Finite element heterogeneous multiscale method for nonlinear monotone parabolic homogenization problems

A. Abdulle[1], M.E. Huber[1]

July 1, 2014

Abstract

We propose a multiscale method based on a finite element heterogeneous multiscale method (in space) and the implicit Euler integrator (in time) to solve nonlinear monotone parabolic problems with multiple scales due to spatial heterogeneities varying rapidly at a microscopic scale. The multiscale method approximates the solution at the scale of interest at computational cost independent of the small scale by performing numerical upscaling (coupling of macro and micro finite element methods). Optimal a priori error estimates in the $L^2(H^1)$ and $\mathcal{C}^0(L^2)$ norm are derived taking into account the error due to time discretization as well as macro and micro spatial discretizations. Further, we present numerical simulations to illustrate the theoretical error estimates and the applicability of the multiscale method to practical problems.

Keywords: nonlinear monotone parabolic problem, multiple scales, heterogeneous multiscale method, finite elements, implicit Euler, fully discrete error, resonance error.

AMS subject classification (2010): 65N30, 65M60, 74Q10, 74D10.

1 Introduction

In this article, we propose a numerical method to solve a class of nonlinear monotone parabolic multiscale problems of the type

$$\partial_t u^\varepsilon - \operatorname{div}(\mathcal{A}^\varepsilon(x, \nabla u^\varepsilon)) = f, \qquad \text{in } \Omega \times (0,T), \tag{1}$$

where $\Omega \subset \mathbb{R}^d$, $d \leq 3$, is a bounded domain, $(0,T)$ a finite time interval and initial data as well as zero Dirichlet boundary conditions are prescribed. We consider maps $\mathcal{A}^\varepsilon(x,\xi)$ that highly oscillate at a small scale ε with respect to the space variable x and are nonlinear and monotone in the second variable ξ. Many physical processes can be modeled by parabolic partial differential equations (PDEs) of the form (1), e.g., non-Newtonian fluids, ferromagnetic materials or composites with nonlinear material laws, see [10, 33].

Using standard numerical methods, like the finite element method (FEM), to discretize the problem (1) in space leads to high computational cost as the small scale ε of the spatial heterogeneities of $\mathcal{A}^\varepsilon$ has to be resolved. Thus, to efficiently approximate the solution of (1) at the scale of interest, effective models for (1) are needed. Homogenization theory, see [11, 30], is the usual framework used to study the solutions u^ε to (1) in the limit $\varepsilon \to 0$ and aims at characterizing a limiting function u^0 as the solution of a homogenized (or effective) equation. The upscaling of (1) has been studied by Pankov and Svanstedt in [34] and [36], respectively, using the notion of parabolic G-convergence (extending the work by Tartar, see [38] and [39, Chapter 11], to parabolic problems). In particular, the homogenized equation (with solution u^0) is again of the same type as (1) with $\mathcal{A}^\varepsilon$ replaced by the homogenized map $\mathcal{A}^0$ for which the small scales are averaged out.

For linear homogenization PDEs, a broad literature about multiscale methods exists nowadays, see [2, 4] (elliptic problems), [8, 31] (parabolic problems) and the references therein. Numerical homogenization methods for nonlinear problems are however less numerous, e.g., see [9] for an overview of numerical methods for multiscale PDEs with a *nonmonotone* nonlinearity (with respect to the solution u^ε). For parabolic multiscale PDEs with *monotone* nonlinearities (with respect to the gradient ∇u^ε) given by (1), Svanstedt et al. proposed in [37] a numerical method for periodically oscillating (in space and time) maps $\mathcal{A}^\varepsilon$ based on an augmented Lagrangian method. In [23], Efendiev et al. applied a generalized

[1]ANMC, Mathematics Section, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland; Assyr.Abdulle@epfl.ch, Martin.Huber@epfl.ch.

multiscale finite element method (MsFEM) – developed in [22] for elliptic monotone problems – to PDE (1) with stochastic heterogeneities. We note that in both [37] and [23] convergence of the numerical solution to the homogenized solution is shown without deriving explicit convergence rates. Further, we note that for *elliptic* monotone multiscale PDEs a sparse tensor FEM and a finite element heterogeneous multiscale method (FE-HMM) has been studied in [28], [27] and [24], respectively, while those concepts for multiscale methods have not yet been applied to the parabolic problem (1). In particular, Gloria studied in [24] numerical homogenization methods (FE-HMM and MsFEM) for a class of elliptic monotone PDEs (associated to minimization problems), proved convergence of their modeling error and derived a priori estimates in the $W^{1,p}$ norm for FE-HMM applied to periodic problems with p-structure with $p \geq 2$. In contrast, our results are valid for monotone maps without assuming that $\mathcal{A}^\varepsilon$ has an associated scalar potential. For example, maps $\mathcal{A}^\varepsilon(x,\xi) = a^\varepsilon(x)\xi$ with a non-symmetric tensor a^ε positive definite and bounded (linear problem) or $\mathcal{A}^\varepsilon(x,\xi) = a^\varepsilon(x)(1 + (1 + \sum_{i=1}^d \xi_i^4)^{-1/4})\xi$ (nonlinear problems) with a^ε positive definite and bounded are allowed.

In this article, we introduce a multiscale method to solve nonlinear monotone parabolic multiscale problems of type (1) following the design principles of the heterogeneous multiscale method (HMM), see [20, 4]. Based on a homogenization result ensuring the existence of an effective model associated to (1), we solve the effective problem using a macroscopic finite element method and the implicit Euler scheme for time integration. While the effective problem is (in general) not available in closed form, we approximate the effective properties of the map $\mathcal{A}^\varepsilon$ by upscaling the available micro informations. This is achieved by solving nonlinear monotone elliptic PDEs (constrained by the macro state) using a microscopic finite element method within micro domains which are of the size of the finest scale ε. The computational complexity of the multiscale method is thus independent of the smallest scale ε. We note, that piecewise affine functions are used for both macro and micro finite element methods ($\mathcal{P}^1$-FEM). For further developments, we refer to [6] where we study a linearized variant of the proposed multiscale method for problems (1) with maps $\mathcal{A}^\varepsilon$ decomposed as $\mathcal{A}^\varepsilon(x,\xi) = a^\varepsilon(x,\xi)\xi$, where $a^\varepsilon(x,\xi) \in \mathbb{R}^{d\times d}$.

Our main contribution in this article is the fully discrete a priori error analysis of the proposed multiscale method. Without any structural assumptions on the spatial heterogeneities of the maps $\mathcal{A}^\varepsilon$, we derive *sharp* error estimates in both $L^2(H^1)$ and $\mathcal{C}^0(L^2)$ norms with respect to the timestep size and the spatial mesh size for macro as well as micro discretizations. Such estimates are crucial to balance temporal, macro *and* micro spatial discretizations to obtain a given precision at minimal computational cost and are not available for the methods proposed in [37, 23]. Further, we note that while our method is formulated for general spatial variations (contrary to [37] where periodicity is assumed), we do not consider maps $\mathcal{A}^\varepsilon$ with rapid oscillations in time (unlike [37, 23]). Finally, in contrast to [24] (for elliptic problems), explicit estimates of the modeling error are proved for nonlinear monotone maps $\mathcal{A}^\varepsilon$ without assuming that $\mathcal{A}^\varepsilon$ has an associated scalar potential (the results in [24] are however valid for elliptic problems in $W^{1,p}$ spaces for general $p \geq 1$, while we assume $p = 2$). Our modeling error results also generalise results obtained in [21] for linear elliptic problems.

To derive the a priori error estimates the total error is decomposed into time discretization error, spatial macro error, modeling error (due to numerical upscaling) and spatial micro error. First, while sharp estimates of the spatial micro error have been derived in literature for linear micro problems (e.g., see [3, 9]), our analysis involves nonlinear micro problems. Second, we note that the estimates of temporal and spatial macro error cannot be shown using standard techniques from FEM analysis for parabolic single scale problems, see [18, 40]. It is known that an elliptic projection is required to obtain sharp estimates of the spatial error in the $\mathcal{C}^0(L^2)$ norm, see [41, 40]. An appropriate projection for parabolic problems $\partial_t u + \operatorname{div}(\mathcal{A}(\nabla u)) = f$ with a monotone nonlinearity in ∇u seems however not to be available in literature. In [16], using an elliptic projection solving a stationary *nonlinear* monotone PDE, Dendy derived $\mathcal{C}^0(L^2)$ estimates for a space-discrete method which are non-optimal for $\mathcal{P}^1$-FEM (convergence rate $2 - d/2$, see [16, Thms 2.2, 2.5]). Optimal convergence of the spatial L^2-error for a space-discrete method with $\mathcal{P}^1$-FEM is however deduced from maximum norm error estimates in [17]. Our proof follows a different strategy which is based on a new *linear* elliptic projection (see (27)) and $W^{1,\infty}$ error estimates for $\mathcal{P}^1$-FEM (for *linear* elliptic PDEs). Compared to [17], we do not derive estimates in the spatial L^∞ norm, but we consider additionally the time discretization error and the influence of variational crimes (due to numerical quadrature and upscaling).

The outline of this article is as follows. In Section 2, we introduce a model problem of type (1) and its associated effective problem. Then, we define in Section 3 a multiscale method based on a numerical upscaling procedure. While the main results of this article, the fully-discrete a priori error estimates for the multiscale algorithm, are presented in Section 4, their proofs are given in Section 5. Further, in Section 6, we discuss an implementation of the proposed method and several numerical tests. In Section 7,

we conclude the article with some remarks about possible generalizations and future research.

Notation 1.1. In what follows, C denotes a generic positive constant, whose value can change at any occurrence. For $D \in \mathbb{R}^m$, we use $\mathcal{C}^k(D,\mathbb{R}^n)$ for the set of k-times continuously differentiable functions $g\colon D \to \mathbb{R}^n$. We consider the usual Sobolev spaces $W^{k,p}(\Omega)$. For $p=2$, we use the notation $H^k(\Omega)$, $H^1_0(\Omega)$ for $p=2$ and $k=1$ with a vanishing trace on the boundary $\partial\Omega$, $H^{-1}(\Omega)$ for the dual space of $H^1_0(\Omega)$ and $W^1_{per}(Y) = \{v \in H^1_{per}(Y) \,|\, \int_Y v(y)dy = 0\}$ where $H^k_{per}(Y)$ is defined as the closure of $\mathcal{C}^\infty_{per}(Y)$ (the subset of $\mathcal{C}^\infty(\mathbb{R}^d)$ of periodic functions in $Y=(0,1)^d$) for the H^k norm. For $g\colon [0,T] \to X$ with Banach space $(X, \|\cdot\|_X)$ the time derivative of g is denoted by $\partial_t g(t)$. The space of L^p functions g and continuous functions g with values in X is denoted by $L^p(0,T;X)$ and $\mathcal{C}^0([0,T],X)$, respectively. Both spaces form a Banach space when endowed with the norm $\|g\|_{L^p(0,T;X)} = (\int_0^T \|g(t)\|_X^p dt)^{1/p}$ and $\|g\|_{\mathcal{C}^0([0,T],X)} = \sup_{t\in[0,T]} \|g(t)\|_X$, respectively. The Euclidean norm for $b \in \mathbb{R}^d$ and the Frobenius norm for $a \in \mathbb{R}^{d\times d}$ are denoted by $|b|$ and $\|a\|_{\mathcal{F}}$, respectively, and the canonical basis of $\mathbb{R}^d$ is given by $e_1,\dots,e_d$.

2 Model problem and homogenization

Let $\Omega \subset \mathbb{R}^d$, $d \le 3$, be a convex polygonal domain and $T>0$. We consider the class of monotone parabolic multiscale problems

$$\begin{aligned} \partial_t u^\varepsilon(x,t) - \operatorname{div}(\mathcal{A}^\varepsilon(x,\nabla u^\varepsilon(x,t))) &= f(x) \text{ in } \Omega\times(0,T), \\ u^\varepsilon(x,t) = 0 \text{ on } \partial\Omega\times(0,T), \quad u^\varepsilon(x,0) &= g(x) \text{ in } \Omega, \end{aligned} \tag{2}$$

with given source $f \in L^2(\Omega)$, initial condition $g \in L^2(\Omega)$ and maps $\mathcal{A}^\varepsilon\colon \Omega\times\mathbb{R}^d \to \mathbb{R}^d$ (indexed by ε) with the property that $\mathcal{A}^\varepsilon(\cdot,\xi)\colon \Omega\to\mathbb{R}^d$ is Lebesgue measurable for every $\xi\in\mathbb{R}^d$. We note that the variable $\varepsilon>0$ represents a small scale in the problem, at which the maps $\mathcal{A}^\varepsilon(\cdot,\xi)$ rapidly vary. We consider homogeneous Dirichlet boundary conditions for simplicity but the results remain valid for other type of boundary conditions. We assume that the maps $\mathcal{A}^\varepsilon$ satisfy the following conditions uniformly in $\varepsilon>0$

($\mathcal{A}_0$) there is some $C_0>0$ such that $|\mathcal{A}^\varepsilon(x,0)| \le C_0$ for almost every (a.e.) $x\in\Omega$;

($\mathcal{A}_1$) the map $\mathcal{A}^\varepsilon(x,\cdot)\colon \mathbb{R}^d\to\mathbb{R}^d$ is Lipschitz continuous, i.e., there exists $L>0$ such that

$$|\mathcal{A}^\varepsilon(x,\xi_1) - \mathcal{A}^\varepsilon(x,\xi_2)| \le L\,|\xi_1-\xi_2|, \qquad \forall\,\xi_1,\xi_2\in\mathbb{R}^d, \text{ a.e. } x\in\Omega;$$

($\mathcal{A}_2$) the map $\mathcal{A}^\varepsilon(x,\cdot)\colon \mathbb{R}^d\to\mathbb{R}^d$ is strongly monotone, i.e., there exists $\lambda>0$ such that

$$(\mathcal{A}^\varepsilon(x,\xi_1) - \mathcal{A}^\varepsilon(x,\xi_2))\cdot(\xi_1-\xi_2) \ge \lambda|\xi_1-\xi_2|^2, \qquad \forall\,\xi_1,\xi_2\in\mathbb{R}^d, \text{ a.e. } x\in\Omega.$$

We note that, hypotheses ($\mathcal{A}_{0-1}$) imply linear growth of $\mathcal{A}^\varepsilon$ with respect to ξ

$$|\mathcal{A}^\varepsilon(x,\xi)| \le L(L_0+|\xi|), \quad \text{where } L_0 = C_0/L, \quad \forall\,\xi\in\mathbb{R}^d, \varepsilon>0, \text{ a.e. } x\in\Omega. \tag{3}$$

Let us give two examples of maps $\mathcal{A}^\varepsilon$ satisfying ($\mathcal{A}_{0-2}$).

Example 1. For linear maps $\mathcal{A}^\varepsilon(x,\xi)$ given by

$$\mathcal{A}^\varepsilon(x,\xi) = a^\varepsilon(x)\xi, \qquad \text{with } a^\varepsilon(x)\in(L^\infty(\Omega))^{d\times d}, \qquad \varepsilon>0,$$

with a uniformly elliptic and bounded family of tensors a^ε the maps $\mathcal{A}^\varepsilon$ satisfy ($\mathcal{A}_{0-2}$) with constants $C_0=0$ and λ given by the ellipticity constant. For such linear data we recover the linear parabolic multiscale problems studied in [8].

Example 2. Next one might consider maps $\mathcal{A}^\varepsilon(x,\xi) = a^\varepsilon(x,\xi)\xi$ where $a^\varepsilon(\cdot,\xi)\in(L^\infty(\Omega))^{d\times d}$ (for $\xi\in\mathbb{R}^d$) is depending on ξ, introducing thus a nonlinearity. We note that adequate conditions have to be imposed on a^ε such that $\mathcal{A}^\varepsilon$ satisfies ($\mathcal{A}_{1-2}$).

As a particular example, let $\mu^\varepsilon\colon \Omega\times\mathbb{R}\to\mathbb{R}_{\ge 0}$ be a continuous function and the maps $\mathcal{A}^\varepsilon$ be given by

$$\mathcal{A}^\varepsilon(x,\xi) = \mu^\varepsilon(x,|\xi|)\xi, \qquad x\in\Omega, \xi\in\mathbb{R}^d,$$

which is an extension of the problems studied in [29] to a multiscale context. If $\mu^\varepsilon(x,t)$ is uniformly (in ε and x) Lipschitz continuous and strongly monotone then the assumption ($\mathcal{A}_{0-2}$) are valid for $\mathcal{A}^\varepsilon$, see [29]. We mention for instance Carreau laws, used to model non-Newtonian fluids, with $\mu^\varepsilon(x,\cdot)\sim 1+(1+t^2)^{\theta-1}$ where $1/2<\theta\le 1$.

Existence and uniqueness of a solution to problem (2) is studied in the Banach space

$$E = \{v \in L^2(0,T;H_0^1(\Omega)) \,|\, \partial_t v \in L^2(0,T;H^{-1}(\Omega))\},$$

endowed with the norm $\|v\|_E = \|v\|_{L^2(0,T;H_0^1(\Omega))} + \|\partial_t v\|_{L^2(0,T;H^{-1}(\Omega))}$ and satisfies the continuous embedding $E \hookrightarrow C^0([0,T],L^2(\Omega))$. Under the assumptions $(\mathcal{A}_{0-2})$ the problem (2) has a unique solution $u^\varepsilon \in E$ for $\varepsilon > 0$, e.g., see [43, Theorem 30.A], which are uniformly bounded

$$\|u^\varepsilon\|_E \leq C(C_0 + \|f\|_{L^2(\Omega)} + \|g\|_{L^2(\Omega)}), \qquad \forall\, \varepsilon > 0.$$

Thus, $\{u^\varepsilon\}$ is a bounded sequence in E and by compactness there exists a subsequence, still denoted by $\{u^\varepsilon\}$, and some $u^0 \in E$, such that

$$u^\varepsilon \rightharpoonup u^0 \ \text{ in } L^2(0,T;H_0^1(\Omega)) \quad \text{ and } \quad \partial_t u^\varepsilon \rightharpoonup \partial_t u^0 \ \text{ in } L^2(0,T;H^{-1}(\Omega)), \quad \text{for } \varepsilon \to 0. \tag{4}$$

The idea of homogenization is to find a limiting equation for u^0. For the problem (2) with $(\mathcal{A}_{0-2})$, this question is studied in terms of G-convergence of parabolic operators, sometimes referred to as PG-convergence or strong G-convergence, see [36, 34]. It can be shown that there exists a subsequence of $\{u^\varepsilon\}$, still denoted by $\{u^\varepsilon\}$, and a map $\mathcal{A}^0 \colon \Omega \times \mathbb{R}^d \to \mathbb{R}^d$, such that u^ε weakly converges to u^0 in the sense of (4) and $\mathcal{A}^\varepsilon(x,\nabla u^\varepsilon) \rightharpoonup \mathcal{A}^0(x,\nabla u^0)$ weakly in $L^2(0,T;(L^2(\Omega))^d)$, where $u^0 \in E$ is the solution of the homogenized or effective problem

$$\begin{aligned} \partial_t u^0(x,t) - \operatorname{div}(\mathcal{A}^0(x,\nabla u^0(x,t))) &= f(x) \ \text{ in } \Omega \times (0,T), \\ u^0(x,t) = 0 \ \text{ on } \partial\Omega \times (0,T), \quad u^0(x,0) &= g(x) \ \text{ in } \Omega, \end{aligned} \tag{5}$$

where $\mathcal{A}^0$ satisfies $(\mathcal{A}_{0-2})$ with possibly different constants C_0 and L. For maps $\mathcal{A}^\varepsilon$ with additional structure, e.g., $\mathcal{A}^\varepsilon(x,\xi) = \mathcal{A}(x/\varepsilon,\xi)$ with $\mathcal{A}(y,\xi)$ a Y-periodic function in y, an explicit representation of $\mathcal{A}^0$ can be derived and thus the whole sequence $\{u^\varepsilon\}$ converges to u^0 in the sense of (4).

3 Multiscale method

In this section, we propose a multiscale method to solve nonlinear monotone parabolic multiscale problems with general spatial heterogeneities. We introduce then a reformulation of that method which is convenient for the analysis and show the existence, uniqueness and boundedness of the numerical solution.

3.1 FE-HMM for nonlinear monotone parabolic problems

The definition of multiscale method studied in this article requires a macroscopic spatial discretization of the domain Ω.

Macro discretization. Let $\mathcal{T}_H$ be a family of macro partitions of Ω consisting of conforming, shape-regular meshes with simplicial elements $K \in \mathcal{T}_H$. We assume that the elements $K \in \mathcal{T}_H$ are open and satisfy $\cup_{K\in\mathcal{T}_H} \bar{K} = \Omega$ (recall that Ω is polygonal). The macro mesh size H is defined by $H = \max_{K\in\mathcal{T}_H} \operatorname{diam} K$, where $\operatorname{diam} K$ denotes the diameter of $K \in \mathcal{T}_H$. Then, we consider the macro finite element space

$$S_0^1(\Omega,\mathcal{T}_H) = \{v^H \in H_0^1(\Omega) \,|\, v^H|_K \in \mathcal{P}^1(K), \forall\, K \in \mathcal{T}_H\}, \tag{6}$$

where $\mathcal{P}^1(K)$ is the space of linear polynomials on $K \in \mathcal{T}_H$. Further, the multiscale method is based on barycentric quadrature

$$\int_\Omega \varphi(x)dx \approx \sum_{K\in\mathcal{T}_H} |K|\varphi(x_K), \qquad \varphi \in \mathcal{C}^0(\Omega,\mathbb{R}), \tag{7}$$

where x_K and $|K|$ denote the barycenter and the measure of $K \in \mathcal{T}_H$, respectively,. We note that the quadrature formula (7) is exact for affine functions φ. Further, for any macro element $K \in \mathcal{T}_H$ we define the sampling domain K_δ located at the quadrature point x_K

$$K_\delta = x_K + \delta\, I, \quad \text{ where } I = (-1/2,1/2)^d \text{ and } \delta \geq \varepsilon.$$

Within the sampling domains, micro simulations are performed to recover the upscaled data.

Multiscale method. Let the time interval $(0,T)$ be uniformly divided into N subintervals of length $\Delta t = T/N$ and define $t_n = n\Delta t$ for $0 \le n \le N$ and $N \in \mathbb{N}_{>0}$. Let $u_0^H \in S_0^1(\Omega, \mathcal{T}_H)$ be a given approximation of the initial condition $g(x)$. To compute an approximation of the effective solution of problem (2) we propose the multiscale method given by the recursion: for $0 \le n \le N-1$, find $u_{n+1}^H \in S_0^1(\Omega, \mathcal{T}_H)$ such that

$$\int_\Omega \frac{u_{n+1}^H - u_n^H}{\Delta t}\, w^H dx + B^H(u_{n+1}^H; w^H) = \int_\Omega f\, w^H dx, \qquad \forall\, w^H \in S_0^1(\Omega, \mathcal{T}_H), \tag{8}$$

with the nonlinear macro map B^H given by

$$B^H(v^H; w^H) = \sum_{K \in \mathcal{T}_H} \frac{|K|}{|K_\delta|} \int_{K_\delta} \mathcal{A}^\varepsilon(x, \nabla v_K^h) dx \cdot \nabla w^H(x_K), \qquad v^H, w^H \in S_0^1(\Omega, \mathcal{T}_H), \tag{9}$$

where v_K^h solve the constrained micro problems (11) on the sampling domains K_δ.

Micro solver. Each sampling domain K_δ, associated to a macro element $K \in \mathcal{T}_H$, is discretized by micro meshes $\mathcal{T}_h$ consisting of simplicial elements $T \in \mathcal{T}_h$. The micro mesh size h is defined by $h = \max_{T \in \mathcal{T}_h} \operatorname{diam} T$ and we consider the micro finite element space

$$S^1(K_\delta, \mathcal{T}_h) = \{v^h \in W(K_\delta) \,|\, v^h|_T \in \mathcal{P}^1(T), \forall\, T \in \mathcal{T}_h\}, \tag{10}$$

where $\mathcal{P}^1(T)$ is the space of linear polynomials on $T \in \mathcal{T}_h$ and $W(K_\delta) \subset H^1(K_\delta)$ is some Sobolev space. We note that the choice of the space $W(K_\delta)$ determines the coupling condition between the macro and micro finite element methods. We consider

- periodic coupling: $W(K_\delta) = W_{per}^1(K_\delta) = \{v \in H_{per}^1(K_\delta) \,|\, \int_{K_\delta} v\, dx = 0\}$;
- Dirichlet coupling: $W(K_\delta) = H_0^1(K_\delta)$.

Let $v^H \in S_0^1(\Omega, \mathcal{T}_H)$ and K_δ be a sampling domain, we define the micro function v_K^h by micro problem: find $v_K^h - v^H \in S^1(K_\delta, \mathcal{T}_h)$ such that

$$\int_{K_\delta} \mathcal{A}^\varepsilon(x, \nabla v_K^h) \cdot \nabla z^h dx = 0, \qquad \forall\, z^h \in S^1(K_\delta, \mathcal{T}_h). \tag{11}$$

We note that v_K^h is the finite element solution to an elliptic nonlinear monotone PDE.

3.2 A useful reformulation of the FE-HMM

First, in what follows, we write the difference quotient with respect to time, like in (8), as $\bar\partial_t v_n = \Delta t^{-1}(v_{n+1} - v_n)$, for a sequence $\{v_n\}_{n\ge 0} \subset L^2(\Omega)$ and $n \ge 0$.

For the analysis of the FE-HMM it is convenient to reformulate the nonlinear map B^H as a standard finite element method applied to a modified macro problem. Let $\xi \in \mathbb{R}^d$ and $K \in \mathcal{T}_H$, we introduce the function $\chi_K^{\xi,h}$ as the solution to the variational problem: find $\chi_K^{\xi,h} \in S^1(K_\delta, \mathcal{T}_h)$ such that

$$\int_{K_\delta} \mathcal{A}^\varepsilon(x, \xi + \nabla \chi_K^{\xi,h}) \cdot \nabla z^h\, dx = 0, \qquad \forall\, z^h \in S^1(K_\delta, \mathcal{T}_h). \tag{12}$$

Similarly, we define $\bar\chi_K^\xi$ by the variational problem: find $\bar\chi_K^\xi \in W(K_\delta)$ such that

$$\int_{K_\delta} \mathcal{A}^\varepsilon(x, \xi + \nabla \bar\chi_K^\xi) \cdot \nabla z\, dx = 0, \qquad \forall\, z \in W(K_\delta). \tag{13}$$

Then, based on the functions $\chi_K^{\xi,h}$ and $\bar\chi_K^\xi$ we define the maps

$$\mathcal{A}_K^{0,h}(\xi) = \frac{1}{|K_\delta|} \int_{K_\delta} \mathcal{A}^\varepsilon(x, \xi + \nabla \chi_K^{\xi,h}) dx, \qquad \bar{\mathcal{A}}_K^0(\xi) = \frac{1}{|K_\delta|} \int_{K_\delta} \mathcal{A}^\varepsilon(x, \xi + \nabla \bar\chi_K^\xi) dx, \tag{14}$$

and the nonlinear map B^H given in (9) can be reformulated using $\mathcal{A}_K^{0,h}(\xi)$

$$B^H(v^H; w^H) = \sum_{K \in \mathcal{T}_H} |K|\, \mathcal{A}_K^{0,h}(\nabla v^H(x_K)) \cdot \nabla w^H(x_K), \qquad v^H, w^H \in S_0^1(\Omega, \mathcal{T}_H).$$

Thus, the modified macro form B^H is obtained by replacing elementwisely the exact effective map $\mathcal{A}^0(x,\xi)$ from the homogenized equation (5) by the approximation $\mathcal{A}_K^{0,h}(\xi)$.

Further, using the effective map $\mathcal{A}^0$ we introduce the map $B^0\colon H_0^1(\Omega)\times H_0^1(\Omega)\to\mathbb{R}$ by

$$B^0(v;w)=\int_\Omega \mathcal{A}^0(x,\nabla v(x))\cdot\nabla w(x)dx,\qquad v,w\in H_0^1(\Omega), \tag{15}$$

and, if $\mathcal{A}^0(\cdot,\xi)\in H^2(\Omega)$ for every $\xi\in\mathbb{R}^d$, we define the nonlinear map $\hat{B}^0$ as its discrete counterpart

$$\hat{B}^0(v^H;w^H)=\sum_{K\in\mathcal{T}_H}|K|\mathcal{A}^0(x_K,\nabla v^H(x_K))\cdot\nabla w^H(x_K),\qquad v^H,w^H\in S_0^1(\Omega,\mathcal{T}_H). \tag{16}$$

3.3 Existence and uniqueness of the numerical solution

First, we analyze the existence and uniqueness of a solution to the micro problem (11).

Lemma 3.1. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$. Let $K\in\mathcal{T}_H$, $v^H\in S_0^1(\Omega,\mathcal{T}_H)$ and $\xi\in\mathbb{R}^d$. For both coupling conditions, i.e., either $W(K_\delta)=H_0^1(K_\delta)$ or $W(K_\delta)=W_{per}^1(K_\delta)$, there exists a unique solution $v_K^h-v^H\in S^1(K_\delta,\mathcal{T}_h)$, $\chi_K^{\xi,h}\in S^1(K_\delta,\mathcal{T}_h)$ and $\bar{\chi}_K^\xi\in W(K_\delta)$ to the micro problems (11), (12) and (13), respectively.*

Proof. We prove the result for the micro problem (11). Consider the map a_K^ξ given by

$$a_K^\xi(z^h;w^h)=\int_{K_\delta}\mathcal{A}^\varepsilon(x,\xi+\nabla z^h)\cdot\nabla w^h\,dx,\qquad z^h,w^h\in S^1(K_\delta,\mathcal{T}_h), \tag{17}$$

which is nonlinear in z^h and linear w^h. As $\mathcal{A}_\xi^\varepsilon(x,\eta)=\mathcal{A}^\varepsilon(x,\xi+\eta)$ satisfies again $(\mathcal{A}_{0-2})$ the map $a_K^\xi(z^h;w^h)$ is continuous in w^h as well as Lipschitz continuous and strongly monotone in z^h. Further, taking $\xi=\nabla v^H(x_K)$, the micro problem (11) can be written as

$$\text{find } v_K^h-v^H\in S^1(K_\delta,\mathcal{T}_h)\text{ such that}\qquad a_K^\xi(v_K^h-v^H;w^h)=0,\qquad \forall w^h\in S^1(K_\delta,\mathcal{T}_h).$$

Thus, the existence and uniqueness of $v_K^h-v^H\in S^1(K_\delta,\mathcal{T}_h)$ follow from [43, Theorem 25.B]. The results for the micro problems (12) and (13) are proved analogously. □

For the analysis of the macro-micro coupling, the following energy equivalence is essential.

Lemma 3.2. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$. Let K_δ be the sampling domain associated to a macro element $K\in\mathcal{T}_H$ and v_K^h be the solution of the micro problem (11) constrained by $v^H\in S_0^1(\Omega,\mathcal{T}_H)$. Then,*

$$\big\|\nabla v^H\big\|_{L^2(K_\delta)}\le\big\|\nabla v_K^h\big\|_{L^2(K_\delta)}\le\frac{L}{\lambda}\big(\sqrt{|K_\delta|}L_0+\big\|\nabla v^H\big\|_{L^2(K_\delta)}\big),$$

where L_0 is defined in Lemma A.1.

Proof. The proof of the first inequality follows the proof of [2, Lemma 3] in the linear case. For the second inequality, we recall that $\int_{K_\delta}\mathcal{A}^\varepsilon(x,\nabla v_K^h)\cdot\nabla q^h dx=0$ for all $q^h\in S^1(K_\delta,\mathcal{T}_h)$. Combining that with the monotonicity $(\mathcal{A}_2)$ yields

$$\begin{aligned}\lambda\big\|\nabla v_K^h-\nabla v^H\big\|_{L^2(K_\delta)}^2&\le\int_{K_\delta}\big[\mathcal{A}^\varepsilon(x,\nabla v_K^h)-\mathcal{A}^\varepsilon(x,\nabla v^H(x_K))\big]\cdot\big(\nabla v_K^h-\nabla v^H(x_K)\big)dx\\&=-\int_{K_\delta}\mathcal{A}^\varepsilon(x,\nabla v^H(x_K))\cdot\big(\nabla v_K^h-\nabla v^H\big)dx\\&\le L\Big(\sqrt{|K_\delta|}L_0+\big\|\nabla v^H\big\|_{L^2(K_\delta)}\Big)\big\|\nabla v_K^h-\nabla v^H\big\|_{L^2(K_\delta)},\end{aligned}$$

where the growth estimate (3) is used. □

Using Lemma 3.2 we prove several properties of the map B^H from (9).

Lemma 3.3. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$. Let the nonlinear map B^H on $S_0^1(\Omega,\mathcal{T}_H)\times S_0^1(\Omega,\mathcal{T}_H)$ be given by (9), then B^H satisfies the bound*

$$\left|B^H(v^H;w^H)\right| \le C_b(L_0+\left\|\nabla v^H\right\|_{L^2(\Omega)})\left\|\nabla w^H\right\|_{L^2(\Omega)},$$

where C_b depends on C_0, λ, L and the measure of Ω. Further, B^H is Lipschitz continuous in its first argument and strongly monotone

$$\begin{aligned}\left|B^H(v^H;w^H)-B^H(z^H;w^H)\right| &\le \frac{L^2}{\lambda}\left\|\nabla v^H-\nabla z^H\right\|_{L^2(\Omega)}\left\|\nabla w^H\right\|_{L^2(\Omega)},\\ B^H(v^H;v^H-w^H)-B^H(w^H;v^H-w^H) &\ge \lambda\left\|\nabla v^H-\nabla w^H\right\|^2_{L^2(\Omega)},\end{aligned}$$

for all $v^H,w^H,z^H\in S_0^1(\Omega,\mathcal{T}_H)$ and where L and λ are given in $(\mathcal{A}_1)$ and $(\mathcal{A}_2)$, respectively.

Proof. Let $v^H,w^H,z^H\in S_0^1(\Omega,\mathcal{T}_H)$. The first estimate is obtained by combining the estimate (3) and the second inequality of Lemma 3.2. For the Lipschitz continuity we observe that using $(\mathcal{A}_2)$, the definition of the micro problems (11) and $(\mathcal{A}_1)$ yields

$$\begin{aligned}\lambda\left\|\nabla v_K^h-\nabla z_K^h\right\|^2_{L^2(K_\delta)} &\le \int_{K_\delta}\left[\mathcal{A}^\varepsilon(x,\nabla v_K^h)-\mathcal{A}^\varepsilon(x,\nabla z_K^h)\right]\cdot\left(\nabla v_K^h-\nabla z_K^h\right)dx\\ &= \int_{K_\delta}\left[\mathcal{A}^\varepsilon(x,\nabla v_K^h)-\mathcal{A}^\varepsilon(x,\nabla z_K^h)\right]\cdot\left(\nabla v^H(x_K)-\nabla z^H(x_K)\right)dx\\ &\le L\sqrt{|K_\delta|}\left\|\nabla v_K^h-\nabla z_K^h\right\|_{L^2(K_\delta)}\left|\nabla v^H(x_K)-\nabla z^H(x_K)\right|.\end{aligned}$$

This estimate at hand, the Lipschitz continuity of B^H follows directly. Finally, the strong monotonicity of B^H is a consequence of the definition of the micro problems (11), the monotonicity $(\mathcal{A}_2)$ and the first inequality of Lemma 3.2. □

Then, the existence and uniqueness of the numerical solution obtained by the multiscale method (8) follows from the nonlinear Lax-Milgram theorem.

Lemma 3.4. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$. Let $z^H\in S_0^1(\Omega,\mathcal{T}_H)$, $\Delta t>0$ and $f\in L^2(\Omega)$ be given. Then, there exists a unique $u^H\in S_0^1(\Omega,\mathcal{T}_H)$ such that*

$$\int_\Omega \frac{u^H-z^H}{\Delta t}w^H dx + B^H(u^H;w^H) = \int_\Omega f\,w^H dx \qquad \forall\, w^H\in S_0^1(\Omega,\mathcal{T}_H), \tag{18}$$

where B^H is given by (9).

Proof. For fixed $z^H\in S_0^1(\Omega,\mathcal{T}_H)$, consider the maps $a^{H,\Delta t}$ and $l^{\Delta t}_{z^H}$ given by

$$a^{H,\Delta t}(v^H;w^H)=\frac{1}{\Delta t}\int_\Omega v^H w^H dx + B^H(v^H;w^H), \qquad l^{\Delta t}_{z^H}(w^H)=\int_\Omega\left(f+\frac{1}{\Delta t}z^H\right)w^H dx,$$

for $v^H,w^H\in S_0^1(\Omega,\mathcal{T}_H)$. Then, the problem (18) can be written as

$$\text{find } u^H\in S_0^1(\Omega,\mathcal{T}_H) \text{ such that} \qquad a^{H,\Delta t}(u^H;w^H)=l_{z^H}(w^H), \qquad \forall\, w^H\in S_0^1(\Omega,\mathcal{T}_H).$$

The Lipschitz continuity and strong monotonicity of $a^{H,\Delta t}(v^H;w^H)$ in v^H is obtained using Lemma 3.3

$$\begin{aligned}\left|a^{H,\Delta t}(v_1^H;w^H)-a^{H,\Delta t}(v_2^H;w^H)\right| &\le C\left(\frac{1}{\Delta t}+1\right)\left\|\nabla v_1^H-\nabla v_2^H\right\|_{L^2(\Omega)}\left\|\nabla w^H\right\|_{L^2(\Omega)},\\ a^{H,\Delta t}(v^H;v^H-w^H)-a^{H,\Delta t}(w^H;v^H-w^H) &\ge \lambda\left\|\nabla v^H-\nabla w^H\right\|^2_{L^2(\Omega)},\end{aligned}$$

where $v_1^H,v_2^H,v^H,w^H\in S_0^1(\Omega,\mathcal{T}_H)$. Further, $a^{H,\Delta t}(v^H;w^H)$ is linear and continuous in its second argument w^H due to the first result of Lemma 3.3. Combining that with the linearity and continuity of $l_{z^H}(\cdot)$ and [43, Theorem 25.B] concludes the proof. □

Finally, the boundedness of the numerical approximations obtained by (9) is proved.

Theorem 3.5. *Assume that $(\mathcal{A}_{0-2})$ hold and that $f \in L^2(\Omega)$, $u_0^H \in S_0^1(\Omega, \mathcal{T}_H)$ are given. Then, for periodic and Dirichlet coupling and any parameter $\Delta t, H, h, \delta > 0$, there exists a unique numerical solution defined by the multiscale method (8). Further, the numerical solution $\{u_n^H\}_{n=1}^N$ satisfies the bound*

$$\max_{1\le n\le N} \|u_n^H\|_{L^2(\Omega)} + \lambda\left(\sum_{n=1}^N \Delta t \|\nabla u_n^H\|_{L^2(\Omega)}^2\right)^{1/2} \le C\Big(L_0 + \|f\|_{L^2(\Omega)} + \|u_0^H\|_{L^2(\Omega)}\Big),$$

where C depends on C_0, λ, L, T, the measure of Ω and the Poincaré constant C_p on Ω.

Proof. The existence and uniqueness of the numerical solution defined by the method (8) is a consequence of Lemma 3.4. To derive the a priori bound we set $w^H = u_{n+1}^H$ in (8) and use the monotonicity bound of B_H from Lemma 3.3 to obtain

$$\begin{aligned}\int_\Omega \bar\partial_t u_n^H \, u_{n+1}^H dx + \lambda\|\nabla u_{n+1}^H\|_{L^2(\Omega)}^2 &\le \int_\Omega f\, u_{n+1}^H dx - B^H(0; u_{n+1}^H)\\ &\le \frac{1}{2\lambda}(C_p\|f\|_{L^2(\Omega)} + C_b L_0)^2 + \frac{\lambda}{2}\|\nabla u_{n+1}^H\|_{L^2(\Omega)}^2,\end{aligned}$$

where C_p is the Poincaré constant on Ω and C_b is the constant from Lemma 3.3. Next, we observe that $1/2\,\bar\partial_t\|u_n^H\|_{L^2(\Omega)}^2 \le \int_\Omega \bar\partial_t u_n^H \, u_{n+1}^H dx$ yielding

$$\|u_{n+1}^H\|_{L^2(\Omega)}^2 - \|u_n^H\|_{L^2(\Omega)}^2 + \lambda\Delta t\|\nabla u_{n+1}^H\|_{L^2(\Omega)}^2 \le \frac{2}{\lambda}\Delta t(C_p^2\|f\|_{L^2(\Omega)}^2 + C_b^2 L_0^2),$$

for any $0 \le n \le N-1$. Summing the last inequality from $n = 0$ to $n = N-1$ concludes the proof. □

4 Main results

In this section we present fully discrete a priori error estimates for the difference between the numerical solution u_n^H defined by the multiscale strategy (8) and the exact homogenized solution $u^0(x,t)$ solving the homogenized problem (5). In particular, we provide error bounds in the $L^2(0,T;H_0^1(\Omega))$ and the $\mathcal{C}^0([0,T],L^2(\Omega))$ norm. In Theorem 4.1, we provide sharp estimates for the temporal and spatial macro error. Then, in Theorem 4.2 and Theorem 4.3, we provide explicit bounds for the upscaling error consisting of the micro and modeling error. We emphasize that the estimates of temporal and spatial (macro and micro) errors are valid without any structural assumptions about the heterogeneities of $\mathcal{A}^\varepsilon$. In contrast, explicit bounds of the modeling error are derived for locally periodic data $\mathcal{A}^\varepsilon$. The multiscale method (8), being defined for general maps $\mathcal{A}^\varepsilon$, is however reasonable if $\mathcal{A}^\varepsilon$ exhibits scale-separation and its good performance is known for stationary data.

4.1 Estimates for temporal and spatial macro error

Focusing first on the error due to time discretization and the macro finite element method, we quantify the upscaling error committed in the multiscale method (8) using the error functional r_{HMM} given by

$$r_{HMM}(\nabla v^H) = \left(\textstyle\sum_{K\in\mathcal{T}_H} |K|\Big|\mathcal{A}^0(x_K, \nabla v^H(x_K)) - \mathcal{A}_K^{0,h}(\nabla v^H(x_K))\Big|^2\right)^{1/2}, \tag{19}$$

for $v^H \in S_0^1(\Omega, \mathcal{T}_H)$. Then, we obtain the following result.

Theorem 4.1. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$. Let u^0 be the solution to the homogenized problem (5) and u_n^H the approximations obtained by the multiscale method (8). Provided that for $\mu = 1$*

$$u^0, \partial_t u^0 \in \mathcal{C}^0([0,T], H^2(\Omega)), \quad \partial_t^2 u^0 \in \mathcal{C}^0([0,T], L^2(\Omega)), \tag{20a}$$

$$\mathcal{A}^0(\cdot,\xi) \in W^{\mu,\infty}(\Omega;\mathbb{R}^d) \quad \textit{with} \quad \|\mathcal{A}^0(\cdot,\xi)\|_{W^{\mu,\infty}(\Omega;\mathbb{R}^d)} \le C(L_0 + |\xi|), \qquad \forall\, \xi \in \mathbb{R}^d, \tag{20b}$$

the following discrete $\mathcal{C}^0(L^2)$ and $L^2(H^1)$ error estimates hold

$$\begin{aligned}\max_{1\le n\le N} \|u^0(\cdot,t_n) - u_n^H\|_{L^2(\Omega)} + \left(\Delta t\sum_{n=1}^N \|\nabla u^0(\cdot,t_n) - \nabla u_n^H\|_{L^2(\Omega)}^2\right)^{1/2}\\ \le C\left[\Delta t + H + \max_{1\le n\le N} r_{HMM}(\nabla \mathcal{I}_H u^0(\cdot,t_n)) + \|g - u_0^H\|_{L^2(\Omega)}\right],\end{aligned}$$

where $\mathcal{I}_H u^0$ denotes the nodal interpolant of u^0 and C is independent of $\Delta t, H$ and r_{HMM}.

If additionally we assume that (20b) *is satisfied for $\mu = 2$ and*

$$u^0 \in \mathcal{C}^0([0,T], W^{2,\infty}(\Omega)), \qquad \mathcal{A}^0(x,\cdot) \in W^{2,\infty}(\mathbb{R}^d;\mathbb{R}^d), \quad a.e.\ x \in \Omega, \tag{21a}$$

$$\mathscr{A}^0_{ij}, \partial_t \mathscr{A}^0_{ij} \in \mathcal{C}^0([0,T], W^{1,\infty}(\Omega)), \qquad 1 \le i,j, \le d, \tag{21b}$$

$$\text{quasi-uniformity of macro meshes } \mathcal{T}_H \text{ and the elliptic regularity } (36), \tag{21c}$$

where $\mathscr{A}^0(x,t) = D_\xi \mathcal{A}^0(x, \nabla u^0(x,t))$, then, there exists an $H_0 > 0$ such that for all $H < H_0$, we get the improved error estimate in the discrete $\mathcal{C}^0(L^2)$ norm

$$\max_{1\le n\le N} \left\| u^0(\cdot,t_n) - u_n^H \right\|_{L^2(\Omega)} \le C \left[\Delta t + H^2 + \max_{1\le n\le N} r_{HMM}(\nabla \tilde{u}^{H,0}(\cdot,t_n)) + \left\| g - u_0^H \right\|_{L^2(\Omega)} \right],$$

where C is independent of $\Delta t, H$ and r_{HMM}.

Let us comment on the hypotheses of Theorem 4.1 in view of results for linear parabolic problems, see [35] and [8] for single scale and multiscale problems, respectively. We recall that the homogenized map $\mathcal{A}^0$ would be given by $\mathcal{A}^0(x,\xi) = a^0(x)\xi$ with $a^0(x) \in \mathbb{R}^{d\times d}$ if problem (2) is linear.

The temporal regularity in (20a) is required to obtain first order global convergence of the implicit Euler scheme. Assumption (20b) allows to estimate the error due to the quadrature formula (7) and reduces to $a^0_{ij} \in W^{\mu,\infty}(\Omega)$ for linear problems, which is likewisely assumed in [35, Theorem 2]. Further, the hypotheses (21) are solely used to show the optimal convergence of the spatial macro error. Condition (21a) is used in combination with (21c) to obtain error estimates in the $W^{1,\infty}$ norm for the elliptic projection (27) and to estimate the Taylor remainder term for the map $\mathcal{A}^0(x,\xi)$ (with respect to ξ). Finally, assumptions (21b) are needed to obtain optimal estimates of $u^0 - \tilde{u}^{H,0}$ and $\partial_t(u^0 - \tilde{u}^{H,0})$ in the L^2 norm, where $\tilde{u}^{H,0}$ is the elliptic projection (27). For linear parabolic problems (with time-dependent data) where $\mathscr{A}^0(x,t) = a^0(x,t)$, assumptions (21b) are comparable to the conditions used in [35, 8].

4.2 Fully discrete space-time a priori error estimates

To derive fully discrete error estimates we decompose the HMM upscaling error r_{HMM} introduced in (19) into micro and modeling error r_{mic} and r_{mod}, respectively, e.g., as in [3], given by

$$r_{mic}(\nabla v^H) = \left(\sum_{K\in\mathcal{T}_H} |K| \left| \bar{\mathcal{A}}^0_K(\nabla v^H(x_K)) - \mathcal{A}^{0,h}_K(\nabla v^H(x_K)) \right|^2 \right)^{1/2}, \tag{22}$$

$$r_{mod}(\nabla v^H) = \left(\sum_{K\in\mathcal{T}_H} |K| \left| \mathcal{A}^0(x_K, \nabla v^H(x_K)) - \bar{\mathcal{A}}^0_K(\nabla v^H(x_K)) \right|^2 \right)^{1/2}, \tag{23}$$

where $v^H \in S^1_0(\Omega,\mathcal{T}_H)$, $\mathcal{A}^0$ is the exact homogenized map and $\bar{\mathcal{A}}^0_K$ and $\mathcal{A}^{0,h}_K$ are given in (14). Then, we have that $r_{HMM}(\nabla v^H) \le r_{mic}(\nabla v^H) + r_{mod}(\nabla v^H)$ for any $v^H \in S^1_0(\Omega,\mathcal{T}_H)$. In particular, r_{mic} accounts for the finite element error committed during micro simulations and r_{mod} quantifies the quality of the micro sampling, i.e.,the influence of the size of the sampling domains K_δ or the boundary conditions in micro problems (11). We recall that estimates for the micro error r_{mic} are derived for general micro structures, while the modeling error r_{mod} is only analyzed for locally periodic maps $\mathcal{A}^\varepsilon$.

First, let us assume that $\bar{\chi}^\xi_K$, the exact solutions to the micro problems (13), satisfy

(H1) $\quad \bar{\chi}^\xi_K \in H^2(K_\delta)$ and $\left|\bar{\chi}^\xi_K\right|_{H^2(K_\delta)} \le C\varepsilon^{-1}(L_0 + |\xi|)\sqrt{|K_\delta|}$,

for $K \in \mathcal{T}_H$, $\xi \in \mathbb{R}^d$. Similar assumptions are used for linear multiscale problems, see [3, Remark 4].

As seen in [19, 9] for non-symmetric linear problems, adjoint micro problems are necessary to derive sharp bounds for the micro error. We introduce a similar adjoint micro problem (41), denote its corresponding solutions by $\bar{X}^{\xi,j}_K$ and assume that

(H1*) $\quad \begin{cases} (i) \quad \bar{X}^{\xi,j}_K \in H^2(K_\delta) \quad \text{and} \quad \left|\bar{X}^{\xi,j}_K\right|_{H^2(K_\delta)} \le C\varepsilon^{-1}\sqrt{|K_\delta|}, \\ (ii) \quad \bar{X}^{\xi,j}_K \in W^{1,\infty}(K_\delta) \quad \text{and} \quad \left|\bar{X}^{\xi,j}_K\right|_{W^{1,\infty}(K_\delta)} \le C, \end{cases}$

for $\xi \in \mathbb{R}^d, 1 \le j \le d$ and $K \in \mathcal{T}_H$. We note that the adjoint cell problem (41) is a linear elliptic problem. Thus, the first hypothesis in (**H1***) follows from classical H^2 regularity results if the data is sufficiently smooth, e.g., see [3, Remark 4].

Theorem 4.2. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$. Let u^0 be the solution to the homogenized problem (5) and u_n^H the approximations obtained by the multiscale method (8). Assume hypotheses (20) and (**H1**) and in addition for $\mu = 2$ that (21) holds. Further, let the multiscale method (8) be initialized with u_0^H such that $\|g - u_0^H\|_{L^2(\Omega)} \leq CH^\mu$. Then we have*

$$\max_{1\leq n\leq N} \left\|u^0(\cdot,t_n) - u_n^H\right\|_{L^2(\Omega)} \leq C\left[\Delta t + H^\mu + \left(\frac{h}{\varepsilon}\right)^\nu + \max_{1\leq n\leq N} r_{mod}(\nabla \mathcal{U}_n^H)\right],$$
$$\left(\Delta t \sum_{n=1}^N \left\|\nabla u^0(\cdot,t_n) - \nabla u_n^H\right\|_{L^2(\Omega)}\right)^{1/2} \leq C\left[\Delta t + H + \left(\frac{h}{\varepsilon}\right)^\nu + \max_{1\leq n\leq N} r_{mod}(\nabla \mathcal{U}_n^H)\right],$$

for $\nu = 1$, where $\mathcal{U}_n^H = \mathcal{I}_H u^0(\cdot,t_n)$ if $\mu = 1$ or $\mathcal{U}_n^H = \tilde{u}^{H,0}(\cdot,t_n)$ and $H < H_0$ (with H_0 from Theorem 4.1) if $\mu = 2$. The constant C is independent of $\Delta t, H, h, \varepsilon, \delta$ and the modeling error r_{mod}.

*If in addition (**H1***) holds and $\mathcal{A}^\varepsilon(x,\cdot) \in W^{2,\infty}(\mathbb{R}^d;\mathbb{R}^d)$ for a.e. $x \in \Omega$, then the above estimates hold for $\nu = 2$.*

As the quadratic micro convergence $(h/\varepsilon)^2$ is optimal for linear homogenization problems, e.g., see [1], Theorem 4.2 provides sharp micro error estimates for sufficiently smooth multiscale problems (2).

Finally, we present explicit estimates for the modeling error r_{mod} supposing that the spatial heterogeneities of the maps $\mathcal{A}^\varepsilon$ are locally periodic and Lipschitz continuous with respect to the macroscopic variable, i.e.,

(H2) the maps $\mathcal{A}^\varepsilon$ are locally periodic, i.e., $\mathcal{A}^\varepsilon(x,\xi) = \mathcal{A}^\varepsilon(x,x/\varepsilon,\xi) = \mathcal{A}(x,y,\xi)$ with $\mathcal{A}(x,y,\xi)$ being Y-periodic in y and satisfying (for $\xi \in \mathbb{R}^d$, a.e. $y \in Y$)

$$|\mathcal{A}(x_1,y,\xi) - \mathcal{A}(x_2,y,\xi)| \leq C|x_1 - x_2|(L_0 + |\xi|), \qquad \forall\, x_1, x_2 \in \Omega.$$

We note, that the collocation $\mathcal{A}^\varepsilon(x_K,x/\varepsilon,\xi)$ at the quadrature nodes x_K is advantageous if the decomposition $\mathcal{A}^\varepsilon(x,\xi) = \mathcal{A}^\varepsilon(x,x/\varepsilon,\xi)$ is explicitely known.

Theorem 4.3. *Assume that the map $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$ and (**H2**). Then, for any $v^H \in S_0^1(\Omega,\mathcal{T}_H)$, the modeling error $r_{mod}(\nabla v^H)$ defined in (23) is bounded by*

$$r_{mod}(\nabla v^H) \leq \begin{cases} 0, & \text{if } W(K_\delta) = W^1_{per}(K_\delta), \delta/\varepsilon \in \mathbb{N} \text{ and} \\ & \mathcal{A}^\varepsilon = \mathcal{A}(x_K,x/\varepsilon,\xi) \text{ collocated at } x_K, \\ C^1_{mod}\,\delta, & \text{if } W(K_\delta) = W^1_{per}(K_\delta), \delta/\varepsilon \in \mathbb{N}, \\ C^2_{mod}(\delta + \sqrt{\varepsilon/\delta}), & \text{if } W(K_\delta) = H^1_0(K_\delta), \delta > \varepsilon, \end{cases}$$

with C^1_{mod} and C^2_{mod} given by

$$C^1_{mod} = C(L_0 + \|\nabla v^H\|_{L^2(\Omega)}), \quad C^2_{mod} = C(C^1_{mod} + \max_{K\in\mathcal{T}_H} \|\chi^{\nabla v^H(x_K)}(x_K,\cdot)\|_{W^{1,\infty}(Y)}),$$

where $\chi^\xi(x_K,\cdot)$, for $\xi \in \mathbb{R}^d$, $K \in \mathcal{T}_H$, denote the exact solutions to the homogenization cell problems (44) and C is independent of $\Delta t, H, h, \varepsilon, \delta$ and v^H.

We note, that combining periodic coupling and collocation is optimal for locally periodic maps $\mathcal{A}^\varepsilon$. Further, the regularity hypothesis $\chi^\xi(x_K,\cdot) \in W^{1,\infty}(Y)$ is a common assumption to bound the resonance error (when using Dirichlet coupling) for linear homogenization problems, see [21, Theorem 1.2].

Refinement strategies. The fully discrete a priori error estimates from Theorem 4.2 reveal that simultaneous refinement of macro and micro meshes is needed for convergence of the spatial errors. For instance, consider a sufficiently smooth problem (2) with locally periodic maps $\mathcal{A}^\varepsilon$. If using periodic coupling and collocation of $\mathcal{A}^\varepsilon$ for the multiscale method (8), then we have, under the conditions from Theorem 4.2 and 4.3,

$$\begin{aligned} \max_{1\leq n\leq N} \left\|u^0(\cdot,t_n) - u_n^H\right\|_{L^2(\Omega)} &\leq C\left[\Delta t + H^2 + \left(\tfrac{h}{\varepsilon}\right)^2\right], \\ \left(\Delta t \sum_{n=1}^N \left\|\nabla u^0(\cdot,t_n) - \nabla u_n^H\right\|^2_{L^2(\Omega)}\right)^{1/2} &\leq C\left[\Delta t + H + \left(\tfrac{h}{\varepsilon}\right)^2\right], \end{aligned} \tag{24}$$

i.e., robust convergence of u_n^H towards the homogenized solution u^0. Further, an efficient decrease of the spatial errors in the $\mathcal{C}^0(L^2)$ norm is obtained when refining the spatial grids $\mathcal{T}_H$ and $\mathcal{T}_h$ according to

$h/\varepsilon \sim H$. Similarly, balancing macro and micro spatial errors in the $L^2(H^1)$ norm leads to $h/\varepsilon \sim \sqrt{H}$. Those refinement strategies allow to obtain convergence at optimal computational cost.

Complexity. We note that the numerical upscaling used in (8) leads to computational cost that are *independent* of the size of the small oscillations ε. For instance, let N_{mac} and N_{mic} denote the number of elements in each dimension for the macro and micro spatial discretization, respectively, using quasi-uniform meshes. Then, the macro and micro mesh sizes H and h scale as $H \sim 1/N_{mac}$ and $h \sim \delta/N_{mic}$, respectively. As the size δ of the sampling domains K_δ is of order $\mathcal{O}(\varepsilon)$, we find that $h/\varepsilon \sim 1/N_{mic}$. Thus, the convergence rates summarized in (24) can be expressed in terms of N_{mac} and N_{mic}, i.e., they are robust with respect to ε, and can be obtained with $\mathcal{O}(N_{mac}^d N_{mic}^d)$ spatial degrees of freedom.

5 Proof of the main results

In this section, we prove Theorem 4.1, 4.2 and 4.3. We split the total error according to $\|u^0 - u_n^H\| \leq \|u^0 - \mathcal{U}_n^H\| + \|\mathcal{U}_n^H - u_n^H\|$ where $\mathcal{U}_n^H$ is an approximation of the exact solution u^0 in $S_0^1(\Omega, \mathcal{T}_H)$. While choosing $\mathcal{U}_n^H$ as the nodal interpolant of u^0 is sufficient to obtain optimal $L^2(H^1)$ estimates, an elliptic projection of u^0 is needed for the optimal $\mathcal{C}^0(L^2)$ estimates. This is well-known for linear problems [41]. As mentioned in the introduction, we will introduce a new elliptic projection for nonlinear monotone problems for our analysis.

5.1 Preliminaries

We next introduce the nodal interpolant and the elliptic projection of the homogenized solution u^0 and derive their approximation properties.

Nodal interpolant. Let $\mathcal{I}_H \colon \mathcal{C}^0(\overline{\Omega}) \to S^1(\Omega, \mathcal{T}_H)$ be the usual nodal interpolant where $S^1(\Omega, \mathcal{T}_H)$ is the FE-space defined as $S_0^1(\Omega, \mathcal{T}_H)$ in (6), but without zero boundary conditions. Then, for $k \in \{1, 2\}$, we have the bounds, see [13, Theorem 3.1.6],

$$\|\mathcal{I}_H z\|_{H^1(\Omega)} \leq C\|z\|_{H^2(\Omega)}, \quad \|\mathcal{I}_H z - z\|_{H^{2-k}(\Omega)} \leq CH^k\|z\|_{H^2(\Omega)}, \quad \forall\, z \in H^2(\Omega), \tag{25}$$

$$\|\mathcal{I}_H z\|_{W^{1,\infty}(\Omega)} \leq C\|z\|_{W^{1,\infty}(\Omega)}, \qquad \forall\, z \in W^{1,\infty}(\Omega). \tag{26}$$

We note that for $z \in \mathcal{C}^0(\overline{\Omega}) \cap H_0^1(\Omega)$ it holds that $\mathcal{I}_H z \in S_0^1(\Omega, \mathcal{T}_H)$.

Remark 5.1. If $u^0, \partial_t u^0 \in \mathcal{C}^0([0,T], H^2(\Omega))$, then the interpolation operator $\mathcal{I}_H$ and the differentiation ∂_t with respect to the time variable can be interchanged, i.e., $\mathcal{I}_H(\partial_t u^0(x,t)) = \partial_t(\mathcal{I}_H u^0(x,t))$ on $\Omega \times [0,T]$. Thus the L^2 error estimate $\|\partial_t u^0(\cdot,t) - \partial_t(\mathcal{I}_H u^0(\cdot,t))\|_{L^2(\Omega)} \leq CH^2\|\partial_t u^0(\cdot,t)\|_{H^2(\Omega)}$ holds.

Elliptic projection. Let $u^0(x,t) \in E$ be the exact solution of the homogenized problem (5). The elliptic projection $\tilde{u}^{H,0}(\cdot,t)$ of $u^0(\cdot,t)$ is given by the variational problem: find $\tilde{u}^{H,0}(\cdot,t) \in S_0^1(\Omega, \mathcal{T}_H)$ such that

$$B_\pi(t; \tilde{u}^{H,0}(\cdot,t), w^H) = B_\pi(t; u^0(\cdot,t), w^H), \qquad \forall\, w^H \in S_0^1(\Omega, \mathcal{T}_H), \tag{27}$$

where, for a.e. $t \in (0,T)$, the bilinear form B_π is defined as, for $v, w \in H_0^1(\Omega)$,

$$B_\pi(t; v, w) = \int_\Omega \mathscr{A}^0(x,t)\nabla v \cdot \nabla w \, dx, \quad \text{with } \mathscr{A}^0(x,t) = D_\xi \mathcal{A}^0(x, \nabla u^0(x,t)), \tag{28}$$

for a.e. $(x,t) \in \Omega \times (0,T)$. The existence and uniqueness of the elliptic projection $\tilde{u}^{H,0}(\cdot,t)$ defined in (27) is studied in Lemma 5.2. Note that for a linear problem we recover the tensor $\mathscr{A}^0(x,t) = a^0(x,t)$ and $\partial_t \mathscr{A}^0(x,t) = \partial_t a^0(x,t)$ as considered in [41, 35, 8].

Lemma 5.2. *Let $\mathcal{A}^0$ satisfy $(\mathcal{A}_{1-2})$ and $\mathcal{A}^0(x,\cdot) \in \mathcal{C}^1(\mathbb{R}^d; \mathbb{R}^d)$ for a.e. $x \in \Omega$. If the homogenized solution u^0 satisfies $u^0 \in L^2(0,T; H_0^1(\Omega))$, then, for a.e. $t \in (0,T)$, the bilinear form B_π given by (28) is uniformly elliptic and bounded and there exists a unique solution $\tilde{u}^{H,0}(\cdot,t)$ to (27). Further, we have*

$$\big\|\nabla \tilde{u}^{H,0}(\cdot,t)\big\|_{L^2(\Omega)} \leq \frac{L}{\lambda}\big\|\nabla u^0(\cdot,t)\big\|_{L^2(\Omega)}, \qquad \text{a.e. } t \in (0,T). \tag{29}$$

Proof. First, due to Lemma A.1 and the regularity of u^0 the tensor $\mathscr{A}^0$ satisfies $\mathscr{A}_{ij}^0 \in L^\infty(0,T; L^\infty(\Omega))$, for $1 \leq i,j \leq d$. Then, using again Lemma A.1 we have that $|B_\pi(t; v, w)| \leq L\|\nabla v\|_{L^2(\Omega)}\|\nabla w\|_{L^2(\Omega)}$ and $\lambda\|\nabla v\|_{L^2(\Omega)}^2 \leq B_\pi(t; v, v)$ for $v, w \in H_0^1(\Omega)$ and a.e. $t \in (0,T)$. Thus, applying the Lax-Milgram theorem concludes the proof. □

Lemma 5.3. *Assume that $\mathcal{A}^0$ satisfies $(\mathcal{A}_{1-2})$ and $\mathcal{A}^0(x,\cdot) \in \mathcal{C}^1(\mathbb{R}^d;\mathbb{R}^d)$ for a.e. $x \in \Omega$. Let the homogenized solution u^0 and $\mathscr{A}^0$ defined in (28) satisfy*

$$u^0, \partial_t u^0 \in \mathcal{C}^0([0,T], H_0^1(\Omega)), \qquad \mathscr{A}^0_{ij}, \partial_t \mathscr{A}^0_{ij} \in \mathcal{C}^0([0,T], L^\infty(\Omega)), \qquad \text{for } 1 \le i,j \le d. \tag{30}$$

Then, the map $t \mapsto \tilde{u}^{H,0}(\cdot,t) \in S_0^1(\Omega,\mathcal{T}_H)$, where $\tilde{u}^{H,0}$ is the elliptic projection (27), is of class $\mathcal{C}^1$.

Proof. For $t \in [0,T]$, we first introduce the auxiliary function $\hat{u}^{H,0}(\cdot,t) \in S_0^1(\Omega,\mathcal{T}_H)$ satisfying

$$B_\pi(t;\hat{u}^{H,0}(\cdot,t), w^H) = F_\pi(t;w^H), \qquad \forall\, w^H \in S_0^1(\Omega,\mathcal{T}_H), \tag{31}$$

where $B'_\pi(t;v,w) = \int_\Omega \partial_t\big(\mathscr{A}^0(x,t)\big)\nabla v \cdot \nabla w\, dx$ (for $v,w \in H_0^1(\Omega)$) and the linear map F_π is given by

$$F_\pi(t;w) = B_\pi(t;\partial_t u^0(\cdot,t), w) + B'_\pi(t;u^0(\cdot,t) - \tilde{u}^{H,0}(\cdot,t), w).$$

Due to (30) the bilinear form B'_π is bounded and problem (31) has a unique solution $\hat{u}^{H,0}(\cdot,t)$ for $t \in [0,T]$.

In what follows, we omit the space variable x for $u^0, \tilde{u}^{H,0}$ and $\hat{u}^{H,0}$. Let $t, t+\tau \in [0,T]$. Then, due to the regularity of $u^0(t)$ and $\mathscr{A}^0(x,t)$ in the time variable t, a simple calculation shows that $\|\tilde{u}^{H,0}(t) - \tilde{u}^{H,0}(t+\tau)\|_{L^2(\Omega)} \to 0$ for $\tau \to 0$, i.e., the map $t \mapsto \tilde{u}^{H,0}(\cdot,t)$ is continuous.

Consider the term $B_\pi(t;\tilde{u}^{H,0}(t+\tau) - \tilde{u}^{H,0}(t), w^H)$ for $w^H \in S_0^1(\Omega,\mathcal{T}_H)$. Using the definition of the elliptic projection (27) it holds that

$$\begin{aligned} B_\pi(t;\tilde{u}^{H,0}(t+\tau) - \tilde{u}^{H,0}(t), w^H) &= B_\pi(t;\tilde{u}^{H,0}(t+\tau), w^H) - B_\pi(t;u^0(t), w^H) \\ &\quad + \underbrace{B_\pi(t+\tau;u^0(t+\tau), w^H) - B_\pi(t+\tau;\tilde{u}^{H,0}(t+\tau), w^H)}_{=0}. \end{aligned} \tag{32}$$

Then, we divide (32) by τ and substract (31) to obtain

$$\begin{aligned} &B_\pi(t;\tau^{-1}\left[\tilde{u}^{H,0}(t+\tau) - \tilde{u}^{H,0}(t)\right] - \hat{u}^{H,0}(t), w^H) \\ &\quad = \tau^{-1}\left[B_\pi(t+\tau;u^0(t+\tau), w^H) - B_\pi(t+\tau;u^0(t), w^H)\right] - B_\pi(t;\partial_t u^0(t), w^H) \\ &\qquad + \tau^{-1}\left[B_\pi(t+\tau;u^0(t), w^H) - B_\pi(t;u^0(t), w^H)\right] - B'_\pi(t;u^0(t), w^H) \\ &\qquad - \big(\tau^{-1}\left[B_\pi(t+\tau;\tilde{u}^{H,0}(t+\tau), w^H) - B_\pi(t;\tilde{u}^{H,0}(t+\tau), w^H)\right] - B'_\pi(t;\tilde{u}^{H,0}(t), w^H)\big). \end{aligned} \tag{33}$$

We next choose $w^H = \tau^{-1}[\tilde{u}^{H,0}(t+\tau) - \tilde{u}^{H,0}(t)] - \hat{u}^{H,0}(t)$ in (33), use the ellipticity and boundedness of B_π and combine that with the regularity assumptions (30) and the boundedness of (29) of $\tilde{u}^{H,0}$ to obtain

$$\left\|\tau^{-1}\left[\tilde{u}^{H,0}(t+\tau) - \tilde{u}^{H,0}(t)\right] - \hat{u}^{H,0}(t)\right\|_{H^1(\Omega)} \to 0, \qquad \text{for } \tau \to 0.$$

Hence, $\partial_t \tilde{u}^{H,0}(t)$ exists and we have $\partial_t \tilde{u}^{H,0} = \hat{u}^{H,0}$ and $\partial_t \nabla \tilde{u}^{H,0} = \nabla \hat{u}^{H,0}$. □

Lemma 5.4. *Assume that $\mathcal{A}^0$ satisfies $(\mathcal{A}_{1-2})$ and $\mathcal{A}^0(x,\cdot) \in \mathcal{C}^1(\mathbb{R}^d;\mathbb{R}^d)$ for a.e. $x \in \Omega$. Let u^0 be the solution of the homogenized problem (5), $\tilde{u}^{H,0}$ its elliptic projection (27) and $\mathscr{A}^0$ the tensor given by (28). Let $k \in \{1,2\}$ and assume*

$$u^0, \partial_t u^0 \in \mathcal{C}^0([0,T], H^2(\Omega)), \qquad \mathscr{A}^0_{ij}, \partial_t \mathscr{A}^0_{ij} \in \mathcal{C}^0([0,T], W^{k-1,\infty}(\Omega)), \qquad \text{for } 1 \le i,j \le d.$$

Then, for any $t \in [0,T]$, we have the error estimates

$$\begin{aligned} &(i) \quad \left\|\tilde{u}^{H,0}(\cdot,t) - u^0(\cdot,t)\right\|_{H^1(\Omega)} \le CH, \qquad &&(iii) \quad \left\|\partial_t(\tilde{u}^{H,0} - u^0)(\cdot,t)\right\|_{H^1(\Omega)} \le CH, \\ &(ii) \quad \left\|\tilde{u}^{H,0}(\cdot,t) - u^0(\cdot,t)\right\|_{L^2(\Omega)} \le CH^k, &&(iv) \quad \left\|\partial_t(\tilde{u}^{H,0} - u^0)(\cdot,t)\right\|_{L^2(\Omega)} \le CH^k, \end{aligned}$$

where C is independent H.

Proof. We omit again the space variable x for u^0 as well as $\tilde{u}^{H,0}$ and let $t \in [0,T]$. First, we note that the estimates *(i)* and *(ii)* follow from standard finite element estimates.
(iii) Using the nodal interpolant $\mathcal{I}_H \partial_t u^0(t)$, we split the error into two terms

$$\left\|\partial_t(\tilde{u}^{H,0}(t) - u^0(t))\right\|_{H^1(\Omega)} \le \left\|\mathcal{I}_H \partial_t u^0(t) - \partial_t u^0(t)\right\|_{H^1(\Omega)} + \left\|\partial_t \tilde{u}^{H,0}(t) - \mathcal{I}_H \partial_t u^0(t)\right\|_{H^1(\Omega)},$$

where the first term can be estimated by (25). For the second term, we use (31) to find

$$\begin{aligned}\lambda\big\|\nabla(\partial_t\tilde{u}^{H,0}(t)-\partial_t\mathcal{I}_H u^0(t))\big\|^2_{L^2(\Omega)} &\le B_\pi(t;\partial_t(\tilde{u}^{H,0}(t)-\mathcal{I}_H u^0(t)),\partial_t(\tilde{u}^{H,0}(t)-\mathcal{I}_H u^0(t)))\\ &= B'_\pi(t;u^0(t)-\tilde{u}^{H,0}(t),\partial_t\tilde{u}^{H,0}(t)-\partial_t u^0(t))\\ &\quad + B_\pi(t;\partial_t u^0(t)-\mathcal{I}_H\partial_t u^0(t),\partial_t\tilde{u}^{H,0}(t)-\mathcal{I}_H\partial_t u^0(t)),\end{aligned}$$

from where estimate *(iii)* is derived using the boundedness of B_π, B'_π, estimate *(i)* and bound (25).

(iv) For $v=\partial_t(\tilde{u}^{H,0}(t)-u^0(t))$, we consider the dual problem

$$\text{find } \varphi_v\in H^1_0(\Omega) \text{ such that} \qquad B_\pi(t;w,\varphi_v)=\langle v,w\rangle_{L^2(\Omega)}, \qquad \forall\, w\in H^1_0(\Omega),$$

where $\langle\cdot,\cdot\rangle_{L^2(\Omega)}$ denotes the $L^2(\Omega)$ inner product. Using the nodal interpolant $\mathcal{I}_H\varphi_v$, the H^2 regularity of φ_v (due to the regularity of $\mathscr{A}^0$ and the convexity of Ω) and equation (31) yields

$$\begin{aligned}&\big\|\partial_t(\tilde{u}^{H,0}(t)-u^0(t))\big\|^2_{L^2(\Omega)}=\big\langle\partial_t(\tilde{u}^{H,0}(t)-u^0(t)),v\big\rangle_{L^2(\Omega)}=B_\pi(t;\partial_t(\tilde{u}^{H,0}(t)-u^0(t)),\varphi_v)\\ &=B_\pi(t;\partial_t(\tilde{u}^{H,0}(t)-u^0(t)),\varphi_v-\mathcal{I}_H\varphi_v)-B'_\pi(t;\tilde{u}^{H,0}(t)-u^0(t),\mathcal{I}_H\varphi_v-\varphi_v)\\ &\quad -B'_\pi(t;\tilde{u}^{H,0}(t)-u^0(t),\varphi_v).\end{aligned}\tag{34}$$

Integrating by parts the last term of (34) and using $\partial_t\mathscr{A}^0(\cdot,t)\in W^{1,\infty}(\Omega)$ leads to

$$B'_\pi(t;\tilde{u}^{H,0}(t)-u^0(t),\varphi_v)\le C\big\|\tilde{u}^{H,0}(t)-u^0(t)\big\|_{L^2(\Omega)}\|\varphi_v\|_{H^2(\Omega)}.\tag{35}$$

Finally, combining identity (34) and inequality (35) with the estimates *(i)*, *(ii)*, *(iii)* and the error (25) bound concludes the proof. □

For the proof of the optimal convergence in the discrete $\mathcal{C}^0(L^2)$ norm we need an estimate of $\tilde{u}^{H,0}(\cdot,t)-u^0(\cdot,t)$ in the $W^{1,\infty}(\Omega)$ norm. Such maximum norm estimates are provided in [12, Chapter 8].

Lemma 5.5. *Assume that $\mathcal{A}^0$ satisfies $(\mathcal{A}_{1-2})$ and $\mathcal{A}^0(x,\cdot)\in\mathcal{C}^1(\mathbb{R}^d;\mathbb{R}^d)$ for a.e. $x\in\Omega$. Let u^0 be the solution of the homogenized problem (5), $\tilde{u}^{H,0}$ its elliptic projection (27), $\mathscr{A}^0(x,t)$ be given by (28) and $u^{0,*}(\cdot,t)$ be solving the dual problem $B_\pi(t;w,u^{0,*}(\cdot,t))=B_\pi(t;u^0(\cdot,t),w)$ for all $w\in H^1_0(\Omega)$. Assume*

$$u^0\in\mathcal{C}^0([0,T],W^{2,\infty}(\Omega)),\quad \mathscr{A}^0_{ij}\in\mathcal{C}^0([0,T],W^{1,\infty}(\Omega)),\qquad 1\le i,j\le d,$$

and the "elliptic regularity", for $t\in[0,T]$ and $1<p<\sigma$ with some $\sigma>d$,

$$\big\|u^0(\cdot,t)\big\|_{W^{2,p}(\Omega)}+\big\|u^{0,*}(\cdot,t)\big\|_{W^{2,p}(\Omega)}\le C\big\|\mathrm{div}(\mathscr{A}^0(\cdot,t)\nabla u^0(\cdot,t))\big\|_{L^p(\Omega)}.\tag{36}$$

If $\{\mathcal{T}_H\}_{H>0}$ is a family of quasi-uniform meshes, e.g., see [13, Condition (3.2.28)], then there exists an $H_0>0$ such that for every $t\in[0,T]$ and $H<H_0$

$$\big\|\tilde{u}^{H,0}(\cdot,t)\big\|_{W^{1,\infty}(\Omega)}\le C\big\|u^0(\cdot,t)\big\|_{W^{1,\infty}(\Omega)},\ \ \big\|u^0(\cdot,t)-\tilde{u}^{H,0}(\cdot,t)\big\|_{W^{1,\infty}(\Omega)}\le CH\big\|u^0(\cdot,t)\big\|_{W^{2,\infty}(\Omega)},$$

where C is independent of H.

Proof. We recall that the elliptic projection $\tilde{u}^{H,0}$ is the finite element solution to a linear elliptic problem, see (27). We can thus apply the maximum norm error estimates provided by [12, Theorem 8.1.11 and Corollary 8.1.12]. □

5.2 Error propagation formula

Let $\mathcal{U}^H(\cdot,t)\in S^1_0(\Omega,\mathcal{T}_H)$ be defined for any $t\in[0,T]$ and $\mathcal{U}^H_n=\mathcal{U}^H(\cdot,t_n)$ for $0\le n\le N$. Further, assume that $u^0,\partial_t u^0\in\mathcal{C}^0([0,T],L^2(\Omega))$. The fundamental tool to derive a priori error estimates using

the energy method is the error propagation formula for the error $\theta_n^H = u_n^H - \mathcal{U}_n^H$, $0 \le n \le N$, given by

$$\begin{aligned}
&\int_\Omega \bar\partial_t \theta_n^H \, w^H \, dx + B^H(u_{n+1}^H; w^H) - B^H(\mathcal{U}_{n+1}^H; w^H) \\
&\qquad = \int_\Omega f \, w^H dx - \int_\Omega \bar\partial_t \mathcal{U}_n^H \, w^H dx - B^H(\mathcal{U}_{n+1}^H; w^H) \\
&\qquad = \int_\Omega \left[\partial_t u^0(x, t_{n+1}) - \bar\partial_t \mathcal{U}_n^H\right] \, w^H dx + B^0(u^0(\cdot, t_{n+1}); w^H) - B^H(\mathcal{U}_{n+1}^H; w^H) \\
&\qquad = \int_\Omega \left[\partial_t u^0(x, t_{n+1}) - \bar\partial_t u^0(x, t_n)\right] \, w^H dx && \text{(37a)} \\
&\qquad\quad + \int_\Omega \left[\bar\partial_t u^0(x, t_n) - \bar\partial_t \mathcal{U}_n^H\right] \, w^H dx && \text{(37b)} \\
&\qquad\quad + B^0(u^0(\cdot, t_{n+1}); w^H) - B^0(\mathcal{U}_{n+1}^H; w^H) && \text{(37c)} \\
&\qquad\quad + B^0(\mathcal{U}_{n+1}^H; w^H) - \hat B^0(\mathcal{U}_{n+1}^H; w^H) && \text{(37d)} \\
&\qquad\quad + \hat B^0(\mathcal{U}_{n+1}^H; w^H) - B^H(\mathcal{U}_{n+1}^H; w^H), && \text{(37e)}
\end{aligned}$$

where $w^H \in S_0^1(\Omega, \mathcal{T}_H)$ is arbitrary, u^0 is the exact solution to the homogenized problem (5) and the forms B^0, $\hat B^0$ and B^H are given by (15), (16) and (9), respectively.

In the error propagation formula (37) we already performed the error decomposition into different components. While the term (37a) accounts for the error due to the time discretization scheme, the terms (37b) and (37c) consists of the finite element error at the discrete time levels t_n. Further, the influence of quadrature formula (7) is captured by (37d). While the components (37a) – (37d) are independent of the multiscale nature of the method (8), i.e., temporal and macro spatial error, the last term (37e) is solely due to the upscaling strategy consisting of micro simulations and averaging techniques. Thus we call term (37e) the HMM error.

In our subsequent analysis, we first estimate in Section 5.3 the different components of the temporal and macro spatial error, for either $\mathcal{U}^H(\cdot, t) = \mathcal{I}_H u^0(\cdot, t)$ or $\mathcal{U}^H(\cdot, t) = \tilde u^{H,0}(\cdot, t)$, and secondly we derive explicit estimates for the HMM error consisting of modeling and micro error in Section 5.4.

5.3 Temporal and macro spatial error

In this section, we provide explicit error bounds for the terms (37a) – (37d).

Time discretization error. We start by estimating the error due to the discretization in time by the backward Euler method, i.e., by estimating term (37a).

Lemma 5.6. *Let u^0 the solution of the homogenized problem (5) satisfy $u^0, \partial_t u^0, \partial_t^2 u^0 \in \mathcal{C}^0([0,T], L^2(\Omega))$. Then, for $w^H \in S_0^1(\Omega, \mathcal{T}_H)$ and $0 \le n \le N-1$, we obtain for term (37a)*

$$\left| \int_\Omega \left[\partial_t u^0(x, t_{n+1}) - \bar\partial_t u^0(x, t_n)\right] w^H dx \right| \le C \Delta t \left\| \partial_t^2 u^0 \right\|_{\mathcal{C}^0([0,T], L^2(\Omega))} \left\| w^H \right\|_{L^2(\Omega)},$$

where C is independent of Δt and H.

Proof. Let $w^H \in S_0^1(\Omega, \mathcal{T}_H)$. As $u^0, \partial_t u^0 \in \mathcal{C}^0([0,T], L^2(\Omega))$ we have

$$\int_\Omega \bar\partial_t u^0(x, t_n) w^H dx = \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \int_\Omega \partial_t u^0(x, s) w^H dx \, ds. \tag{38}$$

Further, due to the regularity $\partial_t u^0, \partial_t^2 u^0 \in \mathcal{C}^0([0,T], L^2(\Omega))$, similar results hold if $\partial_t u^0$ and $\partial_t^2 u^0$ is substitute to u^0 and $\partial_t u^0$, respectively. Hence, combining that with (38) yields

$$\int_\Omega \left[\partial_t u^0(x, t_{n+1}) - \bar\partial_t u^0(x, t_n)\right] w^H dx = \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \int_s^{t_{n+1}} \int_\Omega \partial_t^2 u^0(x, \tau) w^H dx \, d\tau \, ds,$$

from where the result of Lemma 5.6 follows. □

Macro finite element error. Next, we estimate the spatial macro error terms (37b) and (37c) in Lemma 5.7 and Lemma 5.8.

Lemma 5.7. *Let u^0 be the solution of the homogenized problem* (2) *and let either $\mathcal{U}_n^H = \mathcal{I}_H u^0(\cdot, t_n)$ be its nodal interpolant or $\mathcal{U}_n^H = \tilde{u}_n^{H,0}$ its elliptic projection* (27)*, for $0 \le n \le N-1$. Assume that $u^0, \partial_t u^0 \in \mathcal{C}^0([0,T], H^2(\Omega))$ and additionally, if $\mathcal{U}_n^H = \tilde{u}_n^{H,0}$, that $\mathscr{A}^0$ given by* (28) *satisfies* (21b)*. Then,*

$$\left| \int_\Omega \left[\bar{\partial}_t u^0(x, t_n) - \bar{\partial}_t \mathcal{U}_n^H \right] w^H dx \right| \le C H^2 \left\| u^0 \right\|_{\mathcal{C}^0([0,T], H^2(\Omega))} \left\| w^H \right\|_{L^2(\Omega)},$$

for every $w^H \in S_0^1(\Omega, \mathcal{T}_H)$ with a constant C independent of Δt and H.

Proof. As $\mathcal{I}_H u^0, \partial_t \mathcal{I}_H u^0 \in \mathcal{C}^0([0,T], S_0^1(\Omega, \mathcal{T}_H))$ and $\tilde{u}^{H,0}, \partial_t \tilde{u}^{H,0} \in \mathcal{C}^0([0,T], S_0^1(\Omega, \mathcal{T}_H))$, see Remark 5.1 and Lemma 5.3, respectively, equation (38) holds analogously for u^0 substitued by $\mathcal{U}^H$. Thus, for $w^H \in S_0^1(\Omega, \mathcal{T}_H)$, we obtain

$$\left| \int_\Omega \left[\bar{\partial}_t u^0(x, t_n) - \bar{\partial}_t \mathcal{U}_n^H \right] w^H dx \right| \le \frac{1}{\Delta t} \int_{t_n}^{t_{n+1}} \left\| \partial_t u^0(x,s) - \partial_t \mathcal{U}^H(x,s) \right\|_{L^2(\Omega)} \left\| w^H \right\|_{L^2(\Omega)} ds,$$

and the estimate from Remark 5.1 and Lemma 5.4 conclude the proof. □

While in Lemma 5.7 optimal quadratic convergence H^2 is obtained for both $\mathcal{U}_n^H = \mathcal{I}_H u^0(x, t_n)$ and $\mathcal{U}_n^H = \tilde{u}_n^{H,0}$, in Lemma 5.8 the optimal convergence rate is only obtained for $\mathcal{U}_n^H = \tilde{u}_n^{H,0}$ (due to its particular definition (27)).

Lemma 5.8. *Let u^0 be the solution of the homogenized problem* (5)*, $\mathcal{I}_H u^0$ its nodal interpolant, $\tilde{u}^{H,0}$ its elliptic projection* (27) *and B^0 be given by* (15)*. Assume that $\mathcal{A}^0$ satisfies $(\mathcal{A}_1)$ and $u^0 \in \mathcal{C}^0([0,T], H^2(\Omega))$. Further, let $w^H \in S_0^1(\Omega, \mathcal{T}_H)$ and $0 \le n \le N-1$.*

(i) If $\mathcal{U}_{n+1}^H = \mathcal{I}_H u^0(\cdot, t_{n+1})$, then

$$\left| B^0(u^0(\cdot, t_{n+1}); w^H) - B^0(\mathcal{U}_{n+1}^H; w^H) \right| \le C H \left\| u^0 \right\|_{\mathcal{C}^0([0,T], H^2(\Omega))} \left\| \nabla w^H \right\|_{L^2(\Omega)},$$

where C is independent of Δt and H.

(ii) If $\mathcal{U}_{n+1}^H = \tilde{u}_{n+1}^{H,0}$, we additionally assume $(\mathcal{A}_2)$, hypotheses (21a) *for u^0 and $\mathcal{A}^0$, quasi-uniformity and elliptic regularity* (21c) *as well as regularity $\mathscr{A}_{ij}^0 \in \mathcal{C}^0([0,T], W^{1,\infty}(\Omega))$ (for $1 \le i,j \le d$) with $\mathscr{A}^0$ given in* (28)*. Then, there exists an $H_0 > 0$ such that for all $H < H_0$ we have*

$$\left| B^0(u^0(\cdot, t_{n+1}); w^H) - B^0(\mathcal{U}_{n+1}^H; w^H) \right| \le C L_{\mathcal{A}^0} H^2 \left\| u^0 \right\|^2_{\mathcal{C}^0([0,T], W^{2,\infty}(\Omega))} \left\| \nabla w^H \right\|_{L^2(\Omega)},$$

where $L_{\mathcal{A}^0} = \operatorname{ess\,sup}_{x \in \Omega} \left\| \mathcal{A}^0(x, \cdot) \right\|_{W^{2,\infty}(\mathbb{R}^d; \mathbb{R}^d)}$ and C is independent of Δt and H.

Proof. In the case that $\mathcal{U}_{n+1}^H = \mathcal{I}_H u^0(x, t_{n+1})$, the Lipschitz continuity $(\mathcal{A}_1)$ of $\mathcal{A}^0$ yields

$$\left| B^0(u^0(\cdot, t_{n+1}); w^H) - B^0(\mathcal{I}_H u^0(\cdot, t_{n+1})) \right| \le L \left\| \nabla u^0(x, t_{n+1}) - \nabla \mathcal{I}_H u^0(x, t_{n+1}) \right\|_{L^2(\Omega)} \left\| \nabla w^H \right\|_{L^2(\Omega)}.$$

Then, the interpolation estimate (25) leads to the estimate *(i)*.

We turn now to the case $\mathcal{U}_{n+1}^H = \tilde{u}_{n+1}^{H,0}$. Using the Taylor formula (64) and the definition of the elliptic projection (27) we derive

$$\begin{aligned}
B^0(u^0(\cdot, t_{n+1}); w^H) - B^0(\tilde{u}_{n+1}^{H,0}; w^H) &= \int_\Omega \left[\mathcal{A}^0(x, \nabla u^0(x, t_{n+1})) - \mathcal{A}^0(x, \nabla \tilde{u}_{n+1}^{H,0}) \right] \cdot \nabla w^H dx \\
&= \underbrace{\int_\Omega D_\xi \mathcal{A}^0(x, \nabla u^0(x, t_{n+1}))(\nabla \tilde{u}_{n+1}^{H,0} - \nabla u^0(x, t_{n+1})) \cdot \nabla w^H dx}_{B_\pi(t_{n+1}; \tilde{u}_{n+1}^{H,0} - u^0(\cdot, t_{n+1}), w^H) = 0} \\
&\quad + \int_\Omega \int_0^1 \Big\{ D_\xi \mathcal{A}^0\Big(x, \nabla u^0(x, t_{n+1}) + \tau \left[\nabla \tilde{u}_{n+1}^{H,0} - \nabla u^0(x, t_{n+1}) \right]\Big) \\
&\qquad\qquad - D_\xi \mathcal{A}^0\big(x, \nabla u^0(x, t_{n+1})\big) \Big\} d\tau \left[\nabla \tilde{u}_{n+1}^{H,0} - \nabla u^0(x, t_{n+1}) \right] \cdot \nabla w^H dx \\
&\le L_{\mathcal{A}^0} \left\| u^0(\cdot, t_{n+1}) - \tilde{u}_{n+1}^{H,0} \right\|^2_{W^{1,4}(\Omega)} \left\| \nabla w^H \right\|_{L^2(\Omega)},
\end{aligned}$$

where we used the estimate (65). Then, the maximum norm bounds of Lemma 5.5 conclude the proof. □

Quadrature error. Estimating the effect of the barycentric quadrature (7) used in the method (8) is achieved by comparing the maps B^0 and $\hat{B}^0$ given by (15) and (16), respectively. To estimate the term (37d) we first derive the following result.

Lemma 5.9. *Assume that $\mathcal{A}^0$ satisfies the hypothesis* (20b) *for $\mu = 1$ or $\mu = 2$. Let B^0 and $\hat{B}^0$ be given by* (15) *and* (16)*, respectively. Then, the error due to the quadrature* (7) *is bounded by*

$$\left|B^0(v^H; w^H) - \hat{B}^0(v^H; w^H)\right| \leq CH^\mu(L_0 + \left\|\nabla v^H\right\|_{L^2(\Omega)})\left\|\nabla w^H\right\|_{L^2(\Omega)},$$

for any $v^H, w^H \in S_0^1(\Omega, \mathcal{T}_H)$ and where C is independent of H.

Proof. Let $v^H, w^H \in S_0^1(\Omega, \mathcal{T}_H)$ and consider first $\mu = 1$. As the gradients ∇v^H and ∇w^H are piecewise constant, we have from (15) and (16) that

$$B^0(v^H; w^H) - \hat{B}^0(v^H; w^H) = \sum_{K\in\mathcal{T}_H} \int_K \left[\mathcal{A}^0(x, \nabla v^H(x_K)) - \mathcal{A}^0(x_K, \nabla v^H(x_K))\right] \cdot \nabla w^H(x_K) dx.$$

The uniform Lipschitz continuity (20b) of $\mathcal{A}^0$ in the space variable x then yields the desired result.

For $\mu = 2$, an application of [14, Theorem 6] immediately yields

$$\left|B^0(v^H; w^H) - \hat{B}^0(v^H; w^H)\right| \leq CH^2\left\|\mathcal{A}^0(x, \nabla v^H)\right\|_{\bar{H}^2(\Omega)}\left\|\nabla w^H\right\|_{L^2(\Omega)},$$

where $\|\cdot\|^2_{\bar{H}^2(\Omega)} = \sum_{K\in\mathcal{T}_H} \|\cdot\|^2_{H^2(K)}$ denotes a broken Sobolev norm. Let the k-th coordinate function of $\mathcal{A}^0$ be denoted by $\mathcal{A}^0_k$, for $1 \leq k \leq d$. Then, for $1 \leq i, j, k \leq d$ and a.e. $x \in \Omega$, the (weak) derivatives of $\mathcal{A}^0(x, \nabla v^H)$ are given by

$$\partial_{x_i}\left[\mathcal{A}^0_k(x, \nabla v^H(x))\right] = \partial_{x_i}\mathcal{A}^0_k(x, \nabla v^H(x)), \;\; \partial_{x_j x_i}\left[\mathcal{A}^0(x, \nabla v^H(x))\right] = \partial_{x_j x_i}\mathcal{A}^0_k(x, \nabla v^H(x)),$$

as ∇v^H is piecewise constant. We conclude the proof by observing that for any $K \in \mathcal{T}_H$ we have $\|\mathcal{A}^0(x, \nabla v^H(x_K))\|_{H^2(K)} \leq C(L_0 + |\nabla v^H(x_K)|)$ due to (20b). □

Remark. For a linear problem $\mathcal{A}^0(x, \xi) = a^0(x)\xi$, with $a^0 \in (L^\infty(\Omega))^{d\times d}$, the regularity assumption of (20b) becomes $a^0 \in W^{2,\infty}(\Omega)$, which is used for FEM based on numerical integration for linear problems, see [35]. Then, the bounds of (20b) are valid for $L_0 = 0$.

With the Lemma (5.9) at hand, the term (37d) can be estimated immediately.

Corollary 5.10. *Let u^0 be the solution of the homogenized problem* (5)*, $\mathcal{I}^H u^0$ its nodal interpolant, $\tilde{u}^{H,0}$ its elliptic projection* (27) *and consider the maps B^0 and $\hat{B}^0$ given by* (15) *and* (16)*. Let $0 \leq n \leq N-1$. If $\mathcal{U}^H_{n+1} = \mathcal{I}_H u^0(\cdot, t_{n+1})$, let $\mu = 1$. If $\mathcal{U}^H_{n+1} = \tilde{u}^{H,0}_{n+1}$, let $\mu = 2$ and assume that $\mathcal{A}^0$ satisfies $(\mathcal{A}_{0-2})$ as well as $\mathcal{A}^0(x, \cdot) \in \mathcal{C}^1(\mathbb{R}^d; \mathbb{R}^d)$ for a.e. $x \in \Omega$. Then, if $\mathcal{A}^0$ satisfies the hypothesis* (20b) *(depends on μ) and $u^0 \in \mathcal{C}^0([0, T], H^{3-\mu}(\Omega))$, we have for every $w^H \in S_0^1(\Omega, \mathcal{T}_H)$ that*

$$\left|B^0(\mathcal{U}^H_{n+1}; w^H) - \hat{B}^0(\mathcal{U}^H_{n+1}; w^H)\right| \leq CH^\mu(L_0 + \left\|u^0\right\|_{\mathcal{C}^0([0,T],H^{3-\mu}(\Omega))})\left\|\nabla w^H\right\|_{L^2(\Omega)},$$

where C is independent of Δt and H.

HMM upscaling error r_{HMM}. The last term (37e) in the error propagation formula (37) quantifies the HMM error, which is only due to the upscaling procedure intrinsically built into the multiscale method (8) and can be bounded using r_{HMM} introduced in (19). Let $w^H \in S_0^1(\Omega, \mathcal{T}_H)$ and $0 \leq n \leq N-1$. Let either $\mathcal{U}^H_{n+1} = \mathcal{I}_H u^0(\cdot, t_{n+1})$ or $\mathcal{U}^H_{n+1} = \tilde{u}^{H,0}_{N+1}$, we obtain

$$\left|\hat{B}^0(\mathcal{U}^H_{n+1}; w^H) - B^H(\mathcal{U}^H_{n+1}; w^H)\right| \leq r_{HMM}(\nabla\mathcal{U}^H_{n+1})\left\|\nabla w^H\right\|_{L^2(\Omega)}. \tag{39}$$

Explicit error estimates for r_{HMM} are derived in Section 5.4.

Proof of Theorem 4.1. Recall that $\theta^H_n = u^H_n - \mathcal{U}^H_n$ for $0 \leq n \leq N$. First, we provide the proof of Theorem 4.1 for $\mu = 2$ and $\mathcal{U}^H_n = \tilde{u}^{H,0}_n$, for all $0 \leq n \leq N$. Let $0 \leq n \leq N-1$ and use

the error propagation formula (37) with the test function θ_{n+1}^H. Combining that with the inequality $1/2\,\bar{\partial}_t\|\theta_n^H\|_{L^2(\Omega)}^2 \leq \int_\Omega \bar{\partial}_t\theta_n^H\,\theta_{n+1}^H$ and the monotonicity of B^H, see Lemma 3.3, leads to

$$\begin{aligned}\frac{1}{2}\bar{\partial}_t\big\|\theta_n^H\big\|_{L^2(\Omega)}^2 + \lambda\big\|\nabla\theta_{n+1}^H\big\|_{L^2(\Omega)}^2 &\leq \int_\Omega \bar{\partial}_t\theta_n^H\,\theta_{n+1}^H dx + B^H(u_{n+1}^H;\theta_{n+1}^H) - B^H(\tilde{u}_{n+1}^{H,0};\theta_{n+1}^H)\\ &\leq C(\Delta t + H^2 + r_{HMM}(\nabla\tilde{u}_{n+1}^{H,0}))\big\|\nabla\theta_{n+1}^H\big\|_{L^2(\Omega)}\\ &\leq C(\Delta t^2 + H^4 + r_{HMM}(\nabla\tilde{u}_{n+1}^{H,0})^2) + \frac{\lambda}{2}\big\|\nabla\theta_{n+1}^H\big\|_{L^2(\Omega)}^2,\end{aligned}$$

where the terms (37a), (37b), (37c), (37d) and (37e) are estimated using the results of Lemmas 5.6, 5.7, 5.8, Corollary 5.10, and inequality (39), respectively, and r_{HMM} is defined in (19). Then, substracting $\lambda/2\|\nabla\theta_{n+1}^H\|_{L^2(\Omega)}^2$ on both sides and multiplying by $2\Delta t$ yields

$$\big\|\theta_{n+1}^H\big\|_{L^2(\Omega)}^2 - \big\|\theta_n^H\big\|_{L^2(\Omega)}^2 + \lambda\,\Delta t\big\|\nabla\theta_{n+1}^H\big\|_{L^2(\Omega)}^2 \leq C\Delta t(\Delta t^2 + H^4 + r_{HMM}(\nabla\tilde{u}_{n+1}^{H,0})^2). \tag{40}$$

Summing the inequality (40) from $n = 0$ to $n = N-1$ and combining that with Lemma 5.4 concludes the proof for $\mathcal{U}_n^H = \tilde{u}_n^{H,0}$. For $\mu = 1$ and $\mathcal{U}_n^H = \mathcal{I}_H u^0(\cdot, t_n)$, the result follows analogously using (25). □

5.4 Explicit estimates for the HMM upscaling error r_{HMM}

We recall, that the nonlinear error functional r_{HMM} given in (19) can be decomposed using the micro error functional r_{mic} and the modeling error functional r_{mod} defined in (22) and (23). First, the micro error r_{mic}, due to the finite element approximation of the micro problems (11), is estimated in Section 5.4.1. Secondly, the modeling error r_{mod} is studied in Section 5.4.2 for (spatially) locally periodic maps $\mathcal{A}^\varepsilon(x,\xi)$. We emphasize that the estimates for the modeling error in Section 5.4.2 are the only results that rely on specific assumptions about the spatial heterogeneities of $\mathcal{A}^\varepsilon(x,\xi)$.

5.4.1 Micro error

Under classical regularity assumptions on $\bar{\chi}_K^\xi$, which is the exact solution to the micro problem (13), and (**H1**), we derive robust linear convergence, i.e., a convergence rate independent of the small scale ε, see Lemma 5.12. Introducing an auxiliary adjoint problem (42) and assuming regularity (**H1***) of its solution, a robust quadratic convergence can be shown in Lemma 5.13.

First, we derive a preparatory result used later on to estimate the micro error r_{mic} in Lemma 5.12 and Lemma 5.13.

Lemma 5.11. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$. Let $\xi \in \mathbb{R}^d$, $K \in \mathcal{T}_H$ and $\bar{\chi}_K^\xi \in W(K_\delta)$ and $\chi_K^{\xi,h} \in S^1(K_\delta,\mathcal{T}_h)$ be the solution of (13) and (12), respectively, with the same coupling condition (either $W(K_\delta) = H_0^1(K_\delta)$ or $W(K_\delta) = W_{per}^1(K_\delta)$). If $\bar{\chi}_K^\xi$ satisfies (**H1**), then we have the priori estimate*

$$\Big\|\nabla\bar{\chi}_K^\xi - \nabla\chi_K^{\xi,h}\Big\|_{L^2(K_\delta)} \leq C\frac{h}{\varepsilon}(L_0 + |\xi|)\sqrt{|K_\delta|},$$

where C is independent of H, h, δ and ε.

Proof. We recall, that the micro problems (13) and (12) can be reformulated using the map a_K^ξ introduced in (17), which is strongly monotone and Lipschitz continuous in the first variable as well as linear and continuous in the second variable. Thus, the analysis of FEM for monotone elliptic problems applies, e.g., see [13, Theorem 5.3.5]. □

Without any additional tools we can derive a first estimate for the micro error r_{mic}.

Lemma 5.12. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$ and (**H1**) holds. For any $v^H \in S_0^1(\Omega,\mathcal{T}_H)$ and either periodic coupling $W(K_\delta) = W_{per}^1(K_\delta)$ or Dirichlet coupling $W(K_\delta) = H_0^1(K_\delta)$ for the micro problems (11), the micro error $r_{mic}(\nabla v^H)$, see (22), can be estimated by*

$$r_{mic}(\nabla v^H) \leq C\frac{h}{\varepsilon}\Big(L_0 + \big\|\nabla v^H\big\|_{L^2(\Omega)}\Big),$$

where C is independent of H, h, δ and ε.

Proof. Due to the definition (22) of r_{mic}, we estimate the difference $\mathcal{A}_K^{0,h}(\xi) - \bar{\mathcal{A}}_K^0(\xi)$ given by (14) for $\xi \in \mathbb{R}^d$ and $K \in \mathcal{T}_H$. Let $\chi_K^{\xi,h}$ and $\bar{\chi}_K^{\xi}$ be the solutions to the micro problems (12) and (13), respectively, with the same coupling condition (either periodic or Dirichlet coupling) on the sampling domain K_δ (associated to K).

Then, the definition (14), the Lipschitz continuity $(\mathcal{A}_1)$ of $\mathcal{A}^\varepsilon$ and Lemma 5.11 yield

$$\begin{aligned}\left|\bar{\mathcal{A}}_K^0(\xi) - \mathcal{A}_K^{0,h}(\xi)\right| &\leq \frac{1}{|K_\delta|}\int_{K_\delta}\left|\mathcal{A}^\varepsilon(x,\xi+\nabla\bar{\chi}_K^\xi) - \mathcal{A}^\varepsilon(x,\xi+\nabla\chi_K^{\xi,h})\right|dx \\ &\leq L\frac{1}{\sqrt{|K_\delta|}}\left\|\nabla\bar{\chi}_K^\xi - \nabla\chi_K^{\xi,h}\right\|_{L^2(K_\delta)} \leq C\frac{h}{\varepsilon}(L_0+|\xi|),\end{aligned}$$

from where the result follows by using the definition (22) of r_{mic}. □

In [1, 2, 9] a convergence of the order $(h/\varepsilon)^2$ has been shown for linear micro problems (11), i.e., for data $\mathcal{A}^\varepsilon(x,\xi) = a^\varepsilon(x)\xi$. Thus, the estimate of Lemma 5.12 is non-optimal. We note that an adjoint micro problem was used to prove the quadratic convergence for non-symmetric tensors $a^\varepsilon(x)$, see [9, Lemma 4.6] for a short proof. In this view, we introduce the following *linear* auxiliary micro problems: for $\xi \in \mathbb{R}^d$, $1 \leq j \leq d$ and $K \in \mathcal{T}_H$, find $\bar{X}_K^{\xi,j} \in W(K_\delta)$ such that

$$\int_{K_\delta}\left(D_\xi\mathcal{A}^\varepsilon(x,\xi+\nabla\bar{\chi}_K^\xi)\right)^T(e_j+\nabla\bar{X}_K^{\xi,j})\cdot\nabla z\,dx = 0, \qquad \forall\, z \in W(K_\delta), \tag{41}$$

where K_δ is the sampling domain associated to K and $\bar{\chi}_K^\xi$ solves the cell problem (13). We note that problem (41) admits a unique solution if $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$ and $\mathcal{A}^\varepsilon(x,\cdot) \in \mathcal{C}^1(\mathbb{R}^d;\mathbb{R}^d)$, as then the Jacobian $D_\xi\mathcal{A}^\varepsilon$ is uniformly bounded and elliptic, see Lemma A.1.

Remark. We note, that for a linear map $\mathcal{A}^\varepsilon(x,\xi) = a^\varepsilon(x)\xi$ the derivative $D_\xi\mathcal{A}^\varepsilon$ is simply given by $D_\xi\mathcal{A}^\varepsilon(x,\xi) = a^\varepsilon(x)$. Thus, the auxiliary micro problem (41) reduces to

$$\int_{K_\delta} a^\varepsilon(x)^T(e_j+\nabla\bar{X}_K^j)\cdot\nabla z\,dx = 0, \qquad \forall\, z \in W(K_\delta), \tag{42}$$

which is independent of the corrector $\bar{\chi}_K^j$. Indeed, we recover the adjoint micro problem used to analyze linear homogenization problems, e.g., see [9].

Lemma 5.13. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$, $\mathcal{A}^\varepsilon(x,\cdot) \in W^{2,\infty}(\mathbb{R}^d;\mathbb{R}^d)$ and (**H1**), (**H1***) hold. For any $v^H \in S_0^1(\Omega,\mathcal{T}_H)$ and either periodic coupling $W(K_\delta) = W_{per}^1(K_\delta)$ or Dirichlet coupling $W(K_\delta) = H_0^1(K_\delta)$ for the micro problems (11), the micro error $r_{mic}(\nabla v^H)$, see (22), can be estimated by*

$$r_{mic}(\nabla v^H) \leq C\left(\frac{h}{\varepsilon}\right)^2\left(L_0+L_0^2+\left\|\nabla v^H\right\|_{L^2(\Omega)}+\left\|\nabla v^H\right\|_{L^4(\Omega)}^2\right),$$

where C is independent of H, h, δ and ε.

Proof. Again, like in Lemma 5.12, we estimate the difference $\mathcal{A}_K^{0,h}(\xi) - \bar{\mathcal{A}}_K^0(\xi)$ for $\xi \in \mathbb{R}^d$ and $K \in \mathcal{T}_H$ (with associated sampling domain K_δ), where $\bar{\mathcal{A}}_K^0(\xi)$ and $\mathcal{A}_K^{0,h}(\xi)$ are given by (14). They are based on the solutions $\bar{\chi}_K^\xi$ and $\chi_K^{\xi,h}$ to the micro problems (13) and (12), respectively, solved with the same coupling condition. Let $1 \leq j \leq d$, then

$$\begin{aligned}\bar{\mathcal{A}}_K^0(\xi)\cdot e_j - \mathcal{A}_K^{0,h}(\xi)\cdot e_j &= \frac{1}{|K_\delta|}\int_{K_\delta}\left[\mathcal{A}^\varepsilon(x,\xi+\nabla\bar{\chi}_K^\xi) - \mathcal{A}^\varepsilon(x,\xi+\nabla\chi_K^{\xi,h})\right]\cdot e_j\,dx \\ &= \frac{1}{|K_\delta|}\int_{K_\delta}\left[\mathcal{A}^\varepsilon(x,\xi+\nabla\bar{\chi}_K^\xi) - \mathcal{A}^\varepsilon(x,\xi+\nabla\chi_K^{\xi,h})\right]\cdot\left(e_j+\nabla I_h\bar{X}_K^{\xi,j}\right)dx,\end{aligned}$$

where the Galerkin orthogonality for monotone FEM is used and $I_h\bar{X}_K^{\xi,j} \in S^1(K_\delta,\mathcal{T}_h)$ is the nodal

interpolant of $\bar{X}_K^{\xi,j}$ on K_δ. Further, we apply the Taylor formula (64) and use that $\bar{X}_K^{\xi,j}$ solves (41)

$$
\begin{aligned}
[\mathcal{A}_K^{0,h}(\xi) - \bar{\mathcal{A}}_K^0(\xi)] \cdot e_j &= \frac{1}{|K_\delta|}\int_{K_\delta} D_\xi \mathcal{A}^\varepsilon(x, \xi + \nabla\bar{\chi}_K^\xi)(\nabla\chi_K^{\xi,h} - \nabla\bar{\chi}_K^\xi)\cdot(e_j + \nabla I_h \bar{X}_K^{\xi,j})dx \\
&\quad + \frac{1}{|K_\delta|}\int_{K_\delta}\int_0^1 D_\xi\mathcal{A}^\varepsilon(x, \xi + \nabla\bar{\chi}_K^\xi + \tau(\nabla\chi_K^{\xi,h} - \nabla\bar{\chi}_K^\xi)) - D_\xi\mathcal{A}^\varepsilon(x,\xi+\nabla\bar{\chi}_K^\xi)d\tau \\
&\qquad \times (\nabla\chi_K^{\xi,h} - \nabla\bar{\chi}_K^\xi)\cdot(e_j + \nabla I_h\bar{X}_K^{\xi,j})dx \\
&= \frac{1}{|K_\delta|}\int_{K_\delta} D_\xi \mathcal{A}^\varepsilon(x, \xi + \nabla\bar{\chi}_K^\xi)(\nabla\chi_K^{\xi,h} - \nabla\bar{\chi}_K^\xi)\cdot(\nabla I_h \bar{X}_K^{\xi,j} - \nabla\bar{X}_K^{\xi,j})dx \\
&\quad + \frac{1}{|K_\delta|}\int_{K_\delta}\int_0^1 D_\xi\mathcal{A}^\varepsilon(x, \xi + \nabla\bar{\chi}_K^\xi + \tau(\nabla\chi_K^{\xi,h} - \nabla\bar{\chi}_K^\xi)) - D_\xi\mathcal{A}^\varepsilon(x,\xi+\nabla\bar{\chi}_K^\xi)d\tau \\
&\qquad \times (\nabla\chi_K^{\xi,h} - \nabla\bar{\chi}_K^\xi)\cdot(e_j + \nabla I_h\bar{X}_K^{\xi,j})dx.
\end{aligned}
$$

Then, the uniform boundedness and the Lipschitz continuity of $D_\xi\mathcal{A}^\varepsilon(x,\cdot)$ yield

$$
\begin{aligned}
\left|\bar{\mathcal{A}}_K^0(\xi)\cdot e_j - \mathcal{A}_K^{0,h}(\xi)\cdot e_j\right| &\le \frac{L}{|K_\delta|}\left\|\nabla\chi_K^{\xi,h} - \nabla\bar{\chi}_K^\xi\right\|_{L^2(K_\delta)}\left\|\nabla I_h\bar{X}_K^{\xi,j} - \nabla\bar{X}_K^{\xi,j}\right\|_{L^2(K_\delta)} \\
&\quad + \frac{C}{|K_\delta|}\left\|\nabla\chi_K^{\xi,h} - \nabla\bar{\chi}_K^\xi\right\|^2_{L^2(K_\delta)}\left(1 + \left|I_h\bar{X}_K^{\xi,j}\right|_{W^{1,\infty}(K_\delta)}\right) \\
&\le C\left(\frac{h}{\varepsilon}\right)^2(L_0 + L_0^2 + |\xi| + |\xi|^2)\left(1 + \left|I_h\bar{X}_K^{\xi,j}\right|_{W^{1,\infty}(K_\delta)}\right),
\end{aligned}
$$

where we appplied Lemma 5.11 (using assumption (**H1**) and the standard H^1 interpolation error estimate (25) for the nodal interpolation operator $\mathcal{I}_h$ on K_δ (using assumption *(i)* from (**H1***)). Further, the bound (26) for $\mathcal{I}_h$ and hypothesis *(ii)* from (**H1***) yield $|I_h\bar{X}_K^{\xi,j}|_{W^{1,\infty}(K_\delta)} \le C$. Then, the result follows from the definition (22) of r_{mic}. □

Proof of Theorem 4.2. We combine the results of Theorem 4.1 with the estimates of Lemma 5.12 and Lemma 5.13, with $v^H = \mathcal{U}_n^H$ for $1 \le n \le N$, for linear and quadratic micro convergence, respectively. We note that $\|\nabla\mathcal{U}_n^H\|_{L^2(\Omega)}$ and $\|\nabla\mathcal{U}_n^H\|_{L^4(\Omega)}$ are bounded for both $\mathcal{U}_n^H = \mathcal{I}_H u^0(\cdot,t_n)$ the nodal interpolant of the homogenized solution u^0 and $\mathcal{U}_n^H = \tilde{u}_n^{H,0}$ the elliptic projection (27). In particular, we have $\|\nabla\mathcal{I}_H u^0(\cdot,t_n)\|_{L^4(\Omega)} \le C\|u^0(\cdot,t_n)\|_{H^2(\Omega)}$, from classical interpolation results, see [13, Theorem 3.1.6], and Lemma 5.5 yields $\|\nabla\tilde{u}_n^{H,0}\|_{L^4(\Omega)} \le C\|u^0(\cdot,t_n)\|_{W^{1,\infty}(\Omega)}$. □

5.4.2 Modeling error

In this section, we assume that $\mathcal{A}^\varepsilon$ has locally periodic spatial heterogeneities. With this structural assumption, an explicit representation of $\mathcal{A}^0$ can be derived. This representation allows to estimate the modeling error r_{mod} explicitely including the influence of the boundary conditions chosen for $W(K_\delta)$ in (10), the sampling domain size δ and the absence of collocation of $\mathcal{A}(x, x/\varepsilon, \xi)$ in the slow variable x.

Periodic homogenization. As we are considering locally periodic maps $\mathcal{A}^\varepsilon$ independent of the time variable t, we can use the representation of $\mathcal{A}^0$ derived in the case of monotone elliptic problems $-\operatorname{div}(\mathcal{A}^\varepsilon(x,\nabla u^\varepsilon)) = f$ in Ω, see [36, Theorem 8.1]. In particualr, let $\mathcal{A}^\varepsilon$ satisfy conditions ($\mathcal{A}_{0-2}$) uniformly for $\varepsilon > 0$ and assume that $\mathcal{A}^\varepsilon(x,\xi) = \mathcal{A}(x, x/\varepsilon, \xi)$ where $\mathcal{A}(x,y,\xi)$ is Y-periodic in y, i.e., $\mathcal{A}^\varepsilon$ is locally periodic. Then, see [28, Section 3], the homogenized map $\mathcal{A}^0$ is explicitely given by

$$
\mathcal{A}^0(x,\xi) = \int_Y \mathcal{A}(x, y, \xi + \nabla\chi^\xi(x,y))dy, \tag{43}
$$

where $x \in \Omega$, $\xi \in \mathbb{R}^d$ and $\chi^\xi(x,\cdot) \in W^1_{per}(Y)$ solves the cell problem: find $\chi^\xi(x,\cdot) \in W^1_{per}(Y)$ such that

$$
\int_Y \mathcal{A}(x,y,\xi + \nabla\chi^\xi(x,y))\cdot\nabla z\, dy = 0, \qquad \forall\, z \in W^1_{per}(Y). \tag{44}
$$

Collocation in the slow variable. If the explicit decomposition between macro and micro scale is known for a locally periodic map $\mathcal{A}^\varepsilon$, we can collocate the map $\mathcal{A}(x, x/\varepsilon, \xi)$ in the slow variable x at

the quadrature nodes x_K. Then, for $K \in \mathcal{T}_H$ and $\xi \in \mathbb{R}^d$, the collocated micro problem reads as: find $\tilde{\chi}_K^\xi \in W(K_\delta)$ such that

$$\int_{K_\delta} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla \tilde{\chi}_K^\xi) \cdot \nabla z \, dx = 0, \qquad \forall\, z \in W(K_\delta), \tag{45}$$

and, analogously to (14), the homogenized map $\tilde{\mathcal{A}}_K^0$ can be defined

$$\tilde{\mathcal{A}}_K^0(\xi) = \frac{1}{|K_\delta|} \int_{K_\delta} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla \tilde{\chi}_K^\xi) dx. \tag{46}$$

First, we show the boundedness of the solutions to micro problems (13) and (45).

Lemma 5.14. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$. Let $\bar{\chi}_K^\xi$ be the solution to the micro problem (13), then,*

$$\left\| \nabla \bar{\chi}_K^\xi \right\|_{L^2(K_\delta)} \leq \frac{L}{\lambda} (L_0 + |\xi|) \sqrt{|K_\delta|}, \qquad \text{for any } \xi \in \mathbb{R}^d, K \in \mathcal{T}_H,$$

where C is independent of ξ, δ and ε. If additionally $\mathcal{A}^\varepsilon$ is locally periodic, i.e., $\mathcal{A}^\varepsilon(x, \xi) = \mathcal{A}(x, x/\varepsilon, \xi)$, the same bound holds for $\tilde{\chi}_K^\xi$ solving the collocated micro problem (45).

Proof. This result is obtained by combining the monotonicity of $\mathcal{A}^\varepsilon$, the definition of the micro problem (13) and the growth estimate (3) for $\mathcal{A}^\varepsilon$ derived in Lemma A.1. □

If a locally periodic map $\mathcal{A}^\varepsilon$ is not collocated in the slow variable to solve the micro problems (11), an error of order $\mathcal{O}(\delta)$ is introduced.

Lemma 5.15. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$, and (**H2**). Let $\xi \in \mathbb{R}^d$, $K \in \mathcal{T}_H$ and the homogenized maps $\bar{\mathcal{A}}_K^0$ and $\tilde{\mathcal{A}}_K^0$ be given in (14) and (46), respectively. Then, for every $\delta \geq \varepsilon$ and independently of the coupling conditions imposed by the choice of $W(K_\delta)$ (either periodic or Dirichlet coupling) we have*

$$\left| \tilde{\mathcal{A}}_K^0(\xi) - \bar{\mathcal{A}}_K^0(\xi) \right| \leq C\delta(L_0 + |\xi|),$$

where C is independent of ξ, δ and ε.

Proof. Let $\xi \in \mathbb{R}^d$, $K \in \mathcal{T}_H$ and the functions $\bar{\chi}_\xi^K$ and $\tilde{\chi}_\xi^K$ be the solutions of the micro problems (13) and (45), respectively. We use the monotonicity of $\mathcal{A}(x, y, \cdot)$ and the formulas (13) and (45) to obtain

$$\begin{aligned}
\lambda \left\| \nabla \tilde{\chi}_K^\xi - \nabla \bar{\chi}_K^\xi \right\|_{L^2(K_\delta)}^2 &\leq \int_{K_\delta} \left[\mathcal{A}(x, \tfrac{x}{\varepsilon}, \xi + \nabla \tilde{\chi}_K^\xi) - \mathcal{A}(x, \tfrac{x}{\varepsilon}, \xi + \nabla \bar{\chi}_K^\xi) \right] \cdot (\nabla \tilde{\chi}_K^\xi - \nabla \bar{\chi}_K^\xi) dx \\
&= \int_{K_\delta} \left[\mathcal{A}(x, \tfrac{x}{\varepsilon}, \xi + \nabla \tilde{\chi}_K^\xi) - \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla \tilde{\chi}_K^\xi) \right] \cdot (\nabla \tilde{\chi}_K^\xi - \nabla \bar{\chi}_K^\xi) dx \\
&\leq C\delta(L_0 + |\xi|) \sqrt{|K_\delta|} \left\| \nabla \tilde{\chi}_K^\xi - \nabla \bar{\chi}_K^\xi \right\|_{L^2(K_\delta)},
\end{aligned} \tag{47}$$

where we used the Lipschitz continuity from (**H2**) and Lemma 5.14. Then, the result is obtained by combining the definitions of $\bar{\mathcal{A}}_K^0$ and $\tilde{\mathcal{A}}_K^0$ with the Lipschitz continuity from (**H2**), Lemma 5.14 and estimate (47). □

Periodic boundary conditions. We next show that periodic coupling with a sampling domain size δ as an integer multiple of ε is optimal for locally periodic data.

Lemma 5.16. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$ and (**H2**). Let $K \in \mathcal{T}_H$ and the maps $\tilde{\mathcal{A}}_K^0$ and $\mathcal{A}^0(x_K, \cdot)$ be given by (46) and (43), respectively. Then, for periodic coupling $W(K_\delta) = W_{per}^1(K_\delta)$ and a sampling domain size δ such that $\delta/\varepsilon \in \mathbb{N}_{>0}$ it holds that $\mathcal{A}^0(x_K, \xi) = \tilde{\mathcal{A}}_K^0(\xi)$ for all $\xi \in \mathbb{R}^d$.*

Proof. First, we assume that $\delta = \varepsilon$. Using the periodicity of $\chi^\xi(x_k, \cdot)$ and $\mathcal{A}(x, \cdot, \xi)$ we observe that

$$\begin{aligned}
\mathcal{A}^0(x_K, \xi) - \tilde{\mathcal{A}}_K^0(\xi) &= \int_Y \mathcal{A}(x_K, y, \xi + \nabla \chi^\xi(x_K, y)) dy - \frac{1}{|K_\varepsilon|} \int_{K_\varepsilon} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla \tilde{\chi}_K^\xi(x)) dx \\
&= \frac{1}{|K_\varepsilon|} \int_{K_\varepsilon} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla \chi^\xi(x_K, \tfrac{x}{\varepsilon})) - \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla \tilde{\chi}_K^\xi(x)) dx.
\end{aligned}$$

Thus, it is sufficient to show that $\nabla\chi^\xi(x_K, x/\varepsilon) = \nabla\tilde{\chi}_K^\xi(x)$. Indeed, we observe that the function $\eta_{x_K}^\xi(x) = \varepsilon\chi^\xi(x_K, x/\varepsilon) \in W_{per}^1(K_\varepsilon)$ is a solution of (45) and the claim follows by uniqueness of such solutions. In the case of $\delta/\varepsilon \in \mathbb{N}_{>1}$, the proof is totally analogous by using periodic extensions of $\chi^\xi(x_K, \cdot)$ to $K_{n\varepsilon}$, for some $n \in \mathbb{N}$. □

Dirichlet boundary conditions. In contrast to the optimal coupling from Lemma 5.16, using Dirichlet coupling leads to resonance errors due to the artificial boundary conditions.

Lemma 5.17. *Assume that $\mathcal{A}^\varepsilon$ satisfies $(\mathcal{A}_{0-2})$ and (**H2**). Let $\xi \in \mathbb{R}^d$, $K \in \mathcal{T}_H$ and the maps $\tilde{\mathcal{A}}_K^0$ and $\mathcal{A}^0(x_K, \cdot)$ be given by (46) and (43), respectively. Further, assume that the exact corrector $\chi^\xi(x_K, \cdot)$ solving the cell problem (44) satisfies $\chi^\xi(x_K, \cdot) \in W^{1,\infty}(Y)$. Then, for Dirichlet coupling $W(K_\delta) = H_0^1(K_\delta)$ and a sampling domain size $\delta > \varepsilon$ it holds*

$$\left|\mathcal{A}^0(x_K, \xi) - \tilde{\mathcal{A}}_K^0(\xi)\right| \leq C\Big(\frac{\varepsilon}{\delta}\Big)^{1/2}\Big(L_0 + |\xi| + \big\|\chi^\xi(x_K, y)\big\|_{W^{1,\infty}(Y)}\Big),$$

where C is independent of ξ, δ and ε.

Proof. We use the techniques used to analyze the resonance error for linear homogenization problems, see [21, Theorem 1.2]. Let $n \in \mathbb{N}$ be given by $n = \lfloor \delta/\varepsilon \rfloor$ (if $\delta/\varepsilon \notin \mathbb{N}$), or $n = \delta/\varepsilon - 1$ (if $\delta/\varepsilon \in \mathbb{N}_{>0}$). Further, we define $K_\Gamma = K_\delta \setminus K_{n\varepsilon}$ and we observe that $|K_\Gamma| \leq C\varepsilon\delta^{d-1}$. Then we decompose the difference $\tilde{\mathcal{A}}_K^0 - \mathcal{A}^0(x_K, \xi)$ into two terms according to

$$\begin{aligned}\tilde{\mathcal{A}}_K^0(\xi) - \mathcal{A}^0(x_K, \xi) &= \frac{1}{|K_\delta|}\int_{K_\delta} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\tilde{\chi}_K^\xi) - \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon}))dx \\ &\quad + \frac{1}{|K_\delta|}\int_{K_\delta} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon}))dx - \frac{1}{|K_{n\varepsilon}|}\int_{K_{n\varepsilon}} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon}))dx\end{aligned}$$

where $\chi^\xi(x_K, y)$ is extended periodically to $\mathbb{R}^d$ and the first and second line is denoted by I_1 and I_2, respectively. First, we estimate I_2 similarly as for the linear case

$$\begin{aligned}I_2 &= \frac{1}{|K_\delta|}\int_{K_\Gamma} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon}))dx + \left(\frac{1}{|K_\delta|} - \frac{1}{|K_{n\varepsilon}|}\right)\int_{K_{n\varepsilon}} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon}))dx \\ &\leq C\left(\frac{|K_\Gamma|}{|K_\delta|} + \frac{|K_\Gamma|}{|K_\delta||K_{n\varepsilon}|}|K_{n\varepsilon}|\right)(L_0 + |\xi| + |\chi^\xi(x_K, y)|_{W^{1,\infty}(Y)}) \\ &\leq C\frac{\varepsilon}{\delta}(L_0 + |\xi| + |\chi^\xi(x_K, y)|_{W^{1,\infty}(Y)}),\end{aligned} \tag{48}$$

using the estimate (3) for $\mathcal{A}(x, \xi)$ and the assumption $\chi^\xi(x_K, \cdot) \in W^{1,\infty}(Y)$. To estimate the term I_1 we define the function $\theta^\xi(x) = \tilde{\chi}_K^\xi(x) - \varepsilon\chi^\xi(x_K, x/\varepsilon)$ on K_δ (using the periodic extension of $\chi^\xi(x_K, \cdot)$). As $\tilde{\chi}_K^\xi|_{\partial K_\delta} = 0$ (in the sense of traces), we decompose θ^ξ into

$$\theta^\xi(x) = \theta_0^\xi(x) - \varepsilon\chi^\xi(x_K, \tfrac{x}{\varepsilon})(1 - \rho_\varepsilon(x)), \qquad x \in K_\delta, \tag{49}$$

where $\theta_0^\xi \in H_0^1(K_\delta)$ and $\rho_\varepsilon \colon K_\delta \to \mathbb{R}$ is a smooth cut-off function satisfying $\rho_\varepsilon \equiv 1$ in $K_{n\varepsilon}$, $\rho_\varepsilon|_{\partial K_\delta} \equiv 0$ and $|\nabla\rho_\varepsilon| \leq C\varepsilon^{-1}$ in K_Γ (where C is independent of δ and ε). Using the monotonicity $(\mathcal{A}_2)$ of $\mathcal{A}$ and the decomposition (49) of θ^ξ we obtain

$$\begin{aligned}&\lambda\left\|\nabla\tilde{\chi}_K^\xi(x) - \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon})\right\|_{L^2(K_\delta)}^2 \\ &\leq \int_{K_\delta}\Big[\mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\tilde{\chi}_K^\xi(x)) - \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon}))\Big] \cdot \Big(\nabla\tilde{\chi}_K^\xi(x) - \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon})\Big)dx \\ &= \int_{K_\delta} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\tilde{\chi}_K^\xi(x)) \cdot \nabla\theta_0^\xi dx - \int_{K_\delta} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon})) \cdot \nabla\theta_0^\xi dx \\ &\quad + \int_{K_\delta}\Big[\mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\tilde{\chi}_K^\xi(x)) - \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon}))\Big] \cdot \nabla[\varepsilon\chi^\xi(x_K, \tfrac{x}{\varepsilon})(1 - \rho_\varepsilon(x))]dx \\ &=: J_1 + J_2 + J_3.\end{aligned} \tag{50}$$

First, as $\tilde{\chi}_K^\xi$ solves the cell problem (45) in the space $W(K_\delta) = H_0^1(K_\delta)$ we have $J_1 = 0$. To show that the second term J_2 vanishes as well we define $\theta_{per}^\xi \in W_{per}^1(K_{(n+1)\varepsilon})$ by

$$\theta_{per}^\xi(x) = \theta_0^\xi(x) - \int_{K_\delta} \theta_0^\xi(x)dx, \quad (\text{if } x \in K_\delta), \qquad \theta_{per}^\xi(x) = 0, \quad (\text{if } x \in K_{(n+1)\varepsilon} \setminus K_\delta).$$

Thus, we observe that $\nabla\theta_{per}^\xi = \nabla\theta^\xi$ on K_δ and $\nabla\theta_{per}^\xi = 0$ on $K_{(n+1)\varepsilon} \setminus K_\delta$. Hence,

$$J_2 = \int_{K_{(n+1)\varepsilon}} \mathcal{A}(x_K, \tfrac{x}{\varepsilon}, \xi + \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon})) \cdot \nabla\theta_{per}^\xi(x)dx = 0,$$

is obtained by following the proof of Lemma 5.16. Further, using the Lipschitz continuity ($\mathcal{A}_1$) of $\mathcal{A}$, we estimate the term J_3 as

$$\begin{aligned} |J_3| &\le L\Big\|\nabla\tilde{\chi}_K^\xi(x) - \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon})\Big\|_{L^2(K_\delta)} \big\|\nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon})(1-\rho_\varepsilon(x)) + \varepsilon\nabla\rho_\varepsilon(x)\chi^\xi(x_K, \tfrac{x}{\varepsilon})\big\|_{L^2(K_\Gamma)} \\ &\le C\sqrt{|K_\Gamma|}\Big\|\nabla\tilde{\chi}_K^\xi(x) - \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon})\Big\|_{L^2(K_\delta)} \big\|\chi^\xi(x_K, y)\big\|_{W^{1,\infty}(Y)}, \end{aligned} \tag{51}$$

where we used the properties of ρ_ε, in particular, $1 - \rho_\varepsilon(x) \equiv 0$ on $K_{n\varepsilon}$ and $\nabla\rho_\varepsilon \le C\varepsilon^{-1}$. Combining that $J_1 = J_2 = 0$ and the estimate (51) of J_3 with the inequality (50) leads to

$$\Big\|\nabla\tilde{\chi}_K^\xi(x) - \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon})\Big\|_{L^2(K_\delta)} \le C\sqrt{|K_\Gamma|}\big\|\chi^\xi(x_K, y)\big\|_{W^{1,\infty}(Y)}.$$

Thus, I_1 can be estimated by the previous estimate and the Lipschitz continuity ($\mathcal{A}_1$)

$$|I_1| \le \frac{L}{\sqrt{|K_\delta|}}\Big\|\nabla\tilde{\chi}_K^\xi(x) - \nabla\chi^\xi(x_K, \tfrac{x}{\varepsilon})\Big\|_{L^2(K_\delta)} \le C\Big(\frac{\varepsilon}{\delta}\Big)^{1/2}\big\|\chi^\xi(x_K, y)\big\|_{W^{1,\infty}(Y)}. \tag{52}$$

Combining the estimates (52) and (48) for I_1 and I_2, respectively, concludes the proof. □

Proof of Theorem 4.3. The Theorem 4.3 is proved by combining Theorem 4.2 with the estimates from Lemma 5.15 (collocation error), Lemma 5.16 (periodic coupling) and Lemma 5.17 (Dirichlet coupling). □

6 Implementation and numerical results

In this section, we comment on the implementation of the multiscale method (8) and present numerical results for different test problems. In particular, numerical studies of the convergence rates as well as the modeling error are given and the applicability to a test problem from material sciences is shown.

6.1 Implementation

In this section, we briefly discuss an implementation of the multiscale method (8). As the macroscopic equation (8) and the micro problems (11) are both nonlinear and coupled together, some care is needed. We thus describe how $u_{n+1}^H \in S_0^1(\Omega, \mathcal{T}_H)$ solving (8) is obtained for given $u_n^H \in S_0^1(\Omega, \mathcal{T}_H)$ and $n \in \mathbb{N}$.

At the macro level, the unknown u_{n+1}^H is approximated by a sequence $\{u_{n+1}^{H,(j)}\}_{j\in\mathbb{N}}$ obtained by a Newton iteration for the macro equation (8) with the initial guess $u_{n+1}^{H,(0)} = u_n^H$. As the macro equation involves the nonlinear map B^H given in (9), a set of constrained micro problems (11) has to be solved (at each macro iteration) and the Fréchet derivative of $B^H(v^H; w^H)$ with respect to v^H has to be computed. We follow the ideas from [26].

Newton's method for micro problems. Let $v^H \in S_0^1(\Omega, \mathcal{T}_H)$ be a macro function and $K \in \mathcal{T}_H$ with associated sampling domain K_δ. The solution v_K^h to the micro problem (11) is then computed by a Newton's method at microscopic level. In particular, for a given initial guess $v_K^{h,(0)}$, the micro solution v_K^h is approximated by the sequence $\{v_K^{h,(j)}\}_{j\in\mathbb{N}}$ with $v_K^{h,(j)} - v^H \in S^1(K_\delta, \mathcal{T}_h)$ solving

$$\mathcal{N}_K^h(v_K^{h,(j)}; v_K^{h,(j+1)} - v_K^{h,(j)}, w^h) = -B_K^h(v_K^{h,(j)}; w^h), \qquad \forall\, w^h \in S^1(K_\delta, \mathcal{T}_h), j \in \mathbb{N}, \tag{53}$$

where the linear map $B_K^h(z^H+q^h;\cdot)$ and the bilinear map $\mathcal{N}_K^h(z^H+q^h;\cdot,\cdot)$ are given by

$$B_K^h(z^H+q^h;w^h) = \int_{K_\delta} \mathcal{A}^\varepsilon(x,\nabla z^H+\nabla q^h)\cdot\nabla w^h dx, \tag{54}$$

$$\mathcal{N}_K^h(z^H+q^h;v^h,w^h) = \int_{K_\delta} D_\xi\mathcal{A}^\varepsilon(x,\nabla z^H+\nabla q^h)\nabla v^h\cdot\nabla w^h dx, \tag{55}$$

for $z^H\in S_0^1(\Omega,\mathcal{T}_H)$ and $q^h,v^h,w^h\in S^1(K_\delta,\mathcal{T}_h)$.

Further, the local contribution to the Fréchet derivative of B^H is computed via an auxiliary micro problem, see [26]. For $z^H\in S_0^1(\Omega,\mathcal{T}_H)$ and z_K^h its associated micro solution to (11), the auxiliary micro function v_K^{h,z^H} solves: find $v_K^{h,z^H}-v^H\in S^1(K_\delta,\mathcal{T}_h)$ such that

$$\mathcal{N}_K^h(z_K^h;v_K^{h,z^H},w^h)=0, \qquad \forall\, w^h\in S^1(K_\delta,\mathcal{T}_h), \tag{56}$$

where $\mathcal{N}_K^h$ is defined in (55). As the auxiliary micro problem (56) is linear it only leads to additional computational cost comparable to one iteration of the micro Newton's method (53). Finally, both problems (53) and (56) admit a unique solution, as $D_\xi\mathcal{A}^\varepsilon$ is uniformly bounded and elliptic, see Lemma A.1.

Newton's method for macro scheme. For $j\in\mathbb{N}$, the $(j+1)$-th iterate $u_{n+1}^{H,(j+1)}$ of the macro Newton's method to approximate u_{n+1}^H solves

$$\int_\Omega \frac{u_{n+1}^{H,(j+1)}-u_n^H}{\Delta t}\,w^H dx + \mathcal{N}^H(u_{n+1}^{H,(j)};u_{n+1}^{H,(j+1)}-u_{n+1}^{H,(j)},w^H) = \int_\Omega f\,w^H dx - B^H(u_{n+1}^{H,(j)};w^H), \quad \forall\, w^H\in S_0^1(\Omega,\mathcal{T}_H),$$

where B^H is given by (9) and $\mathcal{N}^H$ is defined for $v^H,w^H,z^H\in S_0^1(\Omega,\mathcal{T}_H)$

$$\mathcal{N}^H(z^H;v^H,w^H) = \sum_{K\in\mathcal{T}_H}\frac{|K|}{|K_\delta|}\int_{K_\delta} D_\xi\mathcal{A}^\varepsilon(x,\nabla z_K^h)\nabla v_K^{h,z^H}dx\cdot\nabla w^H(x_K), \tag{57}$$

where z_K^h is the micro solution to (11) associated to z^H and v_K^{h,z^H} is the solution to the auxiliary micro problem (56) constrained by v^H. Again, a unique solution $u_{n+1}^{H,(j+1)}$ to (57) exists as $D_\xi\mathcal{A}^\varepsilon$ is uniformly bounded and elliptic.

6.2 Convergence rates

The aim of this section is to validate the theoretical convergence rates given in Section 4.

Setting. We construct a test problem similar to Hoang [28]. The model problem (2) is considered on the time interval $[0,2]$ with homogeneous Dirichlet boundary conditions on the spatial domain $\Omega=(0,1)^2$. The data $\mathcal{A}^\varepsilon$ and source term f are chosen such that

$$u^0(x,t)=\Phi(t)(x_1^2-x_1)(x_2^2-x_2), \qquad \Phi(t)=21\cdot(10\cos(\tfrac{\pi}{2}t)+11)^{-1}, \tag{58}$$

is the exact solution to the homogenized problem (5). In particular, we consider locally periodic data $\mathcal{A}^\varepsilon(t;x,\xi)=\mathcal{A}(t;x,x/\varepsilon,\xi)=\mathcal{A}(t;x,y,\xi)$ with Y-periodic (with respect to y) map $\mathcal{A}$ decomposed as $\mathcal{A}(t;x,y,\xi)=\mathcal{A}_p(x,y,\xi)+c(t;x,y)$. We then take $\mathcal{A}_p$ as

$$\mathcal{A}_p(x,y,\xi)=\Big[1+\sin(2\pi(y_1+y_2))+(\tfrac{9}{8}+\sin(2\pi y_1+\tfrac{\pi}{3}))(\tfrac{9}{8}+\cos(2\pi y_2))(1+|\xi|^2)^{-1/2}\Big]\,\xi, \tag{59}$$

which indeed satisfies assumptions $(\mathcal{A}_{0-2})$. Further, we derive (using Maple) that

$$f(x,t)=\Phi'(t)(x_1^2-x_1)(x_2^2-x_2), \quad c(t;x,y)=-\mathcal{A}_p(x,y,[e_1(t;x,y),e_2(t;x,y)]^T), \tag{60}$$

where $\Phi'(t)$ is the derivative of $\Phi(t)$ from (58) and $e_i(t;x,y)$, for $i=1,2$, is given by

$$e_i(t;x,y)=\Phi(t)[(2x_i-1)(x_{3-i}^2-x_{3-i})+(x_1+x_2)\cos(2\pi y_i)\sin(2\pi y_{3-i})].$$

We note that the map $\mathcal{A}^\varepsilon(t;x,\xi)$ and the source term f depend on the time t, while the results of Section 4 assumes time-independent $\mathcal{A}^\varepsilon$ and f. As $\mathcal{A}^\varepsilon$ and f are smooth with respect to t, the analysis can however be extended to time-dependent data.

For the numerical method (8), we choose periodic coupling for the micro problems (11), take $\delta = \varepsilon = 10^{-4}$ as size of the sampling domains K_δ and collocate $\mathcal{A}(t; x, x/\varepsilon, \xi)$ in the slow variable x at the barycenters x_K. Thus, the modeling error is identical to zero and we expect the convergence rates (24). Further, we discretize the macro domain $\Omega = (0,1)^2$ and the sampling domains K_δ by uniform triangular meshes with N_{mac} and N_{mic} macro and micro elements in each spatial dimension, respectively.

Error measure. To measure the error between the exact homogenized solution u^0 and the numerical solution u_n^H obtained by (8), we use the relative error measures

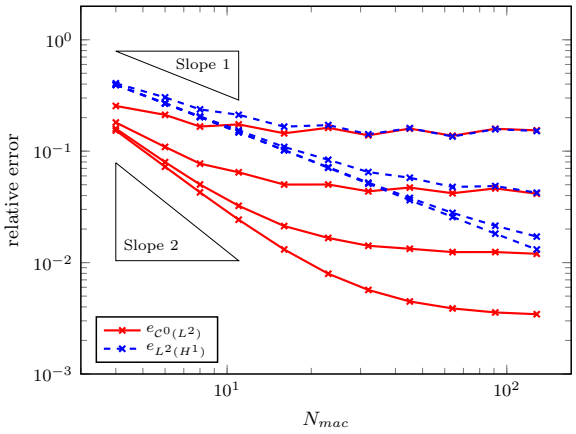
$$e_{\mathcal{C}^0(L^2)} = \max_{0\le n\le N} \left(\sum_{K\in\mathcal{T}_H} \left\| u^0(\cdot,t_n) - u_n^H \right\|^2_{L^2(K)} \right)^{1/2} \left\| u^0 \right\|^{-1}_{\mathcal{C}^0(0,T,L^2(\Omega))}, \tag{61a}$$

$$e_{L^2(H^1)} = \left(\Delta t \sum\nolimits_{n=0}^{'N} \sum_{K\in\mathcal{T}_H} \left\| \nabla u^0(\cdot,t_n) - \nabla u_n^H \right\|^2_{L^2(K)} \right)^{1/2} \left\| u^0 \right\|^{-1}_{L^2(0,T;H_0^1(\Omega))}, \tag{61b}$$

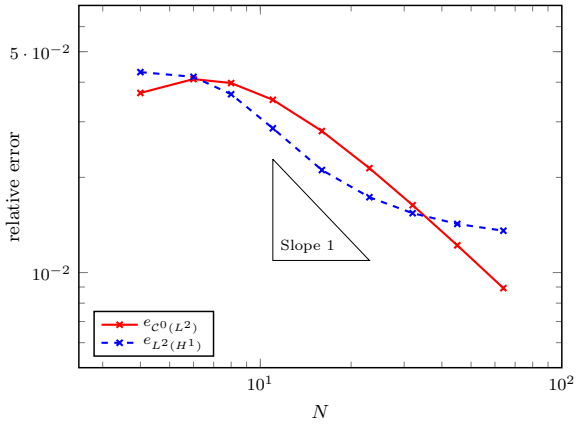
where the $L^2(K)$ norms are evaluated using a higher order quadrature formula and $\sum'$ indicates that the first and the last term of the sum are divided by two (trapezoidal rule).

Numerical results. In Figure 1.(a) we study the space discretization errors from macro and micro FEM. Thus, we take a small time step $\Delta t = 10^{-3}$ and we plot the error measures $e_{\mathcal{C}^0(L^2)}$ and $e_{L^2(H^1)}$ versus N_{mac} behaving like $H \sim 1/N_{mac}$, where H is the macro mesh size. For a fixed micro mesh, with $N_{mic} = 4, 8, 16$ or 32 micro elements in each spatial dimension, we observe quadratic and linear convergence of $e_{\mathcal{C}^0(L^2)}$ and $e_{L^2(H^1)}$ for small N_{mac} and an error saturation for N_{mac} large enough. The saturation levels depend on the micro meshes $\mathcal{T}_h$ and, in particular, those level decrease by a factor around 4 when N_{mic} is doubled, i.e., a quadratic convergence in $h/\varepsilon \sim 1/N_{mic}$ of the micro error. Thus the spatial convergence rates of Theorem 4.2 are confirmed.

In Figure 1.(b) we plot the error measures $e_{\mathcal{C}^0(L^2)}$ and $e_{L^2(H^1)}$ versus the number of time steps N while using fine spatial macro and micro discretizations with $N_{mac} = N_{mic} = 128$. We observe that the error $e_{\mathcal{C}^0(L^2)}$ clearly converges linearly in Δt, while for the $e_{L^2(H^1)}$ already for $N > 16$ a saturation of the error can be found. Further numerical tests show that the spatial macro error is responsible for the saturation of $e_{L^2(H^1)}$ observed in Figure 1.(b). The observations are however still in good agreement with the linear convergence in Δt derived in Theorem 4.2.



(a) Space discretization errors. The different lines correspond to a constant micro mesh $N_{mic} = 4, 8, 16, 32$. Number of time steps $N = 2000$. Macro meshes with $N_{mac} = 4, 6, 8, 11, 16, 23, 32, 45, 64, 91, 128$.



(b) Time discretization error. Macro and micro space discretization with constant meshes $N_{mac} = N_{mic} = 128$. Number of time steps $N = 4, 6, 8, 11, 16, 23, 32, 45, 64$.

Figure 1: Test problem of Section 6.2. Relative error measures $e_{\mathcal{C}^0(L^2)}$ (solid line) and $e_{L^2(H^1)}$ (dashed line), see (61), as a function of N_{mac} (in part (a)) and N (in part (b)), respectively.

6.3 Influence of the sampling domain size δ

In this section, we study the influence of the sampling domain size δ for a nonlinear monotone parabolic test problem of type (2) for Dirichlet and periodic coupling.

For many practical applications, even for periodic data $\mathcal{A}^\varepsilon$, the exact value of the micro period ε is not known exactly. A common strategy is to use Dirichlet coupling $W(K_\delta) = H_0^1(K_\delta)$ and a sampling

domain size $\delta > \varepsilon$ which is larger than some available upper bound of ε. We recall that in this case the modeling error can be estimated (with respect to δ) for locally periodic maps $\mathcal{A}^\varepsilon$, see Theorem 4.3. Further, oversampling techniques can be used to reduce the influence of the wrong boundary conditions, see [25] for an overview. Even for unknown value of ε, one might however still opt to use periodic coupling. While it has not been rigorously proved yet, experimental studies show that periodic coupling yields good results for general $\delta > \varepsilon$ (usually better than Dirichlet coupling), see [42, 7, 5].

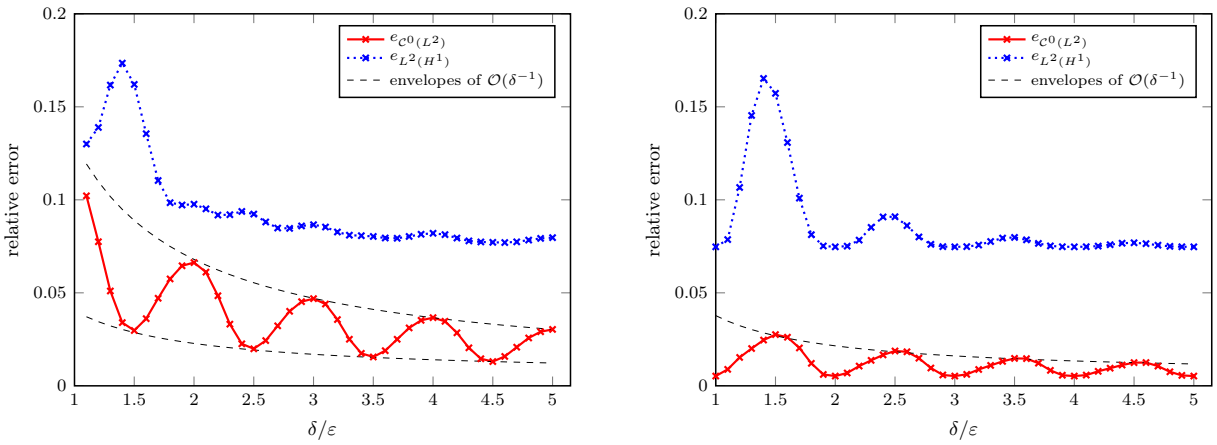
Setting. We modify the test problem of Section 6.2 to get a more significant modeling error in the numerical results by replacing $\Phi(t)$ and $\mathcal{A}_p$ in (58) and (59), respectively, by

$$\Phi(x,t) = \cos(\tfrac{\pi}{2} t x_2), \qquad \mathcal{A}_p(x,y,\xi) = \left[1 + (2 + \sin(2\pi(y_1+y_2)))(1+|\xi|^2)^{-1/2}\right]\xi.$$

The expressions for $c(t;x,y)$ and $f(x,t)$ in (60) are then derived analogously to Section 6.2.

We take $N = 40$ number of time steps and $N_{mac} = 32$ macro elements in each spatial dimension to discretize Ω. We collocate $\mathcal{A}(x, x/\varepsilon, \xi)$ in the slow variable and choose $\varepsilon = 10^{-4}$. To keep the micro error constant for different sampling domain sizes δ we adapt the micro discretization in space such that the micro mesh size $h \sim \delta/N_{mic}$ is constant. For instance we take $N_{mac} = N_{mic}$ for $\delta = \varepsilon$.

Numerical results. In Figure 2.(a), we plot the error measures $e_{\mathcal{C}^0(L^2)}$ and $e_{L^2(H^1)}$ from (61) versus the normalized size δ/ε of the sampling domain K_δ for Dirichlet coupling $W(K_\delta) = H_0^1(K_\delta)$ and sampling domain sizes $\delta_i = (10+i)/10 \cdot \varepsilon$ for $i = 1, \dots, 40$. Globally, both error measures decrease for increasing δ, but locally, an oscillatory behavior with peaks at the resonance values $\delta/\varepsilon \in \mathbb{N}$ can be discovered. Further, the envelopes of order $\mathcal{O}(\delta^{-1})$ for $e_{\mathcal{C}^0(L^2)}$ suggest that the modeling error decays as $\mathcal{O}(\varepsilon/\delta)$ (which is known to be optimal for linear homogenization problems, see [21]) rather than $\mathcal{O}(\sqrt{\varepsilon/\delta})$ as predicted in Theorem 4.3. The result from Theorem 4.3 might thus be non-optimal.



(a) Dirichlet coupling $W(K_\delta) = H_0^1(K_\delta)$ and $1.1 \leq \delta/\varepsilon \leq 5$.

(b) Periodic coupling $W(K_\delta) = W_{per}^1(K_\delta)$ and $1 \leq \delta/\varepsilon \leq 5$.

Figure 2: Test problem of Section 6.3. Relative error measures $e_{\mathcal{C}^0(L^2)}$ (solid line) and $e_{L^2(H^1)}$ (dotted line), see (61), as a function of the sampling domain size δ. Constant number of timesteps $N = 40$ and $N_{mac} = 32$ macro elements per spatial dimension. Microscopic mesh size h chosen such that $h/\varepsilon = H$, i.e., remains constant for different sampling domain sizes δ.

Next, in Figure 2.(b), we present the results for periodic coupling $W(K_\delta) = W_{per}^1(K_\delta)$ and sampling domain size $\delta_i = (10+i)/10 \cdot \varepsilon$ for $i = 0, \dots, 40$. Again, like for the Dirichlet coupling, we discover an oscillating behavior coupled to a global decrease (which is again $\mathcal{O}(\delta^{-1})$ for $e_{\mathcal{C}^0(L^2)}$). In contrast, for periodic coupling, we get optimal accuracy for $\delta/\varepsilon \in \mathbb{N}$ (which is consistent with Theorem 4.3) and the peaks with a locally maximal modeling error at $\delta = (k+1/2)\varepsilon$ for $k \in \mathbb{N}$. Further, comparing the accuracy of the solution obtained for a given δ_i using Dirichlet and periodic coupling reveals that the periodic coupling produces better results for any δ_i, for $1 \leq i \leq 40$. Thus, we observe similar results like for linear elliptic homogenization problems as reported in [42].

6.4 Simulation of a laminated iron core

In this section, we use the multiscale method (8) for a test problem inspired by laminated iron cores, i.e., a layered material. We refer to [32, 33] for more details about the application, where the magnetostatics and magnetodynamics of such laminated cores are studied using a multiscale method based on HMM.

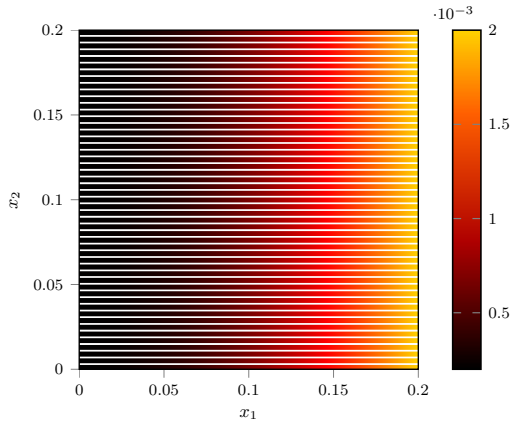
Setting. We consider a layered material in the spatial domain $\Omega = (0, 0.2)^2$ and on the time interval $[0, 2]$. The layered material consists of 51 lamination and 50 insulation layers and is modeled by the locally periodic map $\mathcal{A}^\varepsilon(x, \xi) = \mu^\varepsilon(x, \xi)\xi = \mu(x, x/\varepsilon, \xi)\xi$, see Figure 3.(a), with

$$\mu(x, y, \xi) = \begin{cases} 5000\mu_0(1.03 - \cos(\frac{5\pi}{4}x_1))(1 + |\xi|^2)^{-0.485}, & y \in [0, \frac{3}{4}], \\ \mu_r, & y \in (\frac{3}{4}, 1), \end{cases} \tag{62}$$

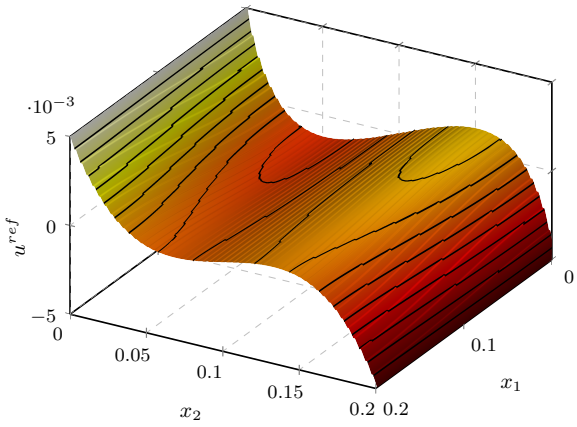
where $\mu_0 = 4\pi \cdot 10^{-7}$ and $\mu_r = 0.05$ represents the permeability of the vacuum and insulation, respectively, and the period is given by $\varepsilon = 1/5 \cdot (50 + 3/4)^{-1}$. Thus, we have constant magnetic permeability μ_r in the insulation layers and a nonlinear constitutive law in the lamination layers. Further, the map $\mathcal{A}^\varepsilon$ is discontinuous in space and degenerates for $|\xi| \to \infty$. We then solve the model problem (2) with Dirichlet conditions on $\Gamma_D = [0, 0.2] \times \{x_2 = 0, 0.2\}$ and homogeneous Neumann conditions on $\Gamma_N = \partial\Omega \setminus \Gamma_D$. The (time-dependent) Dirichlet data $u_D(x_1, x_2, t)$ and the initial conditions $g(x_1, x_2)$ are given by

$$u_D(x_1, 0.2, t) = \tfrac{1}{100}(\cos(\tfrac{\pi}{4}t) - \tfrac{1}{2}), \;\; u_D(x_1, 0, t) = -u_D(x_1, 0.2, t), \qquad g(x_1, x_2) = \tfrac{1}{100}(5x_2 - \tfrac{1}{2}).$$

Reference solution. We compare the results obtained by the multiscale method (8) for the test problem of Section 6.4 to a reference solution u^{ref} calculated by standard FEM combined with the implict Euler integrator. We note that the spatial mesh to compute the reference solution has to resolve the finescale details of the map $\mathcal{A}^\varepsilon$, i.e., the mesh size has to be smaller than the period ε, while the multiscale method (8) solves the problem at cost *independent* of ε. We use a spatial mesh with 10^6 degrees of freedom and 160 equidistant timesteps, see Figure 3.(b) for u^{ref} at final time $T = 2$.



(a) $\mu^\varepsilon(x, \xi)$ for $\xi = (1/5, 1/5)$. Insulation layers depicted as white areas.



(b) Finescale solution at $T = 2$. 10^6 spatial degrees of freedom, 160 timesteps.

Figure 3: Test problem with layered material of Section 6.4. Finescale solution obtained by standard FEM combined with implicit Euler method.

We calculate the error in the spatial L^2 norm (known to be of order $\mathcal{O}(\varepsilon)$ for linear homogenization problems) and compare the energy norms, for $t \in [0, T]$, $1 \le n \le N$,

$$\begin{aligned} \left\|u^{ref}(\cdot, t)\right\|_E^2 &= \tfrac{1}{2}\left\|u^{ref}(\cdot, t)\right\|_{L^2(\Omega)}^2 + \int_0^t \int_\Omega \mathcal{A}^\varepsilon(x, \nabla u^{ref}(x, \tau)) \cdot \nabla u^{ref}(x, \tau) dx d\tau, \\ \left\|u_n^H\right\|_E^2 &= \tfrac{1}{2}\left\|u_n^H\right\|_{L^2(\Omega)}^2 + \Delta t \sum_{k=0}^{'n} \sum_{K \in \mathcal{T}_H} |K| \mathcal{A}_K^{0,h}(\nabla u_k^H(x_K)) \cdot \nabla u_k^H(x_K), \end{aligned}$$

where $\sum'$ indicates again the trapezoidal rule in time. In particular, we use the error measure e_{energy}

$$e_{energy} = \left(\max_{0 \le n \le N} \left|\left\|u^{ref}(\cdot, t_n)\right\|_E - \left\|u_n^H\right\|_E\right|\right)\left(\max_{0 \le n \le N} \left\|u^{ref}(\cdot, t_n)\right\|_E\right)^{-1}. \tag{63}$$

We note, that for linear parabolic homogenization problems, the energy of the finescale solution u^ε converges uniformly on $[0,T]$ to the energy of the homogenized solution, see [15, Section 11.3], while the error in the spatial H^1 norm is at least of order $\mathcal{O}(1)$ due to the small oscillations in u^ε.

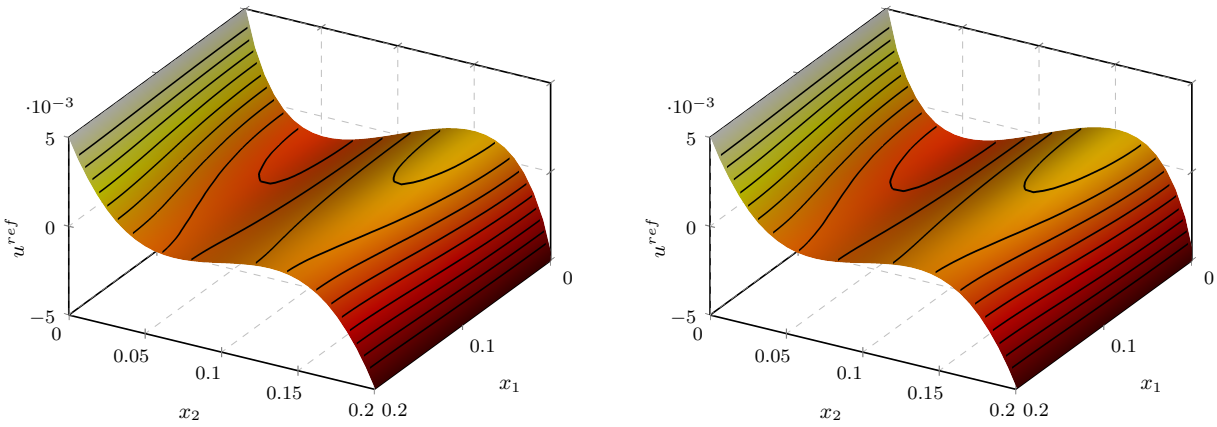
Numerical results. We use the multiscale method (8) on a macro mesh with $N_{mac} = 32$ number of macro elements in each spatial dimension and $N = 160$ equidistant timesteps. As $\mathcal{A}^\varepsilon(x,\xi)$ used in Section 6.4 is locally periodic, we collocate the data (62) in the slow variable x at the quadrature points. We employ Dirichlet coupling $W(K_\delta) = H_0^1(K_\delta)$ for the micro problems (11) for different sampling domain sizes $\delta \geq \varepsilon$. We note that the micro mesh sizes h are chosen such that $h \sim \delta/N_{mic}$ (where N_{mic} is the number of micro elements per spatial dimension) is constant for the different values δ.

In Table 1 we compare the FE-HMM solutions (obtained with Dirichlet coupling and $\delta = 2^k\varepsilon$, $0 \leq k \leq 5$) to the reference solution u^{ref} using $e_{\mathcal{C}^0(L^2)}$ and e_{energy}, introduced in (61a) and (63), respectively. Both error measures decrease monotonically when the sampling domain size δ is increased (the error e_{energy} converges at rate $\mathcal{O}(\delta^{-1})$). Thus, a sufficiently large sampling domain size δ is required for reliable results.

	$\delta = \varepsilon$	$\delta = 2\varepsilon$	$\delta = 4\varepsilon$	$\delta = 8\varepsilon$	$\delta = 16\varepsilon$	$\delta = 32\varepsilon$
$e_{\mathcal{C}^0(L^2)}$	0.4832	0.3224	0.1971	0.1150	0.0682	0.0435
e_{energy}	0.2411	0.1274	0.0657	0.0334	0.0169	0.0083

Table 1: Comparison of the FE-HMM solutions to the standard FEM finescale solution u^{ref} for the test problem of Section 6.4. Study of the influence of the size δ of the sampling domains K_δ for Dirichlet coupling. Error measured by $e_{\mathcal{C}^0(L^2)}$ and e_{energy}, see (61a) and (63), respectively.

For comparison, we then use periodic coupling $W(K_\delta) = W^1_{per}(K_\delta)$ and optimally chosen sampling domain size $\delta = \varepsilon$. In Figure 4, we plot the numerical solutions at final time $T = 2$ for Dirichlet coupling with sampling domain size $\delta = 32\varepsilon$ and for periodic coupling with $\delta = \varepsilon$. Both solutions in Figure 4 capture the effective behavior of the reference solution u^{ref}, see Figure 3.(b). The calculation with periodic coupling however already provides qualitatively good approximations with much less microscopic degrees of freedom.



(a) FE-HMM with Dirichlet coupling, $\delta = 32\varepsilon$ and $N_{mic} = 1024$.

(b) FE-HMM with periodic coupling, $\delta = \varepsilon$ and $N_{mic} = 32$.

Figure 4: Test problem with layered material of Section 6.4. FE-HMM solutions at final time $T = 2$ computed with multiscale method (8) using Dirichlet or periodic coupling. Simulations with $N = 160$ time steps, $N_{mac} = 32$ macro elements per spatial dimension and constant micro error.

7 Conclusion

We have developed and analyzed a multiscale method to solve nonlinear monotone parabolic homogenization problems by combining the implicit Euler integrator (in time) with a numerical homogenization procedure (based on the heterogeneous multiscale method) coupling macro and micro finite element simulations (in space).

Without any structural assumptions on the microscopic heterogeneities, we derived optimal a priori error estimates quantifying the influence of time and space discretization on both macro and micro scale. Further, if we assume local periodicity of the micro structure of the data $\mathcal{A}^\varepsilon$, the modeling error has been explicitely estimated as well. We note that the error analysis can be generalized without any difficulties to different boundary conditions of (2) as well as maps $\mathcal{A}^\varepsilon$ and source terms f smoothly varying in time.

Further, we have shown that the computational cost of the multiscale method is independent of the small characteristic size of the microstructure. Thus, the method is well-suited for practical engineering problems. However, the implementation of the proposed multiscale method still involves systems of nonlinear equations, see Section 6.1. We refer to [6] for an efficient linearized multiscale method in case that the maps $\mathcal{A}^\varepsilon$ can be decomposed as $\mathcal{A}^\varepsilon(x,\xi) = a^\varepsilon(x,\xi)\xi$, with some tensor a^ε.

Acknowledgements. This work was supported in part by the Swiss National Science Foundation under Grant 200021_134716/1.

A Appendix

Lemma A.1. *Let $\mathcal{A}\colon \Omega \times \mathbb{R}^d \to \mathbb{R}^d$ satisfy $\mathcal{A}(x,\cdot) \in \mathcal{C}^1(\mathbb{R}^d;\mathbb{R}^d)$ for a.e. $x \in \Omega$. Then, the map $\mathcal{A}$ satisfies hypotheses $(\mathcal{A}_{1-2})$ if and only if $D_\xi\mathcal{A}(x,\xi)$ is uniformly elliptic and bounded, i.e.,*

$$D_\xi\mathcal{A}(x,\xi)\,\eta\cdot\eta \geq \lambda\,|\eta|^2, \qquad |D_\xi\mathcal{A}(x,\xi)\,\eta| \leq L\,|\eta|, \qquad \forall\,\xi,\eta\in\mathbb{R}^d,\ \text{a.e. } x\in\Omega.$$

Proof. Let $\xi,\eta\in\mathbb{R}^d$. For a.e. $x\in\Omega$, the monotonicity $(\mathcal{A}_2)$ and the regularity of $\mathcal{A}(x,\cdot)$ yield

$$D_\xi\mathcal{A}(x,\xi)\,\eta\cdot\eta = \lim_{t\to 0}\left[(\mathcal{A}(x,\xi+t\eta)-\mathcal{A}(x,\xi))\cdot\eta\right]t^{-1} \geq \lambda\lim_{t\to 0}\left[t\eta\cdot\eta\right]t^{-1} = \lambda|\eta|^2.$$

Similarly, the Lipschitz continuity $(\mathcal{A}_1)$ leads to $|D_\xi\mathcal{A}(x,\xi)\eta| \leq L|\eta|$. The converse is proved using

$$\mathcal{A}(x,\xi)-\mathcal{A}(x,\eta) = \int_0^1 D_\xi\mathcal{A}(x,\eta+\tau(\xi-\eta))(\xi-\eta)d\tau, \qquad \text{a.e. } x\in\Omega. \qquad \square$$

Remark. We note that if $\mathcal{A}(x,\cdot)\in\mathcal{C}^1(\mathbb{R}^d;\mathbb{R}^d)$ the identity

$$\mathcal{A}(x,\xi+\eta) = \mathcal{A}(x,\xi) + D_\xi\mathcal{A}(x,\xi)\eta + \int_0^1 D_\xi\mathcal{A}(x,\xi+\tau\eta) - D_\xi\mathcal{A}(x,\xi)d\tau\,\eta, \tag{64}$$

holds for $\xi,\eta\in\mathbb{R}^d$ and a.e. $x\in\Omega$. If additionally $D_\xi\mathcal{A}(x,\cdot)$ is Lipschitz continuous (with Lipschitz constant L_x), e.g., $\mathcal{A}(x,\cdot)\in W^{2,\infty}(\Omega)$, we obtain

$$\left|\int_\Omega \left[\mathcal{A}(x,\nabla v+\nabla w)-\mathcal{A}(x,\nabla v)-D_\xi\mathcal{A}(x,\nabla v)\cdot\nabla w\right]\cdot\nabla z\,dx\right| \leq L_{\mathcal{A}}\|w\|^2_{W^{1,4}(\Omega)}\|z\|_{H^1(\Omega)}, \tag{65}$$

for $v,w\in W^{1,4}(\Omega)$, $z\in H^1(\Omega)$ and where $L_{\mathcal{A}} = \operatorname{ess\,sup}_{x\in\Omega} L_x$.

References

[1] A. Abdulle, *On a priori error analysis of fully discrete heterogeneous multiscale FEM*, Multiscale Model. Simul., 4 (2005), pp. 447–459.

[2] ——, *The finite element heterogeneous multiscale method: a computational strategy for multiscale PDEs*, in Multiple scales problems in biomathematics, mechanics, physics and numerics, vol. 31 of GAKUTO Internat. Ser. Math. Sci. Appl., Gakkōtosho, Tokyo, 2009, pp. 133–181.

[3] ——, *A priori and a posteriori error analysis for numerical homogenization: a unified framework*, Ser. Contemp. Appl. Math. CAM, 16 (2011), pp. 280–305.

[4] A. Abdulle, W. E, B. Engquist, and E. Vanden-Eijnden, *The heterogeneous multiscale method*, Acta Numer., 21 (2012), pp. 1–87.

[5] A. Abdulle, M. J. Grote, and C. Stohrer, *Finite element heterogeneous multiscale method for the wave equation: long-time effects*. Accepted in Multiscale Model. Simul., 2013.

[6] A. Abdulle, M. E. Huber, and G. Vilmart, *Linearized numerical homogenization methods for nonlinear monotone parabolic multiscale problems.* MATHICSE Technical Report July 2014, École Polytechnique Fédérale de Lausanne.

[7] A. Abdulle and A. Nonnenmacher, *A short and versatile finite element multiscale code for homogenization problems*, Comput. Methods Appl. Mech. Engrg., 198 (2009), pp. 2839–2859.

[8] A. Abdulle and G. Vilmart, *Coupling heterogeneous multiscale FEM with Runge-Kutta methods for parabolic homogenization problems: a fully discrete space-time analysis*, Math. Models Methods Appl. Sci., 22 (2012), pp. 1250002/1–40.

[9] ——, *Analysis of the finite element heterogeneous multiscale method for quasilinear elliptic homogenization problems*, Math. Comp., 83 (2014), pp. 513–536.

[10] J. W. Barrett and W. B. Liu, *Quasi-norm error bounds for the finite element approximation of a non-Newtonian flow*, Numer. Math., 68 (1994), pp. 437–456.

[11] A. Bensoussan, J.-L. Lions, and G. Papanicolaou, *Asymptotic analysis for periodic structures*, North-Holland Publishing Co., Amsterdam, 1978.

[12] S. C. Brenner and L. R. Scott, *The mathematical theory of finite element methods*, vol. 15 of Texts in Applied Mathematics, Springer, New York, third ed., 2008.

[13] P. G. Ciarlet, *The finite element method for elliptic problems*, vol. 4 of Studies in Mathematics and its Applications, North-Holland, 1978.

[14] P. G. Ciarlet and P. A. Raviart, *The combined effect of curved boundaries and numerical integration in isoparametric finite element methods*, in The mathematical foundations of the finite element method with applications to partial differential equations, 1972, pp. 409–474.

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

[16] J. E. Dendy, Jr., *Galerkin's method for some highly nonlinear problems*, SIAM J. Numer. Anal., 14 (1977), pp. 327–347.

[17] M. Dobrowolski, *L^∞-convergence of linear finite element approximation to nonlinear parabolic problems*, SIAM J. Numer. Anal., 17 (1980), pp. 663–674.

[18] J. Douglas, Jr. and T. Dupont, *Galerkin methods for parabolic equations*, SIAM J. Numer. Anal., 7 (1970), pp. 575–626.

[19] R. Du and P. Ming, *Heterogeneous multiscale finite element method with novel numerical integration schemes*, Commun. Math. Sci., 8 (2010), pp. 863–885.

[20] W. E and B. Engquist, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1 (2003), pp. 87–132.

[21] W. E, P. Ming, and P. Zhang, *Analysis of the heterogeneous multiscale method for elliptic homogenization problems*, J. Amer. Math. Soc., 18 (2005), pp. 121–156.

[22] Y. Efendiev and A. Pankov, *Numerical homogenization and correctors for nonlinear elliptic equations*, SIAM J. Appl. Math., 65 (2004), pp. 43–68.

[23] ——, *Numerical homogenization of nonlinear random parabolic operators*, Multiscale Model. Simul., 2 (2004), pp. 237–268.

[24] A. Gloria, *An analytical framework for the numerical homogenization of monotone elliptic operators and quasiconvex energies*, Multiscale Model. Simul., 5 (2006), pp. 996–1043 (electronic).

[25] ——, *Reduction of the resonance error. Part 1: Approximation of homogenized coefficients*, Math. Models Methods Appl. Sci., 21 (2011), pp. 1601–1630.

[26] P. Henning and M. Ohlberger, *A Newton-scheme framework for multiscale methods for nonlinear elliptic homogenization problems*, in Proceedings of the ALGORITMY 2012, 19th Conference on Scientific Computing, Vysoké Tatry, Podbanské, 2012, pp. 65–74.

[27] ——, *Error control and adaptivity for heterogeneous multiscale approximations of nonlinear monotone problems*, Discrete Contin. Dyn. Syst. Ser. S, 8 (2015).

[28] V. H. Hoang, *Sparse finite element method for periodic multiscale nonlinear monotone problems*, Multiscale Model. Simul., 7 (2008), pp. 1042–1072.

[29] P. Houston, J. Robson, and E. Süli, *Discontinuous Galerkin finite element approximation of quasilinear elliptic boundary value problems. I. The scalar case*, IMA J. Numer. Anal., 25 (2005), pp. 726–749.

[30] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik, *Homogenization of differential operators and integral functionals*, Springer-Verlag, Berlin, Heidelberg, 1994.

[31] P. Ming and P. Zhang, *Analysis of the heterogeneous multiscale method for parabolic homogenization problems*, Math. Comp., 76 (2007), pp. 153–177.

[32] I. Niyonzima, R. V. Sabariego, P. Dular, and C. Geuzaine, *Finite element computational homogenization of nonlinear multiscale materials in magnetostatics*, IEEE Transactions on Magnetics, 48 (2012), pp. 587–590.

[33] I. Niyonzima, R. V. Sabariego, P. Dular, F. Henrotte, and C. Geuzaine, *Computational homogenization for laminated ferromagnetic cores in magnetodynamics*, IEEE Transactions on Magnetics, 49 (2013), pp. 2049–2052.

[34] A. Pankov, *G-convergence and homogenization of nonlinear partial differential operators*, vol. 422 of Mathematics and its Applications, Kluwer Academic Publishers, Dordrecht, 1997.

[35] P. A. Raviart, *The use of numerical integration in finite element methods for solving parabolic equations*, in Topics in numerical analysis. Proceedings of the Royal Irish Academy, Conference on Numerical Analysis, 1972, J. J. H. Miller, ed., Academic Press, 1973, pp. 233–264.

[36] N. Svanstedt, *G-convergence of parabolic operators*, Nonlinear Anal., 36 (1999), pp. 807–842.

[37] N. Svanstedt, N. Wellander, and J. Wyller, *A numerical algorithm for nonlinear parabolic equations with highly oscillating coefficients*, Numer. Methods Partial Differential Equations, 12 (1996), pp. 423–440.

[38] L. Tartar, *Cours Peccot*. Collège de France, 1977.

[39] ——, *The general theory of homogenization*, vol. 7 of Lecture Notes of the Unione Matematica Italiana, Springer-Verlag, Berlin; UMI, Bologna, 2009. A personalized introduction.

[40] V. Thomée, *Galerkin finite element methods for parabolic problems*, vol. 25 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, second ed., 2006.

[41] M. F. Wheeler, *A priori L^2 error estimates for Galerkin approximations to parabolic partial differential equations*, SIAM J. Numer. Anal., 10 (1973), pp. 723–759.

[42] X. Yue and W. E, *The local microscale problem in the multiscale modeling of strongly heterogeneous media: effects of boundary conditions and cell size*, J. Comput. Phys., 222 (2007), pp. 556–572.

[43] E. Zeidler, *Nonlinear functional analysis and its applications. II/B*, Springer-Verlag, New York, 1990. Nonlinear monotone operators, Translated from the German by the author and Leo F. Boron.