A hybridized Nitsche method for sign-changing elliptic PDEs

Erik Burman, Janosch Preuss, Alexandre Ern

A HYBRIDIZED NITSCHE METHOD FOR SIGN-CHANGING ELLIPTIC PDES

ERIK BURMAN, ALEXANDRE ERN AND JANOSCH PREUSS

Abstract. We present and analyze a hybridized Nitsche method for acoustic metamaterials. The use of a stabilized primal-dual formulation allows us to cope with the sign-changing nature of the problem and to prove optimal error estimates under a well-posedness assumption. The method can be used on arbitrary shape-regular meshes (fitted to material interfaces) and yields optimal convergence rates for smooth solutions. As an illustration, the method is applied to simulate a realistic acoustic cloaking device.

1. Introduction

Acoustic metamaterials have many interesting applications, for example they can be used to cloak objects from incoming sound waves. We refer the reader to the review paper [25] and to reference [30] for an overview of the mathematical theory behind cloaking. When it comes to simulating wave propagation inside metamaterials, one is faced with the challenge to deal with the sign-changing nature of the material coefficients. This leads to variational problems which are not coercive, even when the wavenumber vanishes, thereby precluding the application of many established numerical methods relying on this property.

To start with, let us briefly outline the current approaches in the numerical analysis literature which have been proposed to cope with sign-changing coefficients. To investigate well-posedness of these problems at the continuous level, the approach of T-coercivity, introduced in [7], has proved to be fruitful. We refer to [4] in which scalar problems are studied using this approach and to references [5, 6] for applications to time-harmonic Maxwell problems. T-coercivity is a reformulation of the inf-sup condition that is necessary and sufficient for well-posedness, here applied to a symmetric problem in Hilbert spaces (see, e.g., [28, Chap. 25]). The T operator allows one to keep track of the maximizer considered in the proof of the inf-sup condition.

It turns out that for problems involving sign-changing coefficients, certain assumptions on the mesh must be respected for the Galerkin discretization to work properly. These assumptions allow one to apply the T-coercivity framework, or, equivalently, to prove the inf-sup condition, at the discrete level, see, e.g., [22, 3]. Furthermore, we refer to [20, 31] for an application to eigenvalue problems

Date: May 7, 2024.

E.B. and J.P. acknowledge funding by EPSRC grant EP/V050400/1.

and to [21] for an application to multi-scale problems using the framework of localized orthogonal decomposition. Unfortunately, the above assumption on the mesh can be very challenging to realize if the interface at which the sign change occurs is geometrically complicated. The assumption even becomes impossible to realize for certain advanced discretization techniques, e.g., geometrically unfitted methods. In the recent preprint [32], an approach was suggested to avoid these stringent assumptions on the mesh. However, its implementation involves an intricate assembly procedure and a delicate construction of adapted quadrature rules, which have so far been neglected in the corresponding error analysis.

We are aware of two other approaches which can be applied on general meshes and are applicable to some more general settings. An optimization-based method was proposed in [1], see also [2], and very recently also an optimal control-based method [23] was devised which overcomes a potentially restrictive regularity condition required in the former reference. These methods can be proven to converge. Yet, it seems that convergence rates have not been proven for either of these methods. Hence, there is the need for more research on this subject.

In this article, we propose a stabilized finite element method for the numerical simulation of acoustic metamaterials. We proceed in the spirit of [8] in which a stabilized primal-dual stabilized framework is introduced to discretize non-coercive or ill-posed problems using the finite element method. This methodology has for example been applied to various unique continuation problems [9, 14, 15, 16, 17] for which it leads to optimal error estimates when combined with appropriate conditional stability estimates for the continuous problem [18]. In this article, we show how to apply this framework to treat problems with sign-changing coefficients. In particular, we derive optimal error estimates in the H^1-norm under the assumption that the continuous problem is well-posed. A hybridized Nitsche method is considered in which an interface variable is introduced to enforce the appropriate zero-jump conditions across the interface. Moreover, the discretization hinges, for both the primal and the dual variable, on continuous finite elements on both subdomains, and discontinuous finite elements for the interface variable. The hybridized Nitsche method has been introduced in [26] for an elliptic interface problem without sign-changes. We also notice that it is possible to employ a hybridized discontinuous Galerkin method in the subdomains, as done, e.g., in [10, 11] for unique continuation problems.

Let us briefly distinguish our method from the ones already proposed in the literature. In contrast to the plain Galerkin discretization, our approach is applicable on arbitrary shape-regular meshes (which are for simplicity assumed to be fitted to the interface). The price to pay is to solve for a larger problem since a dual variable is introduced and needs to be approximated as well. Furthermore, even though there already exist other methods in the literature [1, 23, 2] which can be applied on arbitrary shape-regular meshes, it seems that convergence rates for these methods have not been shown so far. As we are able to do so under the assumption of well-posedness, it thus seems that our method closes a gap in the

literature. Another relevant aspect is that the methods in references [1, 23, 2] appear to be applicable under weaker conditions than well-posedness. If we lower the assumption on the continuous problem to uniqueness, we still obtain convergence, but in a fairly weak norm which does not provide much information of practical interest about the quality of the approximate solution. Thus, the development of numerical methods under weaker assumptions on the continuous problem and the study of their optimal convergence rates remains a topic that deserves additional research.

The remainder of this article is structured as follows. We define the hybridized Nitsche method in Section 2 and prove its convergence in Section 3. In Section 4, we conduct numerical experiments to investigate the performance of the method for academic test cases and an actual metamaterial proposed in the physics literature. We finish in Section 5 with a conclusion and give some perspectives on further research.

2. Continuous and discrete settings

In this section, we present the continuous and discrete settings.

2.1. **Model problem.** We consider a domain $\Omega \subset \mathbb{R}^d$ for $d \in \{2,3\}$ split by an interface Γ into two subdomains $\Omega_\pm$ in such a way that $\overline{\Omega} = \overline{\Omega}_+ \cup \overline{\Omega}_-$ and $\partial\Omega_+ \cap \partial\Omega_- = \Gamma$. For a pair of constants $\sigma_+ > 0$, $\sigma_- < 0$, a pair of functions $\mu_\pm \in L^\infty(\Omega_\pm)$ representing reaction coefficients, and a pair of functions $f_\pm \in L^2(\Omega_\pm)$ representing source terms, we consider the following model problem: Find $u := (u_+, u_-) \in H^1(\Omega_+) \times H^1(\Omega_-)$ such that

$$\mathcal{L}_\pm(u_\pm) := -\nabla\cdot(\sigma_\pm\nabla u_\pm) + \mu_\pm u_\pm = f_\pm \quad \text{in } \Omega_\pm, \tag{2.1a}$$

$$u_\pm = 0 \quad \text{on } \partial\Omega_\pm \setminus \Gamma, \tag{2.1b}$$

together with the jump interface conditions

$$[\![u]\!]_\Gamma = 0 \quad \text{on } \Gamma, \tag{2.2a}$$

$$[\![\sigma\nabla u]\!]_\Gamma\cdot\mathbf{n}_\Gamma = 0 \quad \text{on } \Gamma, \tag{2.2b}$$

where

$$[\![u]\!]_\Gamma := u_+|_\Gamma - u_-|_\Gamma, \qquad [\![\sigma\nabla u]\!]_\Gamma := \sigma_+\nabla u_+|_\Gamma - \sigma_-\nabla u_-|_\Gamma. \tag{2.3}$$

Here, we denote the outward unit normal vector of the subdomain $\Omega_\pm$ by $\mathbf{n}_\pm$ and conventionally set $\mathbf{n}_\Gamma := \mathbf{n}_+$ pointing from Ω_+ into Ω_-. Note that there is no assumption on the sign of $\mu_\pm$, so that (2.1a) covers, in particular, the case of the Helmholtz equation. Owing to the jump condition (2.2a), it is meaningful to consider the function $\tilde{u} \in H^1(\Omega)$ such that $\tilde{u}|_{\Omega_\pm} = u_\pm$, and we notice that $\tilde{u} \in H^1_0(\Omega)$ owing to (2.1b). We slightly abuse the notation and write u instead of $\tilde{u}$. For later purpose, we set $\sigma_\flat := \min(\sigma_+, -\sigma_-)$ and $\sigma_\sharp := \max(\sigma_+, -\sigma_-)$.

To perform the error analysis, we will make one of the following two assumptions on the model problem (2.1)-(2.2). Clearly, Assumption 2 is stronger than Assumption 1.

Assumption 1 (Weak stability)**.** *The model problem* (2.1)-(2.2) *with zero right-hand side admits only the trivial solution in* $H^1_0(\Omega)$*.*

Assumption 2 (Strong stability)**.** *There is* C^{stab} *such that, for all* $f := (f_+, f_-) \in L^2(\Omega)$*, the model problem* (2.1)-(2.2) *admits a unique solution* $u \in H^1_0(\Omega)$*, and it fulfills the stability estimate*

$$\Big\{ \sum_{\pm} |\sigma_\pm| \, \|\nabla u_\pm\|^2_{\Omega_\pm} \Big\}^{\frac{1}{2}} \leqslant \sigma_\flat^{-\frac{1}{2}} C^{\mathrm{stab}} \, \|f\|_{H^{-1}(\Omega)}, \tag{2.4}$$

with $\|f\|_{H^{-1}(\Omega)} := \sup_{y \in H^1_0(\Omega)} \frac{\langle f, y\rangle}{\|\nabla y\|_\Omega}$ *and where the (nondimensional) constant* C^{stab} *can depend on the coefficients of* $\mathcal{L}_\pm$*.*

Generic constants that are independent of the problem parameters (including the size of Ω) will simply be called C in what follows. The value of C can change at each occurrence. We will characterize the dependence of the other named constants on the problem parameters explicitly.

We define the following functional spaces:

$$V_\pm := \{v_\pm \in H^1(\Omega_\pm) \mid v_\pm|_{\partial\Omega_\pm \setminus \Gamma} = 0\}, \quad V := V_+ \times V_-, \quad V_\Gamma := L^2(\Gamma), \quad \hat{V} := V \times V_\Gamma. \tag{2.5}$$

Note that for a function $v := (v_+, v_-) \in V$, we have in general $[\![v]\!]_\Gamma \neq 0$. For later use, we record here the following result.

Lemma 2.1 (Poincaré inequality)**.** *There is a (nondimensional) constant* $C^{\mathrm{P}} > 0$ *(independent of the problem parameters) such that, for all* $(z, z_\Gamma) \in \hat{V}$*, we have*

$$\|z\|^2_{L^2(\Omega)} = \sum_{\pm} \|z_\pm\|^2_{\Omega_\pm} \leqslant C^{\mathrm{P}} \ell_\Omega^2 \sum_{\pm} \Big\{ \|\nabla z_\pm\|^2_{\Omega_\pm} + \ell_\Omega^{-1} \|z_\pm - z_\Gamma\|^2_\Gamma \Big\}, \tag{2.6}$$

where ℓ_Ω *is a length scale associated with* Ω*, e.g., its diameter.*

Proof. Owing to the Peetre–Tartar theorem (see, e.g., [29, Theorem 2.1.3] or [27, Lemma A.20]), we infer that there is a constant C such that, for all $z := (z_+, z_-) \in V$,

$$\|z\|^2_{L^2(\Omega)} \leqslant C \ell_\Omega^2 \left(\sum_{\pm} \|\nabla z_\pm\|^2_{\Omega_\pm} + \ell_\Omega^{-1} \|[\![z]\!]_\Gamma\|^2_\Gamma \right).$$

The claim follows by adding and subtracting z_Γ in the jump term and using the triangle inequality. □

2.2. **Discrete setting.** We assume that Ω and $\Omega_\pm$ are all (Lipschitz) polygons/polyhedra that can be meshed exactly by a matching affine triangulation. Let $\mathcal{T}_h$ be such a triangulation, so that the subtriangulations $\mathcal{T}_h^\pm := \{T \in \mathcal{T}_h \mid T \subset \Omega_\pm\}$ fit the subdomains $\Omega_\pm$, respectively. Moreover, the set of all the facets $\mathcal{F}_h$ of $\mathcal{T}_h$ can be partitioned into

$$\mathcal{F}_h = \mathcal{F}_h^{\partial\Omega} \cup \mathcal{F}_h^\Gamma \cup \mathcal{F}_h^+ \cup \mathcal{F}_h^-, \tag{2.7}$$

where $\mathcal{F}_h^{\partial\Omega}$ and $\mathcal{F}_h^\Gamma$ denote the facets on $\partial\Omega$ and Γ, respectively, and $\mathcal{F}_h^\pm$ are the interior facets of $\Omega_\pm$, i.e., those facets that neither belong to $\partial\Omega$ nor to Γ. For all $F \in \mathcal{F}_h^\pm$, $\mathbf{n}_F$ denotes the unit normal to $F = T_1 \cap T_2$ conventionally pointing from T_1 to T_2, and the jump operator across F is to be interpreted as

$$[\![\nabla u_\pm]\!]_F = \nabla u_\pm|_{T_1}|_F - \nabla u_\pm|_{T_2}|_F.$$

To alleviate technicalities, we assume in what follows that $\mathcal{T}_h$ is quasi-uniform and use a single mesh size h. All what is said henceforth extends to shape-regular triangulations by localizing the mesh size in the error analysis. We use the notation $(v,w)_M := \int_M vw \,\mathrm{d}x$ and $\|v\|_M^2 := (v,v)_M$ to denote the L^2-scalar product and norm (with appropriate Lebesgue measure) over a subset $M \subset \overline{\Omega}$ which can be a collection of either mesh cells or mesh facets.

Let $l \geq 0$ be a polynomial degree, let $\mathbb{P}_l(T)$ be the space of d-variate polynomials of degree at most l on $T \in \mathcal{T}_h$, and let $\mathbb{P}_l(F)$ be the space of $(d-1)$-variate polynomials of degree at most l on $F \in \mathcal{F}_h^\Gamma$. We define the usual continuous finite element spaces on the subdomains:

$$V_{h,\pm}^l := \left\{v_{h,\pm} \in H^1(\Omega_\pm) \mid v_{h,\pm}|_T \in \mathbb{P}_l(T),\ \forall T \in \mathcal{T}_h^\pm \mid v_{h,\pm}|_{\partial\Omega_\pm\setminus\Gamma} = 0\right\} \subset V_\pm, \quad l \geq 1, \tag{2.8}$$

and the discontinuous finite element space on the interface:

$$V_{h,\Gamma}^l := \left\{v_h \in L^2(\Gamma) \mid v_h|_F \in \mathbb{P}_l(F),\ \forall F \in \mathcal{F}_h^\Gamma\right\} \subset V_\Gamma, \quad l \geq 0. \tag{2.9}$$

We will invoke the following trace inequality (see, e.g., [27, Sec. 12.2]):

$$h\,\|\nabla z_\pm\|_\Gamma^2 \leqslant C_\pm^{\mathrm{tr}}\,\|\nabla z_\pm\|_{\Omega_\pm}^2, \qquad \forall z_\pm \in V_{h,\pm}^l. \tag{2.10}$$

For a real number $s \geq 1$, we consider the broken Sobolev spaces $H^s(\mathcal{T}_h^\pm) := \{v_\pm \in L^2(\Omega_\pm) \mid v_\pm|_T \in H^s(T),\ \forall T \in \mathcal{T}_h^\pm\}$.

Let $\Pi_\pm^{h,l}$ and $\Pi_\Gamma^{h,l}$ denote interpolation operators into the spaces $V_{h,\pm}^l$, respectively V_Γ^l, with the expected (local) approximation properties:

$$|v_\pm - \Pi_\pm^{h,l}(v_\pm)|_{H^m(T)} \leqslant Ch^{s-m}|v_\pm|_{H^s(T)}, \qquad \forall T \in \mathcal{T}_h^\pm, \tag{2.11}$$

for all $s \in \{0,\dots,l+1\}$ and all $m \in \{0,\dots,s\}$, and

$$\|v_\pm - \Pi_\Gamma^{h,l}(v_\pm)\|_F \leqslant Ch^{s+\frac{1}{2}}|v_\pm|_{H^{s+1}(T^\pm)}, \qquad \forall F \in \mathcal{F}_h^\Gamma \cup \mathcal{F}_h^\pm, \tag{2.12}$$

for all $s \in \{0, \dots, l\}$, where $T^\pm \in \mathcal{T}_h^\pm$ is a corresponding volume element of which F is a fact. We may take $\Pi_\pm^{h,l}$ to be the Scott–Zhang operator [35] or the L^1-stable quasi-interpolation operators from [27, Chap. 22], and $\Pi_\Gamma^{h,l}$ to be the local L^2-projection.

2.3. **Hybridized Nitsche method.** We use a hybridized Nitsche method to discretize (2.1)–(2.2). We consider polynomial degrees $k, k^* \geq 1$ and $k_\Gamma^* \geq 0$. We assume that

$$k \geq \max(k^*, k_\Gamma^*), \qquad k_\Gamma^* \geq k - 1. \tag{2.13}$$

We define the following bilinear forms on $(V_{h,\pm}^k \times V_{h,\Gamma}^k) \times (V_{h,\pm}^{k^*} \times V_{h,\Gamma}^{k_\Gamma^*})$ (to alleviate the notation, we omit the subscript h for all the arguments):

$$\begin{aligned} a_\pm[(v_\pm, v_\Gamma); (y_\pm, y_\Gamma)] := (\sigma_\pm \nabla v_\pm, \nabla y_\pm)_{\Omega_\pm} + (\mu_\pm v_\pm, y_\pm)_{\Omega_\pm} - (\sigma_\pm \nabla v_\pm \cdot \mathbf{n}_\pm, y_\pm - y_\Gamma)_\Gamma \\ - (\sigma_\pm \nabla y_\pm \cdot \mathbf{n}_\pm, v_\pm - v_\Gamma)_\Gamma + \frac{\lambda_\pm |\sigma_\pm|}{h} (v_\pm - v_\Gamma, y_\pm - y_\Gamma)_\Gamma, \end{aligned} \tag{2.14a}$$

for user-dependent parameters $\lambda_\pm > 0$ to be chosen sufficiently large (see Theorem 3.3 below), as well as the stabilization bilinear forms

$$\begin{aligned} s_\pm[(v_\pm, v_\Gamma); (w_\pm, w_\Gamma)] := \sum_{T \in \mathcal{T}_h^\pm} \gamma_\pm^{\mathrm{LS}} \frac{h^2}{|\sigma_\pm|} \left(\mathcal{L}_\pm(v_\pm), \mathcal{L}_\pm(w_\pm)\right)_T + \sum_{F \in \mathcal{F}_h^\pm} |\sigma_\pm| \, h \left([\![\nabla v_\pm]\!]_F \cdot \mathbf{n}_F, [\![\nabla w_\pm]\!]_F \cdot \mathbf{n}_F\right)_F \\ + \frac{|\sigma_\pm|}{h} (v_\pm - v_\Gamma, w_\pm - w_\Gamma)_\Gamma, \end{aligned} \tag{2.14b}$$

$$s_\pm^*(z_\pm, y_\pm) := \gamma_\pm^* |\sigma_\pm| (\nabla z_\pm, \nabla y_\pm)_{\Omega_\pm} + \tilde{\mu}_\pm (z_\pm, y_\pm)_{\Omega_\pm}, \tag{2.14c}$$

for (nondimensional) stabilization parameters $\gamma_\pm^{\mathrm{LS}} > 0$, $\gamma_-^* = 1$ and $\gamma_+^* = 0$ (it is possible and potentially useful to take $\gamma_+^* \geqslant 0$, e.g., to facilitate the solution of the linear systems). The value of $\gamma_\pm^{\mathrm{LS}}$ is prescribed in Lemma 3.2 below. Moreover, $\tilde{\mu}_\pm := \|\mu_\pm^\ominus\|_{L^\infty(\Omega_\pm)}$ with the negative part operator $\mu_\pm^\ominus := \frac{1}{2}(|\mu_\pm| - \mu_\pm)$. Notice that $\tilde{\mu}_\pm = 0$ if $\mu_\pm \geq 0$. For later purpose, we also define $\mu_\pm^\oplus := \frac{1}{2}(|\mu_\pm| + \mu_\pm)$ and notice that $\mu_\pm = \mu_\pm^\oplus - \mu_\pm^\ominus$, and we set $\mu_{\infty,\pm} := \|\mu_\pm\|_{L^\infty(\Omega_\pm)}$.

To allow for a more compact notation, we define the discrete spaces

$$\hat{V}_h := (V_{h,+}^k \times V_{h,-}^k) \times V_{h,\Gamma}^k, \qquad \hat{V}_h^* := (V_{h,+}^{k^*} \times V_{h,-}^{k^*}) \times V_{h,\Gamma}^{k_\Gamma^*}, \tag{2.15}$$

and we use the notation $\hat{v} := (v, v_\Gamma)$, $v := (v_+, v_-)$ for a generic element of $\hat{V}_h$ and $\hat{z} := (z, z_\Gamma)$, $z := (z_+, z_-)$ for a generic element of $\hat{V}_h^*$ (we omit again the subscript h to simplify the notation). We define the following bilinear forms:

$$a[\hat{v}, \hat{z}] := \sum_\pm a_\pm[(v_\pm, v_\Gamma); (z_\pm, z_\Gamma)], \tag{2.16a}$$

as well as

$$s[\hat{v}, \hat{w}] := \sum_\pm s_\pm[(v_\pm, v_\Gamma); (w_\pm, w_\Gamma)], \qquad s^*[z, y] := \sum_\pm s_\pm^*(z_\pm; y_\pm), \tag{2.16b}$$

for all $\hat{v} \in \hat{V}_h$, all $\hat{w} \in \hat{V}_h$, all $\hat{z} \in \hat{V}_h^*$, and all $\hat{y} \in \hat{V}_h^*$. Putting everything together, we define the bilinear form

$$B[(\hat{v},\hat{z});(\hat{w},\hat{y})] := a[\hat{w},\hat{z}] + a[\hat{v},\hat{y}] + s[\hat{v},\hat{w}] - s^*(z,y). \tag{2.17}$$

The discrete problem consists of finding $(\hat{u}_h,\hat{z}_h) \in \hat{V}_h \times \hat{V}_h^*$ such that, for all $(\hat{w}_h,\hat{y}_h) \in \hat{V}_h \times \hat{V}_h^*$,

$$B[(\hat{u}_h,\hat{z}_h);(\hat{w}_h,\hat{y}_h)] = \sum_{\pm} \Big\{ (f_\pm, y_{h,\pm})_{\Omega_\pm} + \sum_{T\in\mathcal{T}_h^\pm} \gamma_\pm^{\mathrm{LS}} \frac{h^2}{|\sigma_\pm|} (f_\pm, \mathcal{L}_\pm(w_\pm))_T \Big\}. \tag{2.18}$$

Before proceeding with the error analysis, we record here the following integration by parts and consistency results.

Lemma 2.2 (Basic identity)**.** *The following holds for all $v_\pm \in V_{h,\pm}^l + H^s(\mathcal{T}_h^\pm)$, $s > \frac{3}{2}$, and all $w_\pm \in V_\pm$:*

$$\sum_{T\in\mathcal{T}_h^\pm} (\mathcal{L}_\pm(v_\pm), w_\pm)_T = (\sigma_\pm \nabla v_\pm, \nabla w_\pm)_{\Omega_\pm} + (\mu_\pm v_\pm, w_\pm)_{\Omega_\pm} - \sum_{F\in\mathcal{F}_h^\pm} (\sigma_\pm [\![\nabla v_\pm]\!]_F \cdot \mathbf{n}_F, w_\pm)_F - (\sigma_\pm \nabla v_\pm \cdot \mathbf{n}_\pm, w_\pm)_\Gamma. \tag{2.19}$$

Proof. Apply integration by parts. □

Lemma 2.3 (Consistency)**.** *Let $u \in H_0^1(\Omega)$ solve (2.1)-(2.2) and denote $\hat{u} = ((u_+,u_-),u_\Gamma)$ with $u_\pm := u|_{\Omega_\pm}$ and $u_\Gamma := u|_\Gamma$. Assume that $u_\pm \in H^s(\Omega_\pm)$, $s > \frac{3}{2}$. The following holds: For all $(\hat{w}_h,\hat{y}_h) \in \hat{V}_h \times \hat{V}_h^*$,*

$$B[(\hat{u},0),(\hat{w}_h,\hat{y}_h)] = \sum_{\pm} \Big\{ (f_\pm, y_{h,\pm})_{\Omega_\pm} + \sum_{T\in\mathcal{T}_h^\pm} \gamma_\pm^{\mathrm{LS}} \frac{h^2}{|\sigma_\pm|} (f_\pm, \mathcal{L}_\pm(w_\pm))_T \Big\}. \tag{2.20}$$

Proof. We have $B[(\hat{u},0),(\hat{w}_h,\hat{y}_h)] = a[\hat{u},\hat{y}_h] + s[\hat{u},\hat{w}_h]$. Summing (2.14a) over $\pm$, using that $[\![\sigma\nabla u]\!]\cdot\mathbf{n}_\Gamma = 0$, and invoking the basic identity from Lemma 2.2 gives $a[\hat{u},\hat{y}_h] = \sum_\pm (f_\pm, y_{h,\pm})_{\Omega_\pm}$. Moreover, we have $s[\hat{u},\hat{w}_h] = \sum_\pm \sum_{T\in\mathcal{T}_h^\pm} \gamma_\pm^{\mathrm{LS}} \frac{h^2}{|\sigma_\pm|} (f_\pm, \mathcal{L}_\pm(w_\pm))_T$ since $\mathcal{L}_\pm(u_\pm) = f_\pm$ in $\Omega_\pm$. This completes the proof. □

Remark 1 (Stabilization)**.** *It can be useful to scale some of the stabilization terms by some mesh-independent constants to enhance the preasymptotic convergence rate in the numerical experiments. However, for the numerical analysis, these scalings are of no importance and therefore set to unity.*

3. Error analysis

The error analysis proceeds in three steps. In Section 3.1, we prove the (discrete) inf-sup stability of the bilinear form B in a weak triple norm $|||\cdot|||$ defined in (3.2) below. A convergence result in this norm is derived in Section 3.2. The first two steps only require uniqueness of the solution to the continuous problem (2.1)-(2.2), i.e., the weak stability Assumption 1. An error estimate in the H^1-norm is derived in Section 3.3 under the stronger stability Assumption 2.

3.1. **Discrete stability.** Here, we deal with the stability of the discrete problem, and, for ease of notation, we drop the index h on the discrete variables. We define the stabilization (semi-)norm on $\hat{V}_h$ by $|\hat{v}|_s^2 := s[\hat{v},\hat{v}]$, i.e.,

$$|\hat{v}|_s^2 := \sum_{\pm}\Big\{ \sum_{T\in\mathcal{T}_h^\pm} \gamma_\pm^{\mathrm{LS}} \frac{h^2}{|\sigma_\pm|}\,\|\mathcal{L}_\pm(v_\pm)\|_T^2 + \sum_{F\in\mathcal{F}_h^\pm} |\sigma_\pm|\,h\,\|[\![\nabla v_\pm]\!]_F\cdot\mathbf{n}_F\|_F^2 + \frac{|\sigma_\pm|}{h}\,\|v_\pm - v_\Gamma\|_\Gamma^2\Big\}. \tag{3.1}$$

Furthermore, we define the following triple norm on $\hat{V}_h\times\hat{V}_h^*$:

$$\begin{aligned}|||(\hat{v},\hat{z})|||^2 :=\ & |\hat{v}|_s^2 + \sigma_\sharp^{-1} h\,\|[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma\|_\Gamma^2 \\ & + \sum_\pm\Big\{ |\sigma_\pm|\,\|\nabla z_\pm\|_{\Omega_\pm}^2 + \tilde{\mu}_\pm\|z_\pm\|_{\Omega_\pm}^2 + \frac{|\sigma_\pm|}{h}\,\|z_\pm - z_\Gamma\|_\Gamma^2\Big\}.\end{aligned} \tag{3.2}$$

Lemma 3.1 (Triple norm)**.** *Let Assumption 1 hold true. Then $|||\cdot|||$ defines a norm on $\hat{V}_h\times\hat{V}_h^*$.*

Proof. Let $(\hat{v},\hat{z})\in\hat{V}_h\times\hat{V}_h^*$ be such that $|||(\hat{v},\hat{z})||| = 0$. We need to prove that $v_\pm = 0$, $v_\Gamma = 0$, $z_\pm = 0$, and $z_\Gamma = 0$. Owing to the definition of the triple norm, we infer that $\mathcal{L}_\pm(v_\pm)|_T = 0$ for all $T\in\mathcal{T}_h^\pm$, $[\![\nabla v_\pm]\!]_F\cdot\mathbf{n}_F = 0$ for all $F\in\mathcal{F}_h^\pm$, $[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma = 0$, $v_+ = v_\Gamma = v_-$ on Γ, as well as $\nabla z_\pm = 0$ in $\Omega_\pm$ and $z_+ = z_\Gamma = z_-$ on Γ. Since $\hat{z}\in\hat{V}_h^*\subset\hat{V}$, the Poincaré inequality from Lemma 2.1 readily gives $z_\pm = 0$, and thus $z_\Gamma = 0$. Let us now deal with $\hat{v}$. Since $v_+ = v_\Gamma = v_-$, we infer that $[\![v]\!]_\Gamma = 0$, and we also have $[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma = 0$. Moreover, $\mathcal{L}_\pm(v_\pm)|_T = 0$, for all $T\in\mathcal{T}_h^\pm$, and $[\![\nabla v_\pm]\!]_F\cdot\mathbf{n}_F = 0$, for all $F\in\mathcal{F}_h^\pm$, imply that $\mathcal{L}_\pm(v_\pm)|_{\Omega_\pm} = 0$. Hence, $v = (v_+, v_-)$ solves (2.1)-(2.2) with right-hand side $f_\pm = 0$, so that v is zero by Assumption 1. Then, also $v_\Gamma = v_\pm|_\Gamma = 0$, and this completes the proof. □

Lemma 3.2 (Bound on stabilization)**.** *Set the least-squares stabilization parameters so that*

$$\gamma_\pm^{\mathrm{LS}} \le \gamma_{\sharp,\pm}^{\mathrm{LS}} := \big\{\max(1,(\sigma_\flat|\sigma_\pm|)^{-1}h^2\ell_\Omega^2\mu_{\infty,\pm}^2)\big\}^{-1}. \tag{3.3}$$

There is C^{s}, independent of h and the problem parameters, so that, for all $\hat{z}\in\hat{V}_h^*$,

$$|\hat{z}|_s^2 \leqslant C^{\mathrm{s}}\sum_\pm |\sigma_\pm|\Big\{\,\|\nabla z_\pm\|_{\Omega_\pm}^2 + \frac{1}{h}\,\|z_\pm - z_\Gamma\|_\Gamma^2\Big\}. \tag{3.4}$$

Proof. Invoking inverse inequalities, the assumption (3.3) and $h\le\ell_\Omega$, we infer that

$$\begin{aligned}&\sum_\pm\sum_{T\in\mathcal{T}_h^\pm}\gamma_\pm^{\mathrm{LS}}\frac{h^2}{|\sigma_\pm|}\,\|\mathcal{L}_\pm(z_\pm)\|_T^2 \leqslant 2\sum_\pm\sum_{T\in\mathcal{T}_h^\pm}\gamma_\pm^{\mathrm{LS}}\Big\{h^2\,|\sigma_\pm|\,\|\Delta z_\pm\|_T^2 + h^2\tfrac{\mu_{\infty,\pm}^2}{|\sigma_\pm|}\,\|z_\pm\|_T^2\Big\}\\
&\leqslant C\sum_\pm\gamma_\pm^{\mathrm{LS}}\Big\{\,|\sigma_\pm|\,\|\nabla z_\pm\|_{\Omega_\pm}^2 + h^2\tfrac{\mu_{\infty,\pm}^2}{|\sigma_\pm|}\,\|z_\pm\|_{\Omega_\pm}^2\Big\}\\
&\leqslant C\sum_\pm\gamma_\pm^{\mathrm{LS}}\max\Big(1,h^2\ell_\Omega^2\tfrac{\mu_{\infty,\pm}^2}{\sigma_\flat|\sigma_\pm|}\Big)\Big\{\,|\sigma_\pm|\,\|\nabla z_\pm\|_{\Omega_\pm}^2 + \sigma_\flat\ell_\Omega^{-2}\,\|z_\pm\|_{\Omega_\pm}^2\Big\}\\
&\leqslant C\sum_\pm\Big\{\,|\sigma_\pm|\,\|\nabla z_\pm\|_{\Omega_\pm}^2 + \sigma_\flat\ell_\Omega^{-2}\,\|z_\pm\|_{\Omega_\pm}^2\Big\}.\end{aligned}$$

We conclude by observing that

$$\sigma_\flat \ell_\Omega^{-2} \sum_{\pm} \|z_\pm\|_{\Omega_\pm}^2 \leqslant C^{\mathrm{P}} \sum_{\pm} |\sigma_\pm| \Big\{ \|\nabla z_\pm\|_{\Omega_\pm}^2 + \frac{1}{h} \|z_\pm - z_\Gamma\|_\Gamma^2 \Big\},$$

owing to the Poincaré inequality from Lemma 2.1 and $\sigma_\flat \leqslant |\sigma_\pm|$. □

Remark 2 (Stabilization parameter)**.** *Under the mild assumption* $h\ell_\Omega \mu_{\infty,\pm} \leqslant (\sigma_\flat |\sigma_\pm|)^{\frac{1}{2}}$*, one can set* $\gamma_\pm^{\mathrm{LS}} = 1$*.*

We can now state the main stability result of this section.

Theorem 3.3 (Inf-sup stability)**.** *Assume that the polynomial degrees satisfy* (2.13)*. Assume that* $\lambda_\pm \geq 2C_\pm^{\mathrm{tr}} + \frac{1}{2}$ *and that* $\gamma_\pm^{\mathrm{LS}}$ *are prescribed by* (3.3)*. The following holds:*

$$\inf_{(\hat v,\hat z)\in \hat V_h \times \hat V_h^*} \sup_{(\hat w,\hat y)\in \hat V_h \times \hat V_h^*} \frac{B[(\hat v,\hat z);(\hat w,\hat y)]}{|||(\hat v,\hat z)||| |||(\hat w,\hat y)|||} \geqslant \beta > 0, \tag{3.5}$$

where

$$\beta := \frac{1}{4}\big(2(\alpha^2 + C^{\mathrm{s}} + 2\max(C_\pm^{\mathrm{tr}}) + 2)\big)^{-\frac{1}{2}}, \quad \alpha := \max(2, C^{\mathrm{s}} + \tfrac{2}{3}\lambda_\sharp^2 + \tfrac{1}{4}), \quad \lambda_\sharp := \max(\lambda_\pm). \tag{3.6}$$

Consequently, under Assumption 1, the discrete problem (2.18) *is well-posed.*

Proof. (1) Let $\alpha > 0$ be chosen as in (3.6). We have

$$B[(\hat v,\hat z);(\alpha\hat v,-\alpha\hat z)] = \alpha s[\hat v,\hat v] + \alpha s^*[z,z] = \alpha |\hat v|_s^2 + \sum_{\pm} \alpha \Big\{ \gamma_\pm^* |\sigma_\pm| \|\nabla z_\pm\|_{\Omega_\pm}^2 + \tilde\mu_\pm \|z_\pm\|_{\Omega_\pm}^2 \Big\}. \tag{3.7}$$

(2) Since $k \geqslant \max\{k^*, k_\Gamma^*\}$ owing to (2.13), it is legitimate to test with $\hat w := \hat z$. We observe that

$$B[(\hat v,\hat z);(\hat z,0)] = a[\hat z,\hat z] + s[\hat v,\hat z].$$

On the one hand, classical manipulations invoking the trace inequality (2.10) and Young's inequality lead to

$$\begin{aligned} a_\pm[(z_\pm,z_\Gamma);(z_\pm,z_\Gamma)] &\geq \sigma_\pm \|\nabla z_\pm\|_{\Omega_\pm}^2 + (\mu_\pm z_\pm, z_\pm)_{\Omega_\pm} + \frac{\lambda_\pm |\sigma_\pm|}{h} \|z_\pm - z_\Gamma\|_\Gamma^2 \\ &\quad - 2|\sigma_\pm| (C_\pm^{\mathrm{tr}})^{\frac{1}{2}} h^{-\frac{1}{2}} \|\nabla z_\pm\|_{\Omega_\pm} \|z_\pm - z_\Gamma\|_\Gamma \\ &\geq \sigma'_\pm \|\nabla z_\pm\|_{\Omega_\pm}^2 + (\mu_\pm z_\pm, z_\pm)_{\Omega_\pm} + (\lambda_\pm - 2C_\pm^{\mathrm{tr}}) \frac{|\sigma_\pm|}{h} \|z_\pm - z_\Gamma\|_\Gamma^2, \end{aligned}$$

with $\sigma'_+ := \frac{1}{2}\sigma_+$ and $\sigma'_- := \frac{3}{2}\sigma_-$. Summing over both subdomains and using the assumption on $\lambda_\pm$, this gives

$$a[\hat z,\hat z] \geq \sum_{\pm} \Big\{ \sigma'_\pm \|\nabla z_\pm\|_{\Omega_\pm}^2 + (\mu_\pm z_\pm, z_\pm)_{\Omega_\pm} + \frac{1}{2} \frac{|\sigma_\pm|}{h} \|z_\pm - z_\Gamma\|_\Gamma^2 \Big\}. \tag{3.8}$$

On the other hand, owing to Young's inequality and the estimate (3.4), we infer that

$$s[\hat v,\hat z] \geq -C^{\mathrm{s}} |\hat v|_s^2 - \frac{1}{4C^{\mathrm{s}}} |\hat z|_s^2$$

$$\geq -C^{\rm s}|\hat{v}|_s^2 - \frac{1}{4}\sum_{\pm}|\sigma_\pm|\Big\{\|\nabla z_\pm\|_{\Omega_\pm}^2 + \frac{1}{h}\|z_\pm - z_\Gamma\|_\Gamma^2\Big\}.$$

Taking into account (3.8), this gives

$$B[(\hat{v},\hat{z});(\hat{z},0)] \geq -C^{\rm s}|\hat{v}|_s^2 + \sum_{\pm}\Big\{\sigma''_\pm\|\nabla z_\pm\|_{\Omega_\pm}^2 + (\mu_\pm z_\pm, z_\pm)_{\Omega_\pm} + \frac{1}{4}\frac{|\sigma_\pm|}{h}\|z_\pm - z_\Gamma\|_\Gamma^2, \tag{3.9}$$

with $\sigma''_+ := \frac{1}{4}\sigma_+$ and $\sigma''_- := \frac{7}{4}\sigma_-$.

(3) Since $k_\Gamma^* \geq k-1$ owing to (2.13), it is legitimate to test with $\hat{\zeta} := (0,\zeta_\Gamma)$ with $\zeta_\Gamma := \frac{h}{\sigma_\sharp}[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma$. Using the Cauchy–Schwarz inequality followed by Young's inequality and $|\sigma_\pm| \leqslant \sigma_\sharp$, and observing that $s^*(\hat{z},\hat{\zeta}) = 0$, this gives

$$\begin{aligned}
B[(\hat{v},\hat{z});(0,\hat{\zeta})] = a[\hat{v},\hat{\zeta}] &= \sum_{\pm}\Big\{(\sigma_\pm\nabla v_\pm\cdot\mathbf{n}_\pm,\zeta_\Gamma)_\Gamma - \frac{\lambda_\pm|\sigma_\pm|}{h}(v_\pm - v_\Gamma,\zeta_\Gamma)_\Gamma\Big\}\\
&= \frac{h}{\sigma_\sharp}\|[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma\|_\Gamma^2 - \sum_{\pm}\frac{\lambda_\pm|\sigma_\pm|}{\sigma_\sharp}(v_\pm - v_\Gamma,[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma)_\Gamma\\
&\geq \frac{1}{4}\frac{h}{\sigma_\sharp}\|[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma\|_\Gamma^2 - \sum_{\pm}\frac{2\lambda_\pm^2}{3}\frac{|\sigma_\pm|}{h}\|v_\pm - v_\Gamma\|_\Gamma^2.
\end{aligned}\tag{3.10}$$

(4) Combining (3.7), (3.9), and (3.10), we infer that

$$\begin{aligned}
B[(\hat{v},\hat{z});(\hat{w},\hat{y})] &\geq (\alpha - C^{\rm s} - \tfrac{2}{3}\lambda_\sharp^2)|\hat{v}|_s^2 + \frac{1}{4}\frac{h}{\sigma_\sharp}\|[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma\|_\Gamma^2\\
&\quad + \sum_{\pm}\Big\{(\alpha\gamma_\pm^*|\sigma_\pm| + \sigma''_\pm)\|\nabla z_\pm\|_{\Omega_\pm}^2 + \alpha\tilde{\mu}_\pm\|z_\pm\|_{\Omega_\pm}^2 + (\mu_\pm z_\pm, z_\pm)_{\Omega_\pm} + \frac{1}{4}\frac{|\sigma_\pm|}{h}\|z_\pm - z_\Gamma\|_\Gamma^2\Big\},
\end{aligned}$$

with $\hat{w} := \alpha\hat{v} + \hat{z}$, $\hat{y} := -\alpha\hat{z} + \hat{\zeta}$. The condition $\alpha \geq 2 \geq \frac{5}{4}$ ensures that $\alpha\tilde{\mu}_\pm\|z_\pm\|_{\Omega_\pm}^2 + (\mu_\pm z_\pm, z_\pm)_{\Omega_\pm} \geq \frac{1}{4}\tilde{\mu}_\pm\|z_\pm\|_{\Omega_\pm}^2$. Since we also have $\alpha \geq C^{\rm s} + \frac{2}{3}\lambda_\sharp^2 + \frac{1}{4}$, we obtain

$$\begin{aligned}
B[(\hat{v},\hat{z});(\hat{w},\hat{y})] &\geq \frac{1}{4}|\hat{v}|_s^2 + \frac{1}{4}\frac{h}{\sigma_\sharp}\|[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma\|_\Gamma^2\\
&\quad + \sum_{\pm}\Big\{(\alpha\gamma_\pm^*|\sigma_\pm| + \sigma''_\pm)\|\nabla z_\pm\|_{\Omega_\pm}^2 + \frac{1}{4}\tilde{\mu}_\pm\|z_\pm\|_{\Omega_\pm}^2 + \frac{1}{4}\frac{|\sigma_\pm|}{h}\|z_\pm - z_\Gamma\|_\Gamma^2\Big\}.
\end{aligned}$$

Finally, since $\alpha \geq 2$, we have $\alpha\gamma_\pm^*|\sigma_\pm| + \sigma''_\pm \geq \frac{1}{4}|\sigma_\pm|$. We infer that

$$B[(\hat{v},\hat{z});(\hat{w},\hat{y})] \geq \frac{1}{4}|||(\hat{v},\hat{z})|||^2.$$

(5) To conclude the proof of the inf-sup condition, we bound $|||(\hat{w},\hat{y})|||$. We have

$$|||(\hat{w},\hat{y})|||^2 \leq 2\alpha^2|||(\hat{v},\hat{z})|||^2 + 2|||\hat{z},\hat{\zeta}|||^2.$$

Moreover, since $\zeta_\pm = 0$, we have

$$\begin{aligned}
|||(\hat{z},\hat{\zeta})|||^2 &= |\hat{z}|_s^2 + \sigma_\sharp^{-1}h\,\|[\![\sigma\nabla z]\!]_\Gamma\cdot\mathbf{n}_\Gamma\|_\Gamma^2 + \sum_{\pm}\frac{|\sigma_\pm|}{h}\,\|\zeta_\Gamma\|_\Gamma^2\\
&\leq |\hat{z}|_s^2 + \sigma_\sharp^{-1}h\,\|[\![\sigma\nabla z]\!]_\Gamma\cdot\mathbf{n}_\Gamma\|_\Gamma^2 + 2\frac{h}{\sigma_\sharp}\|[\![\sigma\nabla v]\!]_\Gamma\cdot\mathbf{n}_\Gamma\|_\Gamma^2,
\end{aligned}$$

where we used the definition of ζ_Γ and $|\sigma_\pm| \leq \sigma_\sharp$. Invoking (3.4) to bound the first term on the right-hand side and the trace inequality (2.10) to bound the second one, we infer that

$$|||(\hat{z},\hat{\zeta})|||^2 \leq (C^{\mathrm{s}} + 2\max(C_\pm^{\mathrm{tr}}) + 2)|||(\hat{v},\hat{z})|||^2.$$

(6) Since the discrete problem (2.18) amounts to a square linear system, its well-posedness is a direct consequence of the inf-sup condition (3.5) together with Lemma 3.1. □

3.2. **Convergence in the triple norm.** For all $\hat{v} := (v, v_\Gamma)$ with $v_\pm \in V_{h,\pm}^l + H^2(\mathcal{T}_h^\pm)$ and $v_\Gamma \in L^2(\Gamma)$, we define the norm

$$\begin{aligned}
|||\hat{v}|||_\sharp^2 &= |||(\hat{v},0)|||^2 + \sum_\pm |\sigma_\pm| \Big\{ \|\nabla v_\pm\|_{\Omega_\pm}^2 + h\,\|\nabla v_\pm \cdot \mathbf{n}_\Gamma\|_\Gamma^2 + (\sigma_\flat|\sigma_\pm|)^{-1}\ell_\Omega^2\mu_{\infty,\pm}^2\,\|v_\pm\|_{\Omega_\pm}^2 \Big\} \\
&= |\hat{v}|_s^2 + \sigma_\sharp^{-1} h\,\|[\![\sigma\nabla v]\!]_\Gamma \cdot \mathbf{n}_\Gamma\|_\Gamma^2 \\
&\quad + \sum_\pm |\sigma_\pm| \Big\{ \|\nabla v_\pm\|_{\Omega_\pm}^2 + h\,\|\nabla v_\pm \cdot \mathbf{n}_\Gamma\|_\Gamma^2 + (\sigma_\flat|\sigma_\pm|)^{-1}\ell_\Omega^2\mu_{\infty,\pm}^2\,\|v_\pm\|_{\Omega_\pm}^2 \Big\}. \qquad (3.11)
\end{aligned}$$

Lemma 3.4 (Continuity)**.** *For all $\hat{v} = (v, v_\Gamma)$ with $v_\pm \in V_{h,\pm}^l + H^2(\mathcal{T}_h^\pm)$ and $v_\Gamma \in L^2(\Gamma)$, and for all $(\hat{w}_h, \hat{y}_h) \in \hat{V}_h \times \hat{V}_h^*$, the following holds:*

$$\big|B[(\hat{v},0),(\hat{w}_h,\hat{y}_h)]\big| \leqslant C^{\mathrm{bnd}}|||\hat{v}|||_\sharp|||(\hat{w}_h,\hat{y}_h)|||, \qquad (3.12)$$

with a constant C^{bnd} independent of h and the problem parameters.

Proof. We observe that $\big|B[(\hat{v},0),(\hat{w}_h,\hat{y}_h)]\big| \leq \big|a[\hat{v},\hat{y}_h]\big| + \big|s[\hat{v},\hat{w}_h]\big|$ and bound the two terms on the right-hand side.

(1) Bound on $a[\hat{v},\hat{y}_h]$. Since $\hat{y}_h \in \hat{V}_h^*$, we can use the trace inequality (2.10) to obtain

$$\sum_\pm (\sigma_\pm \nabla y_{h,\pm}\cdot\mathbf{n}_\pm, v_\pm - v_\Gamma)_\Gamma \leqslant (C_\pm^{\mathrm{tr}})^{\frac12} \sum_\pm \|\nabla y_{h,\pm}\|_{\Omega_\pm}\,|\sigma_\pm|\,h^{-\frac12}\,\|v_\pm - v_\Gamma\|_\Gamma \leqslant (C_\pm^{\mathrm{tr}})^{\frac12}|||\hat{v}|||_\sharp|||(0,\hat{y}_h)|||.$$

Moreover, we estimate the reaction term as follows:

$$\begin{aligned}
\sum_\pm (\mu_\pm v_\pm, y_{h,\pm})_{\Omega_\pm} &\leqslant \left(\sum_\pm \sigma_\flat^{-1}\ell_\Omega^2\mu_{\infty,\pm}^2\,\|v_\pm\|_{\Omega_\pm}^2\right)^{\frac12} \left(\sigma_\flat\ell_\Omega^{-2}\sum_\pm \|y_{h,\pm}\|_{\Omega_\pm}^2\right)^{\frac12} \\
&\leqslant \left(\sum_\pm \sigma_\flat^{-1}\ell_\Omega^2\mu_{\infty,\pm}^2\,\|v_\pm\|_{\Omega_\pm}^2\right)^{\frac12} (C^{\mathrm{P}})^{\frac12} \left(\sum_\pm |\sigma_\pm| \Big\{ \|\nabla y_{h,\pm}\|_{\Omega_\pm}^2 + h^{-1}\,\|y_{h,\pm} - y_{h,\Gamma}\|_\Gamma^2 \Big\}\right)^{\frac12} \\
&\leqslant \left(\sum_\pm \sigma_\flat^{-1}\ell_\Omega^2\mu_{\infty,\pm}^2\,\|v_\pm\|_{\Omega_\pm}^2\right)^{\frac12} (C^{\mathrm{P}})^{\frac12}|||(0,\hat{y}_h)|||,
\end{aligned}$$

where we used the Poincaré inequality from Lemma 2.1 and $\sigma_\flat \leqslant |\sigma_\pm|$. The remaining terms are easily bounded using the Cauchy-Schwarz inequality, leading to

$$\big|a[\hat{v},\hat{y}_h]\big| \leq C|||\hat{v}|||_\sharp|||(0,\hat{y}_h)|||.$$

(2) Bound on $s[\hat{v},\hat{w}_h]$. The Cauchy–Schwarz inequality readily gives

$$\big|s[\hat{v},\hat{w}_h]\big| \le |\hat{v}|_s|\hat{w}_h|_s \le |||\hat{v}|||_\sharp|||(\hat{w}_h,0)|||.$$

(3) Combining the bounds from Steps 1 and 2 yields the assertion since $|||(\hat{w}_h,\hat{y}_h)|||^2 = |||(\hat{w}_h,0)|||^2 + |||(0,\hat{y}_h)|||^2$. □

The next result demonstrates that the error in the triple norm $|||(\cdot,\cdot)|||$ is bounded by the best-approximation error in the augmented triple norm $|||\cdot|||_\sharp$.

Theorem 3.5 (Convergence in triple norm)**.** *Let $u \in H^1_0(\Omega)$ solve (2.1)-(2.2) and denote $\hat{u} := ((u_+,u_-),u_\Gamma)$ with $u_\pm := u|_{\Omega_\pm}$ and $u_\Gamma := u|_\Gamma$. Assume that $u_\pm \in H^s(\mathcal{T}_h^\pm)$, $s > \frac{3}{2}$. Let $(\hat{u}_h,\hat{z}_h) \in \hat{V}_h \times \hat{V}_h^*$ solve the discrete problem (2.18). Under Assumption 1, we have*

$$|||(\hat{u}-\hat{u}_h,\hat{z}_h)||| \leqslant \left(1+\frac{C^{\rm bnd}}{\beta}\right)\inf_{\hat{v}\in\hat{V}_h}|||\hat{u}-\hat{v}_h|||_\sharp. \tag{3.13}$$

Proof. Let $\hat{v}_h \in \hat{V}_h$ be arbitrary. For all $(\hat{w}_h,\hat{y}_h) \in \hat{V}_h \times \hat{V}_h^*$, the consistency result from Lemma 2.3 gives

$$\begin{aligned}B[(\hat{u}_h-\hat{v}_h,\hat{z}_h);(\hat{w}_h,\hat{y}_h)] &= B[(\hat{u}_h,\hat{z}_h);(\hat{w}_h,\hat{y}_h)] - B[(\hat{v}_h,0);(\hat{w}_h,\hat{y}_h)]\\ &= B[(\hat{u},0);(\hat{w}_h,\hat{y}_h)] - B[(\hat{v}_h,0);(\hat{w}_h,\hat{y}_h)] = B[(\hat{u}-\hat{v}_h,0);(\hat{w}_h,\hat{y}_h)].\end{aligned}$$

Using Lemma 3.4 then yields

$$B[(\hat{u}_h-\hat{v}_h,\hat{z}_h);(\hat{w}_h,\hat{y}_h)] \leqslant C^{\rm bnd}|||\hat{u}-\hat{v}_h|||_\sharp|||(\hat{w}_h,\hat{y}_h)|||.$$

In view of the inf-sup condition from (3.5), this implies

$$|||(\hat{u}_h-\hat{v}_h,\hat{z}_h)||| \leqslant \frac{C^{\rm bnd}}{\beta}|||\hat{u}-\hat{v}_h|||_\sharp.$$

The claim then follows from the triangle inequality $|||(\hat{u}-\hat{u}_h,\hat{z}_h)||| \leqslant |||(\hat{u}-\hat{v}_h,0)||| + |||(\hat{v}_h-\hat{u}_h,\hat{z}_h)|||$, observing that $|||(\hat{u}-\hat{v}_h,0)||| \le |||\hat{u}-\hat{v}_h|||_\sharp$, and recalling that $\hat{v}_h \in \hat{V}_h$ is arbitrary. □

Corollary 3.6 (Convergence rates for smooth solutions)**.** *Under the assumptions of Theorem 3.5 and assuming that $u_\pm \in H^{k+1}(\mathcal{T}_h^\pm)$, we have*

$$|||(\hat{u}-\hat{u}_h,\hat{z}_h)||| \leqslant CC^{\rm app}h^k\sum_\pm |u_\pm|_{H^{k+1}(\mathcal{T}_h^\pm)}, \tag{3.14}$$

with $C^{\rm app} := \sigma_\sharp^{\frac{1}{2}}\max(1,(\sigma_\flat|\sigma_\pm|)^{-\frac{1}{2}}h\ell_\Omega\mu_{\infty,\pm})$.

Proof. Combine the estimate (3.13) with the approximation properties (2.11)-(2.12) (we used $\gamma_\pm^{\rm LS} \le 1$ owing to (3.3) to simplify some expressions). □

3.3. **Convergence in H^1.** To derive convergence rates in the H^1-norm, we will assume well-posedness of the continuous problem (see Assumption 2). In particular, we would like to apply the stability estimate (2.4) to the error $u - u_h$. However, this is not possible since the discrete solution is not continuous across the interface. To overcome this issue, we will need to interpolate the discrete solution into an $H^1(\Omega)$-conforming space. The following lemma ensures that the corresponding interpolation error can be bounded by the jump terms over the interface which are controlled by the triple norm.

Lemma 3.7 (Discontinuous to continuous interpolation)**.** *There exists an interpolation operator Π_h^{c} from $\hat{V}_h$ into a subspace of $H^1(\Omega)$ such that, for all $\hat{w}_h \in \hat{V}_h$,*

$$\sum_{\pm} |\sigma_\pm| \Big\{ h^{-2} \left\| \Pi_h^{\mathrm{c}}(\hat{w}_h) - w_{h,\pm} \right\|_{\Omega_\pm}^2 + \left\| \nabla \left(\Pi_h^{\mathrm{c}}(\hat{w}_h) - w_{h,\pm} \right) \right\|_{\Omega_\pm}^2 \Big\} \leqslant C^{\Pi} \sum_{\pm} \frac{|\sigma_\pm|}{h} \left\| w_{h,\pm} - w_{h,\Gamma} \right\|_{\Gamma}^2, \quad (3.15)$$

with a constant C^{Π} independent of h and the problem parameters.

Proof. The claim follows from [13, Lemma 3.2 & 5.3 and Remark 3.2] with the following minor modification in the construction: the value of $\Pi_h^{\mathrm{c}}(\hat{w}_h)$ at the mesh nodes located on Γ is prescribed by using $w_{h,\Gamma}$. □

We can now state our main error estimate.

Theorem 3.8. *Let $u \in H_0^1(\Omega)$ solve (2.1)-(2.2) and denote $\hat{u} := ((u_+, u_-), u_\Gamma)$ with $u_\pm := u|_{\Omega_\pm}$ and $u_\Gamma := u|_\Gamma$. Assume that $u_\pm \in H^s(\mathcal{T}_h^\pm)$, $s > \frac{3}{2}$. Let $(\hat{u}_h, \hat{z}_h) \in \hat{V}_h \times \hat{V}_h^*$ solve the discrete problem (2.18). Assume that $\gamma_\pm^{\mathrm{LS}} \geq \frac{1}{2}\gamma_{\sharp,\pm}^{\mathrm{LS}}$ with $\gamma_{\sharp,\pm}^{\mathrm{LS}}$ defined in (3.3). Under Assumption 2, we have*

$$\Big\{ \sum_{\pm} |\sigma_\pm| \left\| \nabla(u - u_{h,\pm}) \right\|_{\Omega_\pm}^2 \Big\}^{\frac{1}{2}} \leqslant C C^{\mathrm{E}} \inf_{\hat{v} \in \hat{V}_h} |||\hat{u} - \hat{v}_h|||_\sharp, \quad (3.16)$$

with $C^{\mathrm{E}} := \max(1, C^{\mathrm{stab}}) \left(\frac{\sigma_\sharp}{\sigma_\flat}\right)^{\frac{1}{2}} \max(1, |\sigma_\pm|^{-\frac{1}{2}} \ell_\Omega \mu_{\infty,\pm}^{\frac{1}{2}}, (\sigma_\flat |\sigma_\pm|)^{-\frac{1}{2}} h \ell_\Omega \mu_{\infty,\pm})$ and C^{stab} defined in Assumption 2. Notice that $C^{\mathrm{E}} \leqslant \max(1, C^{\mathrm{stab}}) \left(\frac{\sigma_\sharp}{\sigma_\flat}\right)^{\frac{1}{2}} \max(1, (\sigma_\flat |\sigma_\pm|)^{-\frac{1}{2}} \ell_\Omega^2 \mu_{\infty,\pm})$.

Proof. We define the linear form $r_h \in H^{-1}(\Omega)$ so that, for all $y \in H_0^1(\Omega)$,

$$\langle r_h, y \rangle := \sum_{\pm} \Big\{ \left(\sigma_\pm \nabla \left(\Pi_h^{\mathrm{c}}(\hat{u}_h) - u_\pm \right), \nabla y \right)_{\Omega_\pm} + \left(\mu_\pm (\Pi_h^{\mathrm{c}}(\hat{u}_h) - u_\pm), y \right)_{\Omega_\pm} \Big\}. \quad (3.17)$$

It is shown in Lemma 3.9 below that

$$\|r_h\|_{H^{-1}(\Omega)} \leq C C^{\mathrm{R}} |||(\hat{u} - \hat{u}_h, \hat{z}_h)|||,$$

with $C^{\mathrm{R}} := \sigma_\sharp^{\frac{1}{2}} \max(1, |\sigma_\pm|^{-\frac{1}{2}} \ell_\Omega \mu_{\infty,\pm}^{\frac{1}{2}}, (\sigma_\flat |\sigma_\pm|)^{-\frac{1}{2}} h \ell_\Omega \mu_{\infty,\pm})$. Invoking the stability estimate (2.4) from Assumption 2, we infer that

$$\begin{aligned} \Big\{ \sum_{\pm} |\sigma_\pm| \left\| \nabla(\Pi_h^{\mathrm{c}}(\hat{u}_h) - u) \right\|_{\Omega_\pm}^2 \Big\}^{\frac{1}{2}} &\leqslant \sigma_\flat^{-\frac{1}{2}} C^{\mathrm{stab}} \left\| r_h \right\|_{H^{-1}(\Omega)} \\ &\leqslant C \sigma_\flat^{-\frac{1}{2}} C^{\mathrm{stab}} C^{\mathrm{R}} |||(\hat{u} - \hat{u}_h, \hat{z}_h)|||. \end{aligned}$$

By applying the triangle inequality

$$\Big\{\sum_{\pm}|\sigma_{\pm}|\,\|\nabla(u-u_{h,\pm})\|_{\Omega_\pm}^2\Big\}^{\frac12}\leqslant\Big\{2\sum_{\pm}|\sigma_{\pm}|\,\|\nabla(\Pi_h^{\mathrm{c}}(\hat u_h)-u_{h,\pm})\|_{\Omega_\pm}^2\Big\}^{\frac12}$$
$$+\Big\{2\sum_{\pm}|\sigma_{\pm}|\,\|\nabla(\Pi_h^{\mathrm{c}}(\hat u_h)-u)\|_{\Omega_\pm}^2\Big\}^{\frac12},$$

and using the estimate (3.15) for the first term on the right-hand side, we obtain (we absorb the constant C^{Π} in the generic constant C and use that $\sigma_\flat^{-\frac12}C^{\mathrm{R}}\geq 1$)

$$\Big\{\sum_{\pm}|\sigma_{\pm}|\,\|\nabla(u-u_{h,\pm})\|_{\Omega_\pm}^2\Big\}^{\frac12}\leqslant C\max(1,C^{\mathrm{stab}})\sigma_\flat^{-\frac12}C^{\mathrm{R}}|||(\hat u-\hat u_h,\hat z_h)|||.$$

The claim follows by invoking the error estimate in the triple norm from Theorem 3.5. □

Lemma 3.9 (Bound on r_h)**.** *Let $r_h\in H^{-1}(\Omega)$ be defined in* (3.17)*. The following holds:*

$$\|r_h\|_{H^{-1}(\Omega)}\leq CC^{\mathrm{R}}|||(\hat u-\hat u_h,\hat z_h)|||, \tag{3.18}$$

with $C^{\mathrm{R}}:=\sigma_\sharp^{\frac12}\max(1,|\sigma_\pm|^{-\frac12}\ell_\Omega\mu_{\infty,\pm}^{\frac12},(\sigma_\flat|\sigma_\pm|)^{-\frac12}h\ell_\Omega\mu_{\infty,\pm})$.

Proof. Let $y\in H_0^1(\Omega)$. Set $y_\pm:=y|_{\Omega_\pm}$ and $y_\Gamma:=y|_\Gamma$. We have

$$\langle r_h,y\rangle=\sum_{\pm}\Big\{(\sigma_\pm\nabla(\Pi_h^{\mathrm{c}}(\hat u_h)-u_{h,\pm}),\nabla y)_{\Omega_\pm}+(\mu_\pm(\Pi_h^{\mathrm{c}}(\hat u_h)-u_{h,\pm}),y)_{\Omega_\pm}\Big\}$$
$$+\sum_{\pm}\Big\{(\sigma_\pm\nabla(u_{h,\pm}-u_\pm),\nabla y)_{\Omega_\pm}+(\mu_\pm(u_{h,\pm}-u_\pm),y)_{\Omega_\pm}\Big\}=:I_1+\tilde I.$$

We need to decompose $\tilde I$ further. Invoking the basic identity (2.19) from Lemma 2.2 gives

$$\tilde I=\sum_{\pm}\Big\{\sum_{T\in\mathcal T_h^\pm}(\mathcal L_\pm(u_{h,\pm}-u_\pm),y_\pm)_T+\sum_{F\in\mathcal F_h^\pm}(\sigma_\pm[\![\nabla u_h]\!]_F\cdot\mathbf n_F,y_\pm)_F\Big\}+([\![\sigma\nabla u_h]\!]_\Gamma\cdot\mathbf n_\Gamma,y_\Gamma)_\Gamma.$$

Moreover, using the variational formulation (2.18) with $\hat w_h=0$ and since $\mathcal L_\pm(u_\pm)=f_\pm$, we obtain, for all $\hat y_h\in\hat V_h$,

$$\sum_{\pm}\sum_{T\in\mathcal T_h^\pm}(\mathcal L_\pm(u_\pm),y_{h,\pm})_T=B[(\hat u_h,\hat z_h),(0,\hat y_h)]=a[\hat u_h,\hat y_h]-s^*[\hat z_h,\hat y_h].$$

Invoking again the basic identity (2.19) to transform the expression of $a[\hat u_h,\hat y_h]$, we infer that

$$0=\sum_{\pm}\Big\{\sum_{T\in\mathcal T_h^\pm}(\mathcal L_\pm(u_{h,\pm}-u_\pm),y_{h,\pm})_T+\sum_{F\in\mathcal F_h^\pm}(\sigma_\pm[\![\nabla u_h]\!]_F\cdot\mathbf n_F,y_{h,\pm})_F\Big\}+([\![\sigma\nabla u_h]\!]_\Gamma\cdot\mathbf n_\Gamma,y_{h,\Gamma})_\Gamma$$
$$+\sum_{\pm}\Big\{\frac{\lambda_\pm|\sigma_\pm|}{h}(u_{h,\pm}-u_{h,\Gamma},y_{h,\pm}-y_{h,\Gamma})_\Gamma-(\sigma_\pm\nabla y_{h,\pm}\cdot\mathbf n_\pm,u_{h,\pm}-u_{h,\Gamma})_\Gamma\Big\}-s^*[\hat z_h,\hat y_h].$$

Subtracting the above identity from the above expression for $\tilde I$ and adding to I_1 gives $\langle r_h,y\rangle=\sum_{j\in\{1:7\}}I_j$, with I_1 defined above and

$$I_2:=\sum_{\pm}\sum_{T\in\mathcal T_h^\pm}(\mathcal L_\pm(u_{h,\pm}-u_\pm),y_\pm-y_{h,\pm})_T,\quad I_3:=\sum_{\pm}\sum_{F\in\mathcal F_h^\pm}(\sigma_\pm[\![\nabla(u_h-u)]\!]_F\cdot\mathbf n_F,y_\pm-y_{h,\pm})_F,$$

$$I_4 := (\llbracket \sigma \nabla (u_h - u) \rrbracket_\Gamma \cdot \mathbf{n}_\Gamma, y_\Gamma - y_{h,\Gamma})_\Gamma\,, \qquad I_5 := \sum_\pm \frac{\lambda_\pm\,|\sigma_\pm|}{h}(u_{h,\pm} - u_{h,\Gamma}, y_{h,\Gamma} - y_{h,\pm})_\Gamma$$

$$I_6 := \sum_\pm (\sigma_\pm \nabla y_{h,\pm} \cdot \mathbf{n}_\pm, u_{h,\pm} - u_{h,\Gamma})_\Gamma, \qquad I_7 := s^*[\hat z_h, \hat y_h],$$

where we used that $\llbracket \nabla u \rrbracket_F \cdot \mathbf{n}_F = 0$ in the expression of I_3 and that $\llbracket \sigma \nabla u \rrbracket_\Gamma \cdot \mathbf{n}_\Gamma = 0$ in the expression of I_4. We now choose

$$y_{h,\pm} := \Pi_\pm^{h,k^*}(y_\pm), \qquad y_{h,\Gamma} := \Pi_\pm^{h,k_\Gamma^*}(y_\Gamma)$$

and bound the seven terms I_j in terms of $\|\nabla y\|_\Omega$ and $|||(\hat u - \hat u_h, \hat z_h)|||$. Invoking the Cauchy–Schwarz inequality, the Poincaré inequality for y, and the approximation property (3.15) gives (we absorb the constant C^Π in the generic constant C)

$$\begin{aligned}|I_1| &\leqslant C\left(\sum_\pm |\sigma_\pm|\,\|\nabla(\Pi_h^{\mathrm c}(\hat u_h) - u_{h,\pm})\|_{\Omega_\pm} + \ell_\Omega \mu_{\infty,\pm}\,\|\Pi_h^{\mathrm c}(\hat u_h) - u_{h,\pm}\|_{\Omega_\pm}\right)\|\nabla y\|_\Omega \\ &\leqslant C\sigma_\sharp^{\frac12}\max(1, |\sigma_\pm|^{-1} h \ell_\Omega \mu_{\infty,\pm})|\hat u - \hat u_h|_s\,\|\nabla y\|_\Omega\,,\end{aligned}$$

where the second bound uses that $u_\pm|_\Gamma = u_\Gamma$. Using the Cauchy–Schwarz inequality, the (low-order) approximation properties of $\Pi_\pm^{h,k^*}$ and $\Pi_\pm^{h,k_\Gamma^*}$, and observing that $1 \leq (\gamma_\pm^{\mathrm{LS}})^{-1} \leq 2\max(1, \frac{h^2\ell_\Omega^2\mu_{\infty,\pm}^2}{\sigma_\flat|\sigma_\pm|})$ (see (3.3)), we infer that

$$|I_2 + I_3 + I_4| \leq C\sigma_\sharp^{\frac12}\max(1, (\sigma_\flat|\sigma_\pm|)^{-\frac12} h\ell_\Omega\mu_{\infty,\pm})|||(\hat u - \hat u_h, 0)|||\,\|\nabla y\|_\Omega\,.$$

To bound I_5, we write

$$I_5 = \sum_\pm \frac{\lambda_\pm\,|\sigma_\pm|}{h}(u_{h,\pm} - u_{h,\Gamma} - u_\pm + u_\Gamma, y_\pm - y_\Gamma - y_{h,\pm} + y_{h,\Gamma})_\Gamma,$$

so that, reasoning as above gives (we absorb the factor $\lambda_\sharp^{\frac12}$ in the generic constant C)

$$|I_5| \leq C\sigma_\sharp^{\frac12}|||(\hat u - \hat u_h, 0)|||\,\|\nabla y\|_\Omega\,.$$

From the discrete trace inequality (2.10) (applied now on functions in $V_{h,\pm}^{k^*}$), we have $h\,\|\nabla y_{h,\pm}\|_\Gamma \leqslant C_\pm^{\mathrm{tr}}\,\|\nabla y_{h,\pm}\|_{\Omega_\pm}$. Since $I_6 = \sum_\pm(\sigma_\pm\nabla y_{h,\pm}\cdot\mathbf{n}_\pm, u_{h,\pm} - u_{h,\Gamma} - (u_\pm - u_\Gamma))_\Gamma$, we infer that (we absorb the factor $C_\pm^{\mathrm{tr}}$ in the generic constant C)

$$|I_6| \leqslant C\sigma_\sharp^{\frac12}|||(\hat u - \hat u_h, 0)|||\,\|\nabla y\|_\Omega\,,$$

where we used the H^1-stability of $\Pi_\pm^{h,k^*}$. Finally, the Cauchy–Schwarz inequality the Poincaré inequality, and the H^1-stability of $\Pi_\pm^{h,k^*}$ give

$$|I_7| \leqslant C\sigma_\sharp^{\frac12}\max(1, |\sigma_\pm|^{-\frac12}\ell_\Omega\mu_{\infty,\pm}^{\frac12})|||(0, \hat z_h)|||\,\|\nabla y\|_\Omega\,.$$

Combining the above bounds yields the assertion since $\sigma_\flat \leq |\sigma_\pm|$. □

Finally, convergence rates are inferred by proceeding as in the proof of Corollary 3.6.

Corollary 3.10 (Convergence rates for smooth solutions)**.** *Under the assumptions of Theorem 3.8 and assuming that* $u_\pm \in H^{k+1}(\mathcal{T}_h^\pm)$*, we have*

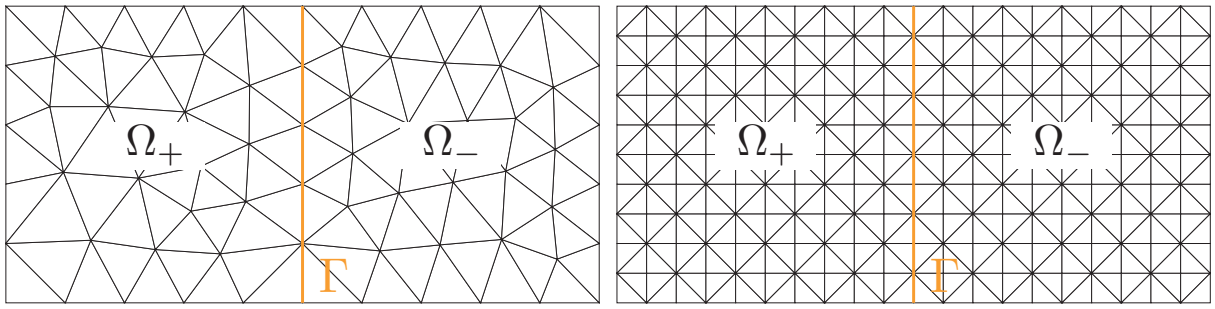
$$\Big\{ \sum_{\pm} |\sigma_\pm| \, \|\nabla(u - u_{h,\pm})\|^2_{\Omega_\pm} \Big\}^{\frac{1}{2}} \leqslant C C^{\mathrm{app}} C^{\mathrm{E}} h^k \sum_{\pm} |u_\pm|_{H^{k+1}(\mathcal{T}_h^\pm)}, \tag{3.19}$$

with C^{app} *defined in Corollary 3.6.*

4. Numerical experiments

We present a series of numerical experiments to investigate the performance of the proposed method. We start in Section 4.1 with an academic test case and proceed in Section 4.2 to an acoustic cloaking device proposed in [36]. Finally, in Section 4.3, we explore the setting in which the stability Assumption 2 is violated.

The numerical experiments have been implemented in the finite element library `NGSolve` [33, 34]. Reproduction material for the presented experiments is avalaible at zenodo in the form of a docker image `https://doi.org/10.5281/zenodo.11067991`.



(A) Unstructured mesh (B) Symmetric mesh

Figure 1. Meshes for the cavity problem.

4.1. **Symmetric cavity.** The symmetric cavity problem is one of the main benchmark tests in the numerical analysis literature on problems with sign-changing coefficients, see e.g. [22, 1, 23]. This is a pure diffusion problem, i.e. $\mu_\pm := 0$, in which the subdomains are given by $\Omega_+ := (-1, 0) \times (0, 1)$ and $\Omega_- := (0, 1) \times (0, 1)$. It is known from [22, Section 3.3] that Assumption 2 holds true for $\sigma_+ + \sigma_- \neq 0$. In this case, the solution is given by

$$u(x, y) = \begin{cases} ((x+1)^2 - (\sigma_+ + \sigma_-)^{-1}(2\sigma_+ + \sigma_-)(x+1)) \sin(\pi y) & \text{in } \Omega_+, \\ (\sigma_+ + \sigma_-)^{-1}\sigma_+(x-1) \sin(\pi y) & \text{in } \Omega_-. \end{cases} \tag{4.1}$$

The problem becomes numerically more difficult to handle if the critical contrast $\sigma_+/\sigma_- = -1$ is approached. Let us first test the hybridized Nitsche method for some contrasts away from the critical

value. The relative H^1-errors on a sequence of unstructured meshes, see Fig. 1a for an example, are displayed in Fig. 2. We have taken here the minimal choice for the dual stabilization, $k^* = 1$ and $k^*_\Gamma = k - 1$. Clearly, the convergence rates predicted by Theorem 3.8 are achieved.

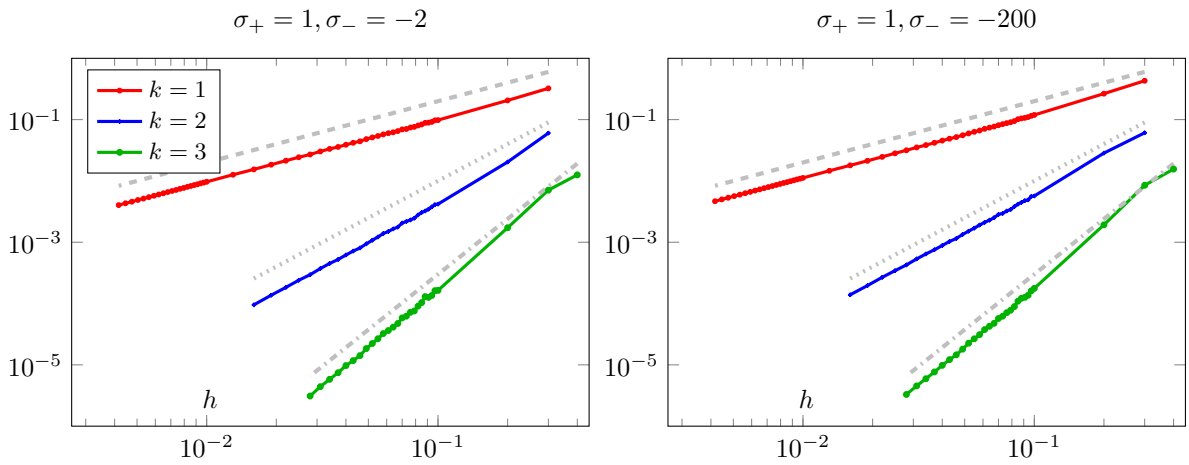


FIGURE 2. Relative error $\|u - u_h\|_{H^1(\Omega_+ \cup \Omega_-)} / \|u\|_{H^1(\Omega_+ \cup \Omega_-)}$ using $u_h|_{\Omega_\pm} = u_{h,\pm}$ for the symmetric cavity in the well-posed case on unstructured meshes.

Now let us approach the critical contrast by setting $\sigma_+ = 1$ and $\sigma_- = -1.001$. It is well-known, see, e.g., [22, 1], that a naive Galerkin discretization obtained from the bilinear form $(u, v) \mapsto \int_\Omega \sigma \nabla u \cdot \nabla v$ with $\sigma|_{\Omega_\pm} = \sigma_\pm$, suffers from instabilities on unstructured meshes, yet yields optimal convergence rates on symmetric meshes of the form shown in Fig. 1b. We compare the performance of our stabilized method with the plain Galerkin discretization in Fig. 3. We observe that the Galerkin method is unstable on unstructured meshes, whereas the stabilized method shows a fairly robust performance and optimal convergence rates. We have taken the full dual order $k^* = k$ and $k^*_\Gamma = k$ for this example since it was observed that this allows to reduce the size of the stabilization parameters without affecting the numerical stability, thereby leading to reduced errors on coarse meshes. Instead, on symmetric meshes, the error for the stabilized method and coarse mesh sizes is slightly higher than that produced by the Galerkin method. As our method contains a built-in Galerkin discretization, it is indeed possible to deactivate the stabilization on symmetric meshes and achieve the same errors as the plain Galerkin method. However, to keep the comparison fair, we did not resort to this device here.

4.2. **Metamaterial.** Let us proceed to a more realistic test case. To this end, we consider the acoustic cloaking device from [36]. The equation for a point source at $x_0 \in \mathbb{R}^2$ takes the form

$$-\nabla \cdot (\sigma(r) \nabla u) - \mu(r) u = \delta_{x_0},$$

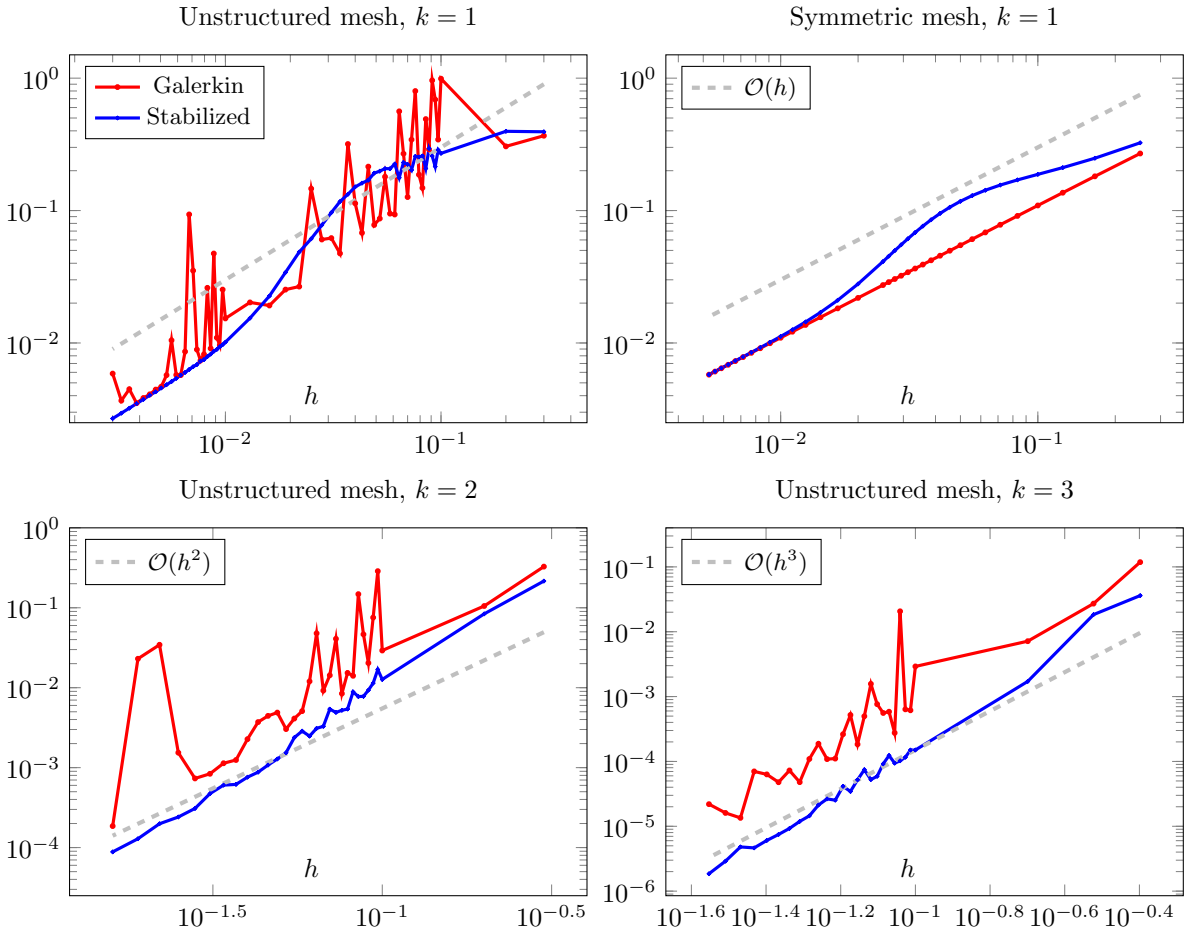


FIGURE 3. Relative error $\|u - u_h\|_{H^1(\Omega_+ \cup \Omega_-)} \,/\, \|u\|_{H^1(\Omega_+ \cup \Omega_-)}$ for the symmetric cavity at contrast $\sigma_+/\sigma_- = -1.001$.

for a piecewise constant σ and a radially varying μ given by

$$\sigma(r) := \begin{cases} 1/\rho_0 & \text{for } 0 < r < a, \\ 1/\rho_1 & \text{for } a < r < b, \\ -1/\rho_1 & \text{for } b < r < c, \\ 1/\rho_0 & \text{for } c < r, \end{cases} \qquad \mu(r) := \begin{cases} (\omega^2/\kappa_0)(b/a)^4 & \text{for } 0 < r < a, \\ -(\omega^2/\kappa_1)(b/r)^4 & \text{for } a < r < b, \\ \omega^2/\kappa_1 & \text{for } b < r < c, \\ \omega^2/\kappa_0 & \text{for } c < r, \end{cases}$$

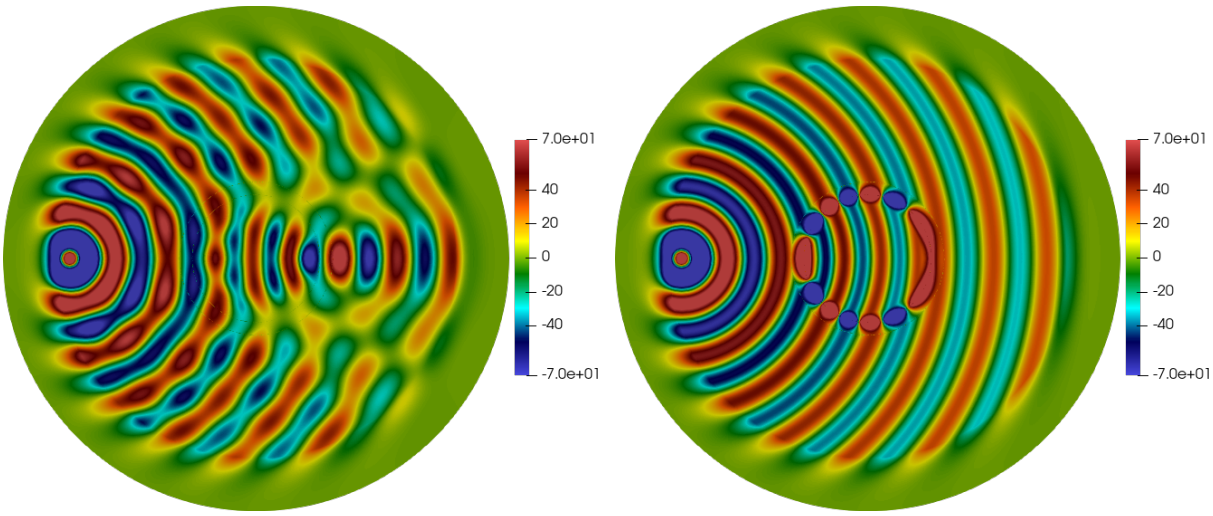
where $r := \sqrt{x^2 + y^2}$. The parameters are as follows:

$$\kappa_0 := 2.19 \text{ GPa}, \ \kappa_1 := 0.48\kappa_0, \ \rho_0 := 998 \text{ kg/m}^3, \ \rho_1 := \rho_0,$$

$$a := 1.0 \text{ m}, \ b := 1.2 \text{ m}, \ c := 1.44 \text{ m}.$$

The source is positioned at $x_0 := (-3.5, 0)$ m, and we truncate the computational domain by a perfectly matched layer (PML), see, e.g., [24], active in the region $\Omega_{\text{PML}} := \{r \in (3.75, 4.75) \text{ m}\}$. To

ensure that the PML will work as usual to attenuate the waves, we take the full dual order and replace the primal stabilization terms in the PML layer by a scaled Tikhonov regularization of the form $h^{2k}(u_{h,+}, v_{h,+})_{\Omega_{\mathrm{PML}}}$. A sketch of the computational setup is displayed in the upper left panel of Fig. 5.



(A) without cloak (B) with cloak

FIGURE 4. Cloaking using a metamaterial. The function values in the cloaking domain are actually very high, but have been truncated here to ± 70 to aid presentation.

Before we embark on convergence studies, let us first study the effectiveness of the metamaterial, which is located in the layer $a < r < c$. The idea of this layer is to cloak the object contained in the region $r < a$. Since $\mu(r)$ differs for $r < a$ and $r > c$, we expect to see traces of the object in the propagating waves if no metamaterial is present, i.e., if $\sigma(r) = 1/\rho_0$ and $\mu(r) = \omega^2/\kappa_0$ uniformly for $r > a$. This is confirmed in Fig. 4a which shows the numerical solution computed with the stabilized method using $\omega = 2\pi \cdot 1481.5$ Hz. The waves are indeed strongly disturbed by the inhomogeneity. However, if the cloak is activated, as shown in Fig. 4b, the object becomes invisible.

If the cloaking works perfectly, we would expect that the solution in the exterior of the cloaked region $r > c$ is given by a spherical wave emanating from the point source at x_0 given by

$$u = \frac{i\rho_0}{4} H_0^{(1)} \left(\omega \sqrt{\frac{\rho_0}{\kappa_0}} \, \|x - x_0\| \right), \tag{4.2}$$

where $H_0^{(1)}$ denotes the Hankel function of the first kind of order zero. We will measure the convergence of the numerical solution against this reference solution in the two circular regions

$$\Omega_i := \{c < r < 1.7\}, \quad \Omega_e := \{1.7 < r < 3.25\},$$

as sketched in the upper left panel of Fig. 5. Note that Ω_i represents a buffer layer between Ω_e and the interface.

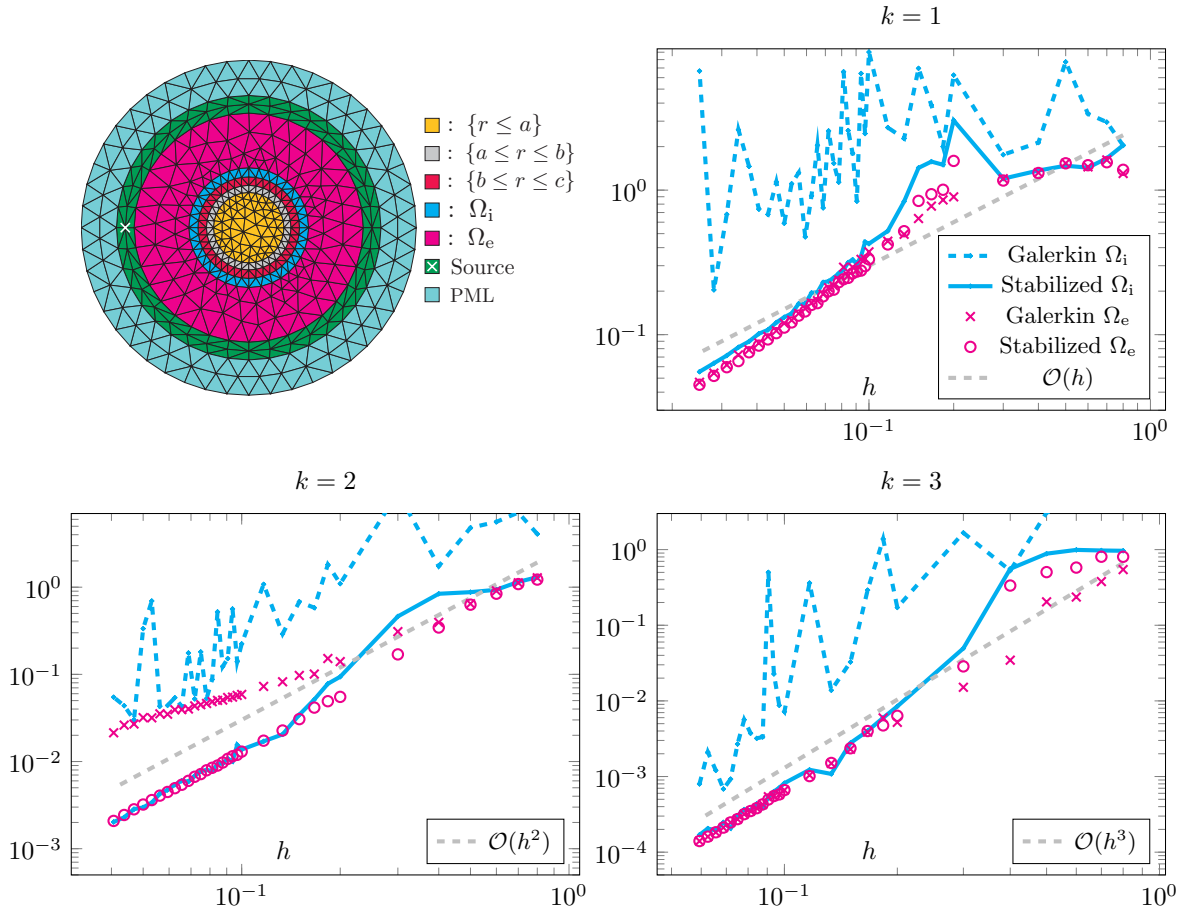


FIGURE 5. Upper left panel: computational setup. Other panels: Relative H^1-error in Ω_i and Ω_e with respect to the reference solution (4.2) as a function of the mesh size. Here, h refers to an upper bound on the mesh size in $\{r < a\} \cup \{r > c\}$. The mesh size within the metamaterial $\{a \leqslant r \leqslant c\}$ is bounded from above by $h/3$.

The relative H^1-erors for the stabilized and the plain Galerkin methods in the various regions are displayed Fig. 5. We observe that the Galerkin method shows severe instabilities in the region Ω_i. Interestingly, for $k = 2$, these errors even pollute the approximation in the exterior region Ω_e, so that the expected convergence rate is not reached there. For $k = 1$ (not shown here), $k = 3$ and $k = 4$, however, the Galerkin solution converges at the optimal rate in Ω_e. In contrast, the stabilized method always converges at the optimal rate in both regions. These results clearly demonstrate the suitability and relevance of the proposed method for the reliable simulation of physically interesting metamaterials.

4.3. **Non-symmetric cavity at super-critical contrast.** Let us finally consider a test case for which it is known, see [22, Section 3.3] and [2, Section 4.2], that the well-posedness Assumption 2 fails. As in

Section 4.1, we set $\Omega_+ := (-1,0)\times(0,1)$, but now $\Omega_- := (0,3)\times(0,1)$ which breaks the symmetry of the cavity. We set $\mu_\pm = 0$ and consider the super-critical contrast $\sigma_+ = 1, \sigma_- = -1$. In this case, the problem is not Fredholm, yet the weaker Assumption 1 of injectivity holds true. We consider then the exact solution from [2, Section 7.2] defined as

$$u(x,y) := \begin{cases} (2(x+1)^2 - 5(x+1))\sin(\pi y) & \text{in } \Omega_+, \\ (x-3)\sin(\pi y) & \text{in } \Omega_-. \end{cases}$$

Notice that the theoretical convergence result derived in Theorem 3.8 cannot be applied here as it

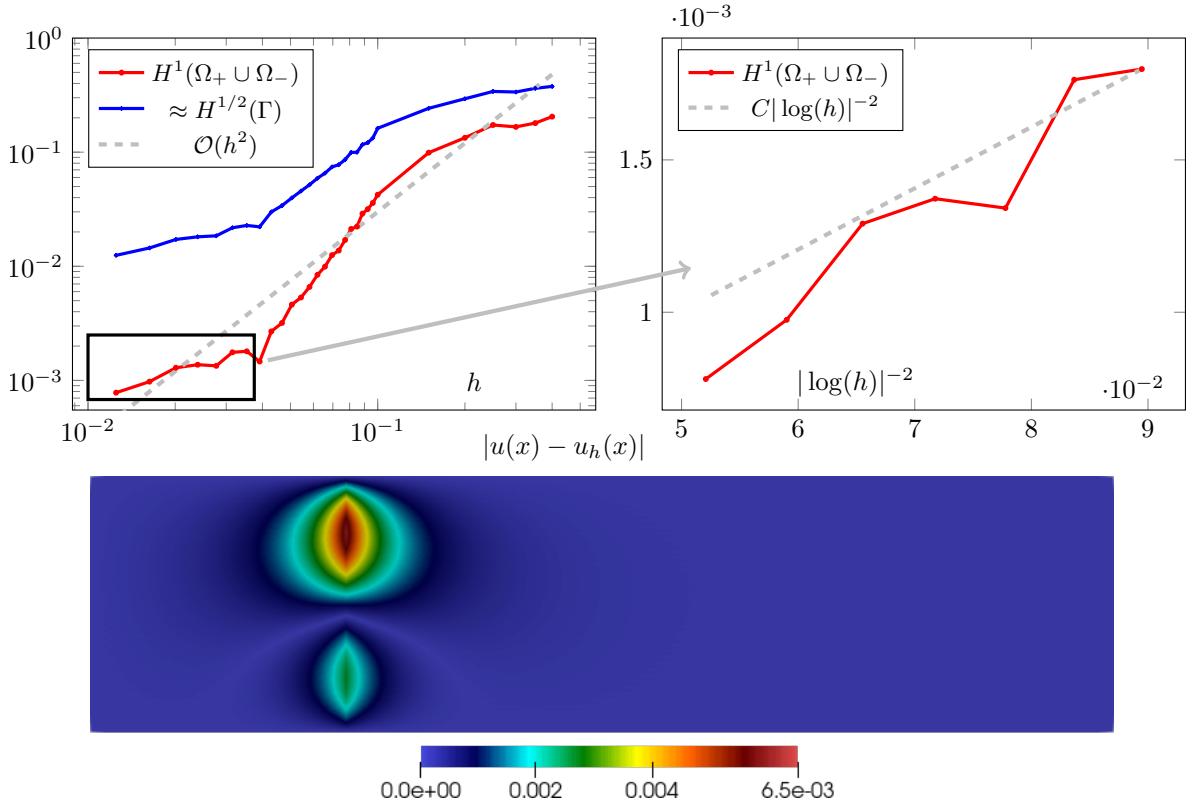


FIGURE 6. Results for the non-symmetric cavity at super-critical contrast obtained with the stabilized method using $k = 2$. The red line displays the relative error in the bulk, while the blue line shows the relative errors for the measure $\|u_\Gamma - u_{h,\Gamma}\|_\Gamma^{\frac{1}{2}} \|\nabla_\Gamma(u_\Gamma - u_{h,\Gamma})\|_\Gamma^{\frac{1}{2}}$. Note that the x-axis in the right plot is given in terms of $|\log(h)|^{-2}$ and that the error looks almost like a straight line in this plot. The figure on the bottom displays the absolute error on a mesh of size $h \approx 0.028$.

relies on the well-posedness of the continuous problem, but we do have the convergence result in the weaker triple norm as established in Theorem 3.5. We would like to investigate whether convergence in H^1 can still be obtained with our method. The relative H^1-errors using polynomial degree $k = 2$ are

displayed in Fig. 6. To enhance the accuracy of the method, we add to the primal stabilization bilinear form $s_\pm(u_\pm, v_\pm)$ the following second-order jump terms:

$$\sum_{F \in \mathcal{F}_h^\pm} h^3 \left([\![D^2 u_\pm]\!]_F, [\![D^2 v_\pm]\!]_F \right)_F ,$$

where $D^2 u_\pm$ denotes the Hessian. This is reasonable since the solutions in the subdomains are smooth. Despite this modification, we were only able to obtain second-order convergence up to a mesh-size of about $h \approx 0.05$. The right plot in Fig. 6 shows that on finer meshes the convergence deteriorates to a logarithmic rate. Let us mention that without adding the second-order jump penalties, the error even seemed to stagnate. The plot of the absolute error in Fig. 6 shows that this behavior seems to stem from a poor approximation close to the interface. To investigate this more qualitatively, we plot additionally the error measure (see blue line in Fig. 6)

$$\left(\frac{\|u_\Gamma - u_{h,\Gamma}\|_\Gamma \, \|\nabla_\Gamma (u_\Gamma - u_{h,\Gamma})\|_\Gamma}{\|u_\Gamma\|_\Gamma \, \|\nabla_\Gamma u_\Gamma\|_\Gamma} \right)^{\frac{1}{2}}$$

which, according to the Gagliardo–Nirenberg inequality, gives an upper bound on $\|u_\Gamma - u_{h,\Gamma}\|_{H^{\frac{1}{2}}(\Gamma)}$. We observe the same logarithmic convergence behavior in the asymptotic regime as for the error in the bulk and a significantly higher error constant. This supports the conclusion that the difficulty of capturing the behavior of the solution at the interface is responsible for the observed degeneration of the convergence rate. Overall, this example shows that our method can still be applied when the well-posedness assumption is violated, but the convergence rates proven in Theorem 3.8 are no longer valid.

5. Conclusion

We presented a stabilized primal-dual finite element method for the numerical approximation of acoustic metamaterials and proved optimal error estimates under a well-posedness assumption on the continuous problem. The method can be applied on general shape-regular meshes and has shown to be reliable and accurate in numerical experiments featuring physically relevant metamaterials. These results motivate to conduct further research on the proposed method. The following extensions seem interesting:

- At the discrete level, the solutions in the subdomains are coupled only via a trace variable defined on the interface Γ. This suggest to solve the linear system efficiently via static condensation, which seems particularly interesting for metamaterials composed of several layers.
- In the analysis and implementation of the method, we have assumed that the mesh fits the interface. However, we expect that the method can be extended to unfitted discretizations by combining the techniques from [12] and [19].

- To prove error estimates in H^1, we assumed well-posedness of the continuous problem. In the numerical experiment of Section 4.3 where this assumption is violated, the method appears to convergence asymptotically at a logarithmic rate. It would be interesting to investigate theoretically whether this rate is optimal.
- We restricted our attention to acoustic metamaterials. An extension of the method to Maxwell's equations, which would be required to capture the electromagnetic characteristics of general metamaterials, is certainly of practical relevance.

References

[1] A. Abdulle, M. E. Huber, and S. Lemaire. An optimization-based numerical method for diffusion problems with sign-changing coefficients. *Comptes Rendus Math.*, 355(4):472–478, 2017.

[2] A. Abdulle and S. Lemaire. An optimization-based method for sign-changing elliptic PDEs. *ESAIM: Math. Model. Numer. Anal.*, 2024. to appear.

[3] A.-S. Bonnet-Ben Dhia, C. Carvalho, and P. Ciarlet, Jr. Mesh requirements for the finite element approximation of problems with sign-changing coefficients. *Numer. Math.*, 138(4):801–838, 2018.

[4] A.-S. Bonnet-Ben Dhia, L. Chesnel, and P. Ciarlet, Jr. T-coercivity for scalar interface problems between dielectrics and metamaterials. *ESAIM: Math. Model. Numer. Anal.*, 46(6):1363–1387, 2012.

[5] A.-S. Bonnet-Ben Dhia, L. Chesnel, and P. Ciarlet, Jr. T-coercivity for the Maxwell problem with sign-changing coefficients. *Comm. Partial Diff. Equ.*, 39(6):1007–1031, 2014.

[6] A.-S. Bonnet-Ben Dhia, L. Chesnel, and P. Ciarlet, Jr. Two-dimensional Maxwell's equations with sign-changing coefficients. *Appl. Numer. Math.*, 79:29–41, 2014. Workshop on Numerical Electromagnetics and Industrial Applications (NELIA 2011).

[7] A.-S. Bonnet-Ben Dhia, P. Ciarlet, Jr., and C. M. Zwölf. Time harmonic wave diffraction problems in materials with sign-shifting coefficients. *J. Comp. Appl. Math.*, 234(6):1912–1919, 2010. Eighth International Conference on Mathematical and Numerical Aspects of Waves (Waves 2007).

[8] E. Burman. Stabilized finite element methods for nonsymmetric, noncoercive, and ill-posed problems. Part I: elliptic equations. *SIAM J. Sci. Comput.*, 35(6):A2752–A2780, 2013.

[9] E. Burman. Error estimates for stabilized finite element methods applied to ill-posed problems. *Comptes Rendus Math.*, 352(7):655–659, 2014.

[10] E. Burman, G. Delay, and A. Ern. A hybridized high-order method for unique continuation subject to the Helmholtz equation. *SIAM J. Numer. Anal.*, 59(5):2368–2392, 2021.

[11] E. Burman, G. Delay, and A. Ern. The unique continuation problem for the heat equation discretized with a high-order space-time nonconforming method. *SIAM J. Numer. Anal.*, 61(5):2534–2557, 2023.

[12] E. Burman, D. Elfverson, P. Hansbo, M. G. Larson, and K. Larsson. Hybridized CutFEM for elliptic interface problems. *SIAM J. Sci. Comput.*, 41(5):A3354–A3380, 2019.

[13] E. Burman and A. Ern. Continuous interior penalty *hp*-finite element methods for advection and advection-diffusion equations. *Math. Comp.*, 76(259):1119–1140, 2007.

[14] E. Burman, P. Hansbo, and M. G. Larson. Solving ill-posed control problems by stabilized finite element methods: an alternative to Tikhonov regularization. *Inverse Problems*, 34(3):035004, jan 2018.

[15] E. Burman, M. G. Larson, and L. Oksanen. Primal-dual mixed finite element methods for the elliptic Cauchy problem. *SIAM J. Numer. Anal.*, 56(6):3480–3509, 2018.

[16] E. Burman, M. Nechita, and L. Oksanen. Unique continuation for the Helmholtz equation using stabilized finite element methods. *J. Math. Pures Appl.*, 129:1–22, 2019.

[17] E. Burman, M. Nechita, and L. Oksanen. A stabilized finite element method for inverse problems subject to the convection–diffusion equation. I: diffusion-dominated regime. *Numer. Math.*, 144(3):451–477, 2020.

[18] E. Burman, M. Nechita, and L. Oksanen. Optimal approximation of unique continuation. To appear in *Found. Comput. Math.*, 2023.

[19] E. Burman and J. Preuss. Unique continuation for an elliptic interface problem using unfitted isoparametric finite elements. arXiv:2307.05210, 2023.

[20] C. Carvalho, L. Chesnel, and P. Ciarlet, Jr. Eigenvalue problems with sign-changing coefficients. *Comptes Rendus Math.*, 355(6):671–675, 2017.

[21] T. Chaumont-Frelet and B. Verfürth. A generalized finite element method for problems with sign-changing coefficients. *ESAIM: Math. Model. Numer. Anal.*, 55(3):939–967, 2021.

[22] L. Chesnel and P. Ciarlet, Jr. T-coercivity and continuous Galerkin methods: application to transmission problems with sign changing coefficients. *Numer. Math.*, 124(1):1–29, 2013.

[23] P. Ciarlet, Jr., D. Lassounon, and M. Rihani. An optimal control-based numerical method for scalar transmission problems with sign-changing coefficients. *SIAM J. Numer. Anal.*, 61(3):1316–1339, 2023.

[24] F. Collino and P. Monk. The Perfectly Matched Layer in curvilinear coordinates. *SIAM J. Sci. Comput.*, 19(6):2061–2090, Nov. 1998.

[25] S. A. Cummer, J. Christensen, and A. Alù. Controlling sound with acoustic metamaterials. *Nature Reviews Materials*, 1(3):1–13, 2016.

[26] H. Egger. A class of hybrid mortar finite element methods for interface problems with non-matching meshes. *preprint AICES-2009-2, Jan*, 2009.

[27] A. Ern and J.-L. Guermond. *Finite Elements I: Approximation and Interpolation*, volume 72 of *Texts in Applied Mathematics*. Springer Nature, Cham, Switzerland, 2021.

[28] A. Ern and J.-L. Guermond. *Finite elements. II. Galerkin approximation, elliptic and mixed PDEs*, volume 73 of *Texts in Applied Mathematics*. Springer, Cham, 2021.

[29] V. Girault and P.-A. Raviart. *Finite element methods for Navier-Stokes equations: Theory and algorithms*, volume 5. 1986.

[30] A. Greenleaf, Y. Kurylev, M. Lassas, and G. Uhlmann. Cloaking devices, electromagnetic wormholes, and transformation optics. *SIAM Rev.*, 51(1):3–33, 2009.

[31] M. Halla. On the approximation of dispersive electromagnetic eigenvalue problems in two dimensions. *IMA J. Numer. Anal.*, 43(1):535–559, 12 2021.

[32] M. Halla, T. Hohage, and F. Oberender. A new numerical method for scalar eigenvalue problems in heterogeneous, dispersive, sign-changing materials, 2024.

[33] J. Schöberl. NETGEN An advancing front 2D/3D-mesh generator based on abstract rules. *Comput. Vis. Sci.*, 1(1):41–52, 1997.

[34] J. Schöberl. C++11 implementation of finite elements in NGSolve. Technical report, ASC-2014-30, Institute for Analysis and Scientific Computing, September 2014.

[35] R. L. Scott and S. Zhang. Finite element interpolation of nonsmooth functions satisfying boundary conditions. *Math. Comp.*, 54(190):483–493, 1990.

[36] X. Zhu, B. Liang, W. Kan, X. Zou, and J. Cheng. Acoustic cloaking by a superlens with single-negative materials. *Phys. Rev. Lett.*, 106(1):014301, 2011.