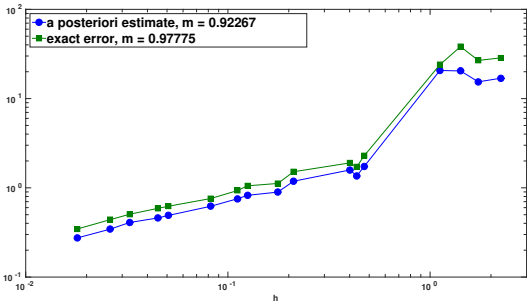
$$F_1 = \{(x, y, z) \in \mathbb{R}^3 : \ z \in (-1, 1), \ y \in (0, 1), x = 0\},$$
$$F_2 = \{(x, y, z) \in \mathbb{R}^3 : \ x \in (-1, 1), \ y \in (0, 1), z = 0\}.$$



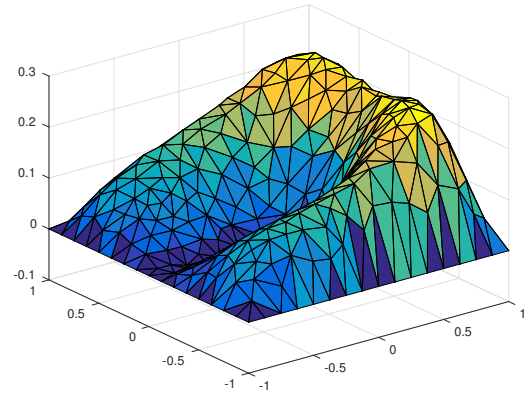
(a) `test1` Error estimate (blue) and error (green) vs. h



(b) `test2` Error estimate (blue) and error (green) vs. h



(c) `7_fract` Error estimate (blue) and error (green) vs. h



(d) `test2` Discrete solution in F_1

Figure A.3

$\max_{T\in\mathcal{T}_\delta} h_T$	ε
0.0047	0.3130
0.0063	0.3251
0.0097	0.3160
0.0112	0.3146
0.0146	0.2907
0.0221	0.3011
0.0305	0.2810
0.0331	0.2595
0.0339	0.2643
0.0773	0.2526
0.0899	0.2892
0.1012	0.2614
0.1250	0.2888
0.2435	0.3507
0.3860	0.3297
0.4373	0.3454
0.4375	0.3518

(a) `test1`

$\max_{T\in\mathcal{T}_\delta} h_T$	ε
0.0049	0.2943
0.0054	0.2672
0.0101	0.2394
0.0103	0.2478
0.0157	0.2324
0.0234	0.2183
0.0317	0.2079
0.0413	0.2075
0.0488	0.2007
0.0815	0.1978
0.0938	0.1933
0.1047	0.1949
0.1415	0.2050
0.2582	0.1980
0.2795	0.2329
0.3953	0.2347
0.5014	0.2146

(b) `test2`

$\max_{T\in\mathcal{T}_\delta} h_T$	ε
0.0180	1.2598
0.0261	1.2769
0.0327	1.2377
0.0450	1.2809
0.0508	1.2676
0.0822	1.2210
0.1115	1.2359
0.1250	1.2780
0.1766	1.2505
0.2111	1.2763
0.4004	1.2003
0.4337	1.2604
0.4719	1.3189
1.1180	1.1613
1.4142	1.8722
1.7321	1.7375
2.2361	1.6907

(c) `7_fract`

Table A.1: Tables of effectivity indexes

We have $\mathcal{M} = \{1\}$ and $\Gamma_1 = \{(x, y, z) \in \mathbb{R}^3 : \ x = 0, z = 0, y \in (0, 1)\}$, we set homogeneous Dirichlet boundary conditions on both fractures. For further details regarding this problem we refer to [6].

Results for this first problem are reported in Table A.1a, where the values of the effectivity indices for different meshsizes are shown. We can see that the effectivity index is almost independent of the meshsize, with values oscillating in a range between 0.2526 and 0.3518 for h spanning two orders of magnitude. In Figure A.3a we plot the true errors and the estimated errors. In the legend of this Figure we report the exponent m of the fitting of these errors with respect to h ($err \sim h^m$ and $est_\delta \sim h^m$). The plots show a good agreement between the error and the estimator.

A.7.2 Problem `test2`. In the second test problem we consider the two fracture DFN displayed in Figure A.2b. In particular, F_1 is not intersected completely by F_2:

$$F_1 = \left\{(x, y, z) \in \mathbb{R}^3 : \ -1 < x < 1, \ -1 < y < 1, \ z = 0\right\},$$
$$F_2 = \left\{(x, y, z) \in \mathbb{R}^3 : \ -1 < x < 0, \ y = 0, \ -1 < z < 1\right\}.$$

Again, we have $\mathcal{M} = \{1\}$ and we set $\Gamma_1 = \left\{(x, y, z) \in \mathbb{R}^3 : \ y = z = 0, \ -1 < x < 0\right\}$, and Dirichlet boundary conditions are set on all the boundaries. In this case we have a less regular solution around the trace tip (see [3, 8]). In Figure A.3d we report a computed solution on F_1. In Table A.1b we report the values of the effectivity indices for different meshsize. We can see that, again, these values are quite stable with respect to the meshsize. In Figure A.3b we plot the true errors and the estimated errors and report the slopes m of the fitting. The plots show a good agreement between the error and the estimator.

A.7.3 Problem `7_fract`. The last test problem considers the DFN of 7 fractures intersecting in 11 traces shown in Figure A.2c. We set a constant Dirichlet boundary condition

$h_D = 10$ on one side of F_6 and an homogeneous Dirichlet boundary condition on one side of F_7 (see Figure A.2d). The stated problem has a piecewise linear solution on each fracture (see [6]), that could be exactly computed by the FEM method if one had meshes totally conforming to traces. This is not the case of our meshes, thus this is a good test for the behaviour of the non-conformity estimators. In Figure A.3c we plot the true errors and the estimated errors and in Table A.1c we report the values of the effectivity indices. We note that for the three coarsest meshes the effectivity indices are larger than the values observed for the other meshes. This shows the influence of non-conformity on the efficiency of the estimate: indeed, these meshes feature a non-conformity indicator $\max_{T\in\mathcal{T}_\delta} h_{NC,T} \approx 0.8571$, which yields $C_{NC} \approx 13$ (Theorem A.3). With mesh refinement, starting from the fourth coarsest mesh, we have $\max_{T\in\mathcal{T}_\delta} h_{NC,T} \leq 0.5$, and the value of C_{NC} critically drops to values lower than or equal to 3 (see Remark A.2) and the effectivity index becomes almost constant.

A.7.4 Estimators characterization. It is interesting to characterize the estimators with respect to the information they can provide about the distribution of the errors on the domain. With this target we define, for all $T \in \mathcal{T}_\delta$, two different indicators:

$$\eta_{res,T} := \begin{cases} \eta_{R,T} + \frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\delta,T}} \xi_{F,\sigma} & \text{if } \forall m \in \mathcal{M},\ |\Gamma_m \cap T| = 0 \\ \eta_{R,T} & \text{otherwhise} \end{cases}$$

$$\eta_{tr,T} := \begin{cases} \eta_{P,T} + \xi_{P,T} + \xi_{NC,T} + \frac{1}{2}\sum_{\sigma\in\mathcal{E}_{\delta,T}} \xi_{F,\sigma} & \text{if } \exists m \in \mathcal{M}:\ |\Gamma_m \cap T| \neq 0 \\ \eta_{P,T} + \xi_{P,T} + \xi_{NC,T} & \text{otherwhise} \end{cases}$$

where $\mathcal{E}_{\delta,T}$ indicates the set of the edges of T. In Figure A.4 we see the behaviour of these two quantities for problem `test2` on F_1, whose solution is depicted in Figure A.3d. The quantity $\eta_{res,T}$ provides information about the error on each fracture that is related to the Finite Element approximation of the solution of (A.9) in the interior of the fractures far from the traces. On the other hand, the quantity $\eta_{tr,T}$ provides information about the non conformity errors and the violation of matching conditions on the traces. In Figure A.4 we plot these two quantities on the elements of two different meshes for F_1, being the coarsest of these meshes the one used for the solution shown in Figure A.3d. On the first two top figures we report $\eta_{res,T}$ (left) and $\eta_{tr,T}$ (right) on the coarse mesh, we see that $\eta_{res,T}$ is larger where the solution displays strong curvatures, i.e. far from the trace and around trace tip. Instead, as expected, the conformity indicator $\eta_{tr,T}$ is concentrated around the trace. A similar behaviour with different order of magnitude of the estimators is obtained on the finer grid.

A.8 Conclusions

In this chapter we have derived residual based "a posteriori" error estimates for the constrained optimization formulation of a simple model for the flow in DFNs. Numerical results have confirmed very weak dependence of the effectivity index on the meshsize, a very good agreement between the estimator and the error distribution, and we have identified a parameter correlating the estimate with the non-conformity of the meshes. The terms of the estimator can be collected in two indicators with a clear meaning: an indicator related to the error inside each fracture, essentially related to the attitude of the Finite Element space to describe the hydraulic head in each fracture, and a second indicator essentially concerning

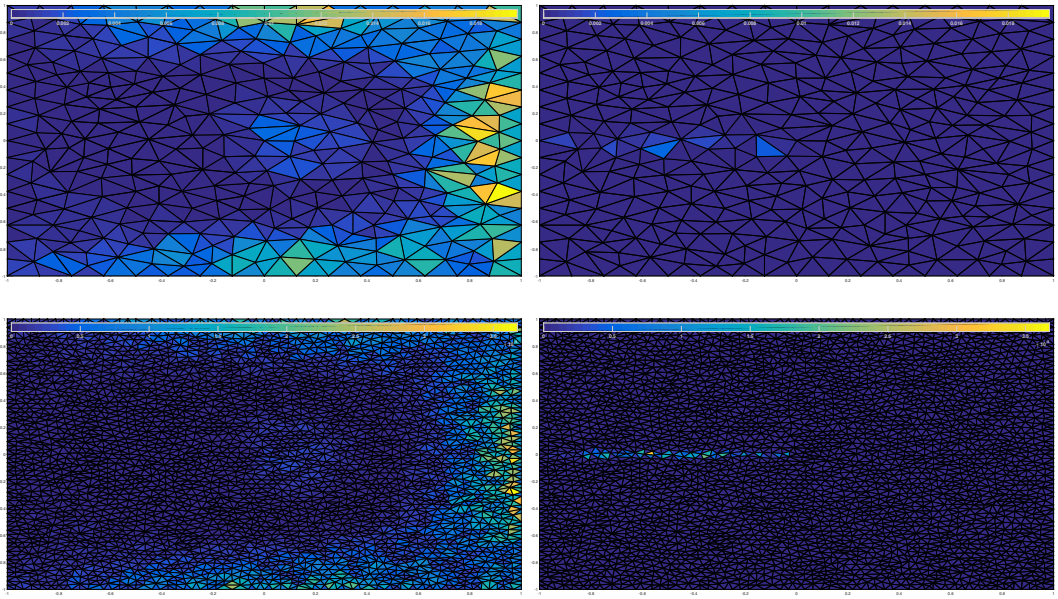


Figure A.4: `test2` Behaviour of $\eta_{res,T}$ (left) and $\eta_{tr,T}$ with two different mesh sizes

the lack of continuity of the hydraulic head, the flux mismatch at the traces and the non-conformity of the meshes of intersecting fractures. The different nature of the estimators is a useful tool for their use in an adaptive algorithm that will be object of future investigations.

References for Appendix A

[1] M. Ainsworth and J. T. Oden. *A posteriori error estimation in finite element analysis*. Vol. 37. New York: John Wiley & Sons, 2011.

[2] M. Benedetto, S. Berrone, A. Borio, S. Pieraccini, and S. Scialò. "A Hybrid Mortar Virtual Element Method For Discrete Fracture Network Simulations". In: *J. Comput. Phys.* 306 (2016), pp. 148–166. DOI: 10.1016/j.jcp.2015.11.034.

[3] M. Benedetto, S. Berrone, S. Pieraccini, and S. Scialò. "The virtual element method for discrete fracture network simulations". In: *Comput. Methods Appl. Mech. Engrg.* 280.0 (2014), pp. 135–156. ISSN: 0045-7825. DOI: 10.1016/j.cma.2014.07.016.

[4] M. Benedetto, S. Berrone, and S. Scialò. "A Globally Conforming Method For Solving Flow in Discrete Fracture Networks Using the Virtual Element Method". In: *Finite Elem. Anal. Des.* 109 (2016), pp. 23–36. DOI: 10.1016/j.finel.2015.10.003.

[5] S. Berrone, A. Borio, and S. Scialò. "A posteriori error estimate for a PDE-constrained optimization formulation for the flow in DFNs". In: *SIAM J. Numer. Anal.* 54.1 (2016), pp. 242–261. DOI: 10.1137/15M1014760.

[6] S. Berrone, S. Pieraccini, and S. Scialò. "A PDE-constrained optimization formulation for discrete fracture network flows". In: *SIAM J. Sci. Comput.* 35.2 (2013), B487–B510. ISSN: 1064-8275. DOI: 10.1137/120865884.

[7] S. Berrone, S. Pieraccini, and S. Scialò. “An optimization approach for large scale simulations of discrete fracture network flows”. In: *J. Comput. Phys.* 256 (2014), pp. 838–853. ISSN: 0021-9991. DOI: 10.1016/j.jcp.2013.09.028.

[8] S. Berrone, S. Pieraccini, and S. Scialò. “On simulations of discrete fracture network flows with an optimization-based extended finite element method”. In: *SIAM J. Sci. Comput.* 35.2 (2013), A908–A935. ISSN: 1064-8275. DOI: 10.1137/120882883.

[9] S. Berrone, S. Pieraccini, S. Scialò, and F. Vicini. “A parallel solver for large scale DFN flow simulations”. In: *SIAM J. Sci. Comput.* 37.3 (2015), pp. C285–C306. DOI: 10.1137/140984014.

[10] P. G. Ciarlet. *Numerical analysis of the finite element method*. Vol. 59. Presses de l’Université de Montréal, 1976.

[11] A. Ern and M. Vohralík. “Flux reconstruction and a posteriori error estimation for discontinuous Galerkin methods on general nonmatching grids”. In: *C. R. Math. Acad. Sci. Paris* 347.7-8 (2009), pp. 441–444. ISSN: 1631-073X. DOI: 10.1016/j.crma.2009.01.017.

[12] P. Jiránek, Z. Strakoš, and M. Vohralík. “A Posteriori Error Estimates Including Algebraic Error and Stopping Criteria for Iterative Solvers”. In: *SIAM Journal on Scientific Computing* 32.3 (2010), pp. 1567–1590. DOI: 10.1137/08073706X. eprint: 10.1137/08073706X.

[13] J. Nečas. “Sur une méthode pour résoudre les équations aux dérivées partielles du type elliptique, voisine de la variationnelle.” French. In: *Ann. Sc. Norm. Super. Pisa, Sci. Fis. Mat., III. Ser.* 16 (1962). Ed. by B. Zanichelli, pp. 305–326. ISSN: 0036-9918.

[14] R. Verfürth. “A posteriori error estimation and adaptive mesh-refinement techniques”. In: *Journal of Computational and Applied Mathematics* 50 (1994), pp. 67–83.

[15] R. Verfürth. *A posteriori error estimation techniques for finite element methods*. Oxford: Oxford University Press, 2013.

[16] R. Verfürth. *A Review of a Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*. Advances in Numerical Mathematics Series. Teubner B.G. GmbH, 1996. ISBN: 9783519026051.

[17] M. Vohralík. “Guaranteed and fully robust a posteriori error estimates for conforming discretizations of diffusion problems with discontinuous coefficients”. In: *Journal of Scientific Computing* 46.3 (2011), pp. 397–438.