# Multiscale modeling for a class of high-contrast heterogeneous sign-changing problems

Changqing Ye, Xingguang Jin, Patrick Ciarlet, Eric T. Chung

# Multiscale modeling for a class of high-contrast heterogeneous sign-changing problems

Changqing Ye*[1], Xingguang Jin[1], Patrick Ciarlet Jr.[2], and Eric T. Chung[1]

[1]Department of Mathematics, The Chinese University of Hong Kong, Shatin, Hong Kong SAR, China.
[2]POEMS, CNRS, INRIA, ENSTA Paris, Institut Polytechnique de Paris, 91120 Palaiseau, France.

July 24, 2024

**Abstract**

The mathematical formulation of sign-changing problems involves a linear second-order partial differential equation in the divergence form, where the coefficient can assume positive and negative values in different subdomains. These problems find their physical background in negative-index metamaterials, either as inclusions embedded into common materials as the matrix or vice versa. In this paper, we propose a numerical method based on the constraint energy minimizing generalized multiscale finite element method (CEM-GMsFEM) specifically designed for sign-changing problems. The construction of auxiliary spaces in the original CEM-GMsFEM is tailored to accommodate the sign-changing setting. The numerical results demonstrate the effectiveness of the proposed method in handling sophisticated coefficient profiles and the robustness of coefficient contrast ratios. Under several technical assumptions and by applying the T-coercivity theory, we establish the inf-sup stability and provide an a priori error estimate for the proposed method.

## 1 Introduction

Metamaterials greatly expand the design space of materials by offering unconventional properties that are not found in nature. Typically, metamaterials are created by assembling common materials periodically. The exotic characteristics of metamaterials manifest at the effective medium level, which can strikingly differ from the intrinsic properties of the constituent materials. Some notable examples include auxetics, exhibiting a negative Poisson's ratio [2, 38]; pentamode materials, characterized by a vanishing shear modulus [42, 37]; and negative-index materials, displaying negative electric and magnetic permeability [48, 46]. From a mathematical perspective, the emergence of metamaterials prompts a reconsideration of the well-posedness theory of Partial Differential Equations (PDEs), as the coefficients in PDEs may fall outside the conventional range for coercivity. In this context, we focus on the so-called sign-changing problem, which is rooted in the background that negative-index materials are embedded into a common medium, or vice versa.

The mathematical nature of the sign-changing problem is a linear second-order PDE in the divergence form, where the coefficient allows both positive and negative values in different subdomains, with a discontinuity across the interface between the subdomains. The well-posedness of the model problem is generally questionable due to the absence of uniform strict positivity of the coefficient. To address this, T-coercivity

*cqye@math.cuhk.edu.hk

was introduced by Bonnet-BenDhia, Ciarlet Jr., and Zwölf in their work [13], providing a systematic approach to study the well-posedness of the model problem. The key idea behind T-coercivity is to devise a bijective map $\mathcal{T}$ that enables the inf-sup (Banach–Nečas–Babuška) condition to hold trivially. The construction of $\mathcal{T}$ only relies on the geometric information of interfaces that separate the subdomains with different signs of the coefficient. The main statement of the T-coercivity theory is that if the contrast ratio between the negative and positive coefficients is sufficiently large, then the model problem becomes well-posed. In [10], the sharpest condition of the contrast ratio ensuring well-posedness on certain simple interface problems was derived. Further extensions of the T-coercivity theory encompass Helmholtz-like problems [24], time-harmonic Maxwell equations [12, 11], eigenvalue problems [18], and mixed problems [6].

Although analytical solutions offer us a profound mathematical insight into sign-changing problems, numerical methods are indispensable for practical applications. The T-coercivity theory justifies the well-posedness of the PDE models, which also guides the design of suitable approximations of the original problem. An intriguing aspect lies in the compatibility between $\mathcal{T}$ at the continuous level and the discretization level. As an immediate result, it is emphasized in [20] that, to obtain an optimal convergence rate, meshes near flat interfaces should be symmetric. Later, a new treatment at the corners of interfaces was proposed in [9], leading to meshing rules for an arbitrary polygonal interface. Considering the low regularity of the solution due to the heterogeneous coefficient, a posteriori error analysis was conducted in [44] and in [25], which provides a reliable error estimator for adaptive mesh refinement routines [47]. Beyond classic continuous Galerkin methods, Chung and Ciarlet Jr. introduced a staggered discontinuous Galerkin method in [21], accompanied by stability and convergence analysis.

The nature of heterogeneous coefficients in sign-changing problems motivates the application of multiscale computational methods. Pioneered by Hou and Wu in [32], the methodology of incorporating model information into the construction of finite element spaces, coined as MsFEMs, has garnered significant attention. As a discretization scheme, an advantage of MsFEMs is that the meshes are not required to resolve the heterogeneity of the coefficient, although the implementation of MsFEMs indeed relies on a pair of nested meshes. To relieve the rigidity from boundary conditions in constructing multiscale bases, the oversampling technique is introduced in [32] and subsequently proved to improve convergence rates (ref. [30, 29]). The accuracy of MsFEMs, to some extent, may deteriorate when the coefficient violates the scale-separation assumption, as evidenced by the convergence theories in [33, 51, 43]. To address this issue, the Generalized Multiscale Finite Element Method (GMsFEM) was proposed by Efendiev, Galvis, and Hou in [28]. GMsFEMs leverage spectral decomposition to perform dimension reduction for the online space, exhibiting superior performance when dealing with high-contrast and channel-like coefficient profiles [23]. The first construction of multiscale bases capable of achieving the theoretically best approximation property for general $L^\infty$ coefficients was credited to Målqvist and Peterseim in their celebrating work [40]. This construction, known as Localized Orthogonal Decomposition (LOD), utilizes quasi-interpolation operators to decompose the solution into macroscopic and microscopic components [3, 41]. The combination of GMsFEMs and LOD led to the development of a CEM-GMsFEM by Chung, Efendiev, and Leung in [22], where "CEM" is the acronym for "Constraint Energy Minimizing". The novelty of CEM-GMsFEMs resides in replacing quasi-interpolation operators in LOD with element-wise eigenspace projections. Moreover, CEM-GMsFEMs introduce a relaxed version of the energy minimization problems to construct multiscale bases, which eliminates the necessity of solving saddle-point linear systems. Our intention here is not to present a comprehensive review of multiscale computational methods from the community, and hence, notable advancements such as heterogeneous multiscale methods [27, 1], generalized finite element methods [4, 5, 39], and variational multiscale methods [34, 35] are not covered.

This article serves as an application of the CEM-GMsFEM to sign-changing problems. While the T-coercivity theory guarantees the well-posedness of the model problem, the heterogeneity of the coefficient leads to a generally low regularity of the solution, resulting in suboptimal convergence rates when using standard finite element methods. The CEM-GMsFEM, being a multiscale computational method, is specifically designed to handle the low regularity of the solution, and the construction of multiscale bases in this paper is tailored to the sign-changing setting. For instance, the auxiliary space forms a core module in the original CEM-GMsFEM and is created by solving generalized eigenvalue problems, where the coefficient enters bilinear forms on both sides. However, this approach is not suitable for sign-changing problems, as

the eigenvalues can be negative, rendering the generalized eigenvalue problems ill-defined. To ensure the positivity of the eigenvalues and the approximation ability of the auxiliary space, we modify the generalized eigenvalue problems by replacing the coefficient with its absolute value counterpart. It is worth noting that in building multiscale bases, we adhere to the relaxed version of the energy minimization problems in this paper, which offers implementation advantages. We conduct numerical experiments to validate the performance of the proposed method, emphasizing that coarse meshes need not align with the sign-changing interfaces. Moreover, we demonstrate that contrast robustness, which is an important feature of the original CEM-GMsFEM, is inherited in the proposed method. Under several assumptions, we prove the existence of multiscale bases in the T-coercivity framework, the exponential decay property for the oversampling layers, and the inf-sup stability of the online space. Moreover, we can provide an apriori estimate for the proposed method, indicating that errors can be bounded in terms of coarse mesh sizes and the number of oversampling layers, while being independent of the regularity of the solution.

Currently, to the best of our knowledge, efforts to apply multiscale computational methods to sign-changing problems are scarce. An exception can be found in the work by Chaumont-Frelet and Verfürth in [19], where they employed the LOD framework. Note that contrast ratios play a significant role in the T-coercivity theory for assessing the well-posedness of the model problem. In comparison, the proposed method presented in this paper is more general, as it can handle a wider range of coefficient profiles, including those with high contrast ratios.

This paper is organized as follows. In Section 2, we introduce the model problem and present the T-coercivity theory. The construction of multiscale bases in the proposed method is detailed in Section 3. To validate the performance of the proposed method, Section 4 presents numerical experiments conducted on four different models. All theoretical analysis for the proposed method is gathered in Section 5. Finally, in Section 6, we conclude the paper.

# 2 Preliminaries

For simplicity, we consider a 2D Lipschitz domain denoted as $\Omega$, which can be divided into two non-overlapping subdomains $\Omega^+$ and $\Omega^-$ with $\Gamma$ as the interfaces. The extension of the following discussion to 3D is straightforward. Let $\sigma$ belong to $L^\infty(\Omega)$ such that the essential infimum of $\sigma$ over $\Omega^+$ is greater than zero ($\operatorname{ess\,inf}_{x\in\Omega^+}\sigma(x)>0$), and the essential supremum of $\sigma$ over $\Omega^-$ is less than zero ($\operatorname{ess\,sup}_{x\in\Omega^-}\sigma(x)<0$). In particular, $\sigma$ is discontinuous across the interface seperating $\Omega^+$ and $\Omega^-$. We further introduce notations: $\sigma^+_{\max}=\operatorname{ess\,sup}_{x\in\Omega^+}\sigma(x)$, $\sigma^+_{\min}=\operatorname{ess\,inf}_{x\in\Omega^+}\sigma(x)$, $\sigma^-_{\max}=\operatorname{ess\,sup}_{x\in\Omega^-}|\sigma(x)|$ and $\sigma^-_{\min}=\operatorname{ess\,inf}_{x\in\Omega^-}|\sigma(x)|$. Hence, we require that $\sigma^+_{\max}\geq\sigma^+_{\min}>0$ and $\sigma^-_{\max}\geq\sigma^-_{\min}>0$. We redefine $V$ as the conventional Hilbert space $H^1_0(\Omega)$, and consider a bilinear form $a(\cdot,\cdot)$ on $V\times V$ given by:

$$a(v,w)=\int_\Omega \sigma\nabla v\cdot\nabla w\,\mathrm{d}x,\ \forall(v,w)\in V\times V.$$

Then for an function $f$ belonging to $L^2(\Omega)$, the following variational form defines the model problem:

$$\text{find } u\in V \text{ s.t. } \forall v\in V,\ a(u,v)=\int_\Omega fv\,\mathrm{d}x. \tag{1}$$

Certainly, (1) corresponds to a PDE of $u$, i.e., $-\operatorname{div}\sigma\nabla u=f$ with a boundary condition $u=0$ on $\partial\Omega$. It is also convenient to introduce a notation for a norm as $\|v\|_{\tilde{a},\omega}:=(\int_\omega|\sigma|\,|\nabla v|^2\,\mathrm{d}x)^{1/2}$, where $\omega$ is a subdomain of $\Omega$. Moreover, we abbreviate $\|\cdot\|_{\tilde{a},\Omega}$ as $\|\cdot\|_{\tilde{a}}$, which is exactly the *energy norm* on $V$.

However, the well-posedness of (1) is generally questionable due to the loss of uniform strict positivity of $\sigma$. T-coercivity is based on a fact: if there exists a *bijective* map $\mathcal{T}\colon V\to V$ and a positive constant $\alpha$ such that for all $v\in V$, $|a(v,\mathcal{T}v)|\geq\alpha\|v\|^2_{\tilde{a}}$, then the Banach–Nečas–Babuška theory confirms the existence and uniqueness of the solution. The novelty of T-coercivity lies in providing a systematic construction of such $\mathcal{T}$ using a "flip" operator, which is essentially dependent on the geometric information of the interfaces between $\Omega^+$ and $\Omega^-$. Let $V^\pm\subset H^1(\Omega^\pm)$ be the space obtained by restricting functions in $V$ in $\Omega^\pm$ and $\Gamma$

be the interface. Assuming the existence of a bounded map $\mathcal{R}\colon V^+ \to V^-$ with $\mathcal{R}v|_\Gamma = v|_\Gamma$ in the sense of Sobolev traces, the construction of $\mathcal{T}$ is given by

$$\mathcal{T}v = \begin{cases} v_1, & \text{in } \Omega^+, \\ -v_2 + 2\mathcal{R}v_1, & \text{in } \Omega^-, \end{cases} \tag{2}$$

where

$$v = \begin{cases} v_1, & \text{in } \Omega^+, \\ v_2, & \text{in } \Omega^-. \end{cases} \tag{3}$$

Certainly, the definition of $\mathcal{T}$ in (2) is bijective and bounded. Furthermore, we can show that

$$\begin{aligned} a(v, \mathcal{T}v) &= \int_{\Omega^+} |\sigma| \nabla v_1 \cdot \nabla v_1 \,\mathrm{d}x - \int_{\Omega^-} |\sigma| \nabla v_2 \cdot \nabla(-v_2 + 2\mathcal{R}v_1) \,\mathrm{d}x \\ &\underset{(\forall \eta > 0)}{\geq} \int_{\Omega^+} |\sigma| \left|\nabla v_1\right|^2 \mathrm{d}x - \frac{1}{\eta} \int_{\Omega^-} |\sigma| \left|\nabla \mathcal{R}v_1\right|^2 \mathrm{d}x + (1-\eta) \int_{\Omega^-} |\sigma| \left|\nabla v_2\right|^2 \mathrm{d}x \quad \text{(by Young's inequality)} \\ &\geq \int_{\Omega^+} |\sigma| \left|\nabla v_1\right|^2 \mathrm{d}x - \frac{\sigma^-_{\max} \|\mathcal{R}\|_1^2}{\eta} \int_{\Omega^+} \left|\nabla v_1\right|^2 \mathrm{d}x + (1-\eta) \int_{\Omega^-} |\sigma| \left|\nabla v_2\right|^2 \mathrm{d}x \\ &\geq \left(1 - \frac{\sigma^-_{\max} \|\mathcal{R}\|_1^2}{\sigma^+_{\min} \eta}\right) \int_{\Omega^+} |\sigma| \left|\nabla v_1\right|^2 \mathrm{d}x + (1-\eta) \int_{\Omega^-} |\sigma| \left|\nabla v_2\right|^2 \mathrm{d}x \\ &\geq \left(1 - \sqrt{\frac{\sigma^-_{\max}}{\sigma^+_{\min}}} \|\mathcal{R}\|_1\right) \|v\|_a^2, \quad \text{(by choosing } \eta = \sqrt{\sigma^-_{\max}/\sigma^+_{\min}} \|\mathcal{R}\|_1) \end{aligned} \tag{4}$$

where $\|\mathcal{R}\|_1$ should be understood as a positive constant such that for all $w \in V^+$,

$$\left(\int_{\Omega^-} \left|\nabla \mathcal{R}w\right|^2 \mathrm{d}x\right)^{1/2} \leq \|\mathcal{R}\|_1 \left(\int_{\Omega^-} \left|\nabla w\right|^2 \mathrm{d}x\right)^{1/2}. \tag{5}$$

We can observe that (1) will always be well-posed if the ratio $\sigma^+_{\min}/\sigma^-_{\max}$ is large enough. Note that the derivation in (4) involves flipping "positive" to "negative" as indicated by the definition of $\mathcal{T}\colon V^+ \to V^-$. It is possible to flip "negative" to "positive", and the well-posedness of (1) can also be achieved if the $\sigma^-_{\min}/\sigma^+_{\max}$ is large, as discussed in [10].

To gain insight into the operator $\mathcal{R}$, let us consider a square domain $\Omega = (0,1) \times (0,1)$ with $\Omega^+ = (0,1) \times (\gamma, 1)$ and $\Omega^- = (0,1) \times (0,\gamma)$, where $0 < \gamma < 1$. For a $w \in V^+$, we can define $\mathcal{R}w$ as follows:

$$(\mathcal{R}w)(x_1, x_2) = w(x_1, 1 - \tfrac{1-\gamma}{\gamma} x_2), \ \forall (x_1, x_2) \in \Omega^-.$$

It is also straightforward to calculate that $\|\mathcal{R}\|_1 = \max\{\sqrt{(1-\gamma)/\gamma}, 1\}$. Note that the construction of $\mathcal{R}$ is not unique. As an alternative approach, we can introduce a smooth cut-off function $\xi$ satisfying $\xi = 1$ on $\Gamma$ and $\operatorname{supp} \xi w \subset (0,1) \times (\gamma, a)$ with $\gamma < a < 1$. Then, we can utilize the same technique to design $\mathcal{R}'$ such that $\operatorname{supp} \mathcal{R}'(\xi w) \subset (0,1) \times (b, \gamma)$ with $0 \leq b < \gamma$. Naturally, as an operator, $\mathcal{R}'\xi\colon V^+ \to V^-$ is also valid.

# 3 Methods

The proposed method relies on a pair of nested meshes $\mathcal{K}_h$ and $\mathcal{K}_H$. The fine mesh $\mathcal{K}_h$ should be capable of resolving the heterogeneity of $\sigma$. Meanwhile, the construction of multiscale bases, known as the offline phase, is performed on $\mathcal{K}_h$. Solving the final linear system is commonly referred to as the online phase, whose computational complexity is associated with the coarse mesh $\mathcal{K}_H$. We reserve the index $i$ for enumerating

all coarse elements on $\mathcal{K}_H$, where $1 \leq i \leq N_{\text{elem}}$. The oversampling region $K_i^m$ of a coarse element $K_i$ holds a vital significance in describing the method. Mathematically, $K_i^m$ is defined by a recursive relation:

$$\forall m \geq 1, \quad K_i^m := \text{int}\Big(\cup\{\overline{K} \mid K \in \mathcal{K}_H \text{ with } \overline{K} \cap \overline{K_i^{m-1}} \neq \varnothing\}\Big)$$

and $K_i^0 := K_i$. Figure 1 serves as an illustration of the nested meshes, and note that two oversampling regions $K_{i'}^2$ and $K_{i''}^2$ are colored in gray.

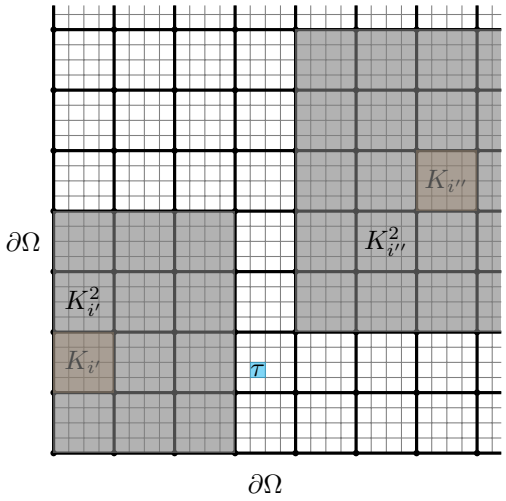


Figure 1: Illustration of the nested meshes $\mathcal{K}_h$ and $\mathcal{K}_H$. A fine element $\tau$, two coarse elements $K_{i'}$ and $K_{i''}$, accompanied by their corresponding oversampling regions $K_{i'}^2$ and $K_{i''}^2$, are colored differently.

In the original CEM-GMsFEM, a generalized eigenvalue problem,

$$\text{find } \lambda \in \mathbb{R} \text{ and } v \in H^1(K_i) \setminus \{0\} \text{ s.t. } \int_{K_i} \sigma \nabla v \cdot \nabla w \, \mathrm{d}x = \lambda \int_{K_i} \mu_{\text{msh}} \, \text{diam}(K_i)^{-2} \sigma v w \, \mathrm{d}x, \tag{6}$$

is solved on each coarse element $K_i$, where $\mu_{\text{msh}}$ is a generic positive constant that depends on the mesh quality. Then the local auxiliary space $V_i^{\text{aux}}$ is formed by collecting leading eigenvectors. However, recalling that $\sigma$ is not uniformly positive, the left-hand bilinear form in (6) is not positive semidefinite, and similarly, the right-hand bilinear form in (6) is not positive definite. Then, determining leading eigenvectors via (6) is problematic since the eigenvalues could be negative. To address this issue, we instead construct the following generalized eigenvalue problem on each $K_i$, which forms the first step of the proposed method:

**First step**

$$\begin{aligned} &\text{find } \lambda \in \mathbb{R} \text{ and } v \in H^1(K_i) \setminus \{0\} \text{ s.t. } \forall w \in H^1(K_i), \\ &\int_{K_i} |\sigma| \, \nabla v \cdot \nabla w \, \mathrm{d}x = \lambda \int_{K_i} \mu_{\text{msh}} \, \text{diam}(K_i)^{-2} \, |\sigma| \, v w \, \mathrm{d}x. \end{aligned} \tag{7}$$

One may argue that if $\sigma$ only takes values in $\{-1, 1\}$, then (7) would yield trivial eigenspaces corresponding to the Laplace operator, thus failing to capture any heterogeneity information of $\sigma$. However, according to T-coercivity, in such cases, the well-posedness of the model problem is violated (ref. [10] section 6). This inherent issue suggests that the original model problem should require a more sophisticated investigation in this case, which falls outside the scope of the proposed method. Upon solving the $l^*$ leading eigenvectors, denoted as $\psi_{i,j}$ with $1 \leq j \leq l^*$, we can construct the local auxiliary space $V_i^{\text{aux}} \subset L^2(K_i)$ as $\text{span}\{\psi_{i,j}\}$

for $1 \leq i \leq N_{\text{elem}}$. The global auxiliary space $V^{\text{aux}} \subset L^2(\Omega)$ is defined as $V^{\text{aux}} = \oplus_{i=1}^{N_{\text{elem}}} \widetilde{V}_i^{\text{aux}}$, where $\widetilde{V}_i^{\text{aux}} \subset L^2(\Omega)$ is the space by performing zero-extension of functions in $V_i^{\text{aux}}$.

We introduce several notations for future reference. Let $\mu$ be a function in $L^\infty(\Omega)$ satisfying

$$\mu|_{K_i} = \mu_{\text{msh}} \operatorname{diam}(K_i)^{-2} \sigma|_{K_i}$$

for all $K_i \in \mathcal{K}_H$. Next, for a subdomain $\omega$ in $\Omega$, we define two bilinear forms:

$$a(v,w)_\omega \coloneqq \int_\omega \sigma \nabla v \cdot \nabla w \,\mathrm{d}x \text{ and } s(v,w)_\omega \coloneqq \int_\omega \mu v w \,\mathrm{d}x.$$

Similarly to the definition of $\|\cdot\|_{\tilde{a},\omega}$, we define $\|v\|_{\tilde{s},\omega}$ as $(\int_\omega |\mu|\,|v|^2 \,\mathrm{d}x)^{1/2}$. Again, we drop the subscript $\omega$ if $\omega = \Omega$. Additionally, we define the orthogonal projection operator $\mathcal{P}_H \colon L^2(\Omega) \to L^2(\Omega)$ under the norm $\|\cdot\|\tilde{s}$, with $V^{\text{aux}} = \operatorname{im} \mathcal{P}_H$. We also use the shorthand notation $V_i^m$ as $H_0^1(K_i^m)$, and we always implicitly identify $V_i^m$ as a subspace of $V$.

Functions in $V^{\text{aux}}$ may not be continuous in $\Omega$, and thus $V^{\text{aux}}$ cannot be used as a conforming finite element space. In the second step, we will construct a multiscale basis $\phi_{i,j}$ in $V_i^m$, corresponding to $\psi_{i,j}$ in $V_i^{\text{aux}}$ that is obtained in the first step. This construction is based on the following variational problem, where the righthand bilinear form is defined on the original coarse element $K_i$:

**Second step**

$$\begin{aligned} &\text{find } \phi_{i,j} \in V_i^m \text{ s.t. } \forall w \in V_i^m, \\ &a(\phi_{i,j}, w)_{K_i^m} + s(\mathcal{P}_H \phi_{i,j}, \mathcal{P}_H w)_{K_i^m} = s(\psi_{i,j}, \mathcal{P}_H w)_{K_i}. \end{aligned} \tag{8}$$

In the original CEM-GMsFEM [22], the variational form in (8) is derived from a "relaxed" constrained energy minimization problem:

$$\phi_{i,j} = \operatorname{argmin}\big\{ a(w,w)_{K_i^m} \mid \ w \in V_i^m,\ \mathcal{P}_H w = \psi_{i,j} \big\}.$$

We note that, in the present situation, the bilinear forms $a(\cdot,\cdot)_{K_i^m}$ and $s(\mathcal{P}_H \cdot, \mathcal{P}_H \cdot)_{K_i^m}$ may lack coercivity, leading to an ill-defined minimization. The multiscale space is formed as

$$V_H^m = \operatorname{span}\{\phi_{i,j} \mid 1 \leq i \leq N_{\text{elem}},\ 0 \leq j < l^*\},$$

and the solution of the model problem is approximated by solving the following variational problem:

$$\text{find } u_H \in V_H^m \text{ s.t. } \forall w_H \in V_H^m,\ a(u_H, w_H) = \int_\Omega f w_H \,\mathrm{d}x. \tag{9}$$

Note that (7) and (8) are practically solved on the fine mesh $\mathcal{K}_h$.

The success of the proposed method hinges on that the multiscale basis decays rapidly w.r.t. $m$, the number of oversampling layers. To demonstrate this, we consider a square domain with a $10 \times 10$ periodic structure, as illustrated in the subplot **(a)** of Fig. 2. Each periodic cell contains a square inclusion that is centered with a negative coefficient imposed, resulting in the union of all inclusions forming the subdomain $\Omega^-$. The length ratio of the inclusion to the periodic cell is set to $1/2$. The material properties are set as $\sigma = 1.0$ in $\Omega^+$ and $\sigma = -0.1$ in $\Omega^-$ such that T-coercivity is satisfied. The coarse mesh $\mathcal{K}_H$ aligns with the periodic structure. We select a coarse element marked with red borders in the subplot **(a)** of Fig. 2 and plot the first three eigenfunctions $\psi_{i,1}$, $\psi_{i,2}$ and $\psi_{i,3}$ calculated via (7) in the subplots **(b)** to **(d)**. It is worth noting that the first eigenvalue is always 0, and the corresponding eigenfunction is always a constant function, as easily derived in (7) and also validated in the subplot **(b)**. The decay of the multiscale basis $\phi_{i,1}$, $\phi_{i,2}$ and $\phi_{i,3}$ are solved by (8) with different oversampling layers is demonstrated in Fig. 3, while $a(\cdot,\cdot)_{K_i^m}$ and $s(\mathcal{P}_H \cdot, \mathcal{P}_H \cdot)_{K_i^m}$ are not coercive. In this figure, the first, second, and third rows correspond to the results of the first, second, and third eigenfunctions, respectively, while the first, second, and third columns

display the plots of the multiscale basis with $m = 1$, 2, and 3, respectively. The position of the selected coarse element determines the maximum value of $m$, which in this case is 8. Consequently, we calculate the relative differences in both the energy and $L^2$ norm of the multiscale bases between $m = 8$ and $m = 1, \ldots, 7$, and present the results in the fourth column of Fig. 3. From the plots of multiscale bases, we observe that multiscale bases vanish away from the selected coarse element. Hence, one may expect to accurately compute multiscale bases with a small number of oversampling layers. Moreover, the relative differences in the logarithmic scale are almost linear w.r.t. $m$, which suggests the exponential decay may still hold in sign-changing problems.

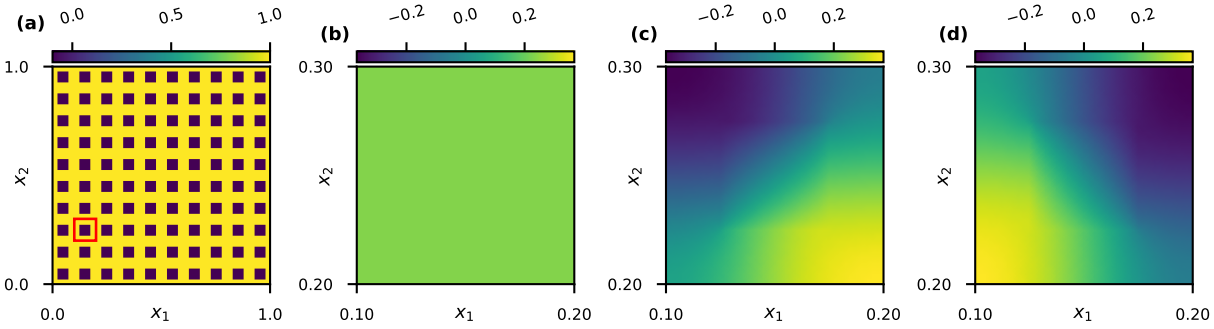


Figure 2: **(a)** The coefficient profile and the marked coarse element. **(b)**–**(d)** The plot of the first/second/third eigenfunction corresponding to the marked coarse element.

# 4 Numerical experiments

We conduct numerical experiments on a square domain $\Omega = (0, 1) \times (0, 1)$. The fine mesh $\mathcal{K}_h$ is generated by dividing $\Omega$ into $400 \times 400$ squares. Consequently, the coefficient profile $\sigma$ is represented by a $400 \times 400$ matrix/image, with each element corresponding to a constant value on a fine element. To investigate the convergence behavior of the proposed method with different coarse mesh sizes $H$, we consider four different coarse meshes $\mathcal{K}_H$: $10 \times 10$, $20 \times 20$, $40 \times 40$, and $80 \times 80$. The reference solution, auxiliary spaces, and multiscale bases are all calculated on the fine mesh $\mathcal{K}_h$ using the $Q_1$ finite element method. We evaluate the convergence of the proposed method using two relative error indices: the relative energy error and the relative $L^2$ error which are defined as follows:

$$\frac{\|e_h\|_{\tilde{a}}}{\|u_h\|_{\tilde{a}}} \text{ and } \frac{\|e_h\|_{L^2(\Omega)}}{\|u_h\|_{L^2(\Omega)}},$$

where $u_h$ is the reference solution calculated by the $Q_1$ FEM on $\mathcal{K}_h$ or the nodal interpolation of the exact solution (if available), and $e_h$ is the error between the reference solution and the numerical solution. For simplicity, we take $\mu$ as

$$\mu|_{K_i} = \mu_{\text{msh}} \operatorname{diam}(K_i)^{-2} \sigma|_{K_i} = 24 H^{-2} \sigma|_{K_i}$$

for all numerical experiments, as suggested in [50]. In the following discussions, we mark statements of direct observations from numerical experiments with a circled number, e.g., ①. We implement the method using the Python libraries NumPy and SciPy[1], and all the codes are hosted on Github[2].

[1] Instead of using the default sparse linear system solver in SciPy, we opted to utilize the `pardiso` solver to enhance efficiency.
[2] `https://github.com/Laphet/sign-changing`

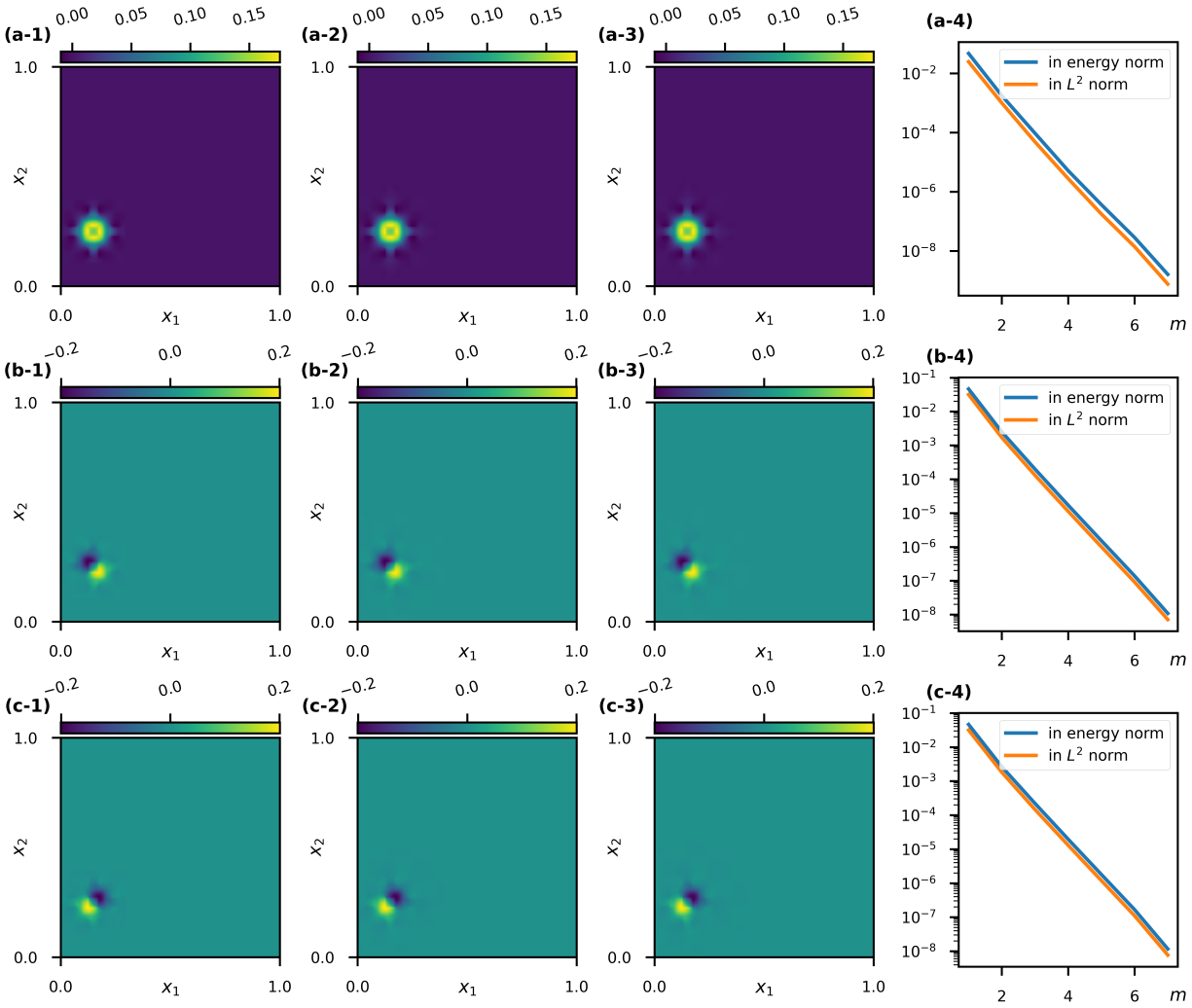


Figure 3: The subplots are marked as **(x-y)**, where **x** can take **a**, **b**, or **c**, corresponding to the results for the first, second, or third eigenfunction, respectively. If **y** is **1**, **2**, or **3**, the subplot displays the multiscale basis with $m$ oversampling layers, $m$ equal to **y**. Alternatively, if **y** is **4**, the subplot shows the relative differences (y-axis) in the energy and $L^2$ norm of the multiscale bases between $m = 8$ and $m = 1, \ldots, 7$ (x-axis).

## 4.1 Flat interface model

We first consider a flat interface model described in the ending part of Section 2, i.e., we define $\Omega^+ = (0,1) \times (\gamma, 1)$ and $\Omega^- = (0,1) \times (0, \gamma)$ with $0 < \gamma < 1$. In $\Omega^+$, we assign a fixed value of $\sigma_*^+$ to $\sigma^+$, while in $\Omega^-$, we take $-\sigma_*^-$ for $\sigma^-$, where $\sigma_*^+$ and $\sigma_*^-$ are both positive. We can devise an exact solution $u$ as

$$u(x_1, x_2) = \begin{cases} -\sigma_*^- x_1(x_1 - 1)x_2(x_2 - 1)(x_2 - \gamma), & \text{in } \Omega^+, \\ \sigma_*^+ x_1(x_1 - 1)x_2(x_2 - 1)(x_2 - \gamma), & \text{in } \Omega^-, \end{cases}$$

which corresponds to a smooth source term $f$ given by

$$f(x_1, x_2) = \sigma_*^- \sigma_*^+ \Big( 2x_2(x_2 - 1)(x_2 - \gamma) + x_1(x_1 - 1)(6x_2 - 2(\gamma + 1)) \Big).$$

According to the T-coercivity theory [20], the problem is well-posed if

$$\frac{\sigma_*^-}{\sigma_*^+} \notin \left[\frac{\gamma}{1-\gamma}, 1\right],$$

provided that $\gamma \leq 1/2$.

**Case I** We set $\gamma = 0.5$, we can check that now the interface $\Gamma = (0,1)\times\{\gamma\}$ is *fully resolved by every coarse mesh*. Therefore, we can expect that applying the $Q_1$ FEM on $\mathcal{K}_H$ directly can yield satisfactory accuracy. we conducted two groups of experiments with $(\sigma_*^+, \sigma_*^-) = (1.01, 1)$ and $(\sigma_*^+, \sigma_*^-) = (1, 1.01)$, both satisfying the T-coercivity condition. For the setting of the proposed multiscale method, we fix $l^* = 3$, indicating that we calculate the first three eigenfunctions in (7) and construct three multiscale bases for each coarse element, while we vary the oversampling layers $m$ from 1 to 4. We refer to Fig. 4 for the numerical results. ① We can observe from subplots **(a)** to **(d)** that the convergence of the $Q_1$ FEM manifests a linear pattern w.r.t. $H$ in the logarithmic scale, consistent with the theoretical expectation. ② We can also see that the number of oversampling layers $m$ has a significant impact on the accuracy of the proposed method. ③ However, for the same $m$, the error decaying w.r.t. $H$ does not always hold, as depicted in subplots **(b)** and **(d)**. ④ Although, for $m = 4$, the proposed method exhibits higher accuracy than the $Q_1$ FEM, the computational cost is significantly higher due to the sophisticated process of constructing multiscale bases. Therefore, the proposed method seems more suitable for scenarios involving intricate coefficient profiles. Interestingly, we notice that subplots **(a)** and **(c)**, as well as **(b)** and **(d)**, are almost identical, implying that the contrast ratio $\sigma_*^-/\sigma_*^+$ crossing the critical value 1 does not generate a significant influence on numerical methods.

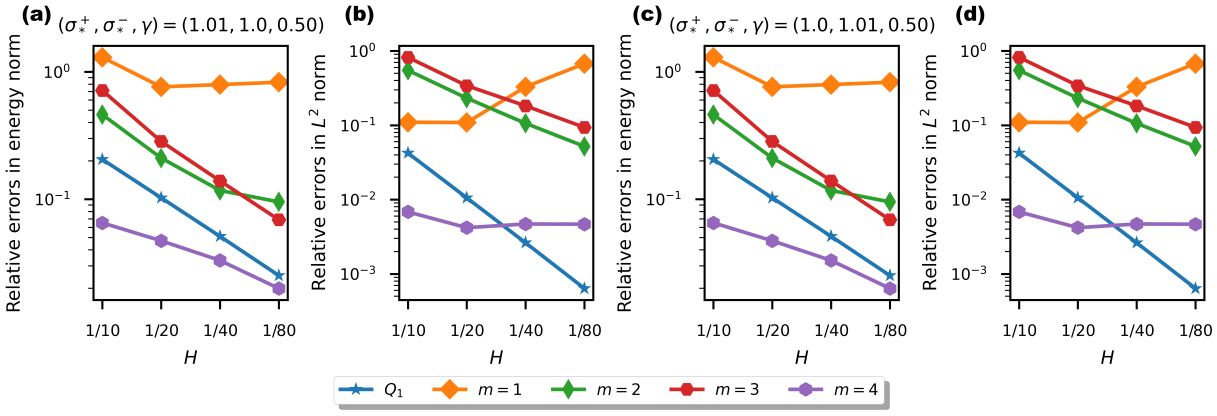


Figure 4: Numerical results for the flat interface model with $\gamma = 1/2$, where the relative errors of the proposed method with different numbers of oversampling layers $m$ and the $Q_1$ FEM are calculated w.r.t. the coarse mesh size $H$. Subplots **(a)** and **(b)** correspond to $(\sigma_*^+, \sigma_*^-) = (1.01, 1)$, which the relative errors are measured in the energy and $L^2$ norm, respectively. Similarly, subplots **(c)** and **(d)** correspond to the setting $(\sigma_*^+, \sigma_*^-) = (1, 1.01)$, following the same manner.

**Case II** Next, we consider the case that $\gamma = 0.49$, resulting in *none of the coarse meshes being capable of resolving the interface*. The results by setting $(\sigma_*^+, \sigma_*^-) = (1, 1.01)$ are reported in Fig. 5, where subplots **(a)** and **(b)** correspond to the relative errors, measured in the energy and $L^2$ norm respectively. Subplots **(c)** and **(d)** display the actual differences between the reference solution and numerical solutions obtained by the $Q_1$ FEM with $H = 1/80$ and the proposed method with $(H, m) = (1/80, 3)$, respectively. ① We observe that a slight change in $\gamma$ from 0.5 to 0.49 disrupts the convergence of the $Q_1$ FEM. ② In contrast, the proposed method can achieve satisfactory accuracy at a level of 1% for $m = 3$ and of 0.1% for $m = 4$ in the energy norm. ③ Again, for a fixed $m$, the errors from the proposed method do not always decay w.r.t.

$H$. As shown in subplots **(c)** and **(d)**, both methods exhibit a concentration of errors near the interface, while the proposed method outperforms the $Q_1$ FEM by an order of magnitude of two in terms of pointwise errors.

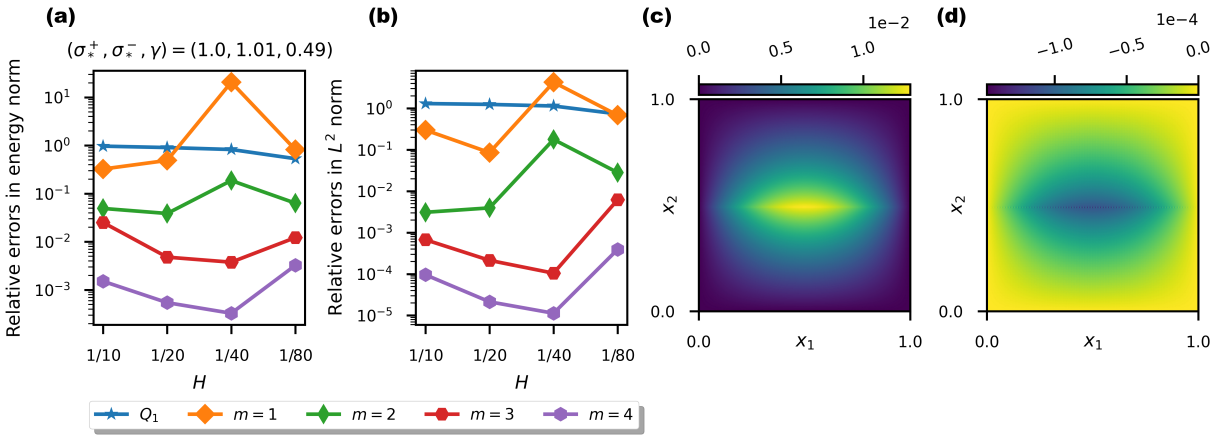


Figure 5: Numerical results for the flat interface model with $(\sigma_*^+, \sigma_*^-, \gamma) = (1, 1.01, 0.49)$. Subplots **(a)** and **(b)** show the relative errors of the proposed method with different numbers of oversampling layers $m$ and the $Q_1$ FEM w.r.t. the coarse mesh size $H$, but measured in different norms. Subplots **(c)** and **(d)** display the actual pointwise differences between the reference solution and numerical solutions obtained by the $Q_1$ FEM with $H = 1/80$ and the proposed method with $(H, m) = (1/80, 3)$.

## 4.2 Periodic square inclusion model

In this subsection, we revisit the periodic square inclusion model described in the ending part of Section 3. The coefficient profile $\sigma$ is determined as Fig. 2-**(a)**. Besides the $10 \times 10$ periodic configuration shown in Fig. 2-**(a)**, we also consider the $20 \times 20$ periodic configurations. The source term $f$ is constructed as the superposition of four 2D Gaussian functions centered at $(0.25, 0.25)$, $(0.75, 0.25)$, $(0.25, 0.75)$, and $(0.75, 0.75)$ with a variance of 0.01, as shown in Fig. 6-**(a)**. The reference solutions corresponding to the $10 \times 10$ and $20 \times 20$ periodic configurations are plotted in Fig. 6-**(b)** and **(c)**, respectively. We can observe that multiscale features emerge in the reference solutions, which may inspire future investigations into extending classical multiscale asymptotic analysis (ref. [7, 26]) to sign-changing problems. Notably, the homogenization theory for the model was recently completed by Bunoiu and Ramdani in [15, 14, 16, 17]. Meanwhile, numerical methods such as LOD [19], along with the proposed method, seek a low-dimensional representation of the solutions and can be regarded as general numerical homogenization techniques beyond the periodic setting.

We conduct numerical experiments using the proposed method on the $10 \times 10$ and $20 \times 20$ periodic configurations. The relative errors in the energy norm and the $L^2$ norm are tabulated in Tables 1 and 2, respectively. We again take $l^* = 3$ for all tests, while varying the number of oversampling layers $m$ from 1 to 4. ① From Tables 1 and 2, when $H = 1/10$, enlarging $m$ from 2 to 4 does not lead to a significant improvement in accuracy, and relative errors in the energy norm reach a saturation level of 1%. ② In comparison, for $H = 1/80$, the numerical solutions achieve a relative error of 0.1% in the energy norm for $m = 3$ and 4. We can conclude that the accuracy of the proposed method is mutually controlled by the number of oversampling layers $m$ and the coarse mesh size $H$, and this convergence pattern is consistent with the flat interface model. ③ Another noteworthy observation is that the scale of models, i.e., 1/10 and 1/20, does not affect the accuracy of the proposed method, as the relative errors in Tables 1 and 2 are almost at the same level with different $H$ and $m$. Hence, we can infer that the proposed method is robust to the scale of models and is not subject to the resonance error phenomenon encountered in the classic MsFEMs [32, 33, 29].

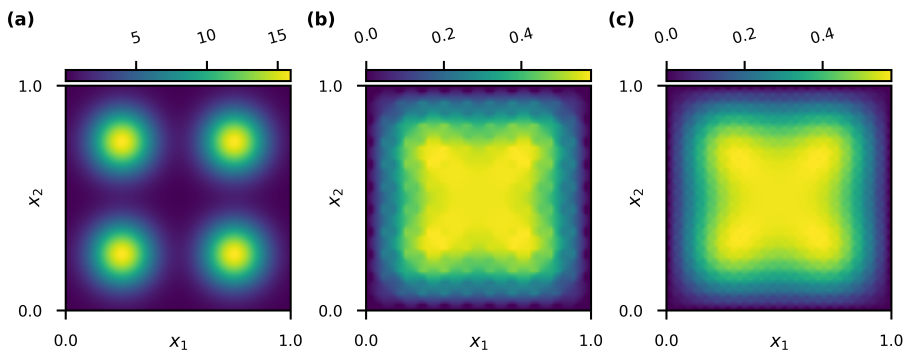


Figure 6: **(a)** The smooth source term $f$ that is constructed as the superposition of four 2D Gaussian functions centered at $(0.25, 0.25)$, $(0.75, 0.25)$, $(0.25, 0.75)$, and $(0.75, 0.75)$ with a variance of 0.01. **(b)-(c)** The reference solutions correspond to the $10 \times 10$ and $20 \times 20$ periodic configurations, respectively.

Table 1: For the $10 \times 10$ periodic square inclusion model, the relative errors in the energy norm (in the columns labeled with $\|\cdot\|_{\tilde{a}}$) and in the $L^2$ norm (in the columns labeled with $\|\cdot\|_{L^2(\Omega)}$).

| $H$ | $m=1$ | | $m=2$ | | $m=3$ | | $m=4$ | |
|---|---|---|---|---|---|---|---|---|
| | $\|\cdot\|_{\tilde{a}}$ | $\|\cdot\|_{L^2(\Omega)}$ | $\|\cdot\|_{\tilde{a}}$ | $\|\cdot\|_{L^2(\Omega)}$ | $\|\cdot\|_{\tilde{a}}$ | $\|\cdot\|_{L^2(\Omega)}$ | $\|\cdot\|_{\tilde{a}}$ | $\|\cdot\|_{L^2(\Omega)}$ |
| $\frac{1}{10}$ | 2.433e−1 | 8.931e−2 | 5.162e−2 | 5.741e−3 | 5.225e−2 | 5.785e−3 | 5.232e−2 | 5.786e−3 |
| $\frac{1}{20}$ | 3.785e−1 | 1.923e−1 | 4.960e−2 | 4.981e−3 | 5.583e−2 | 5.922e−3 | 5.662e−2 | 6.024e−3 |
| $\frac{1}{40}$ | 6.978e−1 | 5.833e−1 | 4.011e−2 | 2.292e−3 | 1.753e−3 | 3.293e−5 | 1.376e−4 | 3.075e−6 |
| $\frac{1}{80}$ | 8.832e−1 | 8.610e−1 | 8.895e−2 | 1.064e−2 | 3.855e−3 | 3.991e−5 | 1.941e−4 | 1.921e−6 |

Table 2: For the $20 \times 20$ periodic square inclusion model, the relative errors in the energy norm (in the columns labeled with $\|\cdot\|_{\tilde{a}}$) and in the $L^2$ norm (in the columns labeled with $\|\cdot\|_{L^2(\Omega)}$).

| $H$ | $m=1$ | | $m=2$ | | $m=3$ | | $m=4$ | |
|---|---|---|---|---|---|---|---|---|
| | $\|\cdot\|_{\tilde{a}}$ | $\|\cdot\|_{L^2(\Omega)}$ | $\|\cdot\|_{\tilde{a}}$ | $\|\cdot\|_{L^2(\Omega)}$ | $\|\cdot\|_{\tilde{a}}$ | $\|\cdot\|_{L^2(\Omega)}$ | $\|\cdot\|_{\tilde{a}}$ | $\|\cdot\|_{L^2(\Omega)}$ |
| $\frac{1}{10}$ | 5.571e−1 | 3.663e−1 | 6.379e−2 | 6.159e−3 | 2.653e−2 | 1.492e−3 | 2.637e−2 | 1.455e−3 |
| $\frac{1}{20}$ | 4.531e−1 | 2.884e−1 | 3.195e−2 | 1.498e−3 | 2.598e−2 | 1.441e−3 | 2.616e−2 | 1.442e−3 |
| $\frac{1}{40}$ | 6.306e−1 | 4.938e−1 | 4.544e−2 | 2.613e−3 | 2.589e−2 | 1.361e−3 | 2.782e−2 | 1.485e−3 |
| $\frac{1}{80}$ | 8.803e−1 | 8.499e−1 | 8.795e−2 | 1.057e−2 | 4.303e−3 | 4.984e−5 | 2.862e−4 | 3.115e−6 |

## 4.3 Periodic cross-shaped inclusion model

In this subsection, we apply the proposed method to a periodic cross-shaped inclusion model. Specifically, in each periodic cell, a cross-shaped inclusion is centered, imposing a negative coefficient, and the width of the cross arms is set to 1/5 of the periodic cell. We consider two periodic configurations, $10 \times 10$ and $20 \times 20$, as shown in Fig. 7-**(a)** and **(b)**, respectively. Note that cross-shaped inclusions are all connected, leading to long channels crossing the domain. Constructing a "flip" operator $\mathcal{R}$ in this model to validate the T-coercivity condition is nontrivial compared with the flat interface and the periodic square inclusion models. Using the theory developed in [9], one can prove a weak T-coercivity property holds as long as $\sigma^+_{\min}/\sigma^-_{\max}$ or $\sigma^-_{\min}/\sigma^+_{\max} > 3$. This model is designed to test the capability of the proposed method to handle long and

high-contrast channels, and we hence set $(\sigma_*^+, \sigma_*^-) = (1, 10^3)$, where $\sigma_*^+$ and $\sigma_*^-$ are defined as in the flat interface model. The source term $f$ is again from Fig. 6-**(a)**, and the reference solutions corresponding to the $10 \times 10$ and $20 \times 20$ periodic configurations are plotted in Fig. 7-**(c)** and **(d)**, respectively. From reference solution plots, we can observe that long and high-contrast channels create more pronounced oscillations compared to the square inclusion model (cf. Fig. 6-**(b)** and **(c)**), challenging numerical methods greatly.

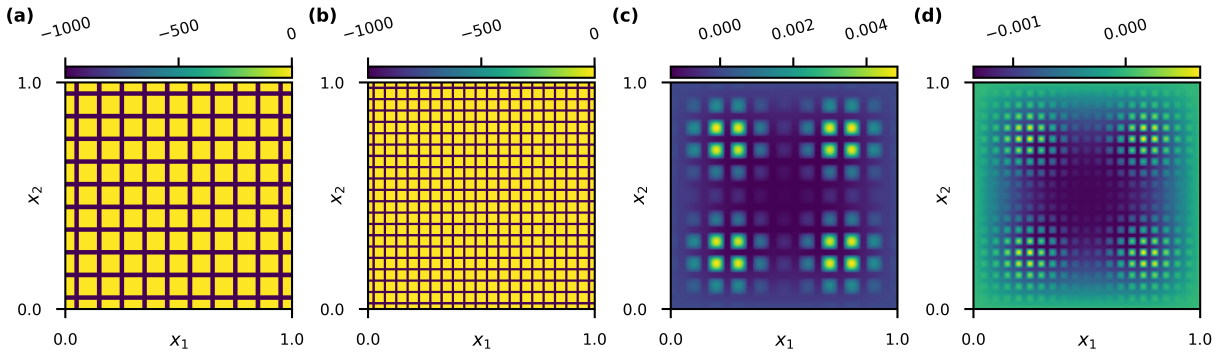


Figure 7: Subplots **(a)** and **(b)** display visualizations for the coefficient $\sigma$ used in the periodic cross-shaped inclusion model with $10 \times 10$ and $20 \times 20$ periodic configurations, respectively. Subplots **(c)** and **(d)** demonstrate the reference solutions for $10 \times 10$ and $20 \times 20$ periodic configurations, respectively.

The numerical results of the $Q_1$ FEM and the proposed method with $m \in \{1, 2, 3, 4\}$ are presented in Fig. 8. Subplots **(a)** and **(b)** in Fig. 8 share the same setting (corresponding to the $10 \times 10$ periodic configuration) but the relative errors are measured in the different norms, and similarly for subplots **(c)** and **(d)**. The $Q_1$ FEM fails to provide satisfactory accuracy. ① Even with a finer coarse mesh, such as $H = 1/80$, there is only a slight improvement, yet the relative errors in the energy norm remain close to 50%. The classical CEM-GMsFEM [22] is proven to be effective in handling long and high-contrast channels, and the proposed method inherits this advantage. ② By setting $m = 3$, the proposed method can achieve a relative error of 1% in the energy norm for both the $10 \times 10$ and $20 \times 20$ periodic configurations. ③ Typically, the relative errors in the $L^2$ norm are significantly smaller by an order of magnitude than those in the energy norm, and the proposed method can achieve a relative error of 0.1% in the $L^2$ norm for $m = 3$ and 4. ④ Furthermore, comparing subplots **(a)** and **(c)**, as well as **(b)** and **(d)**, reveals similar convergence behavior, indicating that the scale of the models does not affect the accuracy of the proposed method. Note that all coarse meshes employed cannot resolve the channels, highlighting the capability of the proposed method to handle complex coefficient profiles.

## 4.4 Random inclusion model

In this subsection, we consider a random inclusion model that is utilized in several multiscale methods as a showcase of the capability of handling nonperiodic coefficient profiles [22, 52, 50, 45]. The subdomains $\Omega^+$ and $\Omega^-$ are demonstrated in Fig. 9-**(a)**, and $\sigma$ is determined again by $(\sigma_*^+, \sigma_*^-)$. We consider two cases: $(\sigma_*^+, \sigma_*^-) = (1, 10^{-3})$ and $(\sigma_*^+, \sigma_*^-) = (1, 10^3)$. By setting the source term $f$ as depicted in Fig. 6-**(a)**, we plot the reference solutions for the two cases, as shown in Fig. 9-**(b)** and **(c)**. We can observe that void-type inclusions ($(\sigma_*^+, \sigma_*^-) = (1, 10^{-3})$) and rigid-type inclusions ($(\sigma_*^+, \sigma_*^-) = (1, 10^3)$) exhibit distinct characteristics in the reference solutions. A recent work [31] discussed the asymptotic behavior when the coefficient in inclusions tends to positive infinity, while the case of negative infinity has not been explored in the literature. Our aim in studying this model is to investigate the effect of choosing different eigenvectors $l^*$.

The approximation of eigenspaces relies on the rapid growth of the eigenvalue $\lambda$ in (7). Ideally for the Laplace operator, the eigenvalue distribution follows Weyl's law [49, 36]. However, as the contrast ratio

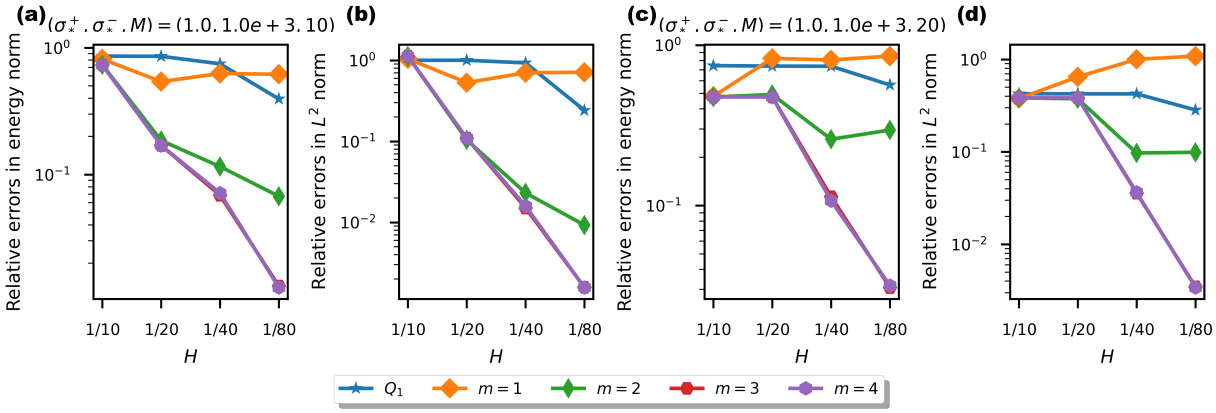


Figure 8: Numerical results for the periodic cross-shaped inclusion model with different periodic configurations. Subplots **(a)** and **(b)** show the relative errors of the proposed method for the $10 \times 10$ configuration with different numbers of oversampling layers $m$ and the $Q_1$ FEM w.r.t. the coarse mesh size $H$, but measured in different norms. Similarly, subplots **(c)** and **(d)** correspond to the $20 \times 20$ configuration, following the same manner.

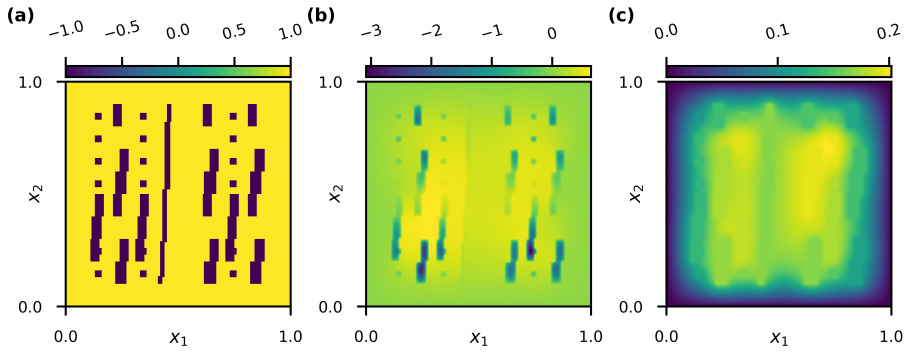


Figure 9: **(a)** The illustration of inclusions (dark regions) of the random inclusion model. **(b)** The plot of the reference solution by setting $(\sigma_*^+, \sigma_*^-) = (1, 10^{-3})$. **(c)** The plot of the reference solution by setting $(\sigma_*^+, \sigma_*^-) = (1, 10^3)$.

increases, the growth rate is expected to slow down, leading to a deterioration of the approximation quality. Surprisingly, this deterioration commonly appears in several leading eigenvalues in (7) where a weighted $L^2$ bilinear form is utilized. To investigate this phenomenon, we examine the first four eigenvalues for $(\sigma_*^+, \sigma_*^-) = (1, 10^{-3})$ and $(\sigma_*^+, \sigma_*^-) = (1, 10^3)$. The results are presented in Tables 3 and 4, where the minimal and maximal values of eigenvalues are calculated over all coarse elements. Since the first eigenvalue $\lambda_1$ is always 0, we omit the results for $\lambda_1$ in Tables 3 and 4. ① We can observe from Table 3 that for $\lambda_2$, there exist small eigenvalues that are on the order of $10^{-4}$. ② However, for $\lambda_3$ and $\lambda_4$, the minimum values are significantly larger, on the order of $10^{-1}$, compared to the minimum values of $\lambda_1$. ③ Interestingly, for $H = 1/80$, the maximum and minimum values for both cases are nearly the same. We conjecture that this is because the coarse meshes are fine enough, and the microstructures on each coarse element are simple. Note that the eigenvalues by setting $(\sigma_*^+, \sigma_*^-) = (1, 10^{-3})$ and $(\sigma_*^+, \sigma_*^-) = (10^3, 1)$ are strictly identical. If some coarse elements exhibit a "symmetrization" pattern between $\Omega^+$ and $\Omega^-$ and also contribute to the extreme values, we can expect to observe the aforementioned phenomenon.

We proceed by presenting the numerical errors of the proposed methods for the two settings in Fig. 10,

Table 3: For the random inclusion model with $(\sigma_*^+, \sigma_*^-) = (1, 10^{-3})$, the minimal and maximal values of the second/third/fourth eigenvalue over total coarse elements.

| $H$ | $\lambda_2$ | | $\lambda_3$ | | $\lambda_4$ | |
|---|---|---|---|---|---|---|
| | min | max | min | max | min | max |
| $\frac{1}{10}$ | 6.939e−4 | 4.114e−1 | 3.011e−1 | 4.114e−1 | 4.121e−1 | 8.229e−1 |
| $\frac{1}{20}$ | 7.806e−4 | 4.121e−1 | 3.066e−1 | 8.404e−1 | 4.129e−1 | 1.278 |
| $\frac{1}{40}$ | 1.008e−3 | 4.146e−1 | 2.617e−1 | 7.579e−1 | 5.091e−1 | 1.291 |
| $\frac{1}{80}$ | 1.051e−1 | 6.755e−1 | 3.255e−1 | 8.499e−1 | 5.228e−1 | 1.350 |

Table 4: For the random inclusion model with $(\sigma_*^+, \sigma_*^-) = (1, 10^{3})$, the minimal and maximal values of the second/third/fourth eigenvalue over total all coarse elements.

| $H$ | $\lambda_2$ | | $\lambda_3$ | | $\lambda_4$ | |
|---|---|---|---|---|---|---|
| | min | max | min | max | min | max |
| $\frac{1}{10}$ | 2.685e−1 | 4.335e−1 | 3.707e−1 | 1.141 | 6.694e−1 | 1.553 |
| $\frac{1}{20}$ | 6.856e−2 | 4.121e−1 | 3.254e−1 | 8.696e−1 | 5.733e−1 | 1.264 |
| $\frac{1}{40}$ | 1.365e−1 | 6.502e−1 | 3.292e−1 | 7.866e−1 | 5.763e−1 | 1.300 |
| $\frac{1}{80}$ | 1.051e−1 | 6.755e−1 | 3.255e−1 | 8.499e−1 | 5.228e−1 | 1.350 |

where we fix $m = 3$ and change $H$ and $l^*$. Subplots **(a)** and **(b)** in Fig. 10 correspond to $(\sigma_*^+, \sigma_*^-) = (1, 10^{-3})$, and subplots **(c)** and **(d)** correspond to $(\sigma_*^+, \sigma_*^-) = (1, 10^{3})$. ① Upon initial observation, we note a decay of numerical errors with increasing $l^*$ for most cases. ② However, this decay is less pronounced with $l^*$ from 3 to 4. Therefore, it is advisable to consider employing a larger number of eigenvectors as a stabilizing strategy rather than solely relying on it for accuracy improvement. ③ To gain further clarity, the results with $H = 1/10$ in Fig. 10-**(a)** and **(b)** do not show a decreasing pattern in errors with $l^*$. ④ However, we can observe a more substantial reduction in errors as $l^*$ increases for $H = 1/10$ in Fig. 10-**(c)** and **(d)**. This implies that the proposed method is more sensitive to the number of eigenvectors when the coarse mesh is not fine enough, while $l^* = 3$ is a recommended choice.

# 5 Analysis

In this section, we present a rigorous analysis of the proposed method. We begin by examining the global version of the method, where we set $K_i^m$ in (8) as $\Omega$. While the global version can yield optimal error estimates, it is impractical due to its high computational cost. Next, we delve into the local version of the method and replicate several estimates found in the original CEM-GMsFEM. Given the sign-changing setting and the T-coercivity framework, certain assumptions seem to be unavoidable due to technical difficulties, and we will explain the rationale behind these assumptions.

## 5.1 Global version

Let $\Upsilon$ be $\sigma_{\min}^+/\sigma_{\max}^-$. As shown in (4), the well-posedness of the model problem can be ensured if $\Upsilon$ is sufficiently large. By replacing $\sigma$ with $-\sigma$, the case that $\sigma_{\min}^-/\sigma_{\max}^+$ is sufficiently large can also be reduced to the case that we mainly focus on. Essentially, we require that the solution to the model problem be at least unique to discuss the convergence of the numerical method. Therefore, we introduce the following stronger assumption, although it is rarely touched in the subsequent analysis.

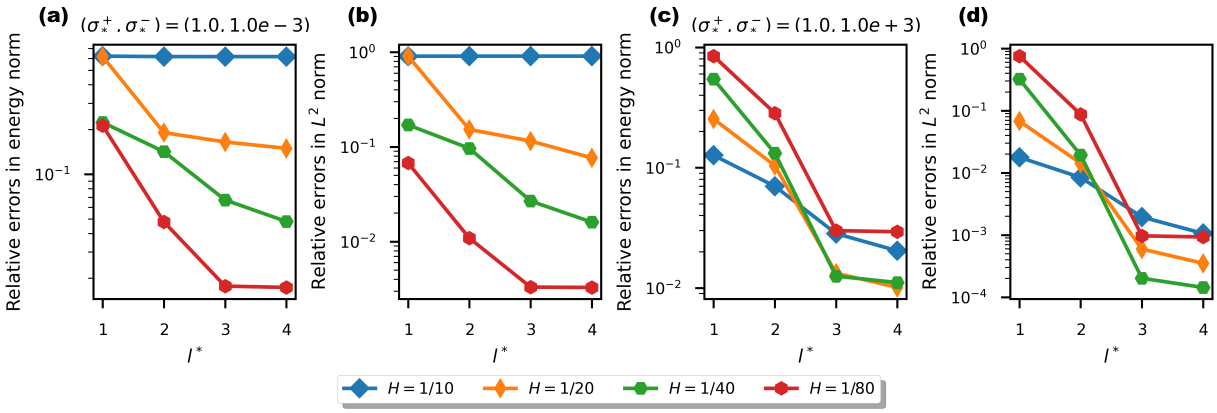


Figure 10: Numerical results for the random inclusion model. Subplots **(a)** and **(b)** show the relative errors of the proposed method for the setting $(\sigma_*^+, \sigma_*^-) = (1, 10^{-3})$ with different coarse mesh size $H$ w.r.t. the number of eigenvectors $l^*$, but measured in different norms. Similarly, subplots **(c)** and **(d)** correspond to $(\sigma_*^+, \sigma_*^-) = (1, 10^3)$, following the same manner.

**Assumption I.** The model problem (1) is well-posed in the sense that there exists a positive constant $C_{\mathrm{wp}}$, independent of $f$, such that

$$\|u\|_{\tilde{a}} \leq C_{\mathrm{wp}} \|f\|_{L^2(\Omega)}.$$

We define the global operator $\mathcal{G}_i^{\infty} \colon L^2(\Omega) \to V = H_0^1(\Omega)$ corresponding to the coarse element $K_i$ via the following variational problem:

$$\text{find } \mathcal{G}_i^{\infty}\psi \in V \text{ s.t. } \forall w \in V,\ a(\mathcal{G}_i^{\infty}\psi, w) + s(\mathcal{P}_H \mathcal{G}_i^{\infty}\psi, \mathcal{P}_H w) = s(\mathcal{P}_H\psi, \mathcal{P}_H w)_{K_i}. \tag{10}$$

We reiterate several important facts here: the bilinear form is $s(\cdot,\cdot)$ defined on $L^2(\Omega) \times L^2(\Omega)$ as

$$s(v,w) = \int_{\Omega} \mu v w \,\mathrm{d}x = \mu_{\mathrm{msh}} \sum_{i=1}^{N_{\mathrm{elem}}} \mathrm{diam}(K_i)^{-2} \int_{K_i} \sigma v w \,\mathrm{d}x,$$

which is not coercive due to the sign-changing property of $\mu$; the operator $\mathcal{P}_H \colon L^2(\Omega) \to L^2(\Omega)$ is an orthogonal projection with $\mathrm{im}\,\mathcal{P}_H = V^{\mathrm{aux}}$, regarding the weighted $L^2$ inner-product $\int_{\Omega} |\mu| \, v w \,\mathrm{d}x$ but not the bilinear form $s(\cdot,\cdot)$ or the standard $L^2$ inner-product. Taking a summation of $\mathcal{G}_i^{\infty}$ gives $\mathcal{G}^{\infty} := \sum_{i=1}^{N_{\mathrm{elem}}} \mathcal{G}_i^{\infty}$, which is interpreted as the global operator corresponding to the whole domain. Certainly, we are required to check the well-posedness of (10), which essentially involves proving the inf-sup stability of the bilinear form:

$$a(v,w) + s(\mathcal{P}_H v, \mathcal{P}_H w),\ \forall (v,w) \in V \times V.$$

Following the T-coercivity approach, we may examine the coercivity of

$$a(v, \mathcal{T}v) + s(\mathcal{P}_H v, \mathcal{P}_H \mathcal{T} v).$$

We encounter a challenge here as the properties of $\mathcal{P}_H\mathcal{T}$ are not clear, as $\mathcal{P}_H$ is associated with the coarse mesh $\mathcal{K}_H$, while $\mathcal{T}$ depends on the geometric information of $\Omega^{\pm}$ and could be rather complicated. To address this issue, we introduce the following assumption.

**Assumption II.** All coarse elements in $\mathcal{K}_H$ can be categorized into two groups $\mathcal{K}_H^+$ or $\mathcal{K}_H^-$, where

$$\mathcal{K}_H^{\pm} := \left\{ K \mid K \in \mathcal{K}_H \text{ and } K \subset \Omega^{\pm} \right\}.$$

This assumption is rather stringent, as it implies that the coarse mesh $\mathcal{K}_H$ is capable of resolving $\Omega^{\pm}$. On the other hand, such an assumption greatly facilitates the analysis once the orthogonal projection $\mathcal{T}_H$ comes into play. Another reason for this assumption is shown in proving Lemma 5.5. We emphasize that several numerical experiments in Section 4 indeed do not align with the assumption, yet the obtained results remain promising. To facilitate the analysis, we also require the following assumption. We mention that similar assumptions are also raised in constructing "flip" operators that satisfy weak T-coercivity (ref. [9] section 4).

**Assumption III.** There exists a positive constant $\|\mathcal{R}\|_0$, such that for any $v \in V(\Omega^+)$,

$$\|\mathcal{R}v\|_{0,\Omega^-} \leq \|\mathcal{R}\|_0 \|v\|_{0,\Omega^+},$$

where $\mathcal{R}$ is the operator in (2).

We introduce this assumption to establish an estimate for $s(\mathcal{P}_H v, \mathcal{P}_H \mathcal{T} v)$. This assumption is valid when the operator $\mathcal{R}$ acts as a change of variables. Denoting $\lambda_i^*$ as the $(l^*+1)$-th eigenvalue of the generalized eigenvalue problem in (7), the following lemma will be frequently utilized in the analysis, while its proof is simply a straightforward application of the properties of eigenspace expansions.

**Lemma 5.1.** *On each coarse element $K_i$ and for any $v \in H^1(K_i)$, the following estimates hold:*

$$\|v - \mathcal{P}_H v\|_{\tilde{s},K_i} \leq \frac{1}{\sqrt{\lambda_i^*}} \|v\|_{\tilde{a},K_i}, \tag{11}$$

$$\|v\|^2_{\tilde{s},K_i} \leq \|\mathcal{P}_H v\|^2_{\tilde{s},K_i} + \frac{1}{\lambda_i^*} \|v\|^2_{\tilde{a},K_i}. \tag{12}$$

We also introduce a notation that $\epsilon \coloneqq 1/(\max_i \lambda_i^*)$. The following estimates pave the way for proving the well-posedness of (10).

**Lemma 5.2.** *It holds that for any $v \in V$,*

$$a(v, \mathcal{T}v) \geq \left(1 - \|\mathcal{R}\|_1/\sqrt{\Upsilon}\right) \|v\|^2_{\tilde{a}}, \tag{13}$$

$$s(\mathcal{P}_H v, \mathcal{P}_H \mathcal{T} v) \geq \left(1 - \|\mathcal{R}\|_0/\sqrt{\Upsilon}\right) \|\mathcal{P}_H v\|^2_{\tilde{s}} - \epsilon \|\mathcal{R}\|_0/\sqrt{\Upsilon} \|v\|^2_{\tilde{a}}, \tag{14}$$

$$\|\mathcal{T}v\|_{\tilde{a}} \leq \max\left\{ \left(1 + 8\|\mathcal{R}\|_1^2/\Upsilon\right)^{1/2}, \sqrt{2} \right\} \|v\|_{\tilde{a}}, \tag{15}$$

$$\|\mathcal{T}v\|_{\tilde{s}} \leq \max\left\{ \left(1 + 8\|\mathcal{R}\|_0^2/\Upsilon\right)^{1/2}, \sqrt{2} \right\} \|v\|_{\tilde{s}}. \tag{16}$$

*Proof.* The proof of (13) has already been given in (4). We hence first prove (14). For any $v \in V$ with $v_1$ and $v_2$ defined as (3), we have

$$\begin{aligned} s(\mathcal{P}_H v, \mathcal{P}_H \mathcal{T} v) &= \int_{\Omega^+} |\mu| \left|\mathcal{P}_H v_1\right|^2 \mathrm{d}x + \int_{\Omega^-} |\mu| \left|\mathcal{P}_H v_2\right|^2 \mathrm{d}x - 2 \int_{\Omega^-} |\mu| \, \mathcal{P}_H v_2 \mathcal{P}_H \mathcal{R} v_1 \, \mathrm{d}x \\ &\underset{(\forall \eta > 0)}{\geq} \int_{\Omega^+} |\mu| \left|\mathcal{P}_H v_1\right|^2 \mathrm{d}x - \frac{1}{\eta} \int_{\Omega^-} |\mu| \left|\mathcal{P}_H \mathcal{R} v_1\right|^2 \mathrm{d}x + (1-\eta) \int_{\Omega^-} |\mu| \left|\mathcal{P}_H v_2\right|^2 \mathrm{d}x, \end{aligned}$$

where we implicitly utilize assumption II. We turn to the estimate of $\int_{\Omega^-} |\mu| \left|\mathcal{P}_H \mathcal{R} v_1\right|^2 \mathrm{d}x$ as follows:

$$\begin{aligned} \int_{\Omega^-} |\mu| \left|\mathcal{P}_H \mathcal{R} v_1\right|^2 \mathrm{d}x &\leq \int_{\Omega^-} |\mu| \left|\mathcal{R} v_1\right|^2 \mathrm{d}x \quad (\mathcal{P}_H \text{ is an orthogonal projection}) \\ &\leq \mu^-_{\max} \int_{\Omega^-} \left|\mathcal{R} v_1\right|^2 \mathrm{d}x \leq \mu^-_{\max} \|\mathcal{R}\|_0^2 \int_{\Omega^+} \left|v_1\right|^2 \mathrm{d}x \quad (\text{by assumption III}) \end{aligned}$$

$$
\begin{aligned}
&\leq \|\mathcal{R}\|_0^2 \frac{\mu_{\max}^-}{\mu_{\min}^+} \int_{\Omega^+} |\mu| \, |v_1|^2 \, \mathrm{d}x \\
&\leq \|\mathcal{R}\|_0^2 \frac{\sigma_{\max}^-}{\sigma_{\min}^+} \left( \int_{\Omega^+} |\mu| \, |\mathcal{P}_H v_1|^2 \, \mathrm{d}x + \epsilon \int_{\Omega^+} |\sigma| \, |\nabla v_1|^2 \, \mathrm{d}x \right). \quad \text{(by (12))}
\end{aligned}
$$

By choosing $\eta = \|\mathcal{R}\|_0 / \sqrt{\Upsilon}$, we derive that

$$
s(\mathcal{P}_H v, \mathcal{P}_H \mathcal{T} v) \geq (1 - \|\mathcal{R}\|_0 / \sqrt{\Upsilon}) \|\mathcal{P}_H v\|_{\tilde{s}}^2 - \epsilon \|\mathcal{R}\|_0 / \sqrt{\Upsilon} \|v\|_{\tilde{a}}^2,
$$

which finishes the proof of (14).

The proofs of (15) and (16) follow a similar procedure, and we only provide the proof for the former:

$$
\begin{aligned}
\|\mathcal{T} v\|_{\tilde{a}}^2 &= \int_{\Omega^+} |\sigma| \, |\nabla v_1|^2 \, \mathrm{d}x + \int_{\Omega^-} |\sigma| \, |-\nabla v_2 + 2 \nabla \mathcal{R} v_1|^2 \, \mathrm{d}x \\
&\leq \int_{\Omega^+} |\sigma| \, |\nabla v_1|^2 \, \mathrm{d}x + 2 \int_{\Omega^-} |\sigma| \, |\nabla v_2|^2 \, \mathrm{d}x + 8 \int_{\Omega^-} |\sigma| \, |\nabla \mathcal{R} v_1|^2 \, \mathrm{d}x \quad \text{(by the basic inequality)} \\
&\leq \int_{\Omega^+} |\sigma| \, |\nabla v_1|^2 \, \mathrm{d}x + 2 \int_{\Omega^-} |\sigma| \, |\nabla v_2|^2 \, \mathrm{d}x + 8 \sigma_{\max}^- \|\mathcal{R}\|_1^2 \int_{\Omega^+} |\nabla v_1|^2 \, \mathrm{d}x \\
&\leq \left( 1 + 8 \|\mathcal{R}\|_1^2 \frac{\sigma_{\max}^-}{\sigma_{\min}^+} \right) \int_{\Omega^+} |\sigma| \, |\nabla v_1|^2 + 2 \int_{\Omega^-} |\sigma| \, |\nabla v_2|^2 \, \mathrm{d}x \\
&\leq \max\left\{ 1 + 8 \|\mathcal{R}\|_1^2 / \Upsilon, 2 \right\} \|v\|_{\tilde{a}}^2.
\end{aligned}
$$

□

*Remark.* According to the proof, we can refine $\|v\|_{\tilde{a}}^2$ in (14) to $\|v\|_{\tilde{a},\Omega^+}^2$. However, we choose to retain the original form for the sake of simplicity.

Now the well-posedness of (10) can be established by combining (13) and (14).

**Proposition 5.3.** *There exist $\Upsilon'$ and $\epsilon'$ such that for any $\Upsilon \geq \Upsilon'$ and $\epsilon \leq \epsilon'$, the operator $\mathcal{G}_i^\infty$ in (10) is well-posed.*

Recalling that $\operatorname{im} \mathcal{P}_H = V^{\mathrm{aux}}$, The following lemma in some sense offers an interpolation operator that maps from $L^2(\Omega)$ to $V$, such that the projections onto $V^{\mathrm{aux}}$ by $\mathcal{P}_H$ are preserved.

**Lemma 5.4** (ref. [22])**.** *There exists a bounded map $\mathcal{Q}_H \colon L^2(\Omega) \to V$ and a positive constant $C_{\mathrm{inv}}$ such that for all $v \in L^2(\Omega)$, it holds that $\mathcal{P}_H \mathcal{Q}_H v = \mathcal{P}_H v$ and $\|\mathcal{Q}_H v\|_{\tilde{a}} \leq C_{\mathrm{inv}} \|\mathcal{P}_H v\|_{\tilde{s}}$. Moreover, for each coarse element $K_i$, $\mathcal{Q}_H v|_{K_i}$ depends only on the data of $v$ in $K_i$ and vanishes on $\partial K_i$.*

We introduce two function spaces as $W \coloneqq \ker \mathcal{P}_H \cap V$ and $V_H^\infty \coloneqq \operatorname{im} \mathcal{G}^\infty \subset V$. According to Lemma 5.1, it is clear that

$$
\|w\|_{\tilde{s}} \leq \sqrt{\epsilon} \|w\|_{\tilde{a}}, \quad \forall w \in W. \tag{17}
$$

The following lemma reveals a relationship of "orthogonality" between $W$ and $V_H^\infty$ concerning the bilinear form $a(\cdot, \cdot)$. However, we must be cautious in using the term "orthogonal" since $a(\cdot, \cdot)$ cannot define an inner product on $V$.

**Lemma 5.5.** *For any $v \in V_H^\infty$ and $w \in W$, it holds that $a(v, w) = 0$. If $w \in V$ and $a(v, w) = 0$ holds for any $v \in V_H^\infty$, then $w \in W$.*

*Proof.* The first argument that $a(v, w) = 0$ for any $v \in V_H^\infty$ and $w \in W$ is a direct result of (10). We then prove the second argument step by step. For simplicity, the notation $\sum_{i,j}$ is used to represent the summation over $i = 1, \ldots, N_{\mathrm{elem}}$ and $j = 1, \ldots, l^*$.

**Step1** We first state that the set $\{\mathcal{G}_i^\infty \psi_{i,j} \mid 1 \le i \le N_{\text{elem}}, 1 \le j \le l^*\}$ is linearly independent, where each $\psi_{i,j}$ is an eigenfunction by solving (7). Suppose there exists coefficients $\alpha_{i,j}$ such that $\sum_{i,j} \alpha_{i,j} \mathcal{G}_i^\infty \psi_{i,j} = 0$. Then for any $w \in V$, applying (10), we have

$$0 = a(\sum_{i,j} \alpha_{i,j} \mathcal{G}_i^\infty \psi_{i,j}, z) + s(\mathcal{P}_H \sum_{i,j} \alpha_{i,j} \mathcal{G}_i^\infty \psi_{i,j}, \mathcal{P}_H z) = s(\sum_{i,j} \alpha_{i,j} \psi_{i,j}, \mathcal{P}_H z), \quad \forall z \in V.$$

Recalling the property of $\mathcal{Q}_H$ from Lemma 5.4, we find $s(\sum_{i,j} \alpha_{i,j} \psi_{i,j}, z') = 0$ for any $z' \in V^{\text{aux}}$. By assumption II, on each $K_i$, we have $\mu = |\mu|$ or $-|\mu|$, which yields that

$$\int_{K_i} |\mu| \sum_j \alpha_{i,j} \psi_{i,j} z'' \, \mathrm{d}x = \int_{K_i} \mu \sum_j \alpha_{i,j} \psi_{i,j} z'' \, \mathrm{d}x = 0, \quad \forall z'' \in V_i^{\text{aux}}.$$

Therefore, we derive that $\alpha_{i,j} = 0$ for any $i$ and $j$, and thus prove the statement.

**Step2** We then state that $\{\psi_{i,j} - \mathcal{P}_H \mathcal{G}_i^\infty \psi_{i,j} \mid 1 \le i \le N_{\text{elem}}, 1 \le j \le l^*\}$ is a linearly independent set. Once again, from (10), if $\sum_{i,j} \alpha_{i,j} (\psi_{i,j} - \mathcal{P}_H \mathcal{G}_i^\infty \psi_{i,j}) = 0$, we have

$$a(\sum_{i,j} \alpha_{i,j} \mathcal{G}_i^\infty \psi_{i,j}, z) = s(\sum_{i,j} \alpha_{i,j} (\psi_{i,j} - \mathcal{P}_H \mathcal{G}_i^\infty \psi_{i,j}), \mathcal{P}_H z) = 0, \quad \forall z \in V.$$

According to assumption I, we conclude that $\sum_{i,j} \alpha_{i,j} \mathcal{G}_i^\infty \psi_{i,j} = 0$, which yields $\alpha_{i,j} = 0$ for any $i$ and $j$ from the previous step.

**Step3** Now we return to the original argument. If $\mathcal{P}_H w \neq 0$, we assert that there exists a set of $\{\alpha_{i,j}\}$ such that

$$s(\sum_{i,j} \alpha_{i,j} (\psi_{i,j} - \mathcal{P}_H \mathcal{G}_i^\infty \psi_{i,j}), \mathcal{P}_H w) = 1.$$

Otherwise, for all choices of $\{\alpha_{i,j}\}$, we must have $s(\sum_{i,j} \alpha_{i,j} (\psi_{i,j} - \mathcal{P}_H \mathcal{G}_i^\infty \psi_{i,j}), \mathcal{P}_H w) = 0$. Taking $\{\psi_{i,j}\}$ as bases for $V^{\text{aux}}$, we can check that the matrix representation of $s(\cdot, \cdot)$ on $V^{\text{aux}}$ is diagonal, which is also nonsingular thanks to assumption II. From the previous step, we have established that $\{\psi_{i,j} - \mathcal{P}_H \mathcal{G}_i^\infty \psi_{i,j}\}$ spans the finite-dimentional space $V^{\text{aux}}$. However, the relation $s(\sum_{i,j} \alpha_{i,j} (\psi_{i,j} - \mathcal{P}_H \mathcal{G}_i^\infty \psi_{i,j}), \mathcal{P}_H v) = 0$ for any $\{\alpha_{i,j}\}$ implies that $\mathcal{P}_H v = 0$, which is a contradiction. Therefore, We have

$$1 = a(\sum_{i,j} \alpha_{i,j} \mathcal{G}_i^\infty \psi_{i,j}, w) = s(\sum_{i,j} \alpha_{i,j} (\psi_{i,j} - \mathcal{P}_H \mathcal{G}_i^\infty \psi_{i,j}), \mathcal{P}_H w),$$

which contradicts to $a(v, w) = 0$ for any $v \in V_H^\infty$. □

The global solution is defined by solving the following variational form:

$$\text{find } u_H^\infty \in V_H^\infty, \ \text{ s.t. } \forall w_H \in V_H^\infty, \ a(u_H, w_H) = \int_\Omega f w_H \, \mathrm{d}x. \tag{18}$$

To guarantee the well-posedness of (18), we need to examine the inf-sup stability of $V_H^\infty$, i.e., proving the existence of a uniform lower bound of

$$\inf_{v_H \in V_H^\infty} \sup_{w_H \in V_H^\infty} \frac{a(v_H, w_H)}{\|v_H\|_{\tilde{a}} \|w_H\|_{\tilde{a}}}.$$

The Fortin trick [8] suggests that it suffices to check that

$$\inf_{v \in W} \sup_{w \in W} \frac{a(v, w)}{\|v\|_{\tilde{a}} \|w\|_{\tilde{a}}} \ge \Lambda_W$$

holds. The technique that we utilize here is introducing a modified $\mathcal{T}$ operator, denoted as $\mathcal{T}_H$, defined on $W$ as follows:

$$\mathcal{T}_H v = \mathcal{T} v - \mathcal{Q}_H \mathcal{P}_H \mathcal{T} v = \begin{cases} v_1 - \mathcal{Q}_H \mathcal{P}_H v_1, & \text{in } \Omega^+, \\ -v_2 + 2\mathcal{R} v_1 + \mathcal{Q}_H \mathcal{P}_H v_2 - 2\mathcal{Q}_H \mathcal{P}_H \mathcal{R} v_1, & \text{in } \Omega^-. \end{cases}$$
$$= \begin{cases} v_1, & \text{in } \Omega^+, \\ -v_2 + 2\mathcal{R} v_1 - 2\mathcal{Q}_H \mathcal{P}_H \mathcal{R} v_1, & \text{in } \Omega^-. \end{cases}$$

Here, the terms $\mathcal{Q}_H \mathcal{P}_H v_1$ and $\mathcal{Q}_H \mathcal{P}_H v_2$ vanish because $\mathcal{P}_H v = 0$ on each coarse element, and meanwhile $\mathcal{Q}_H \mathcal{P}_H v|_{K_i}$ is only dependent on $\mathcal{P}_H v|_{K_i}$ (see Lemma 5.4). Recalling the boundness of $\mathcal{Q}_H$ from Lemma 5.4, we can see that $\mathcal{T}_H$ is a bounded operator that maps from $W$ to $W$. Then, we reduce the proof of the inf-sup stability to establish the coercivity of $a(v, \mathcal{T}_H v)$ for all $v \in W$. The details of this proof are essentially a replication of (4). Taking $\mathcal{R}'_H \coloneqq \mathcal{R} - \mathcal{Q}_H \mathcal{P}_H \mathcal{R}$, we can obtain

$$a(v, \mathcal{T}_H v) \geq \left(1 - \|\mathcal{R}'_H\|_1 / \sqrt{\Upsilon}\right) \|v\|_{\tilde{a}}^2,$$

where $\|\mathcal{R}'_H\|_1$ is defined similarly as $\|\mathcal{R}\|_1$ in (5). Presenting a complete estimate of $\|\mathcal{R}'_H\|_1$ is complicated. However, we note that if $\sigma^-_{\max}/\sigma^-_{\min}$ can be bounded, which is a common scenario in practice, then an apriori estimate of $\|\mathcal{R}'_H\|_1$ could be achieved.

After proving the inf-sup stability of $V_H^\infty$, an error estimate of the global solution can be derived. Taking $e = u - u_H^\infty$, we have $a(e, w_H) = 0$ for all $w_H \in V_H^\infty$. Therefore, Lemma 5.5 gives that $e \in W$. Then, we can show that

$$\begin{aligned} \|e\|_{\tilde{a}} &\leq \frac{1}{\Lambda_W} \sup_{w \in W} \frac{a(e, w)}{\|w\|_{\tilde{a}}} = \frac{1}{\Lambda_W} \sup_{w \in W} \frac{a(u, w)}{\|w\|_{\tilde{a}}} = \frac{1}{\Lambda_W} \sup_{w \in W} \frac{\int_\Omega f w \, \mathrm{d}x}{\|w\|_{\tilde{a}}} \\ &\leq \frac{1}{\Lambda_W} \sup_{w \in W} \frac{\|f\|_{\tilde{s}^*} \|w\|_{\tilde{s}}}{\|w\|_{\tilde{a}}} \quad \text{(by the Cauchy–Schwarz inequality)} \\ &\leq \frac{\sqrt{\epsilon}}{\Lambda_W} \|f\|_{\tilde{s}^*}, \quad \text{(by the estimate (17))} \end{aligned}$$

where

$$\|f\|_{\tilde{s}^*} \coloneqq \left( \int_\Omega |\mu|^{-1} |f|^2 \, \mathrm{d}x \right)^{1/2} \approx H \max\left\{ 1/\left(\sigma^+_{\min}\right)^{1/2}, 1/\left(\sigma^-_{\min}\right)^{1/2} \right\} \|f\|_{L^2(\Omega)}.$$

We summarize all results presented above in the following theorem.

**Theorem 5.6.** *There exist positive constants $\Upsilon'$ and $\epsilon'$ such that for any $\Upsilon \geq \Upsilon'$ and $\epsilon \leq \epsilon'$, the solution $u_H^\infty$ in (18) exists and is unique. Moreover, the following estimate*

$$\|u - u_H^\infty\|_{\tilde{a}} \leq \frac{\sqrt{\epsilon}}{\Lambda_W} \|f\|_{\tilde{s}^*},$$

*holds.*

## 5.2 Local version

When turning to the practical method outlined in Section 3, the initial concern pertains to the existence of multiscale bases as defined in (8). However, a challenge arises due to the incompatibility between the definition of an oversampling region $K_i^m$ and $\mathtt{T}$-coercivity. Consequently, we cannot obtain a result similar to Proposition 5.3 for the local version. Nevertheless, it is worth noting that the numerical experiments presented in Section 4 did not encounter any lack of well-posedness at the discrete level. To proceed with the analysis, we introduce the following assumption.

**Assumption IV.** There exists a list of subdomains $\{\hat{K}_i^1, \hat{K}_i^2, \ldots\}$ that fulfills the requirements listed below:

1. Each $\hat{K}_i^m$ consists of coarse elements from $\mathcal{K}_H$.
2. There exists an inclusion relation $K_i \subset \hat{K}_i^1 \subset \hat{K}_i^2 \subset \cdots \subset$, such that $\operatorname{dist}(\partial\hat{K}_i^m, \partial\hat{K}_i^{m+1}) \geq C_{\mathrm{msh}} H$.
3. For any $v \in V$ with $\operatorname{supp} v \subset \overline{\hat{K}_i^m}$ or $\overline{\Omega \setminus \hat{K}_i^m}$, it holds that $\operatorname{supp} \mathcal{T} v \subset \overline{\hat{K}_i^m}$ or $\overline{\Omega \setminus \hat{K}_i^m}$ accordingly.

A construction of such subdomains can be found in [19], where a method called symmetrization is described. However, implementing such a construction (i.e., replacing $K_i^m$ in (8)) is highly impractical due to the lack of clarity associated with the definition of the operator $\mathcal{T}$. The fulfillment of the first and second requirements allows us to employ conventional cut-off functions. Furthermore, the third requirement suggests that, to some extent, T-coercivity can be preserved within these subdomains. In a sense, this complies with the definition of geometrically-based T-coercivity operators [44, 10, 20]. To distinguish the local problems utilized for analysis from the practical method in Section 3, we add a hat notation "$\hat{\cdot}$" (e.g., $\hat{V}_i^m = H_0^1(\hat{K}_i^m)$) to indicate that this term is associated with $\hat{K}_i^m$ rather than $K_i^m$. Applying the same technique as in establishing Proposition 5.3, we can obtain the well-posedness of the modified local operator $\hat{\mathcal{G}}_i^m$ associated with the coarse element $K_i$:

$$\begin{aligned} &\text{find } \hat{\mathcal{G}}_i^m \psi \in \hat{V}_i^m \text{ s.t. } \forall w \in \hat{V}_i^m, \\ &a(\hat{\mathcal{G}}_i^m \psi, w)_{\hat{K}_i^m} + s(\mathcal{P}_H \hat{\mathcal{G}}_i^m \psi, \mathcal{P}_H w)_{\hat{K}_i^m} = s(\mathcal{P}_H \psi, \mathcal{P}_H w)_{K_i}. \end{aligned} \tag{19}$$

To facilitate the analysis, we introduce the following notations for constants that are from (13) to (16):

$$\begin{aligned} C_0 &\coloneqq \max\Big\{\Big(1 + 8\|\mathcal{R}\|_0^2/\Upsilon\Big)^{1/2},\ \sqrt{2}\Big\}, \\ C_1 &\coloneqq \max\Big\{\Big(1 + 8\|\mathcal{R}\|_1^2/\Upsilon\Big)^{1/2},\ \sqrt{2}\Big\}, \\ C_2 &\coloneqq \min\Big\{1 - \|\mathcal{R}\|_1/\sqrt{\Upsilon} - \epsilon\|\mathcal{R}\|_0/\sqrt{\Upsilon},\ 1 - \|\mathcal{R}\|_0/\sqrt{\Upsilon}\Big\}. \end{aligned}$$

We can see that $C_0$, $C_1$, and $C_2$ can all be bounded from above and below providing that $\Upsilon$ is sufficiently large and $\epsilon$ is sufficiently small, which has also been stated in Proposition 5.3 and Theorem 5.6. A key ingredient employed in the analysis is multiplying by a cut-off function $\hat{\chi}_i^{m-1,m} \in C^{0,1}(\Omega)$ defined as

$$\hat{\chi}_i^{m-1,m} = \begin{cases} 1, \text{ in } \hat{K}_i^{m-1}, \\ 0, \text{ in } \Omega \setminus \hat{K}_i^m, \end{cases}$$

with $0 \leq \hat{\chi}_i^{m-1,m} \leq 1$ in $\hat{K}_i^m \setminus \hat{K}_i^{m-1}$. By carefully controlling the parameter $\mu_{\mathrm{msh}}$, we can derive the inequality

$$\Big|\nabla \hat{\chi}_i^{m-1,m}\Big|^2 |\sigma| \leq |\mu|$$

on $\Omega$ for any $i$ and $m$. Taking $\hat{\mathcal{G}}^m \coloneqq \sum_{i=1}^{N_{\mathrm{elem}}} \hat{\mathcal{G}}_i^m$, We shall prove an estimate of $\mathcal{G}^\infty - \hat{\mathcal{G}}^m$ in Theorem 5.9 below as the first main result in this subsection, while Lemmas 5.7 and 5.8 as presented in [50] are reformulated next to reflect the current sign-changing context.

**Lemma 5.7.** *There exists a positive constant $\theta$ with $\theta < 1$ that depends on $C_0$, $C_1$, and $C_2$ such that for any $m \geq 1$, $1 \leq i \leq N_{\mathrm{elem}}$ and $\psi \in L^2(\Omega)$,*

$$\|\mathcal{G}_i^\infty \psi\|_{\tilde{a},\Omega\setminus\hat{K}_i^m}^2 + \|\mathcal{P}_H \mathcal{G}_i^\infty \psi\|_{\tilde{s},\Omega\setminus\hat{K}_i^m}^2 \leq \theta^m \Big(\|\mathcal{G}_i^\infty \psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H \mathcal{G}_i^\infty \psi\|_{\tilde{s}}^2\Big).$$

**Lemma 5.8.** *It holds that for any $m \geq 1$, $1 \leq i \leq N_{\mathrm{elem}}$ and $\psi \in L^2(\Omega)$,*

$$\|(\mathcal{G}_i^{\infty} - \hat{\mathcal{G}}_i^m)\psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H(\mathcal{G}_i^{\infty} - \hat{\mathcal{G}}_i^m)\psi\|_{\tilde{s}}^2 \leq C\theta^{m-1}\Big(\|\mathcal{G}_i^{\infty}\psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H\mathcal{G}_i^{\infty}\psi\|_{\tilde{s}}^2\Big),$$

*where $\theta$ here is identical to the one in Lemma 5.7, and $C$ is a positive constant that depends on $C_0$, $C_1$, and $C_2$.*

**Theorem 5.9.** *There exist positive constants $\Upsilon'$ and $\epsilon'$ such that for any $\Upsilon \geq \Upsilon'$ and $\epsilon \leq \epsilon'$, it holds that for any $\psi \in L^2(\Omega)$,*

$$\|(\mathcal{G}^{\infty} - \hat{\mathcal{G}}^m)\psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H(\mathcal{G}^{\infty} - \hat{\mathcal{G}}^m)\psi\|_{\tilde{s}}^2 \leq C(m+1)^d\theta^{m-1}\|\mathcal{P}_H\psi\|_{\tilde{s}}^2,$$

*where $\theta$ and $C$ depend on $C_0$, $C_1$, and $C_2$ with $\theta < 1$.*

In the proof of Theorem 5.9, we require an assumption that specifies the growth of the size of $\hat{K}_i^m$ with $m$.

**Assumption V.** The number of coarse elements within $\hat{K}_i^m$ satisfies a relation:

$$\#\Big\{K \in \mathcal{K}_H \mid K \subset \hat{K}_i^m\Big\} \leq C_{\mathrm{ol}} m^d,$$

where $C_{\mathrm{ol}}$ depends solely on the mesh quality.

We first prove Lemma 5.7.

*Proof.* Taking $z_i \coloneqq (1 - \hat{\chi}_i^{m-1,m})\mathcal{G}_i^{\infty}\psi$, we can see that $z_i$ is supported in $\Omega \setminus \hat{K}_i^{m-1}$ and hence $(\operatorname{supp}\mathcal{T}z_i) \cap K_i = \varnothing$ according to assumption IV.3. Substituting $\mathcal{T}z_i$ for $w$ in the variational form (10), we have

$$a(\mathcal{G}_i^{\infty}\psi, \mathcal{T}z_i) + s(\mathcal{P}_H\mathcal{G}_i^{\infty}\psi, \mathcal{P}_H\mathcal{T}z_i) = 0.$$

As a result, we can formulate a decomposition as follows:

$$\begin{aligned} &a(z_i, \mathcal{T}z_i) + s(\mathcal{P}_H z_i, \mathcal{P}_H\mathcal{T}z_i) = -a(\hat{\chi}_i^{m-1,m}\mathcal{G}_i^{\infty}\psi, \mathcal{T}z_i) - s(\mathcal{P}_H\hat{\chi}_i^{m-1,m}\mathcal{G}_i^{\infty}\psi, \mathcal{P}_H\mathcal{T}z_i) \\ &= -\underbrace{\int_{\Omega} \sigma\mathcal{G}_i^{\infty}\psi\nabla\hat{\chi}_i^{m-1,m}\cdot\nabla\mathcal{T}z_i\,\mathrm{d}x}_{J_1} - \underbrace{\int_{\Omega}\sigma\hat{\chi}_i^{m-1,m}\nabla\mathcal{G}_i^{\infty}\psi\cdot\nabla\mathcal{T}z_i\,\mathrm{d}x}_{J_2} - \underbrace{\int_{\Omega}\mu\mathcal{P}_H\Big(\hat{\chi}_i^{m-1,m}\mathcal{G}_i^{\infty}\psi\Big)\mathcal{P}_H\mathcal{T}z_i\,\mathrm{d}x}_{J_3}. \end{aligned}$$

For $J_1$, by the Cauchy-Schwarz inequality, it is evident that

$$\begin{aligned} |J_1| &\leq \left(\int_{\Omega}|\sigma|\left|\nabla\hat{\chi}_i^{m-1,m}\right|^2|\mathcal{G}_i^{\infty}\psi|^2\,\mathrm{d}x\right)^{1/2}\left(\int_{\Omega}|\sigma|\,|\nabla\mathcal{T}z_i|^2\,\mathrm{d}x\right)^{1/2} \\ &\leq \|\mathcal{G}_i^{\infty}\psi\|_{\tilde{s},\hat{K}_i^m\setminus\hat{K}_i^{m-1}}\|\mathcal{T}z_i\|_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} \\ &\leq C_0\Big(\|\mathcal{P}_H\mathcal{G}_i^{\infty}\psi\|_{\tilde{s},\hat{K}_i^m\setminus\hat{K}_i^{m-1}}^2 + \epsilon\|\mathcal{G}_i^{\infty}\psi\|_{\tilde{a},\hat{K}_i^m\setminus\hat{K}_i^{m-1}}^2\Big)^{1/2}\|z_i\|_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}}, \end{aligned}$$

where the last line follows from (15) and Lemma 5.1. For $J_2$, we have

$$\begin{aligned} |J_2| &\leq \left(\int_{\hat{K}_i^m\setminus\hat{K}_i^{m-1}}|\sigma|\left|\hat{\chi}_i^{m-1,m}\right|^2|\nabla\mathcal{G}_i^{\infty}\psi|^2\,\mathrm{d}x\right)^{1/2}\left(\int_{\Omega}|\sigma|\,|\nabla\mathcal{T}z_i|^2\,\mathrm{d}x\right)^{1/2} \\ &\leq \|\mathcal{G}_i^{\infty}\psi\|_{\tilde{a},\hat{K}_i^m\setminus\hat{K}_i^{m-1}}\|\mathcal{T}z_i\|_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} \\ &\leq C_0\|\mathcal{G}_i^{\infty}\psi\|_{\tilde{a},\hat{K}_i^m\setminus\hat{K}_i^{m-1}}\|z_i\|_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}}. \end{aligned}$$

For $J_3$, we can similarly show that

$$\begin{aligned}
|J_3| &\le \left(\int_{\hat{K}_i^m\setminus\hat{K}_i^{m-1}} |\mu| \left|\mathcal{P}_H\left(\hat{\chi}_i^{m-1,m}\mathcal{G}_i^\infty\psi\right)\right|^2 \mathrm{d}x\right)^{1/2} \left(\int_\Omega |\mu| \left|\mathcal{P}_H\mathcal{T}z_i\right|^2 \mathrm{d}x\right)^{1/2} \\
&= \left\|\mathcal{P}_H\left(\hat{\chi}_i^{m-1,m}\mathcal{G}_i^\infty\psi\right)\right\|_{\tilde{s},\hat{K}_i^m\setminus\hat{K}_i^{m-1}} \left\|\mathcal{P}_H\mathcal{T}z_i\right\|_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}} \quad \text{(By assumption IV.3)} \\
&\le \left\|\hat{\chi}_i^{m-1,m}\mathcal{G}_i^\infty\psi\right\|_{\tilde{s},\hat{K}_i^m\setminus\hat{K}_i^{m-1}} \left\|\mathcal{T}z_i\right\|_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}} \\
&\le C_1\left\|\mathcal{G}_i^\infty\psi\right\|_{\tilde{s},\hat{K}_i^m\setminus\hat{K}_i^{m-1}} \left\|z_i\right\|_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}} \\
&\le C_1\Big(\left\|\mathcal{P}_H\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{s},\hat{K}_i^m\setminus\hat{K}_i^{m-1}} + \epsilon\left\|\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{a},\hat{K}_i^m\setminus\hat{K}_i^{m-1}}\Big)^{1/2} \Big(\left\|\mathcal{P}_H z_i\right\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}} + \epsilon\left\|z_i\right\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}}\Big)^{1/2}.
\end{aligned}$$

Recalling (13) and (14), we can obtain that

$$a(z_i,\mathcal{T}z_i) + s(\mathcal{P}_H z_i,\mathcal{P}_H\mathcal{T}z_i) \ge C_2\Big(\left\|z_i\right\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} + \left\|\mathcal{P}_H z_i\right\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}}\Big)$$

Providing that $\epsilon$ is small enough ($\epsilon \le 1$), we have

$$\begin{aligned}
&\Big(\left\|\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^m} + \left\|\mathcal{P}_H\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^m}\Big)^{1/2} \le \Big(\left\|z_i\right\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} + \left\|\mathcal{P}_H z_i\right\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}}\Big)^{1/2} \\
\le\ & \frac{2C_0+C_1}{C_2}\Big(\left\|\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{a},\hat{K}_i^m\setminus\hat{K}_i^{m-1}} + \left\|\mathcal{P}_H\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{s},\hat{K}_i^m\setminus\hat{K}_i^{m-1}}\Big)^{1/2}.
\end{aligned}$$

Note that

$$\begin{aligned}
&\left\|\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} + \left\|\mathcal{P}_H\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}} \\
=\ & \left\|\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^m} + \left\|\mathcal{P}_H\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^m} + \left\|\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{a},\hat{K}_i^m\setminus\hat{K}_i^{m-1}} + \left\|\mathcal{P}_H\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{s},\hat{K}_i^m\setminus\hat{K}_i^{m-1}}.
\end{aligned}$$

We hence derive an iteration relation

$$\begin{aligned}
&\left\|\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^m} + \left\|\mathcal{P}_H\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^m} \\
&\qquad \le \left(1+\left(\frac{C_2}{2C_0+C_1}\right)^2\right)^{-1}\Big(\left\|\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} + \left\|\mathcal{P}_H\mathcal{G}_i^\infty\psi\right\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}}\Big),
\end{aligned}$$

which completes the proof. □

We then prove Lemma 5.8.

*Proof.* We now take $z_i = (\mathcal{G}_i^\infty - \hat{\mathcal{G}}_i^m)\psi$ and introduce a decomposition

$$z_i = \underbrace{(1-\hat{\chi}_i^{m-1,m})\mathcal{G}_i^\infty\psi}_{z_i'} + \underbrace{(\hat{\chi}_i^{m-1,m}-1)\hat{\mathcal{G}}_i^m\psi + \hat{\chi}_i^{m-1,m}z_i}_{z_i''}.$$

We can observe that $z_i'' \in \hat{V}_i^m$, which implies $\mathcal{T}z_i'' \in \hat{V}_i^m$. Therefore, it is easy to see that $a(z_i,\mathcal{T}z_i'') + s(\mathcal{P}_H z_i,\mathcal{P}_H\mathcal{T}z_i'') = 0$ based on the definitions of $\mathcal{G}_i^\infty$ and $\hat{\mathcal{G}}_i^m$ in (10) and (19). Now, our task is to estimate $a(z_i,\mathcal{T}z_i') + s(\mathcal{P}_H z_i,\mathcal{P}_H\mathcal{T}z_i')$, and a starting step is

$$a(z_i,\mathcal{T}z_i') + s(\mathcal{P}_H z_i,\mathcal{P}_H\mathcal{T}z_i') \le \left\|z_i\right\|_{\tilde{a}}\left\|\mathcal{T}z_i'\right\|_{\tilde{a}} + \left\|\mathcal{P}_H z_i\right\|_{\tilde{s}}\left\|\mathcal{P}_H\mathcal{T}z_i'\right\|_{\tilde{s}}.$$

Recalling that $\left\|\mathcal{T}z_i'\right\|_{\tilde{a}} \le C_0\left\|z_i'\right\|_{\tilde{a}}$ and using a similar technique as in the proof of Lemma 5.7, we can obtain that

$$\left\|z_i'\right\|^2_{\tilde{a}} \le 2\int_{\Omega\setminus\hat{K}_i^{m-1}} |\sigma|\left|1-\hat{\chi}_i^{m-1,m}\right|^2 \left|\nabla\mathcal{G}_i^\infty\psi\right|^2 \mathrm{d}x + 2\int_{\hat{K}_i^m\setminus\hat{K}_i^{m-1}} |\sigma|\left|\nabla\hat{\chi}_i^{m-1,m}\right|^2 \left|\mathcal{G}_i^\infty\psi\right|^2 \mathrm{d}x$$

$$\leq 2\|\mathcal{G}_i^\infty\psi\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} + 2\int_{\Omega\setminus\hat{K}_i^{m-1}} |\mu|\,|\mathcal{G}_i^\infty\psi|^2\,\mathrm{d}x \quad \text{(by assumption IV.2)}$$
$$\leq 2(1+\epsilon)\|\mathcal{G}_i^\infty\psi\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} + 4\|\mathcal{P}_H\mathcal{G}_i^\infty\psi\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}}.$$

519 Moreover, we can see that $\|\mathcal{P}_H\mathcal{T}z_i'\|_{\tilde{s}} \leq \|\mathcal{T}z_i'\|_{\tilde{s}} \leq C_1\|z_i'\|_{\tilde{s}}$ and

$$\|z_i'\|^2_{\tilde{s}} = \int_\Omega |\mu|\left|1-\hat{\chi}_i^{m-1,m}\right|^2 |\mathcal{G}_i^\infty\psi|^2\,\mathrm{d}x \leq \|\mathcal{G}_i^\infty\psi\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}}$$
$$\leq \epsilon\|\mathcal{G}_i^\infty\psi\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} + \|\mathcal{P}_H\mathcal{G}_i^\infty\psi\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}}.$$

520 Then, we are close to the target estimate as

$$C_2\Big(\|z_i\|^2_{\tilde{a}} + \|\mathcal{P}_H z_i\|^2_{\tilde{s}}\Big) \leq a(z_i,\mathcal{T}z_i) + s(\mathcal{P}_H z_i,\mathcal{P}_H\mathcal{T}z_i) = a(z_i,\mathcal{T}z_i') + s(\mathcal{P}_H z_i,\mathcal{P}_H\mathcal{T}z_i')$$
$$\leq \Big(\|z_i\|^2_{\tilde{a}} + \|\mathcal{P}_H z_i\|^2_{\tilde{s}}\Big)^{1/2}\Big(\|\mathcal{T}z_i'\|^2_{\tilde{a}} + \|\mathcal{P}_H\mathcal{T}z_i'\|^2_{\tilde{s}}\Big)^{1/2}.$$

521 Suppose that $\epsilon$ is small enough, we can obtain that

$$\|z_i\|^2_{\tilde{a}} + \|\mathcal{P}_H z_i\|^2_{\tilde{s}} \leq \left(\frac{4C_0+C_1}{C_2}\right)^2\Big(\|\mathcal{G}_i^\infty\psi\|^2_{\tilde{a},\Omega\setminus\hat{K}_i^{m-1}} + \|\mathcal{P}_H\mathcal{G}_i^\infty\psi\|^2_{\tilde{s},\Omega\setminus\hat{K}_i^{m-1}}\Big),$$

522 and we hence derive the desired result by utilizing Lemma 5.7. □

523 We are now ready to prove Theorem 5.9.

524 *Proof.* We denote $z_j \coloneqq (\mathcal{G}_j^\infty - \hat{\mathcal{G}}_j^m)\psi$ and $z \coloneqq \sum_{j=1}^{N_{\mathrm{elem}}} z_j$. We take a decomposition of $z$ as

$$z = \underbrace{(1-\hat{\chi}_i^{m,m+1})z_i}_{z'} + \underbrace{\hat{\chi}_i^{m,m+1}z}_{z''},$$

525 where $i$ is arbitrary chosen from $1,\ldots,N_{\mathrm{elem}}$. Again, we can observe that $\mathrm{supp}(\mathcal{T}z')\cap K_i = \varnothing$ and hence
526 have

$$a(\mathcal{G}_i^\infty\psi,\mathcal{T}z') + s(\mathcal{P}_H\mathcal{G}_i^\infty\psi,\mathcal{P}_H\mathcal{T}z') = a(\hat{\mathcal{G}}_i^m\psi,\mathcal{T}z') + s(\mathcal{P}_H\hat{\mathcal{G}}_i^m\psi,\mathcal{P}_H\mathcal{T}z') = 0,$$

527 which leads to

$$a(z_i,\mathcal{T}z) + s(\mathcal{P}_H z_i,\mathcal{P}_H\mathcal{T}z) = a(z_i,\mathcal{T}z'') + s(\mathcal{P}_H z_i,\mathcal{P}_H\mathcal{T}z'') \leq \|\mathcal{T}z''\|_{\tilde{a}}\|z_i\|_{\tilde{a}} + \|\mathcal{P}_H\mathcal{T}z''\|_{\tilde{s}}\|\mathcal{P}_H z_i\|_{\tilde{s}}.$$

528 Similarly, we can estimate $\|\mathcal{T}z''\|_{\tilde{a}}$ and $\|\mathcal{P}_H\mathcal{T}z''\|_{\tilde{s}}$ as

$$\|\mathcal{T}z''\|_{\tilde{a}} \leq C_0\|\hat{\chi}_i^{m,m+1}z\|_{\tilde{a}} \leq C_0\Big(\|z\|_{\tilde{a},\hat{K}_i^{m+1}} + \|z\|_{\tilde{s},\hat{K}_i^{m+1}}\Big) \leq C_0\Big((1+\sqrt{\epsilon})\|z\|_{\tilde{a},\hat{K}_i^{m+1}} + \|\mathcal{P}_H z\|_{\tilde{s},\hat{K}_i^{m+1}}\Big)$$

529 and

$$\|\mathcal{P}_H\mathcal{T}z''\|_{\tilde{s}} \leq \|\mathcal{T}z''\|_{\tilde{s}} \leq C_1\|\hat{\chi}_i^{m,m+1}z\|_{\tilde{s}} \leq C_1\|z\|_{\tilde{s},\hat{K}_i^{m+1}} \leq C_1\Big(\sqrt{\epsilon}\|z\|_{\tilde{a},\hat{K}_i^{m+1}} + \|\mathcal{P}_H z\|_{\tilde{s},\hat{K}_i^{m+1}}\Big).$$

530 To simplify the expression, we assume that $\epsilon \leq 1$ and derive that

$$a(z_i,\mathcal{T}z) + s(\mathcal{P}_H z_i,\mathcal{P}_H\mathcal{T}z) \leq c_0\Big(\|z\|^2_{\tilde{a},\hat{K}_i^{m+1}} + \|\mathcal{P}_H z\|^2_{\tilde{s},\hat{K}_i^{m+1}}\Big)^{1/2}\Big(\|z_i\|^2_{\tilde{a}} + \|\mathcal{P}_H z_i\|^2_{\tilde{s}}\Big)^{1/2},$$

531 where $c_0$ depends on $C_0$, $C_1$, and $C_2$. Based on assumption V, it holds that

$$\sum_{i=1}^{N_{\mathrm{elem}}}\Big(\|z\|^2_{\tilde{a},\hat{K}_i^{m+1}} + \|\mathcal{P}_H z\|^2_{\tilde{s},\hat{K}_i^{m+1}}\Big) \leq C_{\mathrm{ol}}(m+1)^d\Big(\|z\|^2_{\tilde{a}} + \|\mathcal{P}_H z\|^2_{\tilde{s}}\Big).$$

Therefore, applying the Cauchy–Schwarz inequality,

$$
\begin{aligned}
C_2\Big(\|z\|_{\tilde{a}}^2 + \|\mathcal{P}_H z\|_{\tilde{s}}^2\Big) &\le \sum_{i=1}^{N_{\mathrm{elem}}} \big(a(z_i, \mathcal{T}z) + s(\mathcal{P}_H z_i, \mathcal{P}_H \mathcal{T} z)\big) \\
&\le c_0 \left(\sum_{i=1}^{N_{\mathrm{elem}}} \Big(\|z\|_{\tilde{a},\hat{K}_i^{m+1}}^2 + \|\mathcal{P}_H z\|_{\tilde{s},\hat{K}_i^{m+1}}^2\Big)\right)^{1/2} \left(\sum_{i=1}^{N_{\mathrm{elem}}} \Big(\|z_i\|_{\tilde{a}}^2 + \|\mathcal{P}_H z_i\|_{\tilde{s}}^2\Big)\right)^{1/2} \\
&\le c_0 \sqrt{C_{\mathrm{ol}}} (m+1)^{d/2} \Big(\|z\|_{\tilde{a}}^2 + \|\mathcal{P}_H z\|_{\tilde{s}}^2\Big)^{1/2} \left(\sum_{i=1}^{N_{\mathrm{elem}}} \Big(\|z_i\|_{\tilde{a}}^2 + \|\mathcal{P}_H z_i\|_{\tilde{s}}^2\Big)\right)^{1/2}.
\end{aligned}
$$

It has been shown in Lemma 5.8 that, for any $i$,

$$
\|z_i\|_{\tilde{a}}^2 + \|\mathcal{P}_H z_i\|_{\tilde{s}}^2 \le C_* \theta^{m-1} \Big(\|\mathcal{G}_i^\infty \psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H \mathcal{G}_i^\infty \psi\|_{\tilde{s}}^2\Big),
$$

and we turn to provide a bound on $\sum_{i=1}^{N_{\mathrm{elem}}} \|\mathcal{G}_i^\infty \psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H \mathcal{G}_i^\infty \psi\|_{\tilde{s}}^2$. Note that by the definition of (19),

$$
\begin{aligned}
&C_2\Big\{\|\mathcal{G}_i^\infty \psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H \mathcal{G}_i^\infty \psi\|_{\tilde{s}}^2\Big\} \le a(\mathcal{G}_i^\infty \psi, \mathcal{T}\mathcal{G}_i^\infty \psi) + s(\mathcal{P}_H \mathcal{G}_i^\infty \psi, \mathcal{P}_H \mathcal{T} \mathcal{G}_i^\infty \psi) \\
&= s(\mathcal{P}_H \psi, \mathcal{P}_H \mathcal{T} \mathcal{G}_i^\infty \psi)_{K_i} \le \|\mathcal{P}_H \psi\|_{\tilde{s},K_i} \|\mathcal{P}_H \mathcal{T} \mathcal{G}_i^\infty \psi\|_{\tilde{s},K_i} \le \|\mathcal{P}_H \psi\|_{\tilde{s},K_i} \|\mathcal{T} \mathcal{G}_i^\infty \psi\|_{\tilde{s},K_i} \\
&\le \|\mathcal{P}_H \psi\|_{\tilde{s},K_i} \|\mathcal{T} \mathcal{G}_i^\infty \psi\|_{\tilde{s}} \le C_1 \|\mathcal{P}_H \psi\|_{\tilde{s},K_i} \|\mathcal{G}_i^\infty \psi\|_{\tilde{s}} \\
&\le C_1 \|\mathcal{P}_H \psi\|_{\tilde{s},K_i} \Big(\sqrt{\epsilon} \|\mathcal{G}_i^\infty \psi\|_{\tilde{a}} + \|\mathcal{P}_H \mathcal{G}_i^\infty \psi\|_{\tilde{s}}\Big),
\end{aligned}
$$

which gives that

$$
\|\mathcal{G}_i^\infty \psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H \mathcal{G}_i^\infty \psi\|_{\tilde{s}}^2 \le c_1 \|\mathcal{P}_H \psi\|_{\tilde{s},K_i}^2.
$$

where $c_1$ depends on $C_1$ and $C_2$. We finally completed the proof by collecting all the estimates obtained above. □

We similarly denote $\hat{\mathcal{G}}^m \coloneqq \sum_{i=1}^{N_{\mathrm{elem}}} \hat{\mathcal{G}}_i^m$, where we implicitly lift the image of $\hat{\mathcal{G}}_i^m$ to $V$. We now define the multiscale solution in the local version as the solution to the following problem: for $m \ge 1$,

$$
\text{find } \hat{u}_H^m \in \hat{V}_H^m \text{ s.t. } \forall w_H \in \hat{V}_H^m,\ a(\hat{u}_H^m, w_H) = \int_\Omega f w_H \,\mathrm{d}x, \tag{20}
$$

where $\hat{V}_H^m \coloneqq \operatorname{im} \hat{\mathcal{G}}^m \subset V$. The remainder of this section is dedicated to proving an error estimate for the multiscale solution, namely Theorem 5.10 below. Once again, we need to verify the well-posedness of (20), or equivalently, demonstrate that

$$
\inf_{v_H \in \hat{V}_H^m} \sup_{w_H \in \hat{V}_H^m} \frac{a(v_H, w_H)}{\|v_H\|_{\tilde{a}} \|w_H\|_{\tilde{a}}}
$$

can be bounded from below by a positive constant that is independent of $m$ and $H$. Thanks to Theorem 5.6, the inf-sup condition holds for the global problem, while a global basis and a local basis are connected through $\psi$. Those are the main ingredients to prove the inf-sup condition for the local problem. First, for any $v_H \in \hat{V}_H^m$, we can find $\psi$ such that $v_H = \hat{\mathcal{G}}^m \psi$. Next, We choose $v_H' = \mathcal{G}^\infty \psi \in V_H^\infty$. Recalling the inf-sup stability on $V_H^\infty$, we can find $w_H' = \mathcal{G}^\infty \phi \in V_H^\infty$ such that $a(v_H', w_H') \ge \Lambda^\infty \|v_H'\|_{\tilde{a}} \|w_H'\|_{\tilde{a}}$. Similarly, if we denote $w_H = \hat{\mathcal{G}}^m \phi$, we have

$$
\begin{aligned}
a(v_H, w_H) &= a(v_H', w_H') + a(v_H', w_H - w_H') + a(v_H - v_H', w_H) \\
&\ge \Lambda^\infty \|v_H'\|_{\tilde{a}} \|w_H'\|_{\tilde{a}} - \|v_H'\|_{\tilde{a}} \|w_H - w_H'\|_{\tilde{a}} - \|v_H - v_H'\|_{\tilde{a}} \|w_H\|_{\tilde{a}}.
\end{aligned}
$$

Therefore, if we can show that

$$
\|(\mathcal{G}^\infty - \hat{\mathcal{G}}^m)\psi\|_{\tilde{a}} \le \eta \|\mathcal{G}^\infty \psi\|_{\tilde{a}}, \tag{21}
$$

where $\eta$ is a small constant, we can finish the proof. The reason is that we now have $1/2\|v_H\|_{\tilde{a}} \leq \|v_H'\|_{\tilde{a}} \leq 3/2\|v_H\|_{\tilde{a}}$ and $1/2\|w_H\|_{\tilde{a}} \leq \|w_H'\|_{\tilde{a}} \leq 3/2\|w_H\|_{\tilde{a}}$ by the smallness of $\eta$. Consequently, we can deduce that

$$a(v_H, w_H) \geq (\Lambda^{\infty}/4 - 9\eta/4 - 3\eta/2)\|v_H\|_{\tilde{a}}\|w_H\|_{\tilde{a}}.$$

By comparing (21) with Theorem 5.9, our remaining task is to demonstrate $\|\mathcal{P}_H\psi\|_{\tilde{s}}$ can be bounded by $\|\mathcal{G}^{\infty}\psi\|_{\tilde{a}}$. According to Lemma 5.4, it not harmful to assume that $\psi \in V$ with $\|\psi\|_{\tilde{a}} \leq C_{\rm inv}\|\mathcal{P}_H\psi\|_{\tilde{s}}$. Taking $\mathcal{T}\psi$ as a test function, we can obtain that

$$a(\mathcal{G}^{\infty}\psi, \mathcal{T}\psi) + s(\mathcal{P}_H\mathcal{G}^{\infty}\psi, \mathcal{P}_H\mathcal{T}\psi) = s(\mathcal{P}_H\psi, \mathcal{P}_H\mathcal{T}\psi).$$

The estimate (14) says that

$$C'\|\mathcal{P}_H\psi\|_{\tilde{s}}^2 - C''\|\psi\|_{\tilde{a}}^2 \leq s(\mathcal{P}_H\psi, \mathcal{P}_H\mathcal{T}\psi)$$

where $C' \to 1$ and $C'' \to 0$ as $\Upsilon \to \infty$. Therefore, if $\Upsilon$ is large enough, we can show that

$$\begin{aligned} c_0\|\mathcal{P}_H\psi\|_{\tilde{s}}^2 &\leq \big(C' - C''C_{\rm inv}^2\big)\|\mathcal{P}_H\psi\|_{\tilde{s}}^2 \leq C'\|\mathcal{P}_H\psi\|_{\tilde{s}}^2 - C''\|\psi\|_{\tilde{a}}^2 \\ &\leq s(\mathcal{P}_H\psi, \mathcal{P}_H\mathcal{T}\psi) \leq \|\mathcal{G}^{\infty}\psi\|_{\tilde{a}}\|\mathcal{T}\psi\|_{\tilde{a}} + \|\mathcal{P}_H\mathcal{G}^{\infty}\psi\|_{\tilde{s}}\|\mathcal{P}_H\mathcal{T}\psi\|_{\tilde{s}} \\ &\leq c_1\sqrt{1 + C_{\rm inv}^2}\Big(\|\mathcal{G}^{\infty}\psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H\mathcal{G}^{\infty}\psi\|_{\tilde{s}}^2\Big)^{1/2}\Big(\|\psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H\psi\|_{\tilde{s}}^2\Big)^{1/2}, \end{aligned}$$

where $c_0$ and $c_1$ are positive constants and the last line follows the techniques in the previous proofs. We hence obtain that

$$\|\psi\|_{\tilde{a}}/C_{\rm inv} \leq \|\mathcal{P}_H\psi\|_{\tilde{s}} \leq c_2\Big(\|\mathcal{G}^{\infty}\psi\|_{\tilde{a}}^2 + \|\mathcal{P}_H\mathcal{G}^{\infty}\psi\|_{\tilde{s}}^2\Big)^{1/2}.$$

Meanwhile, utilizing Poincaré inequality, we can see that

$$\|\mathcal{P}_H\mathcal{G}^{\infty}\psi\|_{\tilde{s}} \leq \|\mathcal{G}^{\infty}\psi\|_{\tilde{s}} \leq C_{\rm po}H^{-1}\|\mathcal{G}^{\infty}\psi\|_{\tilde{a}}.$$

Hence, in conjunction with Theorem 5.9, we establish the existence of such $\eta$ in (21). Once the inf-sup condition is validated, the existence and uniqueness of the multiscale solution are guaranteed. Moreover, we can employ Céa's lemma to derive the error estimate. Specifically, we obtain

$$\|u - \hat{u}_H^m\|_{\tilde{a}} \leq c_3\|u - v_H\|_{\tilde{a}} \leq c_3\Big(\|u - u_H^{\infty}\|_{\tilde{a}} + \|u_H^{\infty} - v_H\|_{\tilde{a}}\Big),$$

where $v_H = \hat{\mathcal{G}}^m\psi \in \hat{V}_H^m$ if $u_H^{\infty} = \mathcal{G}^{\infty}\psi$. We have already shown that

$$\|u_H^{\infty} - v_H\|_{\tilde{a}} \leq c_4H^{-1}(m+1)^{d/2}\theta^{(m-1)/2}\|u_H^{\infty}\|_{\tilde{a}}.$$

Finally, we present the main theorem of this section.

**Theorem 5.10.** *There exist positive constants $\Upsilon'$ and $\epsilon'$ such that for any $\Upsilon \geq \Upsilon'$ and $\epsilon \leq \epsilon'$, the solution to* (20) *exists and is unique. Moreover, the following error estimate holds:*

$$\|u - \hat{u}_H^m\|_{\tilde{a}} \leq C_*\Big(1 + H^{-2}(m+1)^{d/2}\theta^{(m-1)/2}\Big)\|f\|_{\tilde{s}^*},$$

*where the positive constant $C_*$ and $\theta$ is independent of $H$ and $m$ with $\theta < 1$.*

# 6 Conclusions

We have proposed a multiscale computational method for solving sign-changing problems, utilizing the framework of CEM-GMsFEMs in the construction of multiscale basis functions. However, a direct application of the original CEM-GMsFEM encounters an immediate challenge during the construction of auxiliary spaces, as the generalized spectral problems can become ill-defined due to a non-positive definite right-hand

bilinear form. We have addressed this issue by replacing the coefficient with its absolute value in this step and explained that this modification complies with the T-coercivity theory. Moreover, we have focused on the relaxed version of the CEM-GMsFEM, which is more implementation-friendly as it eliminates the need to solve saddle-point problems. The numerical experiments conducted have highlighted several advantages of the proposed method: (1) the flexibility that coarse meshes do not require to resolve with interfaces, (2) the accuracy that remains stable under low regular exact solutions, and (3) the robustness in high contrast coefficient profiles. Due to technical difficulties, such as the nonlocality of the reflection operator $\mathcal{T}$, the final error estimates have been proved under several stringent assumptions, primarily related to coarse element partitions and coefficient contrast. However, we emphasize that the potential of the proposed method has been demonstrated through numerical experiments, and we firmly believe that the theoretical results can be further improved by relaxing the assumptions and developing more sophisticated analysis techniques.

The current implementation of the proposed method did not exploit the parallelism in constructing multiscale basis functions, and hence the computation time in the offline stage is awkwardly long, even longer than generating reference solutions on the fine mesh. Therefore, we plan to investigate the parallelization in the shared memory architecture of the offline stage in the future. We emphasize that the application of computational multiscale methods to sign-changing problems is much more appealing in large-scale simulations. As a matter of fact, the algebraic linear systems in this situation are no longer positive definite, which renders iterative solvers less efficient, even with preconditioning techniques. Consequently, in this context of direct solvers, the reduction of degrees of freedom by multiscale methods is expected to be more beneficial.

# Acknowledgments

EC's research is partially supported by the Hong Kong RGC General Research Fund (Project numbers: 14305222 and 14304021).

# References

[1] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden, *The heterogeneous multiscale method*, Acta Numerica, 21 (2012), pp. 1–87, https://doi.org/10.1017/S0962492912000025.

[2] R. F. Almgern, *An isotropic three dimensional structure with Poisson's ratio = −1*, Journal of Elasticity, 15 (1985), pp. 427–430, https://doi.org/10.1007/bf00042531.

[3] R. Altmann, P. Henning, and D. Peterseim, *Numerical homogenization beyond scale separation*, Acta Numerica, 30 (2021), pp. 1–86, https://doi.org/10.1017/S0962492921000015.

[4] I. Babuška and R. Lipton, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Modeling & Simulation. A SIAM Interdisciplinary Journal, 9 (2011), pp. 373–406, https://doi.org/10.1137/100791051.

[5] I. Babuška, R. Lipton, P. Sinz, and M. Stuebner, *Multiscale-spectral GFEM and optimal oversampling*, Computer Methods in Applied Mechanics and Engineering, 364 (2020), pp. 112960, 28, https://doi.org/10.1016/j.cma.2020.112960.

[6] M. Barré and P. Ciarlet Jr., *The T-coercivity approach for mixed problems*. In press in Comptes Rendus Mathematique, 2023, https://hal.science/hal-03820910.

[7] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, AMS Chelsea Publishing, Providence, RI, 2011, https://doi.org/10.1090/chel/374. Corrected reprint of the 1978 original [MR0503330].

[8] D. Boffi, F. Brezzi, and M. Fortin, *Mixed finite element methods and applications*, Springer, Berlin, Germany and Heidelberg, Germany, 2013, https://doi.org/10.1007/978-3-642-36519-5.

[9] A.-S. Bonnet-BenDhia, C. Carvalho, and P. Ciarlet Jr., *Mesh requirements for the finite element approximation of problems with sign-changing coefficients*, Numerische Mathematik, 138 (2018), pp. 801–838, https://doi.org/10.1007/s00211-017-0923-5.

[10] A.-S. Bonnet-BenDhia, L. Chesnel, and P. Ciarlet Jr., *T-coercivity for scalar interface problems between dielectrics and metamaterials*, ESAIM. Mathematical Modelling and Numerical Analysis, 46 (2012), pp. 1363–1387, https://doi.org/10.1051/m2an/2012006.

[11] A.-S. Bonnet-BenDhia, L. Chesnel, and P. Ciarlet Jr., *T-coercivity for the Maxwell problem with sign-changing coefficients*, Communications in Partial Differential Equations, 39 (2014), pp. 1007–1031, https://doi.org/10.1080/03605302.2014.892128.

[12] A.-S. Bonnet-BenDhia, L. Chesnel, and P. Ciarlet Jr., *Two-dimensional Maxwell's equations with sign-changing coefficients*, Applied Numerical Mathematics, 79 (2014), pp. 29–41, https://doi.org/10.1016/j.apnum.2013.04.006.

[13] A.-S. Bonnet-BenDhia, P. Ciarlet Jr., and C. M. Zwölf, *Time harmonic wave diffraction problems in materials with sign-shifting coefficients*, Journal of Computational and Applied Mathematics, 234 (2010), pp. 1912–1919, https://doi.org/10.1016/j.cam.2009.08.041.

[14] R. Bunoiu, K. Karim, and C. Timofte, *T-coercivity for the asymptotic analysis of scalar problems with sign-changing coefficients in thin periodic domains*, Electronic Journal of Differential Equations, 2021 (2021), pp. 1–22, https://doi.org/10.58997/ejde.2021.59.

[15] R. Bunoiu and K. Ramdani, *Homogenization of materials with sign changing coefficients*, Communications in Mathematical Sciences, 14 (2016), pp. 1137–1154, https://doi.org/10.4310/cms.2016.v14.n4.a13.

[16] R. Bunoiu, K. Ramdani, and C. Timofte, *T-coercivity for the homogenization of sign-changing coefficients scalar problems with extreme contrasts*, Mathematical Reports, Romanian Academy of Sciences, 24(74) (2022), pp. 113–123.

[17] R. Bunoiu, K. Ramdani, and C. Timofte, *Homogenization of a transmission problem with sign-changing coefficients and interfacial flux jump*, Communications in Mathematical Sciences, 21 (2023), pp. 2029–2049, https://doi.org/10.4310/cms.2023.v21.n7.a13.

[18] C. Carvalho, L. Chesnel, and P. Ciarlet Jr., *Eigenvalue problems with sign-changing coefficients*, Comptes Rendus Mathematique, 355 (2017), pp. 671–675, https://doi.org/10.1016/j.crma.2017.05.002.

[19] T. Chaumont-Frelet and B. Verfürth, *A generalized finite element method for problems with sign-changing coefficients*, ESAIM. Mathematical Modelling and Numerical Analysis, 55 (2021), pp. 939–967, https://doi.org/10.1051/m2an/2021007.

[20] L. Chesnel and P. Ciarlet Jr., *T-coercivity and continuous Galerkin methods: application to transmission problems with sign changing coefficients*, Numerische Mathematik, 124 (2013), pp. 1–29, https://doi.org/10.1007/s00211-012-0510-8.

[21] E. T. Chung and P. Ciarlet Jr., *A staggered discontinuous Galerkin method for wave propagation in media with dielectrics and meta-materials*, Journal of Computational and Applied Mathematics, 239 (2013), pp. 189–207, https://doi.org/10.1016/j.cam.2012.09.033.

[22] E. T. Chung, Y. Efendiev, and W. T. Leung, *Constraint energy minimizing generalized multiscale finite element method*, Computer Methods in Applied Mechanics and Engineering, 339 (2018), pp. 298–319, https://doi.org/10.1016/j.cma.2018.04.010.

[23] E. T. Chung, Y. Efendiev, and G. Li, *An adaptive GMsFEM for high-contrast flow problems*, Journal of Computational Physics, 273 (2014), pp. 54–76, https://doi.org/10.1016/j.jcp.2014.05.007.

[24] P. Ciarlet Jr., *T-coercivity: Application to the discretization of Helmholtz-like problems*, Computers & Mathematics with Applications, 64 (2012), pp. 22–34, https://doi.org/10.1016/j.camwa.2012.02.034.

[25] P. Ciarlet Jr. and M. Vohralík, *Localization of global norms and robust a posteriori error control for transmission problems with sign-changing coefficients*, ESAIM. Mathematical Modelling and Numerical Analysis, 52 (2018), pp. 2037–2064, https://doi.org/10.1051/m2an/2018034.

[26] D. Cioranescu and P. Donato, *An introduction to homogenization*, vol. 17 of Oxford Lecture Series in Mathematics and its Applications, The Clarendon Press and Oxford University Press, New York, NY, 1999.

[27] W. E and B. Engquist, *The heterogeneous multiscale methods*, Communications in Mathematical Sciences, 1 (2003), pp. 87–132.

[28] Y. Efendiev, J. Galvis, and T. Y. Hou, *Generalized multiscale finite element methods (GMsFEM)*, Journal of Computational Physics, 251 (2013), pp. 116–135, https://doi.org/10.1016/j.jcp.2013.04.045.

[29] Y. Efendiev and T. Y. Hou, *Multiscale finite element methods: Theory and applications*, vol. 4 of Surveys and Tutorials in the Applied Mathematical Sciences, Springer, New York, NY, 2009. Theory and applications.

[30] Y. Efendiev, T. Y. Hou, and X.-H. Wu, *Convergence of a nonconforming multiscale finite element method*, SIAM Journal on Numerical Analysis, 37 (2000), pp. 888–910, https://doi.org/10.1137/s0036142997330329.

[31] Y. Gorb and Y. Kuznetsov, *Asymptotic expansions for high-contrast scalar and vectorial PDEs*, SIAM Journal on Applied Mathematics, 81 (2021), pp. 2246–2264, https://doi.org/10.1137/20m1357937.

[32] T. Y. Hou and X.-H. Wu, *A multiscale finite element method for elliptic problems in composite materials and porous media*, Journal of Computational Physics, 134 (1997), pp. 169–189, https://doi.org/10.1006/jcph.1997.5682.

[33] T. Y. Hou, X.-H. Wu, and Z. Cai, *Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients*, Mathematics of Computation, 68 (1999), pp. 913–943, https://doi.org/10.1090/S0025-5718-99-01077-7.

[34] T. J. R. Hughes, *Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods*, Computer Methods in Applied Mechanics and Engineering, 127 (1995), pp. 387–401, https://doi.org/10.1016/0045-7825(95)00844-9.

[35] T. J. R. Hughes and G. Sangalli, *Variational multiscale analysis: the fine-scale Green's function, projection, optimization, localization, and stabilized methods*, SIAM Journal on Numerical Analysis, 45 (2007), pp. 539–557, https://doi.org/10.1137/050645646.

[36] V. Ivrii, *100 years of Weyl's law*, Bulletin of Mathematical Sciences, 6 (2016), pp. 379–452, https://doi.org/10.1007/s13373-016-0089-y.

[37] M. Kadic, T. Bückmann, N. Stenger, M. Thiel, and M. Wegener, *On the practicability of pentamode mechanical metamaterials*, Applied Physics Letters, 100 (2012), https://doi.org/10.1063/1.4709436.

[38] R. S. Lakes, *Negative-Poisson's-ratio materials: Auxetic solids*, Annual Review of Materials Research, 47 (2017), pp. 63–81, https://doi.org/10.1146/annurev-matsci-070616-124118.

[39] C. Ma, R. Scheichl, and T. Dodwell, *Novel design and analysis of generalized finite element methods based on locally optimal spectral approximations*, SIAM Journal on Numerical Analysis, 60 (2022), pp. 244–273, https://doi.org/10.1137/21m1406179.

[40] A. Målqvist and D. Peterseim, *Localization of elliptic multiscale problems*, Mathematics of Computation, 83 (2014), pp. 2583–2603, https://doi.org/10.1090/S0025-5718-2014-02868-8.

[41] A. Målqvist and D. Peterseim, *Numerical homogenization by localized orthogonal decomposition*, vol. 5 of SIAM Spotlights, Society for Industrial & Applied Mathematics (SIAM), Philadelphia, PA, 2021.

[42] G. W. Milton and A. V. Cherkaev, *Which elasticity tensors are realizable?*, Journal of Engineering Materials and Technology, 117 (1995), pp. 483–493, https://doi.org/10.1115/1.2804743.

[43] P. Ming and S. Song, *Error estimate of multiscale finite element method for periodic media revisited*, Multiscale Modeling & Simulation. A SIAM Interdisciplinary Journal, 22 (2024), pp. 106–124, https://doi.org/10.1137/22m1511060.

[44] S. Nicaise and J. Venel, *A posteriori error estimates for a finite element approximation of transmission problems with sign changing coefficients*, Journal of Computational and Applied Mathematics, 235 (2011), pp. 4272–4282, https://doi.org/10.1016/j.cam.2011.03.028.

[45] L. A. Poveda, J. Galvis, and E. T. Chung, *A second-order exponential integration constraint energy minimizing generalized multiscale method for parabolic problems*, Journal of Computational Physics, 502 (2024), p. 112796, https://doi.org/10.1016/j.jcp.2024.112796.

[46] R. A. Shelby, D. R. Smith, and S. Schultz, *Experimental verification of a negative index of refraction*, Science, 292 (2001), pp. 77–79, https://doi.org/10.1126/science.1058847.

[47] R. Verfürth, *A posteriori error estimation eechniques for finite element methods*, Oxford University Press, Oxford, UK, Apr. 2013, https://doi.org/10.1093/acprof:oso/9780199679423.001.0001.

[48] V. G. Veselago, *The electrodynamics of substances with simultaneously negative values of $\epsilon$ and $\mu$*, Soviet Physics Uspekhi, 10 (1968), pp. 509–514, https://doi.org/10.1070/pu1968v010n04abeh003699.

[49] H. Weyl, *Ueber die asymptotische Verteilung der Eigenwerte*, Nachrichten von der Gesellschaft der Wissenschaften zu Göttingen, Mathematisch-Physikalische Klasse, 1911 (1911), pp. 110–117, http://eudml.org/doc/58792.

[50] C. Ye and E. T. Chung, *Constraint energy minimizing generalized multiscale finite element method for inhomogeneous boundary value problems with high contrast coefficients*, Multiscale Modeling & Simulation. A SIAM Interdisciplinary Journal, 21 (2023), pp. 194–217, https://doi.org/10.1137/21m1459113.

[51] C. Ye, H. Dong, and J. Cui, *Convergence rate of multiscale finite element method for various boundary problems*, Journal of Computational and Applied Mathematics, 374 (2020), p. 112754, https://doi.org/10.1016/j.cam.2020.112754.

[52] L. Zhao and E. T. Chung, *An analysis of the NLMC upscaling method for high contrast problems*, Journal of Computational and Applied Mathematics, 367 (2020), p. 112480, https://doi.org/10.1016/j.cam.2019.112480.