Numerical homogenization methods for parabolic monotone problems

Assyr Abdulle[1]

Abstract In this paper we review various numerical homogenization methods for monotone parabolic problems with multiple scales. The spatial discretisation is based on finite element methods and the multiscale strategy relies on the heterogeneous multiscale method. The time discretization is performed by several classes of Runge-Kutta methods (strongly $A-$stable or explicit stabilized methods). We discuss the construction and the analysis of such methods for a range of problems, from linear parabolic problems to nonlinear monotone parabolic problems in the very general $L^p(W^{1,p})$ setting. We also show that under appropriate assumptions, a computationally attractive linearized method can be constructed for nonlinear problems.

1 Introduction

Parabolic problems with multiple scales enter in the modelling of a wide range of problems, e.g., thermal diffusion in composite materials, flow problems in heterogeneous medium, etc. We are interested in problems in which the microscopic heterogeneities occur at a much smaller scale than the macroscopic length scale of interest that describes the physical phenomenon of interest. For such problems mathematical homogenization [18, 42] gives the adequate theoretical framework to describe an effective solution originating from the limit of the fine scale solution when the size of the small scales tends to zero. An effective equation for this effective solution can also be established. However, except for special cases, there are no explicit expressions for the effective coefficients (diffusion tensor) of the upscaled equation. The aim of numerical homogenization is to construct computational strategy to compute an approximation of these effective equations and sometimes to capture fine scale

[1] ANMC, Mathematics Section, École Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, assyr.adulle@epfl.ch

oscillations of the multiscale solution. The theory of homogenization is at the root of two classes of numerical methods that we briefly discuss

- methods based on oscillatory basis functions built into a coarse FE space: this idea goes back to Babuška and Osborn [16] and is based on solving local fine scale problems within each macroscopic element of the coarse FE space. Elaboration and generalization have been developed within the multiscale finite element method (MsFEM) [40, 15];
- methods supplementing upscaled data for resolving the effective equation: this idea has been widely used by engineer (see e.g., the references in [34]) and turned into a general framework in the heterogeneous multiscale method (HMM) [29, 3]. In the finite element context, this latter method is called the finite element heterogeneous multiscale method (FE-HMM) and is based on a macroscopic finite element method with input data given by microscopic sampling of the original fine scale problem in patches of size proportional to the fine scale oscillation.

These two classes of methods use either in their formulation or in their analysis the theory of homogenization in an essential way. Further related to homogenization theory we mention the sparse tensor product FEM based on the two-scale convergence theory and its generalization [50, 14] and the projection based numerical homogenization [20, 33] using successive projection of a fine scale discretization of the multiscale equation into a lower dimensional space and iteratively eliminating the fine scale component of the numerical solution.

We also mention multiscale methods that share some similarities with numerical homogenization methods and have been used for homogenization problems. We start with the variational multiscal method [41]. In this approach one starts from a coarse finite element space that cannot resolve the multiscale structure of the fine scale problem. This coarse space is supplemented by a fine scale space and one seeks a numerical solution in the form of a coarse and fine scale components. The fine scale component is obtained by solving localized fine scale problems. Once these problems solved one can solve the coarse scale approximation. Using local quasi-interpolation and an orthogonal decomposition of the coarse and fine spaces, exponential decay of the localisation error has been first proved in [48] (see also [39]). This new approach of the variational multiscale method is called Localised Orthogonal Decomposition (LOD). Finally we also mention methods based on harmonic coordinates [51]. The idea of this method is to compute an appropriate change of coordinates (based on the full fine scale problem) so its composition with the fine scale problem is a slowly varying function that can be approximated in a coarse space. This approach share some similarity with the MsFEM proposed in [15].

In this article, we review several numerical homogenization methods based on the HMM for the solution of the following class of monotone parabolic multiscale problems in a finite time interval $(0,T)$

$$\begin{aligned} \partial_t u^{\varepsilon}(x,t) - \operatorname{div}(\mathscr{A}^{\varepsilon}(x, \nabla u^{\varepsilon}(x,t))) &= f(x) \text{ in } \Omega \times (0,T), \\ u^{\varepsilon}(x,t) = 0 \text{ on } \partial\Omega \times (0,T), \quad u^{\varepsilon}(x,0) &= g(x) \text{ in } \Omega, \end{aligned} \tag{1}$$

with initial source and initial conditions f and g. The maps $\mathscr{A}^\varepsilon\colon \Omega\times\mathbb{R}^d\to\mathbb{R}^d$ are defined on a domain $\Omega\times\mathbb{R}^d$, where $\Omega\subset\mathbb{R}^d$ $d\leq 3$, and $\mathscr{A}^\varepsilon(\cdot,\xi)\colon \Omega\to\mathbb{R}^d$ are Lebesgue measurable for every $\xi\in\mathbb{R}^d$. The indexing by an (abstract) parameter $\varepsilon>0$ indicates that these maps are subject to rapid variations on a very fine scale relative to the size of the domain Ω. For the finite element method we will assume that Ω is a polygonal domain and we will sometimes assume that it is convex. For simplicity neither time dependent source terms $f(x,t)$ or time-dependent maps of the form $\mathscr{A}^\varepsilon(x,t,\nabla u^\varepsilon(x,t))$ are considered but we note that many of the results presented in this review can be extended for these situations.

Let us briefly review the literature on multiscale methods for the parabolic problems (1). For linear problems, most of the methods described above can be used. We mention [31] for MsFEM type methods, [6, 49] for HMM type methods, [47] for LOD type methods. While most of the numerical method have been analysed for the Euler explicit or implicit time discretization, a fully discrete a priori error analysis in space and time for several classes of implicit and explicit Runge-Kutta methods has been given in [12]. For nonlinear monotone parabolic problems, the literature is much more scarce and only methods supplementing upscaled data for resolving the effective equation have been analyzed. In [32] monotone problems with stochastic heterogeneities have been analysed however without convergence rates and for non-discretized micro-problems. In [8, 9] a priori error analysis (in space and time) for two different types of HMM is established under general assumption on the nonlinearity. We close this review by mentioning that for elliptic problems, a posteriori error estimates have been obtained for an HMM type method in the strongly monotone and Lipschitz case in [38] and a priori error estimates for general numerical quadrature methods have been derived in [7]. Finally in [35] numerical homogenization methods (both of HMM and MsFEM types) for monotone PDEs associated to minimization problems have been studied. We note in contrast that for the class of problems (1) discussed in this review, we make no assumptions of an associated scalar potential for $\mathscr{A}^\varepsilon$.

In this paper we aim at reviewing the numerical homogenization methods based on the HMM that have been developed in [12, 8, 9] for parabolic problems (1). We aim at giving a unified description of various error estimates and numerical discretization variant of the FE-HMM

- for linear problem the spatial discretisation based on the FE-HMM is coupled with general classes of Runge-Kutta methods (strongly $A-$stable and explicit stabilized methods), and fully-discrete space-time analysis is proposed for this family of space-time multiscale solvers [12];
- for nonlinear monotone problems a fully discrete space-time method that couples the FE-HMM in space with the backward Euler method in time is shown to converge in the $L^p(W^{1,p})$ and $\mathscr{C}^0(L^2)$ norms towards the homogenized solution u^0 for Problem 1 under the general assumptions. Space-time convergence rates are established for strongly monotone and Lipschitz maps [8];
- for strongly monotone and Lipschitz maps $\mathscr{A}^\varepsilon$ a new linearized scheme that relies only on linear micro and macro finite element (FE) solvers is proposed

and analyzed. A fully discrete space time analysis is also provided for this scheme [9].

We briefly sketch the type of convergence rates that we aim at deriving in this paper: under appropriate assumptions on the tensor $\mathscr{A}^{\varepsilon}$, the family of solutions u^{ε} converges, up to a subsequence, to a homogenized solution u^0 solution of a homogenized equation similar to (1) but with $\mathscr{A}^{\varepsilon}$ replaced by an effective map $\mathscr{A}^0$ that is unknown explicitly (see Section 2). In the context of an FE-HMM method, the goal is to derive an error estimate of the type

$$\max_{1\le n\le N}\left\|u^0(\cdot,t_n)-u_n^H\right\|_{L^2(\Omega)}+\left(\sum_{n=1}^{N}\Delta t\left\|\nabla u^0(\cdot,t_n)-\nabla u_n^H\right\|_{L^2(\Omega)}^2\right)^{1/2} \le C\left[(\Delta t)^r+H^s+\left(\frac{h}{\varepsilon}\right)^q+r_{mod}+\left\|g-u_0^H\right\|_{L^2(\Omega)}\right], \tag{2}$$

where C is independent of $\Delta t, H, h$ and r_{mod}. Here H is the size of a macroscopic triangulation that is used in the FE-HMM to approximate the effective solution u^0 and h is the mesh size of a microscopic triangulation used on a patch K_δ around macroscopic quadrature points. The diameter of the patch K_δ is of size δ typically $\delta=\mathscr{O}(\varepsilon)$. As h must resolve the fine scale oscillation we have $h<\varepsilon\le\delta$. We notice two important facts

- as $h/\varepsilon = 1/N_{mic}$, where N_{mic} is the number of points per oscillation length and the quantity h/ε in the estimate (2) is thus independent of ε and measure the degrees of freedom used to resolve the oscillation; if $\varepsilon\to 0$, so does the patch K_δ hence we solve the fine scale only on small fraction of the macroscopic computational domain and the overall computational cost is independent of ε;
- the quantities, $\Delta t, H, h$ are discretisation parameters while r_{mod} quantifies the error due to the upscaling procedure, i.e., by replacing the true homogenized map $\mathscr{A}^0$ by a map computed from some microscopic models. The coupling condition (periodic, Dirichlet), the size of the sampling domain enter in this modelling error that is not influenced by the macro or micro discretisation parameter H, h. In the most favourable case (e.g., locally periodic homogenization), r_{mod} can be shown to vanish.

In view of the above prototypical error estimate in this paper we will speak about fully discrete spatial error estimates when we have an estimate in terms of both the macroscopic and microscopic spatial mesh H, h and a fully discrete space-time error estimate when we derive an estimate in terms of H, h and Δt.

Several difficulties arise when analyzing a numerical homogenization method: first as the effective data are only available at quadrature points, we necessarily rely on a FEM with numerical quadrature on the macroscale and have to deal with variational crimes. Second, as the upscaled data are obtained from micro solvers (FEM) one has to precisely quantify the propagation of the errors across scales. Finally the modelling error that originates from the averaging procedure designed to recover the effective data need also to be quantified. To close this introduction, we review several important contributions concerning FE methods for single scale nonlinear monotone problems and contrast these results with the numerical homogenization literature. Using quasi-norm techniques, convergence rates have been derived in [17, 27] in the $L^p(W^{1,p})$ setting for single scale parabolic monotone problems with the following p-structure

$$|\mathscr{A}(\xi)-\mathscr{A}(\eta)| \leq L(\kappa_1+|\xi|+|\eta|)^{p-2}|\xi-\eta|,$$
$$(\mathscr{A}(\xi)-\mathscr{A}(\eta))\cdot(\xi-\eta) \geq \lambda(\kappa_2+|\xi|+|\eta|)^{p-2}|\xi-\eta|^2,\ \forall \xi,\eta\in\mathbb{R}^d$$

including for example the p-Laplacian. Note however that under the most general assumptions on the map $\mathscr{A}^\varepsilon$ under which homogenization results are proved (see e.g. [52]) and under which we can show convergence of an FE-HMM method [8], we have have a p-structure if and only if the map $\mathscr{A}^\varepsilon$ is strongly monotone and Lipschitz.

This review is organized as follows. In Section 2 we briefly review the homogenization theory for the class of parabolic problems considered and introduce the numerical homogenization method. In Section 3 we review the coupling of the FE-HMM with various families of Runge-Kutta methods and explain the techniques used to derive a fully-discrete space-time error analysis. Convergence of a fully-discrete numerical method for general nonlinear monotone parabolic problems is discussed in Section 4 and a linearized method is presented in Section 5.

2 Assumptions and homogenization

We consider Problem 1 and the "evolution triple" $W^{1,p}(\Omega)\subset L^2(\Omega)\subset W^{1,p}(\Omega)'$, $f\in L^{p'}(\Omega)$, $g\in L^2(\Omega)$. Very general hypotheses for the maps $\mathscr{A}^\varepsilon$ under which homogenization for (1) can be established, see [21, 52] are the following assumptions assumed to hold uniformly in $\varepsilon>0$ for all $\xi_1,\xi_2\in\mathbb{R}^d$ and almost every $x\in\Omega$. For $1<p<\infty$ and $p>2d/(d+2)$ we assume

($\mathscr{A}_0$) there is some $C_0\geq 0$ such that $|\mathscr{A}^\varepsilon(x,0)|\leq C_0$ for almost every (a.e.) $x\in\Omega$;
($\mathscr{A}_1$) there exist $\kappa_1\geq 0$, $L>0$ and $0<\alpha\leq\min\{p-1,1\}$ such that

$$|\mathscr{A}^\varepsilon(x,\xi_1)-\mathscr{A}^\varepsilon(x,\xi_2)|\leq L(\kappa_1+|\xi_1|+|\xi_2|)^{p-1-\alpha}|\xi_1-\xi_2|^\alpha;$$

($\mathscr{A}_2$) there exist $\kappa_2\geq 0$, $\lambda>0$ and $\max\{2,p\}\leq\beta<\infty$ such that

$$(\mathscr{A}^\varepsilon(x,\xi_1)-\mathscr{A}^\varepsilon(x,\xi_2))\cdot(\xi_1-\xi_2)\geq\lambda(\kappa_2+|\xi_1|+|\xi_2|)^{p-\beta}|\xi_1-\xi_2|^\beta.$$

Then under the assumptions ($\mathscr{A}_{0-2}$) the problem (1) has a unique solution $u^\varepsilon\in E$ for any $\varepsilon>0$

$$E=\{v\in L^p(0,T;W_0^{1,p}(\Omega))\,|\,\partial_t v\in L^{p'}(0,T;(W_0^{1,p}(\Omega))')\}, \tag{3}$$

endowed with the norm $\|v\|_E = \|v\|_{L^p(0,T;W_0^{1,p}(\Omega))} + \|\partial_t v\|_{L^{p'}(0,T;(W_0^{1,p}(\Omega))')}$ (see e.g., [57, Theorem 30.A]).

The aim of homogenization is to find a limiting effective solution for the family of oscillatory solutions $\{u^\varepsilon\}$ and an equation for this effective solution involving a parabolic PDE, where the small scales have been averaged out. We briefly describe this procedure. First, observe that the solution satisfies the bound

$$\|u^\varepsilon\|^p_{L^p(0,T;W_0^{1,p}(\Omega))} + \|\partial_t u^\varepsilon\|^{p'}_{L^{p'}(0,T;(W_0^{1,p}(\Omega))')} \\ \leq C((L_0+\kappa_1+\kappa_2)^p + \|f\|^{p'}_{L^{p'}(\Omega)} + \|g\|^2_{L^2(\Omega)}),$$

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

$$u^\varepsilon \rightharpoonup u^0 \ \text{ in } L^p(0,T;W_0^{1,p}(\Omega)) \quad \text{and} \quad \partial_t u^\varepsilon \rightharpoonup \partial_t u^0 \ \text{ in } L^{p'}(0,T;(W_0^{1,p}(\Omega))') \tag{4}$$

for $\varepsilon \to 0$.

The question answered in the framework of homogenization theory is that of a limiting equation for u^0. For the above parabolic problems, one refers to the so called *G*-convergence of parabolic operators, sometimes called *PG* for strong *G*-convergence ([54, 52]).

The following compactness result can be shown: there exists a subsequence of $\{u^\varepsilon\}$ (still denoted by $\{u^\varepsilon\}$) and a map $\mathscr{A}^0 \colon \Omega \times \mathbb{R}^d \to \mathbb{R}^d$, such that u^ε weakly converges to u^0 in the sense of (4) and the corresponding maps $\mathscr{A}^\varepsilon(x,\nabla u^\varepsilon) \rightharpoonup \mathscr{A}^0(x,\nabla u^0)$ weakly converges in $L^{p'}(0,T;(L^{p'}(\Omega))^d)$. The homogenized solution $u^0 \in E$ is the solution of the following homogenized problem

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

where $\mathscr{A}^0$ satisfies ($\mathscr{A}_{0-2}$) (with possibly different constants $C_0, \kappa_1, \kappa_2, \lambda$ and L) with Hölder exponent $\gamma = \alpha/(\beta-\alpha)$ in ($\mathscr{A}_1$). We note that $\gamma = \alpha$, if and only if $p=2$, $\alpha=1$, $\beta=2$. Convergence of the whole sequence $\{u^\varepsilon\}$ to u^0 can be obtained under additional structure of the maps $\mathscr{A}^\varepsilon$, for example if $\mathscr{A}^\varepsilon(x,\xi) = \mathscr{A}(x/\varepsilon,\xi)$, where $\mathscr{A}(y,\xi)$ is a $Y=(0,1)^d$-periodic function in y. In this case one can also derive a description of $\mathscr{A}^0$ in terms of the solutions of a boundary value problems in the reference domain Y. When the maps $\mathscr{A}^\varepsilon$ depend on both a slow and a fast variable, i.e. $\mathscr{A}(x,x/\varepsilon,\xi)$, the boundary value problems depends on $x \in \Omega$. For completeness we introduce the weak formulation of the homogenized problem, by introducing the map $B^0 \colon W_0^{1,p}(\Omega) \times W_0^{1,p}(\Omega) \to \mathbb{R}$ given by

$$B^0(v;w) = \int_\Omega \mathscr{A}^0(x,\nabla v(x)) \cdot \nabla w(x)dx, \qquad v,w \in W_0^{1,p}(\Omega), \tag{6}$$

We will also sometimes need a discrete weak form based on a quadrature formula $\{x_{K_j}, \omega_{K_j}\}_{j=1}^J$ defined in the next section that reads

$$\hat{B}^0(v^H;w^H) = \sum_{K\in\mathcal{T}_H}\sum_{j=1}^J \omega_{K_j}\mathscr{A}^0(x_{K_j},\nabla v^H(x_{K_j}))\cdot\nabla w^H(x_{K_j}),\ \ v^H,w^H\in S_0^1(\Omega,\mathcal{T}_H), \tag{7}$$

provided $\mathscr{A}^0(\cdot,\xi)$ has a continuous representative for every $\xi\in\mathbb{R}^d$.

2.1 *Multiscale methods: the finite element heterogeneous multiscale methods*

We give in this section a general formulation of the FE-HMM for parabolic problem. The method relies on

- a macroscopic FE method based on a macroscopic spatial discretization of Ω;
- a microscopic solver defined in sampling domains around sampling points $x\in\Omega$, where an approximation of the map $\mathscr{A}^0(x)$ is required;
- a time discretization method.

Macro discretization. Let $\mathcal{T}_H$ be a family of macro partitions of the polygonal domain Ω consisting of conforming, shape-regular meshes with simplicial elements. [2] The macro elements $K\in\mathcal{T}_H$ are open and such that $\cup_{K\in\mathcal{T}_H}\bar{K}=\overline{\Omega}$. Let $\operatorname{diam}K$ be the diameter of $K\in\mathcal{T}_H$ we define by $H=\max_{K\in\mathcal{T}_H}\operatorname{diam}K$ the macroscopic mesh size and consider the macro finite element space

$$S_0^\ell(\Omega,\mathcal{T}_H)=\{v^H\in W_0^{1,p}(\Omega)\,|\,v^H|_K\in\mathscr{P}^\ell(K),\forall K\in\mathcal{T}_H\}, \tag{8}$$

where $\mathscr{P}^\ell(K)$ is the space of polynomials on $K\in\mathcal{T}_H$ of degree at most ℓ. We also consider within each macro element $K\in\mathcal{T}_H$ quadrature points $x_{K_j}\in K$ and weights ω_{K_j} for $j=1,\dots,J$. We assume that $\{x_{K_j},\omega_{K_j}\}_{j=1}^J$ are obtained from a quadrature formula $\{\hat{x}_j,\hat{\omega}_j\}_{j=1}^J$ by $x_{K_j}=F_K(\hat{x}_j),\ \omega_{K_j}=\hat{\omega}_j|\det(\partial F_K)|, j=1,\dots,J$ where F_K is the affine mapping such that $K=F_K(\hat{K})$. We will make the following assumption on the quadrature formula

(Q1) $\int_{\hat{K}}\hat{p}(\hat{x})d\hat{x}=\sum_{j\in J}\hat{\omega}_j\hat{p}(\hat{x}_j),\ \forall\hat{p}(\hat{x})\in\mathscr{P}^\sigma(\hat{K})$, where $\sigma=\max(2\ell-2,\ell)$.

These requirements on the quadrature formula ensure that the optimal convergence rates for elliptic FEM hold when using numerical integration [23].

Multiscale method. The FE-HMM method for parabolic problems can be defined as follows. Find $u^H\in[0,T]\times S_0^\ell(\Omega,\mathcal{T}_H)\to\mathbb{R}$ such that

---

[2] We concentrate on simplicial elements for simplicity but note that many results presented in this paper can be extended to rectangular elements (see for example [12]).

$$
\begin{aligned}
&(\partial_t u^H, v^H) + B_H(u^H, v^H) = (f, v^H) \;\; \forall v^H \in S_0^\ell(\Omega, \mathscr{T}_H) \\
&u^H = 0 \;\text{ on } \partial\Omega \times (0,T) \\
&u^H(x,0) = u_0^H,
\end{aligned} \tag{9}
$$

where

$$
B_H(v^H; w^H) = \sum_{K\in\mathscr{T}_H} \sum_{j=1}^{J} \omega_{K_j} \mathscr{A}_{K_j}^{0,h}(\nabla v^H(x_{K_j})) \cdot \nabla w^H(x_{K_j}) \quad v^H, w^H \in S_0^\ell(\Omega, \mathscr{T}_H) \tag{10}
$$

and $\mathscr{A}_{K_j}^{0,h}(\cdot)$ is a numerically upscaled tensor defined in (13).

Micro solver. We see that for the map B_H in (9), we need to a procedure to recover the effective data $\mathscr{A}_{K_j}^{0,h}(\nabla v^H(x_{K_j}))$. This rely on micro solvers in each sampling domain K_{δ_j}, $j = 1,\dots,J$, associated to a macro element $K \in \mathscr{T}_H$. Let $K_{\delta_j} = x_{X_j} + \delta I$, $I = (-1/2, 1/2)^d$, $\delta \geq \varepsilon$ be discretized by micro meshes $\mathscr{T}_h$ consisting of simplicial elements $T \in \mathscr{T}_h$, with size h is defined by $h = \max_{T\in\mathscr{T}_h} \operatorname{diam} T$. We then consider the micro finite element spaces

$$
S^q(K_{\delta_j}, \mathscr{T}_h) = \{v^h \in W(K_{\delta_j}) \,|\, v^h|_T \in \mathscr{P}^q(T), \forall T \in \mathscr{T}_h\}, \tag{11}
$$

where $\mathscr{P}^q(T)$ is the space of linear polynomials on $T \in \mathscr{T}_h$ and $W(K_{\delta_j}) \subset W^{1,p}(K_{\delta_j})$ is some Sobolev space. The choice of the space $W(K_{\delta_j})$ sets the coupling condition between the macro and micro solver, e.g.,

- $W(K_{\delta_j}) = \mathscr{W}_{per}^{1,p}(K_{\delta_j}) = \{v \in W_{per}^{1,p}(K_{\delta_j}) \,|\, \int_{K_{\delta_j}} v\,dx = 0\}$ (periodic coupling);
- $W(K_{\delta_j}) = W_0^{1,p}(K_{\delta_j})$ (Dirichlet coupling).

For $\xi \in \mathbb{R}^d$ and $K_{\delta_j} \subset K \in \mathscr{T}_H$, we introduce the function $\chi_{K_j}^{\xi,h}$ as the solution to the variational problem: find $\chi_{K_j}^{\xi,h} \in S^q(K_{\delta_j}, \mathscr{T}_h)$ such that

$$
\int_{K_{\delta_j}} \mathscr{A}^\varepsilon(x, \xi + \nabla\chi_{K_j}^{\xi,h}) \cdot \nabla w^h \, dx = 0, \qquad \forall w^h \in S^q(K_{\delta_j}, \mathscr{T}_h). \tag{12}
$$

Based on the functions $\chi_{K_j}^{\xi,h}$ we can compute the effective data by

$$
\mathscr{A}_{K_j}^{0,h}(\xi) = \frac{1}{\left|K_{\delta_j}\right|} \int_{K_{\delta_j}} \mathscr{A}^\varepsilon(x, \xi + \nabla\chi_{K_j}^{\xi,h}) dx. \tag{13}
$$

We also define an auxiliary flux useful for the analysis

$$
\bar{\mathscr{A}}_{K_j}^{0}(\xi) = \frac{1}{\left|K_{\delta_j}\right|} \int_{K_{\delta_j}} \mathscr{A}^\varepsilon(x, \xi + \nabla\bar{\chi}_{K_j}^{\xi}) dx, \tag{14}
$$

where $\bar{\chi}^{\xi}_{K_j} \in W(K_{\delta_j})$ solve (12) in the infinite dimensional space $W(K_{\delta_j})$.

Upscaling error. We define the upscaling error, called r_{HMM}, as the total error made by approximating the effective flux $\mathscr{A}^0$ by the numerics flux $\mathscr{A}^{0,h}_{K_j}$, precisely for any $v^H \in S^{\ell}_0(\Omega, \mathscr{T}_H)$ we define

$$r_{HMM}(\nabla v^H) = \left(\sum_{K\in\mathscr{T}_H} \sum_{j=1}^{J} \omega_{K_j} \left| \mathscr{A}^0(x_{K_j}, \nabla v^H(x_{K_j})) - \mathscr{A}^{0,h}_{K_j}(\nabla v^H(x_{K_j})) \right|^{p'} \right)^{\frac{1}{p'}}, \tag{15}$$

where $p' = p/(p-1)$ is the dual exponent of $1 < p < \infty$. Thanks to the auxiliary flux, we can further decompose r_{HMM} into two components

$$r_{mic}(\nabla v^H) = \left(\sum_{K\in\mathscr{T}_H} \sum_{j=1}^{J} \omega_{K_j} \left| \bar{\mathscr{A}}^0_{K_j}(\nabla v^H(x_{K_j})) - \mathscr{A}^{0,h}_{K_j}(\nabla v^H(x_{K_j})) \right|^{p'} \right)^{\frac{1}{p'}}, \tag{16a}$$

$$r_{mod}(\nabla v^H) = \left(\sum_{K\in\mathscr{T}_H} \sum_{j=1}^{J} \omega_{K_j} \left| \mathscr{A}^0(x_{K_j}, \nabla v^H(x_{K_j})) - \bar{\mathscr{A}}^0_{K_j}(\nabla v^H(x_{K_j})) \right|^{p'} \right)^{\frac{1}{p'}}. \tag{16b}$$

We observe that using the Minkowski inequality we get $r_{HMM}(\nabla v^H) \leq r_{mic}(\nabla v^H) + r_{mod}(\nabla v^H)$ for every $v^H \in S^1_0(\Omega, \mathscr{T}_H)$. The first term $r_{mic}(\nabla v^H)$ quantifies the error made by solving the micro problems (12) in $S^q(K_{\delta_j}, \mathscr{T}_h)$. The second term r_{mod} quantifies the error due to the upscaling procedure used to replace the true homogenized flux $\mathscr{A}^0$ by (14). The coupling condition (periodic, Dirichlet), the size of the sampling domain enter in this modelling error that is not influenced by the macro or micro discretisation parameter H and h. In the most favourable case (e.g., locally periodic homogenization), when $\delta/\varepsilon \in \mathbb{N}^*$ and periodic coupling is used we can have $r_{mod}(\nabla v^H) = 0$ (see [11]).

Existence and uniqueness of the micro nonlinear problem. The assumptions ($\mathscr{A}_0$)-($\mathscr{A}_2$) are sufficient to guarantee existence and uniqueness of a solution to the nonlinear problem (12). To treat both the exact and the FE approximation of this nonlinear problem we consider the more general problem: find $z \in X$ such that

$$a^{\xi}_{K_j}(z; w) := \int_{K_{\delta_j}} \mathscr{A}^{\varepsilon}(x, \xi + \nabla z) \cdot \nabla w \, dx = 0, \qquad \forall w \in X, \tag{17}$$

where X is any closed linear subspace of the Banach space $W(K_{\delta_j})$.

Lemma 1. *Assume that $\mathscr{A}^{\varepsilon}$ satisfies ($\mathscr{A}_{0-2}$). Then problem* (17) *has a unique solution.*

Sketch of the proof. Unless specified otherwise, all the constants below depend on $\kappa_1, |K_\delta|, \xi, L$ and C_0 (see ($\mathscr{A}_{0-2}$)). Using a Hölder inequality and ($\mathscr{A}_0$) yield for any $z \in X$ the bound $|a^{\xi}_K(z; w)| \leq C(z)\|w\|_{L^p(K_\delta)}$ for a constant C depending on z, hence the nonlinear operator $M : X \to X^*$ defined by $\langle Mz, w\rangle = a_K^{\xi}(z;w)$ is well-defined and the problem (17) is equivalent to the problem $Mz = 0$. We next list the properties of the operator M:

1) Using $(\mathscr{A}_1)$ and a Hölder inequality yields

$$\|Mz - Mw\|_{X^*} \le C(\|z\|_{L^p(K_{\delta_j})}, \|w\|_{L^p(K_{\delta_j})})\|z-w\|^{\alpha}_{L^p(K_{\delta_j})},$$

and M is continuous.

2) Thanks to $(\mathscr{A}_2)$ we have $\langle Mz - Mw, z - w\rangle > 0$ for all $z \neq w$ and the operator M is strictly monotone.

3) Finally the bound [26, Lemma 3.1]

$$\|\nabla z - \nabla w\|_{L^p(K_{\delta_j})} \le \left[\kappa \left|K_{\delta_j}\right|^{\frac{1}{p}} + \|\nabla z\|_{L^p(K_{\delta_j})} + \|\nabla w\|_{L^p(K_{\delta_j})}\right]^{\frac{\beta-p}{\beta}} \left(\int_{K_{\delta_j}} (\kappa + |\nabla z| + |\nabla w|)^{p-\beta} |\nabla z - \nabla w|^{\beta} dx\right)^{\frac{1}{\beta}},$$

for any $z, w \in X$ that holds for $1 < p < \infty$, $\beta \ge p$ and $\kappa \ge 0$ together with $(\mathscr{A}_2)$ yields

$$\langle Mz, z\rangle \ge C_1 \|\nabla z\|^p_{L^p(K_{\delta_j})} - C_2$$

where C_1, C_2 in addition also depends on $\kappa_2, \beta, p, \lambda$ and the operator M is coercive. Hence we can apply the Browder-Minty theorem that ensure that the equation $Mz = 0$ with the operator M that continuous, strictly monotone and coercive, has a unique solution. □

We next list several properties of the map $B_H(v^H; w^H)$ that follows from the assumption $\mathscr{A}^{\varepsilon}$ (we refer to [8] for a detailed derivation).

Lemma 2. *Assume that $\mathscr{A}^{\varepsilon}$ satisfies $(\mathscr{A}_{0-2})$. Let $v^H, w^H, z^H \in S_0^{\ell}(\Omega, \mathscr{T}_H)$ then the nonlinear map B_H defined in* (10) *satisfies the following properties*

$$\left|B_H(v^H; w^H)\right| \le C\left[C_d + \left\|\nabla v^H\right\|_{L^p(\Omega)}\right]^{p-1} \left\|\nabla w^H\right\|_{L^p(\Omega)}, \tag{18}$$

$$\left|B_H(v^H; z^H) - B_H(w^H; z^H)\right| \le C\left[C_d + \left\|\nabla v^H\right\|_{L^p(\Omega)} + \left\|\nabla w^H\right\|_{L^p(\Omega)}\right]^{p-1-\gamma} \left\|\nabla v^H - \nabla w^H\right\|^{\gamma}_{L^p(\Omega)} \left\|\nabla z^H\right\|_{L^p(\Omega)}, \tag{19}$$

$$B_H(v^H; v^H - w^H) - B^H(w^H; v^H - w^H) > 0 \quad \text{for } v^H \neq w^H \tag{20}$$

$$B_H(v^H; v^H) \ge \lambda_c \left\|\nabla v^H\right\|^p_{L^p(\Omega)} - C(C_d)^p, \tag{21}$$

where C may depend on on p, α, β, λ, L and the measure of Ω, with $\lambda_c > 0$ depending only on p, β, λ and L and $C_d = L_0 + \kappa_1 + \kappa_2$, $\gamma = \alpha/(\beta - \alpha)$.

The above properties are sufficient to guarantee the existence and uniqueness of a solution to the problem (9). We note that while (20) is sufficient to ensure the strict monotonicity of $B_H(\cdot,\cdot)$ for the error estimate we will need the following monotonicity estimate

$$\left\|\nabla v^H - \nabla w^H\right\|_{L^p(\Omega)} \leq C\left[1 + \left\|\nabla v^H\right\|_{L^p(\Omega)} + \left\|\nabla w^H\right\|_{L^p(\Omega)}\right]^{\frac{\beta-p}{\beta}} \times \left(B^H(v^H; v^H - w^H) - B^H(w^H; v^H - w^H)\right)^{\frac{1}{\beta}}, \qquad (22)$$

where C depends on $C_d, \lambda_c, |\Omega|, p$ and β.

Theorem 1. *Assume that* $(\mathscr{A}_{0-2})$ *hold and that* $f \in L^{p'}(\Omega)$. *Then, for any parameter* $H, h, \delta > 0$*, there exists a unique numerical solution of* (9) *that satisfies*

$$\|u^H\|_{L^p(W_0^{1,p})} \leq C, \qquad \|u^H\|_{C^0(L^2)} \leq C, \qquad (23)$$

where C is independent of H, h, ε.

Proof. The map $B : S_0^\ell(\Omega, \mathscr{T}_H) \to S_0^\ell(\Omega, \mathscr{T}_H)$, defined by $\left\langle Bv^H, w^H \right\rangle = B_H(v^H; w^H)$ is (strictly) monotone (20), hemicontinuous (the map $v^H \to \left\langle Bv^H, w^H \right\rangle$ is continuous for all $w^H \in S_0^\ell(\Omega, \mathscr{T}_H)$ thanks to (19)), coercive (21) and satisfies a growth condition $\|Bv\|_{(W_0^{1,p})^*} \leq c_1 + c_2\|v^h\|^{p-1}_{W_0^{1,p}}$. Hence the ordinary differential equation (9) satisfies the hypothesis of the Caratheodory theorem that guarantees the existence and uniqueness of a solution [57, Lemma 30.4]. The monotonicity and coercivity of B yield the a priori bound. □

3 Fully discrete space-time error estimates for linear parabolic problems

In this section we consider linear parabolic multiscale problems for which $\mathscr{A}^\varepsilon(x,\xi) = a^\varepsilon(x)\xi$. We assume $a^\varepsilon(x) \in (L^\infty(\Omega))^{d\times d}$ and for all $\xi \in \mathbb{R}^d$ and a.e. $x \in \Omega, t \in (0,T)$ there exists $\lambda, L > 0$ such that, uniformly for all $\varepsilon > 0$

$$\lambda|\xi|^2 \leq a^\varepsilon(x)\xi \cdot \xi, \; |a^\varepsilon(x)\xi| \leq L|\xi|. \qquad (24)$$

The maps $\mathscr{A}^\varepsilon$ then satisfy $(\mathscr{A}_{0-2})$ for $p = 2$, $\alpha = 1$, $\beta = 2$ and with constants $C_0 = 0$ and λ, L given by the ellipticity and continuity constants. For simplicity we consider tensors $a^\varepsilon(x)$ independent of time but all the results of this section can be generalised for time dependent tensors [12]. The numerical method that we consider is still given by (9) but we have now the following explicit expression for the flux

$$\mathscr{A}^{0,h}_{K_j}(\xi) = \frac{1}{\left|K_{\delta_j}\right|}\int_{K_{\delta_j}} a^\varepsilon(x)(\xi + \nabla\chi^{\xi,h}_{K_j})dx.$$

Now since $\nabla v^H = \sum_{i=1}^d \mathbf{e}_i \partial_i v^H$, where $\mathbf{e}_i$ $i = 1,\ldots,d$ is the canonical basis of $\mathbb{R}^d$, it is easy to see that $\mathscr{A}_{K_j}^{0,h}(\nabla v^H(x_{K_j})) = a^{0,h}(x_{K_j})\nabla v^H(x_{K_j})$, where the $i-$th row of the matrix $a^{0,h}(x_{K_j})$ is given by

$$a^{0,h}(x_{K_j}) = \frac{1}{\left|K_{\delta_j}\right|}\int_{K_{\delta_j}} a^{\varepsilon}(x)(I+\nabla\chi_{K_j}^h)dx. \tag{25}$$

Here I is the $d\times d$ identity matrix and $\chi_{K_j}^h$ is a $d\times d$ matrix with column given by $\nabla\chi_{K_j}^{\mathbf{e}_i,h}$, where $\chi_{K_j}^{\mathbf{e}_i,h}$ are the (linear) solution of (12). We can thus rewrite the bilinear form (10) as

$$B_H(v^H,w^H) = \sum_{K\in\mathscr{T}_H}\sum_{j=1}^{J}\omega_{K_j}a^{0,h}(x_{K_j})\nabla v^H(x_{K_j})\cdot\nabla w^H(x_{K_j}), \tag{26}$$

for all $v^H, w^H \in S_0^\ell(\Omega,\mathscr{T}_H)$. We will also use below the bilinear form

$$B_{0,H}(v^H,w^H) = \sum_{K\in\mathscr{T}_H}\sum_{j=1}^{J}\omega_{K_j}a^{0}(x_{K_j})\nabla v^H(x_{K_j})\cdot\nabla w^H(x_{K_j}), \tag{27}$$

where a^0 is the (usually unknown) exact homogenized tensor that is known to satisfy similar bound (24) as a^ε. The solution of the homogenized problem (5) will be denoted by $u^0(t)$ and the corresponding bilinear by

$$B(v,w) = \int_\Omega a^0(x)\nabla v\cdot\nabla w dx. \tag{28}$$

We next mention classical estimates for FEM with numerical quadrature that are needed in the analysis below [23, Thms 4,5]. Assuming (**Q1**) and appropriate regularity of the homogenised solution u^0 we have for all $v^H, w^H \in S_0^\ell(\Omega,\mathscr{T}_H)$ (where $\mu = 0$ or 1),

$$|B(v^H,w^H)-B_{0,H}(v^H,w^H)| \le CH\|v^H\|_{H^1(\Omega)}\|w^H\|_{H^1(\Omega)}, \tag{29}$$

$$|B(\mathscr{I}_H u_0,w^H)-B_{0,H}(\mathscr{I}_H u_0,w^H)| \le CH^\ell\|u_0(t)\|_{W^{\ell+1,p}(\Omega)}\|w^H\|_{H^1(\Omega)}, \tag{30}$$

$$|B(\mathscr{I}_H u_0,w^H)-B_{0,H}(\mathscr{I}_H u_0,w^H)| \le CH^{\ell+\mu}\|u_0(t)\|_{W^{\ell+1,p}(\Omega)} \tag{31}$$

$$\cdot\Big(\sum_{K\in\mathscr{T}_H}\|w^H\|^2_{H^2(K)}\Big)^{1/2}, \tag{32}$$

where $\mathscr{I}_H : C^0(\overline{\Omega}) \to S_0^\ell(\Omega,\mathscr{T}_H)$ is the usual nodal interpolant.

For linear parabolic problems, we can derive fully discrete convergence results in both space and time. Furthermore we can perform this a priori convergence analysis for various class of time integrators including "explicit stabilized Runge-Kutta method". The strategy is to first derive fully discrete error estimates in space. In a second step, using semigroup techniques, fully-discrete space time error estimates can be obtained. In contrast fully discrete space-time estimates could be obtained at once starting directly from a time-discrete numerical method instead of first considering (9). With such a strategy we need however new error estimates for each new time-integrator while with the former approach we can derive error estimates for classes of time integrators "at once". In this section we follow the finding of [12].

3.1 Fully discrete a priori convergence rates in space

The starting point of the analysis is to define an appropriate elliptic projection: for all $t \in (0,T)$, let $\Pi_H u^0(t) \in S_0^\ell(\Omega,\mathscr{T}_H)$ be the solution of the problem

$$B_H(\Pi_H u^0(t), z^H) = B(u^0(t), z^H), \qquad \forall z^H \in S_0^\ell(\Omega,\mathscr{T}_H), \quad t \in (0,T), \tag{33}$$

where $u^0(t)$ is the solution of the homogenized problem (5). Thanks to the ellipticity and continuity of B_H, the above problem is well-posed. Using (33), denoting by $\mathscr{I}_H u^0$ the standard nodal interpolant of u^0 we get for all $z^H \in S_0^\ell(\Omega,\mathscr{T}_H)$,

$$\begin{aligned} B_H(\Pi_H u^0 - \mathscr{I}_H u^0, z^H) &= B(u^0 - \mathscr{I}_H u^0, z^H) \\ &+ B(\mathscr{I}_H u^0, z^H) - B_{0,H}(\mathscr{I}_H u^0, z^H) \\ &+ B_{0,H}(\mathscr{I}_H u^0, z^H) - B_H(\mathscr{I}_H u^0, z^H). \end{aligned} \tag{34}$$

Assuming enough regularity of the homogenized solution, the first two terms of the above inequality are bounded by $CH^\ell \|u_0(t)\|_{W^{\ell+1,p}} \|z^H\|_{H^1(\Omega)}$ using the continuity of B and standard results for nodal interpolant [22] (first term) and (30) (second term). In view of (26) and (27), the definition (15) for $p = p' = 2$ and the assumption (**Q1**) on the quadrature formula we have for the third term

$$|B_{0,H}(\mathscr{I}_H u^0, z^H) - B_H(\mathscr{I}_H u^0, z^H)| \le r_{HMM}(\nabla \mathscr{I}_H u^0) \|\nabla z^H\|_{L^2(\Omega)}. \tag{35}$$

We note that we can further decompose $r_{HMM}(\nabla \mathscr{I}_H u^0)$ as

$$r_{HMM}(\nabla \mathscr{I}_H u^0) \le \sup_{K\in\mathscr{T}_H, x_{K_j}\in K} \|a^0(x_{K_j}) - a^{0,h}(x_{K_j})\|_F \|\nabla \mathscr{I}_H u^0\|_{L^2(\Omega)},$$

where $\|\cdot\|_F$ denotes the Frobenius norm of a matrix. We first have the following lemma.

Lemma 3. *Let $u^0(t)$ be the solution of* (5) *and $\Pi_H u^0(t)$ be the elliptic projection defined in* (33)*. Assume that* ($\mathscr{A}_{0-2}$) *and* (**Q1**) *hold. Assume further that the homogenized tensor satisfies $a^0_{ij} \in \mathscr{C}^0([0,T]\times\bar{K})$ for all $K \in \mathscr{T}_H$ and all $i,j = 1,\ldots,d$. Assume further for $\mu = 0$ or 1 and $\ell > d/p$ that*

$$u_0, \partial_t u_0 \in L^2(0,T;W^{\ell+1,p}(\Omega)),$$
$$a^0_{ij}, \partial_t a^0_{ij} \in L^\infty(0,T;W^{\ell+\mu,\infty}(\Omega)), \quad \forall i,j = 1\dots d.$$

Then we have for $k = 0,1$

$$\|\partial_t^k \left(\Pi_H u^0 - u^0\right)\|_{L^2(0,T;H^1(\Omega))} \le C(H^\ell + r_{HMM}(\nabla \mathscr{I}_H u^0)), \tag{36}$$
$$\|\partial_t^k \left(\Pi_H u^0 - u^0\right)\|_{L^2(0,T;L^2(\Omega))} \le C(H^{\ell+\mu} + r_{HMM}(\nabla \mathscr{I}_H u^0)) \ \mu = 0,1, \tag{37}$$

where we assume that Ω *is convex for the estimates* (37) *with* $\mu = 1$*. The constant* C *is independent of* H, h *and* δ*.*

Proof. In view of (34) and the bound of the different terms of the right-hand side of this equality, taking $z^H = \Pi_H u^0 - \mathscr{I}_H u^0$, using the ellipticity of B_H and integrating from 0 to T we obtain $\|\Pi_H u^0 - \mathscr{I}_H u^0\|_{L^2(0,T;H^1(\Omega))} \le C(H^\ell + r_{HMM}(\nabla \mathscr{I}_H u^0)))$. The estimate (36) for $k = 0$ follows by using the triangle inequality and the estimate $\|u^0 - \mathscr{I}_H u^0\|_{L^2(0,T;H^1(\Omega))} \le CH^\ell$. The estimate (36) for $k = 1$ is obtained by differentiating (34) and following the same arguments.

For the estimate (37) $k = 0$ we use the classical Aubin-Nitsche duality argument and consider for almost every $t \in (0,T)$ the solution $\varphi(t) \in H^1_0(\Omega)$ of the problem

$$B(z, \varphi(t)) = (v(t), z), \quad \forall z \in H^1_0(\Omega). \tag{38}$$

Taking $v(t) = z = \Pi_H u^0 - u^0$ using the elliptic projection (33) yields for all φ^H

$$\begin{aligned}(\Pi_H u^0 - u^0, \Pi_H u^0 - u^0) &= B(\Pi_H u^0 - u^0, \varphi - \varphi^H)\\ &+ B(\Pi_H u^0 - \mathscr{I}_H u^0, \varphi^H) - B_H(\Pi_H u^0 - \mathscr{I}_H u^0, \varphi^H)\\ &+ B(\mathscr{I}_H u^0, \varphi^H) - B_H(\mathscr{I}_H u^0, \varphi^H).\end{aligned} \tag{39}$$

We take $\varphi^H = \mathscr{I}_H \varphi(t)$ use the continuity of B, (29) and (32) to obtain

$$\begin{aligned}(\Pi_H u^0 - u^0, \Pi_H u^0 - u^0) &\le C(H + r_{HMM}(\nabla \mathscr{I}_H u^0))\\ &\cdot \ \|\Pi_H u^0(t) - u_0(t)\|_{H^1(\Omega)} \|\varphi(t)\|_{H^2(\Omega)}\\ &+ \ (H^{\ell+\mu} + r_{HMM}(\nabla \mathscr{I}_H u^0))) \|u(t)\|_{H^{\ell+1}(\Omega)} \|\varphi(t)\|_{H^2(\Omega)}.\end{aligned}$$

Using $\|\varphi\|_{L^2(0,T;H^2(\Omega))} \le C\|\Pi_H u^0 - u^0\|_{L^2(0,T;L^2(\Omega))}$ that holds thanks to the regularity $a \in (L^\infty(0,T;W^{1,\infty}(\Omega)))^{d\times d}$ of the tensor and the convexity of the polygonal domain Ω gives (37) for $k = 0$. The estimate (37) for $k = 1$ is obtained by differentiating (39) and following the same arguments. □

Remark 1. Under the assumptions of Lemma 3 the Sobolev embedding $H^1(0,T;X)$ into $\mathscr{C}^0([0,T];X)$ (for a given Banach space) allows to deduce

$$\|\Pi_H u^0 - u^0\|_{\mathscr{C}^0(0,T;H^1(\Omega))} \le C(H^\ell + r_{HMM}(\nabla \mathscr{I}_H u^0))), \tag{40}$$
$$\|\Pi_H u^0 - u^0\|_{\mathscr{C}^0(0,T;L^2(\Omega))} \le C(H^{\ell+\mu} + r_{HMM}(\nabla \mathscr{I}_H u^0))). \tag{41}$$

We state now fully discrete a priori convergence rate in space for the FE-HMM

Theorem 2. *Let $u^0(t)$ be the solution of* (5) *and u^H the solutions of* (9). *Assume the hypotheses of Lemma 3. Then we have the $L^2(H^1)$ and $\mathscr{C}^0(L^2)$ estimates*

$$\|u^0-u^H\|_{L^2([0,T];H^1(\Omega))} \le C(H^\ell + r_{HMM}(\nabla \mathscr{I}_H u^0) + \|g-u_0^H\|_{L^2(\Omega)}), \quad (42)$$

and if $\mu = 1$

$$\|u^0-u^H\|_{\mathscr{C}^0([0,T];L^2(\Omega))} \le C(H^{\ell+1} + r_{HMM}(\nabla \mathscr{I}_H u^0) + \|g-u_0^H\|_{L^2(\Omega)}). \quad (43)$$

If in addition, the tensor is symmetric, then we have the $\mathscr{C}^0(H^1)$ estimate

$$\|u^0-u^H\|_{\mathscr{C}^0([0,T];H^1(\Omega))} \le C(H^\ell + r_{HMM}(\nabla \mathscr{I}_H u^0) + \|g-u_0^H\|_{H^1(\Omega)}). \quad (44)$$

The constants C are independent of $H, r_{HMM}(\nabla \mathscr{I}_H u^0)$.

Proof. To simplify the notation, we use $r_{HMM} = r_{HMM}(\nabla \mathscr{I}_H u^0)$ in the proof.
Step 1: Estimation of $\|u^H - \Pi_H u^0\|_{L^2(0,T;H^1(\Omega))} + \|u^H - \Pi_H u^0\|_{\mathscr{C}^0([0,T];L^2(\Omega))}$.
We set $\xi^H(t) = u^H(t) - \Pi_H u^0(t), t \in [0,T]$. In view of the elliptic projection (33), (5) and (9) we have for all $z^H \in S_0^\ell(\Omega, \mathscr{T}_H)$,

$$(\partial_t \xi^H, z^H) + B_H(\xi^H, z^H) = (\partial_t u^0, z^H) - (\partial_t \Pi_H u^0, z^H). \quad (45)$$

We set $z^H = \xi^H$ integrate this equality from 0 to t using the coercivity of $B_H(\cdot,\cdot)$

$$\|\xi^H(t)\|_{L^2(\Omega)}^2 + c_1 \int_0^t \|\xi^H(s)\|_{H^1(\Omega)}^2 ds \le \|\xi^H(0)\|_{L^2(\Omega)}^2 \quad (46)$$

$$+ \, c_2 \int_0^t \|\partial_t u^0 - \partial_t \Pi_H u^0\|_{L^2(\Omega)}^2 ds. \quad (47)$$

Using the decomposition $\xi(0) = (u^0 - \Pi_H u^0)(0) + (u_0^H - g)$, (41) and (37) yields

$$\|\xi(0)\|_{L^2(\Omega)} \le C(H^{\ell+\mu} + r_{HMM}) + \|u_0^H - g\|_{L^2(\Omega)}. \quad (48)$$

Using (48) and (37) gives the $L^2(H^1)$ estimate and taking the supremum with respect to t gives the $\mathscr{C}^0(L^2)$ estimate. We thus obtain

$$\begin{aligned}\|u^H - \Pi_H u^0\|_{\mathscr{C}^0([0,T];L^2(\Omega))} \, &+ \, \|u^H - \Pi_H u^0\|_{L^2(0,T;H^1(\Omega))} \\ &\le C(H^{\ell+\mu} + r_{HMM}) + \|u_0^H - g\|_{L^2(\Omega)}.\end{aligned} \quad (49)$$

This last estimate together with Lemma 3 and the triangle inequality gives (42) and (43).

Step 2: Estimation of $\|u^H - \Pi_H u^0\|_{\mathscr{C}^0([0,T];H^1(\Omega))}$.
For $\xi^H(t) = u^H(t) - \Pi_H u^0(t), t \in [0,T]$, we set $z^H = \partial_t \xi^H$ in (45). Using the symmetry of the tensor, and integrating from 0 to t, we obtain for $0 \le t \le T$

$$2\int_0^t \|\partial_t \xi^H(s)\|^2_{L^2(\Omega)} ds + B_H(\xi^H(t),\xi^H(t)) = B_H(\xi^H(0),\xi^H(0)) + 2\int_0^t (\partial_t u^0 - \partial_t \Pi_H u^0, \partial_t \xi^H)\, ds.$$

Similariy to (46) we obtain

$$\int_0^t \|\partial_t \xi^H(s)\|^2_{L^2(\Omega)} ds + c_1 \|\xi^H(t)\|^2_{H^1(\Omega)} \leq c_2(\|\xi^H(0)\|^2_{H^1(\Omega)} + \int_0^t \|\xi^H(s)\|^2_{H^1(\Omega)} ds) + \int_0^t \|\partial_t u^0(s) - \partial_t \Pi_H u^0(s)\|^2_{L^2(\Omega)} ds. \quad (50)$$

As before we set $\xi^H(0) = (u^0 - \Pi_H u^0)(0) + (u_0^H - g)$ and (40) gives

$$\|\xi^H(0)\|_{H^1(\Omega)} \leq C(H^\ell + r_{HMM} + \|u_0^H - g\|_{H^1(\Omega)}). \quad (51)$$

Taking the suppremum with respect to t in (50), and using (51),(49) (36), we deduce

$$\|u^H - \Pi_H u^0\|_{\mathscr{C}^0([0,T];H^1(\Omega))} \leq C(H^\ell + r_{HMM} + \|u_0^H - g\|_{H^1(\Omega)}).$$

This together with (40) concludes the proof of (44). □

The last step to obtain fully discrete estimates in space is to quantify r_{HMM}. Remember the decomposition $r_{HMM} \leq r_{mod} + r_{mic}$ (see (16a),(16b)). We can again rewrite

$$r_{mod}(\nabla \mathscr{I}_H u^0) \leq \sup_{K\in\mathscr{T}_H, x_{K_j}\in K} \|a^0(x_{K_j}) - \bar{a}^0(x_{K_j})\|_F \|\nabla \mathscr{I}_H u^0\|_{L^2(\Omega)}, \quad (52)$$

$$r_{mic}(\nabla \mathscr{I}_H u^0) \leq \sup_{K\in\mathscr{T}_H, x_{K_j}\in K} \|\bar{a}^0(x_{K_j}) - a^{0,h}(x_{K_j})\|_F \|\nabla \mathscr{I}_H u^0\|_{L^2(\Omega)}, \quad (53)$$

where we recall that $\bar{a}^0(x_{K_j})$ is defined similarly as $a^{0,h}(x_{K_j})$ (see (25),(14)) but based on exact micro functions, i.e., when $\chi^{\xi}_{K_j}$ is solution of (12) in $W(K_{\delta_j})$. These terms have first been quantified for elliptic problems in [2] and [30, 11]. Using the definition of the cell problem (12) it is not hard to show (for linear problem) that for symmetric tensors $a^\varepsilon(x)$ one has

$$|(\bar{a}^0(x_{K_j}) - a^{0,h}(x_{K_j}))_{mn}| = \Big| \frac{1}{|K_{\delta_j}|} \int_{K_{\delta_j}} a^\varepsilon(x) \left(\nabla\chi^{\mathbf{e}_n}_{K_j}(x) - \nabla\chi^{\mathbf{e}_n,h}_{K_j}(x)\right)\cdot\left(\nabla\chi^{\mathbf{e}_m}_{K_j}(x) - \nabla\chi^{\mathbf{e}_m,h}_{K_j}(x)\right) dx\Big|. \quad (54)$$

Next assuming $|\chi^{\mathbf{e}_n}_{K_j}|_{H^{q+1}(K_{\delta_j})} \leq C\ \varepsilon^{-q}\sqrt{|K_{\delta_j}|}$, where C is independent of ε, the quadrature points x_{K_j}, and the domain K_{δ_j} one obtains

$$r_{mic}(\nabla \mathscr{I}_H u^0) \leq C\left(\frac{h}{\varepsilon}\right)^{2q}, \quad (55)$$

when using the micro finite element space (11). The justification of the above regularity assumption depends on the boundary conditions used for (11). For Dirichlet boundary conditions and for $q=1$ the regularity assumption can be established using classical H^2 regularity results [44, Chap. 2.6] provided $|a^{\varepsilon}_{mn}|_{W^{1,\infty}(\Omega)} \le C\varepsilon^{-1}$ for $m,n=1,\dots,d$. For periodic boundary conditions the above regularity assumption can be established for any given q, provided $a^{\varepsilon}=a(x,x/\varepsilon)=a(x,y)$ is Y-periodic in y, $\delta/\varepsilon \in \mathbb{N}$, and a^{ε} is sufficiently smooth, by following classical regularity results for periodic problems [19].

We finally come to the modelling error: here we need to assume some structure for the oscillatory tensor such as periodicity or random stationarity. For locally periodic problems assuming $a^{\varepsilon}=a(x,x/\varepsilon)=a(x,y) \quad Y=(0,1)^d$-periodic in y, that the sampling domain size is such that $\delta/\varepsilon \in \mathbb{N}$ and that periodic micro boundary conditions are used, we have $r_{mod} \le C\delta$ [2, 11]. Furthermore, if we assume a tensor $a(x_{K_j},x/\varepsilon)$ collocated in the slow variable $x=x_{K_j}$ for the micro and the macro problem, one can show that $r_{mod}=0$. For Dirichlet boundary condition assuming $\delta>\varepsilon$ the bound $r_{mod} \le C(\delta+\frac{\varepsilon}{\delta})$ can be established [30].

We note that for non-symmetric problems, an expression similar to (54) can still be established [28, 13], replacing the second parenthesis in the right-hand side of (54) by $\left(\nabla\bar{\chi}^{\mathbf{e}_m}_{K_j}(x)-\nabla\bar{\chi}^{\mathbf{e}_m,h}_{K_j}(x)\right)$, where $\nabla\bar{\chi}^{\mathbf{e}_m}_{K_j}, \bar{\chi}^{\mathbf{e}_m,h}_{K_j}$ are exact, respectively FE solutions of the adjoint problem of (12). The rest of the discussion is then similar. Finally we mention that by using a perturbed micro-problem, using a zeroth order term, higher order rates have been obtained in [36] for the modeling error.

3.2 *Fully discrete a priori convergence rates in space and time*

In this section we analyse the time-discretization error, when using various classes of time-integrators for the parabolic problems. We will concentrate on strongly $A(\theta)$-stable implicit Runge-Kutta methods and explicit stabilized (Chebyshev) methods.

Consider a basis $\{\phi_j\}_{j=1}^M$ of $S_0^{\ell}(\Omega,\mathcal{T}_H)$ and denote by U^H the column vector of the coefficients of $u^H=\sum_{j=1}^M U_j(t)\phi_j$ in this basis. This allows to rewrite (9) as an ordinary differential equation

$$\frac{d}{dt}U^H(t)=A_H U^H(t)+G^H(t)=F(t,U^H(t)),\;\; U^H(0)=U_0, \tag{56}$$

where $A_H=M^{-1}\hat{A}_H$ and $G^H(t)=M^{-1}P^H$. The matrice $\hat{A}_H$ is defined by the map $\hat{A}_H: S_0^{\ell}(\Omega,\mathcal{T}_H)\to S_0^{\ell}(\Omega,\mathcal{T}_H)$, where $(-\hat{A}_H v^H, w^H)=B_H(v^H,w^H)$, the mass matrix M is given by $M=((\phi_j,\phi_i))_{i,j=1}^M$ and P^H corresponds to the source term. Of course in practical computations we never invert the matrix M, but instead solve a linear system. In some situation we can also use mass lumping techniques that transform M into a matrix that is trivial to invert [55]. As mentioned in the beginning of Section 3, the FE-HMM method and the spatial convergence results can be generalised for time-dependent tensors $a^{\varepsilon}(t,x)$ and time-dependent right-hand side $f(t,x)$. In this situation we would have $B_H(v_H,w_H)$, $A_H = A_H(t)$ and $P^H = P^H(t)$ (see [12] for details).

Resolvent and α-accretive operator. To apply semi-group techniques to estimate the error when applying a Runge-Kutta method to (56), we need bound on the resolvent of $-A_H$. For the type of ODE (56) originating from a spatial discretisation of a parabolic problem, it can be shown that $-A_H$ (see for example [24]) is a so-called α-accretive operator, i.e., there exist $0 \leq \alpha \leq \pi/2$ and $C > 0$ such that for all $z \notin S_\alpha$, the operator $zI + A_H(t)$ is an isomorphism and

$$\|(zI + A_H(t))^{-1}\|_{L^2(\Omega)\to L^2(\Omega)} \leq \frac{1}{d(z,S_\alpha)} \qquad \text{for all } z \notin S_\alpha, \tag{57}$$

where $d(z,S_\alpha)$ is the distance between z and $S_\alpha = \{\rho e^{i\theta}\ ;\ \rho \geq 0, |\theta| \leq \alpha\}$. We note that the operator A_H can be extended straightforwardly to a complex Hilbert space based on $S_0^{\ell}(\Omega,\mathscr{T}_H)$ equipped with the complex scalar product $(u,v) = \int_\Omega u(x)\overline{v}(x)dx$ which is an extension of the usual L^2 scalar product. If we denote by γ_1,γ_2 the coercivity and continuity constant of the bilinear form $B_H(\cdot,\cdot)$, it can be shown that $\alpha \leq \arccos(\gamma_1/\gamma_2)$. Hence A_H generates an analytic semi-group in S_α (see [43]).

Runge-Kutta methods. For the time discretisation of (56) we consider an s-stage Runge-Kutta method

$$U_{n+1} = U_n + \Delta t \sum_{i=1}^{s} b_i K_{ni}, \quad U_{ni} = U_n + \Delta t \sum_{j=1}^{s} \gamma_{ij} K_{nj}, \tag{58}$$

$$K_{ni} = F(t_n + c_i \Delta t U_{ni}), \quad i = 1 \dots s. \tag{59}$$

where γ_{ij}, b_j, c_i with $i,j = 1\dots s$ are the coefficients of the method (with $\sum_{j=1}^{s}\gamma_{ij} = c_i$) and $t_n = n\Delta t$. We further define

$$\Gamma = (\gamma_{ij})_{i,j=1}^{s}, \quad b = (b_1,\dots,b_s)^T, \quad c = (c_1,\dots,c_s)^T = \Gamma\mathbf{1}, \quad \mathbf{1} = (1,\dots,1)^T.$$

The method is said to have "order r" if the error after one step between the exact and the numerical solutions (with the same initial condition) satisfies

$$U_1 - U(t_1) = \mathscr{O}\left(\Delta t^{r+1}\right), \text{ for} \Delta t \to 0,$$

for all sufficiently differentiable systems of differential equations. We recall that the rational function $R(\Delta t\lambda) = R(z) = 1 + zb^T(I - z\Gamma)^{-1}\mathbf{1}$ obtained after one step Δt of a Runge-Kutta method applied to the scalar problem $dy/dt = \lambda y,\ y(0) = 1,\ \lambda \in \mathbb{C}$ is called the stability function of the method.

Strongly $A(\theta)$-stable methods. We consider a subclass of implicit Runge-Kutta methods which are of order r and whose stage order (the accuracy of the internal stages) is $r-1$. We further recall that a Runge-Kutta method is strongly $A(\theta)$-stable with $0 \leq \theta \leq \pi/2$ if $I - z\Gamma$ is a nonsingular matrix in the sector $|\arg(-z)| \leq \theta$ and the stability function satisfies $|R(z)| < 1$ in $|\arg(-z)| \leq \theta$. Notice that all s-

stage Radau Runge-Kutta methods satisfy the above assumptions (with $\theta \geq \pi/2$). In particular, for $s = 1$, we retrieve the implicit Euler method. We refer to [37, Sect. IV.3,IV.15] for details on the stability concepts mentioned here.

Under the assumptions of Theorem (2) we have the following theorem.

Theorem 3. *Let $u^0(t)$ be the solution of* (5) *and let u_n^H be a strongly $A(\theta)$ stable Runge-Kutta approximation of order r and stage order $r-1$ of* (56) *with time step Δt. Assume the hypotheses of Theorem* (2)*,* (55)*, that $r_{MOD} = 0$ and $a^\varepsilon \in \mathscr{C}^r([0,T], L^\infty(\Omega)^{d\times d})$, $\|\partial_t^r u^H(0)\|_{L^2(\Omega)} \leq C$. Then,*

$$\max_{0\leq n\leq N} \|u_n^H - u^0(t_n)\|_{L^2(\Omega)} \leq C\left(H^{\ell+1} + \left(\frac{h}{\varepsilon}\right)^{2q} + \Delta t^r\right).$$

Assuming in addition $\|u^H(0) - g\|_{H^1(\Omega)} \leq C(H^\ell)$ and a^ε is symmetric, then

$$\sum_{n=0}^{N-1} \Delta t_n \|u_n^H - u^0(t_n)\|_{H^1(\Omega)}^2 \leq C\left(H^\ell + \left(\frac{h}{\varepsilon}\right)^{2q} + \Delta t^r\right)^2.$$

All the above constants C are independent of $H, h, \varepsilon, \Delta t$.

The idea of the proof is to consider the decomposition: $u_n^H - u^0(t_n) = (u_n^H - u^H(t_n)) + (u^H(t_n) - u^0(t_n))$. Then the first term is estimated using semigroup techniques (for time independent operators) + perturbation techniques (following [46]). The second term is estimated using Theorem 2. We note that the analysis for implicit methods covers variable time step methods under some mild assumptions on the sequence of time-steps [12]. Finally we mention that the bound $\|\partial_t^r u^H(0)\|_{L^2(\Omega)} \leq C$ can be established provided that we assume an inverse assumption $\frac{H}{H_K} \leq C$ for all $K \in \mathscr{T}_H$ and all $\mathscr{T}_H$ for the macroscopic finite element mesh and appropriate regularity of $\partial_t^k u^0$, $k = 1,\ldots,r$. We refer to [12] for a detailed proof of the above theorem.

Chebyshev methods. Chebyshev methods are a subclass of explicit Runge-Kutta methods with extended stability along the negative real axis suitable for parabolic (advection-diffusion) problems. Such methods have been constructed for order up to $r = 4$ [10, 1, 56, 45]. They are based on s-stage stability functions satisfying

$$|R_s(x)| \leq 1 \quad \text{for } x \in [-L_s, 0] \tag{60}$$

with $L_s = Cs^2$, where the constant C depends on the order of the method. First order methods are based on

$$R_s(x) = T_s(1 + x/s^2), \tag{61}$$

where $T_s(\cdot)$ denotes the Chebyshev polynomial of degree s and $L_s = 2s^2$. The corresponding Runge-Kutta method can be efficiently implemented by using the three-term recurrence relation of the Chebyshev polynomials [56]. For stiff diffusion problems, such methods are much more efficient than classical explicit methods. Indeed let ρ_H be the spectral radius of the discretized parabolic problem (depending on the macro spatial meshsize H) and let Δt be the stepsize to achieve the desired accuracy. Using a classical explicit method such as the forward Euler method requires a stepsize δt satisfying the CFL constraints $\delta t \leq 2/\rho_H$. The number of function evaluations per time-step Δt (taken here a the measure of the numerical work) is therefore $\Delta t/\delta t \geq (\Delta t \rho_H)/2$. Using a Chebyshev method (of order one) with stability function (61) we choose the number of stages s of the method to ensure stability $\Delta t \rho_H \leq 2s^2$. As for Chebyshev methods, there is one new function evaluation per stage the total number of function evaluations per time-step Δt is given by $s = \sqrt{(\Delta t \rho_H)/2}$.

Chebyshev method are usually used in a slightly modified form obtained by changing the stability function (61) to

$$R_s(z) = \frac{T_s(\omega_0 + \omega_1 z)}{T_s(\omega_0)}, \quad \text{with} \quad \omega_0 = 1 + \frac{\eta}{s^2}, \quad \omega_1 = \frac{T_s(\omega_0)}{T_s'(\omega_0)}, \tag{62}$$

we obtain the "damped form" of the Chebyshev method. For any fixed $\eta > 0$ (called the damping parameter) we obtain a damped stability function satisfying

$$\sup_{z \in [-L_s, -\gamma], s \geq 1} |R_s(z)| < 1, \quad \text{for all } \gamma > 0. \tag{63}$$

This modification also ensure that a strip around the negative real axis is contained in the stability domain $\mathscr{S} := \{z \in \mathbb{C}; |R_s(z)| \leq 1\}$. The growth on the negative real axis for the damped form is reduced but remains quadratic [53],[37, Chap. IV.2]. For the analysis we assume the order of the Chebyshev method is $r \geq 1$ for linear problem, precisely,

$$\lim_{z \to 0} \left| \frac{e^z - R_s(z)}{z^{r+1}} \right| < \infty \quad \text{for all } s \geq 1. \tag{64}$$

We also assume that the stability functions are bounded in a neighbourhood of zero uniformly with respect to s, precisely, there exist $\delta > 0$ and $C > 0$ such that

$$|R_s(z)| \leq C \text{ for all } |z| \leq \delta \text{ and all } s. \tag{65}$$

This can be checked for the Chebyshev methods with stability functions (61), (62).

Theorem 4. *Let $u^0(t)$ be the solution of* (5) *with $f = 0$ and a time-independent symmetric tensor a^ε. Let u_n^H be a Runge-Kutta-Chebyshev approximation of the corresponding discretized problem* (56) *with timestep Δt. Assume that the method satisfies* (64) *(order r),* (63) *(strong stability) and* (65)*. Assume in adddition that the stage number s of the Chebyshev method is chosen such that $\rho_H \Delta t \leq L_s$ holds. Assume the hypotheses of Theorem 2 with $\mu = 1$,* (55) *and that $r_{MOD} = 0$. Then,*

$$\max_{0 \leq n \leq N} \|u_n^H - u^0(t_n)\|_{L^2(\Omega)} \leq C \left(H^{\ell+1} + \left(\frac{h}{\varepsilon}\right)^{2q} + \Delta t^r \right).$$

where C is independent of $H, h, \varepsilon, \Delta t$.

The ideas of the proof are as follows. Consider again the decomposition $u_n^H - u^0(t_n) = (u_n^H - u^H(t_n)) + (u^H(t_n) - u^0(t_n))$. The second term is estimated as before using Theorem 2.

For the first term we follow ideas developed in [25, 24] for implicit methods, adapted here for stabilized methods. Using the symmetry of A_H, there exists an orthonormal basis where the operator A_H is in diagonal form. Define next $\varphi_{n,s}(z) = e^{nz} - R_s(z)^n$. Then we have

$$\|\varphi_{n,s}(\Delta t A_H)\|_{L^2(\Omega)\to L^2(\Omega)} = \sup_{z\in sp(A_H)} |\varphi_{n,s}(\Delta t z)|,$$

where $sp(A_H)$ denotes the spectrum of A_H. Using (63),(64),(65) we show that $|\varphi_{n,s}(z)| \leq C_1 n^{-r}$ for all $z \in [-\delta, 0]$, where C_1 is independent of n and s. For the case $z \in [-L_s, -\delta]$ we denote by $\rho < 1$ the quantity in the left-hand side of (63). We can then estimate

$$|\varphi_{n,s}(z)| \leq e^{-n|z|} + \rho^n \leq e^{-n\gamma} + e^{-n(1-\rho)} \leq \frac{(r/e)^r(\gamma^{-r} + (1-\rho)^{-r})}{n^r} = C_2 n^{-r},$$

where we used twice the estimate $e^{-x} \leq (\frac{r}{ex})^r$ (valid for $x \geq 0$). We have thus $|\varphi_{n,s}(z)| \leq C n^{-r}$ for all $z \in [-L_s, 0]$, hence

$$\|u_n^H - u^H(t_n)\|_{L^2(\Omega)} = \|\varphi_{n,s}(\Delta t A_H) u_0^H\|_{L^2(\Omega)} \leq C n^{-r} \|u_0^H\|_{L^2(\Omega)},$$

where C is independent of n, s. By noting that $n \leq T/\Delta t$ we get the result.

4 Fully discrete space-time error estimates for nonlinear monotone parabolic problem

In this section we describe convergence and error estimates for the numerical method (9) applied to the general problem (1). We focus here on a simple time integrator, namely the implicit Euler method and take piecewise linear macro and micro FEM. We consider a uniform subdivision of the time interval $(0,T)$ with time step $\Delta t = T/N$ and discrete time $t_n = n\Delta t$ for $0 \leq n \leq N$ and $N \in \mathbb{N}_{>0}$. The method then reads as folows: for $0 \leq n \leq N-1$ find $u_{n+1}^H \in S_0^1(\Omega, \mathscr{T}_H)$ such that

$$\int_\Omega \frac{u_{n+1}^H - u_n^H}{\Delta t} w^H dx + B_H(u_{n+1}^H; w^H) = \int_\Omega f w^H dx, \qquad \forall w^H \in S_0^1(\Omega, \mathscr{T}_H), \quad (66)$$

with the nonlinear macro map B_H given by

$$B_H(v^H; w^H) = \sum_{K\in\mathscr{T}_H} |K| \mathscr{A}_K^{0,h}(\nabla v^H(x_K)) \cdot \nabla w^H(x_K), \qquad v^H, w^H \in S_0^1(\Omega, \mathscr{T}_H), \quad (67)$$

where $\mathscr{A}_K^{0,h}$ is given by (13) with micro problems (12) computed in $S^1(K_\delta, \mathscr{T}_h)$. Here we have just one quadrature point and sampling domain K_δ located at the barycenter of each macro element K. We note that we will sometimes use the short-hand notation $\bar{\partial}_t v_n = (v_{n+1} - v_n)/\Delta t$. The proof of the existence and uniqueness of a numerical solution can be establish similarly to the proof of Theorem 1. Further, the numerical solution $\{u_n^H\}_{n=1}^N$ satisfies the bound

$$\max_{1\le n\le N} \left\|u_n^H\right\|_{L^2(\Omega)}^2 + \sum_{n=1}^N \Delta t \left\|\nabla u_n^H\right\|_{L^p(\Omega)}^p \le C(1 + \|f\|_{L^{p'}(\Omega)}^{p'} + \left\|u_0^H\right\|_{L^2(\Omega)}^2), \quad (68)$$

where C only depends on $p, \beta, \lambda L, , L_0, \kappa_1, \kappa_2$, the measure of Ω and the Poincaré constant C_P on Ω.

4.1 General estimates in the $W^{1,p}$ setting

For the scheme (66),(67) in the general nonlinear monotone setting we have the following fully discrete convergence result.

Theorem 5. *Let $u^0 \in E$ be the solution to the homogenized problem* (5) *and u_n^H the HMM solution obtained from* (66) *with initial conditions u_0^H satisfying $\|g - u_0^H\|_{L^2(\Omega)} \to 0$ for $H \to 0$. Assume that $\mathscr{A}^\varepsilon$ satisfies ($\mathscr{A}_{0-2}$). Let $\mathscr{A}^0$ be Hölder continuous in space, i.e., there exists $0 < \tilde{\gamma} \le 1$ such that*

$$\left|\mathscr{A}^0(x_1,\xi) - \mathscr{A}^0(x_2,\xi)\right| \le C|x_1 - x_2|^{\tilde{\gamma}}(1 + (\kappa_1 + |\xi|)^{p-1}), \;\; \forall x_1, x_2 \in \Omega, \forall \xi \in \mathbb{R}^d. \quad (69)$$

Assume in addition that the coupling is such that $r_{mod} = 0$. Then we have the convergence

$$\lim_{(\Delta t, H)\to 0} \lim_{h\to 0} \left[\max_{1\le n\le N} \left\|u^0(\cdot,t_n) - u_n^H\right\|_{L^2(\Omega)} + \left\|\nabla u^0 - \nabla u^H\right\|_{\widetilde{L^p}(L^p(\Omega)),} \right] = 0.$$

where

$$\left\|\nabla u^0 - \nabla u^H\right\|_{\widetilde{L^p}(L^p(\Omega))}^p = \sum_{n=0}^{N-1} \int_{t_n}^{t_{n+1}} \left\|\nabla u^0(\cdot,s) - \nabla u_{n+1}^H\right\|_{L^p(\Omega)}^p ds.$$

We sketch the proof of this result.

Step 1: Approximation by smooth function. Due to the low regularity of the true solution we can only rely on a weak approximation in time. Indeed, for u^0 we can only use the formulation $\int_{t_n}^{t_{n+1}} \left\langle \partial_t u^0(\cdot,s), w^H \right\rangle ds$ instead of $\int_{t_n}^{t_{n+1}} \int_\Omega \partial_t u^0(x,s) w^H(x) dx ds$ that only make sense with additional regularity. We therefore consider $\mathscr{U} \in E$ with $\mathscr{U} \in \mathscr{C}^0([0,T], W_0^{1,p}(\Omega))$ and $\partial_t \mathscr{U} \in \mathscr{C}^0([0,T], L^2(\Omega))$. Further, let $\mathscr{U}^H(\cdot,t) \in S_0^1(\Omega, \mathscr{T}_H)$ be an approximation of $\mathscr{U}(\cdot,t)$ for $t \in [0,T]$ and define $\mathscr{U}_n^H = \mathscr{U}^H(\cdot,t_n)$ for $0 \le n \le N$. We will then decompose the error as

$$\left\|u^0(\cdot,t_n)-u_n^H\right\|_{L^2(\Omega)} \leq \left\|u^0(\cdot,t_n)-\mathscr{U}_n^H\right\|_{L^2(\Omega)}+\left\|\theta_n^H\right\|_{L^2(\Omega)} \tag{70}$$

$$\left\|\nabla u^0-\nabla u^H\right\|_{\widetilde{L^p}(L^p(\Omega)} \leq \left\|\nabla u^0-\mathscr{U}^H\right\|_{\widetilde{L^p}(L^p(\Omega))}+(\sum_{n=0}^{N-1}\Delta t\|\nabla\theta_{n+1}^H\|_{L^p(\Omega)}^p)^{1/p}, \tag{71}$$

where $\theta_n^H=u_n^H-\mathscr{U}_n^H$.

Step 2: Density argument, weak approximation in time. To bound the first terms in (70),(71) we use that $\mathscr{U},\partial_t\mathscr{U}\in\mathscr{C}^0([0,T],W^{1,p}(\Omega))$ to obtain fo $t_n\leq s\leq t_{n+1}$

$$\|\nabla\mathscr{U}(\cdot,t_{n+1})-\nabla\mathscr{U}(\cdot,s)\|_{L^p(\Omega)}=\left\|\int_s^{t_{n+1}}\partial_t\nabla\mathscr{U}(\cdot,\tau)d\tau\right\|_{L^p(\Omega)} \tag{72}$$

$$\leq\Delta t\|\partial_t\nabla\mathscr{U}\|_{\mathscr{C}^0([0,T],L^p(\Omega))}. \tag{73}$$

Now if we take $\mathscr{U}_n^H=\mathscr{I}_H\mathscr{U}(\cdot,t_n)$ the above inequality together with standard interpolation results yields (72) in time we get that for $s\in[t_n,t_{n+1}]$ and $0\leq n\leq N-1$

$$\left\|\nabla\mathscr{U}(\cdot,s)-\nabla\mathscr{U}_{n+1}^H\right\|_{L^p(\Omega)}\leq \quad C(\Delta t+H)\Big(\|\mathscr{U}\|_{\mathscr{C}^0([0,T],W^{2,p^*}(\Omega))} \tag{74}$$

$$+ \quad \|\partial_t\nabla\mathscr{U}\|_{\mathscr{C}^0([0,T],L^p(\Omega))}\Big). \tag{75}$$

We also have

$$\max_{1\leq n\leq N}\left\|u^0(\cdot,t_n)-\mathscr{U}_n^H\right\|_{L^2(\Omega)}\leq C_E\left\|u^0-\mathscr{U}\right\|_E+CH\|\mathscr{U}\|_{\mathscr{C}^0([0,T],W^{2,p^*}(\Omega))}, \tag{76}$$

where we used the embeddings $E\hookrightarrow\mathscr{C}^0([0,T],L^2(\Omega))$ (with operator norm C_E), $W^{1,p}(\Omega)\hookrightarrow L^2(\Omega)$ and standard interpolation estimates. We then choose $\mathscr{U}\in C^\infty(\overline{\Omega}\times[0,T])$ such that $\mathscr{U}(\cdot,t)\in C_0^\infty(\Omega)$ for any $t\in[0,T]$ and $\|u^0-\mathscr{U}\|_E<\eta/2$. Then, using (72),(74) we find that for each $\eta>0$ there exists $D(\eta)$ such that for all $\Delta t,H\leq D_1(\eta)$ we have

$$\left\|\nabla u^0-\nabla\mathscr{U}^H\right\|_{\widetilde{L^p}(L^p(\Omega)}\leq\eta,\quad \max_{1\leq n\leq N}\left\|u^0(\cdot,t_n)-\mathscr{U}_n^H\right\|_{L^2(\Omega)}\leq(C_E+1)\eta. \tag{77}$$

Step 3: Macro discretization error. We next need to estimate $\theta_n^H=u_n^H-\mathscr{U}_n^H$, $0\leq n\leq N$. Hölder inequality and the monotonicity estimate (22) gives

$$\sum_{n=0}^{N-1}\Delta t\left\|\nabla\theta_{n+1}^H\right\|_{L^p(\Omega)}^p\leq\mathscr{R}(u_n^H,\mathscr{U}_n^H)^{\frac{p(\beta-p)}{\beta}} \tag{78}$$

$$\cdot\left(\sum_{n=0}^{N-1}\Delta t(B_H(u_{n+1}^H;\theta_{n+1}^H)-B_H(\mathscr{U}_{n+1}^H;\theta_{n+1}^H))\right)^{\frac{p}{\beta}}, \tag{79}$$

where

$$\mathscr{R}(u_n^H,\mathscr{U}_n^H)=\gamma_c^{-p/\beta}\Big(C+\left(\sum_{n=0}^{N-1}\Delta t\left\|\nabla u_{n+1}^H\right\|_{L^p(\Omega)}^p\right)^{\frac{1}{p}}+\left(\sum_{n=0}^{N-1}\Delta t\left\|\nabla \mathscr{U}_{n+1}^H\right\|_{L^p(\Omega)}^p\right)^{\frac{1}{p}}\Big), \tag{80}$$

with C depending on $C_d,T,|\Omega|$. We observe that

$$\mathscr{R}(u_n^H,\mathscr{U}_n^H)\le C, \tag{81}$$

where C is idependent of $\mathscr{U},\eta,\Delta t,H$ (for small enough discretization parameters). Indeed, using (68) and $\|g-u_0^H\|_{L^2(\Omega)}\to 0$ for $H\to 0$ we can find H_0 such that for all $H\le H_0$ we have $\left(\sum_{n=0}^{N-1}\Delta t\left\|\nabla u_{n+1}^H\right\|_{L^p(\Omega)}^p\right)^{\frac{1}{p}}\le C$ independently of the initial approximation u_0^H. Using (74) we find that $\left(\sum_{n=0}^{N-1}\Delta t\left\|\nabla \mathscr{U}_{n+1}^H\right\|_{L^p(\Omega)}^p\right)^{1/p}\le \left\|u^0\right\|_E+1$ for all $\Delta t,H\le\min\{H_0,D_1(\eta_0)\}$.

We need next to estimate $B^H(u_{n+1}^H;\theta_{n+1}^H)-B^H(\mathscr{U}_{n+1}^H;\theta_{n+1}^H)$. This is done by a decomposition

$$\sum_{n=0}^{N-1}\Delta t\big(B^H(u_{n+1}^H;\theta_{n+1}^H)-B^H(\mathscr{U}_{n+1}^H;\theta_{n+1}^H)\big)=\sum_{n=0}^{N-1}\mathscr{B}_n^{tot}-\sum_{n=0}^{N-1}\Delta t\int_\Omega\bar{\partial}_t\theta_n^H\theta_{n+1}^H dx \tag{82}$$

where $\mathscr{B}_n^{tot}$ contains a number of terms that represent the contribution to the error due to the weak approximation in time, the macroscopic numerical discretization, the time discretization, the quadrature error, the micro and the modelling error [8, Sect. 5.1]. We also have in view of

$$\frac{1}{2}\bar{\partial}_t\left\|\theta_n^H\right\|_{L^2(\Omega)}^2\le\int_\Omega\bar{\partial}_t\theta_n^H\theta_{n+1}^H dx,\qquad \text{for } 0\le n\le N-1, \tag{83}$$

that

$$-\sum_{n=0}^{N-1}\Delta t\int_\Omega\bar{\partial}_t\theta_n^H\theta_{n+1}^H dx\le\frac{1}{2}\left\|\theta_0^H\right\|_{L^2(\Omega)}^2-\frac{1}{2}\left\|\theta_N^H\right\|_{L^2(\Omega)}^2. \tag{84}$$

The initial error θ_0^H in (84) can be bounded by using interpolant estimates and the embedding $E\hookrightarrow\mathscr{C}^0([0,T],L^2(\Omega))$ as

$$\left\|\theta_0^H\right\|_{L^2(\Omega)}\le\left\|g-u_0^H\right\|_{L^2(\Omega)}+C_E\left\|u^0-\mathscr{U}\right\|_E+CH\|\mathscr{U}\|_{\mathscr{C}^0([0,T],W^{2,p^*}(\Omega))}. \tag{85}$$

Next it can be shown, in view of (85) and the properties of $\mathscr{U}$ derived in step 2, that for $\Delta t,H\le D_2(\eta)$, with $D_2(\eta)$ small enough we have (see [8, Sect. 5.2] for details)

$$\left(\sum_{n=0}^{N-1}\mathscr{B}_n^{tot}-\sum_{n=0}^{N-1}\Delta t\int_\Omega\bar{\partial}_t\theta_n^H\theta_{n+1}^H dx\right)^{\frac{1}{\beta}}\le C\eta \tag{86}$$

Step 3: Upscaling error. First as $r_{mod}=0$ we have $r_{HMM}(\nabla\mathscr{U}^H_{n+1}) = r_{mic}(\nabla\mathscr{U}^H_{n+1})$, where r_{mic} is given by (16a). Let the macro mesh size $H>0$, the time step size $\Delta t>0$ and the micro finite element space in (12) be given. Then, assuming that $\mathscr{A}^\varepsilon$ satisfies ($\mathscr{A}_{0-2}$) it can be shown that for any sequence $\{\mathscr{U}^H_n\}_{1\le n\le N}\subset S^1_0(\Omega,\mathscr{T}_H)$ for which $\sum_{n=0}^{N-1}\Delta t\|\nabla\mathscr{U}^H_{n+1}\|^p_{L^p(\Omega)}$ is bounded independently of the micro mesh size h, we have

$$\lim_{h\to 0}\left(\sum_{n=0}^{N-1}\Delta t\, r_{mic}(\nabla\mathscr{U}^H_{n+1})^{p'}\right)^{\frac{1}{p'}} = 0.$$

This result follows from a density argument, classical FE interpolation results and the general estimate obtained from ($\mathscr{A}_2$)

$$r_{mic}(\nabla v^H)\le C\left[C_d+\left\|\nabla v^H\right\|_{L^p(\Omega)}\right]^{p-1-\gamma} \times\left(\sum_{K\in\mathscr{T}_H}\frac{|K|}{|K_\delta|}\inf_{z^h\in S^1(K_\delta,\mathscr{T}_h)}\left\|\nabla\bar{\chi}_K^{\nabla v^H(x_K)}-\nabla z^h\right\|^p_{L^p(K_\delta)}\right)^{\frac{\gamma}{p}},$$

for any $v^H\in S^1_0(\Omega,\mathscr{T}_H)$, where $\bar{\chi}^\xi_K$ solves (12) in$W(K_{\delta_j})$ and C is independent of H, h, δ and ε.
Step 4: Assembling the pieces: convergence in $L^p(W^{1,p})$ and $\mathscr{C}^0(L^2)$ norm. In view of (79),(81),(86) if we set $0<D_3(\eta)\le\min\{D_1(\eta_0),H_0,D_2(\eta)\}$ then for $\Delta t,H\le D_3(\eta)$ we have

$$\lim_{h\to 0}\left(\sum_{n=0}^{N-1}\Delta t\left\|\nabla\theta^H_{n+1}\right\|^p_{L^p(\Omega)}\right)^{\frac{1}{p}}\le C\eta, \tag{87}$$

where C is independent of $\mathscr{U},\eta,H,\Delta t,\delta$ and h. Combining this inequality with (71) and the density estimates of step 2 yields the convergence in the $L^p(W^{1,p})$ norm.

Next to derive a bound in the $\mathscr{C}^0(L^2)$, we first observe that (83) together with the monotonicity estimate (22) yield

$$\frac{1}{2}\bar{\partial}_t\left\|\theta^H_n\right\|^2_{L^2(\Omega)}\le\int_\Omega\bar{\partial}_t\theta^H_n\theta^H_{n+1}dx+B_H(u^H_{n+1};\theta^H_{n+1})-B_H(\mathscr{U}^H_{n+1};\theta^H_{n+1}), \tag{88}$$

Summing this inequality for $n=0,\dots,K-1$, taking the maximum over K, using (83) and the monotonicity of B_H from Lemma 2 we get

$$\frac{1}{2}\left\|\theta^H_K\right\|^2_{L^2(\Omega)}-\frac{1}{2}\left\|\theta^H_0\right\|^2_{L^2(\Omega)}\le\sum_{n=0}^{K-1}\mathscr{B}^{tot}_n,$$

where $\mathscr{B}^{tot}_n$ is defined in (82). Using then (85) and an estimate similar to (86) we find that

$$\lim_{h\to 0}(\max_{1\le n\le N}\left\|\theta_n^H\right\|_{L^2(\Omega)})\le C\eta, \tag{89}$$

for all $\Delta t, H$ small enough, where C is independent of $\mathscr{U}, \eta, H, \Delta t, \delta$ and h. Hence together with (77) this shows the $\mathscr{C}^0(L^2)$ estimate of Theorem 5.

4.2 Convergence for strongly monotone and Lipschitz nonlinear maps

Optimal convergence rates can be derived for $p=2$ and $\alpha=1$, $\beta=2$ in $(\mathscr{A}_{1-2})$, i.e., when the nonlinear map $\mathscr{A}^\varepsilon(x,\xi)$ is Lipschitz continuous with respect to its first argument and strongly monotone. In this case we can derive optimal macroscopic, microscopic and temporal error estimates without any structural assumptions such as local periodicity or random stationarity of $\mathscr{A}^\varepsilon$. Explicit bounds of the modelling error are however derived only for locally periodic data $\mathscr{A}^\varepsilon$.

Theorem 6. *For the case $p=2$ assume that $\mathscr{A}^\varepsilon$ satisfies $(\mathscr{A}_{0-2})$ with $\alpha=1$, $\beta=2$. Let u^0 be the solution to the homogenized problem* (5) *and u_n^H the numerical solution obtained from* (66) *with initial condition u_0^H. Provided in addition that*

$$u^0, \partial_t u^0\in\mathscr{C}^0([0,T],H^2(\Omega)),\ \ \partial_t^2 u^0\in\mathscr{C}^0([0,T],L^2(\Omega)), \tag{90a}$$

$$\mathscr{A}^0(\cdot,\xi)\in W^{1,\infty}(\Omega;\mathbb{R}^d) \text{ with } \left\|\mathscr{A}^0(\cdot,\xi)\right\|_{W^{1,\infty}(\Omega;\mathbb{R}^d)}\le C(L_0+|\xi|),\ \ \forall\xi\in\mathbb{R}^d, \tag{90b}$$

then, the following discrete $\mathscr{C}^0(L^2)$ and $L^2(H^1)$ error estimate holds

$$\max_{1\le n\le N}\left\|u^0(\cdot,t_n)-u_n^H\right\|_{L^2(\Omega)}+\left(\sum_{n=1}^N\Delta t\left\|\nabla u^0(\cdot,t_n)-\nabla u_n^H\right\|_{L^2(\Omega)}^2\right)^{1/2} \tag{91}$$
$$\le C\left[\Delta t+H+\max_{1\le n\le N} r_{HMM}(\nabla\mathscr{I}_H u^0(\cdot,t_n))+\left\|g-u_0^H\right\|_{L^2(\Omega)}\right],$$

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

Remark 2. Under additional regularity assumptions, assuming elliptic regularity and quasi-uniform meshes, one can derive the following improved (discrete) $\mathscr{C}^0(L^2)$ error estimate

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

where $\tilde{u}^{H,0}$ is given by an elliptic projection and C is independent of $\Delta t, H$ and r_{HMM} (see [8, Thm. 4.4]).

We sketch the proof of Theorem 6.

Owing to regularity assumptions for u^0 , we can use a strong formulation in time

$$\int_\Omega \partial_t u^0(x,t)\, w\, dx + B^0(u^0(\cdot,t);w) = \int_\Omega f\, w\, dx, \qquad \forall w \in W_0^{1,2}(\Omega), \forall t \in (0,T].$$

Hence the argument density used in Section 4.1 is not needed here. We can then directly define $\mathscr{U}_n^H = \mathscr{I}_H u^0(\cdot,t_n)$ and with $\theta_n^H = \mathscr{U}_n^H - u_n^H$ we obtain instead of (82) the following error propagation formula

$$\begin{aligned}
&\int_\Omega \bar{\partial}_t \theta_n^H w^H\, dx + \left[B_H(u_{n+1}^H;w^H) - B_H(\mathscr{U}_{n+1}^H;w^H)\right] \\
&= \int_\Omega \left[\partial_t u^0(x,t_{n+1}) - \bar{\partial}_t u^0(x,t_n)\right] w^H dx\, ds && \text{(92a)}\\
&+ \int_\Omega \left[\partial_t u^0(x,t_n) - \bar{\partial}_t \mathscr{U}_n^H\right] w^H dx && \text{(92b)}\\
&+ \left[B^0(u^0(\cdot,t_{n+1});w^H) - B^0(\mathscr{U}_{n+1}^H;w^H)\right] && \text{(92c)}\\
&+ \left[B^0(\mathscr{U}_{n+1}^H;w^H) - \hat{B}^0(\mathscr{U}_{n+1}^H;w^H)\right] && \text{(92d)}\\
&+ \left[\hat{B}^0(\mathscr{U}_{n+1}^H;w^H) - B_H(\mathscr{U}_{n+1}^H;w^H)\right]. && \text{(92e)}
\end{aligned}$$

In the above formula the term (92a) is due to the time discretization error, the terms (92b) and (92c) account for the finite element error at the discrete time levels t_n. The influence of the quadrature formula is captured by (92d). Finally the components (92a) – (92d) are independent of the temporal and macro spatial error, while last term (92e) is only due to the upscaling strategy and averaging techniques used to define and compute numerically the upscaled tensor. All these terms can be estimated quantitatively [8]. If we set $w^H = \theta_{n+1}^H$ use the inequality (83) we obtain

$$\begin{aligned}
\frac{1}{2}\bar{\partial}_t \left\|\theta_n^H\right\|_{L^2(\Omega)}^2 &+ \lambda \left\|\nabla \theta_{n+1}^H\right\|_{L^2(\Omega)}^2 \\
&\le C\Delta t \left\|\partial_t^2 u^0\right\|_{\mathscr{C}^0([0,T],L^2(\Omega))} \left\|\theta_{n+1}^H\right\|_{L^2(\Omega)} \\
&+ CH \left\|u^0\right\|_{\mathscr{C}^0([0,T],H^2(\Omega))} \left\|\theta_{n+1}^H\right\|_{L^2(\Omega)} \\
&+ CH \left\|u^0\right\|_{\mathscr{C}^0([0,T],H^2(\Omega))} \left\|\nabla \theta_{n+1}^H\right\|_{L^2(\Omega)} \\
&+ r_{HMM}(\nabla \mathscr{U}_{n+1}^H) \left\|\nabla \theta_{n+1}^H\right\|_{L^2(\Omega)}. && \text{(93)}
\end{aligned}$$

Multiplying the above inequality by Δt and summing first from $n = 0, \ldots, K-1 \le N-1$ and taking the maximum over K yields

$$\begin{aligned}
&\max_{1\le n\le N} \left\|\theta_N^H\right\|_{L^2(\Omega)}^2 + \lambda \sum_{n=1}^N \Delta t \left\|\nabla \theta_n^H\right\|_{L^2(\Omega)}^2 \\
&\le \left\|\theta_0^H\right\|_{L^2(\Omega)}^2 + C(\Delta t + H + \max_{1\le n\le N} r_{HMM}(\nabla \mathscr{U}_n^H))^2. && \text{(94)}
\end{aligned}$$

The classical estimates for nodal interpolant $\|\mathscr{I}_H z - z\|_{H^1(\Omega)} \leq CH\|z\|_{H^2(\Omega)}$ for $u^0(\cdot,t_n) - \mathscr{U}_n^H$ together with the regularity of (90a) and the triangle inequality gives finally the estimate of Theorem 6.

Fully discrete space-time result. Recall that $r_{HMM}(\cdot) \leq r_{mod}(\cdot) + r_{mic}(\cdot)$ (see (16a),(16b)). Following the results for linear problems (with additional technicalities due to the nonlinear micro-problems) we can estimate both $r_{mod}(\cdot)$ and $r_{mic}(\cdot)$. First under the assumptions of Theorem 6 and assuming that the exact solution of Problem (12) satisfies $\bar{\chi}_K^{\xi} \in H^2(K_\delta)$ and $\left|\bar{\chi}_K^{\xi}\right|_{H^2(K_\delta)} \leq C\varepsilon^{-1}(L_0 + |\xi|)\sqrt{|K_\delta|}$ we have the following error estimate for the micro error

$$r_{mic} \leq C\frac{h}{\varepsilon},$$

where C is independent of $\Delta t, H, h, \varepsilon, \delta$. By defining a appropriate linear adjoint auxiliary problem derived from (12) and assuming $W^{1,\infty}(K_\delta)$ regularity of the solutions of these (linear) problems one can get the optimal micro error

$$r_{mic} \leq C\left(\frac{h}{\varepsilon}\right)^2, \tag{95}$$

with the same rate as for linear problem [2].

For the modelling error we need structural assumptions and assume that $\mathscr{A}^{\varepsilon}(x,\xi) = \mathscr{A}(x,x/\varepsilon,\xi)$ where $\mathscr{A}(x,y,\xi)$ is Y-periodic in y, i.e., $\mathscr{A}^{\varepsilon}$ is locally periodic. Then, for any $v^H \in S_0^1(\Omega,\mathscr{T}_H)$, the modelling error $r_{mod}(\nabla v^H)$ defined in (16b) is bounded by

$$r_{mod}(\nabla v^H) \leq \begin{cases} 0, & \text{if } W(K_\delta) = W^1_{per}(K_\delta), \delta/\varepsilon \in \mathbb{N} \text{ and} \\ & \mathscr{A}^{\varepsilon} = \mathscr{A}(x_K, x/\varepsilon, \xi) \text{ collocated at } x_K, \\ C^1_{mod}\,\delta, & \text{if } W(K_\delta) = W^1_{per}(K_\delta), \delta/\varepsilon \in \mathbb{N}, \\ C^2_{mod}(\delta + \sqrt{\varepsilon/\delta}), & \text{if } W(K_\delta) = H^1_0(K_\delta), \delta > \varepsilon, \end{cases} \tag{96}$$

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

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

where $\chi^{\xi}(x_K,\cdot)$, for $\xi \in \mathbb{R}^d$, $K \in \mathscr{T}_H$, denote the exact solutions to the homogenization cell problems find $\bar{\chi}^{\xi}(x,\cdot) \in W^1_{per}(Y)$ such that

$$\int_Y \mathscr{A}(x,y,\xi + \nabla\bar{\chi}^{\xi}(x,y)) \cdot \nabla z\, dy = 0, \qquad \forall z \in W^1_{per}(Y), \tag{97}$$

and C is independent of $\Delta t, H, h, \varepsilon, \delta$ and v^H. We refer to [8] for a detailed proof of these micro and modelling a priori error estimates. We observe that the first two estimates for the modelling error are similar as for linear problem (see Section 3.1).

A better estimate can however be derived for linear problem for the third case for which it is possible to derive the estimate $(\delta+\varepsilon/\delta)$ (see again Section 3.1).

5 A linearized method

We consider again nonlinear monotone problems of the type (1) with strongly monotone and Lipschitz continuous maps $\mathscr{A}^{\varepsilon}(x,\xi)$, i.e., for the case $p=2$ and $\alpha=1$, $\beta=2$ in ($\mathscr{A}_{1-2}$). We further assume that the nonlinear map is of the form $\mathscr{A}^{\varepsilon}(x,\xi)=a^{\varepsilon}(x,\xi)\xi$. We first rewrite the method (66) in a slightly different form: find $u^H_{n+1}\in S^1_0(\Omega,\mathscr{T}_H)$ such that

$$\int_{\Omega}\frac{u^H_{n+1}-u^H_n}{\Delta t}w^H dx+B_H(u^H_{n+1};w^H)=\int_{\Omega}f\,w^H dx,\;\;\forall w^H\in S^1_0(\Omega,\mathscr{T}_H),\qquad(98)$$

with the nonlinear macro map B_H given by

$$B_H(v^H;w^H)=\sum_{K\in\mathscr{T}_H}\frac{|K|}{|K_\delta|}\int_{K_\delta}a^{\varepsilon}(x,\nabla\hat{v}^h_K)\nabla\hat{v}^h_K dx\cdot\nabla w^H(x_K),\;\;v^H,w^H\in S^1_0(\Omega,\mathscr{T}_H),\qquad(99)$$

and the micro functions v^h_K are given similarly to (12) by the following problem: find $\hat{v}^h_K$ such that $\hat{v}^h_K-v^H=v^h_K\in S^1(K_\delta,\mathscr{T}_h)$ and

$$\int_{K_\delta}a^{\varepsilon}(x,\nabla\hat{v}^h_K)\nabla\hat{v}^h_K\cdot\nabla w^h dx=0,\;\;\forall w^h\in S^1(K_\delta,\mathscr{T}_h).\qquad(100)$$

The equivalence of the above formulation and the one in (66) with micro problems given by (12) is easy to check. Following [5] we propose a linearized scheme. The idea is to decouple the micro-solutions in (99) and to consider

$$B_H(\hat{z};v^H,w^H)=\sum_{K\in\mathscr{T}_H}\frac{|K|}{|K_\delta|}\int_{K_\delta}a^{\varepsilon}(x,\nabla z^h_K)\nabla\hat{v}^{h,z^h_K}_K dx\cdot\nabla w^H(x_K),\;\;v^H,w^H\in S^1_0(\Omega,\mathscr{T}_H)\qquad(101)$$

where for given $\{z^h_K\}\in\prod_{K\in\mathscr{T}_H}S^1(K_\delta,\mathscr{T}_h)$, $\hat{v}^{h,z^h_K}_K$ is such that $\hat{v}^{h,z^h_K}_K-v^H=v^{h,z^h_K}_K\in S^1(K_\delta,\mathscr{T}_h)$ and solution of the *linear micro problem*

$$\int_{K_\delta}a^{\varepsilon}(x,\nabla z^h_K)\nabla\hat{v}^{h,z^h_K}_K\cdot\nabla w^h dx=0,\;\;\forall w^h\in S^1(K_\delta,\mathscr{T}_h).\qquad(102)$$

To formalize the numerical method we consider the product of FE spaces

$$\mathscr{S}^{H,h}=S^1_0(\Omega,\mathscr{T}_H)\times\prod_{K\in\mathscr{T}_H}S^1(K_\delta,\mathscr{T}_h),\qquad(103)$$

and define $\hat{z} = (z^H, \{z^h_K\}) \in \mathscr{S}^{H,h}$. Next for a given $\hat{u}_1 = (u^H_1, \{u^h_{1,K}\}) \in \mathscr{S}^{H,h}$ we define one step of the method as a map $\mathscr{S}^{H,h} \mapsto \mathscr{S}^{H,h}$ given by $\hat{u}_n = (u^H_n, \{u^h_{n,K}\}) \mapsto \hat{u}_{n+1} = (u^H_{n+1}, \{u^h_{n+1,K}\})$. To compute this map we implement the following two steps:

1. *update the macroscopic state:* find $u^H_{n+1} \in S^1_0(\Omega, \mathscr{T}_H)$, the solution of the linear problem

$$\int_\Omega \frac{1}{\Delta t}(u^H_{n+1} - u^H_n) w^H dx + B_H(\hat{u}_n; u^H_{n+1}, w^H) = \int_\Omega f w^H dx, \ \ \forall w^H \in S^1_0(\Omega, \mathscr{T}_H); \tag{104}$$

2. *update the microscopic states:* for each $K \in \mathscr{T}_H$, compute $\hat{v}_K^{h,u^h_{n,K}}$ such that $\hat{v}_K^{h,u^h_{n,K}} - u^H_{n+1} \in S^1(K_\delta, \mathscr{T}_h)$ and solution of (102) with parameter $u^h_{n,K}$ and update $u^h_{n+1,K} := \hat{v}_K^{h,u^h_{n,K}} - u^H_{n+1}$.

To completely describe the algorithm we still need to discuss the initialization procedure, i.e., how to define $\hat{u}_1 = (u^H_1, \{u^h_{1,K}\}) \in \mathscr{S}^{H,h}$ given the approximation $u^H_0 \in S^1_0(\Omega, \mathscr{T}_H)$ of the initial condition $g(x)$ of (5). We suggest to use one step of the nonlinear method (66) to set $\hat{u}_1$. This choice allows to prove optimal convergence rates. In turns out that the trivial initialisation obtained by setting $\hat{u}_0 = (u^H_0, \{0\})$ and using one step of the linearised method to define $\hat{u}_1$ deteriorates the accuracy of the linearised scheme [9]. It is also shown in [9] that the above linearised method is up to 10 times faster than the fully nonlinear method 66-67.

Well-posedness of the linearized method can be proved assuming that $a^\varepsilon(x,\xi)$ is uniformly elliptic and bounded, i.e.,

$$\lambda_a|\eta|^2 \leq a^\varepsilon(x,\xi)\eta\cdot\eta, \ |a^\varepsilon(x,\xi)\eta| \leq \Lambda_a|\eta|, \ \forall \xi,\eta \in \mathbb{R}^d, \text{ a.e. } x \in \Omega, \varepsilon > 0.$$

It then follows from similar argument as for linear elliptic problem [2] that

$$B_H(\hat{z}; v^H, v^H) \geq \lambda_a \|\nabla v^H\|^2_{L^2(\Omega)}, \ \ |B_H(\hat{z}; v^H, w^H)| \leq \frac{\Lambda_a^2}{\lambda_a}\|\nabla v^H\|_{L^2(\Omega)}\|\nabla w^H\|_{L^2(\Omega)}.$$

Combining the above estimate for B_H with the existence and uniqueness of the nonlinear initialisation obtained in Section 4 allows to prove existence and uniqueness of a solution to (104) and an a priori estimate similar to (68) with a right-hand side simply given by $C(\|f\|_{L^2(\Omega)} + \|u^H_0\|_{L^2(\Omega)})$.

5.1 A priori error estimates

Fully discrete a priori error estimates of the linearized method can be established following the steps of Section 4.2, with nontrivial modifications due to the linearisation procedure. It will be convenient in the sequel to introduce the two following semi-norm on the space $\mathscr{S}^{H,h}$. For $\hat{z} = (z^H, \{z_K^h\}) \in \mathscr{S}^{H,h}$ we therefore define

$$\|\nabla \hat{z}\|_{\mathscr{S}^{H,h}} = \left(\sum_{K\in\mathscr{T}_H} \frac{|K|}{|K_\delta|} \left\|\nabla \hat{z}_K^h\right\|_{L^2(K_\delta)}^2 \right)^{1/2}, \qquad \|\nabla \hat{z}\|_{\mathscr{S}_\infty^{H,h}} = \max_{K\in\mathscr{T}_H} \left\|\nabla \hat{z}_K^h\right\|_{L^\infty(K_\delta)},$$

where $\hat{z}_K^h = z_K^h + z^H$ on K_δ. In fact due to the Poincaré (or Poincaré-Wirtinger) inequaliy, $\|\cdot\|_{\mathscr{S}^{H,h}}$ is a norm. Observe that since $\int_{K_\delta} \nabla z_K^h dx \cdot \nabla z^H(x_K) = 0$ for micro spaces $S^1(K_\delta, \mathscr{T}_h)$ with periodic and Dirichlet boundary conditions we have $\left\|\nabla \hat{z}_K^h\right\|_{L^2(K_\delta)}^2 = \left\|\nabla z^H(x_K)\right\|_{L^2(K_\delta)}^2 + \left\|\nabla z_K^h\right\|_{L^2(K_\delta)}^2$, which yield for all $\hat{z} = (z^H, \{z_K^h\}) \in \mathscr{S}^{H,h}$ the inequality $\left\|\nabla z^H\right\|_{L^2(\Omega)} \leq \|\nabla \hat{z}\|_{\mathscr{S}^{H,h}}$.

Next consider the numerical solution obtained by the *linearized* multiscale method (104) $\hat{u}_n = (u_n^H, \{u_{n,K}^h\}) \in \mathscr{S}^{H,h}$ and set $\hat{u}_{n,K}^h = u_n^H + u_{n,K}^h$ on K_δ. We also define the nodal interpolation associated to the homogenized solution $\mathscr{U}_n^H = \mathscr{I}_H u^0(\cdot, t_n)$ and consider $\hat{\mathscr{U}}_n = (\mathscr{U}_n^H, \{\mathscr{U}_{n,K}^h\}) \in \mathscr{S}^{H,h}$ such that $\hat{\mathscr{U}}_{n,K}^h = \mathscr{U}_{n,K}^h + \mathscr{U}_n^H$ is the solution to the *nonlinear* micro problem (100). Define for $0 \leq n \leq N$ and $K \in \mathscr{T}_H$

$$\hat{\theta}_n = \hat{u}_n - \hat{\mathscr{U}}_n, \text{ i.e.,} \quad \theta_n^H = u_n^H - \mathscr{U}_n^H, \ \hat{\theta}_{n,K}^h = \hat{u}_{n,K}^h - \hat{\mathscr{U}}_{n,K}^h. \tag{105}$$

A formula similar to (92) leads to

$$\begin{aligned} &\frac{1}{2}\bar{\partial}_t \left\|\theta_n^H\right\|_{L^2(\Omega)}^2 + \lambda_a \left\|\nabla \hat{\theta}_{n+1}^H\right\|_{\mathscr{S}^{H,h}}^2 \\ &\leq C(\Delta t + H + r_{HMM}(\nabla \mathscr{U}_{n+1}^H)) \left\|\nabla \hat{\theta}_{n+1}^H\right\|_{L^2(\Omega)} + \left|L_n(\nabla \hat{\theta}_{n+1})\right|, \end{aligned} \tag{106}$$

where the additional term involves a function; $L_n \colon \mathscr{S}^{H,h} \to \mathbb{R}$ defined by

$$L_n(\nabla \hat{w}) = \sum_{K\in\mathscr{T}_H} \frac{|K|}{|K_\delta|} \int_{K_\delta} \left[a^\varepsilon(x, \nabla \hat{\mathscr{U}}_{n,K}^h) - a^\varepsilon(x, \nabla \hat{u}_{n,K}^h) \right] \nabla \hat{\mathscr{U}}_{n,K}^h \cdot \nabla \hat{w}_K^h dx. \tag{107}$$

This term arises from the linearization error and it can be bounded by

$$|L_n(\nabla \hat{w})| \leq \mathscr{L}_n \left\|\nabla \hat{\theta}_n\right\|_{\mathscr{S}^{H,h}} \|\nabla \hat{w}\|_{\mathscr{S}^{H,h}}, \tag{108}$$

where $\mathscr{L}_n$ will be discussed below. Hence using Young's inequality we obtain

$$\begin{aligned} &\frac{1}{2}\bar{\partial}_t \left\|\theta_n^H\right\|_{L^2(\Omega)}^2 + \lambda_a \left\|\nabla \hat{\theta}_{n+1}^H\right\|_{\mathscr{S}^{H,h}}^2 \\ &\leq C(\Delta t^2 + H^2 + r_{HMM}(\nabla \mathscr{U}_{n+1}^H)^2) + \frac{\mathscr{L}_n^2}{\lambda_a} \left\|\nabla \hat{\theta}_n\right\|_{\mathscr{S}^{H,h}}^2 + \frac{\lambda_a}{2} \left\|\nabla \hat{\theta}_{n+1}\right\|_{\mathscr{S}^{H,h}}^2. \end{aligned} \tag{109}$$

Recall that we use the fully nonlinear method for the first step. Hence the convergence results of Section 4.2 yield

$$\left\|\theta_1^H\right\|_{L^2(\Omega)}^2-\left\|\theta_0^H\right\|_{L^2(\Omega)}^2+\lambda\Delta t\left\|\nabla\theta_1^H\right\|_{L^2(\Omega)}^2\le C\Delta t\left(\Delta t^2+H^2+r_{HMM}(\nabla\mathscr{U}_1^H)^2\right),\tag{110}$$

where λ is the monotonicity constant of $\mathscr{A}^\varepsilon$.

Similarly to (94) summing (109) from $n=1$ to $n=N-1$, adding the term (110) and using the inequality $\left\|\nabla z^H\right\|_{L^2(\Omega)}\le\|\nabla\hat{z}\|_{\mathscr{S}^{H,h}}$ gives

$$\begin{aligned}&\max_{1\le n\le N}\left\|\theta_n^H\right\|_{L^2(\Omega)}^2+\lambda\Delta t\left\|\nabla\theta_1^H\right\|_{L^2(\Omega)}^2+C_{\mathscr{L}}\Delta t\sum_{n=2}^N\left\|\nabla\theta_n^H\right\|_{L^2(\Omega)}^2 &(111)\\&\le C\left(\Delta t^2+H^2+\max_{1\le n\le N}r_{HMM}(\nabla\mathscr{U}_n^H)^2\right)+\left\|\theta_0^H\right\|_{L^2(\Omega)}^2+\frac{2}{\lambda_a}\Delta t\mathscr{L}_1^2\left\|\nabla\hat{\theta}_1\right\|_{\mathscr{S}^{H,h}}^2.\end{aligned}$$

where $C_{\mathscr{L}}=\lambda_a-\frac{2}{\lambda_a}\max_{2\le n\le N-1}\mathscr{L}_n^2$. Recall that $\hat{\theta}_{1,K}^h=\hat{u}_{1,K}^h-\hat{\mathscr{U}}_{1,K}^h$ where $\hat{u}_{K,1}^h$ and $\hat{\mathscr{U}}_{K,1}^h$ are solutions to the nonlinear micro problem (100) constrained by u_1^H and $\mathscr{U}_1^H$, respectively. The difference of two such micro solutions can be estimated by the difference of their respective macro constraints as

$$\left\|\nabla\hat{\theta}_{1,K}^h\right\|_{L^2(K_\delta)}\le\frac{L}{\lambda}\sqrt{|K_\delta|}\left|\nabla u_1^H(x_K)-\nabla\mathscr{U}_1^H(x_K)\right|,\tag{112}$$

hence $\left\|\nabla\hat{\theta}_1\right\|_{\mathscr{S}^{H,h}}\le\frac{L}{\lambda}\left\|\nabla\theta_1^H\right\|_{L^2(\Omega)}$. Assuming $\mathscr{L}_1$ is bounded and $C_{\mathscr{L}}>0$ we obtain

$$\begin{aligned}&\max_{1\le n\le N}\left\|\theta_n^H\right\|_{L^2(\Omega)}^2+\Delta t\sum_{n=1}^N\left\|\nabla\hat{\theta}_n\right\|_{L^2(\Omega)}^2 &(113)\\&\le C\left(\Delta t^2+H^2+\max_{1\le n\le N}r_{HMM}(\nabla\mathscr{U}_n^H)^2\right)+C\left\|\theta_0^H\right\|_{L^2(\Omega)}^2. &(114)\end{aligned}$$

Finally as $\left\|\theta_0^H\right\|_{L^2(\Omega)}\le\left\|u_0^H-g\right\|_{L^2(\Omega)}+\left\|g-\mathscr{U}_0^H\right\|_{L^2(\Omega)}$ using the bound $\left\|g-\mathscr{U}_0^H\right\|_{L^2(\Omega)}\le CH$ gives an estimate similar to (94). In view of (113), classical estimates for nodal interpolants give under the assumptions of Theorem 6, provided $\mathscr{L}_1$ is bounded and $C_{\mathscr{L}}>0$, the error estimate (91).

We briefly discuss the additional assumptions on $\mathscr{L}_1$ and $C_{\mathscr{L}}$. These assumptions can be derived in two ways [9]. Under some regularity assumptions on the exact solutions of the micro problems (100), assuming that $u^0\in\mathscr{C}^0([0,T],W^{2,\infty}(\Omega))$ and $\max_{t\in[0,T]}\left|u^0(x,t)\right|_{W^{1,\infty}(\Omega)}$ is small enough (smallness assumption), then there exist H_0,h_0 such that for any $H<H_0,h<h_0$, $C_{\mathscr{L}}>0$ and $\mathscr{L}_1$ is bounded. Alternatively we can prove the boundedness of $\mathscr{L}_1$ and the positivity of $C_{\mathscr{L}}$ without a smallness assumption on u^0 and without the additional regularity assumption $\mathscr{C}^0([0,T],W^{2,\infty}(\Omega))$ on u^0 if in addition to the assumptions of Theorem 6 we have

$$\max_{\substack{K\in\mathscr{T}_H\\1\le n\le N-1}}\|e_{n,K}\|_{(L^\infty(K_\delta))^{d\times d}}<\frac{\lambda_a}{2\sqrt{2}},\tag{115}$$

where the error term $e_{n,K} \in (L^\infty(K_\delta))^{d\times d}$ is given by

$$e_{n,K}(x) = a^\varepsilon(x, \nabla \hat{u}^h_{n,K}) - \int_0^1 a^\varepsilon(x, \nabla \hat{u}^h_{n,K} - \tau \nabla \hat{\theta}^h_{n,K}) d\tau, \qquad \text{a.e. } x \in K_\delta. \tag{116}$$

The term (116) represent the linearization error. It has been shown numerically for tensors with various ellipticity constant λ_a that (116) holds if the spatial and temporal discretization parameters are small enough [9]. Optimal (discrete) $\mathscr{C}^0(L^2)$ can also be derived under the same additional assumptions as for the fully nonlinear method. Finally fully discrete results, i.e., quantitative estimates for the component r_{mic} and r_{mod} of r_{HMM} can be obtained similarly as in Section 4.2, with similar rates.

6 Conclusion

We have presented a unified framework and analysis for the FE-HMM applied to monotone parabolic problems. We have shown that under the most general assumptions for which homogenization can be established, we can construct an FE-HMM and establish its convergence. Under more restrictive assumptions, e.g. Lipschitz continuous and strongly monotone maps, fully discrete space time a priori error estimates can be derived and in some situation an efficient linearized scheme can be constructed and analyzed. Finally for linear problems we have shown that the FE-HMM can be coupled with classes of Runge-Kutta methods (Radau or Chebyshev methods) and analyzed by combining fully discrete spatial estimates with semi-group techniques in a Hilbert space framework. We have neither discussed implementation issue nor given numerical experiments. This is carefully documented in [12, 8, 9], where the issue of choosing the right coupling of the micro and macro solvers (i.e., the micro boundary conditions) and the size of the sampling domains are discussed. Numerical experiments for non-periodic problems (e.g., log-normal stochastic field) [12] and degenerate problems [9] illustrate the robustness of the numerical homogenization strategy.

Acknowledgements This research is partially supported by the Swiss National Foundation under Grant 200021_150019.

References

1. A. ABDULLE, *Fourth order Chebyshev methods with recurrence relation*, SIAM J. Sci. Comput., 23 (2002), pp. 2041–2054.
2. ———, *On a priori error analysis of fully discrete heterogeneous multiscale FEM*, Multiscale Model. Simul., 4 (2005), pp. 447–459.
3. ———, *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.
4. ———, *A priori and a posteriori error analysis for numerical homogenization: a unified framework*, Ser. Contemp. Appl. Math. CAM, 16 (2011), pp. 280–305.
5. A. ABDULLE, Y. BAI, AND G. VILMART, *An offline-online homogenization strategy to solve quasilinear two-scale problems at the cost of one-scale problems*, Internat. J. Numer. Methods Engrg., 99 (2014), pp. 469–486.
6. A. ABDULLE AND W. E, *Finite difference heterogeneous multi-scale method for homogenization problems*, J. Comput. Phys., 191 (2003), pp. 18–39.
7. A. ABDULLE AND M. E. HUBER, *Error estimates for finite element approximations of nonlinear monotone elliptic problems with application to numerical homogenization*. to appear in Numer. Methods Partial Differential Equations.
8. ———, *Finite element heterogeneous multiscale method for nonlinear monotone parabolic homogenization problems*. to appear in ESAIM Math. Model. Numer. Anal.
9. A. ABDULLE, M. E. HUBER, AND G. VILMART, *Linearized numerical homogenization method for nonlinear monotone parabolic multiscale problems*, Multiscale Model. Simul., 13 (2015), pp. 916–952.
10. A. ABDULLE AND A. A. MEDOVIKOV, *Second order Chebyshev methods based on orthogonal polynomials*, Numer. Math., 90 (2001), pp. 1–18.
11. A. ABDULLE AND C. SCHWAB, *Heterogeneous multiscale FEM for diffusion problems on rough surfaces*, Multiscale Model. Simul., 3 (2005), pp. 195–220.
12. 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.
13. ———, *Analysis of the finite element heterogeneous multiscale method for quasilinear elliptic homogenization problems*, Math. Comp., 83 (2014), pp. 513–536.
14. G. ALLAIRE AND M. BRIANE, *Multiscale convergence and reiterated homogenisation*, Proc. Roy. Soc. Edinburgh Sect. A, 126 (1996), pp. 297–342.
15. G. ALLAIRE AND R. BRIZZI, *A multiscale finite element method for numerical homogenization*, Multiscale Model. Simul., 4 (2005), pp. 790–812 (electronic).
16. I. BABUŠKA AND J. E. OSBORN, *Generalized finite element methods: their performance and their relation to mixed methods*, SIAM J. Numer. Anal., 20 (1983), pp. 510–536.
17. J. W. BARRETT AND W. B. LIU, *Finite element approximation of the parabolic p-Laplacian*, SIAM J. Numer. Anal., 31 (1994), pp. 413–428.
18. A. BENSOUSSAN, J.-L. LIONS, AND G. PAPANICOLAOU, *Asymptotic analysis for periodic structures*, North-Holland Publishing Co., Amsterdam, 1978.
19. L. BERS, F. JOHN, AND M. SCHECHTER, *Partial differential equations*, vol. Proceedings of the summer seminar of Lectures in Applied Mathematics, Proceeding of the Summer Seminar, Boulder, CO, 1957.
20. M. E. BREWSTER AND G. BEYLKIN, *A multiresolution strategy for numerical homogenization*, Appl. Comput. Harmon. Anal., 2 (1995), pp. 327–349.
21. V. CHIADÒ PIAT, G. DAL MASO, AND A. DEFRANCESCHI, *G-convergence of monotone operators*, Ann. Inst. H. Poincaré Anal. Non Linéaire, 7 (1990), pp. 123–160.
22. P. G. CIARLET, *The finite element method for elliptic problems*, vol. 4 of Studies in Mathematics and its Applications, North-Holland, 1978.
23. 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.
24. M. CROUZEIX, *Approximation of parabolic equations*, 2005. Lecture notes available at `http://perso.univ-rennes1.fr/michel.crouzeix/`.
25. M. CROUZEIX, S. LARSSON, S. PISKARËV, AND V. THOMÉE, *The stability of rational approximations of analytic semigroups*, BIT, 33 (1993), pp. 74–84.
26. G. DAL MASO AND A. DEFRANCESCHI, *Correctors for the homogenization of monotone operators*, Differential Integral Equations, 3 (1990), pp. 1151–1166.

27. L. DIENING, C. EBMEYER, AND M. RŮŽIČKA, *Optimal convergence for the implicit space-time discretization of parabolic systems with p-structure*, SIAM J. Numer. Anal., 45 (2007), pp. 457–472 (electronic).
28. R. DU AND P. MING, *Heterogeneous multiscale finite element method with novel numerical integration schemes*, Commun. Math. Sci., 8 (2010), pp. 863–885.
29. W. E AND B. ENGQUIST, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1 (2003), pp. 87–132.
30. 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.
31. Y. EFENDIEV AND T. Y. HOU, *Multiscale finite element methods. Theory and applications*, vol. 4 of Surveys and Tutorials in the Applied Mathematical Sciences, Springer, New York, 2009.
32. Y. EFENDIEV AND A. PANKOV, *Numerical homogenization of nonlinear random parabolic operators*, Multiscale Model. Simul., 2 (2004), pp. 237–268.
33. B. ENGQUIST AND O. RUNBORG, *Wavelet-based numerical homogenization with applications*, in Multiscale and multiresolution methods, vol. 20 of Lect. Notes Comput. Sci. Eng., Springer, Berlin, 2002, pp. 97–148.
34. M. G. D. GEERS, V. G. KOUZNETSOVA, AND W. A. M. BREKELMANS, *Multi-scale computational homogenization: Trends and challenges*, J. Comput. Appl. Math., 234 (2010), pp. 2175–2182.
35. 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).
36. ———, *Reduction of the resonance error. Part 1: Approximation of homogenized coefficients*, Math. Models Methods Appl. Sci., 21 (2011), pp. 1601–1630.
37. E. HAIRER AND G. WANNER, *Solving ordinary differential equations II. Stiff and differential-algebraic problems*, Springer-Verlag, Berlin and Heidelberg, 1996.
38. P. HENNING AND M. OHLBERGER, *Error control and adaptivity for heterogeneous multiscale approximations of nonlinear monotone problems*, Discrete Contin. Dyn. Syst. Ser. S, 8 (2015), pp. 119–150.
39. P. HENNING AND D. PETERSEIM, *Oversampling for the Multiscale Finite Element Method*, SIAM Multiscale Model. Simul., 11 (2013), pp. 1149–1175.
40. T. Y. HOU, X.-H. WU, AND Z. CAI, *Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients*, Math. Comp., 68 (1999), pp. 913–943.
41. T. J. R. HUGHES, G. R. FEIJÓO, L. MAZZEI, AND J.-B. QUINCY, *The variational multiscale method—a paradigm for computational mechanics*, Comput. Methods Appl. Mech. Engrg., 166 (1998), pp. 3–24.
42. V. V. JIKOV, S. M. KOZLOV, AND O. A. OLEINIK, *Homogenization of differential operators and integral functionals*, Springer-Verlag, Berlin, Heidelberg, 1994.
43. T. KATO, *Perturbation theory for linear operators*, Springer-Verlag, Berlin, 1976.
44. O. A. LADYZHENSKAYA, *The boundary value problems of mathematical physics*, vol. 49 of Applied Mathematical Sciences, Springer-Verlag, New York, 1985.
45. V. LEBEDEV, *How to solve stiff systems of differential equations by explicit methods*, in Numerical methods and applications, G. I. Marčuk, ed., CRC Press, 1994, pp. 45–80.
46. C. LUBICH AND A. OSTERMANN, *Runge-Kutta approximation of quasi-linear parabolic equations*, Math. Comp., 64 (1995), pp. 601–627.
47. A. MÅLQVIST AND A. PERSSON, *Multiscale techniques for parabolic equations*, arXiv:1504.08140, (2015).
48. A. MÅLQVIST AND D. PETERSEIM, *Localization of elliptic multiscale problems*, Math. Comp., 83 (2014), pp. 2583–2603.
49. P. MING AND P. ZHANG, *Analysis of the heterogeneous multiscale method for parabolic homogenization problems*, Math. Comp., 76 (2007), pp. 153–177.
50. G. NGUETSENG, *A general convergence result for a functional related to the theory of homogenization*, SIAM J. Math. Anal., 20 (1989), pp. 608–623.

51. H. Owhadi and L. Zhang, *Metric-based upscaling*, Comm. Pure Appl. Math., 60 (2007), pp. 675–723.
52. A. Pankov, *G-convergence and homogenization of nonlinear partial differential operators*, vol. 422 of Mathematics and its Applications, Kluwer Academic Publishers, Dordrecht, 1997.
53. B. P. Sommeijer and J. G. Verwer, *A performance evaluation of a class of Runge-Kutta-Chebyshev methods for solving semi-discrete parabolic differential equations*, tech. rep., Stichting Mathematisch Centrum. Numerieke Wiskunde, Amsterdam, 1980. Report NW91/80.
54. N. Svanstedt, *G-convergence of parabolic operators*, Nonlinear Anal., 36 (1999), pp. 807–842.
55. V. Thomée, *Galerkin finite element methods for parabolic problems*, vol. 25 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 1997.
56. P. van der Houwen and B. P. Sommeijer, *On the internal stage Runge-Kutta methods for large m-values*, Z. Angew. Math. Mech., 60 (1980), pp. 479–485.
57. 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.