

# ANGEWANDTE MATHEMATIK UND INFORMATIK

A-posteriori error estimation for a heterogeneous  
multiscale method for monotone operators and  
beyond a periodic setting

Patrick Henning, Mario Ohlberger

01/11 - N



UNIVERSITÄT MÜNSTER



# A-POSTERIORI ERROR ESTIMATION FOR A HETEROGENEOUS MULTISCALE METHOD FOR MONOTONE OPERATORS AND BEYOND A PERIODIC SETTING<sup>1</sup>

PATRICK HENNING AND MARIO OHLBERGER

## Abstract

In this work we introduce and analyse a heterogeneous multiscale finite element method (HMM) for monotone elliptic operators with rapid oscillations. The concept of HMM was initially introduced by E and Engquist [E, Engquist, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1(1):87–132, 2003]. We first present a macroscopic limit problem for the oscillating non-linear equations and then prove the convergence of the HMM approximations to the solution of the macroscopic limit equation. On the basis of this identification, we derive an a-posteriori error estimate with duality techniques. The general applicability of the method and its corresponding error estimate is demonstrated in numerical experiments. In particular we state and examine two strategies for adaptive mesh refinement based on the error estimate.

## 1 Introduction

This paper is devoted to a heterogeneous multiscale finite element method for solving nonlinear elliptic problems with fast oscillations, i.e. we are looking for the solution  $u^\epsilon$  of the following type of equations:

$$\begin{aligned} -\nabla \cdot A^\epsilon(x, \nabla u^\epsilon) &= f \text{ in } \Omega, \\ u^\epsilon &= 0 \text{ on } \partial\Omega. \end{aligned} \tag{1}$$

Here,  $A^\epsilon(x, \cdot)$  is a monotone function, whereas  $A^\epsilon(\cdot, \xi)$  contains a fast, microscopic behaviour. The parameter  $\epsilon$  is a (possibly abstract) characterization of the multiscale nature of  $A^\epsilon$ . In periodic settings it typically represents the period length of one microscopic oscillation, i.e. we have  $A^\epsilon(x, \cdot) = A(x, \frac{x}{\epsilon} - \lfloor \frac{x}{\epsilon} \rfloor, \cdot)$  for suitable  $A$ .

Such equations have a wide range of applications especially in hydrology and industrial engineering. However, solving the problems accurately typically results in a tremendous computational demand, since the microstructure has to be resolved completely. In a lot of scenarios it is impossible to approach this issue with standard methods such as pure finite elements. Therefore, the usage of alternative strategies is indispensable. There is a large field of methods which are specifically designed for solving multiscale problems.

A well known analytical approach is homogenization of the original equation. Here, we let  $\epsilon$  tend to zero to identify a coarse-scale limit problem which is cheap to solve. This procedure is primarily restricted to periodic and stochastic settings, even though there are exceptions. The treatment of nonlinear elliptic problems is for instance presented in contributions of Allaire [8] and Wall *et al* [46, 34, 35].

A first example for a suitable numerical method is the Multiscale Finite Element Method (MsFEM) developed by Hou and Wu [27]. Here, local fine scale problems are solved in order to construct a set of multiscale basis functions which contain information about the small scale behaviour. A global variational formulation couples these basis functions so that we obtain an accurate approximation of the original problem. This powerful method can be also applied to heterogeneous structures. Concerning linear elliptic equations we refer to [27] and [28]. Nonlinear

---

<sup>1</sup>This work was supported by the Deutsche Forschungsgemeinschaft under the contract number OH 98/4-1.

problems are treated by Efendiev, Hou and Ginting [19] and two phase flow in porous media by Efendiev and Hou [18]. A multiscale finite element method for elliptic interface problems with high contrast coefficients, i.e. where the computational domain contains inclusion which are highly permeable in comparison to the rest of the domain or the case vice versa, is presented in [11]. A perfect overview on the topic of MsFEM can be found in the book of Efendiev and Hou [20].

The two-scale finite element method developed by Schwab and Matache [36, 37, 44] is based on a discretization of the so-called two-scale homogenized equation (see for instance [8]). Using sparse grids as suggested by Hoang and Schwab [26], the computational demand can be reduced significantly. Formally, this method is restricted to the periodic setting even though extensions are possible.

In [10] Arbogast *et al.* introduce a multiscale mortar mixed finite element discretizations for elliptic problems. Here,  $\Omega$  is subdivided into coarse subdomains, on which the original problem is posed. These subdomains are discretized on a very fine grid scale and are stringed together by a low degree-of-freedom mortar space.

Another approach, basing on the works of Hughes *et al.* [29, 30], is the adaptive variational multiscale method which involves a splitting of the solution into coarse and fine-scale contributions. Here, the fine-scale equations are solved in dependency of the residual of the coarse-scale solution. Adaptive algorithms are constructed on the basis of a corresponding a-posteriori error estimate. In [31, 32], Larson and Målqvist treat diffusion dominated elliptic problems and in [33] they treat stationary advection-diffusion problems. A general framework for adaptive multiscale methods for elliptic problems was suggested by Nolen, Papanicolaou and Pironneau [39].

Under the assumption that a regularized problem is available, Goal-Oriented adaptive strategies are suggested by Oden and Vemaganti [41, 45]. Here, those grid cells are determined where the error between the exact solution and a regularized approximation is still too large. In these cells, fine scale problems are solved for a local correction of the old approximation.

In this contribution we discuss the concept of heterogeneous multiscale finite element methods (HMM). It was originally introduced by E and Engquist in 2003 [14, 15, 16]. The idea is to perform detailed fine-scale computations in a certain number of small cells around quadrature points. An average of these results is passed to a coarse grid discretization of the original problem to obtain effective macroscopic properties.

Linear elliptic problems were treated in contributions of E, Ming and Zhang [17], Abdulle and Schwab [7], Ohlberger [42] and Henning and Ohlberger [24]. A combination of HMM and Newton method for solving nonlinear elliptic problems was presented in [22]. Heterogeneous multiscale methods for parabolic equations are stated by Abdulle and E [4], Ming and Zhang [38] and Henning and Ohlberger [25]. A combination of a HMM and an Orthogonal Runge-Kutta Chebyshev (ROCK) method is suggested by Abdulle [2] for solving advection-diffusion problems. A HMM realization with near optimal computational complexity is given by Abdulle and Engquist [5]. A good overview on the whole topic can be found in [1].

Among others, convergence results concerning HMM in the periodic setting were achieved by E *et al.* [14, 17], Abdulle [2, 3], Ohlberger [42] and Henning and Ohlberger [24, 25].

A-posteriori error estimates for the heterogeneous multiscale method for elliptic problems in the periodic setting can be found in [42] (in the energy norm) and in [24] (in the  $L^2$ -norm). A general a-posteriori estimate in dependency of the homogenized matrix is given by Abdulle and Nonnenmacher [6]. An a-posteriori result obtained by Larson and Målqvist [33] for the variational multiscale method for convection-diffusion problems can be also transferred to the framework of heterogeneous multiscale methods. Error estimation for a HMM for advection-diffusion problems with large expected drift was derived in [23].

However, a missing piece on this field is still the achievement of an a-posteriori error es-

timate, completely independent of any assumptions on the microstructure of  $A^\epsilon$ . The goal of this contribution is to close this gap. Here, we start with formulating a HMM for nonlinear elliptic problems. On the basis of this formulation, we identify the limit  $u^c$  of a sequence of HMM approximations and state a corresponding homogenized problem that is fulfilled by  $u^c$ . Now we can split the error  $\|u_{HMM} - u^\epsilon\|_{L^2(\Omega)}$  into  $\|u_{HMM} - u^c\|_{L^2(\Omega)}$  and  $\|u^c - u^\epsilon\|_{L^2(\Omega)}$ . In this work, we only estimate  $\|u_{HMM} - u^c\|_{L^2(\Omega)}$  and assume that the remaining modeling error remains small. However, in the linear setting, estimates for  $\|u^c - u^\epsilon\|_{L^2(\Omega)}$  are already at hand (see [47, 40, 41, 45] ) and could be transferred to the nonlinear setting. Since this is a work on its own, we shall ignore that part in this paper.

*Outline:* In Section 2 we introduce the heterogeneous multiscale finite element method for monotone operators. In particular, this method can be used to efficiently determine the homogenized solution of a nonlinear elliptic multiscale problem with  $\epsilon$ -periodic coefficients (see [22] for comparison). In Section 3 we state a regularized equation of (1) and prove that a sequence of HMM approximations converges to the solution of this problem. In Section 4 we state an a-posteriori estimate for the error between HMM approximation and regularized solution. The estimate is proved in Section 5. The method and its estimate are numerically validated in Section 6. Here, we also introduce two algorithms for an adaptive mesh refinement of the coarse grid.

## 2 The non-linear elliptic problem and a corresponding HMM

### 2.1 Problem and analytical setting

In the following we are concerned with solving the subsequent nonlinear elliptic multiscale problem: find  $u^\epsilon \in \dot{H}^1(\Omega)$  with

$$\int_{\Omega} A^\epsilon(x, \nabla u^\epsilon) \cdot \nabla \Phi(x) \, dx = \int_{\Omega} f(x) \Phi(x) \, dx \quad \forall \Phi \in \dot{H}^1(\Omega). \quad (2)$$

Here,  $\Omega \subset \mathbb{R}^d$  denotes a polygonal bounded domain with dimension  $d \leq 3$ . To ensure existence and uniqueness of  $u^\epsilon \in \dot{H}^1(\Omega)$  we need some additional assumptions on  $A^\epsilon \in (L^\infty(\Omega \times \mathbb{R}^d))^d$ :

**Assumption 2.1** (Continuity and monotonicity of  $A^\epsilon$ ). Let  $|\cdot|$  denote the Euclidean norm on  $\mathbb{R}^d$ . We assume that there exist two constants  $0 < \alpha \leq \beta < \infty$  such that uniformly  $\forall x \in \Omega$ :

$$\begin{aligned} (A^\epsilon(x, \xi_1) - A^\epsilon(x, \xi_2), \xi_1 - \xi_2) &\geq \alpha |\xi_1 - \xi_2|^2, \\ |A^\epsilon(x, \xi_1) - A^\epsilon(x, \xi_2)| &\leq \beta |\xi_1 - \xi_2| \text{ and} \\ A^\epsilon(x, 0) &= 0. \end{aligned}$$

To formulate the HMM, we need several definitions. By  $Y = (-\frac{1}{2}, \frac{1}{2})^d$  we denote the 0-centered unit cube. Scaling  $Y$  with  $\kappa \in \mathbb{R}_{>0}$  and shifting it by  $x \in \mathbb{R}^d$ , we use the notation  $x + \kappa Y := \{x + \kappa y \mid y \in Y\}$ . We define the space of  $Y$ -periodic  $H^1$ -functions with zero average by

$$\tilde{H}_\#^1(Y) := \left\{ \phi \in C^0(Y) \mid \phi \text{ is } Y\text{-periodic and } \int_Y v(y) \, dy = 0 \right\}^{\|\cdot\|_{H^1(Y)}}.$$

By  $|\cdot|_{H^k(\Omega)}$  we denote the semi-norm on  $H^k(\Omega)$  and by  $\|\cdot\|_{H^k(\Omega)}$  we denote the full norm ( $k \in$

$\mathbb{N}_{\geq 1}$ ). Moreover we introduce the following (semi-)norms on the Bochner-space  $L^2(\Omega, H^k(Y))$ :

$$|\Phi|_{L^2(\Omega, H^k(Y))} := \left( \int_{\Omega} |\Phi(x, \cdot)|_{H^k(Y)}^2 \right)^{\frac{1}{2}} \quad \text{and} \quad \|\Phi\|_{L^2(\Omega, H^k(Y))} := \sum_{l=0}^k |\Phi|_{L^2(\Omega, H^l(Y))}.$$

In order to create a suitable discrete setting where we can formulate the multiscale method in, we require additional definitions. By  $\mathcal{T}_H$  we denote a regular simplicial partition of  $\Omega$  with elements  $T$ .  $x_T$  is the barycenter of  $T \in \mathcal{T}_H$  and

$$\Gamma(\mathcal{T}_H(\Omega)) := \{E \mid E = \overline{T_1} \cap \overline{T_2} \neq \emptyset, \text{codim}(E) = 1 \text{ and } T_1, T_2 \in \mathcal{T}_H\}$$

the set of edges without the boundary edges. Analogously we introduce  $\mathcal{T}_h$ , a regular, periodic partition of  $Y$ .  $\Gamma(\mathcal{T}_h)$  is the corresponding set of edges (including the edges on  $\partial Y$ ) and  $y_K$  the barycenter of  $K \in \mathcal{T}_h$ . For  $\kappa \in \mathbb{R}_{>0}$  and  $T \in \mathcal{T}_H$  we use the notation  $Y_{T,\kappa} := x_T + \kappa Y$ . The mapping  $x_T^\kappa : Y \rightarrow Y_{T,\kappa}$  is given by  $x_T^\kappa(y) := x_T + \kappa y$ . The local mesh sizes  $H_T$  and  $h_K$  are defined by  $H_T := \text{diam}(T)$ ,  $T \in \mathcal{T}_H$  and  $h_K := \text{diam}(K)$ ,  $K \in \mathcal{T}_h$ . Analogously we denote  $H_E := \text{diam}(T \cup \tilde{T})$ , where  $E = T \cap \tilde{T} \in \Gamma(\mathcal{T}_H)$  and  $h_{E_Y} := \text{diam}(K \cup \tilde{K})$ . The subset  $\Omega_{\mathcal{T}_H}^\delta$  of  $\Omega$  is called the *representative part of  $\Omega$* . It is defined by

$$\Omega_{\mathcal{T}_H}^\delta := \bigcup_{T \in \mathcal{T}_H} Y_{T,\delta} \tag{3}$$

and includes only the regions, where fine-scale behaviour is relevant for the HMM-computations (i.e. we only perform computations in  $\Omega_{\mathcal{T}_H}^\delta$ ). Microscopic oscillations in  $\Omega \setminus \Omega_{\mathcal{T}_H}^\delta$  are not captured by the multiscale method, even though extrapolation may be applied. With regard to the representative part of  $\Omega$ , it is sufficient to restrict the discretized diffusion operator on  $\Omega_{\mathcal{T}_H}^\delta$ :

**Definition 2.2** (Discrete elliptic operator). Assuming sufficient regularity for this formulation, we define the (piecewise constant) discretized monotone elliptic operator by:

$$A_h^\epsilon : \Omega_{\mathcal{T}_H}^\delta \times \mathbb{R}^d \rightarrow \mathbb{R}^d \quad \text{and for } T \in \mathcal{T}_H, K \in \mathcal{T}_h, \xi \in \mathbb{R}^d : \\ A_h^\epsilon(\cdot, \xi)|_{x_T^\delta(K)} := A^\epsilon(x_T^\delta(y_K), \xi).$$

The discrete space of HMM approximations is given by

$$V_H := \{\Phi_H \in \tilde{H}^1(\Omega) \cap C^0(\Omega) \mid \Phi_{H|_T} \in \mathbb{P}^1(T) \quad \forall T \in \mathcal{T}_H\}$$

and the solution spaces for the local reconstructions by

$$W_h(Y) := \{\phi_h \in \tilde{H}_\#^1(Y) \cap C^0(Y) \mid \phi_h|_K \in \mathbb{P}^1(K) \quad \forall K \in \mathcal{T}_h\}; \text{ and} \\ W_h(Y_{T,\delta}) := \{\phi_h \in H^1(Y_{T,\delta}) \cap C^0(Y_{T,\delta}) \mid (\phi_h \circ x_T^\delta) \in W_h(Y)\}.$$

**Definition 2.3** (Reconstruction operator).

For  $\delta > 0$  and  $Y_{T,\delta} \subset T$ , we introduce the *local reconstruction operator*

$$R_T : V_H \rightarrow V_H + W_h(Y_{T,\delta}).$$

For  $\Phi_H \in V_H$ , the corresponding reconstruction  $R_T(\Phi_H) \in \Phi_H + W_h(Y_{T,\delta})$  is defined as the solution of

$$\int_{Y_{T,\delta}} A_h^\epsilon(x, \nabla_x R_T(\Phi_H)(x)) \cdot \nabla_x \phi_h(x) \, dx = 0 \quad \forall \phi_h \in W_h(Y_{T,\delta}). \tag{4}$$

Note that  $R_T$  is well defined, since the so-called cell problems (4), always have a unique solution. For instance, if we define an operator  $B : H^1(Y_{T,\delta}) \times H^1(Y_{T,\delta}) \rightarrow \mathbb{R}$  for  $b \in (L^\infty(Y_{T,\delta}))^d$  by

$$B(v_h, \phi_h) := \int_{Y_{T,\delta}} A_h^\epsilon(t, x, b(x) + \nabla_x v_h(x)) \cdot \nabla_x \phi_h(x) \, dx.$$

Then  $B$  is still strongly monotone (due to the properties of  $A_h^\epsilon$ ) and therefore coercive in the sense of  $\lim_{\|v_h\| \rightarrow \infty} \frac{B(v_h, v_h)}{\|v_h\|} = \infty$ . The problem 'find  $v_h \in H^1(Y_{T,\delta})$  with  $B(v_h, \phi_h) = 0$  for all  $\phi_h \in H^1(Y_{T,\delta})$ ', therefore has a unique solution.

## 2.2 The heterogeneous multiscale finite element method for monotone operators

Before stating the method, we want to sketch the idea. We note that the subsequent procedure (for deriving one realization of the HMM) is equal to the approach in the setting of the Variational Multiscale Method (VMM, cf. [32]). Let us assume that we have a separation of coarse- and fine-scale part:  $u^\epsilon \approx u^{\text{coarse}} + \epsilon u^{\text{fine}}$ . If we suppose the same for our test functions (i.e.  $\Phi = \Phi^{\text{coarse}} + \epsilon \phi^{\text{fine}}$ ), we obtain with regard to (2):

$$\int_{\Omega} A^\epsilon(\cdot, \nabla u^{\text{coarse}} + \epsilon u^{\text{fine}}) \cdot \nabla (\Phi^{\text{coarse}} + \epsilon \phi^{\text{fine}}) \approx \int_{\Omega} f(\Phi^{\text{coarse}} + \epsilon \phi^{\text{fine}}).$$

Choosing  $\phi^{\text{fine}} = 0$  we have

$$\int_{\Omega} A^\epsilon(\cdot, \nabla u^{\text{coarse}} + \epsilon u^{\text{fine}}) \cdot \nabla \Phi^{\text{coarse}} \approx \int_{\Omega} f \Phi^{\text{coarse}} \quad (5)$$

and with  $\Phi^{\text{coarse}} = 0$  we get

$$\int_{\Omega} A^\epsilon(\cdot, \nabla u^{\text{coarse}} + \epsilon u^{\text{fine}}) \cdot \nabla \epsilon \phi^{\text{fine}} \approx \epsilon \int_{\Omega} f \phi^{\text{fine}} \approx 0. \quad (6)$$

Discretizing (5) and (6) and restricting the computations on representative cells  $x_{\text{quad}} + \delta Y$  yields the basic concept of HMM. Note that  $\nabla \epsilon \phi^{\text{fine}}$  does not become small, due to the possibly large gradient behaving like  $\frac{1}{\epsilon}$ .

In the following we introduce two slightly different versions of a heterogeneous multiscale finite element method for monotone elliptic problems. We start with the canonical formulation of the method, as initially suggested by E and Engquist [14] for linear elliptic equations. For distinction we refer to *TFR* in this case, which abbreviates *test function reconstruction*.

**Definition 2.4** (TFR - Heterogeneous Multiscale Method for monotone operators).

We define the TFR-HMM approximation  $u_{H,h}^S$  of  $u^\epsilon$  by:  $u_{H,h}^S \in V_H$  solves

$$(f, \Phi_H)_{L^2(\Omega)} = \mathcal{A}_h^S(u_{H,h}^S, \Phi_H) \quad \forall \Phi_H \in V_H, \quad (7)$$

with

$$\mathcal{A}_h^S(u_{H,h}^S, \Phi_H) := \sum_{T \in \mathcal{T}_H} |T| \int_{Y_{T,\epsilon}} A_h^\epsilon(x, \nabla_x R_T(u_{H,h}^S)(x)) \cdot \nabla_x R_T(\Phi_H)(x) \, dx. \quad (8)$$

Here,  $R_T$  denotes the local reconstruction operator, as it has been defined in (4). For the parameter  $\delta$  we furthermore assume  $\delta \geq \epsilon$  and  $Y_{T,\delta} \subset T$  for all  $T \in \mathcal{T}_H$ . An expedient choice for the periodic case (i.e. the diffusion operator has a periodic structure) could be  $\delta = \epsilon$ , for the non-periodic case  $\delta = m\epsilon, m \in \mathbb{N}_{>1}$ .

**Remark 1.** For every  $\delta \geq \epsilon$ , the TFR-HM method produces a unique solution  $u_{H,h}^S \in V_H$ . This is a direct effect of the monotonicity property of  $A^\epsilon$ . In the case that  $A^\epsilon$  is linear in the second variable with  $(A_i^\epsilon(x, \cdot))_j = (A_j^\epsilon(x, \cdot))_i$  for  $1 \leq i, j \leq d$ , (8) results in a symmetric linear system of equations, which is cheap to solve.

At the expense of these obvious advantages, the method incorporates a small inaccuracy: the reconstruction of the test function (TFR). The 'VMM way' of deriving the multiscale method results in a test function without reconstruction. However, if  $\epsilon = \delta$  this additional contribution  $\int_{Y_{T,\epsilon}} A_h^\epsilon \left( \cdot, \nabla_x R_T(u_{H,h}^S) \right) \cdot \nabla_x (R_T(\Phi_H) - \Phi_H)$  is equal to zero, due to the properties of  $R_T$ . In this situation it is reasonable to replace  $\Phi_H$  by  $R_T(\Phi_H)$ . In the general case with  $\delta \geq \epsilon$ , we at least expect the contribution to remain small. But, assuming that  $A^\epsilon$  has a relatively heterogeneous structure on each representative cell  $Y_{T,\epsilon}$ , summation over these small inaccuracies may end up in a perceptible failure. Beside this, the big disadvantage of the method (in the nonlinear setting) is that it is typically far more expensive than the subsequent HMM without TFR.

With regard to Remark 1 it is reasonable to formulate a second version of the multiscale method, which does not make use of test function reconstruction:

**Definition 2.5** (Heterogeneous Multiscale Method for monotone operators).

We define the HMM approximation  $u_{H,h}$  of  $u^\epsilon$  by:  $u_{H,h} \in V_H$  solves

$$(f, \Phi_H)_{L^2(\Omega)} = \mathcal{A}_h(u_{H,h}, \Phi_H) \quad \forall \Phi_H \in V_H, \quad (9)$$

with

$$\mathcal{A}_h(u_{H,h}, \Phi_H) := \sum_{T \in \mathcal{T}_H} |T| \int_{Y_{T,\epsilon}} A_h^\epsilon(x, \nabla_x R_T(u_{H,h})(x)) \cdot \nabla_x \Phi_H(x) \, dx. \quad (10)$$

Again,  $R_T$  denotes the local reconstruction operator. For  $\delta$  and  $\epsilon$  we make the same assumptions as in Definition 2.4. For details on how to solve (9), we refer to the HMM-Newton Scheme in [22].

It may happen that we are in a setting where it is hard or even impossible to identify a unique parameter  $\epsilon$  in our original problem. Nevertheless, (10) requires a selection. In this situation  $\epsilon$  should be chosen such that the behaviour of  $A^\epsilon$  in  $Y_{T,\epsilon}$  is fairly representative for its behaviour in  $Y_{T,\delta}$ .

Note that the method also makes sense for  $H < \epsilon$ . Then it can be interpreted as a domain decomposition scheme with oversampling.

**Remark 2.** The advantage of the multiscale method in Definition 2.5 is apparent. We expect it to be a little more accurate and far cheaper than the TFR-HMM (since we do not have to solve reconstruction problems for the test functions). On the other hand, the big issue of the alternative method (in comparison to the TFR-HMM) is that we cannot easily guarantee existence and uniqueness of a solution of equation (9), unless we set  $\delta = \epsilon$  (but then we skip oversampling and the methods are identical). However, if we are in a reasonable scenario, TFR-HMM and HMM are close together and we will therefore never have a problem with uniqueness and existence. In the next section, we discuss this problem a little further and in Section 6 we make corresponding numerical experiments.

**Remark 3.** To summarize our claim, we can say that we strongly suggest the usage of the TFR-HMM in the linear setting and the usage of the HMM in the nonlinear setting. Furthermore, if  $u_{H,h}^S$  denotes the TFR-HMM solution of (7) and  $u_{H,h}$  the HMM solution of (9) and if  $\delta = \epsilon$  then we have  $u_{H,h}^S = u_{H,h}$ . Since this work is dedicated to the treatment of nonlinear elliptic problems, we primarily focus on analyzing the HMM without test-function reconstruction.

### 3 The HMM limit problem

In Subsection 3.1, we introduce a  $\delta$ - $\epsilon$ -homogenized problem. Furthermore, we state our first main result, which says that the solution of this problem is the limit of our HMM-approximations. The proof is given in Subsection 3.2.

#### 3.1 Model assumption and limit problem

In this subsection we are concerned with analyzing the HMM stated in Definition 2.5 with regard to an analytical limit problem for  $H$  and  $h$  tending to zero. On the basis of a model assumption we state a condition which guarantees the existence of the HMM approximation  $u_{H,h}$  which was given by (9). Before we are prepared to present our results, we need several definitions. If  $\chi_{\frac{\epsilon}{\delta}Y}$  denotes the characteristic function of  $\frac{\epsilon}{\delta}Y$ , we introduce the following scaled cut-off function:

$$w^{\epsilon,\delta}(y) := \frac{\delta^d}{\epsilon^d} \chi_{\frac{\epsilon}{\delta}Y}. \quad (11)$$

Denoting  $B_\delta(\Omega) := \{x \in \mathbb{R}^d \mid \exists \bar{x} \in \Omega : |x - \bar{x}| < \delta\}$ , we suppose that  $A^\epsilon$  keeps its properties if we replace  $\Omega$  by  $B_\delta(\Omega)$ . Then the *two-scale separated diffusion*  $A^{\epsilon,\delta} : \Omega \times Y \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  is given by

$$A^{\epsilon,\delta}(x, y, \xi) := A^\epsilon(x + \delta y, \xi)$$

and the image  $Q(\xi)$  of correction operator  $Q : \mathbb{R}^d \rightarrow L^2(\Omega, \tilde{H}_\#^1(Y))$  by

$$\int_Y A^{\epsilon,\delta}(x, y, \xi + \nabla_y Q(\xi)(x, y)) \cdot \phi(y) dy = 0 \quad \forall \phi \in \tilde{H}_\#^1(Y). \quad (12)$$

Note that  $Q$  is well defined due to the properties of  $A^{\epsilon,\delta}$ .

**Definition 3.1** (The  $\delta$ - $\epsilon$ -homogenized operator). In the following we call  $A^0 : \Omega \times \mathbb{R}^d \rightarrow \mathbb{R}^d$  the  $\delta$ - $\epsilon$ -homogenized operator, with

$$A^0(x, \xi) := \int_Y w^{\epsilon,\delta}(y) A^{\epsilon,\delta}(x, y, \xi + \nabla_y Q(\xi)(x, y)) dy. \quad (13)$$

The essential model assumption in this section is the strong monotonicity of  $A^0$ . Note that this is always the case in reasonable applications with physical background. In particular if  $\delta = \epsilon$  the monotonicity is inherited from  $A^\epsilon$ .

**Assumption 3.2** (Model assumption 1). We suppose that the  $\delta$ - $\epsilon$ -homogenized operator  $A^0$  is strongly monotone with  $\bar{\alpha} > 0$ , i.e.:

$$(A^0(x, \xi_1) - A^0(x, \xi_2)) \cdot (\xi_1 - \xi_2) \geq \bar{\alpha} |\xi_1 - \xi_2|^2 \quad \forall \xi_1, \xi_2 \in \mathbb{R}^d, \quad \forall x \in \Omega. \quad (14)$$

**Conclusion 3.3.** With the preceding assumptions and defining

$$\mathcal{A}(\Phi, \Psi) := \int_{\Omega} A^0(x, \nabla\Phi(x)) \cdot \nabla\Psi(x) dx, \quad (15)$$

we have a unique solution  $u^c \in \dot{H}^1(\Omega)$  of

$$\mathcal{A}(u^c, \Phi) = (f, \Phi)_{L^2(\Omega)} \quad \forall \Phi \in \dot{H}^1(\Omega). \quad (16)$$

In the following we want to show that  $u_{H,h}$  converges to  $u^c$  under the given assumptions. In the definition of the reconstruction operator  $R_T$  we suppose that  $Y_{T,\delta} \subset T$ . Before we can have a look at a limit process for the HMM, we therefore have to extend  $R_T$  to the case  $x_T + \delta Y \not\subset T$ . Now, let  $\Phi_H \in V_H$ . Since  $\nabla\Phi_H$  is a constant on  $Y_{T,\delta}$  if  $Y_{T,\delta} \subset T$ , there are two possibilities for interpreting the reconstruction  $R_T$  for small  $H$ : 1. we still use Definition 2.3 even though  $\nabla\Phi_H$  is no more a constant on  $T$  or 2. we replace  $\Phi_H$  in (4) by the linear extension of  $(\Phi_H)|_T$  to the whole domain  $\Omega$ . Both is equally possible. However, we focus on the second alternative since it is closer to the ideas of homogenization and scale separation. Therefore, we define the following extended reconstruction operator.

**Definition 3.4** (Extended reconstruction operator).

For  $\delta > 0$ , we define the image  $R_T(\Phi_H)$  under  $R_T : V_H \rightarrow V_H + W_h(Y_{T,\delta})$  by

$$\int_{Y_{T,\delta}} A_h^\epsilon(x, \nabla_x R_T(\Phi_H)(x)) \cdot \nabla_x \phi_h(x) dx = 0 \quad \forall \phi_h \in W_h(Y_{T,\delta}), \quad (17)$$

where  $R_T(\Phi_H) \in E_T(\Phi_H) + W_h(Y_{T,\delta})$  and  $E_T : \mathbb{P}^1(T) \rightarrow \mathbb{P}^1(\Omega)$  is the canonic extension operator.

In the following we use the above definition of  $R_T$  without mentioning and we can state the first main result:

**Theorem 3.5** (HMM limit problem). *Suppose that (model) Assumption 3.2 holds true and that  $h$  is sufficiently small, then there exists a solution  $u_{H,h}$  of the HMM problem (9) for any given triangulation  $\mathcal{T}_H$  of  $\Omega$ . If  $u^c$  denotes the solution of the  $\delta$ - $\epsilon$ -homogenized problem, we also have*

$$\lim_{H \rightarrow 0} \lim_{h \rightarrow 0} \|u_{H,h} - u^c\|_{H^1(\Omega)} = 0.$$

## 3.2 Proof of the convergence result

In this section we are concerned with proving Theorem 3.5. Furthermore, we specify the meaning of *sufficiently small  $h$*  as mentioned in this theorem. For simplification, we postulate two nested families of triangulations  $(\mathcal{T}_{H_i})_{i \in \mathbb{N}}$  of  $\Omega$  and  $(\mathcal{T}_{h_i})_{i \in \mathbb{N}}$  of  $Y$  with  $H_i := \max\{\text{diam}(T) | T \in \mathcal{T}_{H_i}\} \rightarrow 0$  for  $i \rightarrow \infty$  and analogously for  $h_i$ .

On the basis of the extended reconstruction operator, we start with reformulating the reconstruction problems (or cell problems) by splitting  $R_T$  in macroscopic and microscopic contributions:

**Lemma 3.6** (Discrete correction operator). *We define the image  $Q_h(\xi)$  of the discrete correction operator  $Q_h : \mathbb{R}^d \rightarrow L^2(\Omega, W_h(Y))$  by*

$$\int_Y A_{H,h}^{\epsilon,\delta}(x, y, \xi + \nabla_y Q_h(\xi)(x, y)) \cdot \nabla_y \phi_h(y) dy = 0, \quad \forall \phi_h \in W_h(Y). \quad (18)$$

Here,  $A_{H,h}^{\epsilon,\delta}$  is a discretization of  $A^{\epsilon,\delta}$  given by:

$$A_{H,h}^{\epsilon,\delta}(x, y, \cdot)|_{T \times K} := A^{\epsilon,\delta}(x_T, y_K, \cdot) \text{ for } T \times K \in \mathcal{T}_H \times \mathcal{T}_h. \quad (19)$$

With the definitions above, the subsequent equation holds true:

$$Q_h(\nabla \Phi_H(x_T))(x, y)|_{T \times Y} = \frac{1}{\delta} (R_T(\Phi_H) - E_T(\Phi_H)) \circ x_T^\delta(y) \text{ for } \Phi_H \in V_H.$$

In particular, we get for all  $\Psi_H, \Phi_H \in V_H$ :

$$\mathcal{A}_h(\Psi_H, \Phi_H) = \int_{\Omega} \int_Y w^{\epsilon,\delta}(y) A_{H,h}^{\epsilon,\delta}(x, y, \nabla_x \Psi_H(x) + \nabla_y Q_h(\nabla_x \Psi_H(x))(x, y)) \cdot \nabla_x \Phi_H(x) dy dx \quad (20)$$

where  $\mathcal{A}_h$  is given by (10). In the same way we also have

$$\begin{aligned} \mathcal{A}_h^S(\Psi_H, \Phi_H) &= \int_{\Omega} \int_Y w^{\epsilon,\delta}(y) A_{H,h}^{\epsilon,\delta}(x, y, \nabla_x \Psi_H(x) + \nabla_y Q_h(\nabla_x \Psi_H(x))(x, y)) \\ &\quad \cdot (\nabla_x \Phi_H(x) + \nabla_y Q_h(\nabla_x \Phi_H(x))(x, y)) dy dx \end{aligned} \quad (21)$$

for  $\mathcal{A}_h^S$  given by (8).

*Proof.* For  $\Phi_H \in V_H$  we define

$$\mathcal{Q}_h(\Phi_H)(x, y)|_{T \times Y} := \frac{1}{\delta} (R_T(\Phi_H) - E_T(\Phi_H)) \circ x_T^\delta(y).$$

We verify  $Q_h(\nabla \Phi_H(\cdot)) = \mathcal{Q}_h(\Phi_H)$  to prove the results. First we note that the following holds for every  $x \in T$ :

$$\begin{aligned} \nabla_y \mathcal{Q}_h(\Phi_H)(x, y) &= \frac{1}{\delta} (\delta \nabla_x R_T(\Phi_H) - \delta \nabla_x E(\Phi_H)) \circ x_T^\delta(y) \\ &= (\nabla_x R_T(\Phi_H)) \circ x_T^\delta(y) - \nabla_x \Phi_H(x_T). \end{aligned} \quad (22)$$

By (17) we therefore obtain for all  $\tilde{\phi}_h \in W_h(Y_{T,\delta})$ :

$$\begin{aligned} 0 &= \int_{Y_{T,\delta}} A_h^\epsilon(x, \nabla_x R_T(\Phi_H)(x)) \cdot \nabla_x \tilde{\phi}_h(x) dx \\ &= \int_Y A_h^\epsilon(x_T^\delta(y), \nabla_x R_T(\Phi_H) \circ x_T^\delta(y)) \cdot \nabla_x \tilde{\phi}_h(x_T^\delta(y)) dy \\ &= \int_Y A_{H,h}^{\epsilon,\delta}(x, y, \nabla_x \Phi_H(x_T) + \nabla_y \mathcal{Q}_h(\Phi_H)(x, y)) \cdot \nabla_x \tilde{\phi}_h(x_T^\delta(y)) dy. \end{aligned}$$

Defining  $\phi_h(x, \cdot) \in W_h(Y)$  by  $\phi_h(x, y) := \frac{1}{\delta} \tilde{\phi}_h(x, x_T^\delta(y))$  yields  $Q_h(\nabla \Phi_H(\cdot)) = \mathcal{Q}_h(\Phi_H)$ . To

derive the equation for  $\mathcal{A}_h^S$  we can use (22) to proceed similarly:

$$\begin{aligned}
& \int_{Y_{T,\epsilon}} A_h^\epsilon(x, \nabla_x R_T(\Psi_H)(x)) \cdot \nabla_x R_T(\Phi_H)(x) dx \\
&= \frac{1}{\epsilon^d} \int_{Y_{T,\delta}} \chi_{Y_{T,\epsilon}}(x) A_h^\epsilon(x, \nabla_x R_T(\Psi_H)(x)) \cdot \nabla_x R_T(\Phi_H)(x) dx \\
&= \frac{\delta^d}{\epsilon^d} \int_Y \chi_{Y_{T,\epsilon}}(x_T^\delta(y)) A_h^\epsilon(x_T^\delta(y), \nabla_x \Psi_H(x_T) + \nabla_y \mathcal{Q}_h(\Psi_H)(x_T, y)) \\
&\quad \cdot (\nabla_x \Phi_H(x_T) + \nabla_y \mathcal{Q}_h(\Phi_H)(x_T, y)) dy \\
&= \frac{\delta^d}{\epsilon^d} \int_Y \chi_{\frac{\epsilon}{\delta} Y}(y) A_{H,h}^{\epsilon,\delta}(x_T, y, \nabla_x \Psi_H(x_T) + \nabla_y \mathcal{Q}_h(\Psi_H)(x_T, y)) \\
&\quad \cdot (\nabla_x \Phi_H(x_T) + \nabla_y \mathcal{Q}_h(\Phi_H)(x_T, y)) dy
\end{aligned}$$

Since the quadrature formula is exact for piecewise constant functions (on  $T \in \mathcal{T}_H$ ), we get the result by summation:

$$\begin{aligned}
\left(\frac{\epsilon}{\delta}\right)^d \mathcal{A}_h^S(\Psi_H, \Phi_H) &= \int_\Omega \int_Y \chi_{\frac{\epsilon}{\delta} Y}(y) A_{H,h}^{\epsilon,\delta}(x, y, \nabla_x \Psi_H(x) + \nabla_y \mathcal{Q}_h(\Psi_H)(x, y)) \\
&\quad \cdot (\nabla_x \Phi_H(x) + \nabla_y \mathcal{Q}_h(\Phi_H)(x, y)) dy dx.
\end{aligned}$$

For  $\mathcal{A}_h$  we proceed in analogy. □

For the sake of shortening, we introduce  $\mathcal{Q}(\Phi)(x, y) := Q(\nabla_x \Phi(x))(x, y)$  for  $\Phi \in \dot{H}^1(\Omega)$  and  $\mathcal{Q}_h(\Phi_H)(x, y) := Q_h(\nabla_x \Phi_H(x))(x, y)$  for  $\Phi_H \in V_H$ . From now on we primarily use  $\mathcal{Q}$  instead of  $Q$ . For simplification and since it yields no further difficulties, we also replace  $A_h^{\epsilon,\delta}$  by  $A^{\epsilon,\delta}$  in our HMM formulation in Definition 2.5. The next lemma is an alternative formulation of the  $\delta$ - $\epsilon$ -homogenized equation.

**Lemma 3.7** (Generalized two-scale equation). *Let Assumption 3.2 be fulfilled. If  $u^c$  denotes the solution of (16) and if we define  $u^f := \mathcal{Q}(u^c)$ , then the tuple  $(u^c, u^f) \in \dot{H}^1(\Omega) \times L^2(\Omega, \tilde{H}_\#^1(Y))$  is the unique solution of the generalized two-scale equation:*

$$\begin{aligned}
& \int_\Omega \int_Y A^{\epsilon,\delta}(x, y, \nabla_x u^c(x) + \nabla_y u^f(x, y)) (w^{\epsilon,\delta}(y) \nabla_x \Phi(x) + \nabla_y \phi(x, y)) dy dx \\
&= \int_\Omega f(x) \Phi(x) dx \quad \forall (\Phi, \phi) \in \dot{H}^1(\Omega) \times L^2(\Omega, \tilde{H}_\#^1(Y)). \tag{23}
\end{aligned}$$

The result is obvious and there is nothing to prove. Now, we require two further lemmata to prove the main result. The first one describes the continuity of the operators  $Q$  and  $Q_h$ :

**Lemma 3.8.** *The following inequality holds almost everywhere in  $x$  and for all  $\xi_1, \xi_2 \in \mathbb{R}^d$ :*

$$|Q(\xi_1)(x, \cdot) - Q(\xi_2)(x, \cdot)|_{H^1(Y)} \leq \left(\frac{\beta^2}{\alpha^2} - 1\right)^{\frac{1}{2}} |\xi_1 - \xi_2| \tag{24}$$

*The same result holds true for  $Q_h$ ,  $h \in \{h_i | i \in \mathbb{N}\}$ .*

*Proof.*

$$\begin{aligned}
& \alpha |\xi_1 - \xi_2 + \nabla_y (Q(\xi_1) - Q(\xi_2))(x, \cdot)|_{L^2(Y)}^2 \\
& \leq \int_Y A^{\epsilon, \delta}(x, y, \xi_1 + \nabla_y Q(\xi_1)(x, y)) \cdot ((\xi_1 - \xi_2) + \nabla_y(Q(\xi_1) - Q(\xi_2))(x, y)) \, dy \\
& \quad - \int_Y A^{\epsilon, \delta}(x, y, \xi_2 + \nabla_y Q(\xi_2)(x, y)) \cdot ((\xi_1 - \xi_2) + \nabla_y(Q(\xi_1) - Q(\xi_2))(x, y)) \, dy \\
& = \int_Y A^{\epsilon, \delta}(x, y, \xi_1 + \nabla_y Q(\xi_1)(x, y)) \cdot (\xi_1 - \xi_2) \, dy \\
& \quad - \int_Y A^{\epsilon, \delta}(x, y, \xi_2 + \nabla_y Q(\xi_2)(x, y)) \cdot (\xi_1 - \xi_2) \, dy \\
& \leq \beta |\nabla_x (\xi_1 - \xi_2) + \nabla_y (Q(\xi_1) - Q(\xi_2))(x, \cdot)|_{L^2(Y)} |\xi_1 - \xi_2|
\end{aligned}$$

This yields:

$$\begin{aligned}
& |\xi_1 - \xi_2|^2 + |Q(\xi_1)(x, \cdot) - Q(\xi_2)(x, \cdot)|_{H^1(Y)}^2 \\
& = |\xi_1 - \xi_2 + \nabla_y (Q(\xi_1) - Q(\xi_2))(x, \cdot)|_{L^2(Y)}^2 \\
& \leq \frac{\beta^2}{\alpha^2} |\xi_1 - \xi_2|^2
\end{aligned}$$

□

Now, we define  $X_z$  which is a nonlinear subspace of  $L^2(\Omega, \tilde{H}_\#^1(Y))$  required for the main proof.

**Definition 3.9.** For

$$K(z) := \bar{\alpha}^{-1} \left( \frac{\beta^2}{\alpha} z + \|f\|_{L^2(\Omega)} \right)$$

we define the nonlinear space  $X_z$  by:

$$X_z := \{\phi \in L^2(\Omega, \tilde{H}_\#^1(Y)) \mid \phi(x, \cdot) = Q(\nabla \Phi_H(x))(x, \cdot), \Phi_H \in V_H, \|\Phi_H\|_{H^1(\Omega)} = K(z)\}$$

The next lemma helps us to give a criterion on the size of  $h$  so that we can guarantee a HMM solution.

**Lemma 3.10.** *For all  $z > 0$ , there exists  $i_0 \in \mathbb{N}$ , such that:*

$$\forall i \geq i_0, \forall v \in X_z \exists \phi_{h_i} \in L^2(\Omega, W_{h_i}(Y)) : |v - \phi_{h_i}|_{L^2(\Omega, H^1(Y))} \leq z. \quad (25)$$

*Proof.* First we note that we do not need the specific structure of  $K(z)$  for this proof.  $K(z)$  could be replaced by an arbitrary positive constant.

Since we have a nested family of triangulations  $\mathcal{T}_{h_i}$ , the negation of (25) reads

$$\forall i \in \mathbb{N} \exists v_i \in X_z : \forall \phi_{h_i} \in L^2(\Omega, W_{h_i}(Y)) : |v_i - \phi_{h_i}|_{L^2(\Omega, H^1(Y))} > z. \quad (26)$$

Since  $v_i \in X_z$  there exist  $\Phi_H^i \in V_H$  with  $\mathcal{Q}(\Phi_H^i) = v_i$  and  $\|\Phi_H^i\|_{H^1(\Omega)} = K(z)$ . So  $(\Phi_H^i)_i$  is a bounded sequence in a finite dimensional Hilbert space. We therefore have a strong  $H^1$  limit of  $\Phi_H^i$  (up to a subsequence). It is denoted by  $\Phi_H$ . On the other hand,  $\mathcal{Q}$  is a continuous operator due to (24), which implies  $v_i = \mathcal{Q}(\Phi_H^i) \rightarrow \mathcal{Q}(\Phi_H)$  strongly in  $L^2(\Omega, H^1(Y))$ .

Since  $\cup_{i \in \mathbb{N}} L^2(\Omega, W_{h_i}(Y))$  is dense in  $L^2(\Omega, H^1(Y))$ , we can find  $i_0 \in \mathbb{N}$  and a sequence  $\phi_{h_i} \in L^2(\Omega, W_{h_i}(Y))$  with  $|\mathcal{Q}(\Phi_H) - \phi_{h_i}|_{L^2(\Omega, H^1(Y))} \leq \frac{z}{2}$  for all  $i \geq i_0$ . All together with sufficiently large  $i_0$ :

$$|v_{i_0} - \phi_{h_{i_0}}|_{L^2(\Omega, H^1(Y))} \leq |v_{i_0} - \mathcal{Q}(\Phi_H)|_{L^2(\Omega, H^1(Y))} + |\mathcal{Q}(\Phi_H) - \phi_{h_{i_0}}|_{L^2(\Omega, H^1(Y))} \leq \frac{z}{2} + \frac{z}{2} = z.$$

This is a contradiction to (26).  $\square$

**Remark 4.** Analogously to the preceding proof we can alternatively show that for all  $C > 0$  and for all  $\epsilon_0 > 0$  there exists  $i_0 \in \mathbb{N}$  such that for all  $i \geq i_0$  and for all

$$v \in \{\phi \in L^2(\Omega, \tilde{H}_\#^1(Y)) \mid \phi(x, \cdot) = \mathcal{Q}(\nabla \Phi_H(x))(x, \cdot), \Phi_H \in V_H, \|\Phi_H\|_{H^1(\Omega)} \leq C\}$$

there exists  $\phi_{h_i} \in L^2(\Omega, W_{h_i}(Y))$  such that  $|v - \phi_{h_i}|_{L^2(\Omega, H^1(Y))} \leq \epsilon_0$ .

Now, we are prepared to prove the convergence result for the HMM. By means of the space  $X_z$ , we can detail Theorem 3.5:

**Theorem 3.11.** *Suppose that Assumption 3.2 holds true. For given  $z > 0$ , we can choose  $i_0$  such that for all  $i \geq i_0$ :*

$$\sup_{v \in X_z} \left( \inf_{\phi_{h_i} \in L^2(\Omega, W_{h_i}(Y))} (|v - \phi_{h_i}|_{L^2(\Omega, H^1(Y))}) \right) \leq z. \quad (27)$$

With this condition, there exists a HMM approximation  $u_{H_j, h_i}$  for all  $j \in \mathbb{N}$  and for all  $i \geq i_0$ . Moreover, we have:

$$\lim_{j \rightarrow \infty} \lim_{i \rightarrow \infty} \|u_{H_j, h_i} - u^c\|_{H^1(\Omega)} = 0.$$

*Proof.* (27) is a direct consequence of Lemma 3.10. In the first part of this proof, we fix  $H \in \{H_j \mid j \in \mathbb{N}\}$  and  $h \in \{h_i \mid i \in \mathbb{N}\}$ . We define  $N := \dim(V_H)$ . Let  $\{\Phi_H^{(1)}, \dots, \Phi_H^{(N)}\}$  denote the Lagrange base of  $V_H$ , i.e.  $\Phi_H^{(k)} \in V_H$  with  $\Phi_H^{(k)}(x_s) = \delta_{ks}$ , where  $x_s$  denotes the  $s$ 'th inner node of the triangulation  $\mathcal{T}_H$ . With this, we introduce a norm on  $\mathbb{R}^N$  by:

$$|\xi|_{V_H} := \left\| \sum_{s=1}^N \xi_s \Phi_H^{(s)} \right\|_{H^1(\Omega)}.$$

Now, we define

$$\begin{aligned} g_h^l(\xi) &:= \int_{\Omega} \int_Y w^{\epsilon, \delta}(y) A^{\epsilon, \delta} \left( x, y, \sum_{s=1}^N \xi_s \nabla_x \Phi_H^{(s)}(x) + \nabla_y \mathcal{Q}_h \left( \sum_{s=1}^N \xi_s \Phi_H^{(s)} \right) (x, y) \right) \cdot \nabla_x \Phi_H^{(l)}(x) dy dx \\ &\quad - \int_{\Omega} f(x) \Phi_H^{(l)}(x) dx. \end{aligned}$$

We want to show: there exists  $\bar{\xi} \in \mathbb{R}^N$ , such that  $g_h^l(\bar{\xi}) = 0$  for all  $1 \leq l \leq N$ .  $g_h^l$  is continuous due to the continuity of  $\mathcal{Q}_h$ . Using  $\Phi_H := \sum_{s=1}^N \xi_s \Phi_H^{(s)}$ , we can define

$$\begin{aligned} G_h(\xi) &:= \sum_{l=1}^N g_h^l(\xi) \xi_l = \int_{\Omega} \int_Y w^{\epsilon, \delta}(y) A^{\epsilon, \delta} \left( x, y, \nabla_x \Phi_H(x) + \nabla_y \mathcal{Q}_h(\Phi_H)(x, y) \right) \cdot \nabla_x \Phi_H(x) dy dx \\ &\quad - \int_{\Omega} f(x) \Phi_H(x) dx \end{aligned}$$

to get

$$\begin{aligned} G_h(\xi) &= (\mathcal{A}_h - \mathcal{A})(\Phi_H, \Phi_H) + \mathcal{A}(\Phi_H, \Phi_H) - (f, \Phi_H)_{L^2(\Omega)} \\ &\geq (\mathcal{A}_h - \mathcal{A})(\Phi_H, \Phi_H) + \bar{\alpha}|\xi|_{V_H}^2 - \|f\|_{L^2(\Omega)}|\xi|_{V_H}. \end{aligned} \quad (28)$$

For  $(\mathcal{A}_h - \mathcal{A})(\Phi_H, \Phi_H)$  we use the Cea-Lemma for monotone operators which reads

$$|\mathcal{Q}(\Phi_H)(x, \cdot) - \mathcal{Q}_h(\Phi_H)(x, \cdot)|_{H^1(Y)} \leq \frac{\beta}{\alpha} \inf_{\phi_h \in W_h(Y)} |\mathcal{Q}(\Phi_H)(x, \cdot) - \phi_h|_{H^1(Y)}, \quad (29)$$

to obtain:

$$\begin{aligned} &(\mathcal{A}_h - \mathcal{A})(\Phi_H, \Phi_H) \\ &= \int_{\Omega} \int_Y w^{\epsilon, \delta}(y) (A^{\epsilon, \delta}(x, y, \nabla_x \Phi_H(x) + \nabla_y \mathcal{Q}_h(\Phi_H)(x, y)) \\ &\quad - A^{\epsilon, \delta}(x, y, \nabla_x \Phi_H(x) + \nabla_y \mathcal{Q}(\Phi_H)(x, y))) \cdot \nabla_x \Phi_H(x) dy dx \\ &\leq \beta |\xi|_{V_H} \frac{\beta}{\alpha} \inf_{\phi_h \in L^2(\Omega, W_h(Y))} |\mathcal{Q}(\Phi_H) - \phi_h|_{L^2(\Omega, H^1(Y))}. \end{aligned}$$

If we restrict ourselves to  $\xi$  with  $|\xi|_{V_H} = K(z)$ , we can use Lemma 3.10 to obtain that there exists  $i_0 \in \mathbb{N}$  ( $i_0$  independent of  $\xi$ ) so that for all  $i \geq i_0$ :

$$(\mathcal{A}_{h_i} - \mathcal{A})(\Phi_H, \Phi_H) \leq \frac{\beta^2}{\alpha} z |\xi|_{V_H} = \frac{\beta^2}{\alpha} z K(z).$$

Combining this with (28) we get for all  $i \geq i_0$  and for all  $\xi \in \mathbb{R}^N$  with  $|\xi|_{V_H} = K(z)$ :

$$G_{h_i}(\xi) \geq \bar{\alpha} K(z)^2 - \frac{\beta^2}{\alpha} z K(z) - \|f\|_{L^2(\Omega)} K(z) = 0.$$

With  $G_{h_i}(\xi) \geq 0$ , we can use a simple conclusion from the Brouwer fixed point theorem to obtain, that we have a solution  $\bar{\xi}^i \in \mathbb{R}^N$  of the problem  $g_{h_i}^l(\bar{\xi}^i) = 0$  for all  $1 \leq l \leq N$  and with  $|\bar{\xi}^i|_{V_H} \leq K(z)$  (see Chapter 1, Lemma 2.26 in [43] for comparison). Our HMM approximation is therefore given by  $u_{H, h_i} = \sum_{s=1}^N \bar{\xi}_s^i \Phi_H^{(s)}$ .

It remains to identify the limit of the HMM approximations. First, we have that  $(u_{H, h_i})_{i \geq i_0}$  is a bounded sequence in  $V_H$  (bounded by  $(K(z))$ ). Due to the finite dimension of  $V_H$ , we have that there exists  $u_H \in V_H$  so that  $u_{H, h_i} \xrightarrow{i \rightarrow \infty} u_H$  strongly in  $H^1(\Omega)$  (convergence of subsequences can be replaced by convergence of the whole sequence due to the characterization of the limit problem which always yields a unique solution). With (24), (29) and Remark 4 (with  $C = K(z)$ ) we furthermore obtain:

$$\begin{aligned} &|\mathcal{Q}_{h_i}(u_{H, h_i}) - \mathcal{Q}(u_H)|_{L^2(\Omega, H^1(Y))} \\ &\leq \frac{\beta}{\alpha} \inf_{\phi_{h_i} \in L^2(\Omega, W_{h_i}(Y))} |\phi_{h_i} - \mathcal{Q}(u_{H, h_i})|_{L^2(\Omega, H^1(Y))} + \left(\frac{\beta^2}{\alpha^2} - 1\right)^{\frac{1}{2}} |u_{H, h_i} - u_H|_{H^1(\Omega)} \rightarrow 0. \end{aligned}$$

Due to the strong convergence of  $u_{H, h_i}$  to  $u_H$  and the strong convergence of  $\mathcal{Q}_{h_i}(u_{H, h_i})$  to  $\mathcal{Q}(u_H)$  we obtain that  $u_H$  solves

$$\int_{\Omega} A^0(x, \nabla_x u_H(x)) \cdot \nabla_x \Phi_H(x) dx = \int_{\Omega} f(x) \Phi_H(x) dx$$

for all  $\Phi_H \in V_H$ . Since  $A^0$  is strongly monotone, we can use the Cea-Lemma to get strong convergence of  $u_H$  in  $H^1(\Omega)$  to  $u^c$  (we have uniqueness of the limit, due to uniqueness in the limit problem). This ends the proof.  $\square$

### 3.3 Justification in the periodic setting

In the preceding subsections we showed that the HMM is capable of approximating  $u^c$  (given by (16)) up to a desired accuracy. However, we did not yet comment on the relation between  $u^c$  and the exact solution  $u^\epsilon$ . For a better understanding, we exemplify the justification of equation (23) in the periodic setting. General comments shall be discussed later.

Let us assume, that  $\delta = \epsilon$  and that  $A^\epsilon(x, \xi) = A(\frac{x}{\epsilon}, \xi)$ , where  $A(\cdot, \xi)$  is a  $Y$ -periodic function. In this situation, it is well known, that the solution  $(u^0, u^1) \in \dot{H}^1(\Omega) \times L^2(\Omega, \tilde{H}_\#^1(Y))$  of the so called two-scale homogenized equation fulfills:

$$\int_{\Omega} \int_Y A(y, \nabla_x u^0(x) + \nabla_y u^1(x, y)) (\nabla_x \Phi(x) + \nabla_y \phi(x, y)) dy dx = \int_{\Omega} f(x) \Phi(x) dx \quad (30)$$

for all  $(\Phi, \phi) \in \dot{H}^1(\Omega) \times L^2(\Omega, \tilde{H}_\#^1(Y))$  and moreover

$$\|u^\epsilon - \left(u^0 + u^1(\cdot, \frac{\cdot}{\epsilon})\right)\|_{H^1(\Omega)} \rightarrow 0 \text{ for } \epsilon \rightarrow 0.$$

See for instance [8], [46] and [35] for details. Now, let us define

$$\bar{u}^c(x) := u^0(x) \text{ and } \bar{u}^f(x, y) := u^1(x, y + \frac{x}{\epsilon}),$$

then we immediately have:

$$\begin{aligned} \int_{\Omega} f(x) \Phi(x) dx &= \int_{\Omega} \int_Y A(y, \nabla_x u^0(x) + \nabla_y u^1(x, y)) \cdot \left(\nabla_x \Phi(x) + \nabla_y \phi(x, y - \frac{x}{\epsilon})\right) dy dx \\ &= \int_{\Omega} \int_Y A\left(y + \frac{x}{\epsilon}, \nabla_x u^0(x) + \nabla_y u^1(x, y + \frac{x}{\epsilon})\right) \cdot (\nabla_x \Phi(x) + \nabla_y \phi(x, y)) dy dx \\ &= \int_{\Omega} \int_Y A^\epsilon\left(x + \epsilon y, \nabla_x \bar{u}^c(x) + \nabla_y \bar{u}^f(x, y)\right) \cdot (\nabla_x \Phi(x) + \nabla_y \phi(x, y)) dy dx. \end{aligned}$$

Due to uniqueness of the solution of problem (23), we obtain  $\bar{u}^c = u^c$  and  $\bar{u}^f = u^f$ . This yields the relation  $\nabla_x u^f(x, 0) = \nabla_x (u^1(x, \frac{x}{\epsilon}))$  and the estimate

$$\|u^\epsilon - (u^c + \epsilon u^f(\cdot, 0))\|_{H^1(\Omega)} = \|u^\epsilon - \left(u^0 + \epsilon u^f(\cdot, \frac{\cdot}{\epsilon})\right)\|_{H^1(\Omega)} \rightarrow 0.$$

In particular, we have by standard homogenization theory and by using the results from the preceding sections:

$$\lim_{H \rightarrow 0} \lim_{h \rightarrow 0} \|u^\epsilon - u_{H,h}\|_{L^2(\Omega)} = \|u^\epsilon - u^c\|_{L^2(\Omega)} = \|u^\epsilon - u^0\|_{L^2(\Omega)} = O(\epsilon).$$

This implies that the HMM is justified in the periodic setting.

In the non-periodic setting we may also assume that  $u^\epsilon(x) \approx u^0(x) + \epsilon u^1(x, \frac{x}{\epsilon})$ , where  $u^0$  describes the coarse-scale part and  $u^1$  the fine-scale part. Using this ansatz, inserting it in (2) and applying local averaging over representative cells  $x + \epsilon Y$  yields a formulation, which is very similar to (23). The fact that we cannot assume that  $u^1$  is periodic in the second variable requires an oversampling technique to erase the wrong (periodic) boundary condition for the micro-scale contribution. This is why the cut-off function  $w^{\epsilon, \delta}$  (see (11)) is an essential part of the generalized two-scale equation above.

Alternatively, we see by the following reformulation

$$\begin{aligned}
& \int_{\Omega} f(x) \Phi(x) dx \\
&= \int_{\Omega} \int_Y A^{\epsilon, \delta}(x, y, \nabla_x u^c(x) + \nabla_y u^f(x, y)) \cdot (w^{\epsilon, \delta}(y) \nabla_x \Phi(x)) dy dx \\
&= \int_{\Omega} \frac{\delta^d}{\epsilon^d} \int_{\frac{\epsilon}{\delta} Y} A^{\epsilon}(x + \delta y, \nabla_x u^c(x) + \nabla_y u^f(x, y)) \cdot \nabla_x \Phi(x) dy dx \\
&= \int_{\Omega} \int_{x+\epsilon Y} A^{\epsilon}\left(y, \nabla_x u^c(x) + \nabla_y \left(\delta u^f\left(x, \frac{y-x}{\delta}\right)\right)\right) \cdot \nabla_x \Phi(x) dy dx,
\end{aligned}$$

that we expect  $u^c + \delta u^f(\cdot, 0)$  to approximate  $u^{\epsilon}$  in  $H^1(\Omega)$ .

With these considerations, we should demand:

$$\|u^{\epsilon} - (u^c + \delta u^f(\cdot, 0))\|_{H^1(\Omega)} = O(\delta^{\frac{1}{2}}), \quad (31)$$

so that the usage of the HMM is justified in a certain scenario. Here,  $(u^c, u^f) \in \mathring{H}^1(\Omega) \times L^2(\Omega, \tilde{H}_{\sharp}^1(Y))$  is the solution of equation (23). To keep the framework as simple as possible, we weaken this condition and introduce a second model assumption. This assumption simply stands for *the HMM is applicable to our problem*:

**Assumption 3.12** (Model assumption 2). We assume that problem (16) has a unique solution denoted by  $u^{\epsilon}$ . Then we presume that  $u^c$  approximates  $u^{\epsilon}$  in  $L^2(\Omega)$  up to  $\epsilon$ -accuracy:

$$\|u^{\epsilon} - u^c\|_{L^2(\Omega)} = O(\epsilon). \quad (32)$$

Note that we might as well say that the approximation of  $u^{\epsilon}$  is up to  $\delta$ -accuracy since  $\delta$  is an integral multiple of  $\epsilon$ .

Note that (32) is a reasonable definition for *applicability of the HMM*. If this error remains large in a certain scenario, then the HMM produces wrong approximations. In several applications, the situation is very comfortable. For instance, if we are dealing with periodic or homogeneous stochastic structures, the useability of the HMM (i.e. (32)) can be proved a-priori. In this spirit, Assumption 3.12 is a condition which summarizes all these possible settings. There are several possibilities to check if it is fulfilled. In the next subsection we comment on general frameworks.

### 3.4 General considerations on the applicability of the HMM

It is obvious that the question about the relation between  $u^c$  and  $u^{\epsilon}$  is directly connected to the question about the applicability of the HMM. If (32) is somehow fulfilled, the method can be used. If it is not fulfilled, the HMM is inappropriate. There is a large variety of scenarios in which the HMM is used, without even having a clear guarantee that it is applicable. Beside restrictions on a periodic or stochastic setting, there are no analytical results answering this question, but there are several numerical experiments which demonstrate a very general usability of the multiscale method in case of scale separation. Therefore, it seems to be sufficient to presume applicability of the HMM and to analyze the method with regard to the limit problem. However, for the sake of completeness we want to present some ideas which help to answer the question about applicability. Carrying out these ideas in detail is a long work on its own, which is why we keep this section short.

In the following we refer to  $\|u^\epsilon - u^c\|_{L^2(\Omega)}$  as the modeling error. A powerful approach for verifying Assumption 3.12 is presented in the works of Oden *et al.* [47, 40, 41, 45] for the linear setting. Here, the sole assumption is the availability of some kind of homogenized equation, independently of how it has been derived. A-posteriori error estimates are given to evaluate the size of the modeling error, indicating the quality of the homogenized problem. The error can be evaluated in the  $L^2$ -norm, in the energy-norm or in so called 'quantities of interest' which regard the physical background. Since we derived a homogenized equation (16) which correlates with the HMM, the whole theory is applicable to our issue once it is transferred to the nonlinear setting. Furthermore, it can be even used to improve the HMM using the *Goal-Oriented Adaptive Local Solution Algorithm* (see [41]) which allows local enhancements of the HMM approximation in those regions where it is not sufficiently accurate. A combination of GOALS and HMM yields good approximations even if Assumption 3.12 is not fulfilled.

Another approach is presented in the following. The access to the nonlinear setting is straight forward and it emphasizes an interesting feature of the solution:

Let us assume that we have  $A^\epsilon \in (H^{1,\infty}(B_\delta(\Omega) \times \mathbb{R}^d))^d$  and therefore  $H^2$ -regularity for  $u^\epsilon$ . The goal is to achieve a computable indicator, which gives us a strong hint on whether we can apply the multiscale method or not. In particular, the indicator ought to signal if Assumption 3.12 can be fulfilled.

Let  $u^{coarse} + \delta u^{fine}$  denote an approximation of  $u^\epsilon$  which consists of a coarse- and a fine-scale contribution. If we want to estimate the error  $\|u^\epsilon - u^{coarse}\|_{L^2(\Omega)}$ , we might use standard a-posteriori theory to obtain the following estimate:

$$\begin{aligned} \|u^\epsilon - u^{coarse}\|_{L^2(\Omega)} &\leq \frac{C_P}{\alpha} \|f + \operatorname{div} A^\epsilon(\cdot, \nabla (u^{coarse} + \delta u^{fine}))\|_{L^2(\Omega)} \\ &\quad + \delta C_{Tr} C_R d \|A^\epsilon\|_{H^{1,\infty}(\Omega)} \|u^{fine}\|_{L^2(\partial\Omega)} + \delta \|u^{fine}\|_{L^2(\Omega)}. \end{aligned} \quad (33)$$

Here,  $C_P$  denotes the Poincaré constant,  $C_R$  a regularity constant and  $C_{Tr}$  the constant in the trace theorem. The first term on the right hand side of (33) is to check the fulfillment of the differential equation ( $\|f + \operatorname{div} A^\epsilon(\cdot, \nabla (u^{coarse} + \delta u^{fine}))\|_{L^2(\Omega)}$  small), the second term is to check the fulfillment of the boundary condition up to  $\delta$ -accuracy ( $\delta \|A^\epsilon\|_{H^{1,\infty}(\Omega)} \|u^{fine}\|_{L^2(\partial\Omega)}$  small) and the last term is to check, whether we have a real scale separation, i.e.  $u^{coarse}$  is only the coarse scale contribution and a good  $L^2$ -approximation of  $u^\epsilon$  (size of  $\delta \|u^{fine}\|_{L^2(\Omega)}$  remains negligible).

Unfortunately we cannot use such a result in real applications, since  $u^{coarse}$  and in particular  $u^{fine}$  are typically not explicitly available. If we want to compute both parts accurately enough (for instance by solving a two-scale equation such as (23)), we are in a highly complex situation. We must resolve the micro-structure in the whole computational domain to use this (extremely expensive) result as a very accurate reference for  $(u^{coarse}, u^{fine})$ . But then, we do not need the multiscale method anymore and any further consideration is redundant.

A second problem becomes apparent, if we consider the periodic setting again (i.e.  $A^\epsilon(x) = A(x, \frac{x}{\epsilon})$ ). Let us presume sufficient regularity. If  $(u^0, u^1)$  denotes the solution of the two-scale homogenized equation (30), then we have

$$\|u^\epsilon - \left(u^0 + \epsilon u^1\left(\cdot, \frac{\cdot}{\epsilon}\right)\right)\|_{H^1(\Omega)} \rightarrow 0$$

but we typically do *not* have

$$\|f + \operatorname{div} A^\epsilon(\cdot, \nabla \left(u^0 + \epsilon u^1\left(\cdot, \frac{\cdot}{\epsilon}\right)\right))\|_{L^2(\Omega)} \rightarrow 0.$$

The right hand side in (33) does not become small if we only use  $u^{fine} = u^1(\cdot, \frac{\cdot}{\epsilon})$ . In particular a second corrector  $u^2 \in L^2(\Omega, \tilde{H}_\#^1(Y))$  is required to get the desired property, i.e.:

$$\|f + \operatorname{div} A^\epsilon(\cdot, \nabla(u^0 + \epsilon u^1(\cdot, \frac{\cdot}{\epsilon}) + \epsilon^2 u^2(\cdot, \frac{\cdot}{\epsilon}))\|_{L^2(\Omega)} \rightarrow 0.$$

Here,  $u^2$  is the unique solution of

$$\begin{aligned} & \int_Y A(x, y) \nabla_y u^2(x, y) \cdot \nabla_y \psi(y) dy \\ &= \int_Y \operatorname{div}_x (A(x, y) (\nabla_x u^0(x) + \nabla_y u^1(x, y))) \psi(y) dy - \int_Y A(x, y) \nabla_y u^1(x, y) \cdot \nabla_y \psi(y) dy \end{aligned} \quad (34)$$

for all  $\psi \in \tilde{H}_\#^1(Y)$ . See for instance ([13], chapter 7, 128-137) for comparison. In this spirit,  $u^{fine}$  should be equal to  $u^1(\cdot, \frac{\cdot}{\epsilon}) + \epsilon u^2(\cdot, \frac{\cdot}{\epsilon})$  in (33).

If we want to derive an indicator to check for Assumption 3.12, all the considerations above must be carried over to the discrete setting. First of all we need to restrict our fine-scale evaluation to a number of representative cells (just like in the HMM setting). There, we obtain localized a-posteriori error estimates for every cell. These estimates are similar to (33) but they contain a non-computable boundary contribution which can be ignored in an indicator. In order to check the fulfillment of the differential equation, we need to perform additional computations that account for the missing part in our corrector function. With other words: we must solve further cell problems similar to (34) in order to obtain the image  $\mathcal{Q}_2(u^c)$  of a second correction operator  $\mathcal{Q}_2$  under  $u^c$ . Then we can compute the size of

$$\|f + \operatorname{div} A^\epsilon(\cdot, \left( u^c + \delta \mathcal{Q}(u^c)(x_T, \frac{\cdot - x_T}{\delta}) + \delta^2 \mathcal{Q}_2(u^c)(x_T, \frac{\cdot - x_T}{\delta}) \right))\|_{L^2(Y_{T,\epsilon})}.$$

Depending on the technique that we use to derive the corresponding discrete estimate, we might deal with gradient jumps or additional reconstructions. There are other factors, like the size of the cells  $Y_{T,\epsilon}$ , that should be considered when speaking about the applicability of the HMM. Furthermore, to make sure that the complete fine-scale behaviour is captured by this strategy, it might become necessary to create a covering of  $\Omega$  by the union of the cells  $Y_{T,\epsilon}$ . Then the HMM can be interpreted as a domain decomposition method with oversampling, which is typically still cheaper than solving the whole fine-scale problem.

However, since there is nothing done so far on this field, a detailed analysis of this part exceeds the scope of this work and will be subject of future research. From now on, we assume the a-priori knowledge that the HMM is applicable in our scenario and that we can therefore solely focus on a comparison with the limit problem.

## 4 The a-posteriori error estimates

In this section we want to state an a-posteriori error estimate for the heterogeneous multiscale finite element method introduced in Definition 2.5. For completeness, we also state an estimate for the TFR-HMM given by Definition 2.4, even though we point out that it does not make sense to use this method in the nonlinear setting. It is too expensive since it involves determining the reconstructions  $R_T(\Phi_i)$  for any base function  $\Phi_i$ . These are  $|\mathcal{T}_H|$  nonlinear problems to solve. Furthermore, we expect it to be a little less accurate.

Therefore, our focus is on the  $L^2$ -error between HMM approximation  $u_{H,h}$  and exact solution  $u^\epsilon$ . For the whole section, we presume that the applicability of HMM in the sense of Assumption 3.12 is verified (for instance by the strategies described in Section 3.4). Beside the general assumptions stated in Section 2 we suppose that we have sufficiently regular data, i.e.  $A^\epsilon \in (H^{1,\infty}(B_\delta(\Omega) \times \mathbb{R}^d))^d$ . With these presumptions, the a-posteriori error estimate stated in Theorem 4.5 is reliable without further restrictions.

All the required notations are given in Subsection 4.1. The main results are stated in Subsection 4.2 and the corresponding proofs can be found in Subsection 5.

## 4.1 Notations

For lucidity we introduce the subsequent definitions and notations.

In the following  $u_{H,h}$  denotes the solution of the HMM given by equation (9) and  $u_{H,h}^S$  denotes the solution of the TFR-HMM given by (7). The discrete scale-separated diffusion operator  $A_{H,h}^{\epsilon,\delta}$  is given by (19). Furthermore, we define the restriction of the triangulation  $\mathcal{T}_h$  of  $Y$  on  $\frac{\epsilon}{\delta}Y$  by:

$$\mathcal{T}_h^{\epsilon,\delta} := \{K \in \mathcal{T}_h \mid K \subset \frac{\epsilon}{\delta}Y\}. \quad (35)$$

For simplification, we assume that  $\mathcal{T}_h^{\epsilon,\delta}$ , on its own, is a regular, periodic triangulation of  $\frac{\epsilon}{\delta}Y$ . The corresponding set of all edges, the set of inner edges and the set of outer edges are denoted by:

$$\begin{aligned} \Gamma(\mathcal{T}_h^{\epsilon,\delta}) &:= \{E_Y \in \Gamma(\mathcal{T}_h) \mid E_Y \subset \frac{\epsilon}{\delta}\bar{Y}\}, \\ \Gamma^{inn}(\mathcal{T}_h^{\epsilon,\delta}) &:= \{E_Y \in \Gamma(\mathcal{T}_h) \mid E_Y \subset \frac{\epsilon}{\delta}Y\} \text{ and} \\ \Gamma^{out}(\mathcal{T}_h^{\epsilon,\delta}) &:= \Gamma(\mathcal{T}_h^{\epsilon,\delta}) \setminus \Gamma^{inn}(\mathcal{T}_h^{\epsilon,\delta}). \end{aligned}$$

When using outer normals and gradient jumps, we make use of the subsequent definition:

**Definition 4.1** (Outer normals and jumps). For any bounded domain  $M$ , we denote the outer normal function by  $n_M : \partial M \rightarrow \mathbb{R}^d$ . For two domains  $M_1$  and  $M_2$ , with  $\Gamma := \bar{M}_1 \cap \bar{M}_2$  and for a function  $g \in (L^\infty(M))^d$  with  $g|_{M_i} \in (C^0(M_i))^d$ ,  $i = 1, 2$ , we define the jump  $[g]_\Gamma : \Gamma \rightarrow \mathbb{R}$  of  $g$  over  $\Gamma$  by

$$[g]_\Gamma(x) := \left| \lim_{m_1 \rightarrow \infty} g(x_{m_1}) \cdot n_{M_1}(x) + \lim_{m_2 \rightarrow \infty} g(x_{m_2}) \cdot n_{M_2}(x) \right|,$$

where  $x_{m_i}$  is a sequence in  $M_i$ , with  $x_{m_i} \rightarrow x$ .

In order to use jumps over pairwise opposite edges of  $\Gamma^{out}(\mathcal{T}_h^\kappa)$ , we also need to introduce the set  $\Gamma(\mathcal{T}_h^\kappa) / \sim_{\kappa Y}$ :

**Definition 4.2** (Equivalence relation on sets of edges). For  $i \in \{1, \dots, d\}$ ,  $m \in \{1, 2\}$  we define the mapping  $g_i^m : \mathbb{R}^d \rightarrow \mathbb{R}^d$  by

$$g_i^m(x_1, \dots, x_n) := \begin{cases} x_j & \text{for } j \in \{1, \dots, d\} \setminus \{i\}, \\ (-1)^m \frac{\kappa}{2} & \text{for } j = i. \end{cases}$$

The set of all these mappings plus identity is given by  $G := \{g_i^m \mid 1 \leq i \leq d; m = 1, 2\} \cup \{id\}$ . Now, let  $\kappa \in \mathbb{R}_{>0}$  and  $\mathcal{T}_h^\kappa$  be a regular, periodic triangulation of  $\kappa Y$ . We define the following equivalence relation  $\sim_{\kappa Y}$  on the set of edges  $\Gamma(\mathcal{T}_h^\kappa)$ . For  $E, \tilde{E} \in \Gamma(\mathcal{T}_h^\kappa)$ , we say

$$E \sim_{\kappa Y} \tilde{E} \iff \exists g \in G \text{ with } g(E) = \tilde{E}.$$

With this definition, we naturally extend the definition of jumps over elements of  $\Gamma(\mathcal{T}_h^\kappa)/\sim_{\kappa Y}$  (see Figure 4.1).

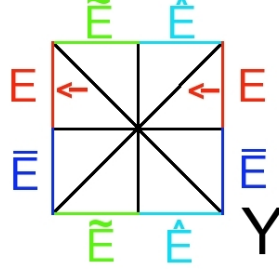


Figure 1: In  $\Gamma(\mathcal{T}_h^\kappa)/\sim_{\kappa Y}$ , pairwise opposite boundary edges are identified with the same element. Due to this identification, jumps over such an edge can be used analogously to Definition 4.1.

The following operators are required for a suitable a-posteriori error estimate for the TFR-HMM. They are not required for the case without test function reconstruction.

**Definition 4.3.** Let  $V_H^0(\Omega)$  denote the space of piecewise constant functions on  $\Omega$  with regard to  $\mathcal{T}_H$ . For  $x \in T \in \mathcal{T}_H$ , we define the non-linear operator  $\tilde{Q}_h : V_H \rightarrow V_H^0(\Omega, W_h(\frac{\epsilon}{\delta}Y))$  by:  $\tilde{Q}_h(\Phi_H)(x_T, \cdot) \in W_h(\frac{\epsilon}{\delta}Y)$  solves

$$\int_{\frac{\epsilon}{\delta}Y} A_{H,h}^{\epsilon,\delta} \left( x_T, y, \nabla_x \Phi_H(x_T) + \nabla_y \tilde{Q}_h(\Phi_H)(x_T, y) \right) \cdot \nabla_y \phi_h(y) dy = 0, \quad (36)$$

for all  $\phi_h \in W_h(\frac{\epsilon}{\delta}Y)$  and all  $\Phi_H \in V_H$ . Furthermore, we introduce associated localized versions  $\tilde{\mathcal{K}}_h^T : \mathbb{P}^1(T)/\mathbb{R} \rightarrow W_h(\frac{\epsilon}{\delta}Y)$  by

$$\tilde{\mathcal{K}}_h^T(\Phi_H) := \tilde{Q}_h(\Phi_H)(x_T, \cdot)$$

and  $\mathcal{K}_h^T : \mathbb{P}^1(T)/\mathbb{R} \rightarrow V_h(\frac{\epsilon}{\delta}Y)$  by

$$\mathcal{K}_h^T(\Phi_H) := Q_h(\Phi_H)(x_T, \cdot)|_{\frac{\epsilon}{\delta}Y},$$

where  $Q_h(\nabla \Phi_H(\cdot)) = Q_h(\Phi_H)$  is given by (18).

For any (non-linear) operator  $G : \mathbb{P}^1(T)/\mathbb{R} \rightarrow H^1(\frac{\epsilon}{\delta}Y)$ , we define a corresponding induced norm by:

$$\|G\|_{Op} := \sup \left\{ \frac{\|G(\Phi_1) - G(\Phi_2)\|_{H^1(\frac{\epsilon}{\delta}Y)}}{|\nabla \Phi_1 - \nabla \Phi_2|} \mid \Phi_1, \Phi_2 \in \mathbb{P}^1(T), \Phi_1 \neq \Phi_2 \right\}.$$

Now, we are prepared to state the local error indicators, which contribute to the later a-posteriori estimates.

**Definition 4.4** (Local error indicators). For  $\Phi_H \in V_H$ , the local *approximation error indicators* are defined by

$$\eta_T^{app}(\Phi_H) := \left\| \int_Y w^{\epsilon,\delta}(y) \left( A^{\epsilon,\delta} - A_h^{\epsilon,\delta} \right) (\cdot, y, \nabla_x \Phi_H(x_T) + \nabla_y Q_h(\Phi_H)(x_T, y)) dy \right\|_{L^2(T)}^2 \quad \text{and}$$

$$\bar{\eta}_T^{app}(\Phi_H) := \left\| \int_Y \left( A^{\epsilon,\delta} - A_h^{\epsilon,\delta} \right) (\cdot, y, \nabla_x \Phi_H(x_T) + \nabla_y Q_h(\Phi_H)(x_T, y)) dy \right\|_{L^2(T)}^2.$$

The local *residual error indicators* are given by

$$\begin{aligned}\eta_E^{res}(\Phi_H) &:= H_E^3 \left\| \int_Y [w^{\epsilon, \delta}(y) A_h^{\epsilon, \delta}(\cdot, y, \nabla_x \Phi_H + \nabla_y \mathcal{Q}_h(\Phi_H)(\cdot, y))]_E dy \right\|_{L^2(E)}^2 \quad \text{and} \\ \bar{\eta}_T^{res}(\Phi_H) &:= \sum_{E_Y \in \Gamma(\mathcal{T}_h)/\sim_Y} h_{E_Y}^3 \left\| [A_h^{\epsilon, \delta}(\cdot, \cdot, \nabla_x \Phi_H + \nabla_y \mathcal{Q}_h(\Phi_H))]_{E_Y} \right\|_{L^2(T \times E_Y)}^2.\end{aligned}$$

For the case that we use the TFR-HMM, we also need local *error indicators for test function reconstruction*:

$$\begin{aligned}\eta_T^{\text{TFR}}(\Phi_H) &:= \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})/\sim_{\frac{\epsilon}{5} Y}} \left\| [A_h^{\epsilon, \delta}(\cdot, \cdot, \nabla_x \Phi_H + \nabla_y \mathcal{Q}_h(\Phi_H))]_{E_Y} \right\|_{L^2(T \times E_Y)}^2, \\ \bar{\eta}_T^{\text{TFR}}(\Phi_H) &:= \|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{\mathcal{O}_p} \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})^{inn}} \left\| \frac{\delta^d}{\epsilon^d} [A_h^{\epsilon, \delta}(\cdot, \cdot, \nabla_x \Phi_H + \nabla_y \mathcal{Q}_h(\Phi_H))]_{E_Y} \right\|_{L^2(T \times E_Y)}^2 \quad \text{and} \\ \bar{\bar{\eta}}_T^{\text{TFR}}(\Phi_H) &:= \|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{\mathcal{O}_p} \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})^{out}} \left\| \frac{\delta^d}{\epsilon^d} A_h^{\epsilon, \delta}(\cdot, \cdot, \nabla_x \Phi_H + \nabla_y \mathcal{Q}_h(\Phi_H)) \cdot n_{\frac{\epsilon}{5} Y} \right\|_{L^2(T \times E_Y)}^2.\end{aligned}$$

In the next subsection we state our main result concerning the a-posteriori error estimates.

## 4.2 The estimates

Using the notations that we introduced in the preceding subsection, we now present two a-posteriori error estimates for the  $L^2$ -difference between HMM approximation ( $u_{H,h}$  or  $u_{H,h}^S$ ) and exact solution  $u^\epsilon$ . Under the given assumptions, these estimates are reliable. The proof of the subsequent theorem is given in Section 5.

**Theorem 4.5** (A-posteriori error estimate). *Let the local error indicators be given by Definition 4.4. Under the assumptions mentioned at the beginning of this section we get the following a-posteriori error estimate for the  $L^2$ -error between HMM approximation  $u_{H,h}$  and  $u^\epsilon$ :*

$$\begin{aligned}\|u_{H,h} - u^\epsilon\|_{L^2(\Omega)} &\leq C \left( \sum_{T \in \mathcal{T}_H} H_T^4 \|f\|_{L^2(T)}^2 \right)^{\frac{1}{2}} + C \left( \sum_{T \in \mathcal{T}_H} \eta_T^{app}(u_{H,h}) + \bar{\eta}_T^{app}(u_{H,h}) \right)^{\frac{1}{2}} \quad (37) \\ &\quad + C \left( \sum_{E \in \Gamma(\mathcal{T}_H)} \eta_E^{res}(u_{H,h}) \right)^{\frac{1}{2}} + C \left( \sum_{T \in \mathcal{T}_H} \bar{\eta}_T^{res}(u_{H,h}) \right)^{\frac{1}{2}}.\end{aligned}$$

For the solution  $u_{H,h}^S$  of the HMM with test function reconstruction we get:

$$\begin{aligned}\|u_{H,h}^S - u^\epsilon\|_{L^2(\Omega)} &\leq C \left( \sum_{T \in \mathcal{T}_H} H_T^4 \|f\|_{L^2(T)}^2 \right)^{\frac{1}{2}} + C \left( \sum_{T \in \mathcal{T}_H} \eta_T^{app}(u_{H,h}^S) + \bar{\eta}_T^{app}(u_{H,h}^S) \right)^{\frac{1}{2}} \quad (38) \\ &\quad + C \left( \sum_{E \in \Gamma(\mathcal{T}_H)} \eta_E^{res}(u_{H,h}^S) \right)^{\frac{1}{2}} + C \left( \sum_{T \in \mathcal{T}_H} \bar{\eta}_T^{res}(u_{H,h}^S) \right)^{\frac{1}{2}} \\ &\quad + C \left( \sum_{T \in \mathcal{T}_H} \eta_T^{\text{TFR}}(u_{H,h}^S) \right)^{\frac{1}{2}} + C \left( \sum_{T \in \mathcal{T}_H} \bar{\eta}_T^{\text{TFR}}(u_{H,h}^S) \right)^{\frac{1}{2}} + C \left( \sum_{T \in \mathcal{T}_H} \bar{\bar{\eta}}_T^{\text{TFR}}(u_{H,h}^S) \right)^{\frac{1}{2}}.\end{aligned}$$

By  $C$  we denote constants not depending on  $\epsilon, \delta, H$  and  $h$ . For further specifications, one might follow the proofs in Subsection 5. If we replace  $\|u_{H,h} - u^c\|_{L^2(\Omega)}$  by  $\|u_{H,h} - u^\epsilon\|_{L^2(\Omega)}$  on the left hand side of (37) (and analogously for  $u_{H,h}^S$  in (38)), we obtain an additional  $O(\delta)$ -term on the right hand side in both estimates.

Note that (38) is only stated for the sake of completeness. The TFR-HMM is far more expensive than the HMM and the corresponding estimated error is far larger, even if the real errors  $\|u_{H,h} - u^\epsilon\|_{L^2(\Omega)}$  and  $\|u_{H,h}^S - u^\epsilon\|_{L^2(\Omega)}$  are of identical size.

In Section 3.3 we saw that the sequence of HMM approximations converges to the homogenized solution  $u^0$  if we are in a periodic setting, i.e. if  $A^\epsilon(x, \xi) = A(x, \frac{x}{\epsilon}, \xi)$  and  $A(x, \cdot, \xi)$   $Y$ -periodic. This implies that we can use (37) to efficiently determine the solution of a nonlinear elliptic homogenization problem. In such a scenario  $u^c$  is the homogenized solution  $u^0$  and (37) is directly applicable.

If we are in a more general setting, the remaining model error  $\|u^c - u^\epsilon\|_{L^2(\Omega)} = O(\delta)$  could be replaced by a reliable indicator if we use for instance the strategies presented in [47, 40, 41, 45].

**Remark 5** (Error contributions). For the heterogeneous multiscale method (9), we can see, that the a-posteriori error estimate (37) consists of two components: one part to estimate the approximation error and one part to estimate the residual error. If we use a sufficiently accurate approximation of the coefficient function  $A^\epsilon$ , the order of convergence of the HMM can be guessed by observing the indicators  $\eta_E^{res}$  and  $\eta_T^{res}$ . Due to the fact that gradient jumps typically yield convergence of order  $O(H^{\frac{1}{2}})$ , we deduce that we can expect a quadratic order of convergence for this method (in  $H$  and  $h$ ). Indeed, proceeding similar to the proofs in Subsection 5, we could derive an a-priori error estimate which formally confirms this. Such a-priori error estimates were carried out in [24] for the linear case following similar ideas.

For the heterogeneous multiscale method with test function reconstruction (7), the a-posteriori error estimate (38) consists of three components. Besides the contributions of approximation and residual error, we also get indicators which account for a possible inaccuracy, produced by the TFR. It is obvious that the size of the a-posteriori bound for  $\|u_{H,h}^S - u^c\|_{L^2(\Omega)}$  strongly depends on  $\eta_T^{TFR}$ ,  $\bar{\eta}_T^{TFR}$  and  $\bar{\bar{\eta}}_T^{TFR}$ . It may remain large, although the rest becomes small. Even if Assumption 3.12 is fulfilled, we cannot guess the order of convergence for the TFR-HMM. More precisely, in order to guarantee that the additional contribution made by test function reconstruction becomes small,  $\mathcal{Q}_h(u_{H,h}^S)(x_T, \cdot)$  needs to hit an almost periodic boundary condition on  $\frac{\epsilon}{\delta}Y$ . If this is not the case, there is formally no reason for assuming that the method is accurate and as a result of this, that  $\eta_T^{TFR}$  and  $\bar{\eta}_T^{TFR}$  remain negligible. Identifying the limit problem of the TFR-HMM would improve the situation only cursorily.

In order to use estimate (38), we need to compute operator norms  $\|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{O_p}$  which occur in  $\bar{\eta}_T^{TFR}$  and  $\bar{\bar{\eta}}_T^{TFR}$ . In the following remark, we comment on these computations:

**Remark 6.** In practical applications we are never concerned with computing an error estimator for the TFR-HMM. The following statements therefore just refer to the treatment of  $\|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{O_p}$  in numerical experiments.

Let  $\mathcal{B}_T$  denote the Lagrange base of  $\mathbb{P}^1(T)/\mathbb{R}$ , then we define for  $G : \mathbb{P}^1(T)/\mathbb{R} \rightarrow H^1(\frac{\epsilon}{\delta}Y)$  the quantity  $\|\cdot\|_{O_p}^*$  by

$$\|G\|_{O_p}^* := \left( \sum_{\Phi \in \mathcal{B}_T} \frac{\|G(\Phi)\|_{H^1(\frac{\epsilon}{\delta}Y)}^2}{|\nabla \Phi|^2} \right)^{\frac{1}{2}}. \quad (39)$$

In the linear setting, we can easily control  $\|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{O_p}$  by using  $\|\cdot\|_{O_p}^*$  (which defines a norm in this situation). We observe that the following estimates hold true

$$\|G\|_{O_p}^* = \left( \sum_{\Phi \in \mathcal{B}_T} \frac{\|G(\Phi)\|_{H^1(\frac{\epsilon}{\delta}Y)}^2}{|\nabla\Phi|^2} \right)^{\frac{1}{2}} \leq |\mathcal{B}_T|^{\frac{1}{2}} \|G\|_{O_p}$$

and

$$\begin{aligned} \|G\|_{O_p} &= \sup \left\{ \frac{\|G(\Psi_1) - G(\Psi_2)\|_{H^1(\frac{\epsilon}{\delta}Y)}}{|\nabla\Psi_1 - \nabla\Psi_2|} \mid \Psi_1, \Psi_2 \in \mathbb{P}^1(T), \Psi_1 \neq \Psi_2 \right\} \\ &= \sup \left\{ \left\| \sum_{\Phi \in \mathcal{B}_T} \frac{\alpha_\Phi}{|\nabla\Phi|} G(\Phi) \right\|_{H^1(\frac{\epsilon}{\delta}Y)} \mid \sum_{\Phi \in \mathcal{B}_T} \frac{\alpha_\Phi}{|\nabla\Phi|} \nabla\Phi = 1 \right\} \\ &\leq |\mathcal{B}_T|^{\frac{1}{2}} \|G\|_{O_p}^* \sup \left\{ |\alpha_\Phi| \mid \sum_{\Phi \in \mathcal{B}_T} \frac{\alpha_\Phi}{|\nabla\Phi|} \nabla\Phi = 1 \right\}. \end{aligned}$$

Since  $\left| \sum_{\Phi \in \mathcal{B}_T} \frac{\alpha_\Phi}{|\nabla\Phi|} \nabla\Phi \right|^2$  is a non-constant polynomial of second order in  $\mathbb{R}^{|\mathcal{B}_T|}$ , there is only a finite number of intersection with the straight line  $p(\alpha) = 1$ . Therefore we have that the term  $\sup \left\{ |\alpha_\Phi| \mid \sum_{\Phi \in \mathcal{B}_T} \frac{\alpha_\Phi}{|\nabla\Phi|} \nabla\Phi = 1 \right\}$  is bounded. Due to these observations, we can compute  $\|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{O_p}^*$  instead of  $\|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{O_p}$ , since they increase with the same rate, independent of meshsizes. Moreover,  $\mathcal{K}_h^T(\Phi)$  is already available for  $\Phi \in \mathcal{B}_T$ .

In the non-linear setting, we do not have the second estimate. Nevertheless, there are three possibilities to overcome the problems that occur with  $\|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{O_p}$ . First of all, we might use the same strategy as in the linear setting, even though it is not reliable. Since  $\|G\|_{O_p}^* \leq |\mathcal{B}_T|^{\frac{1}{2}} \|G\|_{O_p}$ , we know that if  $\|G\|_{O_p}$  increases, then so will  $\|G\|_{O_p}^*$ . But it does not work vice versa. In order to be accurate we can use a direct, very expensive computation of  $\|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{O_p}$  by formulating a corresponding maximization problem in  $\mathbb{R}^{|\mathcal{B}_T|}$ . The best way however, seems to be in completely ignoring this term and setting it equal to zero. This is consequent, if we consider that the essential information about the TFR-error, is already incorporated in the gradient jumps of  $u_{H,h}^S$  over pairwise opposite edges of  $\Gamma^{out}(\mathcal{T}_h^{\frac{\epsilon}{\delta}})$ . In this case,  $\eta_T^{\text{TFR}}(u_{H,h}^S)$  dominates  $\bar{\eta}_T^{\text{TFR}}(u_{H,h}^S)$ , since  $\|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{O_p}$  is merely a speed up for  $\bar{\eta}_T^{\text{TFR}}(u_{H,h}^S)$  getting small (see Lemma 3.8 to verify that  $\|\tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T\|_{O_p}$  is bounded). In  $\bar{\eta}_T^{\text{TFR}}(u_{H,h}^S)$ , we only have a summation over a small number of boundary edges. This suggests that  $\eta_T^{\text{TFR}}(u_{H,h}^S)$  on its own, is already a perfect indicator for the size of the TFR-error.

## 5 Proof of the a-posteriori error estimates

In this section we use the notations introduced in Section 4.1. Furthermore, we suppose that the assumptions stated at the beginning of Section 4 hold true. Exemplarily, all the subsequent proofs are only carried out for the TFR-HMM approximation  $u_{H,h}^S$  since this case contains all the essential ideas.

We start with formulating a dual problem to the  $\delta$ - $\epsilon$ -homogenized problem (23). This is used for deriving an error identity for  $\|u_{H,h}^S - u^c\|_{L^2(\Omega)}$ , which can be estimated in a straight forward way.

To state a linearized dual problem of the  $\delta$ - $\epsilon$ -homogenized problem, we first need to introduce the mean value linearization of  $A^{\epsilon,\delta}$ :

**Definition 5.1.** For  $\zeta_i : \Omega \times Y \rightarrow \mathbb{R}^d$ ,  $i = 1, 2$ , we define the entries of the *mean-value linearization*  $\bar{A}^{\epsilon,\delta}(\zeta_1, \zeta_2)$  of  $A^{\epsilon,\delta}$  by:

$$\bar{a}_{ij}^{\epsilon,\delta}(\zeta_1, \zeta_2)(x, y) := \int_0^1 \partial_{\xi_j} a_i^{\epsilon,\delta}(x, y, \theta \zeta_1(x, y) + (1 - \theta) \zeta_2(x, y)) d\theta.$$

Here,  $a_i^{\epsilon,\delta}$  denotes the  $i$ th component of the operator  $A^{\epsilon,\delta}$  and  $\bar{a}_{ij}^{\epsilon,\delta}(\zeta_1, \zeta_2)$  denotes the  $(i, j)$ -entry of the matrix  $\bar{A}^{\epsilon,\delta}(\zeta_1, \zeta_2)$ .

Due to the following lemma,  $\bar{A}^{\epsilon,\delta}(\zeta_1, \zeta_2)$  is an elliptic matrix with coercivity constant  $\alpha$ .

**Lemma 5.2.** *Let  $(X, \|\cdot\|)$  denote a reflexive Banach space with dual space  $X'$  and let  $B : X \rightarrow X'$  denote a non-linear operator with Frechet derivative  $B' : X \rightarrow L(X, X')$ . If  $A$  is strongly monotone, i.e.*

$$(B(v_1) - B(v_2))(v_1 - v_2) \geq \alpha \|v_1 - v_2\|^2 \quad \forall v_1, v_2 \in X \text{ and } \alpha > 0,$$

then we have

$$B'(u)(v)(v) > \alpha \|v\|^2 \quad \forall u, v \in X.$$

*Proof.* We have

$$B'(u)(v) = \lim_{s \rightarrow 0} \frac{B(u + sv) - B(u)}{s}.$$

This and the strong monotonicity yield:

$$\begin{aligned} B'(u)(v)(v) &= \lim_{s \rightarrow 0} \frac{(B(u + sv) - B(u))(v)}{s} \\ &= \lim_{s \rightarrow 0} \frac{1}{s^2} (B(u + sv) - B(u))(u + sv - u) \\ &\geq \lim_{s \rightarrow 0} \frac{1}{s^2} \alpha \|sv\|^2 = \alpha \|v\|^2. \end{aligned}$$

□

Now, we introduce the mean value linearized correction operator  $\bar{Q}(\zeta_1, \zeta_2)$ :

**Definition 5.3.** For  $\zeta_i : \Omega \times Y \rightarrow \mathbb{R}^d$ ,  $i = 1, 2$  we define the operator

$$\bar{Q}(\zeta_1, \zeta_2) : \mathbb{R}^d \rightarrow L^2(\Omega, \tilde{H}_\#^1(Y)),$$

where  $\bar{Q}(\zeta_1, \zeta_2)(\xi)$  is defined (see (12) for comparison) as the solution of:

$$\int_Y \bar{A}^{\epsilon,\delta}(\zeta_1, \zeta_2)(x, y) (\xi + \nabla_y \bar{Q}(\zeta_1, \zeta_2)(\xi)(x, y)) \cdot \nabla_y \phi(y) dy = 0 \quad \forall \phi \in \tilde{H}_\#^1(Y). \quad (40)$$

Note that  $\bar{Q}(\zeta_1, \zeta_2)$  is well defined because of the coercivity of  $\bar{A}^{\epsilon,\delta}(\zeta_1, \zeta_2)$ .

For formulating the dual problem we choose  $(\zeta_1, \zeta_2) = (\nabla_x u^c + \nabla_y u^f, \nabla_x u_{H,h}^S + \nabla_y \mathcal{Q}_h(u_{H,h}^S))$ :

**Definition 5.4.** Let  $u^c \in \dot{H}^1(\Omega)$  denote the solution of the  $\delta$ - $\epsilon$ -homogenized problem (16) and let  $u_{H,h}^S \in V_H$  denote the TFR-HMM approximation given by (7). We define the entries of the matrix  $\hat{A}^{\epsilon,\delta}$  by:

$$\hat{a}_{ij}^{\epsilon,\delta} := \bar{a}_{ij}^{\epsilon,\delta}(\nabla_x u^c + \nabla_y \mathcal{Q}(u^c), \nabla_x u_{H,h}^S + \nabla_y \mathcal{Q}_h(u_{H,h}^S)),$$

where  $\bar{A}_{ij}^{\epsilon,\delta}$  is given by Definition 5.1. Analogously we define  $\hat{\mathcal{Q}}$  by

$$\hat{\mathcal{Q}}(\Phi) := \bar{\mathcal{Q}}(\nabla_x u^c + \nabla_y \mathcal{Q}(u^c), \nabla_x u_{H,h}^S + \nabla_y \mathcal{Q}_h(u_{H,h}^S))(\nabla \Phi),$$

where  $\Phi \in \dot{H}^1(\Omega)$  and  $\bar{\mathcal{Q}}$  is given by equation (40). The entries of the corresponding linearized  $\delta$ - $\epsilon$ -homogenized matrix  $\hat{A}^0$  are therefore defined (see (13) for comparison) by:

$$\hat{a}_{ij}^0(x) := \int_Y w^{\epsilon,\delta}(y) \hat{A}^{\epsilon,\delta}(x, y) \left( e_i + \nabla_y \hat{\mathcal{Q}}(e_i)(x, y) \right) \cdot e_j dy.$$

**Remark 7.** Observe that the mean-value linearization  $\hat{A}^{\epsilon,\delta}$  fulfills the equation

$$\begin{aligned} & A^{\epsilon,\delta}(x, y, \nabla_x u^c(x) + \nabla_y \mathcal{Q}(u^c)(x, y)) - A^{\epsilon,\delta}(x, y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y)) \\ &= \hat{A}^{\epsilon,\delta}(x, y)(\nabla_x u^c(x) + \nabla_y \mathcal{Q}(u^c)(x, y) - \nabla_x u_{H,h}^S(x) - \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y)), \end{aligned}$$

since

$$\begin{aligned} & \left( \hat{A}^{\epsilon,\delta}(x, y)(\nabla_x u^c(x) + \nabla_y \mathcal{Q}(u^c)(x, y) - \nabla_x u_{H,h}^S(x) - \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y)) \right)_i \\ &= \int_0^1 \sum_{j=1}^d \partial_{\xi_j} a_i^{\epsilon,\delta}(x, y, \theta(\nabla_x u^c(x) + \nabla_y \mathcal{Q}(u^c)(x, y)) + (1-\theta)(\nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y))) \\ & \quad (\partial_{x_j} u^c(x) + \partial_{y_j} \mathcal{Q}(u^c)(x, y) - \partial_{x_j} u_{H,h}^S(x) - \partial_{y_j} \mathcal{Q}_h(u_{H,h}^S)(x, y)) d\theta \\ &= \int_0^1 \frac{d}{d\theta} a_i^{\epsilon,\delta}(x, y, \theta(\nabla_x u^c(x) + \nabla_y \mathcal{Q}(u^c)(x, y)) + (1-\theta)(\nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y))) d\theta. \end{aligned}$$

Now we are prepared to state the final dual problem:

**Definition 5.5** (Dual Problem). The solution  $z^c \in \dot{H}^1(\Omega)$  of the linearized dual problem is defined through

$$\int_{\Omega} \hat{A}^0(x) \nabla_x \Phi(x) \cdot \nabla_x z^c(x) dx = \int_{\Omega} (u_{H,h}^S - u^c)(x) \Phi(x) dx \quad \forall \Phi \in \dot{H}^1(\Omega). \quad (41)$$

Just like for the  $\delta$ - $\epsilon$ -homogenized operator  $A^0$  (given by (13)) coercivity of the corresponding dual operator  $\hat{A}^0$  is only clear for  $\delta = \epsilon$ . For the general case we can simplify the situation and legitimately presume the following in analogy to Assumption 3.2:

**Assumption 5.6** (Model assumption 3). We suppose that the linearized  $\delta$ - $\epsilon$ -homogenized matrix  $\hat{A}^0$  is coercive with  $\bar{\alpha} > 0$ , i.e.:

$$\hat{A}^0(x) \xi \cdot \xi \geq \bar{\alpha} |\xi|^2 \quad \forall \xi \in \mathbb{R}^d, \quad \forall x \in \Omega. \quad (42)$$

From now on, we presume that this assumption holds true.

**Remark 8.** Standard estimates for linear elliptic problems and the regularity estimate from the Friedrichs Theorem (cf. [9], Appendix 10) for convex domains yield that there exist constants  $C_i^c$  independent of  $H$  and  $h$  with:

$$|z^c|_{H^i(\Omega)} \leq C_i^c \|u_{H,h}^S - u^c\|_{L^2(\Omega)}, \quad \text{where } 0 \leq i \leq 2. \quad (43)$$

Note that we have the required regularity due to the Lipschitz-continuity of  $A^\epsilon(\cdot, \xi)$ .

**Lemma 5.7.** For  $z^c \in \dot{H}^1(\Omega)$ , there exists  $z^f \in L^2(\Omega, \tilde{H}_\#^1(Y))$  with:

$$\int_Y \hat{A}^{\epsilon, \delta}(x, y) \nabla_y \phi(y) \cdot (w^{\epsilon, \delta}(y) \nabla_x z^c(x) + \nabla_y z^f(x, y)) dy = 0 \quad \forall \phi \in \tilde{H}_\#^1(Y). \quad (44)$$

Due to the continuous dependency of  $z^f$  on  $z^c$  we obtain

$$|z^f|_{L^2(\Omega, H^i(Y))} \leq C_i^f \|u_{H,h}^S - u^c\|_{L^2(\Omega)}, \quad \text{where } 0 \leq i \leq 2 \quad (45)$$

and  $C_i^f$  independent of  $H$  and  $h$ .

This lemma follows immediately from (41) and the definition of  $\hat{A}^0$ .

Now, we are ready for the proof of our a-posteriori main result, i.e. Theorem 4.5.

*Proof of Theorem 4.5.* Let  $\mathcal{L}_H : C^0(\Omega) \rightarrow V_H$  and  $\mathcal{L}_h : C^0(Y) \rightarrow V_h^1(Y) \cap C_\#^0(Y)$  denote Lagrange interpolation operators (eg. [12], Section 2). Since  $\mathcal{T}_h$  is a periodic partition of  $Y$ , periodicity is preserved by the interpolation operator. Therefore,  $\tilde{\mathcal{L}}_h : L^2(\Omega, C_\#^0(Y)) \rightarrow L^2(\Omega, W_h(Y))$  is well defined by:

$$\tilde{\mathcal{L}}_h(\phi)(x, y) := \mathcal{L}_h(\phi(x, \cdot))(y) - \int_Y \mathcal{L}_h(\phi(x, \cdot))(y) dy \quad \text{for } \phi \in L^2(\Omega, C_\#^0(Y)).$$

$\mathcal{L}_H(z^c) \in V_H(\Omega)$  and  $\tilde{\mathcal{L}}_h(z^f) \in L^2(\Omega, W_h(Y))$  are suitable test functions in (18) and (21) respectively. Using the definitions of  $z^c$ ,  $z^f$  and  $u^f := \mathcal{Q}(u^c)$  we obtain the following error identity:

$$\begin{aligned} & \|u_{H,h}^S - u^c\|_{L^2(\Omega)}^2 \\ &= \int_\Omega \hat{A}^0(x) \nabla_x (u_{H,h}^S - u^c)(x) \cdot \nabla_x z^c(x) dx \\ &= \int_\Omega \int_Y w^{\epsilon, \delta}(y) \hat{A}^{\epsilon, \delta}(x, y) \left( \nabla_x (u_{H,h}^S - u^c)(x) + \nabla_y \hat{\mathcal{Q}}(u_{H,h}^S - u^c)(x, y) \right) \cdot \nabla_x z^c(x) dy dx \\ &\stackrel{(44)}{=} \int_\Omega \int_Y w^{\epsilon, \delta}(y) \hat{A}^{\epsilon, \delta}(x, y) \left( \nabla_x (u_{H,h}^S - u^c)(x) + \nabla_y \hat{\mathcal{Q}}(u_{H,h}^S - u^c)(x, y) \right) \cdot \nabla_x z^c(x) dy dx \\ &\quad + \int_\Omega \int_Y \hat{A}^{\epsilon, \delta}(x, y) \nabla_y (\mathcal{Q}_h(u_{H,h}^S) - \mathcal{Q}(u^c))(x, y) \cdot (w^{\epsilon, \delta}(y) \nabla_x z^c(x) + \nabla_y z^f(x, y)) dy dx \end{aligned}$$

Now, we rearrange the terms on the right hand side and use (44) with  $\phi = \hat{\mathcal{Q}}(u_{H,h}^S - u^c)$  to obtain

$$\begin{aligned} & \|u_{H,h}^S - u^c\|_{L^2(\Omega)}^2 \\ &= \int_\Omega \int_Y \hat{A}^{\epsilon, \delta}(x, y) \left( \nabla_x (u_{H,h}^S - u^c)(x) + \nabla_y (\mathcal{Q}_h(u_{H,h}^S) - \mathcal{Q}(u^c))(x, y) \right) \cdot (w^{\epsilon, \delta}(y) \nabla_x z^c(x)) dy dx \\ &\quad + \int_\Omega \int_Y \hat{A}^{\epsilon, \delta}(x, y) \nabla_y (\mathcal{Q}_h(u_{H,h}^S) - \mathcal{Q}(u^c))(x, y) \cdot \nabla_y z^f(x, y) dy dx \\ &\quad - \int_\Omega \int_Y \hat{A}^{\epsilon, \delta}(x, y) \left( \nabla_y \hat{\mathcal{Q}}(u_{H,h}^S - u^c)(x, y) \right) \cdot \nabla_y z^f(x, y) dy dx. \end{aligned}$$

With this equation, with (40) and with the definition of  $\hat{Q}$  we get:

$$\begin{aligned} & \|u_{H,h}^S - u^c\|_{L^2(\Omega)}^2 \\ &= \int_{\Omega} \int_Y \hat{A}^{\epsilon,\delta}(x,y) (\nabla_x(u_{H,h}^S - u^c)(x) + \nabla_y(\mathcal{Q}_h(u_{H,h}^S) - \mathcal{Q}(u^c))(x,y)) \cdot (w^{\epsilon,\delta}(y) \nabla_x z^c(x)) \, dy \, dx \\ & \quad + \int_{\Omega} \int_Y \hat{A}^{\epsilon,\delta}(x,y) (\nabla_x(u_{H,h}^S - u^c)(x) + \nabla_y(\mathcal{Q}_h(u_{H,h}^S) - \mathcal{Q}(u^c))(x,y)) \cdot \nabla_y z^f(x,y) \, dy \, dx. \end{aligned}$$

In the next step, we use Remark 7:

$$\begin{aligned} & \|u_{H,h}^S - u^c\|_{L^2(\Omega)}^2 \\ &= - \int_{\Omega} \int_Y A^{\epsilon,\delta}(x,y) (\nabla_x u^c(x) + \nabla_y u^f(x,y)) \cdot (w^{\epsilon,\delta}(y) \nabla_x z^c(x) + \nabla_y z^f(x,y)) \, dy \, dx \\ & \quad + \int_{\Omega} \int_Y A^{\epsilon,\delta}(x,y) (\nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x,y)) \cdot (w^{\epsilon,\delta}(y) \nabla_x z^c(x) + \nabla_y z^f(x,y)) \, dy \, dx. \end{aligned}$$

Using (23) and artificially inserting  $A_h^{\epsilon,\delta}$  yields:

$$\begin{aligned} & \|u_{H,h}^S - u^c\|_{L^2(\Omega)}^2 \\ &= - \int_{\Omega} f(x) z^c(x) \, dx \\ & \quad + \int_{\Omega} \int_Y \left( (A^{\epsilon,\delta} - A_h^{\epsilon,\delta})(x,y,\cdot) \right) (\nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x,y)) \cdot (w^{\epsilon,\delta}(y) \nabla_x z^c(x)) \, dy \, dx \\ & \quad + \int_{\Omega} \int_Y \left( (A^{\epsilon,\delta} - A_h^{\epsilon,\delta})(x,y,\cdot) \right) (\nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x,y)) \cdot \nabla_y z^f(x,y) \, dy \, dx \\ & \quad + \int_{\Omega} \int_Y A_h^{\epsilon,\delta}(x,y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x,y)) \cdot (w^{\epsilon,\delta}(y) \nabla_x z^c(x) + \nabla_y z^f(x,y)) \, dy \, dx. \end{aligned}$$

To conclude the identity we require (21) and the definition of  $u_{H,h}^S$  (Galerkin orthogonality). Furthermore, we add and subtract terms depending on  $\tilde{Q}_h$ :

$$\begin{aligned} & \|u_{H,h}^S - u^c\|_{L^2(\Omega)}^2 \\ &= \underbrace{\int_{\Omega} f(x) (\mathcal{L}_H(z^c)(x) - z^c(x)) \, dx}_{=:I} \\ & \quad + \underbrace{\int_{\Omega} \int_Y \left( A^{\epsilon,\delta} - A_h^{\epsilon,\delta} \right) (x,y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x,y)) \cdot (w^{\epsilon,\delta}(y) \nabla_x z^c(x)) \, dy \, dx}_{=:II_1} \\ & \quad + \underbrace{\int_{\Omega} \int_Y \left( A^{\epsilon,\delta} - A_h^{\epsilon,\delta} \right) (x,y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x,y)) \cdot \nabla_y z^f(x,y) \, dy \, dx}_{=:II_2} \\ & \quad + \underbrace{\int_{\Omega} \int_Y A_h^{\epsilon,\delta}(x,y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x,y)) \cdot w^{\epsilon,\delta}(y) \nabla_x (z^c - \mathcal{L}_H(z^c))(x) \, dy \, dx}_{=:III} \\ & \quad + \underbrace{\int_{\Omega} \int_Y A_h^{\epsilon,\delta}(x,y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x,y)) \cdot \left( \nabla_y (z^f - \tilde{\mathcal{L}}_h(z^f)) (x,y) \right) \, dy \, dx}_{=:IV} \end{aligned}$$

$$\begin{aligned}
& - \underbrace{\int_{\Omega} \int_Y w^{\epsilon, \delta}(y) A_h^{\epsilon, \delta}(x, y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y)) \cdot \nabla_y \tilde{\mathcal{Q}}_h(\mathcal{L}_H(z^c)) dy dx}_{=:V} \\
& + \underbrace{\int_{\Omega} \int_Y w^{\epsilon, \delta}(y) A_h^{\epsilon, \delta}(x, y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y)) \cdot \nabla_y (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c)) dy dx}_{=:VI}
\end{aligned}$$

With the properties of the Lagrange interpolation and with estimate (43) we get:

$$\begin{aligned}
\text{I} & \leq \sum_{T \in \mathcal{T}_H} \|f\|_{L^2(T)} \|\mathcal{L}_H(z^c)(x) - z^c\|_{L^2(T)} \leq C_{\mathcal{L}_H} \left( \sum_{T \in \mathcal{T}_H} H_T^4 \|f\|_{L^2(T)}^2 \right)^{\frac{1}{2}} \|z^c\|_{H^2(\Omega)} \\
& \leq C_1 \left( \sum_{T \in \mathcal{T}_H} H_T^4 \|f\|_{L^2(T)}^2 \right)^{\frac{1}{2}} \|u_{H,h}^S - u^c\|_{L^2(\Omega)}.
\end{aligned}$$

For  $\text{II}_1$  we obtain:

$$\begin{aligned}
& \int_{\Omega} \int_Y (A^{\epsilon, \delta} - A_h^{\epsilon, \delta})(x, y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y)) \cdot (w^{\epsilon, \delta}(y) \nabla_x z^c(x)) dy dx \\
& \leq \left( \sum_{T \in \mathcal{T}_H} \left\| \int_Y w^{\epsilon, \delta}(y) (A^{\epsilon, \delta} - A_h^{\epsilon, \delta})(\cdot, y, \nabla_x u_{H,h}^S(x_T) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x_T, y)) dy \right\|_{L^2(T)}^2 \right)^{\frac{1}{2}} |z^c|_{H^1(\Omega)},
\end{aligned}$$

where  $|z^c|_{H^1(\Omega)}$  can be controlled by the error  $\|u_{H,h}^S - u^c\|_{L^2(\Omega)}$ .  $\text{II}_2$  is traded in the same way. For III, we have:

$$\begin{aligned}
& \int_{\Omega} \int_Y A_h^{\epsilon, \delta}(x, y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y)) \cdot w^{\epsilon, \delta}(y) \nabla_x (z^c - \mathcal{L}_H(z^c))(x) dy dx \\
& = \sum_{T \in \mathcal{T}_H} \int_{\partial T} \int_Y (w^{\epsilon, \delta}(y) A_h^{\epsilon, \delta}(x, y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y)) \cdot n_T(x)) (z^c - \mathcal{L}_H(z^c))(x) dy d\sigma(x) \\
& \leq \sum_{E \in \Gamma(\mathcal{T}_H)} \left\| \int_Y [w^{\epsilon, \delta}(y) A_h^{\epsilon, \delta}(\cdot, y, \nabla_x u_{H,h}^S + \nabla_y \mathcal{Q}_h(u_{H,h}^S))(\cdot, y)]_E dy \right\|_{L^2(E)} \|z^c - \mathcal{L}_H(z^c)\|_{L^2(E)} \\
& \leq C_{\mathcal{L}_H} \left( \sum_{E \in \Gamma(\mathcal{T}_H)} \eta_E^{res}(u_{H,h}^S) \right)^{\frac{1}{2}} \cdot C_{\Gamma} |z^c|_{H^2(\Omega)} \\
& \leq C_4 \left( \sum_{E \in \Gamma(\mathcal{T}_H)} \eta_E^{res}(u_{H,h}^S) \right)^{\frac{1}{2}} \cdot \|u_{H,h}^S - u^c\|_{L^2(\Omega)}.
\end{aligned}$$

Here,  $C_{\Gamma}$  denotes the maximum number of edges per simplex  $T \in \mathcal{T}_H$ . Similarly we treat IV:

$$\text{IV} \leq C \left( \sum_{T \in \mathcal{T}_H} \sum_{E_Y \in \Gamma(\mathcal{T}_h)/\sim_Y} h_{E_Y}^3 \|[A_h^{\epsilon, \delta}(\cdot, \cdot, \nabla_x u_{H,h}^S + \nabla_y \mathcal{Q}_h(u_{H,h}^S))]_{E_Y}\|_{L^2(T \times E_Y)}^2 \right)^2 |z^f|_{L^2(\Omega, H^2(Y))}.$$

In order to estimate V, we can proceed like in Lemma 3.8 to obtain:

$$\|\tilde{\mathcal{Q}}_h(\mathcal{L}_H(z^c))\|_{L^2(\Omega, H^1(\frac{\epsilon}{8}Y))} \leq C \|\mathcal{L}_H(z^c)\|_{H^1(\Omega)} \leq C \|z^c\|_{H^1(\Omega)} \leq C \|u_{H,h}^S - u^c\|_{L^2(\Omega)}.$$

With this inequality, we get the remaining estimate for  $V$  in a straightforward way.

For simplification, we now abbreviate:

$$g_{H,h}(x, y) := \frac{\delta^d}{\epsilon^d} A_h^{\epsilon, \delta} (x, y, \nabla_x u_{H,h}^S(x) + \nabla_y \mathcal{Q}_h(u_{H,h}^S)(x, y))$$

This yields the following for VI:

$$\begin{aligned} & \int_{\Omega} \int_{\frac{\delta}{8}Y} g_{H,h}(x, y) \cdot \nabla_y (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c)) dy dx \\ &= \sum_{T \in \mathcal{T}_H} \sum_{K \in \mathcal{T}_h^{\epsilon, \delta}} \int_{T \times \partial K} (g_{H,h}(x, y) \cdot n_K(y)) (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, y) d\sigma(y) dx \\ &= \sum_{T \in \mathcal{T}_H} \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{inn}} \int_T \int_{E_Y} [g_{H,h}(x, \cdot)]_{E_Y}(y) (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, y) d\sigma(y) dx \\ & \quad + \sum_{T \in \mathcal{T}_H} \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{out}} \int_T \int_{E_Y} (g_{H,h}(x, y) \cdot n_{\frac{\delta}{8}Y}(y)) (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, y) d\sigma(y) dx \\ &\leq \sum_{T \in \mathcal{T}_H} \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{inn}} \int_T \| [g_{H,h}(x, \cdot)]_{E_Y} \|_{L^2(E_Y)} \| (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, \cdot) \|_{L^2(E_Y)} dx \\ & \quad + \sum_{T \in \mathcal{T}_H} \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{out}} \int_T \| g_{H,h}(x, \cdot) \cdot n_{\frac{\delta}{8}Y} \|_{L^2(E_Y)} \| (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, \cdot) \|_{L^2(E_Y)} dx \\ &\leq C_{Tr} \sum_{T \in \mathcal{T}_H} \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{inn}} \int_T \| [g_{H,h}(x, \cdot)]_{E_Y} \|_{L^2(E_Y)} \| (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, \cdot) \|_{H^1(\omega_{E_Y})} dx \\ & \quad + C_{Tr} \sum_{T \in \mathcal{T}_H} \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{out}} \int_T \| g_{H,h}(x, \cdot) \cdot n_{\frac{\delta}{8}Y} \|_{L^2(E_Y)} \| (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, \cdot) \|_{H^1(\omega_{E_Y} \cap \frac{\delta}{\epsilon}Y)} dx \\ &\leq C_{Tr} \sum_{T \in \mathcal{T}_H} \int_T \left( \left( \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{inn}} \| [g_{H,h}(x, \cdot)]_{E_Y} \|_{L^2(E_Y)}^2 + \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{out}} \| g_{H,h}(x, \cdot) \cdot n_{\frac{\delta}{8}Y} \|_{L^2(E_Y)}^2 \right)^{\frac{1}{2}} \right. \\ & \quad \cdot \left( \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{inn}} \| (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, \cdot) \|_{H^1(\omega_{E_Y})}^2 \right. \\ & \quad \left. \left. + \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{out}} \| (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, \cdot) \|_{H^1(\omega_{E_Y} \cap \frac{\delta}{\epsilon}Y)}^2 \right)^{\frac{1}{2}} \right) dx \\ &\leq C_{Tr} \sum_{T \in \mathcal{T}_H} \int_T \left( \left( \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{inn}} \| [g_{H,h}(x, \cdot)]_{E_Y} \|_{L^2(E_Y)}^2 + \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\epsilon, \delta})_{out}} \| g_{H,h}(x, \cdot) \cdot n_{\frac{\delta}{8}Y} \|_{L^2(E_Y)}^2 \right)^{\frac{1}{2}} \right. \\ & \quad \left. \cdot C \| (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x, \cdot) \|_{H^1(\frac{\delta}{8}Y)} \right) dx \end{aligned}$$

Since

$$\| (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c))(x_T, \cdot) \|_{H^1(\frac{\delta}{8}Y)}$$

$$= \left\| \left( \tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T \right) (\mathcal{L}_H(z^c)) \right\|_{H^1(\frac{\varepsilon}{3}Y)} \leq \left\| \tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T \right\|_{Op} |\nabla_x \mathcal{L}_H(z^c)|$$

we get:

$$\begin{aligned} & \int_{\Omega} \int_{\frac{\varepsilon}{3}Y} g_{H,h}(x, y) \cdot \nabla_y (\tilde{\mathcal{Q}}_h - \mathcal{Q}_h)(\mathcal{L}_H(z^c)) dy dx \\ & \leq \left( C_5 \left( \sum_{T \in \mathcal{T}_H} \int_T \left\| \tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T \right\|_{Op} \left( \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\varepsilon, \delta})^{inn}} \| [g_{H,h}(x, \cdot)]_{E_Y} \|_{L^2(E_Y)}^2 \right) dx \right)^{\frac{1}{2}} \right. \\ & \quad \left. + C_5 \left( \sum_{T \in \mathcal{T}_H} \int_T \left\| \tilde{\mathcal{K}}_h^T - \mathcal{K}_h^T \right\|_{Op} \left( \sum_{E_Y \in \Gamma(\mathcal{T}_h^{\varepsilon, \delta})^{out}} \| g_{H,h}(x, \cdot) \cdot n_{\frac{\varepsilon}{3}Y} \|_{L^2(E_Y)}^2 \right) dx \right)^{\frac{1}{2}} \right) \\ & \quad \cdot C \|\mathcal{L}_H\|_{L(H^1(\Omega), V_H)} |z^c|_{H^1(\Omega)}. \end{aligned}$$

Putting the estimates together, we obtain the final estimate (38).  $\square$

**Remark 9.** In the case that we only have  $L^\infty$ -regularity for  $A^\varepsilon$ , the Lagrange interpolation operator in the preceding proof of Theorem 4.5 must be substituted by the Clément interpolation operator. Then, we only have linear order of convergence (in  $H$  and  $h$ ), different constants but the indicators remain the same with the exception that we must substitute  $H_T^4$  by  $H_T^2$ ,  $H_E^3$  by  $H_E$  and  $h_{E_Y}^3$  by  $h_{E_Y}$ .

**Remark 10.** Due to Lemma 3.8 we are able to substitute the  $\mathcal{Q}_h(u_{H,h})$  parts in the a-posteriori estimate by terms only depending on  $u_{H,h}$ . This simplifies the computation of the indicators, but it corrupts the estimate extremely.

## 6 Numerical experiments

In this section we are concerned with testing various realizations of the heterogeneous multiscale finite element method for a nonlinear elliptic model problem in which the diffusion operator contains fast, non-periodic oscillations. We apply the HMM to solve this problem efficiently. The elliptic problem reads as follows:

**Model Problem.** Find  $u \in H^1(\Omega)$  with

$$\begin{aligned} -\nabla \cdot A(x, \nabla u(x)) &= \begin{cases} \frac{1}{10} & \text{if } x_2 \leq \frac{1}{10} \\ 1 & \text{else.} \end{cases} \\ u &= 0 \text{ on } \partial\Omega. \end{aligned}$$

Here, the domain  $\Omega \subset [0, 1] \times [0, 1.2]$  is illustrated in Figure 2 and the (rapidly oscillating) nonlinear diffusion operator  $A$  is given by

$$A(x, \xi) := c(x_1, x_2) \begin{pmatrix} \xi_1 + \frac{1}{3}\xi_1^3 \\ \xi_2 + \frac{1}{3}\xi_2^3 \end{pmatrix},$$

where

$$c(x_1, x_2) := \begin{cases} 4 + \frac{18}{5}\sin(40\pi\sqrt{|2x_1|})\sin(90\pi x_2^2) & \text{if } x_2 \leq 0.3 \\ (3 - \frac{10x_2}{3}) \cdot \left( 1 + \frac{9}{10}\sin(40\pi\sqrt{|2x_1|})\sin(90\pi x_2^2) \right) & \text{if } 0.3 < x_2 < 0.6 \\ 1 + \frac{9}{10}\sin(40\pi\sqrt{|2x_1|})\sin(90\pi x_2^2) & \text{if } x_2 \geq 0.6. \end{cases} \quad (46)$$

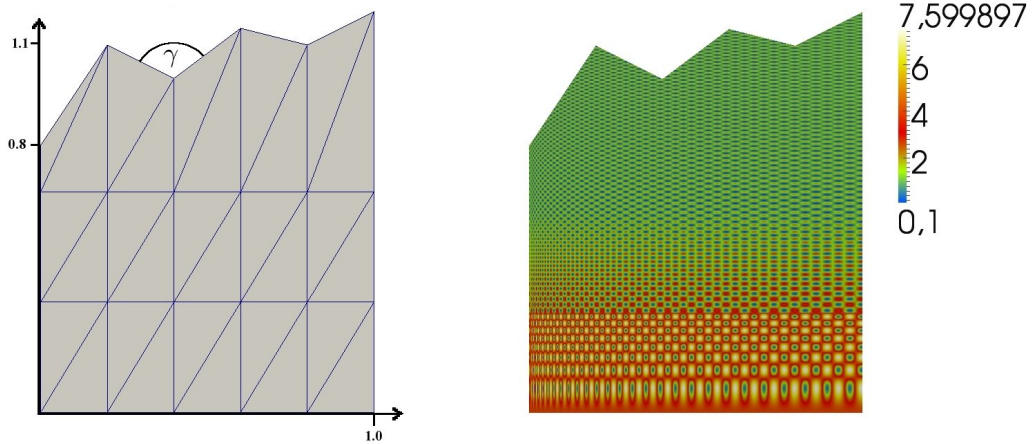


Figure 2: In the left figure we can see the computational domain  $\Omega$ , as well as a corresponding initial triangulation (the coarsest macro grid).  $\gamma \approx 116.57^\circ$  is the smallest angle of a re-entrant corner. In the right figure we display the diffusion coefficient  $c$  (given by (46)) which rapidly takes values between 0.1 and 7.6.

On the left hand side of Figure 2, we can see a plot of  $c$  on  $\Omega$ . Note that we do not have full  $H^2$ -regularity for the solution of this problem because  $\Omega$  is not a convex domain.

This model problem involves a heterogeneous microstructure which makes it impossible to apply standard homogenization techniques. The wavelengths of the oscillations are decreasing the closer they come to the axis  $x = 0$  and they are increasing the closer they come to the axes  $x = 1$  and  $y = 0$ . Furthermore, there is an additional macroscopic contribution. This implies that  $\epsilon$  is only an abstract parameter which does not explicitly occur in our model problem.

In the subsequent realizations of the HMM  $\epsilon$  occurs only as the size of the cells  $Y_{T,\epsilon}$  over which we average the reconstructions. Here, various reasonable choices for  $\epsilon$  are possible. In the following, increasing  $\epsilon$  does not mean that we change the model problem, it simply means that we change the realization of the HMM.

We also want to note that the microstructure in the problem above is still quite coarse in comparison to real applications. We did not use a finer structure since we need a highly accurate reference for the exact solution  $u$ . Due to the heterogeneity, we had to perform an extremely expensive, fine-scale FEM computation which led to a nonlinear algebraic system of equations with almost 8 million unknowns. In the following we just refer to  $u$  as the exact solution. It is plotted in Figure 3.

The model problem above and the corresponding results below can be seen as an example for a large set of other tests that we carried out and which all led to similar outcomes. For further numerical experiments and details on the implementation of the method, we refer to [22].

## 6.1 HMM approximations for different values of $\epsilon$ and $\delta$

In this first subsection we are concerned with various realizations of the HMM. We compare TFR-HMM and HMM, the method with oversampling and the method without oversampling

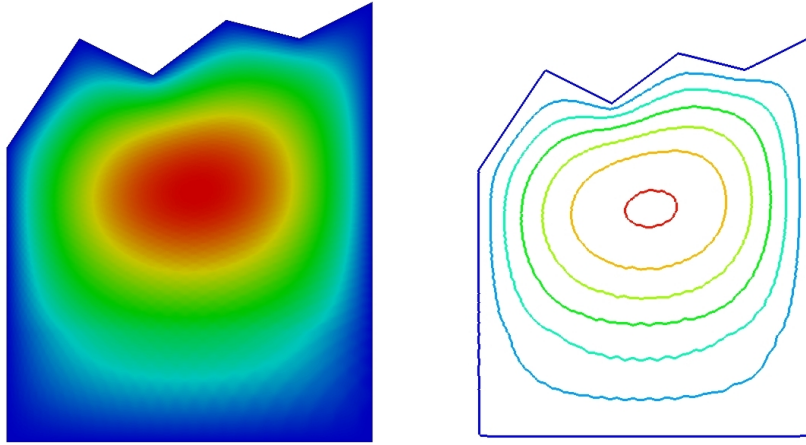


Figure 3: In this figure we can see the exact solution, determined by means of a highly expensive fine-scale computation (left: continuous plot, right: plot of isolines). The color shading is from blue (0.0) to red (0.06). We can still see a very slight micro-structural behavior.

and we have a look at the results for different values of  $\epsilon$  and  $\delta$ .

From now on, the global error indicators  $\xi(u_{H,h})$  and  $\xi(u_{H,h}^S)$  are defined according to the right hand sides in (37) and (38) respectively where the formal  $\mathcal{O}(\delta)$ -dependency is ignored. The error function  $e^{H,h}$  is given by  $e^{H,h} := u_{H,h} - u$ . For  $m \in \mathbb{N}_+$  we also define the experimental order of convergence (EOC) of  $\|e^{H,h}\|_{L^2(\Omega)}$  in  $H$  by

$$\text{EOC}_{((mH,mh) \rightarrow (H,h))}(\|e^{H,h}\|_{L^2(\Omega)}) := \frac{\log\left(\frac{\|e^{m \cdot H, m \cdot h}\|_{L^2(\Omega)}}{\|e^{H,h}\|_{L^2(\Omega)}}\right)}{\log(m)}. \quad (47)$$

Analogously we define the *EOC* with respect to the estimated error  $\xi(u_{H,h})$ . In the following we just refer to *EOC* if it is clear from the context what we mean.

Now, we start to discuss our results. The HMM errors for different values of  $H$ ,  $h$ ,  $\epsilon$  and  $\delta$  are given in Table 1. In the left hand side we fix the macro-grid with a resolution of  $H = 2^{-3}$ . The micro-structure is resolved by means of combinations of  $h$  and  $\delta$ . Note that the effective local resolution is equal to  $\delta \cdot h$ . We see that it seems to be better to choose  $\delta$  not too large, but using small values of  $h$  instead. This claim is emphasized by the results stated in the right list of Table 1, where we also decrease  $H$ . Concerning the balance between error size and computational complexity, the best results are obtained for choosing  $\epsilon = 0.05$  and  $\delta = 0.1$ . This is a bit surprising since the cells become so small in this scenario, that there exists a subregion of  $\Omega$  where the size of a cell falls below the wavelength of an oscillation. However, Table 1 suggests that it is more reasonable to choose a small  $\delta$  and to solve the cell problems accurately. Then we expect good approximations and we save computational demand.

In Table 2 we quantify the convergence behaviour. If we only compare the first three results, we see an average experimental order of convergence of 1.8. This is essentially what we expected,

$H$	$h$	$\epsilon$	$\delta$	$\ e^{H,h}\ _{L^2(\Omega)}$
$2^{-3}$	$2^{-6}$	0.15	0.3	0.00227901
$2^{-3}$	$2^{-5}$	0.05	0.1	0.00166077
$2^{-3}$	$2^{-7}$	0.15	0.3	0.00129459
$2^{-3}$	$2^{-7}$	0.1	0.2	0.00102948
$2^{-3}$	$2^{-8}$	0.15	0.3	0.000970714

$H$	$h$	$\epsilon$	$\delta$	$\ e^{H,h}\ _{L^2(\Omega)}$
$2^{-2}$	$2^{-5}$	0.15	0.3	0.00521319
$2^{-2}$	$2^{-5}$	0.1	0.2	0.00471942
$2^{-2}$	$2^{-6}$	0.15	0.3	0.00407923
$2^{-3}$	$2^{-5}$	0.05	0.1	0.00166077
$2^{-3}$	$2^{-8}$	0.15	0.3	0.000970714
$2^{-4}$	$2^{-6}$	0.05	0.1	0.000433359

Table 1: *Left tabular: HMM error for a fixed macro grid and increasing resolution of the micro-structure. Right tabular: HMM error for various combinations of  $H$ ,  $h$ ,  $\epsilon$  and  $\delta$ .*

since we do not have  $H^2$ -regularity for the solution. More precisely, one can show [21] that the order of convergence is  $(1 + \frac{1}{r})$ , where  $r = \frac{360-\gamma}{180}$  and  $\gamma$  denotes the smallest angle of a re-entrant corner. In our model problem we have  $\gamma \approx 116.57^\circ$ , which yields a theoretical order of convergence of approximately 1.74. For more details on this topic we refer to the book of Grisvard [21].

After the first three results, the EOC is breaking down and we obtain  $\text{EOC}_{((2^{-4}, 2^{-6}) \rightarrow (2^{-5}, 2^{-7}))} = 1.008$ . This is due to the fact that the HMM is reaching maximal accuracy. The exact solution still contains very small oscillations which can be observed in Figure 3. These oscillations cannot be captured by the HMM since the method only converges to a homogenized solution  $u^c$ . The remaining error  $\|u - u^c\|_{L^2(\Omega)}$  is around 0.0002. We can not fall below this value. Nevertheless, we can see by Figure 4 that this accuracy is completely sufficient and that we have a very good matching of the corresponding isolines. Qualitatively, we obtain a perfect approximation by the HMM.

$H$	$h$	$\epsilon$	$\delta$	$\ e^{H,h}\ _{L^2(\Omega)}$
$2^{-2}$	$2^{-4}$	0.05	0.1	0.00503086
$2^{-3}$	$2^{-5}$	0.05	0.1	0.00166077
$2^{-4}$	$2^{-6}$	0.05	0.1	0.000433359
$2^{-5}$	$2^{-6}$	0.05	0.1	0.000301772
$2^{-5}$	$2^{-7}$	0.05	0.1	0.000215419

Table 2: *In this table we can see a listing of  $L^2$ -errors between the exact solution and HMM approximations for decreasing macro- and micro-mesh sizes. In  $(H, h)$ , we observe an average order of convergence of 1.8 but up to a limit accuracy which cannot be exceeded. This is due to the fact that the small remaining oscillations of  $u$  can not be captured by the locally averaged  $u_{H,h}$ . See also Figure 4 for comparison.*

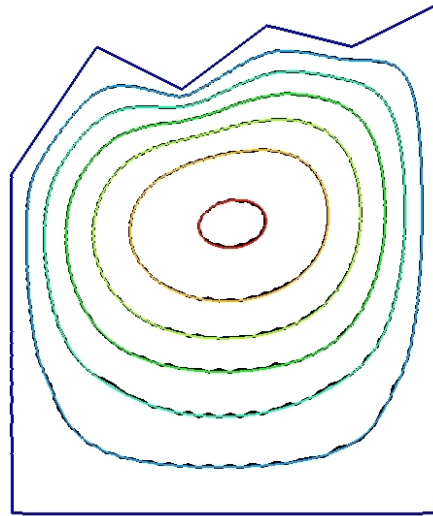


Figure 4: *Comparison of the isolines of the exact solution (black) and the isolines of the HMM approximation for  $(H, h) = (2^{-5}, 2^{-7})$  (colored).*

$H$	$h$	$\epsilon$	$\delta$	$\ e^{H,h}\ _{L^2(\Omega)}$	$\ e_{TFR}^{H,h}\ _{L^2(\Omega)}$	$\xi(u_{H,h})$	$\xi(u_{H,h}^S)$
$2^{-2}$	$2^{-6}$	0.15	0.3	0.00407923	0.00405277	0.0559879	0.43562
$2^{-2}$	$2^{-7}$	0.15	0.3	0.00321042	0.00320094	0.0547529	0.363264
$2^{-3}$	$2^{-6}$	0.15	0.3	0.00227901	0.00227938	0.0162236	0.408395
$2^{-3}$	$2^{-7}$	0.15	0.3	0.00129459	0.00131995	0.0143636	0.331023
$2^{-2}$	$2^{-4}$	0.05	0.1	0.00501086	0.00502993	0.0607383	0.298974
$2^{-3}$	$2^{-5}$	0.05	0.1	0.00166077	0.00167581	0.0164696	0.222514
$2^{-4}$	$2^{-6}$	0.05	0.1	0.000433359	0.000430596	0.00442216	0.179064

Table 3: In this table we list a number of HMM errors and corresponding TFR-HMM error for the same values of  $H, h, \epsilon$  and  $\delta$ . We observe that these two methods yield almost identical approximations for  $u$ .

In Table 3 we compare the approximations obtained by HMM and TFR-HMM. Before discussing the results, we want to point out that the TFR-HMM was significantly slower than the HMM. For the same values of  $H, h, \delta$  and  $\epsilon$ , the TFR-HMM needs up to twice of the CPU time of the HMM. Moreover, we need to save far more information, namely the reconstructions of the base functions. If we purely compare the quality of the approximations, we observe that there is virtually no difference between  $u_{H,h}$  and  $u_{H,h}^S$ . The errors are basically the same. This is what we expect if we consider that the contribution of TFR should be close to zero. For completeness, Table 3 also contains a listing of the corresponding estimated errors. For the HMM, the behaviour of the estimated error is very nice. Its convergence is similar to the convergence of the error itself and we can say that the quotient between  $\xi(u_{H,h})$  and  $\|e^{H,h}\|_{L^2(\Omega)}$  is of order 10. Applications of this error estimator are given in Section 6.2. If we have a look at the estimated error for the TFR-HMM, we see that it is decreasing, but only very slowly. It is clearly dominated by the contributions of  $\bar{\eta}_T^{\text{TFR}}$ , which makes it an improper indicator for a tolerance in adaptive mesh refinement algorithms. Even though HMM and HMM-TFR seem to produce similar results, we can summarize that there is no reason for using a TFR-HMM in a nonlinear setting.

In Table 4, we state the results of the HMM with and without oversampling. Two results are comparable if the macro grid was identical and if  $\delta_1 h_1 = \delta_2 h_2$  (i.e. if they have the same resolution of the micro structure). A little surprisingly we observe that approximations produced by the HMM with oversampling are only slightly better than the results of the HMM without oversampling. A comparison of the isolines in Figure 5 displays that a difference is hardly perceptible. However, the reasons for this seems to lie in the periodic boundary condition for the cell problems. Even though it is a wrong boundary condition, it is very flexible and can easily adapt to the situation. If we use for instance a Dirichlet boundary condition for the cell problems, the impact will be far larger due to strong boundary layer effects. In this case, the lack of oversampling will highly corrupt the approximation.

$H$	$h$	$\ u_{H,h} - u\ _{L^2(\Omega)}$	$\ v_{H,2h} - u\ _{L^2(\Omega)}$
$2^{-2}$	$2^{-4}$	0.00503086	0.00517765
$2^{-3}$	$2^{-5}$	0.00166077	0.00177228
$2^{-4}$	$2^{-6}$	0.000433359	0.000472636
$2^{-5}$	$2^{-7}$	0.000215419	0.00022142

Table 4: In this table,  $u_{H,h}$  denotes the HMM approximation for  $\epsilon = 0.05$  and  $\delta = 0.1$ , whereas  $v_{H,2h}$  denotes a corresponding HMM approximation without oversampling, i.e.  $\epsilon = \delta = 0.05$ . In the table above we compare the quality of both approximations. Note, that we need a resolution of  $2h$  for the HMM without oversampling to get a fair comparison. In the table we only refer to  $h$  as the resolution  $Y$  for the HMM with oversampling. We observe that we gain only small advantages from using oversampling.

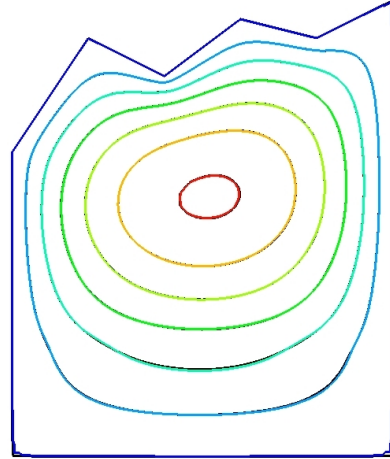


Figure 5: Comparison of the iso-lines of the HMM approximations with and without oversampling for  $H = 2^{-5}$  (almost identical).

## 6.2 Adaptive strategies

In this section we want use the a-posteriori error estimate stated in Theorem 4.5 to construct two algorithms for an adaptive mesh refinement. For this purpose, we define the local error indicators by:

$$\xi_T(u_{H,h}) \quad (48)$$

$$:= \left( H_T^4 \|f\|_{L^2(\Omega)}^2 + \eta_T^{app}(u_{H,h}) + \bar{\eta}_T^{app}(u_{H,h}) + \sum_{\substack{E \in \Gamma(T_H) \\ E \subset \bar{T}}} \left( \frac{1}{2} \eta_E^{res}(u_{H,h}) \right) + \bar{\eta}_T^{res}(u_{H,h}) \right)^{\frac{1}{2}},$$

the average indicator  $\xi^{av}$  by

$$\xi^{av}(u_{H,h}) := \frac{1}{\#\mathcal{T}_H} \sum_{T \in \mathcal{T}_H} \xi_T(u_{H,h}) \quad (49)$$

and the maximum  $\xi^{max}(u_{H,h})$  of the indicators by

$$\xi^{max}(u_{H,h}) := \max_{T \in \mathcal{T}_H} \xi_T(u_{H,h}). \quad (50)$$

In this section,  $\lfloor \cdot \rfloor$  denotes the floor function, which determines the largest previous integer of a real number.

Given a certain error tolerance  $TOL$ , the goal is to quickly determine a corresponding HMM solution so that we fall below  $TOL$ . As far as possible, uniform refinements of the grid shall be avoided to prevent unnecessarily expensive computations. Regions where local refinements are

---

Algorithm 1: adaptiveRefine(  $TOL, \mathcal{T}_H, \mathcal{T}_h$  )

---

```

while  $\xi(u_{H,h}) > TOL$  do
  Compute  $u_{H,h}$  with  $\mathcal{T}_H$  and  $\mathcal{T}_h$ .
  Compute  $\xi(u_{H,h})$ .
  if  $\xi(u_{H,h}) < TOL$  then
    break.
  end
  foreach  $T \in \mathcal{T}_H$  do
    compute  $\xi_T(u_{H,h})$ 
  end
  Compute  $\xi^{av}(u_{H,h})$ .
  foreach  $T \in \mathcal{T}_H$  do
    if  $\xi_T(u_{H,h}) \geq \xi^{av}(u_{H,h})$  then
      mark  $T$  for one refinement.
    else
      do nothing.
    end
  end
  Refine grid.
end

```

---

required are determined by means of the local error indicators  $\xi_T(u_{H,h})$  above. We present the results of two adaptive algorithms.

The first algorithm describes an equal distribution strategy. Here, we solve the HMM for a certain macro grid, where the micro grid remains fixed ( $\epsilon = 0.05$ ,  $\delta = 0.1$ ,  $h = 2^{-6}$ ). This is a reasonable decision with regard to the results from the preceding subsection. After we have  $u_{H,h}$ , we determine the corresponding estimated error. If it is less equal than a certain tolerance the algorithm aborts. If it is greater than  $TOL$  we determine all elements of  $\mathcal{T}_H$  on which the local indicator is larger than the indicator average. These elements are marked for one refinement. After refining the grid, we start again with the algorithm. In Algorithm 1, this strategy is described in detail. Here,  $TOL$  denotes the given error tolerance. The local error indicators are defined by  $\xi_T(u_{H,h})$  and the average indicator by  $\xi^{av}(u_{H,h})$ .

The results of a HMM using Algorithm 1 are given on the left hand side of Table 5. We observe a significant reduction of the error in every cycle of the algorithm till we reach a very high final accuracy. However, we also see that the computation of the HMM approximation is initialized seven times before we find a suitable, adaptively refined grid for our final computation. Since we want to determine the final grid as fast as possible, we suggest a second algorithm for improving the strategy. It is based upon the idea that we start with a very coarse macro grid ( $H = 2^{-1}$ ) to perform a very cheap initial computation. Then we calculate  $\xi(u_{H,h})$ . Since we expect the indicator to show a quadratic order of convergence, we can use the definition of the EOC to guess the size of  $H$  that is required to fall below a desired tolerance. Heuristically,  $\tilde{H}$  with

$$\tilde{H} \leq \left( \frac{\xi(u_{H,h})}{TOL} \right)^{-\frac{1}{2}} \cdot H$$

would be good candidate so that  $\xi(u_{\tilde{H},h}) < TOL$ . Formally, this is only a strategy to determine

---

Algorithm 2: adaptiveRefine2( TOL,  $\mathcal{T}_H$ ,  $\mathcal{T}_h$  )

---

Compute  $u_{H,h}$  with  $\mathcal{T}_H$  and  $\mathcal{T}_h$ . Compute  $\xi(u_{H,h})$ .  
**if**  $\xi(u_{H,h}) < \text{TOL}$  **then**  
     break.  
**end**  
 Compute  $m := \lfloor \frac{1}{2} \sqrt{\frac{\xi(u_{H,h})}{\text{TOL}}} \rfloor$ .  
**foreach**  $T \in \mathcal{T}_H$  **do**  
     mark  $T$  for  $m$  refinements.  
**end**  
 Refine grid.  
**while**  $\xi(u_{H,h}) > \text{TOL}$  **do**  
     Compute  $u_{H,h}$  with  $\mathcal{T}_H$  and  $\mathcal{T}_h$ .  
     Compute  $\xi(u_{H,h})$ .  
     **if**  $\xi(u_{H,h}) < \text{TOL}$  **then**  
         break.  
     **end**  
     **foreach**  $T \in \mathcal{T}_H$  **do**  
         compute  $\xi_T(u_{H,h})$   
     **end**  
     Compute  $\xi^{av}(u_{H,h})$ .  
     **foreach**  $T \in \mathcal{T}_H$  **do**  
         **if**  $\xi_T(u_{H,h}) \geq \xi^{av}(u_{H,h})$  **then**  
             **if**  $\xi_T(u_{H,h}) \geq \frac{\xi^{max}(u_{H,h}) - \xi^{av}(u_{H,h})}{2}$  **then**  
                 mark  $T$  for two refinements.  
             **else**  
                 mark  $T$  for one refinement.  
             **end**  
         **end**  
     **end**  
     Refine grid.  
**end**

---

a required number of uniform refinement steps, but we can also use it for an adaptive algorithm: here we go half the way to  $TOL$  with a uniform refinement (first step), then we have a sufficiently fine macro-grid to apply adaptive mesh refinement (remaining steps till the algorithm aborts). In all these subsequent steps, we subdivide the grid into three regions, which are distinguished according to no refinement, one refinement and two refinements. Details are given in Algorithm 2.

By Table 5 we see that the number of cycles can be halved by using Algorithm 2 instead of Algorithm 1. Indeed, Algorithm 2 also needs only half of the CPU time of Algorithm 1. Nevertheless, we have to admit that the first algorithm reaches a higher accuracy if we set  $TOL = 0.005$ . Both algorithms seem to work out fine. In Figure 6 we depict a highly refined grid, produced by the second adaptive algorithm where the input was a very small error tolerance. This grid is given exemplarily for the behaviour of the refinement procedure, which specifically takes place around the corners at the top of the domain  $\Omega$  where we are dealing with large gradients.

Cycle of Algo.	$\ e^{H,h}\ _{L^2(\Omega)}$	$\xi(u_{H,h})$
1	0.00862357	0.244032
2	0.00614412	0.149392
3	0.00263085	0.0494256
4	0.00175792	0.0267087
5	0.000908581	0.0148526
6	0.000682865	0.0105973
7	0.000394762	0.00525074
8	0.000278383	0.00397427

Cycle of Algo.	$\ e^{H,h}\ _{L^2(\Omega)}$	$\xi(u_{H,h})$
1	0.00862357	0.244032
2	0.000898418	0.0120081
3	0.000642008	0.00717823
4	0.000449121	0.00476699

Table 5: The HMM-computations in both tables refer to a micro-grid size of  $h = 2^{-6}$  and a cell size of  $\delta = 2\epsilon = 0.1$ . The (coarse) initial macro-grid triangulation is illustrated in Figure 2. The tolerance for the estimated error is set to  $TOL = 0.005$ . In left table we can see the results of Algorithm 1 which requires 8 cycles. In the right table we can see the results of Algorithm 2 which only requires 4 cycles (where first cycle involves the initial uniform refinement). Comparing the CPU times, we observed that Algorithm 2 was twice as fast as Algorithm 1.

## 7 Conclusion

In this work we introduced a heterogeneous multiscale finite element method for monotone elliptic operators. In order to derive a corresponding general a-posteriori error estimate, we identified the analytical limit problem of the HMM. On the basis of this identification, the estimate was achieved by means of an error identity deduced from a suitable dual problem. The applicability of the method and associated a-posteriori error estimate was verified in numerical experiments in a heterogeneous setting. We observed a nice convergence behaviour for the  $L^2$ -error, as well as for the estimated  $L^2$ -error. Decomposing the global error indicator into local contributions, we were able to formulate adaptive algorithms. The usability of the algorithms was demonstrated in additional numerical computations.

## References

- [1] A. Abdulle. The finite element heterogeneous multiscale method: a computational strategy for multiscale PDEs. In *Multiple scales problems in biomathematics, mechanics, physics and numerics*, volume 31 of *GAKUTO Internat. Ser. Math. Sci. Appl.*, pages 133–181. Gakkōtoshō, Tokyo, 2009.
- [2] Assyr Abdulle. Multiscale methods for advection-diffusion problems. *Discrete Contin. Dyn. Syst.*, (suppl.):11–21, 2005.
- [3] Assyr Abdulle. On a priori error analysis of fully discrete heterogeneous multiscale FEM. *Multiscale Model. Simul.*, 4(2):447–459 (electronic), 2005.
- [4] Assyr Abdulle and Weinan E. Finite difference heterogeneous multi-scale method for homogenization problems. *J. Comput. Phys.*, 191(1):18–39, 2003.
- [5] Assyr Abdulle and Bjorn Engquist. Finite element heterogeneous multiscale methods with near optimal computational complexity. *Multiscale Model. Simul.*, 6(4):1059–1084, 2007.

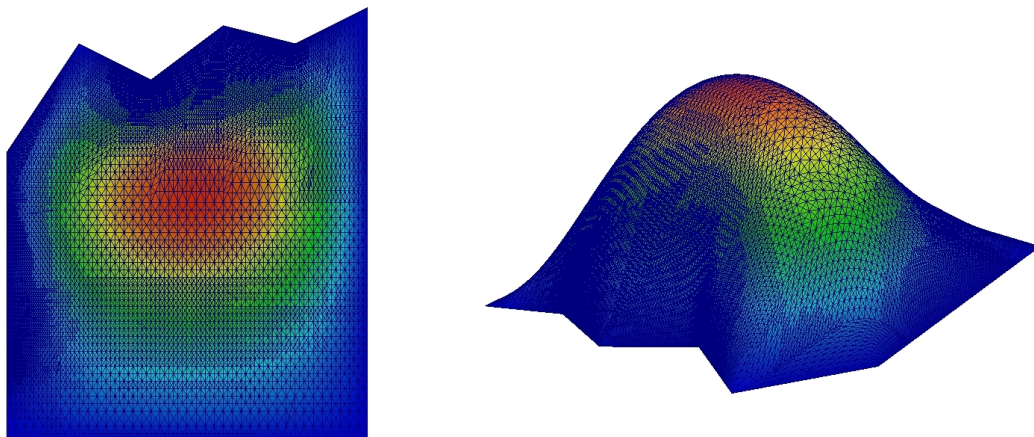


Figure 6: In these Figures we see that the grid is especially refined around the corners at the top of the domain  $\Omega$ . The right figure is a 3D plot of the left figure to emphasize large gradients.

- [6] Assyr Abdulle and Achim Nonnenmacher. A posteriori error analysis of the heterogeneous multiscale method for homogenization problems. *C. R. Math. Acad. Sci. Paris*, 347(17-18):1081–1086, 2009.
- [7] Assyr Abdulle and Christoph Schwab. Heterogeneous multiscale FEM for diffusion problems on rough surfaces. *Multiscale Model. Simul.*, 3(1):195–220 (electronic), 2004/05.
- [8] Grégoire Allaire. Homogenization and two-scale convergence. *SIAM J. Math. Anal.*, 23(6):1482–1518, 1992.
- [9] Hans Wilhelm Alt. <sup>4</sup>*Lineare Funktionalanalysis*. Springer, Berlin, 2002.
- [10] Todd Arbogast, Gergina Pencheva, Mary F. Wheeler, and Ivan Yotov. A multiscale mortar mixed finite element method. *Multiscale Model. Simul.*, 6(1):319–346 (electronic), 2007.
- [11] C.-C. Chu, I. G. Graham, and T.-Y. Hou. A new multiscale finite element method for high-contrast elliptic interface problems. *Math. Comp.*, 79(272):1915–1955, 2010.
- [12] Philippe G. Ciarlet. *The finite element method for elliptic problems*. North-Holland Publishing Co., Amsterdam, 1978. Studies in Mathematics and its Applications, Vol. 4.
- [13] Doina Cioranescu and Patrizia Donato. *An Introduction to Homogenization*. Oxford Lecture Series in Mathematics and its Applications, 1999.
- [14] Weinan E and Bjorn Engquist. The heterogeneous multiscale methods. *Commun. Math. Sci.*, 1(1):87–132, 2003.
- [15] Weinan E and Bjorn Engquist. Multiscale modeling and computation. *Notices Amer. Math. Soc.*, 50(9):1062–1070, 2003.
- [16] Weinan E and Björn Engquist. The heterogeneous multi-scale method for homogenization problems. In *Multiscale methods in science and engineering*, volume 44 of *Lect. Notes Comput. Sci. Eng.*, pages 89–110. Springer, Berlin, 2005.

- [17] Weinan E, Pingbing Ming, and Pingwen Zhang. Analysis of the heterogeneous multiscale method for elliptic homogenization problems. *J. Amer. Math. Soc.*, 18(1):121–156 (electronic), 2005.
- [18] Y. Efendiev and T. Hou. Multiscale finite element methods for porous media flows and their applications. *Appl. Numer. Math.*, 57(5-7):577–596, 2007.
- [19] Y. Efendiev, T. Hou, and V. Ginting. Multiscale finite element methods for nonlinear problems and their applications. *Commun. Math. Sci.*, 2(4):553–589, 2004.
- [20] Yalchin Efendiev and Thomas Y. Hou. *Multiscale finite element methods*, volume 4 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New York, 2009. Theory and applications.
- [21] P. Grisvard. *Singularities in boundary value problems*, volume 22 of *Recherches en Mathématiques Appliquées [Research in Applied Mathematics]*. Masson, Paris, 1992.
- [22] P. Henning and M. Ohlberger. On the implementation of a heterogeneous multiscale finite element method for nonlinear elliptic problems. In *Submitted to the Proceedings of the DUNE-User Meeting 2010*, 2011.
- [23] Patrick Henning and Mario Ohlberger. A-posteriori error estimate for a heterogeneous multiscale finite element method for advection-diffusion problems with rapidly oscillating coefficients and large expected drift. *Preprint 09/09*, 2009.
- [24] Patrick Henning and Mario Ohlberger. The heterogeneous multiscale finite element method for elliptic homogenization problems in perforated domains. *Numer. Math.*, 113(4):601–629, 2009.
- [25] Patrick Henning and Mario Ohlberger. The heterogeneous multiscale finite element method for advection-diffusion problems with rapidly oscillating coefficients and large expected drift. *Netw. Heterog. Media*, 5(4):711–744, 2010.
- [26] Viet Ha Hoang and Christoph Schwab. High-dimensional finite elements for elliptic problems with multiple scales. *Multiscale Model. Simul.*, 3(1):168–194 (electronic), 2004/05.
- [27] Thomas Y. Hou and Xiao-Hui Wu. A multiscale finite element method for elliptic problems in composite materials and porous media. *J. Comput. Phys.*, 134(1):169–189, 1997.
- [28] Thomas Y. Hou, Xiao-Hui Wu, and Zhiqiang Cai. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. *Math. Comp.*, 68(227):913–943, 1999.
- [29] Thomas J. R. Hughes. Multiscale phenomena: Green’s functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods. *Comput. Methods Appl. Mech. Engrg.*, 127(1-4):387–401, 1995.
- [30] Thomas J. R. Hughes, Gonzalo R. Feijóo, Luca Mazzei, and Jean-Baptiste Quinicy. The variational multiscale method—a paradigm for computational mechanics. *Comput. Methods Appl. Mech. Engrg.*, 166(1-2):3–24, 1998.
- [31] Mats G. Larson and Axel Målqvist. Adaptive variational multiscale methods based on a posteriori error estimation: duality techniques for elliptic problems. In *Multiscale methods in science and engineering*, volume 44 of *Lect. Notes Comput. Sci. Eng.*, pages 181–193. Springer, Berlin, 2005.

- [32] Mats G. Larson and Axel Målqvist. Adaptive variational multiscale methods based on a posteriori error estimation: energy norm estimates for elliptic problems. *Comput. Methods Appl. Mech. Engrg.*, 196(21-24):2313–2324, 2007.
- [33] Mats G. Larson and Axel Målqvist. An adaptive variational multiscale method for convection-diffusion problems. *Comm. Numer. Methods Engrg.*, 25(1):65–79, 2009.
- [34] J. L. Lions, D. Lukkassen, L. E. Persson, and P. Wall. Reiterated homogenization of nonlinear monotone operators. *Chinese Ann. Math. Ser. B*, 22(1):1–12, 2001.
- [35] Dag Lukkassen, Gabriel Nguetseng, and Peter Wall. Two-scale convergence. *Int. J. Pure Appl. Math.*, 2(1):35–86, 2002.
- [36] A.-M. Matache. Sparse two-scale FEM for homogenization problems. In *Proceedings of the Fifth International Conference on Spectral and High Order Methods (ICOSAHOM-01) (Uppsala)*, volume 17, pages 659–669, 2002.
- [37] Ana-Maria Matache and Christoph Schwab. Two-scale FEM for homogenization problems. *M2AN Math. Model. Numer. Anal.*, 36(4):537–572, 2002.
- [38] Pingbing Ming and Pingwen Zhang. Analysis of the heterogeneous multiscale method for parabolic homogenization problems. *Math. Comp.*, 76(257):153–177 (electronic), 2007.
- [39] James Nolen, George Papanicolaou, and Olivier Pironneau. A framework for adaptive multiscale methods for elliptic problems. *Multiscale Model. Simul.*, 7(1):171–196, 2008.
- [40] J. Tinsley Oden and Kumar S. Vemaganti. Adaptive modeling of composite structures: Modeling error estimation. *Int. J. Comp. Civil Str. Engrg.*, 1:1–16, 2000.
- [41] J. Tinsley Oden and Kumar S. Vemaganti. Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials. I. Error estimates and adaptive algorithms. *J. Comput. Phys.*, 164(1):22–47, 2000.
- [42] Mario Ohlberger. A posteriori error estimates for the heterogeneous multiscale finite element method for elliptic homogenization problems. *Multiscale Model. Simul.*, 4(1):88–114 (electronic), 2005.
- [43] Michael Růžička. *Nichtlineare Funktionalanalysis*. Springer-Verlag Berlin Heidelberg New York, 2004.
- [44] Christoph Schwab and Ana-Maria Matache. Generalized FEM for homogenization problems. In *Multiscale and multiresolution methods*, volume 20 of *Lect. Notes Comput. Sci. Eng.*, pages 197–237. Springer, Berlin, 2002.
- [45] Kumar S. Vemaganti and J. Tinsley Oden. Estimation of local modeling error and goal-oriented adaptive modeling of heterogeneous materials. II. A computational environment for adaptive modeling of heterogeneous elastic solids. *Comput. Methods Appl. Mech. Engrg.*, 190(46-47):6089–6124, 2001.
- [46] Peter Wall. Some homogenization and corrector results for nonlinear monotone operators. *J. Nonlinear Math. Phys.*, 5(3):331–348, 1998.
- [47] Tarek I. Zohdi, J. Tinsley Oden, and Gregory J. Rodin. Hierarchical modeling of heterogeneous bodies. *Comput. Methods Appl. Mech. Engrg.*, 138(1-4):273–298, 1996.

Preprints  
"Angewandte Mathematik und Informatik"

- 06/08 - N E. Pekalska, B. Haasdonk: Kernel Quadratic Discriminant Analysis with Positive Definite and Indefinite Kernels
- 07/08 - S M. Meiners: Weighted Branching and a Pathwise Renewal Equation
- 08/08 - S M. Ebbers, M. Löwe: Torpid Mixing of the Swapping Chain on Some Simple Spin Glass Models
- 09/08 - I T. Ropinski, I. Viola, M. Biermann, F. Lindemann, R. Leißa, H. Hauser, K. Hinrichs: Multimodal Closeups for Medical Visualization
- 01/09 - I J. Mensmann, T. Ropinski, K. Hinrichs: An Evaluation of the CUDA Architecture for Volume Rendering
- 02/09 - N P. Henning, M. Ohlberger: Advection-diffusion problems with rapidly oscillating coefficients and large expected drift. Part 1: Homogenization – existence, uniqueness and regularity
- 03/09 - N P. Henning, M. Ohlberger: Advection-diffusion problems with rapidly oscillating coefficients and large expected drift. Part 2: The heterogeneous multiscale finite element method
- 04/09 - I J. Meyer-Spradow, T. Ropinski, J. Mensmann, K. Hinrichs: Rapid Prototyping of Volume Visualization in Collaboration with Domain Experts
- 05/09 - N K. Mikula, M. Ohlberger: A New Level Set Method for Motion in Normal Direction Based on a Forward-Backward Diffusion Formulation
- 06/09 - I T. Ropinski, S. Diepenbrock, S. Bruckner, K. Hinrichs, E. Gröller: Volumetric Texturing
- 07/09 - I J.-S. Pražni, J. Mensmann, T. Ropinski, K. Hinrichs: Shape-based Transfer Functions for Volume Visualization
- 08/09 - N A. Dedner, R. Klöforn, M. Nolte, M. Ohlberger: A generic interface for parallel and adaptive scientific computing: Abstraction principles and the DUNE-FEM module
- 09/09 - N P. Henning, M. Ohlberger: A-posteriori error estimate for a heterogeneous multiscale finite element method for advection-diffusion problems with rapidly oscillating coefficients and large expected drift
- 01/10 - N K. Mikula, M. Ohlberger: A New Inflow-Implicit/Outflow-Explicit Finite Volume Method for Solving Variable Velocity Advection Equations
- 02/10 - N M. Drohmann, B. Haasdonk, M. Ohlberger: Reduced Basis Approximation for Nonlinear Parametrized Evolution Equations based on Empirical Operator Interpolation
- 03/10 - N M. Ohlberger, K. Smetana: A new problem adapted hierarchical model reduction technique based on reduced basis methods and dimensional splitting
- 04/10 - I M. Steuwer, P. Kegel, S. Gorlatch: SkelCL – A Portable Multi-GPU Skeleton Library
- 01/11 - N P. Henning, M. Ohlberger: A-posteriori error estimation for a heterogeneous multiscale method for monotone operators and beyond a periodic setting