T-coercivity and continuous Galerkin methods: application to transmission problems with sign changing coefficients

Lucas Chesnel, Patrick Ciarlet

Numer. Math. manuscript No.
(will be inserted by the editor)

T-coercivity and continuous Galerkin methods: application to transmission problems with sign changing coefficients

Lucas Chesnel · Patrick Ciarlet Jr.

Received: Version April 18, 2012 / Accepted: date

Abstract To solve variational indefinite problems, one uses classically the Banach–Nečas–Babuška theory. Here, we study an alternate theory to solve those problems: T-coercivity. Moreover, we prove that one can use this theory to solve the approximate problems, which provides an alternative to the celebrated Fortin lemma. We apply this theory to solve the indefinite problem $\operatorname{div}\sigma\nabla u = f$ set in H_0^1, with σ exhibiting a sign change.

Keywords T-coercivity · metamaterial · negative material · transmission problem ·

1 Introduction

In recent years, some studies have been devoted to the indefinite transmission problem: find $u \in H_0^1(\Omega)$ such that $\operatorname{div}\sigma\nabla u = f$, with a coefficient σ that exhibits a sign change at the crossing of an interface that divides the (bounded) domain Ω. Such is the case of a structure made of a classical dielectrics and of a (negative) metamaterial [21,11,17,13]. This problem is indefinite in the sense that the corresponding sesquilinear form, namely

$$a \ : \ (v,w) \mapsto \int_\Omega \sigma\,\nabla v \cdot \overline{\nabla w}$$

has no fixed sign. One can find v_1, respectively v_2, such that $a(v_1,v_1) > 0$ and $a(v_2,v_2) < 0$. Obviously, it is not coercive so that one can not use the Lax-Milgram theorem to prove that this problem is well-posed. A possible choice is to use the Banach–Nečas–Babuška theory, which relies on the inf-sup condition. Here, we propose instead an alternative choice, the so-called T-coercivity

Lucas Chesnel – Patrick Ciarlet Jr.
Laboratoire POEMS, UMR 7231 CNRS/ENSTA/INRIA
ENSTA ParisTech, 32, boulevard Victor, 75739 Paris Cedex 15, France
E-mail: lucas.chesnel@ensta-paristech.fr – E-mail: patrick.ciarlet@ensta-paristech.fr

theory [3,1] to solve the problem, which relies on the use of explicit inf-sup operators. Interestingly, it can also be used to prove the convergence of finite element discretizations [3,20].

In this paper, we first reformulate the standard well-posedness theory within the T-coercivity framework. Then, we explain how can one use this approach to solve the approximate problems and to prove the convergence of the approximate solutions to the exact solution. Next, we apply these results to the indefinite transmission problem set in $H_0^1(\Omega)$ with a piecewise constant coefficient σ. For the exact problem, we investigate some reference configurations to explain the results we have obtained in terms of the applicability of the method: its well-posedness (possibly in the Fredholm sense) depends critically on the value of the ratio between the positive values and the negative values of σ. We also introduce different approaches to solve numerically the problem using the finite element method. They rely either on the use of special meshes, or on the introduction of some dissipation, which amounts to adding some well-chosen imaginary number to σ. We finally devote our attention to the range of applicability of those discrete approaches, thus complementing the results of [3,20]. In the process, we provide error estimates, which we observe numerically on some examples.

2 General framework

Below, we recall some very standard tools of functional analysis dealing with the well-posedness of an abstract Problem (usually written as a variational formulation), which we reformulate using the theory of T-coercivity [3]. Then, we derive results on a class of indefinite problems by studying their well-posedness via T-coercivity (cf. §3.2).

2.1 Starting point

Let V and W be two Hilbert spaces with inner product $(\cdot,\cdot)_V$ and $(\cdot,\cdot)_W$. We denote $\|\cdot\|_V$ and $\|\cdot\|_W$ the associated norms and by $\mathcal{L}(V,W)$ the vector space of continuous (linear) operators from V to W. Let us introduce $a(\cdot,\cdot)$ a continuous sesquilinear form over $V \times W$ and $f \in W'$. Here, W' refers to the topological dual space of W. The duality pairing is denoted $\langle\cdot,\cdot\rangle$ and the norm is defined by

$$\|f\|_{W'} := \sup_{w\in W\setminus\{0\}} \frac{|\langle f, w\rangle|}{\|w\|_W}.$$

We consider the variational problem

$$\begin{cases} \textit{Find } u \in V \textit{ such that} \\ \forall w \in W,\ a(u,w) = \langle f, w\rangle. \end{cases} \tag{1}$$

First, let us recall a classical definition below.

Definition 1 (Hadamard) Problem (1) is *well-posed* if, and only if, for all f, it has one and only one solution u, with continuous dependence:

$$\exists C > 0,\ \forall f \in W',\ \|u\|_V \leq C\|f\|_{W'}.$$

We define the operator $A \in \mathcal{L}(V, W')$ (the set of bounded operators from V to W') such that $\langle Au, w\rangle = a(u, w)$ for all $w \in W$. It is possible to reformulate Problem (1) as follows

$$\begin{cases} \textit{Find } u \in V \textit{ such that} \\ Au = f \text{ in } W'. \end{cases} \tag{2}$$

Problem (1) is well-posed if, and only if A is an isomorphism from V to W'.

2.2 Well-posedness of the problem: the T-coercivity as a reformulation of the Banach–Nečas–Babuška theorem

To address the solution of Problem (1), one can assume a *stability condition*, also called an *inf-sup condition*.

Definition 2 Let $a(\cdot, \cdot)$ be a continuous sesquilinear form over $V \times W$. It verifies a *stability condition* if

$$\exists \alpha' > 0,\ \forall v \in V,\ \sup_{w \in W\setminus\{0\}} \frac{|a(v, w)|}{\|w\|_W} \geq \alpha'\|v\|_V. \tag{3}$$

Let us now introduce an *a priori* intermediate condition (cf. [3]).

Definition 3 Let $a(\cdot, \cdot)$ be a continuous sesquilinear form over $V \times W$. It is T-*coercive* if

$$\exists \mathtt{T} \in \mathcal{L}(V, W),\ \text{bijective},\ \exists \underline{\alpha} > 0,\ \forall v \in V,\ |a(v, \mathtt{T}v)| \geq \underline{\alpha}\|v\|_V^2. \tag{4}$$

One checks easily that the operator T realizes the inf-sup condition: in (3), for any v in $V \setminus \{0\}$, take $w = \mathtt{T}v \neq 0$.

Theorem 1 (Well-posedness) *Let $a(\cdot, \cdot)$ be a continuous and sesquilinear form. Then the four assertions below are equivalent:*

- *(i) the Problem (1) is well-posed;*
- *(ii) the form a satisfies a stability condition and $R(A) = W'$;*
- *(iii) the form a satisfies a stability condition and the only element $w \in W$ which satisfies $a(v, w) = 0$ for all $v \in V$ is $w = 0$;*
- *(iv) the form a is* T*-coercive.*

Proof The equivalence between the first three assertions is very standard (see theorem 2.6 in [12] and the references therein).

(iv) $\Longrightarrow$ (i): let $\mathtt{T}$ be an isomorphism of $\mathcal{L}(V,W)$ such that $(v,v') \mapsto a(v,\mathtt{T}v')$ is coercive on $V \times V$. Since this form is also sesquilinear and continuous, according to the Lax-Milgram theorem, there exists one, and only one $u \in V$ such that for all $v' \in V$, $a(u,\mathtt{T}v') = \langle f, \mathtt{T}v' \rangle$. Furthermore, since $\mathtt{T}$ is bijective, one remarks that, for all $w \in W$, there holds $a(u,w) = \langle f, w \rangle$, which yields well-posedness of (1).

(i) $\Longrightarrow$ (iv): consider $I_{W' \to W} \in \mathcal{L}(W',W)$ the Riesz bijection, defined by $(I_{W' \to W} w', w)_W = \langle w', w \rangle$, $\forall (w,w') \in W \times W'$. Due to (i), $\mathtt{T} := I_{W' \to W} \circ A$ is a bijective mapping of $\mathcal{L}(V,W)$: $\mathtt{T}^{-1} \in \mathcal{L}(W,V)$ and so $\|v\|_V \leq |||\mathtt{T}^{-1}||| \, \|\mathtt{T}v\|_W$, $\forall v \in V$. We remark that the form a is $\mathtt{T}$-coercive. Indeed, given $v \in V$, we have $a(v,\mathtt{T}v) = \langle Av, \mathtt{T}v \rangle = (I_{W' \to W} \circ Av, \mathtt{T}v)_W = \|\mathtt{T}v\|_W^2 \geq \|v\|_V^2 / |||\mathtt{T}^{-1}|||^2$. □

Remark 1 Assume that $W = V$, then coerciveness of a sesquilinear form implies a stability condition on the same form. Moreover, in this case, a sesquilinear form is coercive if, and only if, it is $\mathtt{I}_V$-coercive.

Remark 2 Assume that $W = V$.
If the form a is *hermitian*, that is if $a(v,w) = \overline{a(w,v)}$ for all $v, w \in V$, the stability condition (3) is *sufficient* to ensure well-posedness.
In the same spirit, for a *hermitian* form a, Definition 3 can be simplified to: $a(\cdot,\cdot)$ is $\mathtt{T}$-*coercive* if

$$\exists \mathtt{T} \in \mathcal{L}(V), \ \exists \underline{\alpha} > 0, \ \forall v \in V, \ |a(v,\mathtt{T}v)| \geq \underline{\alpha} \|v\|_V^2.$$

In other words, the fact that $\mathtt{T}$ be bijective is not required. Indeed, the previous condition implies that $\mathtt{T}$ is injective. Moreover, for all $v \in V \setminus \{0\}$, one has

$$\frac{|a(v,\mathtt{T}v)|}{\|\mathtt{T}v\|_W} \geq \underline{\alpha} \frac{\|v\|_V}{\|\mathtt{T}v\|_W} \|v\|_V \geq \frac{\underline{\alpha}}{|||\mathtt{T}|||} \|v\|_V.$$

Hence condition (3) holds.

To summarize, in the case $W = V$, the Lax-Milgram theorem gives a sufficient condition to ensure well-posedness of Problem (1), whereas theorem 1 provides a necessary and sufficient condition to ensure well-posedness of Problem (1), which writes:

- either the form a is stable and $R(A) = V'$,
- or the form a is $\mathtt{T}$-coercive.

2.3 Approximation of the solution to Problem (1)

Let us turn our attention to the approximation of the solution to Problem (1), which we assume to be well-posed. According to theorem 1, there exists an

operator $\mathtt{T} \in \mathcal{L}(V, W)$ such that the form a is T-coercive. To approximate this Problem, we let $(V_h)_h$ and $(W_h)_h$ be two infinite sequences of finite dimensional vector spaces. The parameter h takes strictly positive values, and it is destined to go to 0: if $n(h)$ denotes the dimension of V_h, then one has $\lim_{h\to 0} n(h) = +\infty$, so that V_h can "approximate" V. This also holds for the sequence of spaces $(W_h)_h$. When, for all h, $V_h \subset V$ and $W_h \subset W$, the approximation is a *conforming* approximation. In the sequel, we will always make this assumption. For an example of non-conforming approximation, see [7].

2.3.1 Natural discretization

The natural discretization of problem (1) writes

$$\begin{cases} \textit{Find } u_h \in V_h \textit{ such that} \\ \forall w_h \in W_h,\ a_h(u_h, w_h) = \langle f_h, w_h \rangle, \end{cases} \tag{5}$$

with discrete forms a_h and f_h (possibly) different respectively from a and f. In operator form, it writes

$$\begin{cases} \textit{Find } u_h \in V_h \textit{ such that} \\ A_h u_h = f_h \text{ in } (W_h)', \end{cases} \tag{6}$$

with $A_h \in \mathcal{L}(V_h, (W_h)')$ defined by $\langle A_h v_h, w_h \rangle = a_h(v_h, w_h)$ for all $(v_h, w_h) \in V_h \times W_h$.
Below, we address the well-posedness of the approximate Problems (5) and we propose error estimates. To be able to solve (5) with uniqueness, a necessary condition is $\dim V_h = \dim W_h$: we make this assumption from now on.

Definition 4 The family of sesquilinear forms $(a_h)_h$ is said to be *uniformly $V_h \times W_h$-stable* if

$$\exists \alpha_\dagger > 0,\ \forall h > 0,\ \forall v_h \in V_h,\quad \sup_{w_h \in W_h \setminus \{0\}} \frac{|a_h(v_h, w_h)|}{\|w_h\|_W} \geq \alpha_\dagger \|v_h\|_V. \tag{7}$$

As for the continuous problem, we give an *a priori* intermediate condition to (7).

Definition 5 The family of sesquilinear forms $(a_h)_h$ is said to be *uniformly $\mathtt{T}_h$-coercive* if

$$\begin{aligned} \exists \alpha^\star, \beta^\star > 0,\ \forall h > 0,\ \exists \mathtt{T}_h \in \mathcal{L}(V_h, W_h),\ \forall v_h \in V_h, \\ |a_h(v_h, \mathtt{T}_h v_h)| \geq \alpha^\star \|v_h\|_V^2 \text{ and } |||\mathtt{T}_h||| \leq \beta^\star. \end{aligned} \tag{8}$$

Next, introduce, for any $h > 0$ and any $v_h \in V_h$,

$$Cons_{f,h} = \sup_{w_h \in W_h \setminus \{0\}} \frac{|\langle f - f_h, w_h \rangle|}{\|w_h\|_V}, \tag{9}$$

$$Cons_{a,h}(v_h) = \sup_{w_h \in W_h \setminus \{0\}} \frac{|(a - a_h)(v_h, w_h)|}{\|w_h\|_V}. \tag{10}$$

These are consistency terms, in the sense that they express the discrepancies between the exact forms (a and f) and approximate forms (resp. a_h and f_h). One can obtain an error estimate including these *consistency terms*. In $V_h \times W_h$, one can apply theorem 1 to prove that Problem (5) is well-posed. When $a_h = a$ for all $h > 0$, classically, one uses the Fortin lemma (see [5,12]) to prove that the family $(a_h)_h$ is uniformly $V_h \times W_h$-stable and to derive error estimates. With our notations, this lemma states: a is uniformly $V_h \times W_h$-stable if, and only if, there is $\beta' > 0$ such that, for all $v \in V$, there is $\Pi_h(v) \in V_h$ such that

$$\forall w_h \in W_h,\ a(\Pi_h(v), w_h) = a(v, w_h) \text{ and } \|\Pi_h(v)\|_V \leq \beta' \|v\|_V.$$

Below, we propose an alternate approach to prove that the family $(a_h)_h$ is uniformly $V_h \times W_h$-stable, based once more on $\mathtt{T}$-coercivity theory.

Theorem 2 (`Well-posedness of the discrete problems`) *Assume that $\dim V_h = \dim W_h$, and that the sesquilinear forms $(a_h)_h$ are uniformly bounded. Then the three assertions below are equivalent:*

(i) *the Problem (5) is well-posed and $(A_h^{-1})_h$ is uniformly bounded;*
(ii) *the family $(a_h)_h$ is uniformly $V_h \times W_h$-stable;*
(iii) *the family $(a_h)_h$ is uniformly $\mathtt{T}_h$-coercive.*

Moreover, if these conditions are satisfied, the error $\|u - u_h\|_V$ is bounded by

$$\|u - u_h\|_V \leq C \inf_{v_h \in V_h} \left(\|u - v_h\|_V + Cons_{f,h} + Cons_{a,h}(v_h) \right), \qquad (11)$$

with $C := \max\left(\frac{1}{\alpha_\dagger}, \frac{|||a|||}{\alpha_\dagger} + 1\right) > 0$ independent of h.

Proof (i) $\Longrightarrow$ (iii): Define $\mathtt{T}_h := I_{W_h' \to W_h} \circ A_h$, where $I_{W_h' \to W_h}$ is the Riesz bijection. One has $|||\mathtt{T}_h||| \leq |||A_h|||$: as the forms $(a_h)_h$ are uniformly bounded, so are the operators $(\mathtt{T}_h)_h$. Due to (i), $\mathtt{T}_h$ is a bijective mapping and moreover $\mathtt{T}_h^{-1} = A_h^{-1} \circ I_{W_h \to W_h'}$ is such that $|||\mathtt{T}_h^{-1}||| \leq \max_h |||A_h^{-1}||| =: C_1 < \infty$. Given $v_h \in V_h$, we find $a_h(v_h, \mathtt{T}_h v_h) = \|\mathtt{T}_h v_h\|_W^2 \geq \|v_h\|_V^2 / C_1^2$. Hence $(a_h)_h$ is uniformly $\mathtt{T}_h$-coercive.

(iii) $\Longrightarrow$ (ii): for $v_h \in V_h$, one has

$$\sup_{w_h \in W_h \setminus \{0\}} \frac{|a_h(v_h, w_h)|}{\|w_h\|_W} \geq \frac{|a_h(v_h, \mathtt{T}_h v_h)|}{\|\mathtt{T}_h v_h\|_W} \geq \alpha^\star \frac{\|v_h\|_V^2}{\|\mathtt{T}_h v_h\|_W} \geq \frac{\alpha^\star}{\beta^\star} \|v_h\|_V.$$

Thus, $(a_h)_h$ is uniformly $V_h \times W_h$-stable.

(ii) $\Longrightarrow$ (i): According to theorem 1, if the family $(a_h)_h$ is uniformly $V_h \times W_h$-stable, Problem (5) is well-posed. Moreover, A_h^{-1} is uniformly bounded. Indeed, $\|A_h^{-1} f\| \leq \|f\| / \alpha_\dagger$.

Now, let us focus on the error estimation. By assumption, (7) holds for some $\alpha_\dagger > 0$. Given any $v_h \in V_h$, there exists $w_h \in W_h$ such that

$$\alpha_\dagger \|u_h - v_h\|_V \|w_h\|_V \leq |a_h(u_h - v_h, w_h)|, \text{ and one can check that}$$
$$a_h(u_h - v_h, w_h) = \langle f_h - f, w_h \rangle + a(u - v_h, w_h) + (a - a_h)(v_h, w_h).$$

It follows that

$$\|u_h - v_h\|_V \leq \frac{1}{\alpha_\dagger}(Cons_{f,h} + |||a||| \, \|u - v_h\|_V + Cons_{a,h}(v_h)),$$

which leads to (11), since $\|u - u_h\|_V \leq \|u - v_h\|_V + \|u_h - v_h\|_V$.

Corollary 1 *Assume there exists an isomorphism $\mathtt{T} \in \mathcal{L}(V, W)$ such that $(v, v') \mapsto a(v, \mathtt{T}v')$ is coercive on $V \times V$. Assume also that $\mathtt{T}V_h \subset W_h$ for all h and $\lim_{h\to 0} |||a_h - a||| = 0$. Then, the family $(a_h)_h$ is uniformly $\mathtt{T}_h$-coercive for h sufficiently small so estimate (11) holds true.*

Proof Indeed one has, with $\mathtt{T}_h = \mathtt{T}_{|V_h}$,

$$\begin{aligned} |a_h(v_h, \mathtt{T}_h v_h)| &= |a(v_h, \mathtt{T}v_h) - (a_h - a)(v_h, \mathtt{T}v_h)| \\ &\geq (\underline{\alpha} - |||a_h - a||| \, |||T|||)\|v_h\|_V^2. \end{aligned}$$

One takes h_0 small enough so that $|||a_h - a||| \, |||T||| < \underline{\alpha}$ for all $h \in (0; h_0]$.

Remark 3 When one is using T-coercivity to solve discrete problems, the assumption $\mathtt{T}V_h \subset W_h$ for all h can be relaxed. See §4.3 below, or [8].

2.3.2 Discretization of the coercive form

We remark that the form $\tilde{a} \ : \ (v, v') \mapsto a(v, \mathtt{T}v')$ is sesquilinear, continuous and coercive over $V \times V$. Therefore, provided that the operator $\mathtt{T}$ is *explicitly known*[1], instead of solving Problem (1) directly, one can solve the *equivalent Problem*

$$\begin{cases} \text{Find } u \in V \text{ such that} \\ \forall v \in V, \ \tilde{a}(u, v) = \langle \tilde{f}, v \rangle \, , \end{cases} \tag{12}$$

where $\tilde{f} \in V'$ is defined by $v \mapsto \langle f, \mathtt{T}v \rangle$. Indeed, given a subspace V_h of V, one solves the approximate Problem

$$\begin{cases} \text{Find } u_h \in V_h \text{ such that} \\ \forall v_h \in V_h, \ \tilde{a}_h(u_h, v_h) = \langle \tilde{f}_h, v_h \rangle. \end{cases} \tag{13}$$

Above, the forms are respectively defined by

$$\forall v_h, w_h \in V_h, \ \tilde{a}_h(v_h, w_h) = a_h(v_h, \mathtt{T}w_h), \ \langle \tilde{f}_h, w_h \rangle = \langle f_h, \mathtt{T}w_h \rangle.$$

[1] By "$\mathtt{T}$ is *explicitly known*", it is understood that the action of $\mathtt{T}$ over elements $v_h \in V_h$ can be computed easily.

Then, one can use Céa's lemma (if $a_h = a_{|V_h \times (\mathtt{T}V_h)}$, $f_h = f_{|(\mathtt{T}V_h)}$)) or more generally the first Strang's lemma to obtain error estimates, which write

$$\|u - u_h\|_V \leq C \inf_{v_h \in V_h} \left\{ \|u - v_h\|_V + Cons_{\tilde{f},h} + Cons_{\tilde{a},h}(v_h) \right\}. \tag{14}$$

Above, $C > 0$ is independent of h and the data f. The consistency terms are respectively defined, for any h and any $v_h \in V_h$, by

$$Cons_{\tilde{f},h} = \sup_{w_h \in V_h \setminus \{0\}} \frac{|\langle \tilde{f} - \tilde{f}_h, w_h \rangle_V|}{\|w_h\|_V}, \; Cons_{\tilde{a},h}(v_h) = \sup_{w_h \in V_h \setminus \{0\}} \frac{|(\tilde{a} - \tilde{a}_h)(v_h, w_h)|}{\|w_h\|_V}.$$

Remark 4 In this simple case, note that one automatically approximates Problem (1) in $V_h \times (\mathtt{T}V_h)$.

2.3.3 Comparison between the two methods of approximation

From a practical point of view, there is a fundamental difference between what we call the "natural discretization" and the discretization of the coercive form. Indeed, for the natural discretization, the isomorphism $\mathtt{T}$ is just a theoretical tool and its action is not implemented. In the contrary, the discretization of the coercive form requires the discretization of $\mathtt{T}$. The advantage of this latter approach is that the convergence of the method is easily proved.

3 Application to $\mathrm{div}\,\sigma\nabla\cdot$: study of the continuous problem

3.1 Notations

For the ease of exposition, Ω will be a bounded domain of $\mathbb{R}^2$ with $\overline{\Omega} = \overline{\Omega_1} \cup \overline{\Omega_2}$, where Ω_1 and Ω_2 are two domains such that $\Omega_1 \cap \Omega_2 = \emptyset$. For extensions to $3D$ polyhedral domains, see [1]. We suppose that the boundaries $\partial\Omega$, $\partial\Omega_1$ and $\partial\Omega_2$ are (connected) polygons. The interface separating the two domains is called $\Sigma := \overline{\Omega_1} \cap \overline{\Omega_2}$. Last the boundaries $\partial\Omega_k$, $k = 1, 2$ are split as $\partial\Omega_k = \Gamma_k \cup \Sigma$, with $\Gamma_k := \partial\Omega \cap \partial\Omega_k$.
In short, if $\mathcal{O}$ is an open subset of $\mathbb{R}^2$, we denote $(\cdot,\cdot)_{\mathcal{O}}$ the scalar products of $L^2(\mathcal{O})$ and $(L^2(\mathcal{O}))^2$, and $\|\cdot\|_{\mathcal{O}}$ the associated norms. Let us define our background by making the following assumptions:

$$\begin{cases} V = W := H_0^1(\Omega) \text{ with norm } \|v\|_V := \|\nabla v\|_\Omega \text{ and } V' = H^{-1}(\Omega), \\ k = 1, 2, \quad V_k := \{v_{|\Omega_k} \,|\, v \in H_0^1(\Omega)\} \text{ with semi-norm } \|v\|_{V_k} := \|\nabla v\|_{\Omega_k}; \end{cases}$$

$$\begin{cases} \forall v, w \in H_0^1(\Omega), \; a(v,w) := (\sigma\nabla v, \nabla w)_\Omega, \\ \sigma_1 := \sigma_{|\Omega_1} \text{ is a constant such that } \sigma_1 > 0, \\ \sigma_2 := \sigma_{|\Omega_2} \text{ is a constant such that } \sigma_2 < 0. \end{cases}$$

The problem we address is

$$\begin{cases} \textit{Find } u \in H_0^1(\Omega) \textit{ s.t.} \\ \forall w \in H_0^1(\Omega), \; (\sigma\nabla u, \nabla w)_\Omega = \langle f, w \rangle \end{cases} \Leftrightarrow \begin{cases} \textit{Find } u \in H_0^1(\Omega) \textit{ s.t.} \\ -\operatorname{div}\sigma\nabla u = f \text{ in } H^{-1}(\Omega) \end{cases}. \tag{15}$$

Remark 5 Classically, Problem (15) is well-posed if σ_2/σ_1 belongs to $\mathbb{C} \setminus \mathbb{R}_-$.

Definition 6 The ratio $\kappa_\sigma := \sigma_2/\sigma_1$ is called the *contrast*.

Under the above assumptions, given any $v \in H_0^1(\Omega)$, we can use the notations $v_k = v_{|\Omega_k}$, $k = 1, 2$, without confusion.
The geometry of the interface, and more precisely, the value of the angles of the corners, plays a major role in the nature of Problem (15). Below, we precise this geometry focusing on the corners and endpoints:

- The set of *interior vertices* (if it is not empty) of the interface Σ is denoted $\mathcal{S}_{int} := \{\boldsymbol{x}^i,\ 1 \leq i \leq N_{int}\}$. And, for $1 \leq i \leq N_{int}$, the subsets Ω_1 and Ω_2 coincide with open cones in a neighbourhood $\mathcal{V}^i$ of $\boldsymbol{x}^i$, locally in Ω:
$$\begin{array}{l} \Omega_1 \cap \mathcal{V}^i = \mathcal{K}_1(\boldsymbol{x}^i) \cap \mathcal{V}^i \ \text{ and } \ \Omega_2 \cap \mathcal{V}^i = \mathcal{K}_2(\boldsymbol{x}^i) \cap \mathcal{V}^i, \\ \text{where } \mathcal{K}_1(\boldsymbol{x}^i) \text{ and } \mathcal{K}_2(\boldsymbol{x}^i) \text{ are open cones, centered at } \boldsymbol{x}^i. \end{array} \tag{16}$$
- There are either 0 or 2 endpoints, called *exterior vertices*: $\mathcal{S}_{ext} := \Sigma \cap \partial\Omega = \{\boldsymbol{x}^i,\ N_{int}+1 \leq i \leq N_{int}+N_{ext}\}$, with $N_{ext} \in \{0, 2\}$. And, for $N_{int}+1 \leq i \leq N_{int}+N_{ext}$, the subsets Ω_1 and Ω_2 coincide with open cones in a neighbourhood $\mathcal{V}^i$ of $\boldsymbol{x}^i$, locally in Ω: i.e., (16) holds.

For each index i, we define the apertures $\alpha_k^i \in (0; 2\pi)$ of the cones $\mathcal{K}_k(\boldsymbol{x}^i)$, $k = 1, 2$. We introduce $\gamma^i := \alpha_1^i + \alpha_2^i$ and $\alpha^i := \min(\alpha_1^i,\, \alpha_2^i)$. Evidently, one has $\gamma^i = 2\pi$ for interior vertices, and $\gamma^i < 2\pi$ for boundary vertices. On the other hand, at an interior vertex $\boldsymbol{x}^i$, Σ has a corner, so $0 < \alpha^i < \pi$.
We let $\mathcal{S}_{ext}^1 := \{\boldsymbol{x}^i \in \mathcal{S}_{ext} \,|\, \alpha_1^i \leq \alpha_2^i\}$ and $\mathcal{S}_{ext}^2 := \{\boldsymbol{x}^i \in \mathcal{S}_{ext} \,|\, \alpha_2^i < \alpha_1^i\}$. The

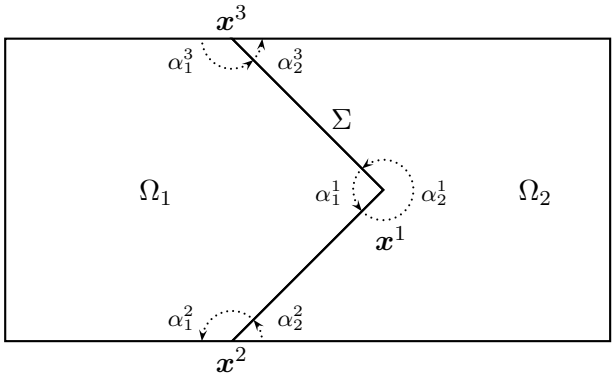

Fig. 1 A sample geometry: $\mathcal{S}_{int} = \{\boldsymbol{x}^1\}$, $\mathcal{S}_{ext}^1 = \emptyset$, $\mathcal{S}_{ext}^2 = \{\boldsymbol{x}^2, \boldsymbol{x}^3\}$.

cardinality of $\mathcal{S}_{int} \cup \mathcal{S}_{ext}$ is denoted by N. Finally, we define
$$I_{\alpha^i} := \frac{\gamma^i - \alpha^i}{\alpha^i} \text{ for } 1 \leq i \leq N.$$

Remark 6 Given any interior vertex, there holds $I_{\alpha^i} > 1$. The same is true for any boundary vertex of $\mathcal{S}_{ext}^2$. On the other hand, for a boundary vertex of $\mathcal{S}_{ext}^1$, one has only $I_{\alpha^i} \geq 1$ (it can happen that $I_{\alpha^i} = 1$).

3.2 Study of the continuous problem

We follow [1] in this subsection. We recall the definition below [18].

Definition 7 An operator $A \in \mathcal{L}(V,W)$ is Fredholm when $\dim(\ker(A)) < \infty$ and $\dim(W/R(A)) < \infty$. When the operator A is Fredholm, its index is equal to $\dim(\ker(A)) - \dim(W/R(A))$.

First, we state a result whose proof relies on localized $\mathtt{T}$-coercivity. Define

$$\hat{R}_\Sigma := \max\left(\max_{\boldsymbol{x}^i \in \mathcal{S}_{int} \cup \mathcal{S}^1_{ext}} I_{\alpha^i}, 1\right), \quad \check{R}_\Sigma := \max\left(\max_{\boldsymbol{x}^i \in \mathcal{S}_{int} \cup \mathcal{S}^2_{ext}} I_{\alpha^i}, 1\right).$$

There holds the

Theorem 3 (CONSTANT COEFFICIENTS) *Assume that the contrast satisfies* $\kappa_\sigma \in (-\infty, 0) \setminus \left[-\hat{R}_\Sigma; -1/\check{R}_\Sigma\right]$. *Then, the operator* $A : u \mapsto -div(\sigma \nabla u)$, *from* $V = H^1_0(\Omega)$ *to* $V' = H^{-1}(\Omega)$, *is Fredholm of index* 0.

Remark 7 In particular, under the assumption of theorem 3, the Problem (15) is well-posed if and only if A is injective. In this case, A is an isomorphism from $H^1_0(\Omega)$ to $H^{-1}(\Omega)$. Still under the assumption of theorem 3, when A is not injective, $\ker(A)$ is of finite dimension so one can write $\ker(A) = \mathrm{span}(\varphi_1, \ldots, \varphi_p)$, for some finite $p \geq 1$. Then Problem (15) has a solution (unique up to a linear combination of the $\varphi_1, \ldots, \varphi_p$) if, and only if, the source term satisfies the compatibility conditions $\langle f, \varphi_k \rangle = 0$ for $k = 1 \ldots p$ (see theorem 2.27 in [18]).

Remark 8 Let us underline that, if the assumption of theorem 3 is not met, there are situations for which (15) is ill-posed in the sense that A is no longer Fredholm (see [4,22,1,2] for more details). Actually, A is never Fredholm if $\kappa_\sigma \in \left(-\hat{R}_\Sigma; -1/\check{R}_\Sigma\right)$.

Now, we prove a result, with a stronger assumption on κ_σ, to assert that A is an isomorphism from $H^1_0(\Omega)$ to $H^{-1}(\Omega)$ (that is to assert that A is Fredholm of index 0 *and* injective).
To obtain some practical results, consider an operator $R_1 \in \mathcal{R}_1$, where $\mathcal{R}_1$ is defined by

$$\mathcal{R}_1 := \{R_1 \in \mathcal{L}(V_1, V_2) \,|\, R_1 v_{1|\Sigma} = v_{1|\Sigma},\ \forall v_1 \in V_1\}.$$

Here, the notation $\cdot_{|\Sigma}$ refers to the trace operator on Σ. With this operator R_1, define $\mathtt{T}$ acting on elements of V as below. For all $v \in V$, let

$$\mathtt{T}v := \begin{cases} v_1 & \text{in } \Omega_1 \\ -v_2 + 2R_1 v_1 & \text{in } \Omega_2 \end{cases}. \tag{17}$$

Since R_1 fulfills the required matching condition on the interface, we check that $\mathtt{T}v \in V$, and $\mathtt{T} \in \mathcal{L}(V)$. Furthermore, one finds that $\mathtt{T} \circ \mathtt{T} = \mathtt{I}_V$. Indeed,

$$(\mathtt{T} \circ \mathtt{T})v = \begin{cases} (\mathtt{T}v)_1 = v_1 & \text{in } \Omega_1 \\ -(\mathtt{T}v)_2 + 2R_1(\mathtt{T}v)_1 = -(-v_2 + 2R_1 v_1) + 2R_1 v_1 = v_2 & \text{in } \Omega_2 \end{cases}.$$

It follows that T is a bijection. Let us now perform a study of the T-coercivity of $a(\cdot,\cdot)$, namely whether the conditions in Definition 3 can be met. Let $v \in V$, and $\eta > 0$:

$$\begin{array}{l} |a(v,\mathtt{T}v)| \\ = |(\sigma_1\nabla v_1, \nabla v_1)_{\Omega_1} + (|\sigma_2|\nabla v_2, \nabla v_2)_{\Omega_2} - 2(|\sigma_2|\nabla v_2, \nabla(R_1 v_1))_{\Omega_2}| \\ \geq \sigma_1\|v_1\|^2_{V_1} + |\sigma_2|\|v_2\|^2_{V_2} - 2|(|\sigma_2|\nabla v_2, \nabla(R_1 v_1))_{\Omega_2}| \\ \geq \sigma_1\|v_1\|^2_{V_1} + |\sigma_2|\|v_2\|^2_{V_2} - \eta|\sigma_2|\|v_2\|^2_{V_2} - \eta^{-1}|\sigma_2|\|R_1 v_1\|^2_{V_2} \\ \geq (\sigma_1 - \eta^{-1}|\sigma_2|\,|||R_1|||^2)\|v_1\|^2_{V_1} + |\sigma_2|(1-\eta)\|v_2\|^2_{V_2} \\ \geq \min\left((\sigma_1 - \eta^{-1}|\sigma_2|\,|||R_1|||^2), |\sigma_2|(1-\eta)\right)\|v\|^2_V. \end{array} \tag{18}$$

Above, we used Young's inequality or, more precisely, its generalization to a positive hermitian form to bound $-2|a_2(v_2, R_1 v_1)|$ from below. Suppose $\sigma_1/|\sigma_2| > |||R_1|||^2$. Taking η such that $|\sigma_2|\,|||R_1|||^2/\sigma_1 < \eta < 1$, we derive T-coercivity for the form. This condition might be *optimized* minimizing the norm of $R_1 \in \mathcal{R}_1$. More precisely, one derives T-coercivity as soon as $\sigma_1/|\sigma_2| > \left(\inf_{R_1\in\mathcal{R}_1} |||R_1|||^2\right)$.
It is also possible to choose an operator $R_2 \in \mathcal{R}_2$, , where $\mathcal{R}_2$ is now defined by

$$\mathcal{R}_2 := \{R_2 \in \mathcal{L}(V_2, V_1)\,|\, R_2 v_{2|\Sigma} = v_{2|\Sigma},\ \forall v_2 \in V_2\},$$

and then to define T as below: for all $v \in V$, let

$$\mathtt{T}v := \begin{cases} v_1 - 2R_2 v_2 & \text{in } \Omega_1 \\ -v_2 & \text{in } \Omega_2 \end{cases}. \tag{19}$$

In this case, we derive T-coercivity as soon as $|\sigma_2|/\sigma_1 > \left(\inf_{R_2\in\mathcal{R}_2} |||R_2|||^2\right)$.
Let us summarize these results with the

Theorem 4 *Assume that the contrast $\kappa_\sigma \in (-\infty, 0)$ satisfies $\kappa_\sigma < -\left(\inf_{R_2\in\mathcal{R}_2} |||R_2|||^2\right)$ or $\kappa_\sigma > -1/\left(\inf_{R_1\in\mathcal{R}_1} |||R_1|||^2\right)$. Then, the operator $A : u \mapsto -div(\sigma\nabla u)$ is an isomorphism from $V = H^1_0(\Omega)$ to $V' = H^{-1}(\Omega)$.*

3.3 Examples

◇ Example of the cavity. We illustrate below, in a practical case, the difference between the results provided by theorems 3 and 4.
Let us consider the cavity (see figure 2) defined by $\Omega := \{(x,y) \in (-a;b) \times (0;1)\}$, $\Omega_1 := (-a;0) \times (0;1)$ and $\Omega_2 := (0;b) \times (0;1)$ with $a > 0$ and $b > 0$. The interface Σ is then equal to the segment $\{0\} \times [0;1]$. Without loss of generality, we suppose $a \geq b$. One handles the case $a < b$ exchanging the roles of Ω_1 and Ω_2.

- According to theorem 3 (here $\mathcal{S}_{int} = \mathcal{S}^2_{ext} = \emptyset$), the operator A is Fredholm of index 0 as soon as $\kappa_\sigma = \sigma_2/\sigma_1 \neq -1$.

- When $\kappa_\sigma = -1$, the operator A is no longer Fredholm (see [1]). In particular, if $a = b$, the authors prove in [1] that $\dim(\ker(A)) = \infty$. Now, suppose that $a \neq b$. Let us prove that A is injective. Consider u an element of $\ker(A)$. Define $e := u_1 - u_2 \circ s$ on $(-b, 0) \times (0, 1)$ with $s(x, y) = (-x, y)$. This element e satisfies the following equations:

$$\Delta e = 0 \text{ in } (-b, 0) \times (0, 1);\ e = 0 \text{ on } \Sigma \text{ and } \partial_x e = 0 \text{ on } \Sigma.$$

  Remark that one has $\sigma_1 \partial_x u_1 = \sigma_2 \partial_x u_2$ on Σ so we can claim $\partial_x e = 0$ on Σ only because $\kappa_\sigma = -1$. The unique continuation principle ((see lemma 4.15 in [19] and the references therein)) implies $e = 0$ in $(-b, 0) \times (0, 1)$. Since $u_2 = 0$ on $\{b\} \times (0; 1)$, one finds $u_1 = 0$ on $\{-b\} \times (0; 1)$. Define $\tilde{\Omega} := (-a; -b) \times (0; 1)$. One notices that $\Delta u_1 = 0$ in $\tilde{\Omega}$ and $u_1 = 0$ on $\partial\tilde{\Omega}$. Consequently, $u_1 = 0$ on $\tilde{\Omega}$. According to the unique continuation principle, it yields $u = 0$ in Ω. Thus, when $\kappa_\sigma = -1$ and when $a \neq b$, A is injective. Since A is not Fredholm, it follows that $\dim\,(H^{-1}(\Omega)/R(A)) = \infty$.
- Let us study now in which cases A is an isomorphism. For that, introduce the operators

$$\begin{array}{l} R_1 : V_1 \to V_2 \\ \qquad v_1 \mapsto R_1 v_1 \text{ with } (R_1 v_1)(x, y) = v_1(-a\,x/b, y)\,; \end{array} \tag{20}$$

$$R_2 : V_2 \to V_1$$
$$v_2 \mapsto R_2 v_2 \text{ with } (R_2 v_2)(x, y) = \begin{cases} v_2(-x, y) & \text{if } -b \leq x \\ 0 & \text{else} \end{cases}\,; \tag{21}$$

$$\mathtt{T}_1 v = \begin{cases} v_1 & \text{in } \Omega_1 \\ -v_2 + 2R_1 v_1 & \text{in } \Omega_2 \end{cases};\ \mathtt{T}_2 v = \begin{cases} v_1 - 2R_2 v_2 & \text{in } \Omega_1 \\ -v_2 & \text{in } \Omega_2 \end{cases}. \tag{22}$$

  One has $R_1 \in \mathcal{R}_1$, $R_2 \in \mathcal{R}_2$, $|||R_1|||^2 = a/b$ and $|||R_2|||^2 = 1$. Consequently, according to theorem 4, A is an isomorphism from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$ as soon as $\kappa_\sigma \notin [-1; -b/a]$.
- For this particular geometry, one can study more precisely the question of the injectivity of A when $\kappa_\sigma \in (-1; -b/a]$ $(a \neq b)$. Let u be an element of $H_0^1(\Omega)$ such that $Au = 0$. The couple (u_1, u_2) satisfies the equations

$$\begin{array}{llll} \Delta u_1 = 0 & \text{in } \Omega_1; & u_1 - u_2 & = 0 \text{ on } \Sigma; \\ \Delta u_2 = 0 & \text{in } \Omega_2; & \sigma_1 \partial_x u_1 - \sigma_2 \partial_x u_2 & = 0 \text{ on } \Sigma. \end{array}$$

  Decomposing u_1 and u_2 in Fourier series (the family $\{y \mapsto \sin(n\pi y)\}_{n=1}^{\infty}$ is a basis of $L^2((0; 1))$), one obtains

$$\begin{array}{r} u_1(x, y) = \sum_{n=1}^{\infty} u_1^n \sinh(n\pi(x + a)) \sin(n\pi y) \\ \text{and } \ u_2(x, y) = \sum_{n=1}^{\infty} u_2^n \sinh(n\pi(x - b)) \sin(n\pi y), \end{array}$$

  where u_1^n and u_2^n are constants. Besides, the transmission conditions imply,

$$\forall n \in \mathbb{N}^*, \quad \left| \begin{array}{ll} u_1^n \sinh(n\pi a) & = -u_2^n \sinh(n\pi b) \\ u_1^n \sigma_1 \cosh(n\pi a) & = u_2^n \sigma_2 \cosh(n\pi b). \end{array} \right. \tag{23}$$

For each $n \in \mathbb{N}^*$, there exists a non trivial solution to the system (23) (in (u_1^n, u_2^n)) if and only if

$$\begin{aligned} &\sigma_2 \sinh(n\pi a)\cosh(n\pi b) + \sigma_1 \sinh(n\pi b)\cosh(n\pi a) = 0 \\ \Leftrightarrow\ &-\frac{\tanh(n\pi b)}{\tanh(n\pi a)} = \kappa_\sigma. \end{aligned}$$

Consequently, A is an isomorphism from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$ if and only if $\kappa_\sigma \notin \{-\tanh(n\pi b)/\tanh(n\pi a),\, n \in \mathbb{N}^*\} \cup \{-1\}$.

Remark 9 The map $g : z \mapsto -\dfrac{\tanh(z\pi b)}{\tanh(z\pi a)}$ is continuous, strictly decreasing on $\mathbb{R}_+$ and $g(1) = -\tanh(\pi b)/\tanh(\pi a) < -b/a$ whereas $\lim_{z\to+\infty} g(z) = -1$.

$\diamond$ Example of the interior corner. Let us consider now the geometry of figure 3. More precisely, define $\Omega := (-1;1)\times(-1;1)$, $\Omega_2 := (0;1)^2$ and $\Omega_1 := \Omega \setminus \overline{\Omega_2}$. According to theorem 3, the operator A is Fredholm of index 0 as soon as $\kappa_\sigma = \sigma_2/\sigma_1 \notin [-3;-1/3]$. As in [20], introduce the operators

$$\begin{aligned} &R_1 : V_1 \to V_2 \\ &\quad v_1 \mapsto R_1 v_1 \text{ with } (R_1 v_1)(x,y) = v_1(-x,y) + v_1(x,-y) - v_1(-x,-y)\,; \end{aligned} \tag{24}$$

$$\begin{aligned} &R_2 : V_2 \to V_1 \\ &\quad v_2 \mapsto R_2 v_2 \text{ with } (R_2 v_2)(x,y) = \begin{cases} v_2(-x,y) & \text{on } (-1;0)\times(0;1) \\ v_2(x,-y) & \text{on } (0;1)\times(-1;0) \\ v_2(-x,-y) & \text{on } (-1;0)^2 \end{cases}; \end{aligned} \tag{25}$$

$$\mathtt{T}_1 v = \begin{cases} v_1 & \text{in } \Omega_1 \\ -v_2 + 2R_1 v_1 & \text{in } \Omega_2 \end{cases};\ \mathtt{T}_2 v = \begin{cases} v_1 - 2R_2 v_2 & \text{in } \Omega_1 \\ -v_2 & \text{in } \Omega_2 \end{cases}. \tag{26}$$

One has $R_1 \in \mathcal{R}_1$, $R_2 \in \mathcal{R}_2$, $|||R_1|||^2 = 3$ and $|||R_2|||^2 = 3$. Consequently, according to theorem 4, A is actually an isomorphism from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$ as soon as $\kappa_\sigma = \sigma_2/\sigma_1 \notin [-3;-1/3]$. This matches the results obtained in [10, 4, 22].

3.4 Regularity of the solution

Up to the end of this document, we suppose that Problem (15) is well-posed and we consider the case of an L^2 source term. So, we focus on the problem

$$\begin{cases} \textit{Find } u \in H_0^1(\Omega) \textit{ such that} \\ \forall w \in H_0^1(\Omega),\ a(u,w) = (f,w)_\Omega. \end{cases} \tag{27}$$

Let us start by recalling some results on the regularity of the solution $u \in H_0^1(\Omega)$ to problem (27). Classically (see chapter 2, volume 1 of [16], theorem 2.1.3 of [14] and, for the study around exterior corners, theorem 2.1.4 of [14]), the following interior regularity result holds.

Proposition 1 *Let $\mathcal{O}$ be an open subset of Ω such that $\overline{\mathcal{O}}$ does not intersect the interface Σ. Then the solution u to problem (27) belongs to $H^{1+s}(\mathcal{O})$, with estimate*

$$\|u\|_{H^{1+s}(\mathcal{O})} \leq C \|f\|_{\Omega},$$

where the constant C is independent of f, and $s \in (0;1]$ only depends on the aperture of the corners located on the boundary[2]*.*

Around the interface, the operator $v \mapsto \operatorname{div} \sigma \nabla v$ is no longer elliptic and the regularity results are less classical. However, usual techniques based on Fourier and Mellin transforms still apply (see [10,4,22,6]). In particular, in the neighbourhood of the smooth part of the interface, one can prove that u is locally H^2 on each side of Σ (see also [9] for methods based on integral representation). More precisely, one has the

Proposition 2 *Assume that $\kappa_\sigma = \sigma_2/\sigma_1 \neq -1$ and consider an open subset $\mathcal{O}$ of Ω such that $\overline{\mathcal{O}} \subset \Omega$ and $\overline{\mathcal{O}}$ does not meet any of the corners of Σ. Then the solution u to problem (27) is such that $u_k \in H^2(\mathcal{O} \cap \Omega_k)$, $k = 1, 2$, with the estimate*

$$\|u_1\|_{H^2(\mathcal{O}\cap\Omega_1)} + \|u_2\|_{H^2(\mathcal{O}\cap\Omega_2)} \leq C \|f\|_{\Omega}.$$

In the neighbourhood of the corners of Σ, the regularity of u depends both on the geometry and on the value of the contrast. To sum up, there exists $s \in (0;1]$ such that $u_k \in H^{1+s}(\mathcal{O} \cap \Omega_k)$, $k = 1, 2$, with the estimate

$$\|u_1\|_{H^{1+s}(\mathcal{O}\cap\Omega_1)} + \|u_2\|_{H^{1+s}(\mathcal{O}\cap\Omega_2)} \leq C \|f\|_{\Omega}.$$

It is important to note that s can be arbitrary small, depending on the contrast and on the geometry of the interface.

4 Application to $\operatorname{div} \sigma \nabla \cdot$: approximation of the solution with hypothesis on the mesh

Below, we present a simple approximation of Problem (15), based on P_1 Lagrange Finite Elements, and we derive error estimates. It is understood that one could use mesh refinement and/or higher order Finite Elements to improve the error estimates.

4.1 Approximability

Let us consider $(\mathcal{T}_h)_h$ a regular family of meshes of $\overline{\Omega}$, made of triangles. Moreover, for all partitions of $\overline{\Omega}$ and for all triangles τ, one has either $\tau \subset \overline{\Omega_1}$ or $\tau \subset \overline{\Omega_2}$.
Define the family of finite element spaces

$$V_h := \left\{ v \in H^1_0(\Omega) \,|\, v_{|\tau} \in \mathbb{P}_1(\tau),\ \forall \tau \in \mathcal{T}_h \right\},$$

[2] If Ω is convex or if $\overline{\mathcal{O}}$ does not meet any of the corners of $\partial\Omega$, one can take $s = 1$.

where $\mathbb{P}_1(\tau)$ is the space of polynomials of degree at most 1 on the triangle τ. Let us consider the family of problems (indexed by h)

$$\begin{cases} \textit{Find } u_h \in V_h \textit{ such that} \\ \forall w_h \in V_h,\ a(u_h, w_h) = (f, w_h)_\Omega. \end{cases} \tag{28}$$

Definition 8 The sequence $(V_h)_h$ fulfills the *basic approximability property* if

$$\forall v \in H_0^1(\Omega),\ \lim_{h\to 0}\left(\inf_{v_h\in V_h} \|v - v_h\|_{H_0^1(\Omega)}\right) = 0.$$

Definition 9 Given $\mathtt{T} \in \mathcal{L}(H_0^1(\Omega))$, the meshes $(\mathcal{T}_h)_h$ are *T-conform* if $\mathtt{T}V_h \subset V_h$ for all h.

4.2 Numerical approximation: T-conform mesh

We would like to apply corollary 1 to derive error estimates. To that aim, we need T-coercivity with an isomorphism $\mathtt{T}$ such that $\mathtt{T}V_h \subset V_h$ for all h.

$\diamond$ Example of the cavity. We consider here the geometry of figure 2: $\Omega := \{(x, y) \in (-2; 1) \times (0; 1)\}$, $\Omega_1 := (-2; 0) \times (0; 1)$ and $\Omega_2 := (0; 1) \times (0; 1)$, where the meshes are symmetric with respect to $\Sigma := \{0\} \times [0; 1]$. Suppose first $\kappa_\sigma < -1$. The operator $\mathtt{T}_2$ defined in (22) is such that $\mathtt{T}_2 V_h \subset V_h$. Consequently, according to corollary 1, Problem (28) is well-posed for each $h > 0$. Moreover, one has the error estimate

$$\|u - u_h\|_{H_0^1(\Omega)} \leq Ch\|f\|_\Omega,$$

because, in this situation, u is of H^2 regularity on both sides of the interface. The same result can be obtained when $-1/2 < \kappa_\sigma < 0$ using the obvious *ad hoc* mesh, using this time $\mathtt{T}_1$.
However, we are not able to conclude when $\kappa_\sigma \in (-1; -1/2]\backslash\{-\tanh(n\pi)/\tanh(2n\pi),\ n \in \mathbb{N}^*\}$ because we do not have at our disposal an *explicit* operator $\mathtt{T}$ such that a is T-coercive.
$\diamond$ Example of the interior corner. Here again, $\Omega := (-1; 1) \times (-1; 1)$, $\Omega_2 := (0; 1)^2$ and $\Omega_1 := \Omega \setminus \overline{\Omega_2}$. Working with the mesh of figure 3, one proves that Problem (28) is well-posed for each $h > 0$ as soon as $\kappa_\sigma = \sigma_2/\sigma_1 \notin [-3; -1/3]$. Moreover, one has the error estimate

$$\|u - u_h\|_{H_0^1(\Omega)} \leq Ch^s\|f\|_\Omega, \tag{29}$$

with $0 < s \leq 1$ which only depends on the contrast (because the angle of the corner has been fixed).

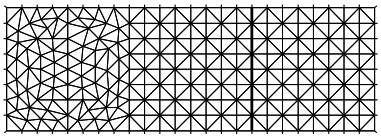

Fig. 2 "Symmetric" mesh for the cavity.

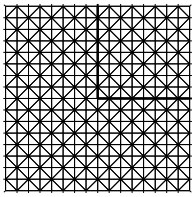

Fig. 3 "Symmetric" mesh for the corner.

4.3 Numerical approximation: locally T-conform mesh

In the preceding paragraph, we have been working with operators $\mathtt{T}$ of the form

$$\mathtt{T}_1 v = \begin{cases} v_1 & \text{in } \Omega_1 \\ -v_2 + 2R_1 v_1 & \text{in } \Omega_2 \end{cases} ; \mathtt{T}_2 v = \begin{cases} v_1 - 2R_2 v_2 & \text{in } \Omega_1 \\ -v_2 & \text{in } \Omega_2 \end{cases} ;$$

with $R_1 \in \mathcal{R}_1$, $R_2 \in \mathcal{R}_2$. In this section, we will further impose the condition that operators R_1, R_2 are bounded in L^2 norm (as it is the case for the geometric transfer operators we introduced previously). The question we would like to consider here is: what happens when corollary 1 does not apply, i.e. when $\mathtt{T}_1 V_h \not\subset V_h$ or $\mathtt{T}_2 V_h \not\subset V_h$? It turns out that one can still obtain convergence when the mesh is locally $\mathtt{T}_k$-conform, $k = 1$ or 2. Let us clarify this notion. Introduce I_h the classical interpolation operator such that $I_h(v) = \sum_{i=1}^{m(h)} v(a_i)\varphi_i$ for all $v \in \mathscr{C}^0(\overline{\Omega})$. Here, $(a_i)_{i=1..m(h)}$ are the nodes (including the nodes of the mesh located on the boundary) and φ_i, $i = 1..m(h)$, are the so-called "hat" functions which satisfy $\varphi_i(a_j) = \delta_{ij}$. Define

$$\mathtt{T}_{1h}^{\mathrm{loc}} v := \begin{cases} v_1 & \text{in } \Omega_1 \\ -v_2 + 2I_h(\chi)R_1 v_1 & \text{in } \Omega_2 \end{cases} ; \mathtt{T}_{2h}^{\mathrm{loc}} v := \begin{cases} v_1 - 2I_h(\chi)R_2 v_2 & \text{in } \Omega_1 \\ -v_2 & \text{in } \Omega_2 \end{cases} ,$$

where $\chi \in \mathscr{C}^\infty(\overline{\Omega}, [0;1])$ is a cut-off function such that $\chi = 1$ in a neighbourhood of Σ (that is there exists an open subset $\mathcal{V}$ of $\mathbb{R}^2$ such that $\Sigma \subset \mathcal{V}$ and $\chi = 1$ on $\mathcal{V}$).

Definition 10 For $k = 1, 2$, we will say that the meshes are *locally $\mathtt{T}_k$-conform* if $\mathtt{T}_{kh}^{\mathrm{loc}} V_h \subset V_h$ for all h smaller than a given $h_0 > 0$.

Proposition 3 *Assume that the form a is $\mathtt{T}_k$-coercive, that the meshes are locally $\mathtt{T}_k$-conform and that the basic approximability property holds. Then, for h small enough, there exists one and only one solution u_h to the problem (28) with the estimate*

$$\|u - u_h\|_{H_0^1(\Omega)} \leq C \inf_{v_h \in V_h} \|u - v_h\|_{H_0^1(\Omega)}, \tag{30}$$

where $C > 0$ is a constant which does not depend on h and f.

Proof Suppose that a is $\mathtt{T}_1$-coercive and that the mesh is locally $\mathtt{T}_1$-conform. Let us prove that the family $(a_h)_h$ defined by $a_h(v_h, w_h) = a(v_h, w_h)$ for all $v_h, w_h \in V_h$ is uniformly $V_h \times V_h$-stable, for h small enough.
To that aim, we will first prove the estimate, for h small enough,

$$|a(u_h, \mathtt{T}_{1h}^{\mathrm{loc}} u_h)| \geq C_1 \|u_h\|^2_{H_0^1(\Omega)} - C_2 \|u_h\|_{H_0^1(\Omega)} \|u_h\|_{\Omega}, \, \forall u_h \in V_h, \tag{31}$$

where $C_1 > 0$ and $C_2 > 0$ are two constants independent of h. Define the intermediate operator of $\mathcal{L}(V)$

$$\mathtt{T}_1^{\mathrm{loc}} v := \begin{cases} v_1 & \text{in } \Omega_1 \\ -v_2 + 2\chi R_1 v_1 & \text{in } \Omega_2 \end{cases}.$$

For $v \in H_0^1(\Omega)$, one has

$$\begin{aligned} & a(v, \mathtt{T}_1^{\mathrm{loc}} v) \\ &= (|\sigma| \nabla v, \nabla v)_{\Omega} - 2(|\sigma_2| \nabla v_2, \nabla(\chi R_1 v_1))_{\Omega_2} \\ &= (|\sigma| \nabla v, \nabla v)_{\Omega} - 2(|\sigma_2| \chi \nabla v_2, \nabla(R_1 v_1))_{\Omega_2} - 2(|\sigma_2| \nabla v_2, (R_1 v_1) \nabla \chi)_{\Omega_2}. \end{aligned} \tag{32}$$

Since $0 \leq \chi \leq 1$ and since a is $\mathtt{T}_1$-coercive, using (18), one finds there exists $C_3 > 0$ such that

$$(|\sigma| \nabla v, \nabla v)_{\Omega} - 2(|\sigma_2| \chi \nabla v_2, \nabla(R_1 v_1))_{\Omega_2} \geq C_3 \|v\|^2_{H_0^1(\Omega)}. \tag{33}$$

One the other hand, since R_1 is bounded for the L^2 norm, one obtains

$$2(|\sigma_2| \nabla v_2, (R_1 v_1) \nabla \chi)_{\Omega_2} \leq C_4 \|v\|_{H_0^1(\Omega)} \|v\|_{\Omega}. \tag{34}$$

Plugging (33) and (34) into (32), one finds

$$a(v, \mathtt{T}_1^{\mathrm{loc}} v) \geq C_3 \|v\|^2_{H_0^1(\Omega)} - C_4 \|v\|_{H_0^1(\Omega)} \|v\|_{\Omega}.$$

Then, observe that, for $v_h \in V_h$, there holds

$$\begin{aligned} |a(v_h, \mathtt{T}_1^{\mathrm{loc}} v_h) - a(v_h, \mathtt{T}_{1h}^{\mathrm{loc}} v_h)| &\leq C_5 \|\chi - I_h(\chi)\|_{W^{1,\infty}(\Omega)} \|v_h\|^2_{H_0^1(\Omega)} \\ &\leq C_6 h |\chi|_{W^{2,\infty}(\Omega)} \|v_h\|^2_{H_0^1(\Omega)}, \end{aligned}$$

according to corollary 1.109 of [12]. Thus,

$$|a(v_h, \mathtt{T}_{1h}^{\mathrm{loc}} v_h)| \geq (C_3 - C_6 |\chi|_{W^{2,\infty}(\Omega)} h) \|v_h\|^2_{H_0^1(\Omega)} - C_4 \|v_h\|_{H_0^1(\Omega)} \|v_h\|_{\Omega},$$

and (31) holds for h small enough.
Now, by contradiction, suppose that the family $(a_h)_h$ is not uniformly $V_h \times V_h$-stable: there exists a sequence of subspaces $(V_h)_h$ together with a sequence of elements $(v_h)_h$, with $v_h \in V_h$, such that

$$\|v_h\|_{H_0^1(\Omega)} = 1 \quad \text{and} \quad \sup_{w_h \in V_h \setminus \{0\}} \frac{|a(v_h, w_h)|}{\|w_h\|_{H_0^1(\Omega)}} < \mu_h, \text{ with } \lim_{h \to 0} \mu_h = 0. \tag{35}$$

Since $(v_h)_h$ is bounded in $H_0^1(\Omega)$ and since the injection of $H_0^1(\Omega)$ in $L^2(\Omega)$ is compact, there exists v in $H_0^1(\Omega)$ such that $(v_h)_h$ converges strongly in $L^2(\Omega)$

and weakly in $H^1(\Omega)$ to v. Classically, thanks to the the basic approximability property, one finds that v satisfies the homogeneous problem which implies that $v = 0$. Using (31) and the uniform continuity of the family $(\mathtt{T}_{1h}^{\mathrm{loc}})_h$, one deduces that, for h small enough, there holds

$$C_1 - C_2\|v_h\|_\Omega \leq C_7\mu_h,$$

where C_1, C_2 and C_7 are three strictly positive constants independent of h. As $\lim_{h\to 0}\|v_h\|_\Omega = \lim_{h\to 0}\mu_h = 0$, this leads to a contradiction. Thus, the family $(a_h)_h$ is uniformly $V_h \times V_h$-stable for h small enough and theorem 2 ensures that the problems (28) are well-posed with the estimate (30). One proceeds exactly in the same way working with $\mathtt{T}_2$ when a is $\mathtt{T}_2$-coercive and the mesh is locally $\mathtt{T}_2$-conform.

Remark 10 It suffices to have $\lim_{h\to 0}\left(|\chi|_{W^{2,\infty}(\Omega)}h\right) = 0$ in the proof of proposition 3. Consequently, we can allow the function χ to change with h. Thus, one can weaken the condition of $\mathtt{T}$-conformity for the mesh: we just need the mesh to be $\mathtt{T}$-conform in a neighbourhood of the interface whose area goes to zero like h^t for some $t \in (0, 1/2)$.

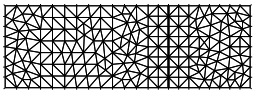

Fig. 4 Locally symmetric mesh for the cavity.

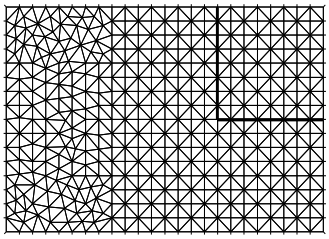

Fig. 5 Locally symmetric mesh for the corner–bis.

◇ Example of the cavity with a locally symmetric mesh (figure 4). Consider a family of meshes, as described on figure 4, which are symmetric with respect to Σ, in the region $(-0.25; 0.25) \times (0; 1)$. Here again, $\Omega := \{(x, y) \in (-2; 1) \times (0; 1)\}$, $\Omega_1 := (-2; 0) \times (0; 1)$ and $\Omega_2 := (0; 1) \times (0; 1)$. According to proposition 3, Problem (28) is well-posed for h small enough as soon as $\kappa_\sigma \notin [-1; -1/2]$. Moreover, in this case, one has the error estimate

$$\|u - u_h\|_{H_0^1(\Omega)} \leq Ch\|f\|_\Omega.$$

◇ Example of the corner–bis (figure 5). Let us consider now the geometry of figure 5. More precisely, define $\Omega := (-2; 1) \times (-1; 1)$, $\Omega_2 := (0; 1)^2$ and $\Omega_1 := \Omega \setminus \overline{\Omega_2}$. According to theorem 3, the operator A is Fredholm of index 0 as soon as $\kappa_\sigma = \sigma_2/\sigma_1 \notin [-3; -1/3]$.
Extending the operator R_2 defined in (25) by 0 on $(-2; -1) \times (-1; 1)$, one

finds that A is an isomorphism from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$ as soon as $\kappa_\sigma < -3$, so that Problem (28) is well-posed for h small enough.
Now, suppose $-1/3 < \kappa_\sigma < 0$ and that A is injective. Introduce $\chi_0 \in \mathscr{C}^\infty(\mathbb{R}, [0;1])$ a cut-off function such that $\chi_0(x) = 1$ for $x \geq -1/2$ and $\chi_0(x) = 0$ for $x \leq -1$. Define $\chi : (x, y) \mapsto \chi_0(x)$. According to proposition 3, Problem (28) is well-posed for h small enough.
Moreover, in these two cases ($\kappa_\sigma < -3$ and $-1/3 < \kappa_\sigma < 0$, A injective), error estimate (29) is valid.

5 Application to $\mathrm{div}\,\sigma\nabla\cdot$: approximation of the solution without hypothesis on the mesh

5.1 Numerical approximation: general mesh

In the present subsection, we would like to consider the case of a general mesh which is neither T-conform nor locally T-conform. In this situation, corollary 1 and proposition 3 fail to justify the well-posedness of the discrete problems. The question to be addressed is how to build a family $(\mathtt{T}_h)_h$ of discrete operators such that the form a is uniformly $\mathtt{T}_h$-coercive, at least for h small enough. Some methods have already been proposed in [3] and [20]. The first one relies on a lifting of the trace on the interface. The second one is based on taking $R_h = \Pi_h^{SZ} R$ where Π_h^{SZ} is the Scott-Zhang interpolation operator [23]. More precisely, the authors apply the Scott-Zhang interpolation operator respectively to $(R_1 u_h)_{|\Omega_2}$ and $(R_2 u_h)_{|\Omega_1}$. Since this operator preserves the boundary conditions, it follows that $\Pi_h^{SZ}(R_1 u_h)_{|\Omega_2} = u_h$ and $\Pi_h^{SZ}(R_2 u_h)_{|\Omega_1} = u_h$ on the interface Σ. The main limitation of theses two approaches is that their range of applicability is not clear *a priori*: in a general situation, for a given value of the contrast and a general mesh, we can not ensure that the discrete problem (28) is well-posed, even for h small enough. Let us explain briefly where the difficulty arises. For that, we propose below an alternate approach to [3,20]. Define

$$V_{1\,h} := \left\{ v_{h|\Omega_1} \,|\, v_h \in V_h \right\}; \quad V_{2\,h} := \left\{ v_{h|\Omega_2} \,|\, v_h \in V_h \right\};$$

$$V_{1\,h}^0 := H_0^1(\Omega_1) \cap V_{1\,h}; \quad V_{2\,h}^0 := H_0^1(\Omega_2) \cap V_{2\,h}.$$

For all $v_{1\,h} \in V_{1\,h}$, let $R_{1\,h} v_{1\,h}$ be defined as the unique solution to problem

$$\begin{cases} \textit{Find } R_{1\,h} v_{1\,h} \in V_{2\,h} \textit{ such that } R_{1\,h} v_{1\,h} = v_{1\,h} \textit{ on } \Sigma \textit{ and} \\ \forall w_h \in V_{2\,h}^0,\ (\sigma\nabla(R_{1\,h} v_{1\,h}), \nabla w_h)_{\Omega_2} = (\sigma\nabla(R_1 v_{1\,h}), \nabla w_h)_{\Omega_2}. \end{cases} \tag{36}$$

For all $h > 0$, one has $|||R_{1\,h}||| \leq C$ where C is a constant independent of h. On the other hand, there is no guarantee that $\inf_{R_{1\,h}} |||R_{1\,h}|||$ is equal to $\inf_{R_1} |||R_1|||$. So, in the spirit of theorem 4, well-posedness of the discrete problems (28) is guaranteed, however under a more restrictive condition on the contrast than $\kappa_\sigma > -1/\inf_{R_1} |||R_1|||$.

Remark 11 Let $v_{1\,h} \in V_{1\,h}$. By construction (cf. (36)), one has $R_{1\,h}v_{1\,h} - R_1 v_{1\,h} \in H_0^1(\Omega_2)$. So, if in addition $R_1 v_{1\,h}$ belongs to $V_{2\,h}$, one obtains $R_{1\,h}v_{1\,h} = R_1 v_{1\,h}$. For this property to hold for all $v_{1\,h} \in V_{1\,h}$, it is sufficient that the mesh be T-conform. According to theorem 4, to recover the same applicability as the continuous Problem (15), one needs that this property be fulfilled for R_1 with minimal norm.

5.2 Numerical approximation: using dissipation

Given $\gamma > 0$, let $\sigma^\gamma := (1 + \imath\,\mathrm{sign}(\sigma)\gamma)\sigma$, and define the approximate problem

$$\begin{cases} \textit{Find } u^\gamma \in H_0^1(\Omega) \textit{ such that} \\ \forall v \in H_0^1(\Omega),\ (\sigma^\gamma \nabla u^\gamma, \nabla v)_\Omega = (f, v). \end{cases} \tag{37}$$

First, one can check easily that

$$\forall v \in H_0^1(\Omega),\ |(\sigma^\gamma \nabla v, \nabla v)_\Omega| \geq \min(\sigma_1, |\sigma_2|)\gamma \|v\|^2_{H_0^1(\Omega)}. \tag{38}$$

In other words, this approximate problem is *always* well-posed for $\gamma > 0$. Below, we let γ go to 0.
We define the operator $A^\gamma \in \mathcal{L}(H_0^1(\Omega), H^{-1}(\Omega))$ such that $\langle A^\gamma u^\gamma, v\rangle = a^\gamma(u^\gamma, v)$ for all $v \in H_0^1(\Omega)$. One has

$$Au = A^\gamma u^\gamma \Leftrightarrow A(u - u^\gamma) = (A^\gamma - A)u^\gamma \Leftrightarrow u - u^\gamma = A^{-1}(A^\gamma - A)u^\gamma.$$

For the last equation, we used the fact that the Problem (15) is well-posed. Noticing that $|((\sigma - \sigma^\gamma)\nabla u, \nabla v)_\Omega| \leq \max(\sigma_1, |\sigma_2|)\gamma \|u\|_{H_0^1(\Omega)} \|v\|_{H_0^1(\Omega)}$ for all $u, v \in H_0^1(\Omega)$, it yields $|||A^\gamma - A||| \leq \max(\sigma_1, |\sigma_2|)\gamma$. Consequently, $\|u - u^\gamma\|_{H_0^1(\Omega)} \leq C_1\gamma \|u^\gamma\|_{H_0^1(\Omega)}$ with $C_1 = |||A^{-1}|||\max(\sigma_1, |\sigma_2|)$. Ones deduces $(1 - C_1\gamma)\|u^\gamma\|_{H_0^1(\Omega)} \leq \|u\|_{H_0^1(\Omega)}$ which proves that $(u^\gamma)_\gamma$ is bounded. Moreover, there holds the estimate

$$\|u - u^\gamma\|_{H_0^1(\Omega)} \leq C_2\gamma \|u\|_{H_0^1(\Omega)} \leq C_3\gamma \|f\|_\Omega.$$

Next, one builds a finite dimensional approximation of Problem (37), which writes

$$\begin{cases} \textit{Find } u_h^\gamma \in V_h \textit{ such that} \\ \forall v_h \in V_h,\ (\sigma^\gamma \nabla u_h^\gamma, \nabla v_h)_\Omega = (f, v_h). \end{cases} \tag{39}$$

According to (38), Problem (39) is always well-posed: discussions on applicability are superfluous, i.e. the applicability of the approximation with dissipation is the same as for the continuous Problem (15)! Using Céa's lemma, we find

$$\|u^\gamma - u_h^\gamma\|_{H_0^1(\Omega)} \leq \frac{C_4}{\gamma} \inf_{v_h \in V_h} \|u^\gamma - v_h\|_{H_0^1(\Omega)},$$

where C_4 is independent of γ and h. Applying the triangular inequality leads to

$$\|u - u_h^\gamma\|_{H_0^1(\Omega)} \leq \|u - u^\gamma\|_{H_0^1(\Omega)} + \|u^\gamma - u_h^\gamma\|_{H_0^1(\Omega)} \leq C_3\gamma \|f\|_\Omega + \frac{C_4}{\gamma} \inf_{v_h \in V_h} \|u^\gamma - v_h\|_{H_0^1(\Omega)}.$$

To conclude, one has to estimate $\inf_{v_h \in V_h} \|u^\gamma - v_h\|_{H_0^1(\Omega)}$. Let us assume that a uniform approximability property like $|u^\gamma|_{H^s(\Omega_1)} + |u^\gamma|_{H^s(\Omega_2)} \leq C_5 \|f\|_\Omega$ holds for some $s > 0$ and γ small enough (elements of proof are given in the annex). Then, one has

$$\inf_{v_h \in V_h} \|u^\gamma - v_h\|_{H_0^1(\Omega)} \leq C_6 h^s \|f\|_\Omega \,.$$

Finally, one can optimize the error estimate by choosing $\gamma = \sqrt{C_4 C_6 / C_3}\, h^{s/2}$, leading to

$$\|u - u_h^\gamma\|_{H_0^1(\Omega)} \leq 2\sqrt{C_3 C_4 C_6}\, h^{s/2} \|f\|_\Omega.$$

This estimate holds as soon as Problem (15) is well-posed and $1 > C_1\gamma$. As $\gamma \sim h^{s/2}$, the latter holds for h "small" enough.

Remark 12 In the above analysis, we assumed that $1 > C_1\gamma$, where $C_1 = |||A^{-1}||| \max(\sigma_1, |\sigma_2|)$. It can happen that the norm $|||A^{-1}|||$ is very "large", so it is important in practice to choose the parameter γ like $\gamma = C_8\, h^{s/2}$ with C_8 "small". In this case, one has $1/|||A^{-1}||| > C_8 \max(\sigma_1, |\sigma_2|)\, h^{s/2}$ even for coarse meshes, with an error in $O(h^{s/2}\|f\|_\Omega)$.

◇ Example of the cavity with a general mesh. In this example, we do not assume that the mesh of the cavity $\Omega := \{(x, y) \in (-2; 1) \times (0; 1)\}$ is locally symmetric. Recall that $\Omega_1 := (-2; 0) \times (0; 1)$ and $\Omega_2 := (0; 1) \times (0; 1)$. Suppose that $\kappa_\sigma \in (-1; -1/2] \setminus \{-\tanh(n\pi)/\tanh(2n\pi),\, n \in \mathbb{N}^*\}$. We know that in such a case, A is an isomorphism from $H_0^1(\Omega)$ to $H^{-1}(\Omega)$.
Then, according to proposition 4 (see §A), the only solution u^γ to problem (37) satisfies $|u^\gamma|_{H^2(\Omega_1)} + |u^\gamma|_{H^2(\Omega_2)} \leq C\|f\|_\Omega$, for γ small enough. Consequently, for a family of general meshes of Ω, we can approximate the unique solution u to problem (27) by the sequence $(u_h^\gamma)_h$. Moreover, there holds the error estimate

$$\|u - u_h^\gamma\|_{H_0^1(\Omega)} \leq C\sqrt{h}\|f\|_\Omega,$$

for h small enough if we take $\gamma \sim \sqrt{h}$.

6 Numerical experiments: influence of the mesh for the cavity example

Let us consider the symmetric cavity defined by $\Omega := \{(x, y) \in (-1; 1)\times(0; 1)\}$, $\Omega_1 := (-1; 0) \times (0; 1)$ and $\Omega_2 := (0; 1) \times (0; 1)$. See figure 6 for different kinds of meshes.
Consider $u \in H_0^1(\Omega)$ defined by

$$u(x, y) := \begin{vmatrix} ((x+1)^2 - (\sigma_1+\sigma_2)^{-1}(2\sigma_1+\sigma_2)(x+1))\sin(\pi y) & \text{on } \Omega_1; \\ (\sigma_1+\sigma_2)^{-1}\sigma_1(x-1)\sin(\pi y) & \text{on } \Omega_2; \end{vmatrix}$$

and $f := -\operatorname{div}\sigma\nabla u \in L^2(\Omega)$. We set σ_1 to 1. According to the results of §3.3, the problem (27) is well-posed as soon as $\kappa_\sigma \neq -1 \Leftrightarrow \sigma_1 + \sigma_2 \neq 0$. Moreover, according to the results of §4.2 and §4.3, we know that discrete problems (28)

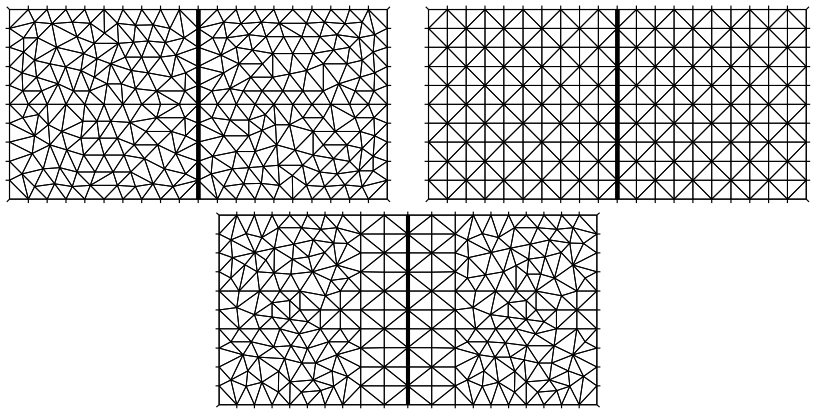

Fig. 6 Meshes for the cavity: Non symmetric mesh (top left) - Symmetric mesh (top right) - Locally symmetric mesh (center).

are well-posed (at least for h small enough), for the symmetric mesh and for the locally symmetric mesh. However, up to now, we have not been able to prove that (28) was well-posed, even for h small enough, for the non symmetric mesh. On the other hand, using dissipation, one recovers automatically well-posed discrete problems, such as (39). According to remark 12, we choose a small dissipation coefficient.

We show on figures 7 and 8 numerical results for a value of the constrast $\kappa_\sigma = \sigma_2/\sigma_1 = -1.001$, with a meshsize $h \in (10^{-0.8}; 10^{-2.2})$. The relative errors, plotted respectively in H^1 semi norm and L^2 norm, are reported in log–log scale, with a the order of convergence. We obtain that all approaches:

- natural discretization with symmetric meshes;
- natural discretization with locally symmetric meshes;
- natural discretization with non symmetric meshes;
- discretization with dissipation with non symmetric meshes;

converge to the exact solution, even though the chosen constrast is very close to the critical value -1. The lowest convergence order is observed for the discretization with dissipation, as expected. Furthermore, it behaves like $O(\sqrt{h})$ as predicted by the theory. On the other hand, the natural discretizations with either symmetric meshes or locally symmetric meshes converge with the expected rates, namely $O(h)$ in H^1 semi norm and $O(h^2)$ in L^2 norm, where the latter estimate is a consequence of the Aubin-Nitsche lemma (cf. [12]) applied to our problem.

To improve the convergence order of the discretization with dissipation, one can increase the discretization order (for instance, P_2 or P_3 Finite Elements),

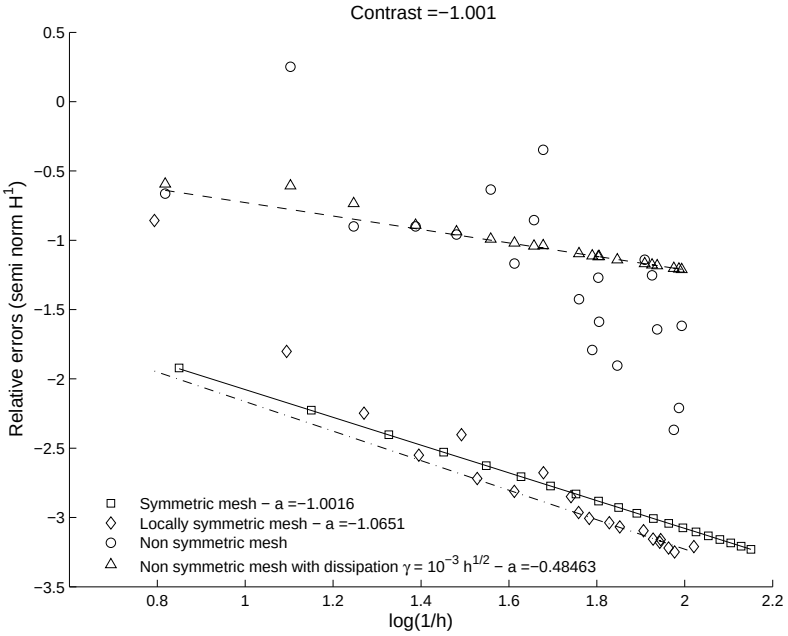

Fig. 7 Errors (H^1 semi norm) for different meshes of the cavity.

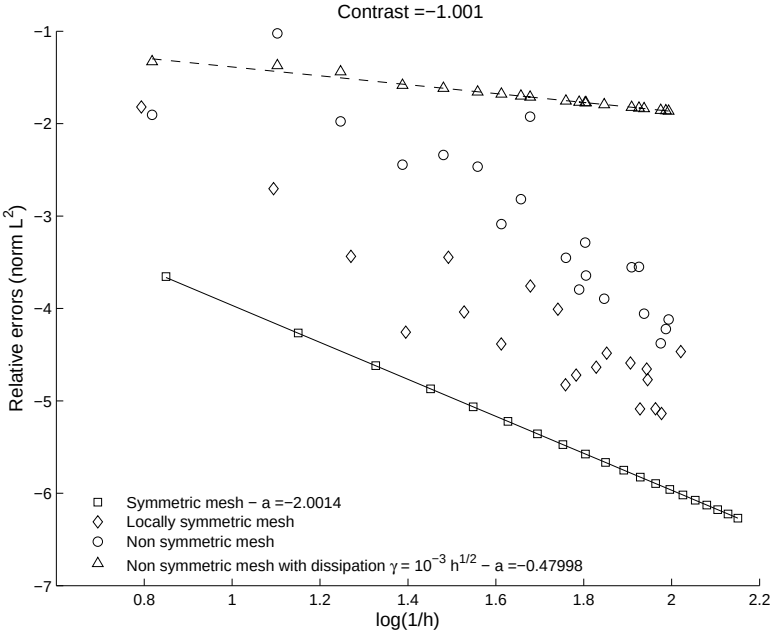

Fig. 8 Errors (L^2 norm) for different meshes of the cavity.

using an appropriately tuned dissipation coefficient. Denoting by $m \in \{1, 2, 3\}$ the order of the finite element, one chooses $\gamma_m \sim h^{m/2}$ to recover a convergence rate in $O(h^{m/2})$ (here the solution is piecewise smooth). The results are shown in figure 9: again, the method behaves as expected.

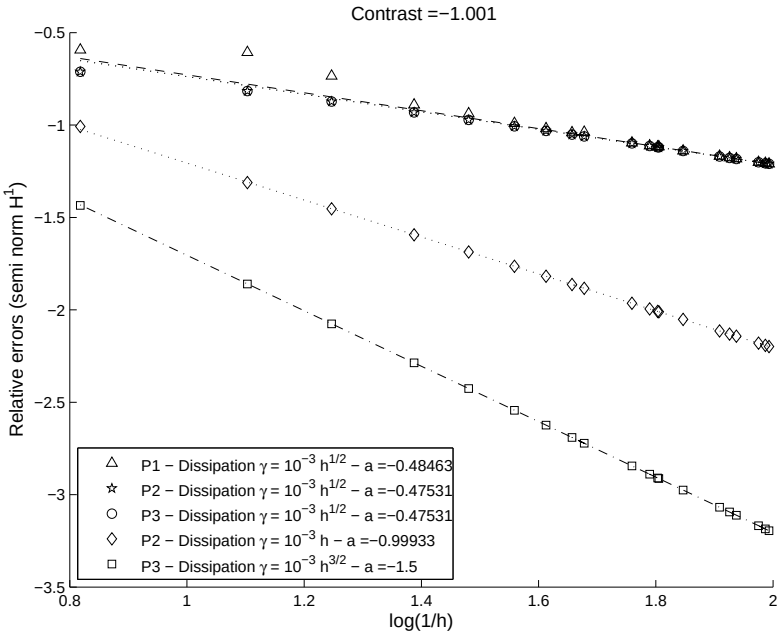

Fig. 9 Comparison of errors (H^1 semi norm) for different finite elements and dissipations.

A Annex

In this section, we consider the geometry of figure 2 for which $\Omega := \{(x, y) \in (-2; 1)\times(0; 1)\}$, $\Omega_1 := (-2; 0) \times (0; 1)$ and $\Omega_2 := (0; 1) \times (0; 1)$ and we suppose that

$$\kappa_\sigma \in (-1; -1/2] \setminus \{-\tanh(n\pi)/\tanh(2n\pi),\, n \in \mathbb{N}^*\}.$$

We let u denote the unique solution to Problem (27).

Remark 13 The results of the Annex hold if $\kappa_\sigma \notin \{-\tanh(n\pi)/\tanh(2n\pi),\, n \in \mathbb{N}^*\}\cup\{-1\}$.

One has the

Proposition 4 *For $\gamma > 0$, let u^γ be the unique solution of problem (37). Then, there holds*

$$|u^\gamma|_{H^2(\Omega_1)} + |u^\gamma|_{H^2(\Omega_2)} \leq C\|f\|_\Omega,$$

for γ small enough, with $C > 0$ independent of γ.

Proof Using a partition of unity, it is sufficient to prove this result locally.
Introduce ζ_1, $\zeta_2 \in \mathscr{C}^\infty(\mathbb{R}^2,[0;1])$ two cut-off functions independent of y such that:

$$\zeta_1(x,y) = 1 \text{ for } x \leq -0.5 \quad \text{and} \quad \zeta_1(x,y) = 0 \text{ for } x \geq -0.25;$$
$$\zeta_2(x,y) = 0 \text{ for } x \leq -0.75 \quad \text{and} \quad \zeta_2(x,y) = 1 \text{ for } x \geq -0.5.$$

◇ `Approximation away from the interface:`

Lemma 1 *There exists a constant $C > 0$, independent of γ and f, such that $|\zeta_1(u - u^\gamma)|_{H^2(\Omega)} \leq C\gamma\|f\|_\Omega$ for γ small enough.*

Proof Define $\mathcal{O} := (-2;-0.25)\times(0;1)$. Since σ is constant in $\mathcal{O}$, simple computations yield to

$$-\Delta(\zeta_1(u-u^\gamma)) = g,$$

with $g = \zeta_1(f/\sigma - f/\sigma^\gamma) - 2\nabla\zeta_1\cdot\nabla(u-u^\gamma) - \Delta\zeta_1(u-u^\gamma) \in L^2(\mathcal{O})$. Now, $\|g\|_\mathcal{O} \leq C\gamma\|f\|_\Omega$ and we know that the Laplacian with homogeneous Dirichlet boundary condition is an isomorphism from $H^2(\mathcal{O})\cap H^1_0(\mathcal{O})$ to $L^2(\mathcal{O})$, so one can write

$$|\zeta_1(u-u^\gamma)|_{H^2(\Omega)} = |\zeta_1(u-u^\gamma)|_{H^2(\mathcal{O})} \leq C\|g\|_\mathcal{O} \leq C\gamma\|f\|_\Omega.$$

□

◇ `Approximation near the interface:`

Lemma 2 *There exists a constant $C > 0$, independent of γ and f, such that $|\zeta_2(u - u^\gamma)|_{H^2(\Omega_1)} + |\zeta_2(u-u^\gamma)|_{H^2(\Omega_2)} \leq C\gamma\|f\|_\Omega$ for γ small enough.*

Proof Introduce the infinite strips $\mathcal{I} := I\times\mathbb{R}$, $\mathcal{I}_j := I_j\times\mathbb{R}$, $j = 1,2$, with $I := (-1;1)$, $I_1 := (-1;0)$ and $I_2 := (0;1)$. By odd reflection, on $\tilde{\mathcal{O}} := (-1;1)\times(-1;2)$, define the functions $\tilde{u}$ and $\tilde{u}^\gamma$ such that,

$$\tilde{u}(x,y) := \begin{cases} -\,u(x,2-y) & \text{for } 1 \leq y \leq 2 \\ \quad u(x,y) & \text{for } 0 \leq y \leq 1 \\ -\,u(x,-y) & \text{for } -1 \leq y \leq 0 \end{cases},$$

$$\tilde{u}^\gamma(x,y) := \begin{cases} -\,u^\gamma(x,2-y) & \text{for } 1 \leq y \leq 2 \\ \quad u^\gamma(x,y) & \text{for } 0 \leq y \leq 1 \\ -\,u^\gamma(x,-y) & \text{for } -1 \leq y \leq 0 \end{cases},$$

for all $x \in (-1;1)$. Define also, again for $x \in (-1;1)$,

$$\tilde{f}(x,y) := \begin{cases} -f(x,2-y) & \text{for } 1 \leq y \leq 2 \\ f(x,y) & \text{for } 0 \leq y \leq 1 \\ -f(x,-y) & \text{for } -1 \leq y \leq 0 \end{cases}.$$

Introduce $\chi \in \mathscr{C}^\infty(\mathbb{R}^2,[0;1])$ a cut-off function independent of x such that:

$$\chi(x,y) = 1 \text{ for } 0 \leq y \leq 1 \quad \text{and} \quad \chi(x,y) = 0 \text{ for } y \leq -0.5 \text{ and } y \geq 1.5.$$

Now, we localize the study of regularity with the help of χ. In the sequel, we make no distinction between elements of $H^1_0(\tilde{\mathcal{O}})$ or $L^2(\tilde{\mathcal{O}})$ and their extension by 0 to $\mathcal{I}$. Consider

$$p := \sigma\,(\tilde{u}\Delta(\chi\zeta_2) + 2\nabla\tilde{u}\cdot\nabla(\chi\zeta_2)) + \tilde{f}\chi\zeta_2 \quad \text{and} \quad p^\gamma := \sigma^\gamma\,(\tilde{u}^\gamma\Delta(\chi\zeta_2) + 2\nabla\tilde{u}^\gamma\cdot\nabla(\chi\zeta_2)) + \tilde{f}\chi\zeta_2\,.$$

These two elements belong to $L^2(\mathcal{I})$ and have compact support. According to their definition, $v := \chi\zeta_2\tilde{u}$ and $v^\gamma := \chi\zeta_2\tilde{u}^\gamma$ satisfy respectively the transmission problem in the infinite strip $\mathcal{I}$

$$(\mathcal{P}_{strip}) \left| \begin{array}{ll} \sigma_j\Delta v_j = p_j & \text{in } \mathcal{I}_j,\ j = 1,2 \\ v_j = 0 & \text{on } \partial\mathcal{I}_j\cap\partial\mathcal{I},\ j = 1,2 \\ v_1 - v_2 = 0 & \text{on } \partial\mathcal{I}_1\cap\partial\mathcal{I}_2 \\ \sigma_1\partial_x v_1 - \sigma_2\partial_x v_2 = 0 & \text{on } \partial\mathcal{I}_1\cap\partial\mathcal{I}_2, \end{array} \right.$$

$$(\mathcal{P}^\gamma_{strip})\begin{vmatrix}\sigma^\gamma_j \Delta v^\gamma_j = p^\gamma_j & \text{in } \mathcal{I}_j,\ j=1,2\\ v^\gamma_j = 0 & \text{on } \partial\mathcal{I}_j \cap \partial\mathcal{I},\ j=1,2\\ v^\gamma_1 - v^\gamma_2 = 0 & \text{on } \partial\mathcal{I}_1 \cap \partial\mathcal{I}_2\\ \sigma^\gamma_1 \partial_x v^\gamma_1 - \sigma^\gamma_2 \partial_x v^\gamma_2 = 0 & \text{on } \partial\mathcal{I}_1 \cap \partial\mathcal{I}_2.\end{vmatrix}$$

Applying the Fourier transform with respect to y to the equations of $(\mathcal{P}_{strip})$ and $(\mathcal{P}^\gamma_{strip})$ for $\lambda \in \mathbb{R}i$, one finds that $x \mapsto \hat{v}(x,\lambda) := \int_{-\infty}^{+\infty} e^{-\lambda y} v(x,y)\, dy$ and $x \mapsto \hat{v}^\gamma(x,\lambda) := \int_{-\infty}^{+\infty} e^{-\lambda y} v^\gamma(x,y)\, dy$ are respectively governed by

$$(\hat{\mathcal{P}}_{strip})\begin{vmatrix}\sigma_j \left(\partial_x^2 + \lambda^2\right) \hat{v}_j(x,\lambda) = \hat{p}_j(x,\lambda) & \text{in } \mathcal{I}_j,\ j=1,2\\ \hat{v}_1(-1,\lambda) = \hat{v}_2(1,\lambda) = 0\\ \hat{v}_1(0,\lambda) = \hat{v}_2(0,\lambda)\\ \sigma_1 \partial_x \hat{v}_1(0,\lambda) = \sigma_2 \partial_x \hat{v}_2(0,\lambda),\end{vmatrix}$$

$$(\hat{\mathcal{P}}_{strip})\begin{vmatrix}\sigma^\gamma_j \left(\partial_x^2 + \lambda^2\right) \hat{v}^\gamma_j(x,\lambda) = \hat{p}^\gamma_j(x,\lambda) & \text{in } \mathcal{I}_j,\ j=1,2\\ \hat{v}^\gamma_1(-1,\lambda) = \hat{v}^\gamma_2(1,\lambda) = 0\\ \hat{v}^\gamma_1(0,\lambda) = \hat{v}^\gamma_2(0,\lambda)\\ \sigma^\gamma_1 \partial_x \hat{v}^\gamma_1(0,\lambda) = \sigma^\gamma_2 \partial_x \hat{v}^\gamma_2(0,\lambda).\end{vmatrix}$$

Lemma 3 *There exists a constant C independent of $\lambda \in \mathbb{R}i$ and γ such that*

$$\textstyle\sum_{j=1}^2 |\hat{v}_j - \hat{v}^\gamma_j|_{H^2(I_j)} + |\lambda|^2 \|\hat{v}_j - \hat{v}^\gamma_j\|_{L^2(I)} \leq C\gamma \|\hat{p}\|_{L^2(I)},$$

for γ small enough.

Proof Denote respectively $(\cdot,\cdot)$, $(\cdot,\cdot)_1$, $(\cdot,\cdot)_2$ the scalar products of $L^2(I)$, $L^2(I_1)$ and $L^2(I_2)$. Define $\tau := i\lambda \in \mathbb{R}$. Introduce the sesquilinear forms defined by, for $\varphi, \psi \in H^1_0(I)$,

$$\begin{aligned} d(\varphi,\psi) &:= \textstyle\sum_{j=1}^2 \left(\sigma_j(\varphi'_j,\psi'_j)_j + \tau^2 \sigma_j(\varphi_j,\psi_j)_j\right);\\ d^\gamma(\varphi,\psi) &:= \textstyle\sum_{j=1}^2 \left(\sigma^\gamma_j(\varphi'_j,\psi'_j)_j + \tau^2 \sigma^\gamma_j(\varphi_j,\psi_j)_j\right). \end{aligned}$$

Let us first study the form d. Since d is not coercive on $H^1_0(I)\times H^1_0(I)$, we use the T-coercivity method in 1D. Introduce the operator R^{1D}_1 such that $(R^{1D}_1\varphi_1)(x) = \varphi_1(-x)$ ($|||R^{1D}_1||| = 1$) and the isomorphism ($\mathtt{T}^{1D}_1 \circ \mathtt{T}^{1D}_1 = \mathtt{I}$) of $H^1_0(I)$ defined by

$$\mathtt{T}^{1D}_1 \varphi := \begin{cases} \varphi_1 & \text{on } I_1\\ -\varphi_2 + 2R^{1D}_1\varphi_1 & \text{on } I_2\end{cases}.$$

For all $\varphi \in H^1_0(I)$, one can write, using Young's inequality, for all $\eta > 0$,

$$\begin{aligned} |\sigma_1^{-1} d(\varphi, \mathtt{T}^{1D}_1\varphi)| &= |(\varphi'_1,\varphi'_1)_1 + \tau^2(\varphi_1,\varphi_1)_1 + |\sigma_2/\sigma_1|((\varphi'_2,\varphi'_2)_2 + \tau^2(\varphi_2,\varphi_2)_2)\\ &\quad +2(\sigma_2/\sigma_1)((\varphi'_2,(R^{1D}_1\varphi_1)')_2 + \tau^2(\varphi_2,(R^{1D}_1\varphi_1))_2)|\\ &\geq (1-\eta^{-1}|\sigma_2/\sigma_1|)((\varphi'_1,\varphi'_1)_1 + \tau^2(\varphi_1,\varphi_1)_1)\\ &\quad +|\sigma_2/\sigma_1|(1-\eta)((\varphi'_2,\varphi'_2)_2 + \tau^2(\varphi_2,\varphi_2)_2). \end{aligned}$$

Thus, as $|\sigma_2/\sigma_1| = |\kappa_\sigma| < 1$, taking η such that $|\sigma_2/\sigma_1| < \eta < 1$, one infers the existence of a constant C independent of τ such that

$$|d(\varphi, \mathtt{T}^{1D}_1\varphi)| \geq C((\varphi',\varphi') + \tau^2(\varphi,\varphi)),\ \forall \varphi \in H^1_0(I). \tag{40}$$

One deduces

$$\begin{aligned} &C((\hat{v}' - \hat{v}^{\gamma\prime}, \hat{v}' - \hat{v}^{\gamma\prime}) + \tau^2(\hat{v} - \hat{v}^\gamma, \hat{v} - \hat{v}^\gamma))\\ &\leq |d(\hat{v} - \hat{v}^\gamma, \mathtt{T}^{1D}_1(\hat{v} - \hat{v}^\gamma))|\\ &= |d(\hat{v}, \mathtt{T}^{1D}_1(\hat{v} - \hat{v}^\gamma)) - d(\hat{v}^\gamma, \mathtt{T}^{1D}_1(\hat{v} - \hat{v}^\gamma))|\\ &= |d(\hat{v}, \mathtt{T}^{1D}_1(\hat{v} - \hat{v}^\gamma)) - d^\gamma(\hat{v}^\gamma, \mathtt{T}^{1D}_1(\hat{v} - \hat{v}^\gamma)) + (d^\gamma - d)(\hat{v}^\gamma, \mathtt{T}^{1D}_1(\hat{v} - \hat{v}^\gamma))|\\ &\leq |(\hat{p} - \hat{p}^\gamma, \mathtt{T}^{1D}_1(\hat{v} - \hat{v}^\gamma))| + |(d^\gamma - d)(\hat{v}^\gamma, \mathtt{T}^{1D}_1(\hat{v} - \hat{v}^\gamma))|. \end{aligned}$$

Since, $\|\hat{p}-\hat{p}^\gamma\|_{L^2(I)} \leq C\gamma\|\hat{p}\|_{L^2(I)}$, one finds that the sequence $(\|\hat{v}^\gamma\|_{H^1(I)})_\gamma$ is bounded, and then that

$$|\hat{v}-\hat{v}^\gamma|_{H^1(I)} + |\lambda|^2 \|\hat{v}-\hat{v}^\gamma\|_{L^2(I)} \leq C\gamma\|\hat{p}\|_{L^2(I)}, \tag{41}$$

for γ small enough. Noticing that $\sigma_j(\lambda^2\hat{v}+(\hat{v})'') = \hat{p}_j$ and $\sigma_j^\gamma(\lambda^2\hat{v}^\gamma+(\hat{v}^\gamma)'') = \hat{p}_j^\gamma$, $j=1,2$, lemma 3 is proved. □

With the help of the Parseval identity (see the lemma 5.2.4 of [15]), one finds $|v-v^\gamma|_{H^2(\mathcal{I}_1)} + |v-v^\gamma|_{H^2(\mathcal{I}_2)} \leq C\gamma\|p\|_{\mathcal{I}}$. Since $v=\zeta_2 u$ and $v^\gamma=\zeta_2 u^\gamma$ for $(x,y)\in(-1;1)\times(0;1)$, one obtains the result of lemma 2 noticing that $\|p\|_{\mathcal{I}} \leq C\|f\|_\Omega$. □

◇ `Conclusion of the proof of proposition 4:` According to lemmas 1 and 2, one has

$$\begin{aligned} &|u-u^\gamma|_{H^2(\Omega_1)} + |u-u^\gamma|_{H^2(\Omega_2)} \\ &\leq |\zeta_1(u-u^\gamma)|_{H^2(\Omega)} + |\zeta_2(u-u^\gamma)|_{H^2(\Omega_1)} + |\zeta_2(u-u^\gamma)|_{H^2(\Omega_2)} \leq C\gamma\|f\|_\Omega. \end{aligned}$$

Consequently, using proposition 2 yields

$$\begin{aligned} |u^\gamma|_{H^2(\Omega_1)} + |u^\gamma|_{H^2(\Omega_2)} &\leq C\gamma\|f\|_\Omega + |u|_{H^2(\Omega_1)} + |u|_{H^2(\Omega_2)} \\ &\leq C\|f\|_\Omega. \end{aligned}$$

This concludes the proof. □

References

1. Bonnet-Ben Dhia, A.-S., Chesnel, L., Ciarlet Jr., P.: T-coercivity for scalar interface problems between dielectrics and metamaterials. Math. Mod. Num. Anal. **46**, 1363–1387 (2012)
2. Bonnet-Ben Dhia, A.-S., Chesnel, L., Claeys, X.: Radiation condition for a non smooth interface between dielectric and metamaterial (submitted)
3. Bonnet-Ben Dhia, A.-S., Ciarlet Jr., P., Zwölf, C.-M.: Time harmonic wave diffraction problems in materials with sign-shifting coefficients. J. Comput. Appl. Math **234**, 1912–1919 (2010). Corrigendum J. Comput. Appl. Math., 234:2616, 2010. Eight International Conference on Mathematical and Numerical Aspects of Waves (Waves 2007)
4. Bonnet-Ben Dhia, A.-S., Dauge, M., Ramdani, K.: Analyse spectrale et singularités d'un problème de transmission non coercif. C. R. Acad. Sci. Paris, Ser. I **328**, 717–720 (1999)
5. Brezzi, F., Fortin, M.: Mixed and hybrid finite element methods. Springer Series In Computational Mathematics, New York (1991)
6. Chesnel, L., Ciarlet, Jr., P.: Compact imbeddings in electromagnetism with interfaces between classical materials and meta-materials. SIAM J. Math. Anal. **43**, 2150–2169 (2011)
7. Chung, E.T., Ciarlet, Jr., P.: Scalar transmission problems between dielectrics and metamaterials: T-coercivity for the Discontinuous Galerkin approach. Technical Report, CUHK-2011-01, Chinese University of Hong Kong, Hong Kong (2011)
8. Ciarlet Jr., P.: T-coercivity: Application to the discretization of Helmholtz-like problems. Computers and Mathematics with Applications (to appear)
9. Costabel, M., Stephan, E.: A direct boundary integral method for transmission problems. J. of Math. Anal. and Appl. **106**, 367–413 (1985)
10. Dauge, M., Texier, B.: Problèmes de transmission non coercifs dans des polygones. http://hal.archives-ouvertes.fr/docs/00/56/23/29/PDF/BenjaminT_arxiv.pdf (2010)
11. Engheta, N.: An idea for thin subwavelength cavity resonator using metamaterials with negative permittivity and permeability. IEEE Antennas Wireless Propag. Lett. **1**, 10–13 (2002)
12. Ern, A., Guermond, J.-L.: Theory and practice of finite elements. Springer-Verlag, Berlin (2004)
13. Genov, D.A., Zhang, S., Zhang, X.: Mimicking celestial mechanics in metamaterials. Nature Physics **5**(9), 687 (2009)

14. Grisvard, P.: Singularities in Boundary Value Problems. RMA 22. Masson, Paris (1992)
15. Kozlov, V.A., Maz'ya, V.G., Rossmann, J.: Elliptic Boundary Value Problems in Domains with Point Singularities, *Mathematical Surveys and Monographs*, vol. 52. AMS, Providence (1997)
16. Lions, J.L., Magenes, E.: Problèmes aux limites non homogènes et applications. Dunod (1968)
17. Maystre, D., Enoch, S.: Perfect lenses made with left-handed materials: Alice's mirror? J. Opt. Soc. Amer. **A 21**, 122–131 (2004)
18. McLean, W.: Strongly elliptic systems and boundary integral equations. Cambridge University Press, Cambridge (2000)
19. Monk, P.: Finite element methods for Maxwell's equations. Oxford University Press, New York (2003)
20. Nicaise, S., Venel, J.: A posteriori error estimates for a finite element approximation of transmission problems with sign changing coefficients. J. Comput. Appl. Math. **235**, 4272–4282 (2011)
21. Pendry, J.B.: Negative refraction makes a perfect lens. Physical Review Letters **85**(18), 3966–3969 (2000)
22. Ramdani, K.: Lignes supraconductrices: analyse mathématique et numérique. Ph.D. thesis, Université Paris 6 (1999)
23. Scott, R., Zhang, S.: Finite element interpolation of nonsmooth functions satisfying boundary conditions. Math. Comp. **54**(190), 483–493 (1990)