Research Article

Philip Freese, Dietmar Gallistl, Daniel Peterseim and Timo Sprekeler*

Computational Multiscale Methods for Nondivergence-Form Elliptic Partial Differential Equations

https://doi.org/10.1515/cmam-2023-0040
Received February 12, 2023; revised May 1, 2023; accepted June 22, 2023

Abstract: This paper proposes novel computational multiscale methods for linear second-order elliptic partial differential equations in nondivergence form with heterogeneous coefficients satisfying a Cordes condition. The construction follows the methodology of localized orthogonal decomposition (LOD) and provides operator-adapted coarse spaces by solving localized cell problems on a fine scale in the spirit of numerical homogenization. The degrees of freedom of the coarse spaces are related to nonconforming and mixed finite element methods for homogeneous problems. The rigorous error analysis of one exemplary approach shows that the favorable properties of the LOD methodology known from divergence-form PDEs, i.e., its applicability and accuracy beyond scale separation and periodicity, remain valid for problems in nondivergence form.

Keywords: Nondivergence-Form Elliptic PDE, Localized Orthogonal Decomposition, Numerical Homogenization, Finite Element Methods

MSC 2020: 35J15, 65N12, 65N30

1 Introduction

In this work, we consider linear second-order elliptic partial differential equations of the form

$$A : D^2 u + b \cdot \nabla u - cu := \sum_{i,j=1}^{n} a_{ij} \partial_{ij}^2 u + \sum_{k=1}^{n} b_k \partial_k u - cu = f \quad \text{in } \Omega, \tag{1.1}$$

posed on a bounded convex polyhedral domain $\Omega \subset \mathbb{R}^n$, $n \geq 2$, with a right-hand side $f \in L^2(\Omega)$, subject to the homogeneous Dirichlet boundary condition

$$u = 0 \quad \text{on } \partial\Omega, \tag{1.2}$$

where $A = (a_{ij})_{1\leq i,j\leq n} \in L^\infty(\Omega; \mathbb{R}^{n\times n}_{\mathrm{sym}})$, $b = (b_k)_{1\leq k\leq n} \in L^\infty(\Omega; \mathbb{R}^n)$, and $c \in L^\infty(\Omega)$ are heterogeneous coefficients such that A is uniformly elliptic, $c \geq 0$ almost everywhere in Ω, and the triple (A, b, c) satisfies a (generalized) Cordes condition. Our main objective in this paper is to propose and rigorously analyze a novel finite element scheme for the accurate numerical approximation of the solution to the multiscale problem (1.1)–(1.2), a task we will refer to as numerical homogenization, by following the methodology of localized orthogonal decomposition (LOD) [31, 32]. It is worth mentioning that we are working in a framework beyond periodicity and separation of scales.

***Corresponding author: Timo Sprekeler,** Department of Mathematics, National University of Singapore, 10 Lower Kent Ridge Road, Singapore 119076, Singapore, e-mail: timo.sprekeler@nus.edu.sg. https://orcid.org/0000-0003-2934-8126
Philip Freese, Institut für Mathematik, Technische Universität Hamburg, Am Schwarzenberg-Campus 3, 21073 Hamburg, Germany, e-mail: philip.freese@tuhh.de. https://orcid.org/0000-0002-9838-6321
Dietmar Gallistl, Institut für Mathematik, Friedrich-Schiller-Universität Jena, Ernst-Abbe-Platz 2, 07743 Jena, Germany, e-mail: dietmar.gallistl@uni-jena.de. https://orcid.org/0000-0001-8954-4175
Daniel Peterseim, Institut für Mathematik & Centre for Advanced Analytics and Predictive Sciences (CAAPS), Universität Augsburg, Universitätsstr. 12a, 86159 Augsburg, Germany, e-mail: daniel.peterseim@uni-a.de. https://orcid.org/0000-0001-7213-556X

The motivation for investigating (1.1)–(1.2) stems from engineering, physics, and mathematical areas such as stochastic analysis. Notably, such equations arise in the linearization of Hamilton–Jacobi–Bellman (HJB) equations from stochastic control theory. A distinguishing feature of nondivergence-form problems such as (1.1)–(1.2) is the absence of a natural variational formulation. However, due to the Cordes condition, there exists a unique strong solution to (1.1)–(1.2) which can be equivalently characterized as the unique solution to the following Lax–Milgram-type problem:

$$\text{seek } u \in V := H^2(\Omega) \cap H^1_0(\Omega) \text{ such that } a(u, v) = \langle F, v\rangle \text{ for all } v \in V \tag{1.3}$$

with some suitably defined $F \in V^*$ and bounded coercive bilinear form $a\colon V \times V \to \mathbb{R}$.

In the presence of coefficients that vary on a fine scale, e.g., when $A(x) = \tilde{A}(\frac{x}{\varepsilon})$ with some $(0,1)^n$-periodic $\tilde{A} \in L^\infty(\mathbb{R}^n; \mathbb{R}^{n\times n}_{\mathrm{sym}})$ and $\varepsilon > 0$ small, classical finite element methods are being outperformed by multiscale finite element methods such as developed in this paper. For periodic coefficients, periodic homogenization has been proposed for linear elliptic equations in nondivergence form; cf. [3, 5, 21, 22, 26, 28, 39, 40]. Numerical homogenization of such problems has not been studied extensively so far. A finite element numerical homogenization scheme for the periodic setting has been proposed and analyzed in [9], which is based on an approximation of the solution to the homogenized problem via a finite element approximation of an invariant measure (see also [39]). Further, there has been some previous study on finite difference approaches for such problems in the periodic setting; see [2, 15]. Concerning fully nonlinear HJB and Isaacs equations, finite element approaches for the numerical homogenization in the periodic setting have been suggested in [20, 27], and some finite difference schemes have been studied in [8, 12, 13].

The case of arbitrarily rough coefficients has not yet been addressed beyond periodicity and scale separation. For divergence-form PDEs, several numerical homogenization methods have been developed in the last decade, which are based on the construction of operator-adapted basis functions and are applicable without such structural assumptions. We highlight the LOD [25, 29, 31, 33], the Generalized Finite Element Method [4, 11, 30], Rough Polyharmonic Splines and Gamblets [34, 36], as well as the recently proposed Super-Localized Orthogonal Decomposition [7, 14, 24].

The aim of this paper is to transfer such modern numerical homogenization methods to the case of nondivergence-form problems, and to provide a proof of concept that this framework also applies to this class of equations. The only existing link between numerical homogenization and nondivergence-form problems is the metric-based upscaling proposed in [35] which exploits nondivergence-form problems for a problem-dependent change of metric as part of the numerical homogenization of divergence-form problems. Our construction of a practical finite element method for the nondivergence-form problem (1.1)–(1.2) in the presence of multiscale data follows the abstract LOD framework for numerical homogenization methods for divergence-form problems presented in [1]. It is based on problem (1.3) as starting point, a-orthogonal decompositions of the solution space V and the test space V into a fine-scale space (defined as the intersection of the kernels of suitably chosen quantities of interest $q_1, \dots, q_N \in V^*$) and some coarse scale space, and a localization argument. In our exemplary approach, the choice of quantities of interest is inspired by the degrees of freedom of the nonconforming Morley finite element.

The remainder of this work is organized as follows. In Section 2, we present the problem setting as well as the theoretical foundation including the well-posedness of (1.1)–(1.2) based on a Cordes condition. In Section 3, we introduce the numerical homogenization scheme for the approximation of the solution to (1.1)–(1.2) based on LOD theory. The proposed numerical homogenization scheme is rigorously analyzed and error bounds are proved. The numerical implementation is based on a H^2-conforming Birkhoff–Mansfield element and is introduced in Section 4.1. In Section 4, we illustrate the theoretical findings by several numerical experiments, and finally, in Section 5, we discuss an alternative discretization based on mixed finite element theory.

2 Problem Setting and Well-Posedness

2.1 Framework

For a bounded convex polyhedral domain $\Omega \subset \mathbb{R}^n$ in dimension $n \geq 2$, and a right-hand side $f \in L^2(\Omega)$, we consider the problem

$$\begin{cases} Lu := A : D^2u + b \cdot \nabla u - cu = f & \text{in } \Omega, \\ u = 0 & \text{on } \partial\Omega, \end{cases} \tag{2.1}$$

where we assume that

$$A \in L^\infty(\Omega; \mathbb{R}^{n\times n}_{\mathrm{sym}}), \quad b \in L^\infty(\Omega; \mathbb{R}^n), \quad c \in L^\infty(\Omega) \text{ with } c \geq 0 \text{ a.e. in } \Omega,$$

that A is uniformly elliptic, i.e., there exist constants $\zeta_1, \zeta_2 > 0$ such that,

$$\text{for all } \xi \in \mathbb{R}^n\backslash\{0\}, \quad \zeta_1 \leq \frac{A\xi \cdot \xi}{|\xi|^2} \leq \zeta_2 \quad \text{a.e. in } \Omega, \tag{2.2}$$

and that the triple (A, b, c) satisfies the Cordes condition, that is, we make the following assumption.

(C1) If $|b| = c = 0$ a.e. in Ω, we assume that there exists a constant $\delta \in (0, 1)$ such that

$$\frac{|A|^2}{(\mathrm{tr}(A))^2} \leq \frac{1}{n - 1 + \delta} \quad \text{a.e. in } \Omega. \tag{2.3}$$

Further, in this case we set $\gamma := \frac{\mathrm{tr}(A)}{|A|^2}$ and $\lambda := 0$.

(C2) Otherwise, we assume that there exist constants $\delta \in (0, 1)$ and $\lambda \in (0, \infty)$ such that

$$\frac{|A|^2 + \frac{1}{2\lambda}|b|^2 + \frac{1}{\lambda^2}c^2}{\left(\mathrm{tr}(A) + \frac{1}{\lambda}c\right)^2} \leq \frac{1}{n + \delta} \quad \text{a.e. in } \Omega. \tag{2.4}$$

Further, in this case, we set

$$\gamma := \frac{\mathrm{tr}(A) + \frac{1}{\lambda}c}{|A|^2 + \frac{1}{2\lambda}|b|^2 + \frac{1}{\lambda^2}c^2}.$$

Here, we have used the notation $|M| := \sqrt{M : M}$ to denote the Frobenius norm of $M \in \mathbb{R}^{n\times n}$.

Remark 2.1. When $n = 2$, uniform ellipticity (2.2) guarantees that condition (2.3) is satisfied for some $\delta \in (0, 1)$; see, e.g., [38].

Remark 2.2 (Properties of γ). Note that $\gamma \in L^\infty(\Omega)$ and that there exist constants $\gamma_0, \Gamma > 0$ depending only on $n, \zeta_1, \zeta_2, \lambda, \|b\|_{L^\infty(\Omega)}, \|c\|_{L^\infty(\Omega)}$ such that $\gamma_0 \leq \gamma \leq \Gamma$ a.e. in Ω.

2.2 Well-Posedness

We introduce the Hilbert space $(V, (\cdot, \cdot)_V)$ by setting

$$V := H^2(\Omega) \cap H^1_0(\Omega), \quad (\cdot, \cdot)_V := (\cdot, \cdot)_{H^2(\Omega)}, \tag{2.5}$$

and we write $\|\cdot\|_V := \|\cdot\|_{H^2(\Omega)}$ and $\|\cdot\|_{V,\omega} := \|\cdot\|_{H^2(\omega)}$ for any subdomain $\omega \subset \Omega$. Then, introducing the bilinear form

$$a\colon V \times V \to \mathbb{R}, \quad a(v_1, v_2) := (\gamma L v_1, \Delta v_2 - \lambda v_2)_{L^2(\Omega)},$$

the linear operator

$$\mathcal{A}\colon V \to V^*, \quad v \mapsto \mathcal{A}v := a(v, \cdot),$$

and the linear functional

$$F\colon V \to \mathbb{R}, \quad v \mapsto \langle F, v\rangle := (\gamma f, \Delta v - \lambda v)_{L^2(\Omega)},$$

it is well known that we have existence and uniqueness of a strong solution to (2.1); see [37, 38].

Theorem 2.1 (Well-Posedness). *The following assertions hold true.*
(i) *A function $u \in V$ is a strong solution to* (2.1) *if, and only if,*

$$a(u, v) = \langle F, v\rangle \quad \text{for all } v \in V. \tag{2.6}$$

(ii) *There exists a unique $u \in V$ such that $\mathcal{A}u = F$ in V^*, i.e., $a(u, v) = \langle F, v\rangle$ for all $v \in V$.*
In particular, problem (2.1) *has a unique strong solution $u \in V$.*

Note that assertion (i) of Theorem 2.1 follows immediately from the fact that, for any $g \in L^2(\Omega)$, there exists a unique $v \in V$ such that $\Delta v - \lambda v = g$, and the positivity of the renormalization function γ. Assertion (ii) of Theorem 2.1 is shown by a standard Lax–Milgram argument using the properties of a and F listed below.

Lemma 2.1 (Properties of the Maps a and F). *The following assertions hold true.*
(i) *Local boundedness of a: There exists a constant $C_a > 0$ depending only on n, ζ_1, ζ_2, λ,$\|b\|_{L^\infty(\Omega)}$,$\|c\|_{L^\infty(\Omega)}$ such that, for any subdomains $\omega_1, \omega_2 \subset \Omega$, we have that,*

$$\text{for all } v_1, v_2 \in V, \quad \operatorname{supp}(v_1) \subset \omega_1, \operatorname{supp}(v_2) \subset \omega_2 \implies |a(v_1, v_2)| \le C_a \|v_1\|_{V,\omega_1\cap\omega_2}\|v_2\|_{V,\omega_1\cap\omega_2}.$$

(ii) *Coercivity of a: There exists a constant $\alpha > 0$ depending only on* $\operatorname{diam}(\Omega)$, *$n$, δ such that*

$$a(v, v) \ge (1 - \sqrt{1-\delta})\|\Delta v - \lambda v\|^2_{L^2(\Omega)} \ge \alpha\|v\|^2_V \quad \text{for all } v \in V.$$

(iii) *$F \in V^*$: There exists a constant $\mu > 0$ depending only on n, ζ_1, ζ_2, λ, $\|b\|_{L^\infty(\Omega)}$, $\|c\|_{L^\infty(\Omega)}$ such that*

$$|\langle F, v\rangle| \le \mu\|f\|_{L^2(\Omega)}\|v\|_V \quad \text{for all } v \in V,$$

or equivalently, $\|F\|_{V^*} \le \mu\|f\|_{L^2(\Omega)}$.

The proofs of assertions (i) and (iii) of Lemma 2.1 are straightforward. A proof of assertion (ii) of Lemma 2.1 can be found in [37, 38], relying on the observation that the Cordes condition implies that, for any subdomain $\omega \subset \Omega$, we have that (see [38, Lemma 1]),

$$\text{for all } v \in H^2(\omega), \quad |\gamma Lv - (\Delta v - \lambda v)| \le \sqrt{1-\delta}\sqrt{|D^2 v|^2 + 2\lambda|\nabla v|^2 + \lambda^2 v^2} \quad \text{a.e. in } \omega,$$

and using the Miranda–Talenti-type estimates (see [38, Theorem 2]),

$$\begin{aligned} \|D^2 v\|^2_{L^2(\Omega)} + 2\lambda\|\nabla v\|^2_{L^2(\Omega)} + \lambda^2\|v\|^2_{L^2(\Omega)} \le \|\Delta v - \lambda v\|^2_{L^2(\Omega)} \qquad &\text{for all } v \in V, \\ \|v\|_V \le C_{\mathrm{MT}}\|\Delta v - \lambda v\|_{L^2(\Omega)} \qquad &\text{for all } v \in V, \end{aligned}$$

with a constant $C_{\mathrm{MT}} > 0$ depending only on $\operatorname{diam}(\Omega)$ and n.

Remark 2.3. In view of Theorem 2.1 and Lemma 2.1, the unique strong solution $u \in V$ to (2.1) satisfies the bound

$$\|u\|_V \le \alpha^{-1}\|F\|_{V^*} \le \alpha^{-1}\mu\|f\|_{L^2(\Omega)},$$

where $\alpha, \mu > 0$ are the constants from Lemma 2.1 (ii)–(iii).

It is worth emphasizing that, in the setting of periodic homogenization, i.e., $A = \tilde{A}(\frac{\cdot}{\varepsilon})$, $b = \tilde{b}(\frac{\cdot}{\varepsilon})$, $c = \tilde{c}(\frac{\cdot}{\varepsilon})$ for some small parameter $\varepsilon > 0$ and $(0,1)^n$-periodic $\tilde{A}, \tilde{b}, \tilde{c} \in L^\infty(\mathbb{R}^n)$ satisfying the Cordes condition in $\mathbb{R}^n$, we have that the $H^2(\Omega)$-norm of the solution to (2.1) is uniformly bounded in $\varepsilon \in (0,1)$, while generically the $H^{2+s}(\Omega)$-norm is unbounded as $\varepsilon \searrow 0$ for any $s > 0$. Note that this is different to the usual periodic homogenization setting for divergence-form equations where generically the $H^2(\Omega)$-norm is unbounded as $\varepsilon \searrow 0$.

3 Numerical Homogenization Scheme

For simplicity, we only work in dimension $n = 2$ and give some remarks on numerical homogenization in higher dimensions in Section 6.2.

3.1 Fine-Scale Space

We start by introducing a triangulation of the bounded convex polygon $\Omega \subset \mathbb{R}^2$. Thereafter, we define a certain closed linear subspace W of V (recall the definition of V from (2.5)) which will be referred to as the fine-scale space.

3.1.1 Triangulation

Let $\mathcal{T}_H$ be a regular quasi-uniform triangulation of Ω into closed triangles with mesh size $H > 0$ and shape-regularity parameter $\rho > 0$ given by

$$H := \max_{T\in\mathcal{T}_H} \operatorname{diam}(T), \quad \rho := H^{-1} \min_{T\in\mathcal{T}_H} \rho_T,$$

where ρ_T denotes the diameter of the largest ball which can be inscribed in the element T. We introduce the piecewise constant mesh-size function $h_{\mathcal{T}_H}\colon \bar{\Omega} \to \mathbb{R}$ given by $h_{\mathcal{T}_H}|_T := H_T := |T|^{\frac{1}{2}}$ for $T \in \mathcal{T}_H$. Let $\mathcal{F}_H$ denote the set of edges, $\mathcal{N}_H^{\mathrm{int}}$ the set of interior vertices, $\mathcal{N}_H^{\partial}$ the set of boundary vertices, and define

$$N_1 := |\mathcal{F}_H|, \quad N_2 := |\mathcal{N}_H^{\mathrm{int}}|, \quad N := N_1 + N_2.$$

We label the edges $F_1, \dots, F_{N_1}$ and the interior vertices $z_1, \dots, z_{N_2}$ so that

$$\mathcal{F}_H = \{F_1, F_2, \dots, F_{N_1}\}, \quad \mathcal{N}_H^{\mathrm{int}} = \{z_1, z_2, \dots, z_{N_2}\}.$$

For each edge $F \in \mathcal{F}_H$, we associate a fixed choice of unit normal ν_F, where we often drop the subscript and only write ν for simplicity. Finally, for a subset $S \subset \Omega$, we define $N^0(S) := S$ and $N^\ell(S) := \bigcup\{T \in \mathcal{T}_H \mid T \cap N^{\ell-1}(S) \neq \emptyset\}$ for $\ell \in \mathbb{N}$.

3.1.2 Quantities of Interest and the Space of Fine-Scale Functions

First, let us note that $V \subset C(\bar{\Omega})$ by Sobolev embeddings. For $i \in \{1, \dots, N\}$, we define the quantity of interest $q_i \in V^*$ by

$$q_i\colon V \to \mathbb{R}, \quad v \mapsto \langle q_i, v\rangle := \begin{cases} \fint_{F_i} \nabla v \cdot \nu & \text{if } 1 \le i \le N_1, \\ v(z_{i-N_1}) & \text{if } N_1 + 1 \le i \le N. \end{cases}$$

The quantities of interest $\{q_1, \dots, q_N\} \subset V^*$ are linearly independent as can be seen from the fact that there exist functions $u_1, \dots, u_N \in V$ such that $\langle q_i, u_j\rangle = \delta_{ij}$ for any $i, j \in \{1, \dots, N\}$; see Section 3.3.1. We define the fine-scale space

$$W := \bigcap_{i\in\{1,\dots,N\}} \ker(q_i) = \{v \in V \mid \langle q_i, v\rangle = 0 \text{ for all } i \in \{1, \dots, N\}\}, \tag{3.1}$$

which is a closed linear subspace of V.

3.1.3 Connection to the Morley Finite Element Space

We consider the Morley finite element space

$$V_H^{\mathrm{Mor}} := \{v \in \mathcal{P}_2(\mathcal{T}_H) \mid v \text{ is continuous at } \mathcal{N}_H^{\mathrm{int}} \text{ and vanishes at } \mathcal{N}_H^{\partial};\\ D_{\mathrm{NC}} v \text{ is continuous at the interior edges' midpoints}\},$$

whose local degrees of freedom are the evaluation of the function at each vertex and the evaluation of the normal derivative at the edges' midpoints. Here, the piecewise action of the differential operator D is indicated by

the subscript NC, i.e., we define $(D_{\mathrm{NC}}v)|_T := D(v|_T)$ for any $T \in \mathcal{T}_H$. Then, letting $\{\phi_1, \dots, \phi_N\} \subset V_H^{\mathrm{Mor}}$ denote Morley basis functions satisfying $\langle q_i, \phi_j \rangle = \delta_{ij}$ for all $i, j \in \{1, \dots, N\}$ (note $\langle q_i, \phi_j \rangle$ is well-defined although $\phi_j \notin V$), we have that the Morley interpolation operator is given by

$$\Pi^{\mathrm{Mor}} \colon V \to V_H^{\mathrm{Mor}}, \quad v \mapsto \Pi^{\mathrm{Mor}} v := \sum_{i=1}^{N} \langle q_i, v \rangle \phi_i, \tag{3.2}$$

and we observe that

$$W = \ker(\Pi^{\mathrm{Mor}}) = \{v \in V \mid \Pi^{\mathrm{Mor}} v = 0\}.$$

In particular, using Morley interpolation bounds (see, e.g., [42]), we have the local estimate

$$\|w\|_{L^2(T)} + H_T \|\nabla w\|_{L^2(T)} \lesssim H_T^2 \|D^2 w\|_{L^2(T)} \quad \text{for all } T \in \mathcal{T}_H \text{ and all } w \in W, \tag{3.3}$$

and the global bound

$$\|w\|_{L^2(\Omega)} + H \|\nabla w\|_{L^2(\Omega)} \lesssim H^2 \|D^2 w\|_{L^2(\Omega)} \quad \text{for all } w \in W. \tag{3.4}$$

3.1.4 Projectors onto the Fine-Scale Space

We introduce the maps

$$\mathcal{C} \colon V \to W, \quad v \mapsto \mathcal{C}v, \qquad \mathcal{C}^* \colon V \to W, \quad v \mapsto \mathcal{C}^* v,$$

where, for $v \in V$, we define $\mathcal{C}v \in W$ to be the unique function in W that satisfies

$$a(\mathcal{C}v, w) = a(v, w) \quad \text{for all } w \in W,$$

and we define $\mathcal{C}^* v \in W$ to be the unique function in W that satisfies

$$a(w, \mathcal{C}^* v) = a(w, v) \quad \text{for all } w \in W.$$

Remark 3.1. In view of Lemma 2.1, we have by the Lax–Milgram theorem that the maps $\mathcal{C}$, $\mathcal{C}^*$ are well-defined, and we have the bounds

$$\|\mathcal{C}v\|_V \le \alpha^{-1} C_a \|v\|_V, \quad \|\mathcal{C}^* v\|_V \le \alpha^{-1} C_a \|v\|_V \quad \text{for all } v \in V.$$

Further, the maps $\mathcal{C}$, $\mathcal{C}^*$ are surjective and continuous projectors onto W, and we have that

$$W = \ker(1 - \mathcal{C}) = \ker(1 - \mathcal{C}^*).$$

3.2 Ideal Numerical Homogenization Scheme

3.2.1 *a*-Orthogonal Decompositions of *V*

We define the trial space $\tilde{U}_H \subset V$ and the test space $\tilde{V}_H \subset V$ by

$$\tilde{U}_H := (1 - \mathcal{C})V, \quad \tilde{V}_H := (1 - \mathcal{C}^*)V. \tag{3.5}$$

In view of Remark 3.1, we then have the following decompositions of the space V:

$$V = (1 - \mathcal{C})V \oplus \mathcal{C}V = \tilde{U}_H \oplus W, \quad V = (1 - \mathcal{C}^*)V \oplus \mathcal{C}^* V = \tilde{V}_H \oplus W. \tag{3.6}$$

We state a few observations below.

Lemma 3.1 (Properties of $\tilde{U}_H$ and $\tilde{V}_H$). *The following assertions hold true.*
(i) *We have that* $\dim(\tilde{U}_H) = \dim(\tilde{V}_H) = N$.
(ii) *Decompositions* (3.6) *are a-orthogonal in the sense that* $a(\tilde{U}_H, W) = 0$ *and* $a(W, \tilde{V}_H) = 0$.

(iii) *We can equivalently characterize the spaces $\tilde{U}_H$ and $\tilde{V}_H$ via*

$$\tilde{U}_H = \{v \in V \mid a(v, w) = 0 \text{ for all } w \in W\},$$
$$\tilde{V}_H = \{v \in V \mid a(w, v) = 0 \text{ for all } w \in W\}.$$

(iv) *We have that $\tilde{U}_H = \operatorname{span}(\mathcal{A}^{-1}q_1, \dots, \mathcal{A}^{-1}q_N)$.*

Proof. (i) By the Riesz representation theorem, there exist $\hat{q}_1, \dots, \hat{q}_N \in V$ such that $q_i = (\cdot, \hat{q}_i)_V$ in V^* for $i \in \{1, \dots, N\}$. Set $S := \operatorname{span}(\hat{q}_1, \dots, \hat{q}_N)$ and note that $\dim(S) = N$ as the quantities of interest $q_1, \dots, q_N$ are linearly independent. Then, in view of (3.1), we have that $W = S^\perp$ and there holds $V = W^\perp \oplus W = S \oplus W$. The claim follows.

(ii) This follows immediately from the definition of the spaces $\tilde{U}_H$ and $\tilde{V}_H$ from (3.5), and the definitions of the projectors $\mathcal{C}$ and $\mathcal{C}^*$ from Section 3.1.4.

(iii) By the properties of the projectors $\mathcal{C}$ and $\mathcal{C}^*$ from Section 3.1.4, we have that

$$\{v \in V \mid a(v, w) = 0 \text{ for all } w \in W\} = \{v \in V \mid \mathcal{C}v = 0\} = (1 - \mathcal{C})V = \tilde{U}_H,$$
$$\{v \in V \mid a(w, v) = 0 \text{ for all } w \in W\} = \{v \in V \mid \mathcal{C}^*v = 0\} = (1 - \mathcal{C}^*)V = \tilde{V}_H.$$

(iv) First, note that $\dim(\tilde{U}_H) = N = \dim(\operatorname{span}(\mathcal{A}^{-1}q_1, \dots, \mathcal{A}^{-1}q_N))$. Next, we observe that, for $i \in \{1, \dots, N\}$, we have that $a(\mathcal{A}^{-1}q_i, w) = \langle q_i, w\rangle = 0$ for all $w \in W$, i.e., $\mathcal{A}^{-1}q_1, \dots, \mathcal{A}^{-1}q_N \in \tilde{U}_H$ holds. It follows that

$$\tilde{U}_H = \operatorname{span}(\mathcal{A}^{-1}q_1, \dots, \mathcal{A}^{-1}q_N). \qquad \square$$

3.2.2 Ideal Numerical Homogenization

The ideal discrete problem is the following:

$$\text{find } \tilde{u}_H \in \tilde{U}_H \text{ such that } a(\tilde{u}_H, \tilde{v}_H) = \langle F, \tilde{v}_H\rangle \text{ for all } \tilde{v}_H \in \tilde{V}_H. \tag{3.7}$$

Theorem 3.1 (Analysis of the Ideal Discrete Problem). *There exists a unique solution $\tilde{u}_H \in \tilde{U}_H$ to the ideal discrete problem* (3.7)*. Further, denoting the unique strong solution to* (2.1) *by $u \in V$, the following assertions hold true.*

(i) *We have the bound*

$$\|\tilde{u}_H\|_V \le \alpha^{-2}C_a\|F\|_{(\tilde{V}_H)^*} \le \alpha^{-2}\mu C_a\|f\|_{L^2(\Omega)}.$$

(ii) *We have that $u - \tilde{u}_H = \mathcal{C}u \in W$ and hence*

$$\langle q_i, \tilde{u}_H\rangle = \langle q_i, u\rangle \quad \text{for all } i \in \{1, \dots, N\},$$

i.e., the quantities of interest are conserved.

(iii) *We have the error bound*

$$\|u - \tilde{u}_H\|_{L^2(\Omega)} + H\|\nabla(u - \tilde{u}_H)\|_{L^2(\Omega)} \le H^2\|f\|_{L^2(\Omega)}$$

for the approximation of the true solution $u \in V$ by $\tilde{u}_H \in \tilde{U}_H$.

Proof. First, recall the properties of a and F from Lemma 2.1. Next, we note that, for any $\tilde{u}_H \in \tilde{U}_H$, we have that

$$\sup_{\tilde{v}_H \in \tilde{V}_H \setminus \{0\}} \frac{|a(\tilde{u}_H, \tilde{v}_H)|}{\|\tilde{v}_H\|_V} \ge \frac{|a(\tilde{u}_H, (1 - \mathcal{C}^*)\tilde{u}_H)|}{\|(1 - \mathcal{C}^*)\tilde{u}_H\|_V} \ge \frac{|a(\tilde{u}_H, \tilde{u}_H)|}{\alpha^{-1}C_a\|\tilde{u}_H\|_V} \ge \alpha^2 C_a^{-1}\|\tilde{u}_H\|_V,$$

where we have used Lemma 3.1 (ii), the fact that $\|1 - \mathcal{C}^*\| = \|\mathcal{C}^*\|$ (see [41]), and that $\|\mathcal{C}^*\| \le \alpha^{-1}C_a$ by Remark 3.1. Similarly, for any $\tilde{v}_H \in \tilde{V}_H$, we have that

$$\sup_{\tilde{u}_H \in \tilde{U}_H \setminus \{0\}} \frac{|a(\tilde{u}_H, \tilde{v}_H)|}{\|\tilde{u}_H\|_V} \ge \frac{|a((1 - \mathcal{C})\tilde{v}_H, \tilde{v}_H)|}{\|(1 - \mathcal{C})\tilde{v}_H\|_V} \ge \frac{|a(\tilde{v}_H, \tilde{v}_H)|}{\alpha^{-1}C_a\|\tilde{v}_H\|_V} \ge \alpha^2 C_a^{-1}\|\tilde{v}_H\|_V.$$

By the Babuška–Lax–Milgram theorem, there exists a unique solution $\tilde{u}_H \in \tilde{U}_H$ to the ideal discrete problem (3.7), and we obtain (i). It only remains to show (ii) and (iii).

(ii) We show that $u - \tilde{u}_H = \mathcal{C}u \in W$. Observing that we have the Galerkin orthogonality (recall $\tilde{V}_H \subset V$)

$$a(u - \tilde{u}_H, \tilde{v}_H) = 0 \quad \text{for all } \tilde{v}_H \in \tilde{V}_H,$$

we find that $u - \tilde{u}_H \in W$ by Lemma 3.1 (ii) and (3.6). Finally, as $u - \tilde{u}_H \in W$, we have that

$$u - \tilde{u}_H = \mathcal{C}(u - \tilde{u}_H) = \mathcal{C}u - \mathcal{C}\tilde{u}_H = \mathcal{C}u.$$

Here, we have used that $\mathcal{C}\tilde{u}_H = 0$ by the definition of $\tilde{U}_H$ from (3.5) and the properties of $\mathcal{C}$ from Remark 3.1.

(iii) First, we note that, by Remarks 3.1 and 2.3, we have the bound $\|\mathcal{C}u\|_V \le \alpha^{-1}C_a\|u\|_V \le \alpha^{-2}\mu C_a\|f\|_{L^2(\Omega)}$. In view of the fact that $u - \tilde{u}_H = \mathcal{C}u \in W$ (see (ii)) and using the bound (3.4), we deduce that

$$\|u - \tilde{u}_H\|_{L^2(\Omega)} + H\|\nabla(u - \tilde{u}_H)\|_{L^2(\Omega)} \lesssim H^2\|\mathcal{C}u\|_V \lesssim H^2\|f\|_{L^2(\Omega)},$$

which concludes the proof. □

3.3 Construction of a Coarse-Scale Space

3.3.1 Construction of a Local Basis

We are going to construct functions $u_1, \dots, u_N \in V$ with local support that satisfy

$$\langle q_i, u_j \rangle = \delta_{ij} \quad \text{for any } i, j \in \{1, \dots, N\}.$$

To this end, we introduce the Hsieh–Clough–Tocher (HCT) finite element space

$$V_H^{\mathrm{HCT}} := \{v \in V \mid v|_T \in \mathcal{P}_3(\mathcal{K}_H(T)) \text{ for all } T \in \mathcal{T}_H\},$$

where $\mathcal{K}_H(T)$ denotes the triangulation of the triangle T into three sub-triangles with shared vertex $\mathrm{mid}(T)$, and we make use of the HCT enrichment operator $E_H\colon V_H^{\mathrm{Mor}} \to V_H^{\mathrm{HCT}}$ defined in [16, Proposition 2.5]. We then define the operator

$$\tilde{E}_H\colon V_H^{\mathrm{Mor}} \to V, \quad v \mapsto \tilde{E}_H v := E_H v + \sum_{i=1}^{N_1} \Big[\fint_{F_i} \nabla(v - E_H v) \cdot \nu \Big] \zeta_{F_i},$$

where $\zeta_{F_i} \in V$ is the function from [16, proof of Proposition 2.6] which satisfies

$$\fint_{F_i} \nabla\zeta_{F_i} \cdot \nu = 1, \quad \zeta_{F_i}(z) = 0 \quad \text{for all } z \in \mathcal{N}_H^{\mathrm{int}},$$

and $\mathrm{supp}(\zeta_{F_i}) \subset \bar{\omega}_{F_i}$, where $\bar{\omega}_{F_i}$ denotes the closure of the union of the two elements that share the edge F_i. For any $v_H^{\mathrm{Mor}} \in V_H^{\mathrm{Mor}}$, we have that

$$\begin{aligned} \fint_F \nabla(\tilde{E}_H v_H^{\mathrm{Mor}}) \cdot \nu &= \fint_F \nabla v_H^{\mathrm{Mor}} \cdot \nu \quad \text{for all } F \in \mathcal{F}_H, \\ (\tilde{E}_H v_H^{\mathrm{Mor}})(z) &= v_H^{\mathrm{Mor}}(z) \qquad \text{for all } z \in \mathcal{N}_H^{\mathrm{int}}, \end{aligned} \tag{3.8}$$

i.e., $\tilde{E}_H$ preserves the quantities of interest $q_1, \dots, q_N$. Further, we have the bound

$$\begin{aligned} &\|h_{\mathcal{T}_H}^{-2}(v_H^{\mathrm{Mor}} - E_H v_H^{\mathrm{Mor}})\|_{L^2(\Omega)} + \|h_{\mathcal{T}_H}^{-1}\nabla_{\mathrm{NC}}(v_H^{\mathrm{Mor}} - E_H v_H^{\mathrm{Mor}})\|_{L^2(\Omega)} + \|D_{\mathrm{NC}}^2(v_H^{\mathrm{Mor}} - E_H v_H^{\mathrm{Mor}})\|_{L^2(\Omega)} \\ &\qquad \lesssim \min_{v \in V} \|D_{\mathrm{NC}}^2(v_H^{\mathrm{Mor}} - v)\|_{L^2(\Omega)} \quad \text{for all } v_H^{\mathrm{Mor}} \in V_H^{\mathrm{Mor}}, \end{aligned}$$

where the subscript NC indicates the piecewise action of a differential operator with respect to the triangulation $\mathcal{T}_H$, and we have that

$$\begin{aligned} &\|h_{\mathcal{T}_H}^{-2}(v_H^{\mathrm{Mor}} - \tilde{E}_H v_H^{\mathrm{Mor}})\|_{L^2(\Omega)} + \|h_{\mathcal{T}_H}^{-1}\nabla_{\mathrm{NC}}(v_H^{\mathrm{Mor}} - \tilde{E}_H v_H^{\mathrm{Mor}})\|_{L^2(\Omega)} + \|D_{\mathrm{NC}}^2(v_H^{\mathrm{Mor}} - \tilde{E}_H v_H^{\mathrm{Mor}})\|_{L^2(\Omega)} \\ &\qquad \lesssim \min_{v \in V} \|D_{\mathrm{NC}}^2(v_H^{\mathrm{Mor}} - v)\|_{L^2(\Omega)} \quad \text{for all } v_H^{\mathrm{Mor}} \in V_H^{\mathrm{Mor}}. \end{aligned} \tag{3.9}$$

The proofs of [16, Propositions 2.5–2.6] furthermore show the quasi-local bound

$$\begin{aligned}\|h_{\mathcal{T}_H}^{-2}(v_H^{\mathrm{Mor}} - \tilde{E}_H v_H^{\mathrm{Mor}})\|_{L^2(T)} + \|h_{\mathcal{T}_H}^{-1}\nabla_{\mathrm{NC}}(v_H^{\mathrm{Mor}} - \tilde{E}_H v_H^{\mathrm{Mor}})\|_{L^2(T)} + \|D_{\mathrm{NC}}^2(v_H^{\mathrm{Mor}} - \tilde{E}_H v_H^{\mathrm{Mor}})\|_{L^2(T)} \\ \lesssim \min_{v\in V}\|D_{\mathrm{NC}}^2(v_H^{\mathrm{Mor}} - v)\|_{L^2(N^1(T))} \quad \text{for all } v_H^{\mathrm{Mor}} \in V_H^{\mathrm{Mor}},\end{aligned} \tag{3.10}$$

for any $T \in \mathcal{T}_H$. We define the functions

$$u_i := \tilde{E}_H \phi_i \in V, \quad i \in \{1, \dots, N\}, \tag{3.11}$$

where $\phi_1, \dots, \phi_N \in V_H^{\mathrm{Mor}}$ are the Morley basis functions from Section 3.1.3, and we set

$$U_H := \operatorname{span}(u_1, \dots, u_N) \subset V.$$

By (3.8) and the definition of the Morley basis functions, there holds

$$\langle q_i, u_j\rangle = \delta_{ij} \quad \text{for all } i, j \in \{1, \dots, N\}, \tag{3.12}$$

and we have that $\Omega_i := \operatorname{supp}(u_i) \subset N^1(\operatorname{supp}(\phi_i))$ for any $i \in \{1, \dots, N\}$. Further, we have the following result.

Lemma 3.2 (Stability of Basis Representation). *For any $u_H = \sum_{i=1}^N c_i u_i \in U_H$ with $c_i = \langle q_i, u_H\rangle$ for $i \in \{1, \dots, N\}$, we have that*

$$\sum_{i=1}^N c_i^2 \|u_i\|_V^2 \lesssim H^{-4}\|u_H\|_V^2.$$

Proof. Let $u_H = \sum_{i=1}^N c_i u_i \in U_H$ with $c_i = \langle q_i, u_H\rangle$ for $i \in \{1, \dots, N\}$. Then, by the definition (3.11) of u_i, the bound (3.9) for $\tilde{E}_H$, and inverse estimates for Morley functions, we have that

$$\sum_{i=1}^N c_i^2 \|u_i\|_V^2 = \sum_{i=1}^N c_i^2 \|\tilde{E}_H\phi_i\|_{H^2(\Omega)}^2 \lesssim \sum_{i=1}^N c_i^2 \|\phi_i\|_{H^2(\Omega;\mathcal{T}_H)}^2 \lesssim H^{-4} \sum_{T\in\mathcal{T}_H} \sum_{\substack{i\in\{1,\dots,N\}\\ \operatorname{supp}(\phi_i)\cap T\neq\emptyset}} \langle q_i, u_H\rangle^2 \|\phi_i\|_{L^2(T)}^2,$$

where we have used the notation $H^2(\Omega;\mathcal{T}_H) := \{\phi \in L^2(\Omega) \mid \phi|_T \in H^2(T) \text{ for all } T \in \mathcal{T}_H\}$ to denote the broken H^2-space, and $\|\cdot\|_{H^2(\Omega;\mathcal{T}_H)} := \sqrt{\sum_{T\in\mathcal{T}_H}\|\cdot\|_{H^2(T)}^2}$ to denote the broken H^2-norm. We deduce that

$$\sum_{i=1}^N c_i^2 \|u_i\|_V^2 \lesssim H^{-4} \sum_{T\in\mathcal{T}_H} \Big\| \sum_{\substack{i\in\{1,\dots,N\}\\ \operatorname{supp}(\phi_i)\cap T\neq\emptyset}} \langle q_i, u_H\rangle \phi_i \Big\|_{L^2(T)}^2 \lesssim H^{-4} \sum_{T\in\mathcal{T}_H} \|\Pi^{\mathrm{Mor}} u_H\|_{L^2(T)}^2 \lesssim H^{-4}\|u_H\|_V^2.$$

In the final step, we have used that $\Pi^{\mathrm{Mor}} u_H = (\Pi^{\mathrm{Mor}} - 1)u_H + u_H$ and a Morley interpolation estimate; see [42]. □

3.3.2 Projector onto U_H

We introduce the projector

$$P_H \colon V \to U_H, \quad v \mapsto P_H v := \sum_{i=1}^N \langle q_i, v\rangle u_i.$$

Remark 3.2. We can equivalently characterize P_H as follows.

(i) By (3.11) and (3.2), we have that

$$P_H v = \sum_{i=1}^N \langle q_i, v\rangle \tilde{E}_H\phi_i = \tilde{E}_H\Big(\sum_{i=1}^N \langle q_i, v\rangle \phi_i\Big) = \tilde{E}_H(\Pi^{\mathrm{Mor}} v) \quad \text{for all } v \in V,$$

that is, $P_H = \tilde{E}_H \circ \Pi^{\mathrm{Mor}}$.

(ii) In view of (i) and introducing $I_H := E_H \circ \Pi^{\mathrm{Mor}}$, we have that

$$P_H v = I_H v + \sum_{i=1}^{N_1} \langle q_i, v - I_H v\rangle \zeta_{F_i} \quad \text{for all } v \in V.$$

We list some stability properties of the projector P_H below.

Lemma 3.3 (Stability of P_H). *There exist constants $C_{P_H}, C_{P_H}^{\mathrm{loc}} > 0$ independent of H such that we have the stability bound*

$$\|P_H v\|_V \leq C_{P_H}\|v\|_V \quad \text{for all } v \in V$$

and the local stability bound

$$\|P_H v\|_{V,S} \leq C_{P_H}^{\mathrm{loc}}\|v\|_{V,N^1(S)} \quad \text{for all } v \in V,$$

for any element patch S.

Proof. The global stability bound follows from the fact that $P_H = \tilde{E}_H \circ \Pi^{\mathrm{Mor}}$ and estimate (3.9). The local stability bound follows from the decomposition $P_H = (\tilde{E}_H - 1) \circ \Pi^{\mathrm{Mor}} + \Pi^{\mathrm{Mor}}$, the triangle inequality, and the local bound (3.10). □

Remark 3.3 (Properties of P_H). We make the following observations.
(i) For any $u_H \in U_H$, we have that $P_H(1 - \mathcal{C})u_H = P_H u_H = u_H$ and $P_H(1 - \mathcal{C}^*)u_H = u_H$.
(ii) There holds $\ker(P_H) = W$.
(iii) For any $v \in V$, we have that $(1 - P_H)v \in W$ and hence $(1 - \mathcal{C})v = (1 - \mathcal{C})P_H v$ and $(1 - \mathcal{C}^*)v = (1 - \mathcal{C}^*)P_H v$.

3.3.3 Connection of $\tilde{U}_H$ and $\tilde{V}_H$ to the Space U_H

First, let us note that, in view of (3.12), we have that $\dim(U_H) = N$ and $U_H \cap W = \{0\}$. Recalling that

$$W = \ker(1 - \mathcal{C}) = \ker(1 - \mathcal{C}^*),$$

we see that $\dim((1 - \mathcal{C})U_H) = \dim((1 - \mathcal{C}^*)U_H) = N$, and we deduce that

$$(1 - \mathcal{C})U_H = \tilde{U}_H, \quad (1 - \mathcal{C}^*)U_H = \tilde{V}_H.$$

Note that $\{(1 - \mathcal{C})u_1, \dots, (1 - \mathcal{C})u_N\}$ is a basis of $\tilde{U}_H$ and that $\{(1 - \mathcal{C}^*)u_1, \dots, (1 - \mathcal{C}^*)u_N\}$ is a basis of $\tilde{V}_H$. It can be checked that the function $(1 - \mathcal{C})u_i$ is independent of the particular choice of u_i as indicated below.

Remark 3.4. Using the arguments presented in [1, Section 3.4], it can be seen that, for any $i \in \{1, \dots, N\}$ there exists a unique pair $(\tilde{u}_i, \tilde{\lambda}) \in V \times \mathbb{R}^N$ such that

$$\begin{cases} a(\tilde{u}_i, v) + \sum_{j=1}^{N} \tilde{\lambda}_j \langle q_j, v\rangle = 0 & \text{for all } v \in V, \\ \langle q_j, \tilde{u}_i\rangle = \delta_{ij} & \text{for all } j \in \{1, \dots, N\}. \end{cases}$$

Further, there holds $\tilde{u}_i = (1 - \mathcal{C})u_i$ and $\tilde{\lambda}_j = -a((1 - \mathcal{C})u_i, (1 - \mathcal{C}^*)u_j)$ for all $i, j \in \{1, \dots, N\}$.

3.4 Construction of Localized Correctors

3.4.1 Exponential Decay of Correctors

The following lemma sets the foundation for the construction of a practical/localized numerical homogenization scheme.

Lemma 3.4 (Exponential Decay of Correctors). *There exists a constant $\beta > 0$ such that, for any $v \in V$ and any $\ell \in \mathbb{N}_0$, we have*

$$\|\mathcal{C}v\|_{V,\Omega\setminus N^\ell(\Omega_v)} \lesssim \mathrm{e}^{-\frac{1}{25}|\log(\beta)|\ell}\|\mathcal{C}v\|_V,$$

where $\Omega_v := \bigcup\{T \in \mathcal{T}_H \mid T \cap \mathrm{supp}(v) \neq \emptyset\}$.

Proof. First, let us note that $\operatorname{supp}(P_H v) \subset N^1(S)$ for any $v \in V$ with $\operatorname{supp}(v) \subset S$, where S is an element patch in $\mathcal{T}_H$. Let $v \in V$ and let $\ell \in \mathbb{N}$ with $\ell \geq 5$.

Let $\eta \in W^{2,\infty}(\Omega)$ be a cut-off function with

$$0 \leq \eta \leq 1, \quad \eta|_{N^{\ell-1}(\Omega_v)} \equiv 0, \quad \eta|_{\Omega \setminus N^{\ell}(\Omega_v)} \equiv 1, \quad \|\nabla \eta\|_{L^\infty(\Omega)} + H\|D^2 \eta\|_{L^\infty(\Omega)} \lesssim H^{-1},$$

and let $\tilde{\eta} := 1 - \eta \in W^{2,\infty}(\Omega)$. We introduce

$$w := (1 - P_H)[\eta \mathcal{C} v], \quad \tilde{w} := (1 - P_H)[\tilde{\eta} \mathcal{C} v] \tag{3.13}$$

and note that $w, \tilde{w} \in W$, there holds $\operatorname{supp}(w) \subset \Omega \setminus N^{\ell-2}(\Omega_v)$, $\operatorname{supp}(\tilde{w}) \subset N^{\ell+1}(\Omega_v)$, and we have that

$$\|\mathcal{C} v\|^2_{V, \Omega \setminus N^{\ell+1}(\Omega_v)} = \|(1 - P_H)[\mathcal{C} v]\|^2_{V, \Omega \setminus N^{\ell+1}(\Omega_v)} = \|w + \tilde{w}\|^2_{V, \Omega \setminus N^{\ell+1}(\Omega_v)} \leq \|w\|^2_V \leq \alpha^{-1} a(w, w), \tag{3.14}$$

where we have successively used that $P_H[\mathcal{C} v] = 0$ as $\ker(P_H) = W$, the definition (3.13) of the functions w and $\tilde{w}$, the fact that $\operatorname{supp}(\tilde{w}) \subset N^{\ell+1}(\Omega_v)$, and coercivity of a from Lemma 2.1 (ii). Next, we observe that

$$a(w, w) + a(\tilde{w}, w) = a((1 - P_H)[\mathcal{C} v], w) = a(\mathcal{C} v, w) = a(v, w) = 0, \tag{3.15}$$

where we have used bilinearity of a, the fact that $P_H[\mathcal{C} v] = 0$, the definition of $\mathcal{C}$, and the observation that, in view of Lemma 2.1 (i), there holds $a(v, w) = 0$ as $\operatorname{supp}(v) \subset \Omega_v$ and $\operatorname{supp}(w) \subset \Omega \setminus N^{\ell-2}(\Omega_v)$. Combining (3.14) and (3.15), and using Lemma 2.1 (i), we find that

$$\|\mathcal{C} v\|^2_{V, \Omega \setminus N^{\ell+1}(\Omega_v)} \leq \alpha^{-1} a(w, w) = -\alpha^{-1} a(\tilde{w}, w) \leq \alpha^{-1} C_a \|\tilde{w}\|_{V,R} \|w\|_{V,R}, \tag{3.16}$$

where $R := \operatorname{supp}(\tilde{w}) \cap \operatorname{supp}(w) = N^{\ell+1}(\Omega_v) \setminus N^{\ell-2}(\Omega_v)$. We proceed by noting that, by Lemma 3.3, we have that

$$\|w\|_{V,R} \lesssim \|\eta \mathcal{C} v\|_{V, N^1(R)} \lesssim \|\mathcal{C} v\|_{V, N^1(R)}. \tag{3.17}$$

Here, the final bound in (3.17) follows from the fact that, for any $T \in N^1(R)$, there holds

$$\begin{aligned} \|\eta \mathcal{C} v\|_{V,T} &\lesssim \|\eta\|_{L^\infty(\Omega)} \|\mathcal{C} v\|_{L^2(T)} + \|\nabla \eta\|_{L^\infty(\Omega)} \|\mathcal{C} v\|_{L^2(T)} + \|\eta\|_{L^\infty(\Omega)} \|\nabla[\mathcal{C} v]\|_{L^2(T)} \\ &\quad + \|D^2 \eta\|_{L^\infty(\Omega)} \|\mathcal{C} v\|_{L^2(T)} + \|\nabla \eta\|_{L^\infty(\Omega)} \|\nabla[\mathcal{C} v]\|_{L^2(T)} + \|\eta\|_{L^\infty(\Omega)} \|D^2[\mathcal{C} v]\|_{L^2(T)} \\ &\lesssim \|\mathcal{C} v\|_{V,T}, \end{aligned}$$

where we have used the properties of the cut-off function η and the bound (3.3) for the function $\mathcal{C} v \in W$. Similarly to (3.17), we find that

$$\|\tilde{w}\|_{V,R} \lesssim \|\tilde{\eta} \mathcal{C} v\|_{V, N^1(R)} \lesssim \|\mathcal{C} v\|_{V, N^1(R)}. \tag{3.18}$$

Combining (3.17)–(3.18) with (3.16), we obtain that there exists a constant $C > 0$ such that

$$\|\mathcal{C} v\|^2_{V, \Omega \setminus N^{\ell+2}(\Omega_v)} \leq \|\mathcal{C} v\|^2_{V, \Omega \setminus N^{\ell+1}(\Omega_v)} \leq C^2 \|\mathcal{C} v\|^2_{V, N^1(R)} = C^2 (\|\mathcal{C} v\|^2_{V, \Omega \setminus N^{\ell-3}(\Omega_v)} - \|\mathcal{C} v\|^2_{V, \Omega \setminus N^{\ell+2}(\Omega_v)}),$$

and hence, setting $\beta := \frac{C}{\sqrt{1 + C^2}} \in (0, 1)$, we have that

$$\|\mathcal{C} v\|_{V, \Omega \setminus N^{\ell+2}(\Omega_v)} \leq \beta \|\mathcal{C} v\|_{V, \Omega \setminus N^{\ell-3}(\Omega_v)}.$$

Setting $k := \lfloor \frac{\ell}{5} \rfloor$ and recalling that $\ell \geq 5$, a repeated application of this bound yields

$$\|\mathcal{C} v\|_{V, \Omega \setminus N^{\ell}(\Omega_v)} \leq \beta^k \|\mathcal{C} v\|_V = \mathrm{e}^{-\frac{k}{\ell} |\log(\beta)| \ell} \|\mathcal{C} v\|_V \leq \mathrm{e}^{-\frac{\ell - 4}{5\ell} |\log(\beta)| \ell} \|\mathcal{C} v\|_V \leq \mathrm{e}^{-\frac{1}{25} |\log(\beta)| \ell} \|\mathcal{C} v\|_V,$$

proving the claim for the case $\ell \geq 5$. Finally, note that, for $\ell \in \mathbb{N}_0$ with $\ell < 5$, we have

$$\|\mathcal{C} v\|_{V, \Omega \setminus N^{\ell}(\Omega_v)} \leq \|\mathcal{C} v\|_V \leq \mathrm{e}^{\frac{1}{5} |\log(\beta)|} \mathrm{e}^{-\frac{1}{25} |\log(\beta)| \ell} \|\mathcal{C} v\|_V,$$

which concludes the proof. □

Using similar arguments, one obtains an analogous exponential decay result for the corrector $\mathcal{C}^*$.

3.4.2 Localized Correctors

Motivated by the fact that, for any $u_H \in U_H$, we have that

$$\mathcal{C}u_H = \sum_{i=1}^{N} \langle q_i, u_H \rangle \varphi_i, \quad \mathcal{C}^* u_H = \sum_{i=1}^{N} \langle q_i, u_H \rangle \psi_i, \quad \text{where } \varphi_i := \mathcal{C}u_i,\ \psi_i := \mathcal{C}^* u_i,$$

we define for $\ell \in \mathbb{N}_0$ the localized correctors

$$\mathcal{C}_\ell \colon U_H \to W, \quad u_H \mapsto \mathcal{C}_\ell u_H := \sum_{i=1}^{N} \langle q_i, u_H \rangle \varphi_i^\ell,$$
$$\mathcal{C}_\ell^* \colon U_H \to W, \quad u_H \mapsto \mathcal{C}_\ell^* u_H := \sum_{i=1}^{N} \langle q_i, u_H \rangle \psi_i^\ell.$$

Here, for $i \in \{1, \dots, N\}$, the functions $\varphi_i^\ell, \psi_i^\ell$ are defined as the unique $\varphi_i^\ell, \psi_i^\ell \in W(N^\ell(\Omega_i))$ that satisfy

$$a(\varphi_i^\ell, w) = a(u_i, w) \quad \text{for all } w \in W(N^\ell(\Omega_i)),$$
$$a(w, \psi_i^\ell) = a(w, u_i) \quad \text{for all } w \in W(N^\ell(\Omega_i)),$$

where we write $W(N^\ell(\Omega_i)) := \{w \in W \mid \operatorname{supp}(w) \subset N^\ell(\Omega_i)\}$. Note that $\mathcal{C}_\ell$ and $\mathcal{C}_\ell^*$ are well-defined by the properties of a from Lemma 2.1.

3.4.3 Localization Error

We can quantify the error committed in approximating the true correctors $\mathcal{C}$, $\mathcal{C}^*$ by their localized counterparts $\mathcal{C}_\ell, \mathcal{C}_\ell^*$.

Theorem 3.2 (Localization Error for Corrector). *There exists a constant $s > 0$ such that, for any $u_H \in U_H$ and $\ell \in \mathbb{N}_0$, there holds*

$$\|(\mathcal{C} - \mathcal{C}_\ell) u_H\|_V \lesssim H^{-2} \sqrt{N} \mathrm{e}^{-s\ell} \|u_H\|_V.$$

Proof. First, suppose $\ell \geq 4$. Note that the functions $\varphi_i = \mathcal{C}u_i$ and φ_i^ℓ are uniquely characterized as solutions to the following problems:

$$\begin{aligned} &\varphi_i \in W, && a(\varphi_i, w) = a(u_i, w) && \text{for all } w \in W, \\ &\varphi_i^\ell \in W(N^\ell(\Omega_i)), && a(\varphi_i^\ell, w) = a(u_i, w) && \text{for all } w \in W(N^\ell(\Omega_i)). \end{aligned}$$

Therefore, as $W(N^\ell(\Omega_i)) \subset W$, we can use the properties of a from Lemma 2.1 and Galerkin orthogonality to find that

$$\|\varphi_i - \varphi_i^\ell\|_V \leq \alpha^{-1} C_a \inf_{w \in W(N^\ell(\Omega_i))} \|\varphi_i - w\|_V. \tag{3.19}$$

Let $\eta \in W^{2,\infty}(\Omega)$ be a cut-off function with

$$0 \leq \eta \leq 1, \quad \eta|_{N^{\ell-2}(\Omega_i)} \equiv 1, \quad \eta|_{\Omega \setminus N^{\ell-1}(\Omega_i)} \equiv 0, \quad \|\nabla \eta\|_{L^\infty(\Omega)} + H \|D^2 \eta\|_{L^\infty(\Omega)} \lesssim H^{-1}.$$

Then, setting $w^\ell := (1 - P_H)[\eta \varphi_i] \in W(N^\ell(\Omega_i))$, we have that

$$\|\varphi_i - \varphi_i^\ell\|_V \leq \alpha^{-1} C_a \|\varphi_i - w^\ell\|_V = \alpha^{-1} C_a \|(1 - P_H)[(1 - \eta)\varphi_i]\|_{V, \Omega \setminus N^{\ell-3}(\Omega_i)} \lesssim \|\varphi_i\|_{V, \Omega \setminus N^{\ell-4}(\Omega_i)},$$

where we have used (3.19), the fact that $\ker(P_H) = W$, and an argument analogous to (3.18) for the final bound. By the exponential decay property for $\mathcal{C}$ from Lemma 3.4, we obtain that

$$\|\varphi_i - \varphi_i^\ell\|_V \lesssim \|\mathcal{C}u_i\|_{V, \Omega \setminus N^{\ell-4}(\Omega_i)} \lesssim \mathrm{e}^{-s\ell} \|\mathcal{C}u_i\|_V \lesssim \mathrm{e}^{-s\ell} \|u_i\|_V \tag{3.20}$$

for some constant $s > 0$, where we have used Remark 3.1 in the final step. Using the triangle inequality, the bound (3.20), and the Cauchy–Schwarz inequality, we find that

$$\|(\mathcal{C} - \mathcal{C}_\ell)u_H\|_V = \left\|\sum_{i=1}^N \langle q_i, u_H\rangle(\varphi_i - \varphi_i^\ell)\right\|_V \lesssim \mathrm{e}^{-s\ell}\sum_{i=1}^N |\langle q_i, u_H\rangle|\|u_i\|_V \lesssim \sqrt{N}\,\mathrm{e}^{-s\ell}\sqrt{\sum_{i=1}^N |\langle q_i, u_H\rangle|^2\|u_i\|_V^2},$$

and hence, by Lemma 3.2,

$$\|(\mathcal{C} - \mathcal{C}_\ell)u_H\|_V \lesssim H^{-2}\sqrt{N}\mathrm{e}^{-s\ell}\|u_H\|_V.$$

Finally, in the case $\ell < 4$, we have from (3.19) and Remark 3.1 that

$$\|\varphi_i - \varphi_i^\ell\|_V \le \alpha^{-1}C_a\|\mathcal{C}u_i\|_V \le \alpha^{-2}C_a^2\|u_i\|_V \le \alpha^{-2}C_a^2\mathrm{e}^{4s}\mathrm{e}^{-s\ell}\|u_i\|_V \lesssim \mathrm{e}^{-s\ell}\|u_i\|_V,$$

and we can conclude as before. □

Using similar arguments, one obtains an analogous result for the corrector $\mathcal{C}^*$ and its localized counterpart $\mathcal{C}_\ell^*$.

3.5 Localized Numerical Homogenization Scheme

We are now in a position to state and analyze the practical numerical homogenization method.

3.5.1 The Localized Numerical Homogenization Scheme

We define the N-dimensional spaces

$$\tilde{U}_H^\ell := (1 - \mathcal{C}_\ell)U_H, \quad \tilde{V}_H^\ell := (1 - \mathcal{C}_\ell^*)U_H.$$

Then the numerical homogenization scheme reads as follows:

$$\text{find } \tilde{u}_H^\ell \in \tilde{U}_H^\ell \text{ such that } a(\tilde{u}_H^\ell, \tilde{v}_H^\ell) = \langle F, \tilde{v}_H^\ell\rangle \text{ for all } \tilde{v}_H^\ell \in \tilde{V}_H^\ell. \tag{3.21}$$

3.5.2 Analysis of the Localized Numerical Homogenization Scheme

The following theorem provides well-posedness of (3.21) as well as error bounds.

Theorem 3.3 (Analysis of the Localized Numerical Homogenization Scheme). *Assume that $\ell \gtrsim \log(H^{-2}\sqrt{N})$ is sufficiently large. Then there exists a unique solution $\tilde{u}_H^\ell \in \tilde{U}_H^\ell$ to* (3.21)*. Further, denoting the unique strong solution to* (2.1) *by $u \in V$ and the unique solution to the ideal discrete problem* (3.7) *by $\tilde{u}_H \in \tilde{U}_H$, there exists a constant $s > 0$ such that the following assertions hold true.*

(i) *There holds*

$$\|P_H(\tilde{u}_H - \tilde{u}_H^\ell)\|_V \lesssim H^{-2}\sqrt{N}\mathrm{e}^{-s\ell}\|u\|_V.$$

(ii) *We have the bound*

$$|\langle q_i, u - \tilde{u}_H^\ell\rangle| = |\langle q_i, \tilde{u}_H - \tilde{u}_H^\ell\rangle| \lesssim H^{-2}\sqrt{N}\mathrm{e}^{-s\ell}\|u\|_V \quad \text{for all } i \in \{1, \dots, N\},$$

for the error in the quantities of interest.

(iii) *We have the error bound*

$$\|u - \tilde{u}_H^\ell\|_{H^1(\Omega)} \lesssim (H + H^{-2}\sqrt{N}\mathrm{e}^{-s\ell})\|f\|_{L^2(\Omega)}.$$

Proof. The well-posedness of (3.21) and the bounds from (i)–(ii) can be shown using identical arguments as in [1]. It remains to prove assertion (iii). To this end, we first use the triangle inequality, Theorem 3.1, and (i) to find that

$$\begin{aligned}\|u - \tilde{u}_H^\ell\|_{H^1(\Omega)} &\le \|u - \tilde{u}_H\|_{H^1(\Omega)} + \|P_H(\tilde{u}_H - \tilde{u}_H^\ell)\|_{H^1(\Omega)} + \|(1 - P_H)(\tilde{u}_H - \tilde{u}_H^\ell)\|_{H^1(\Omega)} \\ &\lesssim H\|f\|_{L^2(\Omega)} + H^{-2}\sqrt{N}\mathrm{e}^{-s\ell}\|f\|_{L^2(\Omega)} + \|(1 - P_H)\tilde{e}_H^\ell\|_{H^1(\Omega)}\end{aligned} \tag{3.22}$$

for some constant $s > 0$, where $\tilde{e}_H^\ell := \tilde{u}_H - \tilde{u}_H^\ell$. Next, using the triangle inequality, Remark 3.2 (i), the bound (3.9), and a Morley interpolation estimate, we obtain that

$$\begin{aligned}\|(1 - P_H)\tilde{e}_H^\ell\|_{H^1(\Omega)} &\leq \|(1 - \tilde{E}_H)\Pi^{\mathrm{Mor}}\tilde{e}_H^\ell\|_{H^1(\Omega)} + \|(1 - \Pi^{\mathrm{Mor}})\tilde{e}_H^\ell\|_{H^1(\Omega)} \\ &\lesssim H\|D_{\mathrm{NC}}^2(\Pi^{\mathrm{Mor}}\tilde{e}_H^\ell - \tilde{e}_H^\ell)\|_{L^2(\Omega)} + H\|\tilde{e}_H^\ell\|_V \lesssim H\|\tilde{e}_H^\ell\|_V. \end{aligned} \tag{3.23}$$

Finally, by the triangle inequality, Theorem 3.1 (ii), Remark 3.1, quasi-optimality of the Petrov–Galerkin scheme (3.21), and Remark 2.3, we have that

$$\|\tilde{e}_H^\ell\|_V \leq \|u - \tilde{u}_H\|_V + \|u - \tilde{u}_H^\ell\|_V = \|\mathcal{C}u\|_V + \|u - \tilde{u}_H^\ell\|_V \lesssim \|u\|_V \lesssim \|f\|_{L^2(\Omega)}. \tag{3.24}$$

Combining (3.22)–(3.24) yields the desired bound. □

4 Numerical Experiments

In this section, we numerically investigate the proposed numerical homogenization scheme for nondivergence-form PDEs, which we abbreviate as LOD, due to its origin. We compare it to a conforming Birkhoff–Mansfield finite element method on the respective coarse mesh with mesh size H, denoted as FEM in the convergence plots. To simplify the presentation, H and h denote the minimal side lengths of the elements instead of their diameters in the remainder of this section.

4.1 Conforming Discretization

The method presented in Section 3 is not yet discrete as it relies on the solution of the localized version of the saddle-point problem presented in Remark 3.4, which is still in continuous form. For a finite element discretization, we choose a finite-dimensional subspace $V_h \subset V$. Here we choose V_h to be the H^2-conforming reduced Birkhoff–Mansfield element space [6, 10]. Given a triangle T, we define the space $X(T)$ to be the sum of the space of tricubic polynomials over T (the polynomials that are cubic when restricted to any line parallel to one of the triangle's edges) and the three rational functions $\lambda_1^2\lambda_2/(1 - \lambda_3)$ (with cyclic permutation of the indices 1, 2, 3). The local shape function space $\mathrm{BM}(T)$ is then the nine-dimensional subspace of $X(T)$ consisting of those functions whose normal derivative on any of the three edges is affine. The space V_h is defined as the subspace of $H^2(\Omega) \cap H_0^1(\Omega)$ of functions that belong to $\mathrm{BM}(T)$ when restricted to any triangle T. The nine local degrees of freedom are the point evaluation of the function and of its gradient in the three vertices of any triangle. Fore more details on the method and its variants, we refer to [6].

4.2 Implementation

In all our experiments, we consider the computational domain

$$\Omega := (-1, 1)^2 \subset \mathbb{R}^2$$

and use a mesh $\mathcal{T}_h$ for the fine-scale discretization that resolves all small oscillations of the coefficients. We evaluate relative errors in the L^2-norm, the H^1-seminorm, and the H^2-seminorm with respect to a reference solution originating from the fine mesh $\mathcal{T}_h$ with $h = 2^{-8}$. For the implementation, we used Matlab and extended the code provided in [32].

For the demonstration of the multiscale method, we consider three choices of heterogeneous coefficients in combination with a right-hand side $f \in \{f^{(1)}, f^{(2)}, f^{(3)}\}$, where $f^{(2)} := f^{(1)}$ and the functions $f^{(1)}, f^{(3)}$ are given by

$$f^{(1)}(x) := (x_1 + \cos(3\pi x_1))x_2^3, \quad f^{(3)}(x) := f^{(1)}(x) + 2\Theta(x_1)$$

for $x = (x_1, x_2) \in \Omega$, where $\Theta(t) := \mathbb{1}_{t>0}$ for $t \in \mathbb{R}$. Note that $f^{(3)} \in L^2(\Omega)\backslash H^1(\Omega)$.

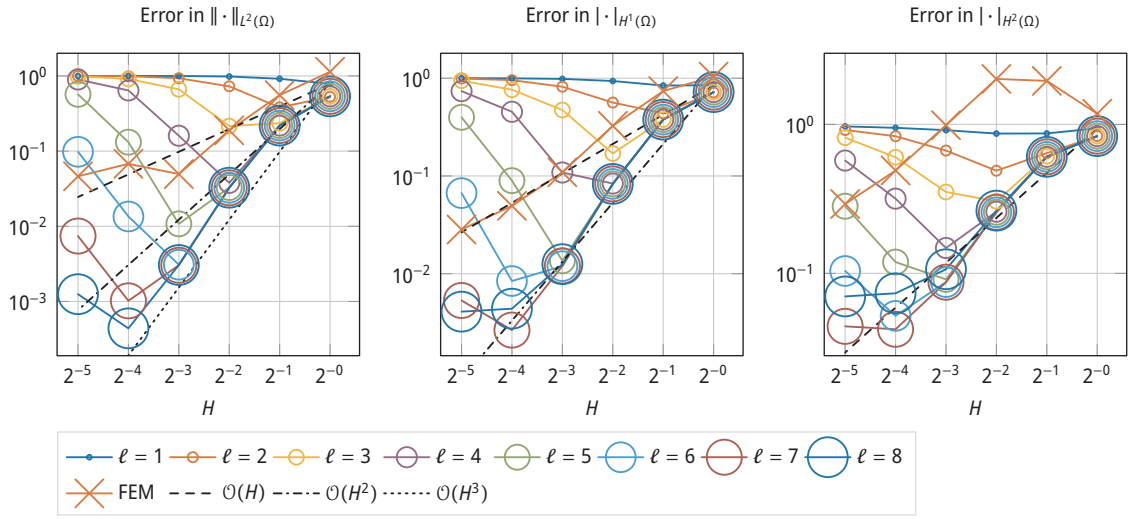

Figure 1: Relative errors for the periodic problem with vanishing lower-order terms ($A = A^{(1)}$, $b = 0$, $c = 0$, $f = f^{(1)}$).

4.3 Periodic Coefficient Example

We begin by considering a configuration with periodic coefficients. The coefficient A is chosen as $A := A^{(1)}$, where $A^{(1)} := \tilde{A}^{(1)}(\frac{\cdot}{\varepsilon})|_\Omega$ with $\varepsilon := 2^{-6}$ and $\tilde{A}^{(1)} \colon \mathbb{R}^2 \to \mathbb{R}^{2\times 2}_{\mathrm{sym}}$ is defined as

$$\tilde{A}^{(1)}(x) := \begin{pmatrix} \frac{11}{4} + \frac{1}{4}\sin(\pi x_1)\cos(\pi x_2) & \operatorname{sign}(\sin(\pi x_1)\sin(\pi x_2)) \\ \operatorname{sign}(\sin(\pi x_1)\sin(\pi x_2)) & \frac{7}{2} + \frac{1}{2}\cos^2(\pi x_1) \end{pmatrix}$$

for $x = (x_1, x_2) \in \mathbb{R}^2$. We perform two numerical experiments with this periodic coefficient $A = A^{(1)}$. The right-hand side is chosen as $f := f^{(1)}$.

4.3.1 Experiment 1: Vanishing Lower-Order Terms

Figure 1 shows the corresponding errors for the case of vanishing lower-order terms ($|b| = c = 0$ a.e. in Ω). Note that the Cordes condition (2.3) is satisfied by Remark 2.1.

4.3.2 Experiment 2: Non-vanishing Lower-Order Terms

For the case of non-vanishing lower-order terms, we choose $b := b^{(1)}$ and $c := c^{(1)}$, where $b^{(1)} := \tilde{b}^{(1)}(\frac{\cdot}{\varepsilon})|_\Omega$, $c^{(1)} := \tilde{c}^{(1)}(\frac{\cdot}{\varepsilon})|_\Omega$ with $\varepsilon = 2^{-6}$ and $\tilde{b}^{(1)} \colon \mathbb{R}^2 \to \mathbb{R}^2$, $\tilde{c}^{(1)} \colon \mathbb{R}^2 \to \mathbb{R}$ are given by

$$\tilde{b}^{(1)}(x) := \begin{pmatrix} \frac{3}{5}\operatorname{sign}(\sin(\pi x_1)\sin(\pi x_2)) \\ \arcsin(\sin^2(\pi x_1)) - \frac{4}{5} \end{pmatrix}, \quad \tilde{c}^{(1)}(x) := \frac{29}{10} + \frac{1}{10}\operatorname{sign}(\sin(\pi x_1)\sin(\pi x_2))$$

for $x = (x_1, x_2) \in \mathbb{R}^2$. Note that the Cordes condition (2.4) is satisfied with $\lambda = 1$ since

$$\frac{|A^{(1)}|^2 + \frac{1}{2}|b^{(1)}|^2 + (c^{(1)})^2}{(\operatorname{tr}(A^{(1)}) + c^{(1)})^2} \leq \frac{27 + \frac{1}{2} + 9}{(6 + \frac{14}{5})^2} = \frac{1}{2 + \frac{222}{1825}} \quad \text{a.e. in } \Omega.$$

The corresponding errors are depicted in Figure 2.

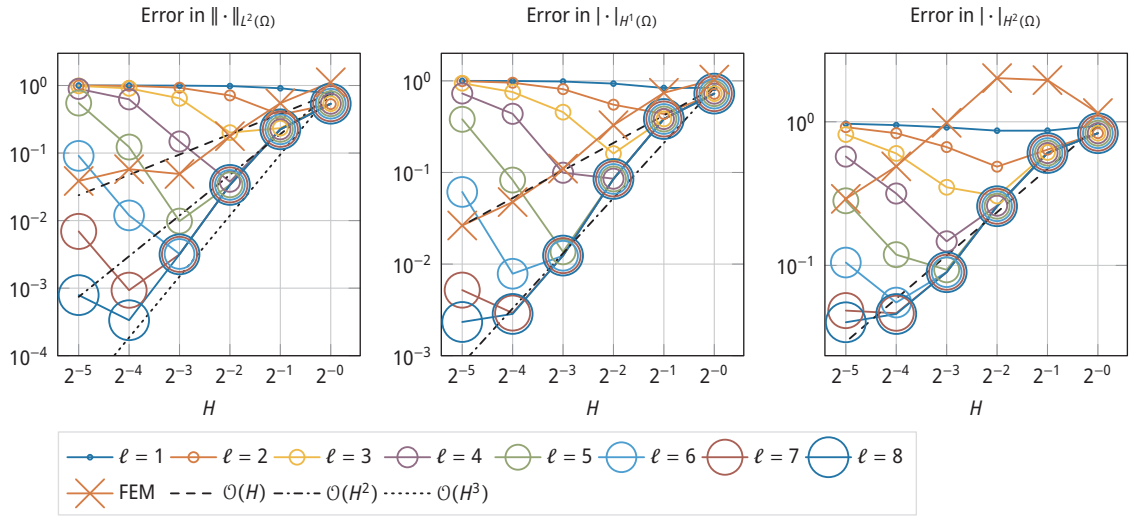

Figure 2: Relative errors for the periodic problem with non-vanishing lower-order terms ($A = A^{(1)}$, $b = b^{(1)}$, $c = c^{(1)}$, $f = f^{(1)}$) and $\lambda = 1$.

4.4 Crack Coefficient Example

Next, we consider an example with $A := A^{(2)}$, where

$$A^{(2)} := \begin{pmatrix} 2 & a_{12}^{(2)} \\ a_{12}^{(2)} & 2 \end{pmatrix}$$

and $a_{12}^{(2)}$ is the realization of a background random field taking piecewise constant values on $\mathcal{T}_{2^{-6}}$, which are independent and identically distributed in the interval $[-1, -0.9]$, which is combined with a channel taking values close to 1 (note $|a_{12}^{(2)}| \leq 1$); see Figure 3 (b). We perform two numerical experiments with this "crack coefficient" $A = A^{(2)}$. The right-hand side is chosen as $f := f^{(2)}$.

4.4.1 Experiment 1: Vanishing Lower-Order Terms

Figure 4 shows the corresponding errors for the case of vanishing lower-order terms ($|b| = c = 0$ a.e. in Ω). Note that the Cordes condition (2.3) is satisfied by Remark 2.1.

4.4.2 Experiment 2: Non-vanishing Lower-Order Terms

For the case of non-vanishing lower-order terms, we choose $b := b^{(2)} := (b_1^{(2)}, b_2^{(2)})$ and $c := c^{(2)}$ that contain cracks at a different position than $a_{12}^{(2)}$. The function $b_1^{(2)}$ consists of a background random field taking values in the interval $[-0.1, 0.1]$ with a crack that varies in the interval $[-1, 1]$. For $b_2^{(2)}$, the random background is identical, whereas the crack varies in $[-0.6, 0.6]$. Finally, $c^{(2)}$ consists of a random background taking values in $[3, 3.1]$ with a crack varying in $[3, 4]$. Figure 3 depicts plots of these coefficients. Note that the Cordes condition (2.4) is satisfied with $\lambda = 2$ since

$$\frac{|A^{(2)}|^2 + \frac{1}{4}|b^{(2)}|^2 + \frac{1}{4}(c^{(2)})^2}{(\operatorname{tr}(A^{(2)}) + \frac{1}{2}c^{(2)})^2} \leq \frac{10 + \frac{17}{50} + 4}{(4 + \frac{3}{2})^2} = \frac{1}{2 + \frac{157}{1434}} \quad \text{a.e. in } \Omega.$$

The corresponding errors are depicted in Figure 5.

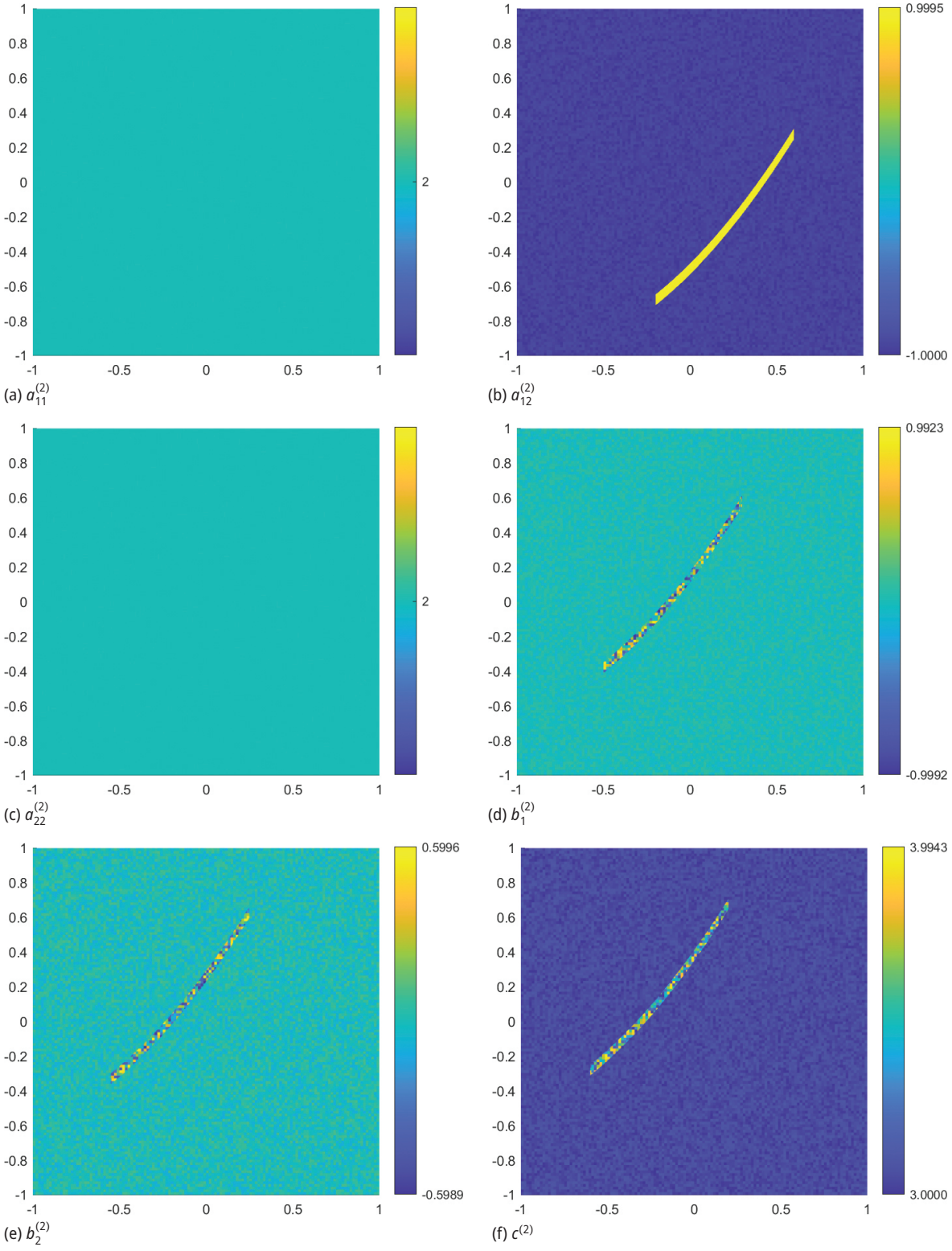

Figure 3: Illustration of the coefficients chosen in Section 4.4.

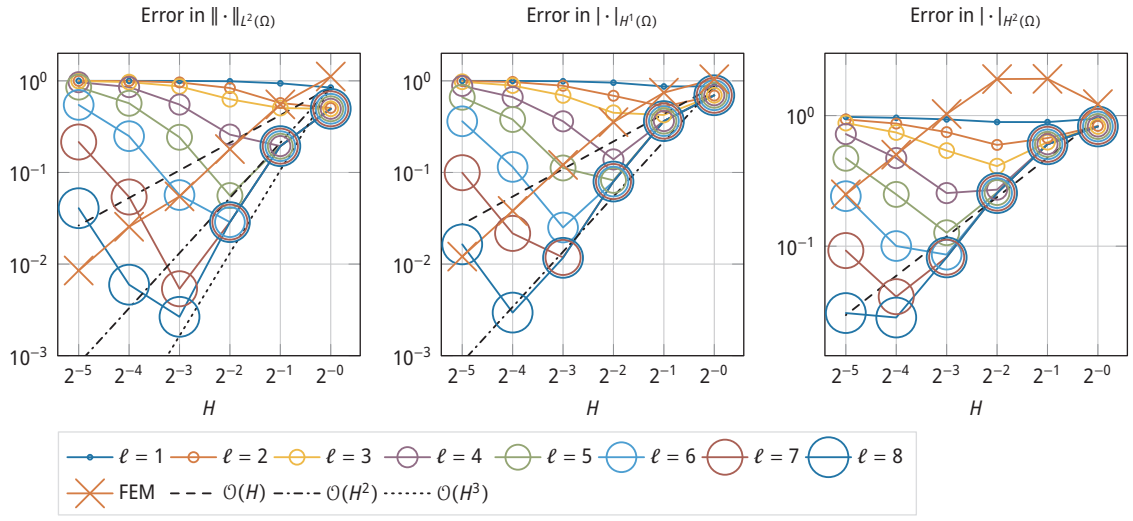

Figure 4: Relative errors for the crack problem with vanishing lower-order terms ($A = A^{(2)}$, $b = 0$, $c = 0$, $f = f^{(2)}$).

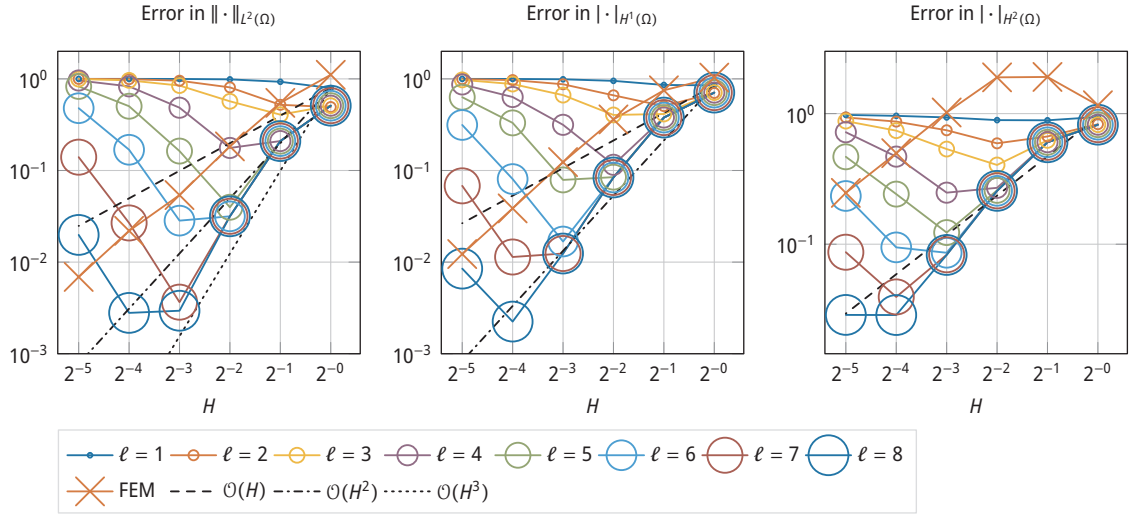

Figure 5: Relative errors for the crack problem with non-vanishing lower-order terms ($A = A^{(2)}$, $b = b^{(2)}$, $c = c^{(2)}$, $f = f^{(2)}$) and $\lambda = 2$.

4.5 Combined Example

The final example combines various types of heterogeneities, and the right-hand side is chosen as

$$f := f^{(3)} \in L^2(\Omega) \backslash H^1(\Omega).$$

A plot of the chosen coefficients $A := A^{(3)}$, $b := b^{(3)}$, and $c := c^{(3)}$ is given in Figure 6. Note that the Cordes condition (2.4) is satisfied with $\lambda = 1$ since

$$\frac{|A^{(3)}|^2 + \frac{1}{2}|b^{(3)}|^2 + (c^{(3)})^2}{(\mathrm{tr}(A^{(3)}) + c^{(3)})^2} \leq \frac{27 + \frac{1}{2} + 9}{(6 + \frac{14}{5})^2} = \frac{1}{2 + \frac{222}{1825}} \quad \text{a.e. in } \Omega.$$

The corresponding errors are depicted in Figure 7.

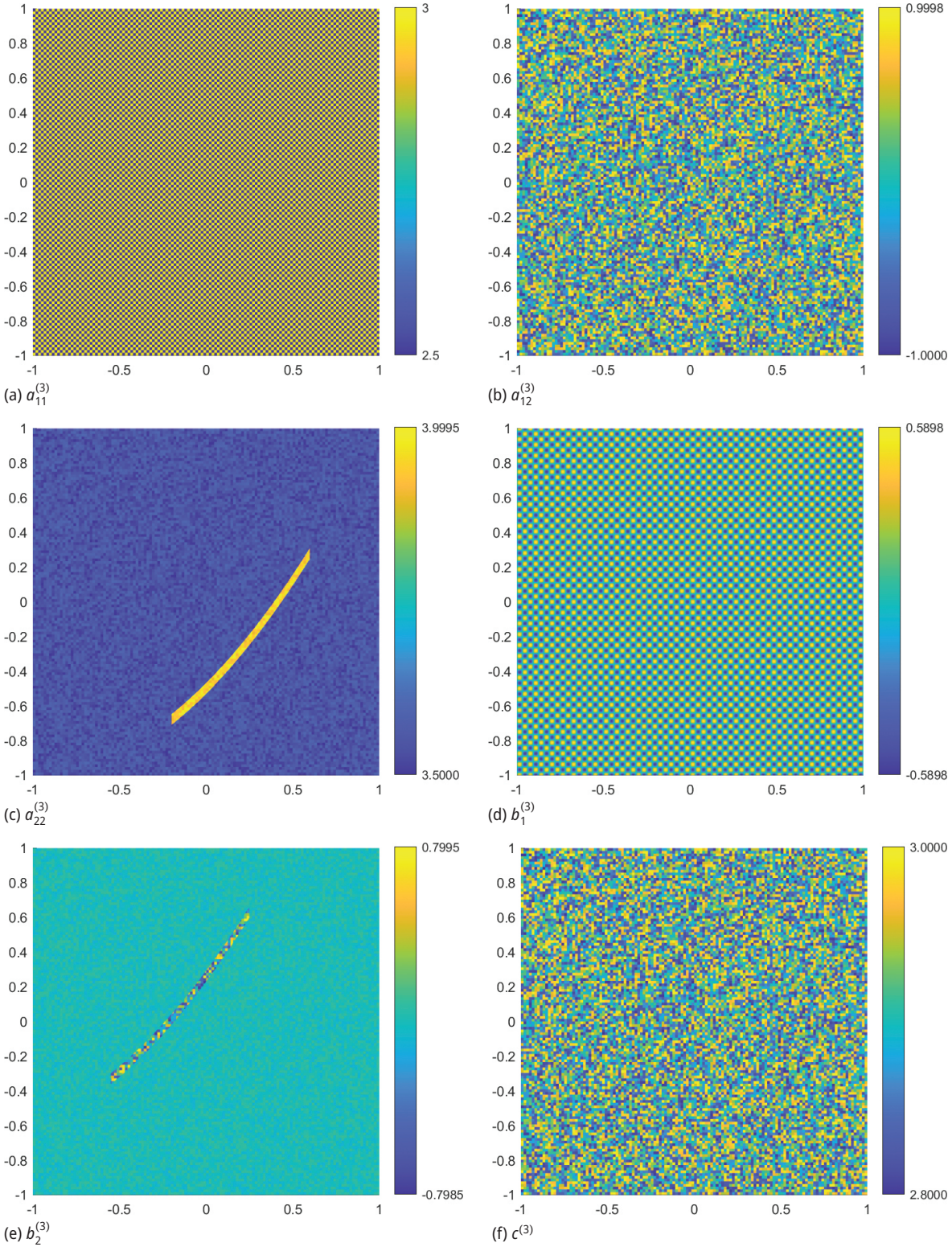

Figure 6: Illustration of the coefficients chosen in Section 4.5.

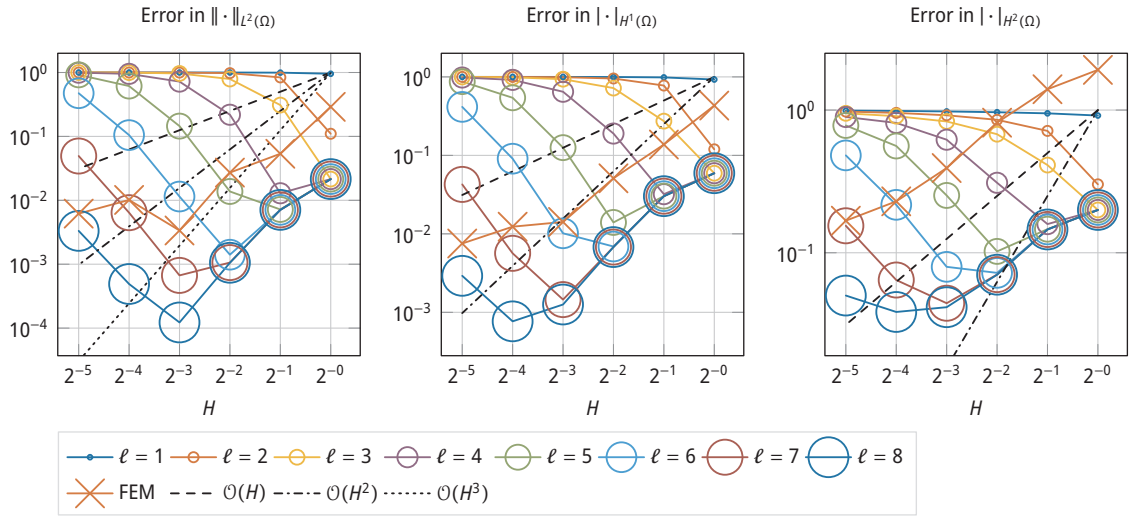

Figure 7: Relative errors for the combined problem ($A = A^{(3)}$, $b = b^{(3)}$, $c = c^{(3)}$, $f = f^{(3)}$) and $\lambda = 1$.

4.6 Conclusions from the Numerical Experiments

In our numerical experiments, we observe that the approximation errors for the LOD scheme in the L^2-norm, the H^1-seminorm, and the H^2-seminorm stay below the corresponding errors for the classical H^2-conforming finite element method and decay at a faster rate with respect to H (order 3/2/1 for LOD as compared to order 1-2/1-2/1 for classical FEM) for $\ell \approx |\log(H^2)|$. The numerical homogenization scheme further avoids the pre-asymptotic behavior of the classical H^2-conforming finite element scheme in the H^2-norm. Let us note that the increase of the approximation error of the LOD method for fixed ℓ and small H is as expected and that it can be explained with the negative powers of H in the error bound from Theorem 3.3 (iii). As compared to the result of Theorem 3.3, the numerical evidence suggests that the proposed method also converges in H^2 and that an improved H^1-error bound might be possible to achieve for right-hand sides $f \in H^1(\Omega)$. The results from Section 4.5 indicate that the $\mathcal{O}(H)$ term in our estimate from Theorem 3.3 (iii) is sharp for $f \in L^2(\Omega)\backslash H^1(\Omega)$.

5 Alternative Discretization Using a Mixed FEM

We emphasize that many other discretizations and combinations with quantities of interest are possible. The method from Section 3 is one particular example and represents the proof of concept that numerical homogenization for nondivergence-form PDEs is possible. During the work on this project, many other possible discretizations were at hand, and we want to present one particularly simple example, which is also quite efficient in the computations and allows using standard finite element techniques. The idea is to use a mixed formulation introduced in [17, 18] and improved in [19]. For the detailed derivation of the scheme, we refer to the respective references. Hence, in this section, we use the mixed method for the solution of the global problem (2.6) and the local problems as given in Remark 3.4 subject to a uniformly refined mesh $\mathcal{T}_h$ that resolves all fine scales. It should be emphasized that the admissible right-hand sides when using a mixed system instead of the problem from Remark 3.4 reduce to a strict subspace of V^*. In particular, point evaluations are unbounded functionals in the mixed setting. Therefore, in our computations, the quantities of interest q_j in the problem from Remark 3.4 are replaced by averaging operations, also known as quasi-interpolation.

We briefly illustrate the performance of the mixed method for the coefficients from the previous section with vanishing lower-order terms. The results can be found in Figures 8–10. In general, we observe that a significantly lower number of oversampling layers is sufficient to achieve similar errors. It is worth noting that, in the case of $f \in L^2(\Omega)\backslash H^1(\Omega)$, the order reduction for the error in the H^1-seminorm does not seem to be present.

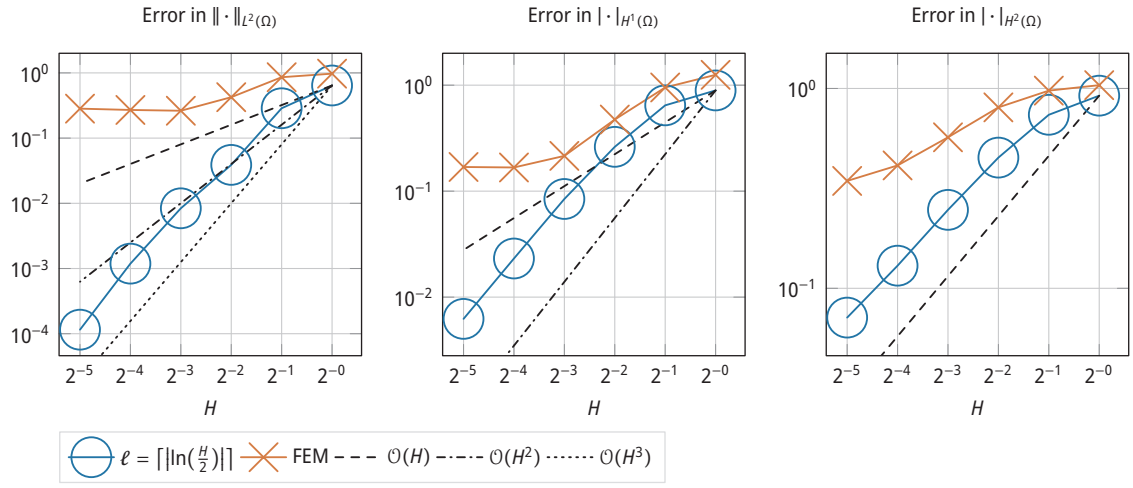

Figure 8: Relative errors for the periodic problem with vanishing lower-order terms ($A = A^{(1)}$, $b = 0$, $c = 0$, $f = f^{(1)}$) using the mixed method.

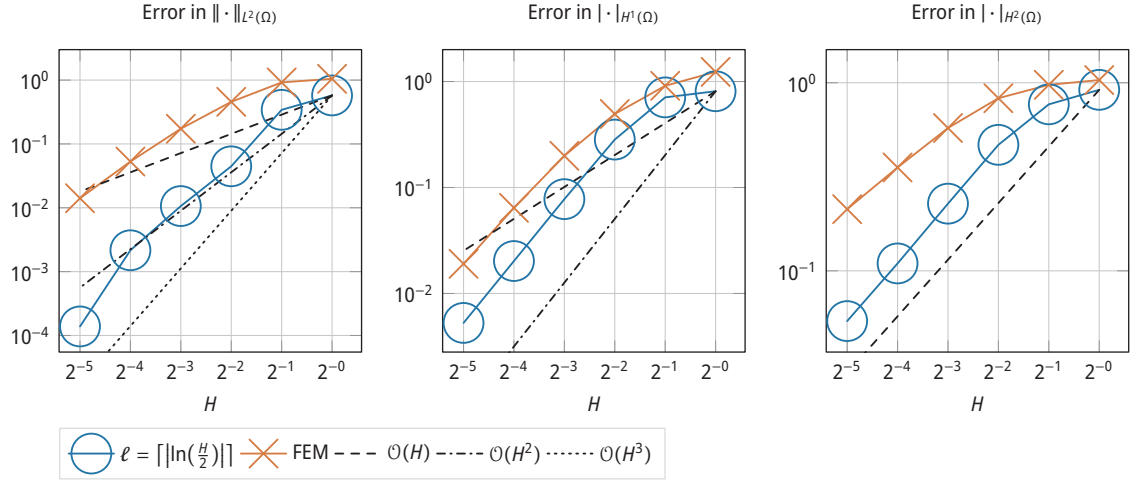

Figure 9: Relative errors for the crack problem with vanishing lower-order terms ($A = A^{(2)}$, $b = 0$, $c = 0$, $f = f^{(2)}$) using the mixed method.

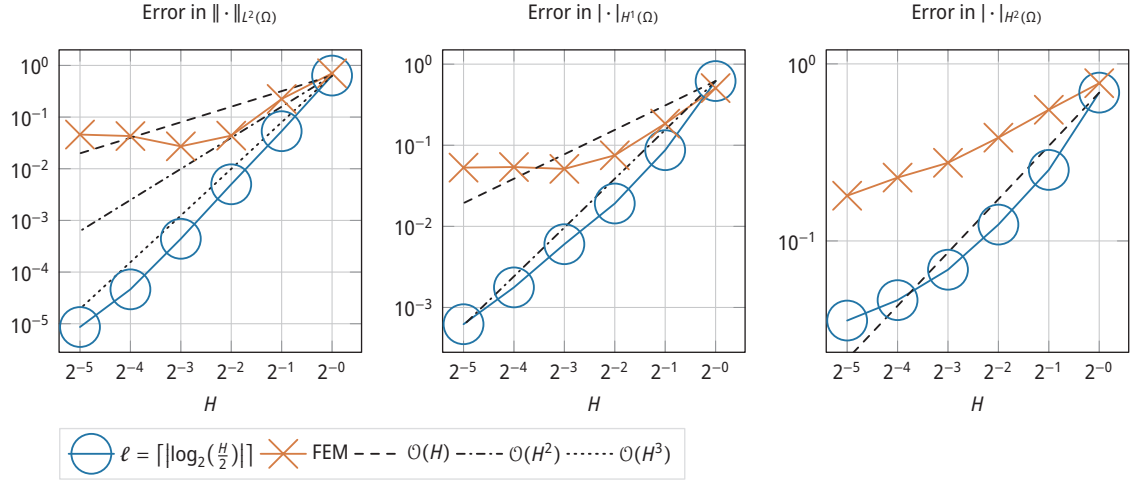

Figure 10: Relative errors for the combined problem with vanishing lower-order terms ($A = A^{(3)}$, $b = 0$, $c = 0$, $f = f^{(3)}$) using the mixed method.

Note that a full error analysis of this scheme is not contained in this work but follows along the same lines of what was presented in this work, with modifications for the mixed formulation and the above-mentioned quasi-interpolation.

6 Concluding Remarks

6.1 Review

In this work, we presented a novel numerical homogenization scheme for linear second-order elliptic PDEs in nondivergence form with coefficients that satisfy a (generalized) Cordes condition. Motivated by the degrees of freedom of the nonconforming Morley finite element, our approach using the LOD framework provides a proof of concept that this method is also applicable to the class of nondivergence-form PDEs. The error analysis revealed that numerical homogenization is applicable to problems with coefficients that do not exhibit any scale separation, even beyond periodicity. Moreover, the favorable accuracy properties of the classical LOD for divergence-form PDEs are preserved. Various numerical experiments have been performed that support the theoretical findings.

6.2 Extensions and Future Work

Finally, we give some remarks regarding extensions of our results and we address future work.

6.2.1 The Case $n \geq 3$

In Section 3, we assumed that $n = 2$ for simplicity of the presentation of the methodology. It is straightforward to adapt this to dimensions $n \geq 3$ by defining the quantities of interest corresponding to the degrees of freedom of the Morley element in dimension n; see [42].

6.2.2 H^2-Convergence of the Numerical Homogenization Scheme

The numerical experiments suggest that the numerical homogenization scheme presented in this work converges not only in the H^1-norm, but also in the H^2-norm. A proof of an H^2-error bound is subject of future work.

6.2.3 Improved Localization

We emphasize that, using an improved localization technique proposed in [23], increasing errors for refinements in H for fixed ℓ can be cured. A potential application of the improved localization using the Super-Localized Orthogonal Decomposition [7, 14, 24] might also be possible. Moreover, we see the potential to use the proposed scheme as a preconditioner.

6.2.4 Different Problem Classes

The method presented in this paper can be applied to any Lax–Milgram-type problem over $V = H^2(\Omega) \cap H^1_0(\Omega)$ of the form (1.3) with $F \in V^*$ and a locally bounded and coercive bilinear form $a\colon V \times V \to \mathbb{R}$.

Funding: The work of P. Freese and D. Gallistl is part of projects that have received funding from the European Research Council ERC under the European Union's Horizon 2020 research and innovation program (Grant agreements No. 865751 and No. 891734). D. Peterseim acknowledges funding by the Deutsche Forschungsgemeinschaft within the research project *Convexified Variational Formulations at Finite Strains based on Homogenised Damaged Microstructures* (PE 2143/5-1).

References

[1] R. Altmann, P. Henning and D. Peterseim, Numerical homogenization beyond scale separation, *Acta Numer.* **30** (2021), 1–86.
[2] D. Arjmand and G. Kreiss, An equation-free approach for second order multiscale hyperbolic problems in non-divergence form, *Commun. Math. Sci.* **16** (2018), no. 8, 2317–2343.
[3] M. Avellaneda and F.-H. Lin, Compactness methods in the theory of homogenization. II. Equations in nondivergence form, *Comm. Pure Appl. Math.* **42** (1989), no. 2, 139–172.
[4] I. Babuska and R. Lipton, Optimal local approximation spaces for generalized finite element methods with application to multiscale problems, *Multiscale Model. Simul.* **9** (2011), no. 1, 373–406.
[5] A. Bensoussan, J.-L. Lions and G. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, AMS Chelsea, Providence, 2011.
[6] G. Birkhoff and L. Mansfield, Compatible triangular finite elements, *J. Math. Anal. Appl.* **47** (1974), 531–553.
[7] F. Bonizzoni, P. Freese and D. Peterseim, Super-localized orthogonal decomposition for convection-dominated diffusion problems, preprint (2022), https://arxiv.org/abs/2206.01975.
[8] F. Camilli and C. Marchi, Rates of convergence in periodic homogenization of fully nonlinear uniformly elliptic PDEs, *Nonlinearity* **22** (2009), no. 6, 1481–1498.
[9] Y. Capdeboscq, T. Sprekeler and E. Süli, Finite element approximation of elliptic homogenization problems in nondivergence-form, *ESAIM Math. Model. Numer. Anal.* **54** (2020), no. 4, 1221–1257.
[10] P. G. Ciarlet, *The Finite Element Method for Elliptic Problems*, Stud. Math. Appl. 4, North-Holland, Amsterdam, 1978.
[11] Y. Efendiev, J. Galvis and T. Y. Hou, Generalized multiscale finite element methods (GMsFEM), *J. Comput. Phys.* **251** (2013), 116–135.
[12] C. Finlay and A. M. Oberman, Approximate homogenization of convex nonlinear elliptic PDEs, *Commun. Math. Sci.* **16** (2018), no. 7, 1895–1906.
[13] C. Finlay and A. M. Oberman, Approximate homogenization of fully nonlinear elliptic PDEs: Estimates and numerical results for Pucci type equations, *J. Sci. Comput.* **77** (2018), no. 2, 936–949.
[14] P. Freese, M. Hauck and D. Peterseim, Super-localized orthogonal decomposition for high-frequency Helmholtz problems, preprint (2021), https://arxiv.org/abs/2112.11368.
[15] B. D. Froese and A. M. Oberman, Numerical averaging of non-divergence structure elliptic operators, *Commun. Math. Sci.* **7** (2009), no. 4, 785–804.
[16] D. Gallistl, Morley finite element method for the eigenvalues of the biharmonic operator, *IMA J. Numer. Anal.* **35** (2015), no. 4, 1779–1811.
[17] D. Gallistl, Stable splitting of polyharmonic operators by generalized Stokes systems, *Math. Comp.* **86** (2017), no. 308, 2555–2577.
[18] D. Gallistl, Variational formulation and numerical analysis of linear elliptic equations in nondivergence form with Cordes coefficients, *SIAM J. Numer. Anal.* **55** (2017), no. 2, 737–757.
[19] D. Gallistl, Numerical approximation of planar oblique derivative problems in nondivergence form, *Math. Comp.* **88** (2019), no. 317, 1091–1119.
[20] D. Gallistl, T. Sprekeler and E. Süli, Mixed finite element approximation of periodic Hamilton–Jacobi–Bellman problems with application to numerical homogenization, *Multiscale Model. Simul.* **19** (2021), no. 2, 1041–1065.
[21] X. Guo, T. Sprekeler and H. V. Tran, Characterizations of diffusion matrices in homogenization of elliptic equations in nondivergence-form, preprint (2022), https://arxiv.org/abs/2201.01974.
[22] X. Guo, H. V. Tran and Y. Yu, Remarks on optimal rates of convergence in periodic homogenization of linear elliptic equations in non-divergence form, *Partial Differ. Equ. Appl.* **1** (2020), no. 4, Paper No. 15.
[23] M. Hauck and D. Peterseim, Multi-resolution localized orthogonal decomposition for Helmholtz problems, *Multiscale Model. Simul.* **20** (2022), no. 2, 657–684.
[24] M. Hauck and D. Peterseim, Super-localization of elliptic multiscale problems, *Math. Comp.* **92** (2023), no. 341, 981–1003.
[25] P. Henning and D. Peterseim, Oversampling for the multiscale finite element method, *Multiscale Model. Simul.* **11** (2013), no. 4, 1149–1175.
[26] V. V. Jikov, S. M. Kozlov and O. A. Oleĭnik, *Homogenization of Differential Operators and Integral Functionals*, Springer, Berlin, 1994.
[27] E. L. Kawecki and T. Sprekeler, Discontinuous Galerkin and C^0-IP finite element approximation of periodic Hamilton–Jacobi–Bellman–Isaacs problems with application to numerical homogenization, *ESAIM Math. Model. Numer. Anal.* **56** (2022), no. 2, 679–704.
[28] S. Kim and K.-A. Lee, Higher order convergence rates in theory of homogenization: Equations of non-divergence form, *Arch. Ration. Mech. Anal.* **219** (2016), no. 3, 1273–1304.

[29] R. Kornhuber, D. Peterseim and H. Yserentant, An analysis of a class of variational multiscale methods based on subspace decomposition, *Math. Comp.* **87** (2018), no. 314, 2765–2774.
[30] C. Ma, R. Scheichl and T. Dodwell, Novel design and analysis of generalized finite element methods based on locally optimal spectral approximations, *SIAM J. Numer. Anal.* **60** (2022), no. 1, 244–273.
[31] A. Målqvist and D. Peterseim, Localization of elliptic multiscale problems, *Math. Comp.* **83** (2014), no. 290, 2583–2603.
[32] A. Målqvist and D. Peterseim, *Numerical Homogenization by Localized Orthogonal Decomposition*, SIAM Spotlights 5, Society for Industrial and Applied Mathematics, Philadelphia,2021.
[33] R. Maier, A high-order approach to elliptic multiscale problems with general unstructured coefficients, *SIAM J. Numer. Anal.* **59** (2021), no. 2, 1067–1089.
[34] H. Owhadi, Multigrid with rough coefficients and multiresolution operator decomposition from hierarchical information games, *SIAM Rev.* **59** (2017), no. 1, 99–149.
[35] H. Owhadi and L. Zhang, Metric-based upscaling, *Comm. Pure Appl. Math.* **60** (2007), no. 5, 675–723.
[36] H. Owhadi, L. Zhang and L. Berlyand, Polyharmonic homogenization, rough polyharmonic splines and sparse super-localization, *ESAIM Math. Model. Numer. Anal.* **48** (2014), no. 2, 517–552.
[37] I. Smears and E. Süli, Discontinuous Galerkin finite element approximation of nondivergence form elliptic equations with Cordès coefficients, *SIAM J. Numer. Anal.* **51** (2013), no. 4, 2088–2106.
[38] I. Smears and E. Süli, Discontinuous Galerkin finite element approximation of Hamilton–Jacobi–Bellman equations with Cordes coefficients, *SIAM J. Numer. Anal.* **52** (2014), no. 2, 993–1016.
[39] T. Sprekeler, Homogenization of nondivergence-form elliptic equations with discontinuous coefficients and finite element approximation of the homogenized problem, preprint (2023), https://arxiv.org/abs/2305.19833.
[40] T. Sprekeler and H. V. Tran, Optimal convergence rates for elliptic homogenization problems in nondivergence-form: Analysis and numerical illustrations, *Multiscale Model. Simul.* **19** (2021), no. 3, 1453–1473.
[41] D. B. Szyld, The many proofs of an identity on the norm of oblique projections, *Numer. Algorithms* **42** (2006), no. 3–4, 309–323.
[42] M. Wang and J. Xu, Minimal finite element spaces for $2m$-th-order partial differential equations in R^n, *Math. Comp.* **82** (2013), no. 281, 25–43.