Reduced order modeling techniques for numerical homogenization methods applied to linear and nonlinear multiscale problems

THÈSE N^{O} 5922 (2013)

PRÉSENTÉE LE 20 SEPTEMBRE 2013

À LA FACULTÉ DES SCIENCES DE BASE

CHAIRE D'ANALYSE NUMÉRIQUE ET MATHÉMATIQUES COMPUTATIONNELLES

PROGRAMME DOCTORAL EN MATHÉMATIQUES

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Yun BAI

acceptée sur proposition du jury:

Prof. F. Eisenbrand, président du jury
Prof. A. Abdulle, directeur de thèse
Prof. S. Bartels, rapporteur
Prof. B. Engquist, rapporteur
Prof. A. Quarteroni, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2013

A wise man is never confused;
A virtuous man is never worried;
A courageous man is never afraid.

— Confucius

To my parents

Acknowledgements

My journey of the PhD years has come to an end and new adventure will soon start. Here I want to thank everyone who made this journey a valuable and unforgettable memory to me.

Specially, I would like to express my deepest respect and gratitude to my thesis advisor Prof. Assyr Abdulle for bringing me to this interesting research subject and guiding me with great insight and patience. The knowledge I gained from our numerous discussions on resolving research obstacles, improving my writing and presentation skills as well as planning future career, is such a precious asset that will be beneficial to my following life journey.

My acknowledgement goes to the members of the jury, Prof. Sören Bartels, Prof. Björn Engquist and Prof. Alfio Quarteroni for reading my work and for their comments. I further thank Prof. Friedrich Eisenbrand for presiding the jury.

Also, I am very grateful to my master advisor Prof. Xingye Yue for guiding my first attempt to do research and his kind encouragement and recommendation for my move to EPFL.

My gratitude goes to my colleagues at EPFL, Achim, Gilles, Ondrej, Orane, Martin, Adrian and Timothée for many inspiring discussions and building up such a friendly environment. The company of many friends in Lausanne makes my life very enjoyable and I would like to address special acknowledgement to Yuxuan Ji, Na Li, Yu Lin, Le Chen.

Furthermore, I want to thank my boyfriend for his love and support. Finally, I would like to express my greatest gratitude to my parents for all their love, care and effort.

Lausanne, September 2013 Yun Bai

Abstract

The characteristic of effective properties of physical processes in heterogeneous media is a basic modeling and computational problem for many applications. As standard numerical discretization of such multiscale problems (e.g. with classical finite element method (FEM)) is often computationally prohibitive, there is a need for a novel computational algorithm able to capture the effective behavior of the physical system without resolving the finest scale in the system on the whole computational domain.

In this thesis we propose and analyze a new class of numerical methods that combine the so-called finite element heterogeneous multiscale method (FE-HMM) with reduced order modeling techniques for linear and nonlinear multiscale problems.

In the first part of the thesis we generalize the FE-HMM to elliptic problems with an arbitrary number of well-separated scales. We provide a rigorous a priori error analysis of this method that generalizes previous work restricted to two-scale problems. In the second part of the thesis, we develop our new reduced order multiscale method that combines the FE-HMM with reduced basis (RB) method. This method, the reduced basis finite element heterogeneous multiscale method (RB-FE-HMM) provides a significant improvement in computational efficiency compared to the FE-HMM, specially for high dimensional problems or high order methods. A priori and a posteriori error analyses are derived for linear elliptic problems, as well as goal oriented adaptivity techniques. The RB-FE-HMM is then generalized to a class of nonlinear elliptic and parabolic problems. A priori error analysis and extensive computational results for nonlinear problems are also provided.

Keywords: Multiscale method, reduced basis, homogenization, finite element, adaptivity, goal oriented, nonlinear problems.

Résumé

La caractéristique des propriétés effectives d'un processus physique agissant dans un milieu hétérogène est un problème de modélisation et d'implémentation inhérent à beaucoup d'applications. Les méthodes de discrétisations numériques standards appliquées à de tels problèmes multi-échelles (par exemple la méthode des éléments finis classique (FEM)) avec un maillage très fin étant souvent d'un coût prohibitif, il est nécessaire de trouver un nouvel algorithme numérique capable de capturer le comportement macroscopique d'un système physique sans avoir à résoudre les micro-échelles du système sur l'ensemble du domaine discrétisé.

Dans cette thèse, nous proposons et analysons une nouvelle classe de méthodes numériques qui combinent la méthode d'éléments finis hétérogène multi-échelles (FE-HMM) avec des techniques de modélisation à ordres réduits appliquées à des problèmes multi-échelles linéaires et non-linéaires.

Dans la première partie de la thèse, nous généralisons la méthode FE-HMM aux problèmes elliptiques avec un nombre arbitraires d'échelles explicitement séparées. Nous donnons une analyse a priori de la vitress de convergence de la méthode, qui généralise des travaux antérieures pour les problèmes à double échelles.

Dans la seconde partie de la thèse, nous développons notre nouvelle méthode multi-échelles à ordres réduits qui combine FE-HMM avec la méthode des bases réduites (RB). Comparée à la méthode FE-HMM, la méthode d'éléments finis hétérogène multi-échelles à bases réduites (RB-FE-HMM) permet une amélioration significative en termes d'efficacité, notamment pour les problèmes à dimension élevée ou pour des éléments finis ordre élevé. L'analyse a priori et a posteriori de l'erreur de la méthode est dérivée pour les problèmes elliptiques linéaires ainsi que pour les techniques adaptives de type goal-oriented. La méthode RB-FE-HMM est ensuite généralisée à une classe de problèmes elliptiques et paraboliques non-linéaires. Une analyse a priori de l'erreur et de nombreux résultats numériques sont aussi présentés pour les problèmes non-linéaires.

Mots Clés : Méthode multi-échelles, bases réduites, homogénéisation, éléments finis, méthode adaptive, goal-oriented, problèmes non-linéaires.

Contents

List of Figures

List of Tables

List of notations

ε	the small scale in the multiscale partial diffierential equations (PDE).
Ω	the macro domain where the multiscale PDE is considered and $\varepsilon << \|\Omega\|$.
$\mathbb{R}$	the real number field.
Y	a unit cube in $\mathbb{R}^d$.
$\mathbf{e}_i$	the i-th canonical basis of $\mathbb{R}^d$.
$D^\alpha u$	For $\alpha = \{\alpha_1, \cdots, \alpha_d\} \in \mathbb{N}^d$, $D^\alpha u = \partial_1^{\alpha_1} \cdots \partial_d^{\alpha_d} u$, and denote $\|\alpha\| = \sum_{i=1}^d \alpha_i$.
$\mathscr{W}^{k,p}(\Omega)$	a Sobolev space defined as $\mathscr{W}^{k,p}(\Omega) = \left\{u \in \mathscr{L}^p(\Omega) \middle\| D^\alpha u \in \mathscr{L}^p(\Omega),\ \forall \|\alpha\| \le k\right\}$.
$\|\cdot\|_{\mathscr{W}^{k,p}(\Omega)}$	$\|u\|_{\mathscr{W}^{k,p}(\Omega)} := \begin{cases} \left(\sum_{\|\alpha\|\le k} \|D^\alpha u\|^p_{\mathscr{L}^p(\Omega)}\right)^{\frac{1}{p}}, & 1 \le p < +\infty \\ \max_{\|\alpha\|\le k} \|D^\alpha u\|_{\mathscr{L}^\infty(\Omega)}, & p = +\infty \end{cases}$.
$\mathscr{H}^k(\Omega)$	a Hilbert space with the norm $\|\cdot\|_{\mathscr{W}^{k,2}(\Omega)}$.
$\mathscr{H}_0^1(\Omega)$	the closure of $\mathscr{C}_0^\infty(\Omega) := \left\{u \in \mathscr{C}^\infty(\Omega) \middle\| u = 0 \text{ on } \partial\Omega\right\}$.
$\mathscr{H}^{-1}(\Omega)$	the dual space to $\mathscr{H}_0^1(\Omega)$.
$\mathscr{H}^1_{per}(Y)$	the closure of $\mathscr{C}^\infty_{per}(Y) := \left\{u \in \mathscr{C}^\infty(Y) \middle\| u \text{ periodic in } Y\right\}$.
$\mathscr{W}^1_{per}(Y)$	$\mathscr{W}^1_{per}(Y) := \left\{u \in \mathscr{H}^1_{per}(Y) \middle\| \int_Y u dy = 0\right\}$.
$\|\cdot\|_{\mathscr{F}}$	the Frobenius matrix norm $\|A\|_{\mathscr{F}} := \sqrt{\sum_i \sum_j \|a_{ij}\|^2}$.

Introduction

Why do we need multiscale methods?

Multiscale problems are considered in almost all the branches of modern applied science, describing different features appearing in different space or temporal scales, for example the macro properties versus micro structures of composite materials or volatility of high frequency financial data versus long term variation. The short history of our exploration of the micro world (the emergence of nanotechnology can be traced back to the 1980s) leave us many questions on the mechanisms of the micro structures and the transition from microscopic phenomena to macrosopic bulk properties. For the problems for which we are able to explain the physical mechanisms, modeling still remains a difficult task due to the variability and uncertainty of the microscopic heterogeneity. For instance, the permeability of aquifer for groundwater flow depends on the sizes and connectivity of the microscopic pores which are completely irregular while the groundwater transport occurs at a scale that makes the accurate modeling of all the micro structures accurately impossible. However, we cannot just simply ignore the heterogeneity of the micro feature. As shown in many physical experiments, the microscopic heterogeneity in composite materials induces significantly influence macroscopic properties (e.g. thermal or electric conductivity or elastic properties). Assuming that the multiscale problems are modeled by partial differential equations (PDEs), another challenge comes from solving those equations which is often prohibitive for classical numerical methods, i.e. finite element method (FEM), finite difference method (FDM) or finite volume method (FVM). For example, if a multiscale problem is modeled on micro scale at $10^{-6}m$ and the macroscopic domain is of the order of $1m^3$, to reasonably with a classical method resolve the micro scale, we need about 10^{18} number of degrees of freedom (DOF). The most powerful supercomputer "Tianhe-2" (National University of Defense Technology, China, ranking No. 1 in Top 500, June 2013) with storage 12.4PB (12.4^{16} bytes) can hardly handle this case [1].

Fortunately, for many applications, a macroscopic description is often enough (e.g. conductivity, effective flow, temperature, electricity, etc.). For those applications, the microscopic information needs only to be sampled on patches of the whole computational domain. In turn multiscale methods which can bridge the small scale heterogeneity the large scale properties without solving the whole multiscale problem on the small scale become attractive.

[1] Large energy consumption and intense noise are the extra price paid to gain the powerful computation, BBC archive, January 20, 2013.

Numerical homogenization methods

Consider a general family of PDEs $L_\varepsilon(u^\varepsilon) = f$ in a macro domain Ω that depend on a small parameter ε. The micro solution over the whole domain Ω is denoted by u^ε. Here ε can be either modeled (for two-scale problems) as the micro scale (a scalar) that is order of magnitude smaller than the size of the macro domain or (for $N+1$ scale problems) a series of meso and micro scales i.e. $\big(\varepsilon_1(\varepsilon), \cdots, \varepsilon_N(\varepsilon)\big)$, where there is a scale separation between any two scales. As studied in mathematical homogenization theory [42, 74], the family of micro solutions converge to a macro limit u^0, when $\varepsilon \to 0$ and this u^0 is the solution of the so-called homogenized (also macro or effective) equation $L_0(u^0) = f$ which is independent of the micro scale ε (see Fig. 1). Classical numerical methods can be applied to the homogenized equation but the effective data in the homogenized equations are not explicitly available in general and can only be computed from explicit cell problems in some special cases. More details on homogenization theory are presented in Chapter 1.

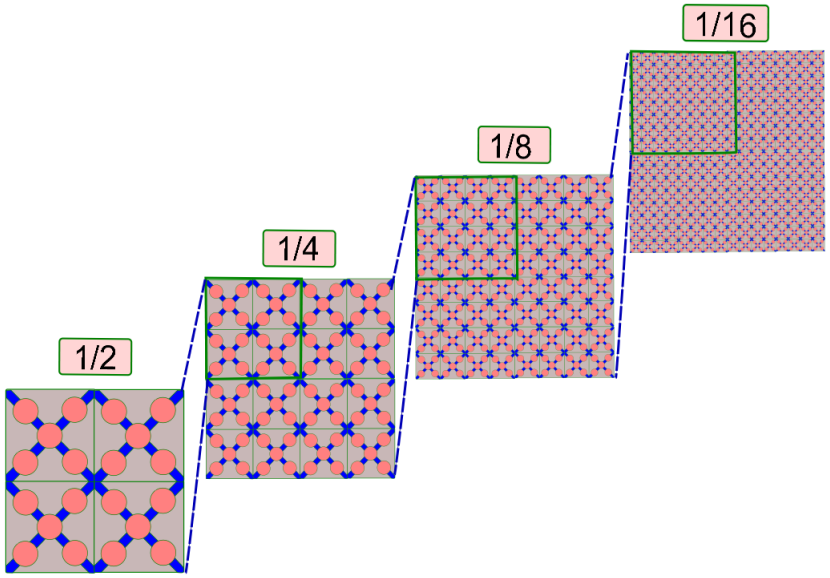


Figure 1: Heterogeneous domain with periodic hetergeneities of size $\varepsilon \to 0$.

Throughout this thesis, we consider a class of multiscale problems modeled by highly oscillating elliptic PDEs. Both linear and nonlinear problems will be considered as well as problems with more than two scales. In its simplest form (linear, two-scale) the elliptic multiscale problem reads as follows: Find $u^\varepsilon \in V(\Omega)$ such that

$$B(u^\varepsilon, v) := \int_\Omega a^\varepsilon(x)\nabla u^\varepsilon \cdot \nabla v dx = \int_\Omega f v dx := (f, v), \ \forall v \in V(\Omega). \tag{0.0.1}$$

where a^ε is a microscopic tensor oscillating on the small scale ε and $V(\Omega)$ is a Sobolev space that generally we assume to be $\mathscr{H}_0^1(\Omega)$. By homogenization theory, we have the homogenized

equation written as follows

$$B_0(u^0, v) := \int_\Omega a^0(x)\nabla u^0 \cdot \nabla v dx = (f, v),\ \forall v \in V(\Omega). \tag{0.0.2}$$

As pointed out in [34], solving (0.0.1) by classical numerical methods with coarse discretization (with meshsize $H >> \varepsilon$) gives non-converging numerical solution. However, the lack of explicit expressions for $a^0(x)$ prevents the possibility to apply classical numerical methods directly. Numerical homogenization methods must therefore be constructed. In what follows, we give a brief overview of several main numerical homogenization methods which we classify into three categories (following the recent review [6]):

- Type 1: methods based on a reduced model generated from the original fine scale problem.
- Type 2: methods that sample the original fine scale problem on patches to recover effective data of a macroscopic model and use correctors to reconstruct the fine scale solution.
- Type 3: methods based on the two-scale convergence homogenization theory that solve the full limit problem in a sparse tensor product FE space.

We focus on the multiscale methods based on FEMs but we note that the algorithm described below can often be implemented (and analyzed) for other methods, e.g. FDM, FVM.

Type 1 multiscale methods: VMM, RFB, MsFEM

In 1983, Babŭska and Osborn developed in [35] the so-called generalized finite element method (GFEM) for one dimensional PDEs with rough coefficients. The main idea of the GFEM is to modify the coarse FE space applied to (0.0.1) by adding fine scale information. Inspired by the GFEM, various methods were developed later for high dimensional problems.

Variational multiscale method (VMM) and residual free bubble (RFB) method. Hughes and collaborators proposed in [70, 71] the variational multiscale method (VMM) for an efficient approximation of mulitscale problem. The basic idea is to decompose the numerical approximation u^h of u^ε into $u^h = u^H + \tilde{u}$, where u^H is computed in a coarse FE space $V_H(\Omega)$ and $\tilde{u}$ is computed in a fine space $\tilde{V}(\Omega)$ (i.e. a FE space with a fine mesh or higher order polynomial space). We then use the following decomposition to approximate the solution of problem (0.0.1): Find $u^h = u^H + \tilde{u} \in V_H \oplus \tilde{V}$ satisfying

$$\begin{aligned} B(u^H + \tilde{u}, v^H) &= (f, v^H) \quad \forall v^H \in V_H(\Omega), & (0.0.3)\\ B(\tilde{u}, \tilde{v}) &= (f, \tilde{v}) - B(u^H, \tilde{v}) \quad \forall \tilde{v} \in \tilde{V}(\Omega). & (0.0.4) \end{aligned}$$

By Riesz representation theorem, there exists a mapping $\mathscr{L} : V_H \to \tilde{V}$ such that $(\mathscr{L}(u^H), \tilde{v}) =$

$B(u^H, \tilde{v})$ and thus we have $B(\tilde{u}, \tilde{v}) = (f - \mathscr{L}(u^H), \tilde{v})$. We can further write $\tilde{u} = \mathscr{M}(f - \mathscr{L}(u^H))$ and reformulate (0.0.3) as

$$B(u^H, v^H) + B(\mathscr{M}(f - \mathscr{L}(u^H)), v^H) = (f, v^H), \ \forall v^H \in V_H,$$

where $\mathscr{M} : V_H \to \tilde{V}$ is a bounded linear operator. For the VMM, the key for efficiency is to apply various strategies to localize the operator $\mathscr{M}$. A simple way for the localization is to set $\tilde{u}$ to be zero on the boundary of each coarse element K, i.e. solving $\tilde{u}$ in space $\tilde{V} := \{v \in V_h(\Omega); v|_{\partial K} = 0, \ \forall K\}$ where $V_h(\Omega)$ is an FE space with a partition on Ω that resolves the fine scale.

In the residual free bubble (RFB) method [47] one starts with the coarse FE space and enriches it by adding the so-called bubble functions on each coarse element which are computed in a fine localized FE space and which vanish on the boundary of each coarse FE. The RFB method can be seen as a specific realization of the VMM.

Multiscale finite element method (MsFEM). The multiscale finite element method (MsFEM) proposed in [69] is a development of the GFEM from a different point of view. Instead of adding micro structure information into the macro FE space, the MsFEM constructs a special finite element space with localized oscillating basis functions. In the MsFEM, we first set a macro partition of a macro FE space V_H and write $\{\phi^H_{K,1}, \cdots, \phi^H_{K,M}\} \subset V_H$ as the basis functions of a macro element K. We then construct the local oscillating basis functions by solving the following cell problems: Find $\phi^h_{K,i} - \phi^H_{K,i} \in V_h(K)$ (where $V_h(K)$ is a fine scale FE space defined on K with meshsize $h < \varepsilon$) such that

$$\int_K a^\varepsilon(x) \nabla \phi^h_{K,i} \cdot \nabla z^h dx = 0, \ \ \forall z^h \in V_h(K), \ i = 1, \cdots, M,$$

with the boundary condition

$$\phi^h_{K,i}|_{\partial K} = \phi^H_{K,i}. \tag{0.0.5}$$

The original oscillating PDE (0.0.1) is then solved in the multiscale FE space $V_{MsFEM} := span\{\phi^h_{K,1}, \cdots, \phi^h_{K,N}\}$ and the MsFEM approximation reads as follows: Find $u^{hH} \in V_{MsFEM}$ such that

$$B(u^{hH}, v^{hH}) = (f, v^{hH}), \ \ \forall v^{hH} \in V_{MsFEM}.$$

Convergence results for locally periodic linear problems i.e. $a^\varepsilon(x) = a(x, \frac{x}{\varepsilon}) = a(x, y)$ y-periodic in the reference cell Y, are proved in [31, 69] (assuming that V_H and V_h consist of piecewise linear polynomials): There exists constants C_1, C_2 independent of H, h, ε such that

$$\| u^\varepsilon - u^{hH} \|_{\mathscr{H}^1(\Omega)} \leq C_1 \big(H + (\frac{h}{\varepsilon}) \big) + C_2 (\frac{\varepsilon}{H})^{1/2}. \tag{0.0.6}$$

The term $C_2(\frac{\varepsilon}{H})^{1/2}$ is the so called resonance error, appearing because of artificial boundary conditions (0.0.5) for the micro problem. The resonance error can be reduced to $\mathcal{O}(\frac{\varepsilon}{H})$ by using an oversampling technique. We refer to [61] for a general review. Higher order MsFEMs have been developed in [31, 91] based on harmonic coordinates. As pointed out in [99], the VMM and RFB share similarities with the MsFEM.

Now we have a look at the complexity of the MsFEM. Observed from (0.0.6), it requires that $\frac{h}{\varepsilon} \approx H = \frac{1}{N_{mac}}$ to obtain optimal convergence rate, where N_{mac} is the macro DOF in one direction. Therefore the total complexity is $\mathcal{O}\big((N_{mac})^d \cdot (\frac{H}{h})^d\big) = \mathcal{O}\big((N_{mac})^d \cdot \varepsilon^{-d}\big)$. Since the cell problems are independent of each other, parallel computation can be applied to improve the efficiency.

Type 2 multiscale methods: RVE, HMM

For this type of methods, one considers the homogenized equation for example (0.0.2) and uses local cell problems to construct the unknown data in the homogenized equation in order to obtain an approximation of the homogenized solution u^0. A classical engineering method, the so-called representative volume element (RVE) method, recovers the effective data (i.e. a^0) by a precomputed micro problem on a sampling cell domain where the size of the sampling domain is sufficiently large to be statistically representative of all the heterogeneity of the micro structure. However, the choice of the sampling cell (the location, the size) is somehow unrelated to the the macro solver for the homogenized problem. In turn, convergence results for the numerical solution cannot be proved except for uniformly periodic problems (i.e. $a^\varepsilon(x) = a(\frac{x}{\varepsilon} = a(y)$ y-periodic in Y). Moreover for non-periodic problems, the size of the sampling domains, their location and the propagation of information between micro and macro solvers are delicate issues. For nonlinear problems, these methods are usually very expensive, without theoretical foundation ensuring convergence.

The heterogeneous multiscale method (HMM) proposed in [59], provides an efficient strategy to overcome the limitation of the RVE method. For the HMM, one applies a macro solver to the homogenized equation in which the missing numerical data are located and estimated by localized microscopic problems. The link between micro and macro solvers is built in the methodology and macroscopic method can be seen as a numerical method with quadrature for a modified effective problem. In turn, available technique for single scale problems can be re-used. The HMM also provides large flexibility to choose macro solvers for different goals (e.g. adaptivity, local conservation properties, etc.). We mention here the finite difference HMM [13], the spectral HMM [15], the finite element HMM (FE-HMM) [60, 1], the discontinuous FE-HMM for second order elliptic equations in [5] and for convection diffusion equations in [16], the adaptive FE-HMM in [18, 20, 21], etc.

In this thesis, we focus on the FE-HMM that will be described in Chapter 1. The basic idea of the FE-HMM is to use an FEM with numerical quadrature formulas (QF) for the macro solver on a macroscopic triangulation of the physical domain Ω with meshsize H. The unknown data on each quadrature point is recovered on-the-fly by solving cell problems with a micro FEM on sampling domains centered at the corresponding quadrature points, see Fig. 2. The sampling

domains are usually cubes of size δ and we have $\varepsilon \leq \delta << H$. The triangulation on the sampling domain resolves the ε-scale and the micro mesh size h therefore satisfies $h < \varepsilon$.

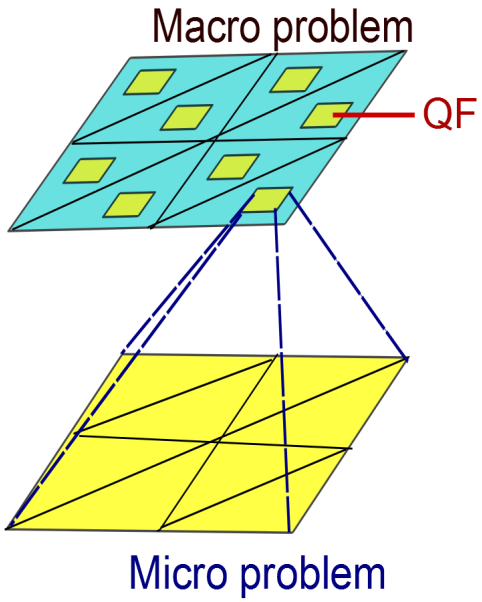


Figure 2: The FE-HMM.

As pointed out in [60, 3], the fine scale solution of (0.0.1), can be reconstructed by adding an oscillating corrector to the the macro numerical solution of (0.0.2), where the corrector is a linear combination of the cell solutions (similar idea also appeared in other contexts [89]). For locally periodic problems, using periodic constraints for the micro problem and sampling domains of size $\delta = N\varepsilon,\ N \in \mathbb{N}$, we have the following convergence results (taking piecewise linear finite element for both macro and micro problems for simplicity) [60, 1]

$$\|u^0 - u^H\|_{\mathscr{H}^1(\Omega)} \leq C_1\big(H + (\frac{h}{\varepsilon})^2\big) + C_2\varepsilon,$$
$$\|u^\varepsilon - u^{hH}\|_{\mathscr{H}^1(\Omega)} \leq C_1\big(H + \frac{h}{\varepsilon}\big) + C_2\sqrt{\varepsilon},$$

where u^H is the FE-HMM solution and u^{hH} is the corresponding reconstruction for u^ε based on u^H. To obtain the optimal convergence rate for both macro solution and micro reconstruction, we set $H = \frac{h}{\varepsilon} := \frac{1}{N_{mac}}$ (similar to the MsFEM). Therefore, the total complexity is $\mathscr{O}(N_{mac}^{2d})$ scaled independent of ε. We refer to [14, 58] for an overview of the HMM framework and [3, 4] for recent developments of the FE-HMM.

Type 3 multiscale method: Sparse FEM

This method focuses on the limit equation based on the two-scale convergence theory and its generalization [30, 81]. Solving this limit equation by using the tensor product FEM leads to a complexity comparable to solving the original fine scale problem. The sparse tensor FEM introduced in [68] based on hierarchical sequences of FE spaces allows to significantly reduce the cost of the method. Indeed, it can be shown that the complexity becomes comparable to single scale problems. However, the implementation of this method is a challenging task especially for problems with complex geometries. Furthermore the optimal numerical linear algebra solvers for such a method are still under investigation.

Main contribution

The main contributions of this thesis address the following two issues.

First issue. Most of the proposed methods for elliptic homogenization problems have only been analyzed and implemented for two separated scales (a macro and a micro scale) (see [61, 3, 58, 14] and the references therein). Therefore it is of interest to have a generalized FE-HMM for problems with more than two scales. In Chapter 2, we propose a rigorous analysis of the FE-HMM for elliptic problems with an arbitrary numbers of well separated scales. While the design of the generalized FE-HMM is straightforward from the method for two-scale problems proposed in [59], the analysis is considerably more difficult than the two-scale analysis [60, 1] due to the multi-level hierarchy structure of the method and the numerical integration involved in each level.

Second issue. Complexity of the two-scale FE-HMM due to the simultaneous refinement of macro and micro problems is a serious issue for: High-dimensional problems, high-order methods, adaptive procedures and nonlinear problems.

Therefore we propose to couple the FE-HMM methodology with a reduced order modeling strategy, the reduced basis (RB) method, to address the complexity issue of the FE-HMM. After designing a new method, namely the reduced basis finite element heterogeneous multiscale method (RB-FE-HMM), we derive

i an a priori error analysis for linear problems;

ii energy norm based and goal oriented a posteriori error analysis for linear problems;

iii an a priori error analysis for nonlinear nonmonotone problems.

In our approach the RB-FE-HMM, we select by a greedy procedure a number of representative sampling domains on which we solve accurately micro problems. Their corresponding solutions span the RB space. This procedure is called the offline stage, in the RB terminology, and is usually only done once, as a pre-processing step. In a so-called online stage, the effective solution is obtained from the macro solver of the FE-HMM with effective coefficients recovered from

micro problems solved in the RB space. The required data at the macro integration points are now obtained from the solutions of small dimensional linear problems involving suitable interpolations of the precomputed RB space which is independent of the macro discretization. In turn, optimal convergence rates in the RB-FE-HMM can be obtained just by refining the macro mesh. Thus, expensive micro FE computations as required by the FE-HMM are avoided. High order macroscopic methods can be designed with the same set of RB as used for linear macro FE. A priori error analysis including macro error, micro error, resonance error and error coming from the use of the RB is derived. As demonstrated in the numerical experiments, the online time cost of the RB-FE-HMM can be comparable to the single scale FEM. For problems with large macro meshes or with iteration steps in macro solvers, the total cost of the RB-FE-HMM is often only a few percents of the cost for the FE-HMM with the same macro mesh.

The first step of our work focuses on designing the RB-FE-HMM for inear elliptic multiscale problems, presented in Chapter 4. The further exploration on this method discussed in Chapter 5, focuses on the adaptive RB-FE-HMM where two kinds of multiscale adaptive a posteriori error estimators are considered: The energy-norm based estimator and the goal-oriented estimator. In Chapter 6, we extend the RB-FE-HMM for quasi-linear multiscale problems, where a new a posteriori error estimator for the RB offline is designed in order to guarantee the convergence of Newton method used for the macro solver in the online stage. We emphasize that though all of those methods are analyzed only for elliptic multiscale PDEs in this thesis, they can be easily applied to time dependent problems just with small modifications. In Chapter 7, we present the implementation details of the quasilinear RB-FE-HMM for both static and time dependent problems, and present extensive numerical simulations for two and three dimensional problems.

Part I

The finite element hetergeneous multiscale method and its generalization

In this part we consider the following elliptic multiscale problem

$$\begin{aligned} -\nabla\cdot(a^\varepsilon(x)\nabla u^\varepsilon(x)) &= f \text{ in } \Omega, \\ u^\varepsilon(x) &= 0 \text{ on } \partial\Omega, \end{aligned} \tag{1.0.1}$$

in a domain $\Omega \in \mathbb{R}^d$, $d \leq 3$.

We first shortly review the finite element heterogeneous multiscale method (FE-HMM) for the two scale (macro-micro scale) problems i.e. for $a^\varepsilon = a(x, \frac{x}{\varepsilon})$. We then generalize the method for problems with more than two scales i.e., $a^\varepsilon = a(x, \frac{x}{\varepsilon_1}, \cdots, \frac{x}{\varepsilon_N})$ and provide a complete error analysis.

Rigorous averaging theory for problems such as (1.0.1) has been derived in the framework of homogenization theory. We mention homogenization techniques using the multiple scale expansion [42, 74]. Convergence of the fine scale solution towards the homogenized solution can be studied using the energy method due to Tartar [104]. Another approach is the two-scale convergence proposed by Nguentseng [81], developed and generalized by Allaire and Briane [29, 30] for problems with more than two-scales. While numerous numerical methods have been proposed for two-scale problems, only a few numerical strategies have been proposed and analyzed for problems with more than two scales. We mention an analysis of the multiscale finite element method by Efendiev et al. [62], and the numerical method based on high-dimensional FEM and sparse tensor-product approximation [68] based on the limit problem obtained from the reiterated homogenization proposed in [30]. Here, we propose the generalized FE-HMM for problems with more than two scales and study the fully discrete error analysis.

Outline of Part I.

- Chapter 1: Gives a short review on homogenization theory, the FEM and the FE-HMM for two-scale problems.
- Chapter 2: Presents the generalized FE-HMM and the fully discrete error analysis for $(N+1)$-scale problems. This chapter is essentially taken from [9].

1 Homogenization and the finite element heterogeneous multiscale method

We consider the second-order elliptic problem (1.0.1) in a bounded polyhedron domain $\Omega \subset \mathbb{R}^d$, $d \leq 3$. Here, for simplicity we choose a zero Dirichlet boundary condition and $f \in \mathscr{L}^2(\Omega)$ (note that this condition can be relaxed to $f \in \mathscr{H}^{-1}(\Omega)$). The $d \times d$ tensor $a^\varepsilon(x)$ is assumed to be symmetric uniformly elliptic and bounded, i.e.,

$$\lambda|\xi|^2 \leq a^\varepsilon(x)\xi \cdot \xi, \ |a^\varepsilon(x)\xi| \leq \Lambda|\xi|, \ \forall \xi \in \mathbb{R}^d, \forall \varepsilon > 0, \tag{1.0.1}$$

for a.e $x \in \Omega$. The tensor varies on a small spatial scale denoted by ε. Homogenization theory provides the theoretical foundation for the numerical homogenization methods discussed in Introduction. Here we present the basic results of the homogenization theory in Section 1.1. For simplicity of the presentation, we will assume that the tensor a^ε is symmetric, but the numerical method presented below and its analysis can be generalized to nonsymmetric tensors.

1.1 Homogenization

Mathematical homogenization aims at describing "averaged" (i.e., homogenized) solutions of PDEs with rapidly oscillating coefficients varying over multiple spatial or temporal scales. For a function $\phi(x,y)$, where $x \in \Omega$ is called the slow variable and $y = \frac{x}{\varepsilon} \in Y$ (the domain of periodicity, e.g., $Y = (-\frac{1}{2}, \frac{1}{2})^d$) is called the fast variable, we consider $\phi(x, \frac{x}{\varepsilon})$ defined in Ω.

Formal asymptotic expansion. In the locally periodic case, i.e., $a^\varepsilon(x) = a(x, \frac{x}{\varepsilon}) = a(x,y)$ is Y-periodic in the y variable, a formal approach relies on the multiple scale expansion (see [42]). Therefore, we look for a solution u^ε of (1.0.1) in the form $u^\varepsilon(x) = u^0(x, \frac{x}{\varepsilon}) + \varepsilon u^1(x, \frac{x}{\varepsilon}) + \varepsilon^2 u^2(x, \frac{x}{\varepsilon}) + \ldots$. Upon inserting the ansatz in (1.0.1) and rewriting the operator $\nabla := \nabla_x + \frac{1}{\varepsilon}\nabla_y$, one first obtains (formally) by identifying the terms in ε^{-2} to zero that $u^0(x, \frac{x}{\varepsilon}) = u^0(x)$. Then identifying the terms in ε^{-1} to zero gives the so-called cell problem

$$-\nabla_y \cdot \big(a(x,y)\nabla_y \chi_i(x,y)\big) = \nabla_y \cdot (a(x,y)\mathbf{e}_i), \ \ y \in Y, \ i = 1, \cdots, d, \tag{1.1.2}$$

where $\chi_i(\cdot, y) \in \mathscr{H}^1_{per}(Y) := \{g \in \mathscr{H}^1(Y) |\ g \text{ periodic in } Y\}$ is unique up to a constant and $\mathbf{e}_i$ is the ith canonical basis of $\mathbb{R}^d$. We thus write the second term of the above ansatz as $u^1(x,y) = \sum_{i=1}^d \chi_i(x,y)\frac{\partial u^0}{\partial x_i}$. By identifying the coefficient in ε^0 one can deduce that $u^0(x)$ is a solution of the following "homogenized equation"

$$\begin{aligned} -\nabla\cdot\left(a^0(x)\nabla u^0(x)\right) &= f(x) \ \text{ in } \Omega \\ u^0(x) &= 0 \ \text{ on } \ \partial\Omega, \end{aligned} \tag{1.1.3}$$

and the homogenized tensor $a^0(x)$ can be expressed as

$$a^0(x) = \frac{1}{|Y|}\int_Y a(x,y)(I + \nabla_y \chi(x,y))\,dy, \tag{1.1.4}$$

where $\chi(x,y) := (\chi_1(x,y), \cdots, \chi_d(x,y))$ is a vector function, I is a $d \times d$ identity matrix and $|Y|$ denotes the volume of Y. By the expression (1.1.4), one can show that $a^0(x)$ again is uniformly bounded and elliptic. We note that for global periodic tensors, i.e. $a^\varepsilon(x) = a(\frac{x}{\varepsilon}) = a(y)$ $Y-$periodic in y, the corresponding homogenized tensor a^0 is constant.

Convergence analysis. These formal computations do not provide a converge result of u^ε towards u^0. Using Tartar's method of oscillating test functions [103] (see also [42]) it is possible to show that for locally periodic problem $u^\varepsilon \rightharpoonup u^0$ weakly in $\mathscr{H}^1_0(\Omega)$, $a^\varepsilon \nabla u^\varepsilon \rightharpoonup a^0 \nabla u^0$ weakly in $(\mathscr{L}^2(\Omega))^d$, where u^0 is the solution of (1.1.3).

Furthermore, as proved in [74, Chapter 1] assuming appropriate regularity of oscillating tensor and appropriate domain Ω (convex with smooth boundaries), strong $\mathscr{L}^2$ convergence result

$$\|u^\varepsilon - u^0\|_{\mathscr{L}^2(\Omega)} \leq C\varepsilon \tag{1.1.5}$$

can be obtained. Based on the strong $\mathscr{L}^2$ convergence, one can only obtain for the $\mathscr{H}^1$ norm that $\|u^\varepsilon - u^0\|_{\mathscr{H}^1(\Omega)} \sim \mathscr{O}(1)$, which indicates that the oscillation in the gradient of u^ε cannot be captured by u^0. But the $\mathscr{H}^1$ convergence estimate can be improved by adding the corrector $\varepsilon u^1(x, \frac{x}{\varepsilon})$ to the homogenized solution u^0 and we have

$$\|u^\varepsilon - (u^0 + \varepsilon u^1)\|_{\mathscr{H}^1(\Omega)} \leq C\sqrt{\varepsilon}, \tag{1.1.6}$$

again assuming appropriate regularity of the oscillating tensor and the domain.

The problem gets more involved if a^ε is not locally periodic. On the theoretical side, one can rely on $H-$convergence [103] which ensures the existence of a subsequence of the matrices a^ε and a homogenized tensor a^0 (again uniformly elliptic and bounded) such that for the corresponding subsequence, u^ε and $a^\varepsilon \nabla u^\varepsilon$ enjoy the same convergence properties as described above for the

locally periodic case. However for non-periodic oscillating tensors, the homogenized tensors are in general unknown (in an explicit form). For numerical homogenization one usually assumes scale separation between fast and slow variables and relies on local problems to compute the homogenized tensor for a given value of $x \in \Omega$.

A typical example we consider for numerical homogenization is the locally periodic tensor. However even for such a simplified situation, we have an infinite number of cell problems (for each $x \in \Omega$), whose solutions must usually be computed numerically. The task in numerical homogenization is thus to design an algorithm capable of computing an approximation of the homogenized solution $u^0(x)$, relying on a finite number of cell problems chosen in such a way that the overall computation is efficient and reliable.

1.2 The finite element heterogeneous multiscale method (FE-HMM)

As discussed in the introduction, the FE-HMM belongs to the class of numerical homogenization methods. It is based on a macroscopic FEM with QF defined on a macroscopic partition of the domain Ω. As $a^0(x)$ is usually unknown, the method is supplemented by microscopic FEMs defined on sampling domains centered at the macroscopic quadrature points of the QF, relying only on the data given in (1.0.1). A suitable averaging of the solutions of the microscopic FEMs allows to recover the missing macroscopic tensor at the macroscopic quadrature points.

The method is applicable to general problems and does not rely on a specific structure of the oscillating tensor such as periodicity. We however assume that there is a well defined homogenized problem with certain smoothness properties and a scale separation between fast and slow variables, which we will make precise in the following. We emphasize that although we consider the model equation (1.0.1) with homogeneous Dirichlet boundary condition , the numerical method presented below (as well as the methods presented in later chapters) can be readily generalized to other boundary conditions , e.g.,

$$\begin{aligned} &u^\varepsilon(x) = g_D(x) \quad \text{on } \partial\Omega_D, \\ &n \cdot (a^\varepsilon(x)\nabla u^\varepsilon(x)) = g_N(x) \quad \text{on } \partial\Omega_N, \end{aligned}$$

where $\partial\Omega = \partial\Omega_D \cup \partial\Omega_N$, with $\partial\Omega_D$ having non-zero measure, and $g_D \in \mathscr{H}^{\frac{1}{2}}(\partial\Omega_D)$, $g_N \in \mathscr{L}^2(\partial\Omega_N)$. We also emphasize that all the methods in the following chapters (for both of the FE-HMM and the RB-FE-HMM) can be easily extended to non-symmetric problems.

1.2.1 Preliminary: The single scale finite element method (FEM)

In this subsection, we briefly introduce the standard FEM for single scale problems and various a priori error estimates that are used in the analysis of the FE-HMM and refer to [95] for the

detailed introduction of FEM. We consider the following single scale problem

$$\begin{aligned} -\nabla\cdot\big(a(x)\nabla u(x)\big) &= f \quad \text{in } \Omega, \\ u(x) &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{1.2.7}$$

where we assume that the right hand side function $f \in \mathscr{L}^2(\Omega)$ and that $a(x)$ is symmetric, uniformly bounded and elliptic (hence ensuring a unique solution for (1.2.7)). We then write this single scale problem (1.2.7) into the following weak form

$$B_s(u,v) := \int_\Omega a\nabla u\cdot\nabla v\,dx = \int_\Omega f v\,dx, \quad \forall v \in \mathscr{H}_0^1(\Omega).$$

Here we use the subscript s in this subsection to denote the single scale problem in order distinguish from the bilinear forms defined later for multiscale problems. Let $\mathscr{T}_H$ is a family of partition of Ω in simplicial or quadrilateral elements K of diameter H_K, and $\mathscr{R}^\ell(K)$ is the space $\mathscr{P}^\ell(K)$ of polynomials on K of total degree at most ℓ if K is a simplicial FE, or the space $\mathscr{Q}^\ell(K)$ of polynomials on K of degree at most ℓ in each variable if K is a quadrilateral FE. For a given domain partition, denote $H := \max_{K\in\mathscr{T}_H} H_K$. We then define the finite element space as the following

$$S_0^\ell(\Omega,\mathscr{T}_H) = \{v^H \in \mathscr{H}_0^1(\Omega);\ v^H|_K \in \mathscr{R}^\ell(K),\ \forall K \in \mathscr{T}_H\}, \tag{1.2.8}$$

which is a finite dimensional subspace of $\mathscr{H}_0^1(\Omega)$.

FEM with numerical quadrature is a basic building block of the FE-HMM described later. Consider therefore a quadrature formula (QF) $\{\hat{x}_j,\hat{\omega}_j\}_{j=1}^J$ on a reference element $\hat{K}$. We equip each element K with a corresponding QF $\{x_{K_j},\omega_{K_j}\}_{j=1}^J$ by using a C^1-diffeomorphism. We will make the following assumptions on the quadrature formula (see [54]),
(**Q1**) $\hat{\omega}_j > 0,\ j = 1,\dots,J, \quad \sum_{j=1}^J \hat{\omega}_j|\nabla\hat{p}(\hat{x}_j)|^2 \geq \hat{\lambda}\|\nabla\hat{p}\|^2_{\mathscr{L}^2(\hat{K})},\ \forall\hat{p}(\hat{x}) \in \mathscr{R}^\ell(\hat{K})$, with $\hat{\lambda} > 0$;
(**Q2**) $\int_{\hat{K}}\hat{p}(\hat{x})d\hat{x} = \sum_{j\in J}\hat{\omega}_j\hat{p}(\hat{x}_j),\ \forall\hat{p}(\hat{x}) \in \mathscr{R}^\sigma(\hat{K})$, where $\sigma = \max(2\ell-2,\ell)$ if $\hat{K}$ is a simplicial FE, or $\sigma = \max(2\ell-1,\ell+1)$ if $\hat{K}$ is a rectangular FE.
The assumptions (**Q1**) and (**Q2**) will be repeatedly used in this thesis.

The FEM with numerical quadrature for Problem (1.2.7) reads: Find $u^H \in S_0^\ell(\Omega,\mathscr{T}_H)$ such that

$$B_{s,H}(u^H,v^H) = \int_\Omega f v^H dx, \quad \forall v^H \in S_0^\ell(\Omega,\mathscr{T}_H),$$

where

$$B_{s,H}(v^H,w^H) := \sum_{K\in\mathscr{T}_H}\sum_{x_{K_j}\in K}\omega_{K_j}a(x_{K_j})\nabla v^H(x_{K_j})\cdot\nabla w^H(x_{K_j})$$

The following classical error estimates for FEM with numerical quadrature can be found in [53, Chapter 3] and the related references within.

Theorem 1.2.1. *Assume that the family of partition $\mathcal{T}_H$ is regular and that all the elements $K \in \mathcal{T}_H$ are affine-equivalent and of class C^0. Furthermore we assume that the solution u of* (1.2.7) *has the regularity of $\mathcal{H}^{\ell+1}(\Omega)$. In addition suppose that $a \in \mathcal{W}^{\ell,\infty}(\Omega)$ for $\mathcal{H}^1$ estimate and $a \in \mathcal{W}^{\ell+1,\infty}(\Omega)$ for $\mathcal{L}^2$ estimate and that assumptions (**Q1**) (**Q2**) hold for the QF coupled with FEM with numerical quadrature, then we have for the FEM solution $u^H \in S_0^\ell(\Omega,\mathcal{T}_H)$ that*

$$\|u-u^H\|_{\mathcal{H}^1(\Omega)} \leq CH^\ell |u|_{\mathcal{H}^{\ell+1}(\Omega)}, \quad \|u-u^H\|_{\mathcal{L}^2(\Omega)} \leq CH^{\ell+1}|u|_{\mathcal{H}^{\ell+1}(\Omega)}, \tag{1.2.9}$$

where $|\cdot|_{\mathcal{H}^{\ell+1}(\Omega)}$ is the $\mathcal{H}^{\ell+1}$ semi-norm.

1.2.2 The FE-HMM

We next describe the FE-HMM and note that for the method proposition we do not need to assume the oscillating tensor is locally periodic. Here we denote $S_0^\ell(\Omega,\mathcal{T}_H)$ as the macro FE space for the macro problem defined in (1.2.8) where we note that H in the macro discretization is allowed to be much larger than ε. Ideally we would like to apply the FEM to (1.1.3) to obtain the numerical macro solution, i.e. find $u^{0,H} \in S_0^\ell(\Omega,\mathcal{T}_H)$ such that

$$B_{0,H}(u^{0,H},v^H) = \int_\Omega f v^H dx, \quad \forall\, v^H \in S_0^\ell(\Omega,\mathcal{T}_H), \tag{1.2.10}$$

where for $v^H, w^H \in S_0^\ell(\Omega,\mathcal{T}_H)$

$$B_{0,H}(v^H,w^H) := \sum_{K\in\mathcal{T}_H}\sum_{x_{K_j}\in K} \omega_{K_j} a^0(x_{K_j})\nabla v^H(x_{K_j})\cdot\nabla w^H(x_{K_j}) \tag{1.2.11}$$

and $a^0(x_{K_j})$ is the homogenized tensor of problem (1.1.3) at the quadrature point x_{K_j}.

In order to estimate the unknown data in (1.2.11), we need to introduce micro cell problems. We first define for each macro element $K \in \mathcal{T}_H$ and each integration point $x_{K_j} \in K$, $j = 1,\ldots,J$, the sampling domains $K_{\delta_j} = x_{K_j} + (-\delta/2,\delta/2)^d$, $(\delta \geq \varepsilon)$. For a sampling domain K_{δ_j}, we then define a micro FE space $S^q(K_{\delta_j},\mathcal{T}_h) \subset W(K_{\delta_j})$ with simplicial or quadrilateral FEs and piecewise polynomial of degree q for simplicial FEs (or piecewise polynomial of maximum degree q in each variable for quadrilateral FEs) where $\mathcal{T}_h$ is a conformal and shape regular family of partition. The space $\mathcal{W}(K_{\delta_j})$ is either the Sobolev space

$$\mathcal{W}(K_{\delta_j}) = \mathcal{W}^1_{per}(K_{\delta_j}) \tag{1.2.12}$$

for a periodic coupling or the Sobolev space

$$\mathscr{W}(K_{\delta_j}) = \mathscr{H}^1_0(K_{\delta_j}) \tag{1.2.13}$$

for a coupling with Dirichlet boundary conditions.

We first apply the FEM at the macroscopic level: Find $u^H \in S^\ell_0(\Omega, \mathscr{T}_H)$ such that

$$B_H(u^H, v^H) = \int_\Omega f v^H dx, \ \forall v^H \in S^\ell_0(\Omega, \mathscr{T}_H), \tag{1.2.14}$$

where

$$B_H(v^H, w^H) := \sum_{K\in\mathscr{T}_H} \sum_{j=1}^{J} \frac{\omega_{K_j}}{|K_{\delta_j}|} \int_{K_{\delta_j}} a^\varepsilon(x) \nabla v^h_{K_j}(x) \cdot \nabla w^h_{K_j}(x) dx. \tag{1.2.15}$$

In (1.2.15) $v^h_{K_j}$ (respectively $w^h_{K_j}$) denotes the solution of the following micro problem (computed on-the-fly): Find $v^h_{K_j}$ such that $v^h_{K_j} - v^H_{lin,j}(x) \in S^q(K_{\delta_j}, \mathscr{T}_h)$ and

$$\int_{K_{\delta_j}} a^\varepsilon(x) \nabla v^h_{K_j}(x) \cdot \nabla z^h(x) dx = 0 \quad \forall z^h \in S^q(K_{\delta_j}, \mathscr{T}_h), \tag{1.2.16}$$

where we used the notation $v^H_{lin,j}(x) := v^H(x_{K_j}) + (x - x_{K_j}) \cdot \nabla v^H(x_{K_j})$.

1.2.3 A priori estimates for the FE-HMM

The following estimates hold provided suitable regularity assumptions on $a^0(x)$, $f(x)$ and $u^0 \in H^{\ell+1}(\Omega)$ (see [3, 1, 60]):

$$\begin{aligned} \left\| u^0 - u^H \right\|_{\mathscr{H}^1(\Omega)} &\le C\left(H^\ell + r_{MIC} + r_{MOD}\right), \\ \left\| u^0 - u^H \right\|_{\mathscr{L}^2(\Omega)} &\le C\left(H^{\ell+1} + r_{MIC} + r_{MOD}\right). \end{aligned} \tag{1.2.17}$$

The term r_{MOD} (called modeling error) quantifies the error induced by artificial micro boundary conditions or non-optimal sampling of the micro structure. It does not depend on the macro or micro meshsizes and can be analyzed for locally periodic tensor [60, 22, 57]. The term r_{MIC} (called micro error) quantifies the error due to the micro FEM.

To estimate the micro error, it is convenient to make the following assumption on $\psi^i_{K_j}$, the exact solution of Problem (1.2.16), i.e., the solution of $\int_{K_{\delta_j}} a^\varepsilon(x) \nabla(\psi^i_{K_j}(x) + x_i) \cdot \nabla z(x) dx = 0 \quad \forall z \in \mathscr{W}(K_{\delta_j}),\ i = 1, \dots, d$, [4, 5]

(**H1**) Given $q \in \mathbb{N}$, the cell functions $\psi^i_{K_j}$ satisfy

$$|\psi^i_{K_j}|_{\mathscr{H}^{q+1}(K_{\delta_j})} \leq C\varepsilon^{-q}\sqrt{|K_{\delta_j}|},$$

with C independent of $i = 1 \dots d$, ε, the quadrature point x_{K_j} and the domain K_{δ_j}.

Remark 1.2.2. *If $\mathscr{W}(K_{\delta_j}) = \mathscr{H}^1_0(K_{\delta_j})$ then (**H1**) holds for $q = 1$ provided $a^\varepsilon|_K \in \mathscr{W}^{1,\infty}(K)$ and $|a^\varepsilon_{ij}|_{\mathscr{W}^{1,\infty}(K)} \leq C_K\varepsilon^{-1}$ for $i, j = 1, \dots d$. If K_{δ_j} covers an integer number of periods and $a^\varepsilon(x) = a(x, \frac{x}{\varepsilon}) = a(x, y)$, then (**H1**) holds for higher order q provided that periodic boundary conditions are used for the micro problems and $a^\varepsilon = a(\cdot, y)$ is smooth enough.*

Then, the microscopic error r_{MIC} can be bounded by [1, 4, 5]

$$r_{MIC} \leq C\left(\frac{h}{\varepsilon}\right)^{2q}. \tag{1.2.18}$$

Finally, assuming

$$a^\varepsilon(x) = a(x, \frac{x}{\varepsilon}) = a(x, y) \text{ Y-periodic in } y, \tag{1.2.19}$$

the modeling error r_{MOD} can be bounded as follows [60, 22]

$$r_{MOD} = 0 \qquad \text{if } \mathscr{W}(K_{\delta_j}) = \mathscr{W}^1_{per}(K_{\delta_j}) \text{ and } \delta/\varepsilon \in \mathbb{N},^1 \tag{1.2.20}$$

$$r_{MOD} \leq C(\delta + \frac{\varepsilon}{\delta}) \qquad \text{if } \mathscr{W}(K_{\delta_j}) = \mathscr{H}^1_0(K_{\delta_j}) \ (\delta > \varepsilon). \tag{1.2.21}$$

1.2.4 Complexity and optimal macro-micro refinement

Taking N_{mic} elements in each space dimension for the discretization of the sampling domain K_{δ_j}, we have $h = \delta/N_{mic}$ and thus $\hat{h} = (\delta/\varepsilon) \cdot (1/N_{mic})$. Since δ scales with ε, typically $\delta = C\varepsilon$ (where C is a constant of moderate size), we obtain $\hat{h} = (C/N_{mic})$, independent of ε.

We denote by $M_{mac} = \mathscr{O}(\hat{h}^{-d})$ the number of degrees of freedom (DOF) for the micro FEM and by M_{mac}, the number of DOF of the macro FEM. For quasi-uniform macro meshes, the macro meshsize H and the micro meshsize $\hat{h}$ are related to M_{mac} and M_{mic} as

$$H = \mathscr{O}(M^{-1/d}_{mac}), \qquad \hat{h} = \mathscr{O}(M^{-1/d}_{mic}).$$

In view of (1.2.17) and (1.2.18), optimal macroscopic convergence rates (up to a modeling error

[1] For this estimate to hold, one needs to consider a suitable modification of the bilinear form (1.2.15) and the micro problem (1.2.16), namely one has to collocate the term $a(x, \frac{x}{\varepsilon})$ in the slow variable $a(x_{K_j}, \frac{x}{\varepsilon})$ in both (1.2.15) and (1.2.16).

r_{MOD} independent of H, h) are obtained for quasi-uniform microscopic meshsizes given by

$$\hat{h} \simeq H^{\frac{\ell}{2q}} \quad \text{for the } \mathscr{H}^1 \text{ norm}, \quad \hat{h} \simeq H^{\frac{\ell+1}{2q}} \quad \text{for the } \mathscr{L}^2 \text{ norm}.$$

The corresponding complexity in term of macro DOF reads

$$\underbrace{H^{-d}}_{M_{mac}} \cdot \underbrace{H^{\frac{-d\ell}{2q}}}_{M_{mic}} \cdot n_s = (M_{mac})^{1+\frac{\ell}{2q}} \cdot n_s \qquad \text{for the } \mathscr{H}^1 \text{ norm},$$
$$\underbrace{H^{-d}}_{M_{mac}} \cdot \underbrace{H^{\frac{-d(\ell+1)}{2q}}}_{M_{mic}} \cdot n_s = (M_{mac})^{1+\frac{\ell+1}{2q}} \cdot n_s \qquad \text{for the } \mathscr{L}^2 \text{ norm},$$

where n_s denotes the number of sampling domains per macro element $K \in \mathscr{T}_H$. We first observe that in contrast to numerical methods of type 1 (see the introduction) the complexity is *independent* of ε. Second, as can be seen above and as first noticed in [1] the overall complexity of the method is a function of M_{mac} and M_{mic} and in general superlinear with respect to the macro DOF. For example, using piecewise linear polynomials on simplicial FEs, assuming quasi-uniform macro and micro meshes and that the complexity is proportional to the total DOF we obtain a cost of $\mathscr{O}(M_{mac}^{3/2})$ ($\mathscr{H}^1$ norm) and $\mathscr{O}(M_{mac}^2)$ ($\mathscr{L}^2$ norm). [2] In contrast, the memory demand is proportional to $M_{mac} + M_{mic}$ only as the micro problems, being independent of one another, can be solved one at a time. Finally we note that by using spectral methods or $p-FEM$ for the micro solvers can reduce the complexity of the overall FE-HMM (up to log-linear complexity). This was investigated in [15]. Such an approach requires however high regularity in the oscillating tensor a^ε which may not hold for some applications as for example in material science.

[2] Notice that as the micro problems are solved independently, the method is well suited for parallel implementation which can reduce significantly the complexity of the FE-HMM.

2 Fully discrete analysis of the heterogeneous multiscale method for elliptic problems with multiple scales

In this chapter we consider (1.0.1) the multiscale tensor in a generalized form oscillating on $N+1$ separate scales, i.e., $a^\varepsilon = a(x, \frac{x}{\varepsilon_1}, \cdots, \frac{x}{\varepsilon_N})$, and $\varepsilon_1, \ldots, \varepsilon_N$ are N positive functions $\varepsilon_i(\varepsilon)$ that converge to 0 when $\varepsilon \to 0$ and that are well-separated in the sense that $\lim_{\varepsilon\to 0} \frac{\varepsilon_{i+1}(\varepsilon)}{\varepsilon_i(\varepsilon)} = 0$ for $i = 1, \ldots N-1$. The above tensor has thus one macroscopic scale and N microscopic scales, i.e., it varies over $N+1$ scales (for simplicity of indexing the scales in our numerical scheme, we will refer to the macro scale as the zero-th scale). Here, we still assume homogeneous Dirichlet boundary conditions for simplicity and take $f \in \mathscr{L}^2(\Omega)$.

Here we generalize the FE-HMM, and propose a fully discrete analysis for $(N+1)$-scale problems. Recall that the FE-HMM can be seen as a FEM with numerical quadrature for a modified effective problem, as seen in Chapter 1. The data actually recovered by the micro FEMs are a perturbed version of the true effective data, because the computed data depends on the accuracy of the micro solver and the modeling error as we discussed in Chapter 1. For the analysis we have thus to deal with variational crimes (as we have a FEM with numerical quadrature) and modeling error (as the recovered effective problem differs from the true effective model). Yet for two-scale problems, the micro scale was assumed to be solved by standard FEM (the microscopic data are given by model equation (1.0.1)) in the analysis [1, 2], while for problems with more than two scales, such assumption cannot be made as the problems at intermediate scales (called mesoscales in what follows) depend on effective coefficients computed around quadrature points of the meso FE meshes. Hence, we have a cascade of interdepending FEMs with numerical quadrature and a cascade of variational crimes. In turn we need on one hand to generalize the analysis for FEM with numerical quadrature for single scale problem given by [54], on the other hand characterize the propagation of numerical discretization and modeling errors from micro to meso and macro scales. This precise characterization allows to set up the optimal meshes at each scale in order to obtain the desired convergence rate at the macro scale with minimal computational complexity. We note that a complexity analysis of the FE-HMM for $N+1$ scales shows that the method even though much cheaper than the fine scale problem (intractable in general) can be costly, due to the cascade of cell problems to be solved. Nevertheless, the analysis of the FE-HMM for $(N+1)$-scale problems will be an important cornerstone for further research using in addition model reduction technique to the FE-HMM.

This chapter is organized as follows. In Section 2.1, we briefly discuss the $(N+1)$-scale homogenization problem. The FE-HMM for $N+1$ scales is defined in Section 2.2. Existence, uniqueness and a priori error analysis are presented in Section 2.3. The modeling error of the FE-HMM for $N+1$ scales is discussed in Section 2.4, while the estimates for numerical quadrature used in the proof of the main theorems are given in Section 2.5. We conclude this chapter with numerical experiments in Section 2.6.

2.1 Model problem and homogenization

We consider the model equation (1.0.1). Let $\Omega \subset \mathbb{R}^d$ be a bounded polyhedron subset in $\mathbb{R}^d$ and $f \in \mathscr{L}^2(\Omega)$. Assume that $a^\varepsilon(x) \in \mathscr{L}^\infty(\Omega)^{d\times d}$ is uniformly bounded and elliptic (1.0.1). By Lax-Milgram theorem, there exists for all fixed $\varepsilon > 0$ a unique solution u^ε of (1.0.1) which is bounded in $\mathscr{H}^1_0(\Omega)$ uniformly in ε. Hence, by a standard compactness argument, one can show that there exists a subsequence of $\{u^\varepsilon\}$ that converges weakly in $\mathscr{H}^1_0(\Omega)$. Invoking h_0 convergence [80] (or G convergence [56] for the symmetric case) on can show that there exists a tensor $a^0(x) \in \mathscr{L}^\infty(\Omega)^{d\times d}$ that is again elliptic and bounded and a subsequence of $\{u^\varepsilon\}$ that weakly converges in $\mathscr{H}^1_0(\Omega)$ to $u_0 \in \mathscr{H}^1_0(\Omega)$ that is the solution of the problem

$$\begin{aligned} -\nabla\cdot(a^0\nabla u^0) &= f \quad \text{in } \Omega, \\ u^0 &= 0 \quad \text{on } \partial\Omega. \end{aligned} \tag{2.1.1}$$

However in the general case, the limit tensor a^0 is difficult to characterize and might not be unique. If one assumes that

$$a^\varepsilon = a(x, \frac{x}{\varepsilon_1}, \cdots, \frac{x}{\varepsilon_N}) = a(x, y_1, \ldots, y_N) \in \mathscr{L}^\infty\big(\Omega, \mathscr{C}^0(\mathbb{R}^{dN})\big)^{d\times d}, \tag{2.1.2}$$

where

$$a(x, y_1, \ldots, y_N) \text{ is periodic with respect to } y_s \text{ with period } Y = [0,1]^d \text{ for each } s = 1, \ldots, N, \tag{2.1.3}$$

and that $\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_N$ are N positive functions $\varepsilon_s(\varepsilon)$ that converge to 0 when $\varepsilon \to 0$ and are well-separated

$$\lim_{\varepsilon\to 0} \frac{\varepsilon_{s+1}(\varepsilon)}{\varepsilon_s(\varepsilon)} = 0 \text{ for } s = 1, \ldots N-1, \tag{2.1.4}$$

then as proved in [30], then the whole sequence $\{u^\varepsilon\}$ weakly converges in $\mathscr{H}^1_0(\Omega)$ and the homogenized solution and homogenized tensor, u^0, a^0, respectively, are unique. Furthermore, $a^0(x)$ can be obtained by an inductive homogenization formula by computing a cascade of periodic micro functions and related homogenized tensors at the successive mesoscales (see [30, Corollary 2.12]). The FE-HMM for $N+1$ scales will be defined for a general tensor assuming (1.0.1),(2.1.2) and (2.1.4). For a full characterization of the fine scale successive numerical errors we will in addition use (2.1.3). Finally assuming the periodicity of the tensor $a^\varepsilon(x)$ facilitates the

analysis and will also be assumed here. We however note that this assumption could be removed at the cost of introducing dual problems to recover optimal convergence rates for the macro and micro scales (see [57] and [26, Lemma 4.6]).

2.2 FE-HMM for $(N+1)$-scale problems

In this section, we extend and analyze the FE-HMM discussed in Chapter 1 (for two-scale problems) to $(N+1)$-scale problems.

The main idea is to apply a macro FEM to (2.1.1) and introduce mesoscopic FEMs on meso sampling domains to solve the meso cell problems recursively (recovering the unknown homogenized tensor at scale s by cell problems at appropriate quadrature points at scale $s+1$ until the scale N is reached where we use the given oscillatory data) in order to recover the unknown data in the macro solver (we emphasize that a^0 is unknown in general). This is illustrated in Fig. 1 for the three scale FE-HMM.

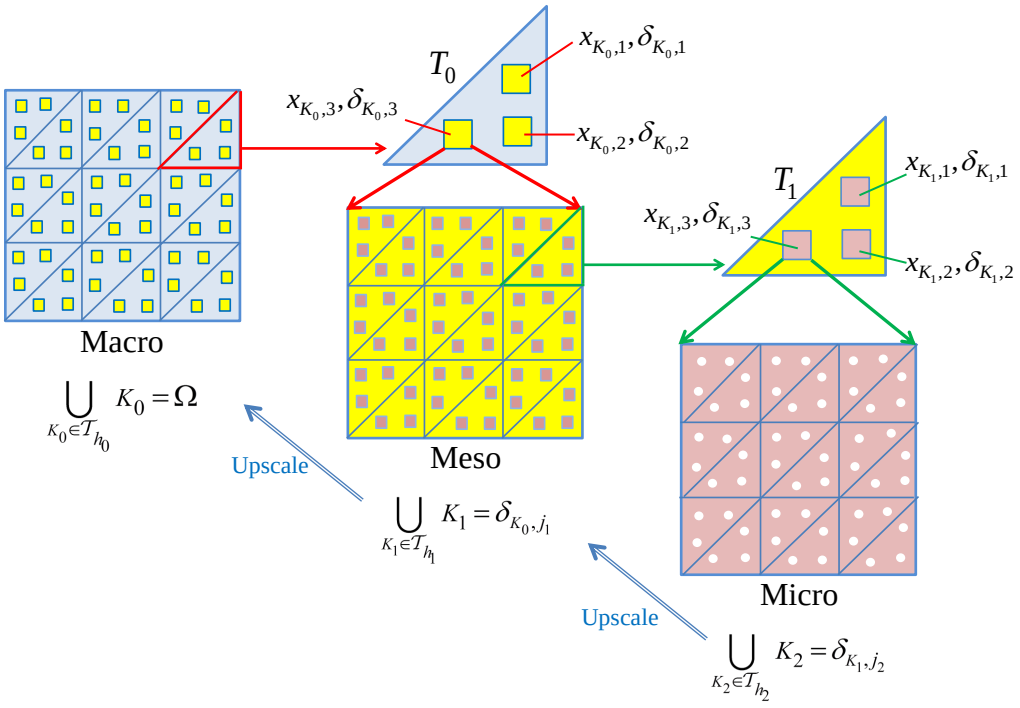


Figure 1: Illustration for the three scale FE-HMM

The FE-HMM is based on a macro finite element (FE) space

$$S_0^{q_0}(\Omega, \mathcal{T}_{h_0}) = \{v^{h_0} \in \mathcal{H}_0^1(\Omega);\ v^{h_0}|_{K_0} \in \mathcal{R}^{q_0}(K_0),\ \forall K_0 \in \mathcal{T}_{h_0}\},$$

where $\mathcal{T}_{h_0}$ is a family of (macro) partition of Ω in simplicial or rectangular (parallelogram) elements K_0 of diameter h_{K_0}, and $\mathscr{R}^{q_0}(K_0)$ is the space $\mathscr{P}^{q_0}(K_0)$ of polynomials on K_0 of total degree at most q_0 if K_0 is a simplicial FE, or the space $\mathscr{Q}^{q_0}(K_0)$ of polynomials on K_0 of degree at most q_0 in each variable if K_0 is a rectangular (parallelogram) FE. For a given macro partition, we define as usual $h_0 := \max_{K_0 \in \mathcal{T}_{h_0}} h_{K_0}$ and assume that the family of triangulation $\mathcal{T}_{h_0}$ is shape regular We note that h_0 in our discretization is allowed to be much larger than ε_1.

Since the quadrature formula (QF) is the one of the main concerns in this chapter, we restate here the definition of QF for convenience: Define a QF $\{\hat{x}_j, \hat{\omega}_j\}_{j=1}^J$ on a reference element $\hat{K}$ coupled with a general FE space $S_D^\ell(\mathscr{D}, \mathcal{T}_h)$ [1]. We make the following assumptions on the quadrature formula, see [53]

(**Q1**) $\hat{\omega}_j > 0,\ j = 1, \ldots, J, \sum_{j=1}^J \hat{\omega}_j |\nabla \hat{p}(\hat{x}_j)|^2 \geq \hat{\lambda} \|\nabla \hat{p}\|^2_{\mathscr{L}^2(\hat{K})},\ \forall \hat{p}(\hat{x}) \in \mathscr{R}^\sigma(\hat{K})$, with $\hat{\lambda} > 0$;

(**Q2**) $\int_{\hat{K}} \hat{p}(\hat{x}) d\hat{x} = \sum_{j=1}^J \hat{\omega}_j \hat{p}(\hat{x}_j),\ \forall \hat{p}(\hat{x}) \in \mathscr{R}^\sigma(\hat{K})$.

where $\sigma = \max(2\ell - 2, \ell)$ if $\hat{K}$ is a simplicial FE, or $\sigma = \max(2\ell - 1, \ell + 1)$ if $\hat{K}$ is a rectangular FE. We choose J in such a way that (**Q2**) is guaranteed.

Remark 2.2.1. *Assumption (**Q1**) is critical for the uniqueness and existence of the numerical solution. Assumption (**Q2**) ensures the precision of the QF. For $\hat{K}$ is a simplicial FE, assumption (**Q2**) indicates that $\sum_{j=1}^J \hat{\omega}_j |\nabla \hat{p}(\hat{x}_j)|^2 = \|\nabla \hat{p}\|^2_{\mathscr{L}^2(\hat{K})}$ and thus (**Q1**) is not needed. Whereas this equality does not hold for rectangular $\hat{K}$ so that* (**Q1**) *is required while one can still show for rectangular $\hat{K}$: $\sum_{j=1}^J \hat{\omega}_j |\nabla \hat{p}(\hat{x}_j)|^2 \leq \hat{C} \|\nabla \hat{p}\|_{\mathscr{L}^2(\hat{K})}$.*

For a given $J_1 \in \mathbb{N}$ and a given QF $\{\hat{x}_{j_1}, \hat{\omega}_{j_1}\}_{j_1=1}^{J_1}$, define the quadrature nodes on each macro element $K_0 \in \mathcal{T}_{h_0}$ by the affine map $x_{K_0, j_1} = F_K(\hat{x}_{j_1}),\ j_1 = 1, \cdots, J_1$ and corresponding quadrature weights $\omega_{K_0, j_1} = |K_0| \hat{\omega}_{j_1}$. Here and in what follows, we denote the measure of a subset $\mathscr{D} \subset \Omega$ by $|\mathscr{D}|$.

We define the first meso scale sampling domains δ_{K_0, j_1} around each x_{K_0, j_1} by

$$\delta_{K_0, j_1} = x_{K_0, j_1} + (-\delta_1/2, \delta_1/2)^d, \text{ with } \delta_1 \geq \varepsilon_1.$$

We then consider a partition $\mathcal{T}_{h_1}$ of δ_{K_0, j_1} with elements K_1 of size h_1. Likewise, we define recursively a sequence of sampling domains

$$\delta_{K_{s-1}, j_s} = x_{K_{s-1}, j_s} + (-\delta_s/2, \delta_s/2)^d, \text{ with } \delta_s \geq \varepsilon_s,\ s = 1, \cdots, N,$$

where x_{K_{s-1}, j_s} are the quadrature nodes on the element $K_{s-1} \in \mathcal{T}_{h_{s-1}}$ and $\mathcal{T}_{h_{s-1}}$ is a partition of

[1] $S_D^\ell(\mathscr{D}, \mathcal{T}_h)$ is the FE space with certain boundary condition. The QF is applied in different FE problems in this chapter. Therefore we define the QF in a general fashion.

$\delta_{K_{s-2},j_{s-1}}$. Here the quadrature nodes are defined through an affine map

$$x_{K_{s-1},j_s} = F_{K_{s-1},j_s}(\hat{x}_{j_s}),\ j_s = 1,\cdots,J_s$$

from the reference quadrature nodes $\{\hat{x}_{j_s}\}_{j_s=1}^{J_s}$ on $\hat{K}$.

Now we define on a scale s ($s = 1,\cdots,N$) an FE space $S^{q_s}(\delta_{K_{s-1},j_s},\mathcal{T}_{h_s}) \subset \mathcal{W}(\delta_{K_{s-1},j_s})$ with simplicial or rectangular FEs and piecewise polynomials of degree q_s. The space $\mathcal{W}(\delta_{K_{s-1},j_s})$ is the Sobolev space

$$\mathcal{W}(\delta_{K_{s-1},j_s}) = \mathcal{W}^1_{per}(\delta_{K_{s-1},j_s}) \tag{2.2.5}$$

for a periodic coupling or

$$\mathcal{W}(\delta_{K_{s-1},j_s}) = \mathcal{H}^1_0(\delta_{K_{s-1},j_s}) \tag{2.2.6}$$

for a coupling with Dirichlet boundary conditions.

We next introduce meso and micro problems to recover an approximation of the unknown data $a^0(x)$ at suitable quadrature macro nodes.

Meso and micro problems. Assuming $a^s_{K_s}(x_{K_s,j_{s+1}})$ is available (the subscript of $a^s_{K_s}(x_{K_s,j_{s+1}})$ indicates that $a^s_{K_s}(x_{K_s,j_{s+1}})$ is an FE-HMM tensor estimated in element K_s). Associated to each sampling domain $\delta_{K_{s-1},j_s},\ s = 1,\cdots,N,\ 1 \le j_m \le J_m,\ m \le s$, we define an effective numerical tensor $a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})$ on the quadrature nodes x_{K_{s-1},j_s} by

$$a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s}) = \frac{1}{|\delta_{K_{s-1},j_s}|}\sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}} \omega_{K_s,j_{s+1}}\, a^s_{K_s}(x_{K_s,j_{s+1}})(I + J^T_{\chi^{h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})}), \tag{2.2.7}$$

where $J_{\chi^{h_s}_{K_{s-1},j_s}(x)}$ is a $d\times d$ Jacobian matrix with entries $(J_{\chi^{h_s}_{K_{s-1},j_s}(x)})_{ik} = \partial_k\big(\chi^{i,h_s}_{K_{s-1},j_s}(x)\big)$ ($\partial_k := \partial/\partial x_k$) and $|\delta_{K_{s-1},j_s}|$ denotes the measure of the sampling domain δ_{K_{s-1},j_s}. We use the superscript T to denote the transpose of the matrix. Here we define by $\chi^{i,h_s}_{K_{s-1},j_s}(x) \in S^{q_s}(\delta_{K_{s-1},j_s},\mathcal{T}_{h_s})$ the solution of the following cell problem: For any test function $z^{h_s} \in S^{q_s}(\delta_{K_{s-1},j_s},\mathcal{T}_{h_s})$, $\chi^{i,h_s}_{K_{s-1},j_s}$ satisfies

$$\textbf{(Ps)}\quad \sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}} \omega_{K_s,j_{s+1}}\, a^s_{K_s}(x_{K_s,j_{s+1}})\Big(\nabla\chi^{i,h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}}) + \mathbf{e}_i\Big)\cdot\nabla z^{h_s}(x_{K_s,j_{s+1}}) = 0,$$

where $\mathbf{e}_i,\ i = 1,\cdots,d$ denote the canonical basis of $\mathbb{R}^d$. We note that (**Ps**) (an FEM with numerical quadrature) is defined recursively as the data $a^s_{K_s}(x_{K_s,j_{s+1}})$ depends on $a^{s+1}_{K_{s+1}}(x_{K_{s+1},j_{s+2}})$ [2]. In the last level of the recursion we use $a^N_{K_N}(x) = a^\varepsilon(x)$, the available microscopic tensor from (1.0.1) and observe that the symmetry of a^ε implies the symmetry of $a^{N-1}_{K_{N-1}}(x_{K_{N-1},j_N})$ and iteratively the

[2] It will be proved in Section 2.3.1 that each of the problems (**Ps**) has a unique solution.

symmetry of $a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})$ $(s=1,\cdots,N)$.

Remark 2.2.2. *From* (2.2.7) *and* (**Ps**)*, we obtain*

$$a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})\mathbf{e}_i\cdot\mathbf{e}_k = \tag{2.2.8}$$
$$\frac{1}{|\delta_{K_{s-1},j_s}|}\sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}}\omega_{K_s,j_{s+1}}a^s_{K_s}(x_{K_s,j_{s+1}})(\chi^{i,h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})+\mathbf{e}_i)\cdot(\chi^{k,h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})+\mathbf{e}_k),$$

from which we see that

$$a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s}) = \tag{2.2.9}$$
$$\frac{1}{|\delta_{K_{s-1},j_s}|}\sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}}\omega_{K_s,j_{s+1}}a^s_{K_s}(x_{K_s,j_{s+1}})(I+J^T_{\chi^{h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})})(I+J^T_{\chi^{h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})}).$$

Macro problem. The macro problem for the FE-HMM is defined as follows: find $u^{h_0}\in S^{q_0}_0(\Omega,\mathcal{T}_{h_0})$ such that

$$(\mathbf{P0})\quad B_{h_0}(u^{h_0},v^{h_0}) = \int_\Omega f v^{h_0}dx\quad \forall v^{h_0}\in S^{q_0}_0(\Omega,\mathcal{T}_{h_0}),$$

where

$$B_{h_0}(u^{h_0},v^{h_0}) = \sum_{K_0\in\mathcal{T}_{h_0}}\sum_{j_1=1}^{J_1}\omega_{K_0,j_1}a^0_{K_0}(x_{K_0,j_1})\nabla u^{h_0}(x_{K_0,j_1})\cdot\nabla v^{h_0}(x_{K_0,j_1}), \tag{2.2.10}$$

and $a^0_{K_0}(x_{K_0,j_1})$ is the macro numerical homogenized tensor given by

$$a^0_{K_0}(x_{K_0,j_1}) := \frac{1}{|\delta_{K_0,j_1}|}\sum_{K_1\in\mathcal{T}_{h_1}}\sum_{j_2=1}^{J_2}\omega_{K_1,j_2}\,a^1_{K_1}(x_{K_1,j_2})(I+J^T_{\chi^{h_1}_{K_0,j_1}(x_{K_1,j_2})}). \tag{2.2.11}$$

Here $(J^T_{\chi^{h_1}_{K_0,j_1}(x)})_{ik} = \partial_k\big(\chi^{i,h_1}_{K_0,j_1}(x)\big)$ and $\chi^{i,h_1}_{K_0,j_1}$ is the solution of cell problem (**Ps**) $(s=1)$. We emphasize again that the symmetry of $a^0_{K_0}(x_{K_0,j_1})$ can be deduced from the symmetry of $a^1_{K_1}(x_{K_1,j_2})$.

2.3 A priori error analysis for the $(N+1)$-scale FE-HMM

Our aim is to obtain the a priori errors $\|u^{h_0}-u^0\|_{\mathscr{L}^2(\Omega)}$ and $\|u^{h_0}-u^0\|_{\mathscr{H}^1(\Omega)}$ for the $(N+1)$-scale FE-HMM $(N\geq 2)$. Let us first show that the $(N+1)$-scale FE-HMM is well-defined.

2.3.1 Existence and uniqueness of the FE-HMM solution.

With the assumption (1.0.1), we have the following lemma.

Lemma 2.3.1. *Assume that the multiscale tensor a^ε is symmetric, uniformly elliptic and bounded, i.e.,* (1.0.1) *holds. Furthermore for each scale s ($s=1,\cdots,N$), we assume that the QF is chosen such that (***Q1***) (***Q2***) hold for $\ell = q_s$ where q_s is the degree of the polynomials in $S^{q_s}(\delta_{K_{s-1},j_s},\mathcal{T}_{h_s})$. Then we have*

$$\sqrt{|\delta_{K_{s-1},j_s}|} \leq \|\mathbf{e}_i + \nabla\chi^{i,h_s}_{K_{s-1},j_s}\|_{\mathscr{L}^2(\delta_{K_{s-1},j_s})} \leq C\sqrt{|\delta_{K_{s-1},j_s}|}. \tag{2.3.12}$$

Moreover for $s=1,\cdots,N$, we have

$$a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})\xi\cdot\xi \geq \tilde{\lambda}|\xi|^2,\ |a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})\xi| \leq \tilde{C}|\xi|,\ \forall\xi\in\mathbb{R}^d, \tag{2.3.13}$$

for any quadrature points x_{K_{s-1},j_s} used at the scale $s-1$ and $\tilde{\lambda},C,\tilde{C}>1$ depend on λ,Λ,d,s,N and the shape regularity of the triangulation. In particular, for $s=1$ we have

$$a^0_{K_0}(x_{K_0,j_1})\xi\cdot\xi \geq \tilde{\lambda}|\xi|^2,\ |a^0_{K_0}(x_{K_0,j_1})\xi| \leq \tilde{C}|\xi|,\ \forall\xi\in\mathbb{R}^d. \tag{2.3.14}$$

Proof. The inequalities (2.3.12) and (2.3.13) need to be proved recursively. We show if (2.3.13) holds for s, then there exists an unique solution $\chi^{i,h_s}_{K_{s-1},j_s}$ of (**Ps**) (for $i=1,\cdots,d$) and further (2.3.12) and (2.3.13) hold for $s-1$. As (2.3.13) is true for $s=N$, Lemma 2.3.1 holds then by induction for all $s=1,\cdots,N$.

Assume therefore that (2.3.13) holds for $a^s_{K_s}(x_{K_s,j_{s+1}})$ and let us first prove (2.3.12) for $a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})$. The lower bound of $\|\mathbf{e}_i+\nabla\chi^{i,h_s}_{K_{s-1},j_s}\|_{\mathscr{L}^2(\delta_{K_{s-1},j_s})}$ is straightforward from the following equality obtained by using the boundary conditions of the cell problem

$$\|e_i+\nabla\chi^{i,h_s}_{K_{s-1},j_s}\|^2_{\mathscr{L}^2(\delta_{K_{s-1},j_s})} = |\delta_{K_{s-1},j_s}| + \|\nabla\chi^{i,h_s}_{K_{s-1},j_s}\|^2_{\mathscr{L}^2(\delta_{K_{s-1},j_s})}.$$

For the upper bound, using assumption (**Q1**), we can write

$$\begin{aligned}
&\tilde{\lambda}\|\nabla\chi^{i,h_s}_{K_{s-1},j_s}(x)+\mathbf{e}_i\|^2_{\mathscr{L}^2(\delta_{K_{s-1},j_s})}\\
&\leq \sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}}\omega_{K_s,j_{s+1}}a^s_{K_s}(x_{K_s,j_{s+1}})(\nabla\chi^{i,h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})+\mathbf{e}_i)\cdot(\nabla\chi^{i,h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})+\mathbf{e}_i)\\
&= \sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}}\omega_{K_s,j_{s+1}}a^s_{K_s}(x_{K_s,j_{s+1}})\mathbf{e}_i\cdot\mathbf{e}_i\\
&- \sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}}\omega_{K_s,j_{s+1}}a^s_{K_s}(x_{K_s,j_{s+1}})\nabla\chi^{i,h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})\cdot\nabla\chi^{i,h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})\\
&\leq C|\delta_{K_{s-1},j_s}| - C_1\|\nabla\chi^{i,h_s}_{K_{s-1},j_s}\|^2_{\mathscr{L}^2(\delta_{K_{s-1},j_s})} \leq C|\delta_{K_{s-1},j_s}|,
\end{aligned}$$

where the constant $C_1>0$.

In what follows we only prove the first inequality of (2.3.13) (the second inequality can be simply obtained by using the upper bound in inequality (2.3.12)). We consider the definition of $a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})$ written in the form of (2.2.9), the symmetry of $a^s_{K_s}(x_{K_s,j_{s+1}})$ and assumption (**Q2**)

we have

$$
\begin{aligned}
&a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})\xi\cdot\xi\\
&=\frac{1}{|\delta_{K_{s-1},j_s}|}\sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}}\omega_{K_s,j_{s+1}}a^s_{K_s}(x_{K_s,j_{s+1}})\big((I+J^T_{\chi^{h_s}_{K_{s-1},j_s}(x_{K_{s-1},j_s})})\xi\big)\cdot\big((I+J^T_{\chi^{h_s}_{K_{s-1},j_s}(x_{K_{s-1},j_s})})\xi\big)\\
&\geq\frac{\lambda}{|\delta_{K_{s-1},j_s}|}\sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}}\omega_{K_s,j_{s+1}}\big((I+J^T_{\chi^{h_s}_{K_{s-1},j_s}(x_{K_{s-1},j_s})})\xi\big)\cdot\big((I+J^T_{\chi^{h_s}_{K_{s-1},j_s}(x_{K_{s-1},j_s})})\xi\big)\\
&\geq\frac{\tilde{\lambda}}{|\delta_{K_{s-1},j_s}|}\int_{\delta_{K_{s-1},j_s}}\big((I+J^T_{\chi^{h_s}_{K_{s-1},j_s}(x)})\xi\big)\cdot\big((I+J^T_{\chi^{h_s}_{K_{s-1},j_s}(x)})\xi\big)dx\\
&\geq\tilde{\lambda}|\xi|^2+\frac{\tilde{\lambda}}{|\delta_{K_{s-1},j_s}|}\int_{\delta_{K_{s-1},j_s}}(J^T_{\chi^{h_s}_{K_{s-1},j_s}(x)}\xi)\cdot(J^T_{\chi^{h_s}_{K_{s-1},j_s}(x)}\xi)\\
&\geq\tilde{\lambda}|\xi|^2.
\end{aligned}
$$

where we have used assumption (Q1) (notice that $\tilde{\lambda}=C\lambda$ when using simplicial FE, see Remark 2.2.1. We note that (2.3.14) can be similarly proved using the FE-HMM reformulation (2.2.11). □

Remark 2.3.2. *For fully discrete error analysis of two scale (macro-micro) problems, the effect of numerical quadrature on the micro solutions is often not considered since the microscopic tensor a^ϵ is given, see [1]. However in practice, we need to apply numerical quadrature to the micro FE problems and in this chapter, we consider the FEM with the numerical quadrature at all scales.*

Using Lemma 2.3.1, we obtain the existence and uniqueness of the numerical solution u^{h_0} of problem (**P0**) by applying the Lax-Milgram theorem.

Theorem 2.3.3. *Assume that the hypothesis of Lemma 2.3.1 holds, then all the cell problems (**Ps**) for $s=1,\cdots,N$ have unique solutions. Furthermore, problem (**P0**) also has a unique solution.*

2.3.2 Error estimates for the $(N+1)$-scale FE-HMM.

Assume u^0 is the exact homogenized solution of (2.1.1) and a^0 is the exact homogenized tensor. For the error analysis, we need to consider the quantity

$$
r_{HMM}:=\sup_{K_0\in\mathcal{T}_{h_0},\ x_{K_0,j_1}\in\Omega}\|a^0(x_{K_0,j_1})-a^0_{K_0}(x_{K_0,j_1})\|_{\mathcal{F}},\tag{2.3.15}
$$

where $a^0_{K_0}(x_{K_0,j_1})$ is defined in (2.2.11). It is also convenient to introduce the FEM with numerical quadrature for the (exact) homogenized problem (2.1.1): Find $u^{0,h_0}\in S^{q_0}_0(\Omega,\mathcal{T}_{h_0})$ such that

$$
B_{0,h_0}(u^{0,h_0},v^{h_0})=\int_\Omega fv^{h_0}dx,\ \forall v^{h_0}\in S^{q_0}_0(\Omega,\mathcal{T}_{h_0}),\tag{2.3.16}
$$

where

$$B_{0,h_0}(v^{h_0}, w^{h_0}) := \sum_{K_0 \in \mathcal{T}_{h_0}} \sum_{j_1=1}^{J_1} \omega_{K_0,j_1} a^0(x_{K_0,j_1}) \nabla v^{h_0}(x_{K_0,j_1}) \cdot \nabla w^{h_0}(x_{K_0,j_1}) dx. \tag{2.3.17}$$

We emphasize that as the homogenized tensor $a^0(x)$ is unknown (and depends on $N+1$ scales) the equation (2.3.16) cannot be used in practice. It is nevertheless useful for the analysis of the FE-HMM.

Theorem 2.3.4. *Assume that (**Q1**),(**Q2**) hold for the macro QF with $\ell = q_0$ and $\sigma = \max(2q_0-2, q_0)$ for simplicial macro elements or $\sigma = \max(2q_0-1, q_0)$ for rectangular macro elements. Assume that $u^0 \in \mathcal{H}^{q_0+1}(\Omega)$ and that either $a^0(x) \in \mathcal{W}^{q_0,\infty}(\Omega)$ for the $\mathcal{H}^1$ norm estimate or $a^0(x) \in \mathcal{W}^{q_0+1,\infty}(\Omega)$ for the $\mathcal{L}^2$ norm estimate. Then we have*

$$\|u^0 - u^{h_0}\|_{\mathcal{H}^1(\Omega)} \leq C({h_0}^{q_0} + r_{HMM}), \tag{2.3.18}$$

$$\|u^0 - u^{h_0}\|_{\mathcal{L}^2(\Omega)} \leq C({h_0}^{q_0+1} + r_{HMM}), \tag{2.3.19}$$

where the constant C is independent of h_0, h_s, ε_s for $s = 1, \cdots, N$.

Proof. We first decompose the error term

$$\|u^0 - u^{h_0}\| \leq \|u^0 - u^{0,h_0}\| + \|u^{0,h_0} - u^{h_0}\|, \tag{2.3.20}$$

where $\|\cdot\|$ stands for the $\mathcal{H}^1$ norm or $\mathcal{L}^2$ norm and u^{0,h_0} is the FE solution of (2.3.16). By the standard error analysis (see [54]) of FE method with numerical quadrature owning to the assumption (**Q1**), (**Q2**) and the regularity of u^0 and a^0, the first term of (2.3.20) can be bounded by

$$\|u^0 - u^{0,h_0}\| \leq C{h_0}^p,$$

where $p = q_0$ for the $\mathcal{H}^1$ norm and $p = q_0 + 1$ for the $\mathcal{L}^2$ norm.

Furthermore, by the ellipticity of $a^0_{K_0}$ (Lemma 2.3.1) and the bound $\|\nabla u^{0,h_0}\|_{\mathcal{L}^2(\Omega)} \leq C\|f\|_{\mathcal{L}^2(\Omega)}$, using the reformulation (2.2.10), we have

$$\begin{aligned}
\lambda\|\nabla u^{0,h_0} - \nabla u^{h_0}\|^2_{\mathcal{L}^2(\Omega)} &\leq B_{h_0}(u^{0,h_0} - u^{h_0}, u^{0,h_0} - u^{h_0}) \\
&= B_{h_0}(u^{0,h_0}, u^{0,h_0} - u^{h_0}) - \int_\Omega f(u^{0,h_0} - u^{h_0}) dx \\
&= B_{h_0}(u^{0,h_0}, u^{0,h_0} - u^{h_0}) - B_{0,h_0}(u^{0,h_0}, u^{0,h_0} - u^{h_0}) \\
&\leq C \sup_{K_0 \in \mathcal{T}_{h_0}, x_{K_0,j_1} \in \Omega} \|a^0(x_{K_0,j_1}) - a^0_{K_0}(x_{K_0,j_1})\|_{\mathscr{F}} \|\nabla u^{0,h_0} - \nabla u^{h_0}\|_{\mathcal{L}^2(\Omega)}.
\end{aligned}$$

Therefore, we obtain

$$\|\nabla u^{0,h_0} - \nabla u^{h_0}\|_{\mathscr{L}^2(\Omega)} \leq C r_{HMM}.$$

Using the last inequality and the Poincaré inequality gives the stated results. □

2.4 A priori error estimates for r_{HMM}

The quantity r_{HMM} comprises two types of errors: the error originating from the meso and the micro discretization and the modeling error (mismatch of boundary conditions in the various coupling, resonance errors, etc.). The contribution to the term r_{HMM} coming from the micro and the meso discretization error can be quantified assuming appropriate regularity of a^ε. We assume (2.1.2),(2.1.3) and (2.1.4) and for convenience we will denote the these assumptions as assumption (**H1**).

Under the assumption (**H1**), as mentioned in Section 2.1, homogenization results [30] give an explicit characterization of the $s-1$ scale homogenized tensor $a^{s-1}(x, y_1, \cdots, y_{s-1})$ which can be proved provided the knowledge of $a^s(x, y_1, \cdots, y_s)$, where $s = 1, \cdots, N$ (we denote $a^N(x, y_1, \cdots, y_N) = a(x, y_1, \cdots, y_N) = a^\varepsilon(x)$) and periodicity of a^{s-1} with respect to $y_i, i = 1, \cdots, s-1$. Assuming $\delta_s/\varepsilon_s \in \mathbb{N}$, we denote $\underline{x}_{s-1} := (x_{K_0,j_1}, \frac{x_{K_1,j_2}}{\varepsilon_1}, \cdots, \frac{x_{K_{s-1},j_s}}{\varepsilon_{s-1}})$ and $\underline{x}_0 = x_{K_0,j_1}$ to shorten the notation. We define

$$a^{s-1}(\underline{x}_{s-1}) = \frac{1}{|\delta_{K_{s-1},j_s}|} \int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})(I + J^T_{\chi_{K_{s-1},j_s}(x)})\, dx, \tag{2.4.21}$$

where $(J^T_{\chi_{K_{s-1},j_s}(x)})_{ik} = \partial_k(\chi^i_{K_{s-1},j_s}(x))$ and $\chi^i_{K_{s-1},j_s}(x) \in \mathscr{W}^1_{per}(\delta_{K_{s-1},j_s})$, for $k = 1, \cdots, d$ is the exact solution of the s scale cell problem

$$\int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})(\nabla \chi^i_{K_{s-1},j_s}(x) + \mathbf{e}_i) \cdot \nabla z\, dx = 0, \quad \forall z \in \mathscr{W}^1_{per}(\delta_{K_{s-1},j_s}). \tag{2.4.22}$$

In turn, the exact macro homogenized tensor $a^0(x_{K_0,j_1})$ at the macro quadrature point x_{K_0,j_1} can be computed as (here we require $\delta_1/\varepsilon_1 \in \mathbb{N}$)

$$a^0(x_{K_0,j_1}) = \frac{1}{|\delta_{K_0,j_1}|} \int_{\delta_{K_0,j_1}} a^1(x_{K_0,j_1}, \frac{x}{\varepsilon_1})(I + J^T_{\chi_{K_0,j_1}(x)})\, dx, \tag{2.4.23}$$

where $(J_{\chi_{K_0,j_1}(x)})_{ik} = \partial_k\big(\chi^i_{K_0,j_1}(x)\big)$ and $\chi^i_{K_0,j_1}(x) \in \mathscr{W}^1_{per}(\delta_{K_0,j_1})$ is the solution of the cell problem (2.4.22) (with s=1).

As discussed in [4], appropriate regularity assumptions are required for the functions $\chi^i_{K_{s-1},j_s}$ defined in (2.4.22) $s = 1, \cdots, N$. We assume

(**H2**) for given positive integers q_s, the cell solutions $\chi^i_{K_{s-1},j_s}, s = 1,\cdots,N$ satisfy

$$|\chi^i_{K_{s-1},j_s}|_{\mathscr{H}^{q_s+1}(\delta_{K_{s-1},j_s})} \leq C\varepsilon_s^{-q_s}\sqrt{|\delta_{K_{s-1},j_s}|}.$$

Remark 2.4.1. *Similarly to* (2.3.12)*, one can show (without using (***H2***))*

$$\|\nabla\chi^i_{K_{s-1},j_s} + \mathbf{e}_i\|_{\mathscr{L}^2(\delta_{K_{s-1},j_s})} \leq C\sqrt{|\delta_{K_{s-1},j_s}|} \tag{2.4.24}$$

where the constant C only depends on λ and Λ introduced in (1.0.1).

In this subsection, we assume that the meso FE space $S^{q_s}(\delta_{K_{s-1},j_s},\mathscr{T}_{h_s})$ is a subspace of $\mathscr{W}_{per}(\delta_{K_{s-1},j_s})$ or $\mathscr{H}^1_0(\delta_{K_{s-1},j_s})$ (see (2.2.5) and (2.2.6)). In order to distinguish the FE spaces with different boundary conditions, we denote the meso and micro FE spaces by $S^{q_s}(\delta_{K_{s-1},j_s},\mathscr{T}_{h_s})$ when included in $\mathscr{W}_{per}(\delta_{K_{s-1},j_s})$ (periodic boundary coupling) and $S^{q_s}_0(\delta_{K_{s-1},j_s},\mathscr{T}_{h_s})$ when included in $\mathscr{H}^1_0(\delta_{K_{s-1},j_s})$ (Dirichlet boundary coupling).

Before starting the analysis, we need to define several notations. For $s = 1,\cdots,N$, we denote $\bar{\chi}^{i,h_s}_{K_{s-1},j_s}(x) \in S^{q_s}(\delta_{K_{s-1},j_s},\mathscr{T}_{h_s})$ or $S^{q_s}_0(\delta_{K_{s-1},j_s},\mathscr{T}_{h_s})$ the solution of

$$\sum_{K_s\in\mathscr{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}}\omega_{K_s,j_{s+1}}a^s(\underline{x}_{s-1},\frac{x_{K_s,j_{s+1}}}{\varepsilon_s})(\nabla\bar{\chi}^{i,h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})+\mathbf{e}_i)\cdot\nabla z^{h_s} = 0, \tag{2.4.25}$$

where z^{h_s} is an arbitrary test function in $S^{q_s}(\delta_{K_{s-1},j_s},\mathscr{T}_{h_s})$ or $S^{q_s}_0(\delta_{K_{s-1},j_s},\mathscr{T}_{h_s})$. We then define the tensor $\bar{a}^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})$ as

$$\bar{a}^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s}) = \frac{1}{|\delta_{K_{s-1},j_s}|}\sum_{K_s\in\mathscr{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}}\omega_{K_s,j_{s+1}}a^s(\underline{x}_{s-1},\frac{x_{K_s,j_{s+1}}}{\varepsilon_s})(I+J^T_{\bar{\chi}^{h_s}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})}). \tag{2.4.26}$$

We observe from (2.4.26) that $\bar{a}^{N-1}_{K_{N-1}}(x_{K_{N-1},j_N}) = a^{N-1}_{K_{N-1}}(x_{K_{N-1},j_N})$, where $a^{N-1}_{K_{N-1}}(x_{K_{N-1},j_N})$ is defined in (2.2.7) since $a^N = a^\varepsilon$ is the tensor given in (1.0.1).

We will use the following lemma in the proof of Theorem 2.4.6.

Lemma 2.4.2. *Assume (***H2***) holds, then we have the following upper bound for $\bar{\chi}^{i,h_s}_{K_{s-1},j_s}$, $s = 1,\cdots,N$ (defined in* (2.4.25)*)*

$$\|\nabla\bar{\chi}^{i,h_s}_{K_{s-1},j_s} + \mathbf{e}_i\|_{\mathscr{L}^2(\delta_{K_{s-1},j_s})} \leq C\sqrt{|\delta_{K_{s-1},j_s}|}, \tag{2.4.27}$$

where the constant C in (2.4.27) *is independent of H, h_s, ε_s.*

The proof of this lemma follows the proof of Lemma 2.3.1, by using the ellipticity of $a^\varepsilon = a(x, y_1,\cdots,y_N)$ (see assumption (**H1**)).

Remark 2.4.3. *Using the uniform boundedness and ellipticity of the tensor $a(x, y_1, \cdots, N)$, one can show that $a^{s-1}(x, y_1, \cdots, y_{s-1}), \bar{a}^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s}), s = 1, \cdots, N$ are also elliptic and bounded.*

Lemma 2.4.4. *Assume (**H1**), (**H2**) and that a^ε is symmetric, uniformly bounded and elliptic. Assume further that the FE space for (**Ps**) is $S^{q_s}(\delta_{K_{s-1},j_s}, \mathcal{T}_{h_s})$ and that $\delta_s/\varepsilon_s \in \mathbb{N}$. In addition, we assume that $a^s(\cdot, y_s) \in \mathcal{W}^{2q_s}(Y_s)$ and that assumption (**Q2**) (with $\ell = q_s$) holds for the QF coupled with the FE space $S^{q_s}(\delta_{K_{s-1},j_s}, \mathcal{T}_{h_s})$. We further require either Condition (1) or (2) in the following for all the meso and micro FEs:*

*(1) $\sigma = \max(2q_s - 1, q_s)$ in (**Q2**) if K_s are simplicial FEs, or $\sigma = \max(2q_s - 1, q_s + 1)$ if K_s are rectangular (parallelogram) FEs and the quasi-uniform mesh is applied, i.e., $\frac{h_s}{h_{K_s}} \le C$, for all $K_s \in \mathcal{T}_{h_s}$;*

*(2) $\sigma = \max(4q_s - 3, q_s)$ in (**Q2**) if K_s are simplicial FEs, or $\sigma = \max(4q_s - 3, q_s + 1)$ if K_s are rectangular (parallelogram) FEs.*

Then we have for $s = 1, \cdots, N$

$$\| a^{s-1}(\underline{x}_{s-1}) - \bar{a}^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s}) \|_{\mathscr{F}} \le C(\frac{h_s}{\varepsilon_s})^{2q_s}, \tag{2.4.28}$$

and in particular for $s = 1$ we have

$$\| a^0(x_{K_0,j_1}) - \bar{a}^0_{K_0}(x_{K_0,j_1}) \|_{\mathscr{F}} \le C(\frac{h_1}{\varepsilon_1})^{2q_1}. \tag{2.4.29}$$

Proof. We first define an auxiliary tensor

$$\hat{a}^{s-1}_{K_s}(x_{K_{s-1},j_s})\mathbf{e}_i \cdot \mathbf{e}_k = \frac{1}{|\delta_{K_{s-1},j_s}|} \int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})(\nabla \bar{\chi}^{i,h_s}_{K_{s-1},j_s}(x) + \mathbf{e}_i) \cdot (\nabla \bar{\chi}^{k,h_s}_{K_{s-1},j_s}(x) + \mathbf{e}_k) dx,$$

where $\bar{\chi}^{i,h_s}_{K_{s-1},j_s}$ (respectively $\bar{\chi}^{k,h_s}_{K_{s-1},j_s}$) is solution of (2.4.25). Then we consider the decomposition

$$\begin{aligned} \| a^{s-1}(\underline{x}_{s-1}) - \bar{a}^{s-1}_{K_s}(x_{K_{s-1},j_s}) \|_{\mathscr{F}} &\le \| a^{s-1}(\underline{x}_{s-1}) - \hat{a}^{s-1}_{K_s}(x_{K_{s-1},j_s}) \|_{\mathscr{F}} \\ &+ \| \hat{a}^{s-1}_{K_s}(x_{K_{s-1},j_s}) - \bar{a}^{s-1}_{K_s}(x_{K_{s-1},j_s}) \|_{\mathscr{F}}. \end{aligned} \tag{2.4.30}$$

Using (2.4.22) and the symmetry of $a^s(x, y_1, \cdots, y_s)$, we have

$$\begin{aligned} &\big(a^{s-1}(\underline{x}_{s-1}) - \hat{a}^{s-1}_{K_s}(x_{K_{s-1},j_s})\big)\mathbf{e}_i \cdot \mathbf{e}_k \\ &= \frac{1}{|\delta_{K_{s-1},j_s}|} \int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})(\nabla \chi^i_{K_{s-1},j_s}(x) - \nabla \bar{\chi}^{i,h_s}_{K_{s-1},j_s}(x)) \cdot (\nabla \chi^k_{K_{s-1},j_s}(x) + \mathbf{e}_k) dx \\ &+ \frac{1}{|\delta_{K_{s-1},j_s}|} \int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})(\nabla \bar{\chi}^{i,h_s}_{K_{s-1},j_s}(x) + \mathbf{e}_i) \cdot (\nabla \chi^k_{K_{s-1},j_s}(x) - \nabla \bar{\chi}^{k,h_s}_{K_{s-1},j_s}(x)) dx \end{aligned}$$

$$= \frac{1}{|\delta_{K_{s-1},j_s}|}\int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})(\nabla\bar{\chi}^{i,h_s}_{K_{s-1},j_s}(x)-\nabla\chi^i_{K_{s-1},j_s}(x))\cdot(\nabla\chi^k_{K_{s-1},j_s}(x)-\nabla\bar{\chi}^{k,h_s}_{K_{s-1},j_s}(x))dx.$$

Applying Cauchy-Schwarz inequality, the standard FE a priori error estimate for FEM with numerical quadrature gives $\|\nabla\chi^i_{K_{s-1},j_s}(x)-\nabla\bar{\chi}^{i,h_s}_{K_{s-1},j_s}(x)\|_{\mathscr{L}^2(\delta_{K_{s-1},j_s})} \le Ch_s^{q_s}|\chi^i_{K_{s-1},j_s}(x)|_{\mathscr{H}^{q_s+1}(\delta_{K_{s-1},j_s})}$ and using assumption (**H2**), we obtain

$$\|a^{s-1}(\underline{x}_{s-1})-\hat{a}^{s-1}_{K_s}(x_{K_{s-1},j_s})\|_{\mathscr{F}} \le C(\frac{h_s}{\varepsilon_s})^{2q_s}.$$

The second term of (2.4.30) is estimated using Theorem 2.5.3 for simplicial elements or Theorem 2.5.9 for parallelogram elements provided Condition (1) (or Theorem 2.5.5 for simplicial elements and similar arguments for parallelogram elements, see Remark 2.5.10 provided Condition (2)) given in Section 2.5. We obtain [3]

$$\|\hat{a}^{s-1}_{K_s}(x_{K_{s-1},j_s})-\bar{a}^{s-1}_{K_s}(x_{K_{s-1},j_s})\|_{\mathscr{F}} \le C(\frac{h_s}{\varepsilon_s})^{2q_s}. \tag{2.4.31}$$

The proof of (2.4.28) is complete. □

In order to address a corresponding lemma for the solution of the micro problem using Dirichelet boundary coupling, we first define $\xi^i_{K_{s-1},j_s}(x)\in\mathscr{H}^1_0(\delta_{K_{s-1},j_s})$ the exact solution of (2.4.22) with test function in $\mathscr{H}^1_0(\delta_{K_{s-1},j_s})$ (note that for Dirichlet boundary coupling we do not assume $\delta_{s-1}/\varepsilon_{s-1}\in\mathbb{N}$. We state the following assumption which will be used in Lemma 2.4.5,

(**H3**) for the sampling domain δ_{K_{s-1},j_s}, $s=1,\cdots,N$, assume $\xi^i_{K_{s-1},j_s}(x)\in\mathscr{W}^{1,\infty}(\delta_{K_{s-1},j_s})$, $i=1,\cdots,d$, where $\xi^i_{K_{s-1},j_s}(x)$ is defined above and $\chi^i_{K_{s-1},j_s}(x)$ is defined in (2.4.22).

Lemma 2.4.5. *Assume (**H1**), (**H2**), (**H3**) and that a^ε is symmetric, uniformly bounded and elliptic. Assume further that the FE space for (**Ps**) is $S^{q_s}_0(\delta_{K_{s-1},j_s},\mathscr{T}_{h_s})$ and that $\delta_s>\varepsilon_s$. In addition, we assume that $a^s(\cdot,y_s)\in\mathscr{W}^{2q_s}(Y_s)$ and that assumption (**Q2**) ($\ell=q_s$) holds for the QF coupled with the FE space $S^{q_s}(\delta_{K_{s-1},j_s},\mathscr{T}_{h_s})$. We further require either Condition (1) or (2) in the following for all the meso and micro FEs:*

*(1) $\sigma=\max(2q_s-1,q_s)$ in (**Q2**) if K_s are simplicial FEs, or $\sigma=\max(2q_s-1,q_s+1)$ if K_s are rectangular (parallelogram) FEs and the quasi-uniform mesh is applied, i.e., $\frac{h_s}{h_{K_s}}\le C$, for all $K_s\in\mathscr{T}_{h_s}$;*

*(2) $\sigma=\max(4q_s-3,q_s)$ in (**Q2**) if K_s are simplicial FEs, or $\sigma=\max(4q_s-3,q_s+1)$ if K_s are rectangular (parallelogram) FEs.*

[3] Notice that in Theorem 2.5.3 (or Theorem 2.5.5) or Theorem 2.5.9 from Section 2.5, we let $\int_{\mathscr{D}} a\nabla u\cdot\nabla v dx = \int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})\nabla u\cdot\nabla v dx$, $F_1(v)=\int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})\mathbf{e}_i\cdot\nabla v dx$ and $F_2(v)=\int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})\mathbf{e}_k\cdot\nabla v dx$.

Then we have for $s = 1, \cdots, N$,

$$\| a^{s-1}(\underline{x}_{s-1}) - \bar{a}^{s-1}_{K_s}(x_{K_{s-1},j_s}) \|_{\mathscr{F}} \leq C\big((\frac{h_s}{\varepsilon_s})^{2q_s} + \frac{\varepsilon_s}{\delta_s}\big), \tag{2.4.32}$$

where $\bar{a}^{s-1}_{K_s}(x_{K_{s-1},j_s})$ *is defined in* (2.4.26).

Proof. In view of (2.4.21) and Remark 2.2.2 we have

$$a^{s-1}(\underline{x}_{s-1})\mathbf{e}_i \cdot \mathbf{e}_k = \frac{1}{|\varepsilon_{K_{s-1},j_s}|} \int_{\varepsilon_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})\big(\chi^i_{K_{s-1},j_s}(x) + \mathbf{e}_i\big) \cdot \big(\chi^k_{K_{s-1},j_s}(x) + \mathbf{e}_k\big) dx, \tag{2.4.33}$$

where $\varepsilon_{K_{s-1},j_s}$ a meso sampling domain centered at x_{K_{s-1},j_s} which covers the maximum number of ε_s period in each direction contained in the domain δ_{K_{s-1},j_s}, i.e. $|\varepsilon_{K_{s-1},j_s}| = (N\varepsilon_s)^d$, $N \in \mathbb{N}$ (note that $\delta_{s-1}/\varepsilon_{s-1}$ may not belong to $\mathbb{N}$). We define the following tensor

$$\hat{a}^{s-1}(\underline{x}_{s-1})\mathbf{e}_i \cdot \mathbf{e}_k = \frac{1}{|\delta_{K_{s-1},j_s}|} \int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})\big(\xi^i_{K_{s-1},j_s}(x) + \mathbf{e}_i\big) \cdot \big(\xi^k_{K_{s-1},j_s}(x) + \mathbf{e}_k\big) dx \tag{2.4.34}$$

based on the cell functions $\xi^i_{K_{s-1},j_s}(x)$ defined above, and further decompose

$$\begin{aligned} \| a^{s-1}(\underline{x}_{s-1}) - \bar{a}^{s-1}_{K_{s-1},j_s}(x_{K_{s-1},j_s}) \|_{\mathscr{F}} \quad &\leq \quad \| a^{s-1}(\underline{x}_{s-1}) - \hat{a}^{s-1}(\underline{x}_{s-1}) \|_{\mathscr{F}} \\ &+ \quad \| \hat{a}^{s-1}(\underline{x}_{s-1}) - \bar{a}^{s-1}_{K_{s-1},j_s}(x_{K_{s-1},j_s}) \|_{\mathscr{F}}. \end{aligned} \tag{2.4.35}$$

Similarly to the proof of Lemma 2.4.4 (see e.g. (2.4.30)), one can deduce that the second term of (2.4.35) can be bounded as

$$\| \hat{a}^{s-1}(\underline{x}_{s-1}) - \bar{a}^{s-1}_{K_s}(x_{K_{s-1},j_s}) \|_{\mathscr{F}} \leq C(\frac{h_s}{\varepsilon_s})^{2q_s}. \tag{2.4.36}$$

For the first term of (2.4.35), one needs to apply a boundary corrector because of the mismatch between the Dirichlet and the periodic boundary conditions of the cell problem. This has first been studied in [60] for the FE-HMM. We give here a short proof for completeness.

We first write

$$\begin{aligned} &|\big(a^{s-1}(\underline{x}_{s-1}) - \hat{a}^{s-1}(\underline{x}_{s-1})\big)\mathbf{e}_i \cdot \mathbf{e}_k| \\ &\leq \Big| \frac{1}{|\delta_{K_{s-1},j_s}|} \int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})(\nabla\chi^i_{K_{s-1},j_s}(x) + \mathbf{e}_i) \cdot (\nabla\xi^k_{K_{s-1},j_s}(x) + \mathbf{e}_k) dx - \hat{a}^s(\underline{x}_{s-1})\mathbf{e}_i \cdot \mathbf{e}_k \Big| \\ &+ \Big| a^{s-1}(\underline{x}_{s-1})\mathbf{e}_i \cdot \mathbf{e}_k - \frac{1}{|\delta_{K_{s-1},j_s}|} \int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})(\nabla\chi^i_{K_{s-1},j_s}(x) + \mathbf{e}_i) \cdot (\nabla\xi^k_{K_{s-1},j_s}(x) + \mathbf{e}_k) dx \Big| \\ &:= I + II. \end{aligned}$$

Let $\Gamma_{K_{s-1},j_s} = \delta_{K_{s-1},j_s} \setminus \varepsilon_{K_{s-1},j_s}$ be the boundary layer with $|\Gamma_{K_{s-1},j_s}| = C\delta_s^{d-1}\varepsilon_s$ and $\frac{|\Gamma_{K_{s-1},j_s}|}{|\delta_{K_{s-1},j_s}|} \leq C\frac{\varepsilon_s}{\delta_s}$. By assumption (**H3**), one can derive

$$\|\nabla\chi^i_{K_{s-1},j_s}(x) + \mathbf{e}_i\|_{\mathscr{L}^2(\Gamma_{K_{s-1},j_s})} \leq C\sqrt{|\Gamma_{K_{s-1},j_s}|}, \ \ \|\nabla\xi^k_{K_{s-1},j_s}(x) + \mathbf{e}_k\|_{\mathscr{L}^2(\Gamma_{K_{s-1},j_s})} \leq C\sqrt{|\Gamma_{K_{s-1},j_s}|}. \tag{2.4.37}$$

We next define function $\rho^{\varepsilon_s}_{K_{s-1},j_s} \in \mathscr{C}^\infty(\delta_{K_{s-1},j_s})$; $0 \leq \rho^{\varepsilon_s}_{K_{s-1},j_s} \leq 1$ with the following properties

$$\rho^{\varepsilon_s}_{K_{s-1},j_s}(x) = \begin{cases} 1 & dist(x,\partial\delta_{K_{s-1},j_s}) > 2\varepsilon_s \\ 0 & dist(x,\partial\delta_{K_{s-1},j_s}) < \varepsilon_s \end{cases} \tag{2.4.38}$$

and $\varepsilon_s\|\nabla\rho^{\varepsilon_s}_{K_{s-1},j_s}\|_{\mathscr{L}^\infty(\delta_{K_{s-1},j_s})} \leq C$, where C is independent of ε_s. For the construction of functions with such properties we refer for example to [74].

We then introduce the boundary corrector $\theta^i_{K_{s-1},j_s} := \xi^i_{K_{s-1},j_s} - \chi^i_{K_{s-1},j_s}$ which satisfies $\theta^i_{K_{s-1},j_s} + (1-\rho^{\varepsilon_s}_{K_{s-1},j_s})\chi^i_{K_{s-1},j_s} \in \mathscr{H}^1_0(\delta_{K_{s-1},j_s})$. Thus, by noticing that

$$\hat{a}^{s-1}(\underline{x}_{s-1})\mathbf{e}_i\cdot\mathbf{e}_k = \frac{1}{|\delta_{K_{s-1},j_s}|}\int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})(\xi^i_{K_{s-1},j_s}(x)+\mathbf{e}_i)\cdot(\xi^k_{K_{s-1},j_s}(x)+\mathbf{e}_k)dx,$$

we have

$$\begin{aligned} I \leq & \frac{1}{|\delta_{K_{s-1},j_s}|}\Big|\int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})\nabla(\theta^i_{K_{s-1},j_s} + \chi^i_{K_{s-1},j_s}(1-\rho^{\varepsilon_s}_{K_{s-1},j_s}))\cdot(\nabla\xi^k_{K_{s-1},j_s}+\mathbf{e}_k)dx\Big| \\ & + \frac{1}{|\delta_{K_{s-1},j_s}|}\Big|\int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})\nabla(\chi^i_{K_{s-1},j_s}(1-\rho^{\varepsilon_s}_{K_{s-1},j_s}))\cdot(\nabla\xi^k_{K_{s-1},j_s}+\mathbf{e}_k)dx\Big|. \end{aligned}$$

The first term on the right-hand side of the above inequality above vanishes since $\theta^i_{K_{s-1},j_s} + \chi^i_{K_{s-1},j_s}(1-\rho^{\varepsilon_s}_{K_{s-1},j_s}) \in \mathscr{H}^1_0(\delta_{K_{s-1},j_s})$, $a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})$ is symmetric and $\xi^k_{K_{s-1},j_s}$ is the solution of (2.4.22) in $\mathscr{H}^1_0(\delta_{K_{s-1},j_s})$. Then we have

$$\begin{aligned} I \leq & \frac{1}{|\delta_{K_{s-1},j_s}|}\Big|\int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})\big(\nabla\chi^i_{K_{s-1},j_s}(1-\rho^{\varepsilon_s}_{K_{s-1},j_s})\big)\cdot(\nabla\xi^k_{K_{s-1},j_s}+\mathbf{e}_k)dx\Big| \\ & + \frac{1}{|\delta_{K_{s-1},j_s}|}\Big|\int_{\delta_{K_{s-1},j_s}} a^s(\underline{x}_{s-1},\frac{x}{\varepsilon_s})(\chi^i_{K_{s-1},j_s}\nabla\rho^{\varepsilon_s}_{K_{s-1},j_s})\cdot(\nabla\xi^k_{K_{s-1},j_s}+\mathbf{e}_k)dx\Big| \\ & \leq C\frac{1}{|\delta_{K_{s-1},j_s}|}\|\nabla\xi^k_{K_{s-1},j_s}+\mathbf{e}_k\|_{\mathscr{L}^2(\Gamma_{K_{s-1},j_s})}\Big(\|\nabla\chi^i_{K_{s-1},j_s}\|_{\mathscr{L}^2(\Gamma_{K_{s-1},j_s})} + \frac{1}{\varepsilon_s}\|\chi^i_{K_{s-1},j_s}\|_{\mathscr{L}^2(\Gamma_{K_{s-1},j_s})}\Big) \\ & \leq C\frac{\varepsilon_s}{\delta_s}, \end{aligned}$$

where the last inequality is obtained by (2.4.37) and the fact that $\chi^i_{K_{s-1},j_s}(x) = \chi^i_{K_{s-1},j_s}(x_{T_{s-1},j_s} + \varepsilon_s y_s) = \varepsilon_s\tilde{\chi}^i_{K_{s-1},j_s}(y_s)$ where $\tilde{\chi}^i_{K_{s-1},j_s}(y_s)$ is the solution of (2.4.22) on the reference cell Y obtained by the affine mapping defined as $y_s = (x - x_{T_{s-1},j_s})/(N\varepsilon_s)$ from $\varepsilon_{K_{s-1},j_s}$ to Y (observe that $\|\tilde{\chi}^i_{K_{s-1},j_s}\|_{L^2(Y)} \leq C$).

For the term II, we have

$$\begin{aligned} II \leq & \frac{|\delta_{K_{s-1},j_s}| - |\varepsilon_{K_{s-1},j_s}|}{|\delta_{K_{s-1},j_s}|} \frac{1}{|\varepsilon_{K_{s-1},j_s}|} | \int_{\varepsilon_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})(\nabla \chi^i_{K_{s-1},j_s}(x) + \mathbf{e}_i) \cdot \nabla \theta^k_{K_{s-1},j_s} dx| \\ & + \frac{1}{|\delta_{K_{s-1},j_s}|} | \int_{\Gamma_{K_{s-1},j_s}} a^s(\underline{x}_{s-1}, \frac{x}{\varepsilon_s})(\nabla \chi^i_{K_{s-1},j_s}(x) + \mathbf{e}_i) \cdot (\nabla \xi^k_{K_{s-1},j_s}(x) + \mathbf{e}_k) dx| \\ & := (a) + (b). \end{aligned}$$

Finally we have

$$\begin{aligned} (a) \quad & \leq \quad C \frac{|\delta_{K_{s-1},j_s}| - |\varepsilon_{K_{s-1},j_s}|}{|\delta_{K_{s-1},j_s}|} \frac{1}{|\varepsilon_{K_{s-1},j_s}|} \|\nabla \chi^i_{K_{s-1},j_s} + \mathbf{e}_i\|_{\mathscr{L}^2(\varepsilon_{K_{s-1},j_s})} \|\nabla \theta^k_{K_{s-1},j_s}\|_{\mathscr{L}^2(\varepsilon_{K_{s-1},j_s})} \leq C \frac{\varepsilon_s}{\delta_s}, \\ (b) \quad & \leq \quad C \frac{1}{|\delta_{K_{s-1},j_s}|} |\Gamma_{K_{s-1},j_s}| \leq C \frac{\varepsilon_s}{\delta_s}. \end{aligned}$$

We have thus shown $\|a^{s-1}(\underline{x}_{s-1}) - \hat{a}^{s-1}(\underline{x}_{s-1})\|_{\mathscr{F}} \leq C \frac{\varepsilon_s}{\delta_s}$. □

In fact, Lemma 2.4.4 and 2.4.5 give the error estimates at scale $s-1$ between the homogenized tensor and the FE-HMM tensor defined using numerical integration provided the tensor at scale s is accurate, where $s = 1, \cdots, N$. To address the fully discrete error estimates, we first restrict to three scale problems in order to explain the idea of the analysis. We will give the general result for $N+1$ scales at the end of this section We thus denote our three scale tensor as $a(x, \frac{x}{\varepsilon_1}, \frac{x}{\varepsilon_2}) = a(x, y_1, y_2)$ where $y_1 = \frac{x}{\varepsilon_1}$, $y_2 = \frac{x}{\varepsilon_2}$ and we will assume periodicity at the meso and micro scales, i.e., assumption (2.1.3). With the help of Lemma 2.4.4 and 2.4.5, we are able to state our main theorem.

Theorem 2.4.6. *Assume for $s = 1, 2$ that the assumptions of Lemma 2.4.4 (or Lemma 2.4.5 for Dirichlet boundary coupling) hold. Then*

$$r_{HMM} \leq C((\frac{h_1}{\varepsilon_1})^{2q_1} + (\frac{h_2}{\varepsilon_2})^{2q_2} + r_{MOD}), \tag{2.4.39}$$

where r_{MOD} stands for the HMM modeling error which is estimated as follows.

If $\mathscr{W}(\delta_{K_0,j_1}) = \mathscr{W}^1_{per}(\delta_{K_0,j_1})$ and $\mathscr{W}(\delta_{K_1,j_2}) = \mathscr{W}^1_{per}(\delta_{K_1,j_2})$ with $\delta_1/\varepsilon_1 \in \mathbb{N}$, $\delta_2/\varepsilon_2 \in \mathbb{N}$, then

$$r_{MOD} = 0. \tag{2.4.40}$$

*If $\mathscr{W}(\delta_{K_0,j_1}) = \mathscr{H}^1_0(\delta_{K_0,j_1})$ and $\mathscr{W}(\delta_{K_1,j_2}) = \mathscr{H}^1_0(\delta_{K_1,j_2})$ with $\delta_1 > \varepsilon_1$, $\delta_2 > \varepsilon_2$ and assume (**H3**) holds*

then

$$r_{MOD} \leq C\Big(\frac{\varepsilon_1}{\delta_1} + \frac{\varepsilon_2}{\delta_2}\Big). \tag{2.4.41}$$

Proof of Theorem 2.4.6. We first split r_{HMM} given in (2.3.15) into,

$$\begin{aligned}\|a^0(x_{K_0,j_1}) - a^0_{K_0}(x_{K_0,j_1})\|_{\mathscr{F}} &\leq \|a^0(x_{K_0,j_1}) - \bar{a}^0_{K_0}(x_{K_0,j_1})\|_{\mathscr{F}} + \|\bar{a}^0_{K_0}(x_{K_0,j_1}) - a^0_{K_0}(x_{K_0,j_1})\|_{\mathscr{F}} \\ &:= I_1 + I_2 \end{aligned} \tag{2.4.42}$$

where $a^0(x_{K_0,j_1})$ is defined in (2.4.23), $a^0_{K_0}(x_{K_0,j_1})$ is defined in (2.2.11) for $N = 2$, $\bar{a}^0_{K_0}(x_{K_0,j_1})$ is defined in (2.4.26).

Lemma 2.4.4 and Lemma 2.4.5 give respectively the following estimates for the term I_1 in (2.4.42),

$$I_1 \leq C(\frac{h_1}{\varepsilon_1})^{2q_1}$$

for periodic boundary coupling,

$$I_1 \leq C(\frac{h_1}{\varepsilon_1})^{2q_1} + \frac{\varepsilon_1}{\delta_1}$$

for Dirichlet boundary coupling.

For the term I_2, using (2.4.27) and Cauchy-Schwarz inequality we have

$$\begin{aligned}&|\big(\bar{a}^0_K(x_{K_0,j_1}) - a^0_{K_0}(x_{K_0,j_1})\big)\mathbf{e}_i \cdot \mathbf{e}_k| \\ &= \Big|\frac{1}{|\delta_{K_0,j_1}|}\sum_{K_1\in\mathscr{T}_{h_1}}\sum_{j_2=1}^{J_2}\omega_{K_1,j_2}\Big((a^1(x_{K_0,j_1},\frac{x_{K_1,j_2}}{\varepsilon_1}) - a^1_{K_0,j_1}(x_{K_1,j_2}))(\nabla\bar{\chi}^{i,h_1}_{K_0,j_1}(x_{K_1,j_2}) + \mathbf{e}_i)\cdot\mathbf{e}_k \\ &+ a^1_{K_0,j_1}(x_{K_1,j_2})(\nabla\bar{\chi}^{i,h_1}_{K_0,j_1} - \nabla\chi^{i,h_1}_{K_0,j_1})\cdot\mathbf{e}_k\Big)\Big| \\ &\leq C\Big(\max_{x_{K_1,j_2}\in\delta_{K_0,j_1}}\|a^1(x_{K_0,j_1},\frac{x_{K_1,j_2}}{\varepsilon_1}) - a^1_{K_0,j_1}(x_{K_1,j_2})\|_{\mathscr{F}} + \frac{1}{\sqrt{|\delta_{K_0,j_1}|}}\|\nabla\bar{\chi}^{i,h_1}_{K_0,j_1} - \nabla\chi^{i,h_1}_{K_0,j_1}\|_{\mathscr{L}^2(\delta_{K_0,j_1})}\Big).\end{aligned} \tag{2.4.43}$$

Using the ellipticity of $\bar{a}^1_{K_0,j_1}(x)$ (see Remark 2.4.3), we have

$$\begin{aligned}&C\tilde{\lambda}\|\nabla\bar{\chi}^{i,h_1}_{K_0,j_1} - \nabla\chi^{i,h_1}_{K_0,j_1}\|^2_{\mathscr{L}^2(\delta_{K_0,j_1})} \\ &\leq \sum_{K_1\in\mathscr{T}_{h_1}}\sum_{j_2=1}^{J_2}\omega_{K_1,j_2}a^1_{K_0,j_1}(x_{K_1,j_2})\cdot \\ &\big(\nabla\bar{\chi}^{i,h_1}_{K_0,j_1}(x_{K_1,j_2}) - \nabla\chi^{i,h_1}_{K_0,j_1}(x_{K_1,j_2})\big)\cdot\big(\nabla\bar{\chi}^{i,h_1}_{K_0,j_1}(x_{K_1,j_2}) - \nabla\chi^{i,h_1}_{K_0,j_1}(x_{K_1,j_2})\big) \\ &= \sum_{K_1\in\mathscr{T}_{h_1}}\sum_{j_2=1}^{J_2}\omega_{K_1,j_2}\big(a^1_{K_0,j_1}(x_{K_1,j_2}) - a^1(x_{K_0,j_1},\frac{x_{K_1,j_2}}{\varepsilon_1})\big)\cdot\end{aligned}$$

$$\begin{aligned}&\big(\nabla\bar{\chi}^{i,h_1}_{K_0,j_1}(x_{K_1,j_2})+\mathbf{e}_i\big)\cdot\big(\nabla\bar{\chi}^{i,h_1}_{K_0,j_1}(x_{K_1,j_2})-\nabla\chi^{i,h_1}_{K_0,j_1}(x_{K_1,j_2})\big)\\&\leq C\sqrt{|\delta_{K_0,j_1}|}\max_{x_{K_1,j_2}\in\delta_{K_0,j_1}}\|a^1(x_{K_0,j_1},\frac{x_{K_1,j_2}}{\varepsilon_1})-a^1_{K_0,j_1}(x_{K_1,j_2})\|_{\mathscr{F}}\|\nabla\bar{\chi}^{i,h_1}_{K_0,j_1}-\nabla\chi^{i,h_1}_{K_0,j_1}\|_{\mathscr{L}^2(\delta_{K_0,j_1})},\end{aligned}\tag{2.4.44}$$

where we have used that $\chi^{i,h_1}_{K_0,j_1}$ is the solution of (**Ps**) with $s=1$ and that $\bar{\chi}^{i,h_1}_{K_0,j_1}$ is the solution of (2.4.25) with $s=1$, as well as Cauchy-Schwarz inequality and (2.4.27). Hence, we obtain

$$\frac{1}{\sqrt{|\delta_{K_0,j_1}|}}\|\nabla\bar{\chi}^{i,h_1}_{K_0,j_1}-\nabla\chi^{i,h_1}_{K_0,j_1}\|_{\mathscr{L}^2(\delta_{K_0,j_1})}\leq C\max_{x_{K_1,j_2}\in\delta_{K_0,j_1}}\|a^1(x_{K_0,j_1},\frac{x_{K_1,j_2}}{\varepsilon_1})-a^1_{K_0,j_1}(x_{K_1,j_2})\|_{\mathscr{F}}.\tag{2.4.45}$$

Finally, combining (2.4.43) and (2.4.45), we obtain $I_2\leq C\max_{x_{K_1,j_2}\in\delta_{K_0,j_1}}\|a^1(x_{K_0,j_1},\frac{x_{K_1,j_2}}{\varepsilon_1})-a^1_{K_0,j_1}(x_{K_1,j_2})\|_{\mathscr{F}}$ and the upper bound

$$I_2\leq C(\frac{h_2}{\varepsilon_2})^{2q_2}\tag{2.4.46}$$

for periodic boundary coupling (using Lemma 2.4.4), or

$$I_2\leq C\big((\frac{h_2}{\varepsilon_2})^{2q_2}+\frac{\varepsilon_2}{\delta_2}\big)\tag{2.4.47}$$

for Dirichlet boundary coupling (using Lemma 2.4.5). □

Remark 2.4.7. *We notice that in the modeling error for two scale problems as analyzed in [60, 1], there is an additional term. This term vanishes here as all our numerical tensors are only evaluated at quadrature points for all fast and slow variables.*

Generalization to $(N+1)$-scale problems. Following the idea of Theorem 2.4.6, one can generalize the result for $N+1$ scales.

Theorem 2.4.8. *Suppose that assumption* (2.1.3) *for $a^\varepsilon=a(x,\frac{x}{\varepsilon_1},\cdots,\frac{x}{\varepsilon_N})$ holds. In addition suppose that the assumptions of Lemma 2.4.4 (or Lemma 2.4.5 for Dirichlet boundary coupling for the cell problems) hold for $s=1,\cdots,N$. Then we have the following error estimate*

$$r_{HMM}\leq C\big((\frac{h_1}{\varepsilon_1})^{2q_1}+\cdots+(\frac{h_N}{\varepsilon_N})^{2q_N}+r_{MOD}\big).\tag{2.4.48}$$

The term r_{MOD} can be analyzed as follows.

If the cell problems at each scale are coupled with periodic boundary conditions and $\delta_s/\varepsilon_s\in\mathbb{N}$, $s=1,\cdots,N$, then we have

$$r_{MOD}=0.\tag{2.4.49}$$

If the cell problems at each scale are coupled with Dirichlet boundary conditions with $\delta_s > \varepsilon_s$, $s = 1, \cdots, N$, then

$$r_{MOD} \leq C(\frac{\varepsilon_1}{\delta_1} + \cdots + \frac{\varepsilon_N}{\delta_N}). \tag{2.4.50}$$

Proof. Besides the FE-HMM tensor $a^{s-1}_{K_{s-1}}(x_{K_{s-1},j_s})$ in (2.2.7), we need to define the tensor

$$a^{s-1,t}_{K_{s-1}}(x_{K_{s-1},j_s}) = \frac{1}{|\delta_{K_{s-1},j_s}|} \sum_{K_s \in \mathcal{T}_{h_s}} \sum_{j_{s+1}=1}^{J_{s+1}} \omega_{K_s,j_{s+1}} a^{s,t}_{K_s}(x_{K_s,j_{s+1}})(I + J^T_{\chi^{h_s,t}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})}), \tag{2.4.51}$$

for $1 \leq s \leq t \leq N$ where $(J_{\chi^{h_s,t}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})})_{ik} = \partial_k(\chi^{i,h_s,t}_{K_{s-1},j_s}(x_{K_s,j_{s+1}}))$ and for $s = t$ we denote

$$a^{t,t}_{K_t}(x_{K_t,j_{t+1}}) = a^t(\underline{x}_t), \tag{2.4.52}$$

where $a^t(\underline{x}_t)$ is the exact tensor at the t-th scale evaluated at $\underline{x}_t = (x_{K_0,j_1}, \frac{x_{K_1,j_2}}{\varepsilon_1}, \cdots, \frac{x_{K_t,j_{t+1}}}{\varepsilon_t})$. The function $\chi^{i,h_s,t}_{K_{s-1},j_s} \in S^{q_s}(\delta_{K_{s-1},j_s}, \mathcal{T}_{h_s})$ (or $S^{q_s}_0(\delta_{K_{s-1},j_s}, \mathcal{T}_{h_s})$) is the solution of a problem similar to (**Ps**), namely

$$\sum_{K_s \in \mathcal{T}_{h_s}} \sum_{j_{s+1}=1}^{J_{s+1}} \omega_{K_s,j_{s+1}} a^{s,t}_{K_s}(x_{K_s,j_{s+1}}) \Big(\nabla \chi^{i,h_s,t}_{K_{s-1},j_s}(x_{K_s,j_{s+1}}) + \mathbf{e}_i \Big) \cdot \nabla z^{h_s}(x_{K_s,j_{s+1}}) = 0 \tag{2.4.53}$$

where z^{h_s} is the test function in $S^{q_s}(\delta_{K_{s-1},j_s}, \mathcal{T}_{h_s})$ or $S^{q_s}_0(\delta_{K_{s-1},j_s}, \mathcal{T}_{h_s})$.

We note that for $s = 1, t = N$ we have $a^0_{K_0}(x_{K_0,j_1}) = a^{0,N}_{T_0}(x_{K_j})$, where $a^0_{K_0}(x_{K_0,j_1})$ is defined in (2.2.11). Following the proof of Lemma 2.3.1, one can show that $a^{s-1,t}_{K_{s-1}}(x_{K_{s-1},j_s})$ is also symmetric, bounded and elliptic and that

$$\|\nabla \chi^{i,h_s,t}_{K_{s-1},j_s}(x_{K_s,j_{s+1}}) + \mathbf{e}_i\|_{\mathscr{L}^2(\delta_{K_{s-1},j_s})} \leq C\sqrt{|\delta_{K_{s-1},j_s}|} \tag{2.4.54}$$

In view of Theorem 2.3.4 and the definition of r_{HMM}, we need to estimate $r_{HMM} = \|a^0(x_{K_0,j_1}) - a^0_{K_0}(x_{K_0,j_1})\|_{\mathscr{F}}$ for $(N+1)$-scale problems, where $a^0(x_{K_0,j_1})$ is the exact homogenized tensor, defined in (2.4.23). Similarly to (2.4.42), we first have the following decomposition

$$\|a^0(x_{K_0,j_1}) - a^0_{K_0}(x_{K_0,j_1})\|_{\mathscr{F}} \leq \sum_{t=1}^{N} \|a^{0,t-1}_{K_0}(x_{K_0,j_1}) - a^{0,t}_{K_0}(x_{K_0,j_1})\|_{\mathscr{F}} \tag{2.4.55}$$

where t ranges from $2, \cdots, N-1$ and $a^{0,t}_{K_0}(x_{K_0,j_1})$ is defined in (2.4.51) with $s = 1$. It is easy to observe that the analysis for the first term and the last term in (2.4.55) is identical to the proof of Theorem 2.4.6. To estimate (2.4.55) we proceed as follows.

Assume for the time being that for each $t = 1, \cdots, N$ and $s = 1, \cdots, t$

$$\|a^{s-1,t-1}_{K_{s-1}}(x_{K_{s-1},j_s}) - a^{s-1,t}_{K_{s-1}}(x_{K_{s-1},j_s})\|_{\mathscr{F}} \leq C \max_{x_{K_s,j_{s+1}} \in K_s} \|a^{s,t-1}_{K_s}(x_{K_s,j_{s+1}}) - a^{s,t}_{K_s}(x_{K_s,j_{s+1}})\|_{\mathscr{F}}, \tag{2.4.56}$$

where $a^{0,0}_{K_0}(x_{K_0,j_1}) = a^0(x_{K_0,j_1})$ according to the definition (2.4.52) and where C is independent of ε or the meshsizes.

Then by (2.4.56), we obtain recursively

$$\begin{aligned}
&\|a^{0,t-1}_{K_0}(x_{K_0,j_1}) - a^{0,t}_{K_0}(x_{K_0,j_1})\|_{\mathscr{F}} \\
&\leq C \max_{x_{K_1,j_2} \in K_1} \|a^{1,t-1}_{K_1}(x_{K_1,j_2}) - a^{1,t}_{K_1}(x_{K_1,j_2})\|_{\mathscr{F}} \leq \cdots \\
&\leq C \max_{x_{K_1,j_2} \in K_1} \cdots \max_{x_{K_{t-1},j_t} \in K_{t-1}} \|a^{t-1,t-1}_{K_{t-1}}(x_{K_{t-1},j_t}) - a^{t-1,t}_{K_{t-1}}(x_{K_{t-1},j_t})\|_{\mathscr{F}}
\end{aligned}$$

where $a^{t-1,t-1}_{K_{t-1}}(x_{K_{t-1},j_t}) = a^{t-1}(\underline{x}_{t-1})$ and where C can be different in each inequality.

It follows then by Lemma 2.4.4 and 2.4.5

$$\|a^{t-1}(\underline{x}_{t-1}) - a^{t-1,t}_{K_t}(x_{K_{t-1},j_t})\|_{\mathscr{F}} \leq C(\frac{h_t}{\varepsilon_t})^{2q_t}, \tag{2.4.57}$$

with periodic boundary coupling, or

$$\|a^{t-1}(\underline{x}_{t-1}) - a^{t-1,t}_{K_2}(x_{K_{t-1},j_t})\|_{\mathscr{F}} \leq C(\frac{h_t}{\varepsilon_t})^{2q_t} + \frac{\varepsilon_t}{\delta_t}, \tag{2.4.58}$$

with Dirichlet boundary coupling.

Combining the (2.4.56)-(2.4.58) with (2.4.55) gives (2.4.48), (2.4.49) and (2.4.50), hence the claims of the theorem. It thus remains to prove (2.4.56). By (2.4.51) in view of Remark 2.2.2, we have

$$\begin{aligned}
&|\big(a^{s-1,t-1}_{K_{s-1}}(x_{K_{s-1},j_s}) - a^{s-1,t}_{K_{s-1}}(x_{K_{s-1},j_s})\big)\mathbf{e}_i \cdot \mathbf{e}_k| \\
&= \Big|\frac{1}{|\delta_{K_{s-1},j_s}|} \sum_{K_s \in \mathscr{T}_{h_s}} \sum_{j_{s+1}=1}^{J_{s+1}} \omega_{K_s,j_{s+1}} \Big((a^{s,t-1}_{K_s}(x_{K_s,j_{s+1}}) - a^{s,t}_{K_s}(x_{K_s,j_{s+1}}))(\nabla \chi^{i,h_s,t-1}_{K_{s-1},j_s}(x_{K_s,j_{s+1}}) + \mathbf{e}_i) \cdot \mathbf{e}_k \\
&+ a^{s,t}_{K_s}(x_{K_s,j_{s+1}})(\nabla \chi^{i,h_s,t-1}_{K_{s-1},j_s}(x_{K_s,j_{s+1}}) - \nabla \chi^{i,h_s,t}_{K_{s-1},j_s}(x_{K_s,j_{s+1}})) \cdot \mathbf{e}_k \Big)\Big| \\
&\leq C\Big(\max_{x_{K_s,j_{s+1}} \in \delta_{K_{s-1},j_s}} \|a^{s,t-1}_{K_s}(x_{K_s,j_{s+1}}) - a^{s,t}_{K_s}(x_{K_s,j_{s+1}})\|_{\mathscr{F}} \\
&+ \frac{1}{\sqrt{|\delta_{K_{s-1},j_s}|}} \|\nabla \chi^{i,h_s,t-1}_{K_{s-1},j_s} - \nabla \chi^{i,h_s,t}_{K_{s-1},j_s}\|_{\mathscr{L}^2(\delta_{K_{s-1},j_s})}\Big).
\end{aligned} \tag{2.4.59}$$

Using the ellipticity of $a^{s,t-1}_{K_s}$ and (2.4.53), we have

$$\tilde{\lambda} \|\nabla \chi^{i,h_s,t-1}_{K_{s-1},j_s} - \nabla \chi^{i,h_s,t}_{K_{s-1},j_s}\|^2_{\mathscr{L}^2(\delta_{K_{s-1},j_s})}$$

$$
\begin{aligned}
&\leq \sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}} \omega_{K_s,j_{s+1}} a_{K_s}^{s,t-1}(x_{K_s,j_{s+1}})\nabla(\chi_{K_{s-1},j_s}^{i,h_s,t-1}-\chi_{K_{s-1},j_s}^{i,h_s,t})\cdot\nabla(\chi_{K_{s-1},j_s}^{i,h_s,t-1}-\chi_{K_{s-1},j_s}^{i,h_s,t})\\
&= \sum_{K_s\in\mathcal{T}_{h_s}}\sum_{j_{s+1}=1}^{J_{s+1}} \omega_{K_s,j_{s+1}} (a_{K_s}^{s,t}(x_{K_s,j_{s+1}})-a_{K_s}^{s,t-1}(x_{K_s,j_{s+1}}))(\nabla\chi_{K_{s-1},j_s}^{i,h_s,t}+\mathbf{e}_i)\cdot\nabla(\chi_{K_{s-1},j_s}^{i,h_s,t-1}-\chi_{K_{s-1},j_s}^{i,h_s,t})\\
&\leq C\sqrt{|\delta_{K_{s-1},j_s}|}\max_{x_{K_s,j_{s+1}}\in\delta_{K_{s-1},j_s}} \|a_{K_s}^{s,t-1}(x_{K_s,j_{s+1}})-a_{K_s}^{s,t}(x_{K_s,j_{s+1}})\|_{\mathcal{F}}\|\nabla\chi_{K_{s-1},j_s}^{i,h_s,t-1}-\nabla\chi_{K_{s-1},j_s}^{i,h_s,t}\|_{\mathscr{L}^2(\delta_{K_{s-1},j_s})}.
\end{aligned}
$$

The last inequality above is obtained by using Cauchy-Schwarz inequality and (2.4.27). Hence, we obtain

$$
\frac{1}{\sqrt{|\delta_{K_{s-1},j_s}|}}\|\chi_{K_{s-1},j_s}^{i,h_s,t-1}-\nabla\chi_{K_{s-1},j_s}^{i,h_s,t}\|_{\mathscr{L}^2(\delta_{K_{s-1},j_s})}\leq C\max_{x_{K_s,j_{s+1}}\in\delta_{K_{s-1},j_s}}\|a_{K_s}^{s,t}(x_{K_s,j_{s+1}})-a_{K_s}^{s,t-1}(x_{K_s,j_{s+1}})\|_{\mathcal{F}}. \tag{2.4.60}
$$

Inserting (2.4.60) in (2.4.59) proves (2.4.56). Hence the proof is complete. □

Complexity. Let M_0 be the the number of DOF of the macro FEM. We write $\hat{h}_s=\frac{h_s}{\varepsilon_s}$ and denote $M_s=\mathcal{O}(\hat{h}_s^{-d})$ the number of degrees of freedom (DOF) for cell problems at the scale s, where $s=1,\cdots,N$. We emphasize that M_s is independent of ε_s since $|\delta_{K_{s-1},j_s}|=\delta_s^d=(\frac{\delta_s}{\varepsilon_s})^d\varepsilon_s^d=C_s\varepsilon_s^d$ with C_s a moderate constant (recall our assumption that $\delta_s=\hat{C}_s\varepsilon_s$). Using quasi-uniform meshes for each scale, we have the following relations

$$
h_0=\mathcal{O}(M_0^{-1/d}),\ \ \hat{h}_s=\mathcal{O}(M_s^{-1/d}),\ s=1,\cdots,N.
$$

In view of (2.3.18), (2.3.19) and (2.4.48), optimal convergence rates (up to a modeling error r_{MOD} independent of h_0,h_s, where $s=1,\cdots,N$) can be obtained for quasi-uniform meshes given by,

$$
\hat{h}_s\approx h_0^{\frac{q_0}{2q_s}}\ \text{for the } \mathscr{H}^1 \text{ norm},\ \hat{h}_s\approx h_0^{\frac{q_0+1}{2q_s}}\ \text{for the } \mathscr{L}^2 \text{ norm}.
$$

The corresponding complexity in term of macro DOF reads

$$
\begin{aligned}
&\mathcal{O}\big(\underbrace{h_0^{-d}}_{M_0}\cdot\underbrace{h_0^{\frac{-dq_0}{2q_1}}}_{M_1}\cdots\cdots\underbrace{h_0^{\frac{-dq_0}{2q_N}}}_{M_N}\big)=\mathcal{O}\big(M_0^{1+\frac{q_0}{2q_1}+\cdots+\frac{q_0}{2q_N}}\big)\ \text{ for the } \mathscr{H}^1 \text{ norm},\\
&\mathcal{O}\big(\underbrace{h_0^{-d}}_{M_0}\cdot\underbrace{h_0^{\frac{-d(q_0+1)}{2q_1}}}_{M_1}\cdots\cdots\underbrace{h_0^{\frac{-d(q_0+1)}{2q_N}}}_{M_N}\big)=\mathcal{O}\big(M_0^{1+\frac{(q_0+1)}{2q_1}+\cdots+\frac{(q_0+1)}{2q_N}}\big)\ \text{ for the } \mathscr{L}^2 \text{ norm}.
\end{aligned}
$$

2.5 Some extensions of convergence result for numerical quadrature

2.5.1 Numerical integration error analysis for simplicial elements

In this section, we analyse the numerical integration error which we use in Section 2.3 for the estimation of (2.4.45).

Theorem 2.5.1. *(Bramble-Hilbert lemma [53]) Let $\mathscr{D}$ be a domain in $\mathbb{R}^d$, let $k \geq 0$ be an integer, let p be a number satisfying $1 \leq p \leq \infty$, and let L be a continuous linear form on the space $\mathscr{W}^{m+1,p}(\mathscr{D})$ with the property that*

$$L(p) = 0 \ \text{ for all } p \in \mathscr{P}^m(\mathscr{D}). \tag{2.5.61}$$

Then there exists a constant $C(\mathscr{D})$ such that

$$|L(v)| \leq C(\mathscr{D}) \|L\|^*_{\mathscr{W}^{m+1,p}(\mathscr{D})} |v|_{\mathscr{W}^{m+1,p}(\mathscr{D})}, \ \ \forall v \in \mathscr{W}^{m+1,p}(\mathscr{D}), \tag{2.5.62}$$

where $\|\cdot\|^*_{\mathscr{W}^{m+1,p}(\mathscr{D})}$ is the dual norm of space $\mathscr{W}^{m+1,p}(\mathscr{D})$.

Let $E_K(\phi)$ denote the numerical integration error on the element K defined as

$$E_K(\phi) = \int_K \phi dx - \sum_{j=1}^{J} \omega_{K,j} \phi(x_{K,j}), \tag{2.5.63}$$

and $E_{\hat{K}}(\hat{\phi})$ is the corresponding error on the reference element $\hat{K}$

$$E_{\hat{K}}(\hat{\phi}) = \int_{\hat{K}} \hat{\phi} d\hat{x} - \sum_{j=1}^{J} \hat{\omega}_j \phi(\hat{x}_j). \tag{2.5.64}$$

Theorem 2.5.2. *(First theorem for local high order numerical integration error) Assume that K is simplicial and $a \in \mathscr{W}^{2m,\infty}(K)$. Assume*

$$E_{\hat{K}}(\hat{\phi}) = 0, \ \ \forall \hat{\phi} \in \mathscr{P}^{2m-1}(\hat{K}). \tag{2.5.65}$$

Then,

$$E_K(avw) \leq Ch^{2m} \|a\|_{\mathscr{W}^{2m,\infty}(K)} \|v\|_{\mathscr{H}^{m-1}(K)} \|w\|_{\mathscr{H}^{m-1}(K)}, \ \ \forall v, w \in \mathscr{P}^{m-1}(K). \tag{2.5.66}$$

Proof. Since $E_K(\phi) = 2|K| E_{\hat{K}}(\hat{\phi})$, then $E_K(avw) = 2|K| E_{\hat{K}}(\hat{a}\hat{v}\hat{w})$. Let $\hat{\phi} = \hat{a}\hat{v}\hat{w} \in \mathscr{W}^{2m,\infty}(\hat{K})$. We

have

$$|E_{\hat{K}}(\hat{\phi})| = |\int_{\hat{K}} \hat{\phi} - \sum_j \hat{\omega}_j \hat{\phi}| \leq \hat{C}\|\hat{\phi}\|_{\mathscr{L}^\infty(\hat{K})} \leq \hat{C}\|\hat{\phi}\|_{\mathscr{W}^{2m,\infty}(\hat{K})}.$$

The above inequality shows that $E_{\hat{K}}(\hat{\phi})$ is a continuous linear functional on $\mathscr{W}^{2m,\infty}(\hat{K})$. Furthermore, we have the assumption on the quadrature (2.5.63) that $E_{\hat{K}}(\hat{\psi}) = 0,\ \forall \hat{\psi} \in \mathscr{P}^{2m-1}(\hat{K})$. Therefore, Bramble-Hilbert Lemma (see [53, Theorem 28.1]) can be applied to $E_{\hat{K}}$, i.e.

$$|E_{\hat{K}}(\hat{\phi})| \leq \hat{C}|\hat{\phi}|_{\mathscr{W}^{2m,\infty}(\hat{K})}. \tag{2.5.67}$$

As a result, we have

$$\begin{aligned}|E_{\hat{K}}(\hat{a}\hat{v}\hat{w})| &\leq \hat{C}|\hat{a}\hat{v}\hat{w}|_{\mathscr{W}^{2m,\infty}(\hat{K})} \\ &\leq \hat{C}\sum_{i=0}^{m-1}\sum_{k=0}^{2m-i}|\hat{a}|_{\mathscr{W}^{2m-i-k,\infty}(\hat{K})}|\hat{v}|_{\mathscr{W}^{i,\infty}(\hat{K})}|\hat{w}|_{\mathscr{W}^{k,\infty}(\hat{K})}.\end{aligned}$$

By the affine transformation, we have

$$\begin{aligned}&|\hat{a}|_{\mathscr{W}^{2m-i-k,\infty}(\hat{K})} \leq Ch^{2m-i-k}|a|_{\mathscr{W}^{2m-i-k,\infty}(K)}, \\ &|\hat{v}|_{\mathscr{W}^{i,\infty}(\hat{K})} \leq Ch^i|2K|^{-1/2}|v|_{\mathscr{H}^i(K)}.\end{aligned}$$

Therefore,

$$|E_K(avw)| \leq Ch^{2m}\|a\|_{\mathscr{W}^{2m,\infty}(K)}\|v\|_{\mathscr{H}^{m-1}(K)}\|w\|_{\mathscr{H}^{m-1}(K)}.$$

□

Theorem 2.5.3. *(First theorem for global higher order numerical integration error) Assume that $a_{ij} \in \mathscr{W}^{2m,\infty}(\mathscr{D})$ and $u, w \in \mathscr{W}(\mathscr{D}) \cap \mathscr{H}^{m+1}(\mathscr{D})$ $\big(\mathscr{W}(\mathscr{D})$ can be either defined as in* (2.2.5) *provided that $\mathscr{D}$ is a cube or defined as in* (2.2.6)$\big)$ *are the solutions of the following problems: for $\forall\, v \in \mathscr{W}(\mathscr{D})$,*

$$\begin{aligned}&\int_{\mathscr{D}} a\nabla u \cdot \nabla v dx = F_1(v), \\ &\int_{\mathscr{D}} a\nabla w \cdot \nabla v dx = F_2(v),\end{aligned}$$

where F_1, F_2 are two linear functional mapping $\mathscr{W}(\mathscr{D}) \cap \mathscr{H}^{m+1}(\mathscr{D}) \to \mathbb{R}$ and u^h, w^h are the FEM solutions with numerical quadrature in $S^m(\mathscr{D}, \mathscr{T}_h)$ where $\mathscr{T}_h$ is a shape regular partition of $\mathscr{D}$ and elements $K \in \mathscr{T}_h$ are quasi-uniform. Assume the numerical quadrature satisfies

$$E_{\hat{K}}(\hat{\phi}) = 0,\ \ \forall \hat{\phi} \in \mathscr{P}^{2m-1}(\hat{K}). \tag{2.5.68}$$

Then we have

$$\begin{aligned}&|\int_{\mathscr{D}}\sum_{ij}a_{ij}\partial_i u^h\partial_j w^h dx-\sum_{K\in\mathscr{T}_h}\sum_{j=1}^{J}\omega_{K,j}\sum_{ij}(a_{ij}\partial_i u^h\partial_j w^h)(x_{K,j})|\\&\leq Ch^{2m}(\sum_{ij}\|a_{ij}\|_{\mathscr{W}^{2m,\infty}(\mathscr{D})})\|u\|_{\mathscr{H}^m(\mathscr{D})}\|w\|_{\mathscr{H}^m(\mathscr{D})}.\end{aligned}\tag{2.5.69}$$

Proof. We first write

$$E_K(a_{ij}\partial_i u^h\partial_j w^h)=\int_K\sum_{ij}a_{ij}\partial_i u^h\partial_j w^h dx-\sum_{j}^{J}\omega_{K,j}\sum_{ij}(a_{ij}\partial_i u^h\partial_j w^h)(x_{K,j}).$$

Using Theorem 2.5.2,we can obtain the following local estimate

$$|E_K(a_{ij}\partial_i u^h\partial_j w^h)|\leq Ch^{2m}(\sum_{ij}\|a_{ij}\|_{\mathscr{W}^{2m,\infty}(K)})\|u^h\|_{\mathscr{H}^m(K)}\|w^h\|_{\mathscr{H}^m(K)}.$$

Thus, we obtain

$$\begin{aligned}&|\int_{\mathscr{D}}\sum_{ij}a_{ij}\partial_i u^h\partial_j w^h dx-\sum_{K\in\mathscr{T}_h}\sum_{j=1}^{J}\omega_{K,j}\sum_{ij}(a_{ij}\partial_i u^h\partial_j w^h)(x_{K,j})|\\&\leq Ch^{2m}(\sum_{ij}\|a_{ij}\|_{\mathscr{W}^{2m,\infty}(\mathscr{D})})\sum_{K\in\mathscr{T}_h}\|u^h\|_{H^m(K)}\|w^h\|_{\mathscr{H}^m(K)}\\&=Ch^{2m}(\sum_{ij}\|a_{ij}\|_{\mathscr{W}^{2m,\infty}(\mathscr{D})})\sum_{K\in\mathscr{T}_h}\Big((\sum_{\ell=1}^{m}|u^h|^2_{\mathscr{H}^\ell(K)})^{1/2}(\sum_{\ell=1}^{m}|w^h|^2_{\mathscr{H}^\ell(K)})^{1/2}\Big)\\&\leq Ch^{2m}(\sum_{ij}\|a_{ij}\|_{\mathscr{W}^{2m,\infty}(\mathscr{D})})\Big(\sum_{K\in\mathscr{T}_h}\sum_{\ell=1}^{m}|u^h|^2_{\mathscr{H}^\ell(K)}\Big)^{1/2}\Big(\sum_{K\in\mathscr{T}_h}\sum_{\ell=1}^{m}|w^h|^2_{\mathscr{H}^\ell(K)}\Big)^{1/2}.\end{aligned}\tag{2.5.70}$$

Now we define $\Pi_K^u\in\mathscr{P}^m(K)$ which is an interpolation polynomial of u on element K (respectively Π_K^w the interpolation polynomial of w). By classical interpolation error estimate (see [53, Chapter 3]), we have

$$\|u-\Pi_K^u\|_{\mathscr{H}^\ell(K)}\leq Ch^{m+1-\ell}|u|_{\mathscr{H}^{m+1}(K)},\ \text{for } \ell\leq m.$$

Furthermore, we can derive that for $\ell\leq m$

$$\|\Pi_K^u\|_{\mathscr{H}^\ell(K)}\leq\|u-\Pi_K^u\|_{\mathscr{H}^\ell(K)}+\|u\|_{\mathscr{H}^\ell(K)}\leq C\|u\|_{\mathscr{H}^m(K)}.$$

Using inverse inequality (assume that the triangulation $\mathscr{T}_h$ is quasi-uniform), we have

$$\sum_{\ell=1}^{m}|u^h|^2_{\mathscr{H}^\ell(K)}\leq\sum_{\ell=1}^{m}\big(|u^h-\Pi_K^u|^2_{\mathscr{H}^\ell(K)}+|\Pi_K^u|^2_{\mathscr{H}^\ell(K)}\big)$$

$$
\begin{aligned}
&\leq \|\Pi_K^u\|^2_{\mathscr{H}^m(K)} + C\sum_{\ell=1}^{m} h^{-2\ell+2}|u^h - \Pi_K^u|^2_{\mathscr{H}^1(K)} \\
&\leq C\Big(\|u\|^2_{\mathscr{H}^m(K)} + \sum_{\ell=1}^{m} h^{-2\ell+2}\big(|u^h - u|^2_{\mathscr{H}^1(K)} + |u - \Pi_K^u|^2_{\mathscr{H}^1(K)}\big)\Big) \\
&\leq C\big(\|u\|^2_{\mathscr{H}^m(K)} + h^{-2m+2}\|u^h - u\|^2_{\mathscr{H}^1(K)}\big).
\end{aligned}
$$

Sum up the above inequality with respect to element $K \in \mathscr{T}_h$ and we obtain

$$
\sum_{K\in\mathscr{T}_h}\sum_{\ell=1}^{m} |u^h|^2_{\mathscr{H}^\ell(K)} \leq C\Big(\|u\|^2_{\mathscr{H}^m(\mathscr{D})} + h^{-2m+2}\|u - u^h\|^2_{\mathscr{H}^1(\mathscr{D})}\Big). \tag{2.5.71}
$$

Using standard FEM error analysis with numerical quadrature (see for example [53]), we have the following result

$$
\|u - u^h\|_{\mathscr{H}^1(\mathscr{D})} \leq Ch^{m-1}|u|_{\mathscr{H}^m(\mathscr{D})}. \tag{2.5.72}
$$

Combining (2.5.71) with (2.5.72), we have

$$
\sum_{K\in\mathscr{T}_h}\sum_{\ell=1}^{m} |u^h|^2_{\mathscr{H}^\ell(K)} \leq C\|u\|_{\mathscr{H}^m(\mathscr{D})}. \tag{2.5.73}
$$

Taking (2.5.73) into (2.5.70), we have proved (2.5.69). □

Note that in Theorem 2.5.3 we have to assume quasi-uniform meshes for FEM. However, this assumption can be restrictive in some situations, e.g. adaptivity methods or complex domains. Therefore in the following, we give a similar theorem relaxing the quasi-uniform assumption. We however need higher order quadrature rules.

Theorem 2.5.4. *(Second theorem for local high order numerical integration error) Assume that K is simplicial and $a \in \mathscr{W}^{2m,\infty}(K)$. Assume further*

$$
E_{\hat{K}}(\hat{\phi}) = 0, \ \ \forall \hat{\phi} \in \mathscr{P}^{4m-3}(\hat{K}). \tag{2.5.74}
$$

Then,

$$
E_K(avw) \leq Ch^{2m}|a|_{\mathscr{W}^{2m,\infty}(K)}\|v\|_{\mathscr{L}^2(K)}\|w\|_{\mathscr{L}^2(K)}, \ \ \forall v, w \in \mathscr{P}^{m-1}(K). \tag{2.5.75}
$$

Proof. First we write $E_K(avw) = 2|K|E_{\hat{K}}(\hat{a}\hat{v}\hat{w})$. Then we have

$$
|E_{\hat{K}}(\hat{a}\hat{v}\hat{w})| = \Big|\int_{\hat{K}} \hat{a}\hat{v}\hat{w} - \sum_j \hat{\omega}_j \hat{a}\hat{v}\hat{w}\Big| \leq \hat{C}\|\hat{a}\hat{v}\hat{w}\|_{\mathscr{L}^\infty(\hat{K})}
$$

$$\leq \hat{C}\|\hat{a}\|_{\mathscr{W}^{2m,\infty}(\hat{K})}\|\hat{v}\|_{\mathscr{L}^{\infty}(\hat{K})}\|\hat{w}\|_{\mathscr{L}^{\infty}(\hat{K})} \leq \hat{C}\|\hat{a}\|_{\mathscr{W}^{2m,\infty}(\hat{K})}\|\hat{v}\|_{\mathscr{L}^{2}(\hat{K})}\|\hat{w}\|_{\mathscr{L}^{2}(\hat{K})}.$$

The above inequality shows that $f(\cdot) := E_{\hat{K}}(\cdot\hat{v}\hat{w}))$ is a continuous linear functional on $\mathscr{W}^{2m,\infty}(\hat{K})$ and $\|f\|^{*}_{\mathscr{W}^{2m,\infty}(\hat{K})} \leq \hat{C}\|\hat{v}\|_{\mathscr{L}^{2}(\hat{K})}\|\hat{w}\|_{\mathscr{L}^{2}(\hat{K})}$. Furthermore, using the assumption (2.5.74), we have that $f(\hat{\psi}) = 0,\ \forall\hat{\psi} \in \mathscr{P}^{2m-1}(\hat{K})$ (note that $\hat{v}\hat{w} \in \mathscr{P}^{2m-2}(\hat{K})$). Therefore, applying Bramble-Hilbert to f gives

$$|E_{\hat{K}}(\hat{a}\hat{v}\hat{w})| = |f(\hat{a})| \leq \hat{C}\|f\|^{*}_{\mathscr{W}^{2m,\infty}(\hat{K})}|\hat{a}|_{\mathscr{W}^{2m,\infty}(\hat{K})} \leq \hat{C}\|\hat{v}\|_{\mathscr{L}^{2}(\hat{K})}\|\hat{w}\|_{\mathscr{L}^{2}(\hat{K})}|\hat{a}|_{\mathscr{W}^{2m,\infty}(\hat{K})}.$$

By the affine transformation, we obtain

$$|\hat{a}|_{\mathscr{W}^{2m,\infty}(\hat{K})} \leq Ch^{2m}|a|_{\mathscr{W}^{2m,\infty}(K)},$$
$$\|\hat{v}\|_{\mathscr{L}^{2}(\hat{K})} \leq |2K|^{-1/2}\|v\|_{\mathscr{L}^{2}(K)}.$$

Therefore,

$$|E_K(avw)| \leq Ch^{2m}|a|_{\mathscr{W}^{2m,\infty}(K)}\|v\|_{\mathscr{L}^{2}(K)}\|w\|_{\mathscr{L}^{2}(K)}.$$

□

Based on Theorem 2.5.4, it is easy to obtain the following result without using the assumption of quasi-uniform meshes.

Theorem 2.5.5. *(Second theorem for global higher order numerical integration error) Assume that $a_{ij} \in \mathscr{W}^{2m,\infty}(\mathscr{D})$ and $u, w \in \mathscr{W}(\mathscr{D}) \cap \mathscr{H}^2(\mathscr{D})$ $\big(\mathscr{W}(\mathscr{D})$ can be either defined as in* (2.2.5) *provided that $\mathscr{D}$ is a cube or defined as in* (2.2.6)$\big)$ *are the solutions of the following problems: for $\forall v \in \mathscr{W}(\mathscr{D})$,*

$$\int_{\mathscr{D}} a\nabla u\cdot\nabla v dx = F_1(v),$$
$$\int_{\mathscr{D}} a\nabla w\cdot\nabla v dx = F_2(v),$$

where F_1, F_2 are two linear functional mapping $\mathscr{W}(\mathscr{D}) \cap \mathscr{H}^2(\mathscr{D}) \to \mathbb{R}$ and u^h, w^h are the FEM solutions with numerical quadrature in $S^m(\mathscr{D}, \mathscr{T}_h)$ where $\mathscr{T}_h$ is a shape regular partition of $\mathscr{D}$. Assume the numerical quadrature satisfies

$$E_{\hat{K}}(\hat{\phi}) = 0,\ \ \forall\hat{\phi} \in \mathscr{P}^{4m-3}(\hat{K}). \tag{2.5.76}$$

Then we have

$$\Big|\int_{\mathscr{D}}\sum_{ij} a_{ij}\partial_i u^h\partial_j w^h dx - \sum_{K\in\mathscr{T}_h}\sum_{j=1}^{J}\omega_{K,j}\sum_{ij}(a_{ij}\partial_i u^h\partial_j w^h)(x_{K,j})\Big|$$

$$\leq Ch^{2m}(\sum_{ij}|a_{ij}|_{\mathscr{W}^{2m,\infty}(\mathscr{D})})\|u\|_{\mathscr{H}^2(\mathscr{D})}\|w\|_{\mathscr{H}^2(\mathscr{D})}. \tag{2.5.77}$$

Proof. A direct application of Theorem 2.5.4 gives

$$\begin{aligned} &|\int_{\mathscr{D}}\sum_{ij}a_{ij}\partial_i u^h\partial_j w^h dx - \sum_{K\in\mathscr{T}_h}\sum_{j=1}^{J}\omega_{K,j}\sum_{ij}(a_{ij}\partial_i u^h\partial_j w^h)(x_{K,j})| \\ \leq\ & Ch^{2m}(\sum_{ij}|a_{ij}|_{\mathscr{W}^{2m,\infty}(\mathscr{D})})\|\nabla u^h\|_{\mathscr{L}^2(\mathscr{D})}\|\nabla w^h\|_{\mathscr{L}^2(\mathscr{D})}. \end{aligned}$$

Note that u^h, w^h are the FE approximations of u, w respectively and thus we have

$$\|\nabla u^h\|_{\mathscr{L}^2(\mathscr{D})} \leq \|\nabla u\|_{\mathscr{L}^2(\mathscr{D})} + \|\nabla u - \nabla u^h\|_{\mathscr{L}^2(\mathscr{D})} \leq C\|u\|_{\mathscr{H}^2(\mathscr{D})}.$$

Therefore the proof is complete. □

2.5.2 Numerical integration error analysis for parallelogram elements

In order to obtain the same error result for parallelogram elements, one needs to introduce the following semi-norm for space $\mathscr{W}^{m,p}(\mathscr{D})$

$$\{v\}_{\mathscr{W}^{m,p}(\mathscr{D})} = \Big(\int_{\mathscr{D}}\sum_{i=1}^{d}\big(\partial_i^m v\big)^p\Big)^{1/p}, \tag{2.5.78}$$

where $\partial_i := (\partial)/(\partial x_i)$ (see for example [53], where such norms have been discussed). Based on this semi-norm, a corresponding Bramble-Hilbert lemma can be derived for parallelogram elements (see [53, Chapter 11]). For the completeness, we present here the proofs of the equivalent norm theorem and Bramble-Hilbert lemma for parallelogram elements.

Theorem 2.5.6. *(Equivalent norm theorem) Let $m \geq 1$ be an integer and $p \in [1,\infty]$, then there exist a constant C which depends on domain $\mathscr{D}$ such that,*

$$\inf_{q(x)\in\mathscr{Q}^{m-1}(\mathscr{D})}\|v+q\|_{\mathscr{W}^{m,p}(\mathscr{D})} \leq C\{v\}_{\mathscr{W}^{m,p}(\mathscr{D})},\ \ \forall v\in\mathscr{W}^{m,p}(\mathscr{D}).$$

Proof. Let $N = dim\ \mathscr{Q}^{m-1}(\mathscr{D})$ and $\{f_i,\ 1\leq i\leq N\}$ be a set of basis of the dual space of $\mathscr{Q}^{m-1}(\mathscr{D})$. Thus by Hahn-Banach extension theorem [4], there exists a set of continuous linear functionals over the space $\mathscr{W}^{m,p}(\mathscr{D})$ again denoted by $f_i,\ 1\leq i\leq N$ such that $q\in\mathscr{Q}^{m-1}(\mathscr{D})$ and $f_i(q)=0,\ 1\leq i\leq N$

[4] Hahn-Banach extension theorem: If $\mathscr{X}$ is a normed vector space and $\mathscr{X}^0$ is a subspace of $\mathscr{X}$. Let f_0 is a bounded linear functional on $\mathscr{X}^0$. Then there exists a bounded linear functional f on $\mathscr{X}$ satisfying: (1). $f(v) = f_0(v), \forall v \in \mathscr{X}^0$, (2). $\|f\|_{\mathscr{X}} = \|f_0\|_{\mathscr{X}^0}$.

imply $q = 0$. If the following inequality holds:

$$\|v\|_{\mathscr{W}^{m,p}(\mathscr{D})} \leq C\Big(\{v\}_{\mathscr{W}^{m,p}(\mathscr{D})} + \sum_{i=1}^{N} |f_i(v)|\Big) \ \ \forall v \in \mathscr{W}^{m,p}(\mathscr{D}), \tag{2.5.79}$$

then for any given $v \in \mathscr{W}^{m,p}(\mathscr{D})$, we can find $q \in \mathscr{Q}^{m-1}(\mathscr{D})$ such that $f_i(v+q) = 0,\ 1 \leq i \leq N$ and we can then obtain

$$\inf_{q(x)\in\mathscr{Q}^{m-1}(\mathscr{D})} \|v+q\|_{\mathscr{W}^{m,p}(\mathscr{D})} \leq \|v+q\|_{\mathscr{W}^{m,p}(\mathscr{D})} \leq C\{v\}_{\mathscr{W}^{m,p}(\mathscr{D})}. \tag{2.5.80}$$

Now we prove (2.5.79). We first assume (2.5.79) is false, then there exists a sequence $v_\ell \in \mathscr{W}^{m,p}(\mathscr{D}),\ \ell \geq 1$ such that

(1) $\|v_\ell\|_{\mathscr{W}^{m,p}(\mathscr{D})} = 1 \ \ \forall \ell \geq 1$,

(2) $\lim_{\ell\to\infty}\Big(\{v_\ell\}_{\mathscr{W}^{m,p}(\mathscr{D})} + \sum_{i=1}^{N} |f_i(v_\ell)|\Big) = 0$.

Since v_ℓ is bounded in $\mathscr{W}^{m,p}(\mathscr{D})$, there exists a subsequence again denoted as v_ℓ that converges to a limit $v \in \mathscr{W}^{m,p}(\mathscr{D})$. Since $\lim_{\ell\to\infty}\{v_\ell\}_{\mathscr{W}^{m,p}(\mathscr{D})} = 0$, then

$$|\partial_i^m v| = \lim_{\ell\to\infty} |v_\ell| = 0,$$

Thus we have $v \in \mathscr{Q}^{m-1}(\mathscr{D})$. Furthermore, we have

$$|f_i(v)| = \lim_{\ell\to\infty} |f_i(v_\ell)| = 0, \ \ 1 \leq i \leq N.$$

Therefore we can conclude that $v = 0$ which contradicts the assumption that $\|v_\ell\|_{\mathscr{W}^{m,p}(\mathscr{D})} = 1,\ \forall \ell \leq 1$, so that (2.5.79) holds. □

Remark 2.5.7. *Since $\mathscr{P}^{m-1}(\mathscr{D}) \subset \mathscr{Q}^{m-1}(\mathscr{D})$, then* (2.5.79) *also holds for $q \in \mathscr{P}^{m-1}(\mathscr{D})$. In fact, by the quotient space embedding $\mathscr{W}^{m,p}(\mathscr{D})/\mathscr{Q}^{m-1}(\mathscr{D}) \subset \mathscr{W}^{m,p}(\mathscr{D})/\mathscr{P}^{m-1}(\mathscr{D})$, we have*

$$\inf_{p\in\mathscr{P}^{m-1}(\mathscr{D})} \|v+p\|_{\mathscr{W}^{m,p}(\mathscr{D})} \leq \inf_{q\in\mathscr{Q}^{m-1}(\mathscr{D})} \|v+q\|_{\mathscr{W}^{m,p}(\mathscr{D})}.$$

With the help of Theorem 2.5.6, we can correspondingly have a Bramble-Hilbert lemma for quadrilateral elements.

Theorem 2.5.8. *(Bramble-Hilbert lemma for quadrilateral elements) Let $\mathscr{D}$ be a domain in $\mathbb{R}^d$, let $k \geq 0$ be an integer, let p be a number satisfying $1 \leq p \leq \infty$, and let $\mathscr{L}$ be a continuous linear*

functional on the space $\mathscr{W}^{m,p}(\mathscr{D})$ *with the property that*

$$L(q) = 0 \ \text{for all } q \in \mathscr{Q}^{m-1}(\mathscr{D}). \tag{2.5.81}$$

Then there exists a constant C *which depends on* $\mathscr{D}$ *such that*

$$|L(v)| \le C\|L\|^*_{\mathscr{W}^{m,p}(\mathscr{D})}\{v\}_{\mathscr{W}^{m,p}(\mathscr{D})}, \ \ \forall v \in \mathscr{W}^{m,p}(\mathscr{D}), \tag{2.5.82}$$

where $\|\cdot\|^*_{\mathscr{W}^{m,p}(\mathscr{D})}$ *is the dual norm of space* $\mathscr{W}^{m,p}(\mathscr{D})$).

Proof. Since for any given $v \in \mathscr{W}^{m,p}(\mathscr{D})$ we have $\mathscr{L}(v+q) = \mathscr{L}(v)\ \forall q \in \mathscr{Q}^{m-1}(\mathscr{D})$, then we can write

$$|L(v)| = |L(v+q)| \le \|L\|^*_{\mathscr{W}^{m,p}(\mathscr{D})}\|v+q\|_{\mathscr{W}^{m,p}(\mathscr{D})}.$$

We note that the above inequality stands for $\forall q \in \mathscr{Q}^{m-1}(\mathscr{D})$ so that we have

$$|L(v)| \le \|L\|^*_{\mathscr{W}^{m,p}(\mathscr{D})} \inf_{q\in\mathscr{Q}^{m-1}(\mathscr{D})} \|v+q\|_{\mathscr{W}^{m,p}(\mathscr{D})} \le C\|L\|^*_{\mathscr{W}^{m,p}(\mathscr{D})}\{v\}_{\mathscr{W}^{m,p}(\mathscr{D})},$$

where the last inequality is obtained by using Theorem 2.5.6. □

With the help of Theorem 2.5.8, a numerical integration error estimate can be obtained following the proof of Theorem 2.5.3.

Theorem 2.5.9. *(Global higher order numerical integration error theorem for parallelogram elements) Assume that* $a_{ij} \in \mathscr{W}^{2m,\infty}(\mathscr{D})$ *and* $u, w \in \mathscr{W}(\mathscr{D}) \cap \mathscr{H}^{m+1}(\mathscr{D})$ $\big(\mathscr{W}(\mathscr{D})$ *can be either defined as in* (2.2.5) *provided that* $\mathscr{D}$ *is a cube or defined as in* (2.2.6)$\big)$ *are the solutions of the following problems: for* $\forall v \in \mathscr{W}(\mathscr{D})$,

$$\int_{\mathscr{D}} a\nabla u\cdot\nabla v dx = F_1(v),$$
$$\int_{\mathscr{D}} a\nabla w\cdot\nabla v dx = F_2(v),$$

where F_1, F_2 *are two linear functional mapping* $\mathscr{W}(\mathscr{D}) \cap \mathscr{H}^{m+1}(\mathscr{D}) \to \mathbb{R}$ *and* u^h, w^h *are the FEM solutions with numerical quadrature in* $S^m(\mathscr{D}, \mathscr{T}_h)$ *where* $\mathscr{T}_h$ *is a shape regular partition of* $\mathscr{D}$ *and elements* $K \in \mathscr{T}_h$ *are parallelogram and quasi-uniform. Assume the numerical quadrature satisfies*

$$E_{\hat{K}}(\hat{\phi}) = 0, \ \ \forall \hat{\phi} \in \mathscr{Q}^{2m-1}(\hat{K}). \tag{2.5.83}$$

Then we have

$$\begin{aligned}&|\int_{\mathscr{D}}\sum_{ij}a_{ij}\partial_i u^h\partial_j w^h dx-\sum_{K\in\mathscr{T}_h}\sum_{j=1}^{J}\omega_{K,j}\sum_{ij}(a_{ij}\partial_i u^h\partial_j w^h)(x_{K,j})|\\&\leq Ch^{2m}(\sum_{ij}\|a_{ij}\|_{\mathscr{W}^{2m,\infty}(\mathscr{D})})\|u\|_{\mathscr{H}^{m+1}(\mathscr{D})}\|w\|_{\mathscr{H}^{m+1}(\mathscr{D})}.\end{aligned}$$

Remark 2.5.10. *If we assume here $E_{\hat{K}}(\hat{\phi})=0,\ \forall\hat{\phi}\in\mathscr{Q}^{4m-3}(\hat{K})$, then the assumption of the quasi-unform mesh in Theorem 2.5.9 can be removed. The proof can be easily obtained following the proofs of Theorem 2.5.4 and Theorem 2.5.5.*

2.6 Numerical Experiment

We consider problem (1.0.1) on the domain $\Omega=[0,1]^2$ with $f=1$ and the multiscale tensor

$$a^{\varepsilon}(x)=(x_1^2+x_2^2+1)\begin{pmatrix}\sin(2\pi\frac{x_1}{\varepsilon_1}+2)\cos(2\pi\frac{x_1}{\varepsilon_2}+2) & 0\\ 0 & \sin(2\pi\frac{x_2}{\varepsilon_1}+2)\cos(2\pi\frac{x_2}{\varepsilon_2}+2)\end{pmatrix}.$$

where $\varepsilon_1=5\times10^{-3}$, $\varepsilon_2=5\times10^{-5}$. The corresponding homogenized tensor is

$$a^0(x)=3(x_1^2+x_2^2+1)I_2.$$

The implementation of this experiment is a generalization of the two scale FE-HMM code provided in [19]. We use FEM with piecewise linear basis functions on triangle elements (called P1-FEM) as the solver for problems on all three scales. For the QFs, we choose the barycenter of each element as the quadrature node for both macro and meso scales which satisfies the assumptions on the QFs stated in both Theorem 2.3.4 and 2.4.6 for P1 triangle elements. We take the P1-FE solution u^{0,h_0} of (2.1.1) as the reference solution which is computed on a 1024×1024 uniform triangular mesh. We use uniform triangulation in the FE-HMM procedure and denote N_{mac},N_{mes},N_{mic} as the degrees of freedom (DOF) of one direction in the macro, meso and micro partitions respectively, i.e. $h_0=1/N_{mac}$, $h_1/\varepsilon_1=1/N_{mes}$, $h_2/\varepsilon_2=N_{mic}$. In Fig. 2, we observe that the $\mathscr{H}^1$ error decays with a rate of $\mathscr{O}(h_0)$ and the $\mathscr{L}^2$ error decays as $\mathscr{O}(h_0^2)$, which confirms the results in Theorem 2.3.4 and 2.4.6.

In Fig. 3(a)(b), we show the error behavior with N_{mes} is fixed at $2,4,8,16$ while macro and micro meshes are refined simultaneously. In Fig. 3(c)(d) N_{mic} is fixed at $2,4,8,16$ and macro, meso meshes are refined simultaneously. We conclude that the theoretical convergence rate can only be obtained when the meshes are refined simultaneously and that our error estimates (at least for three scale problems) are sharp.

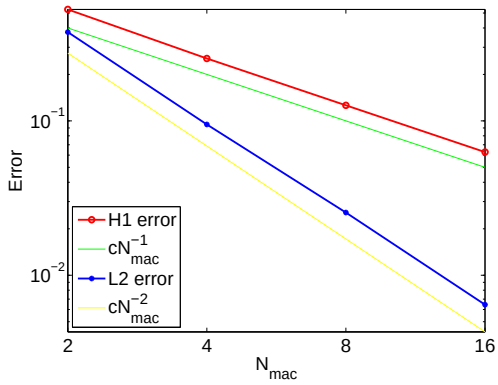
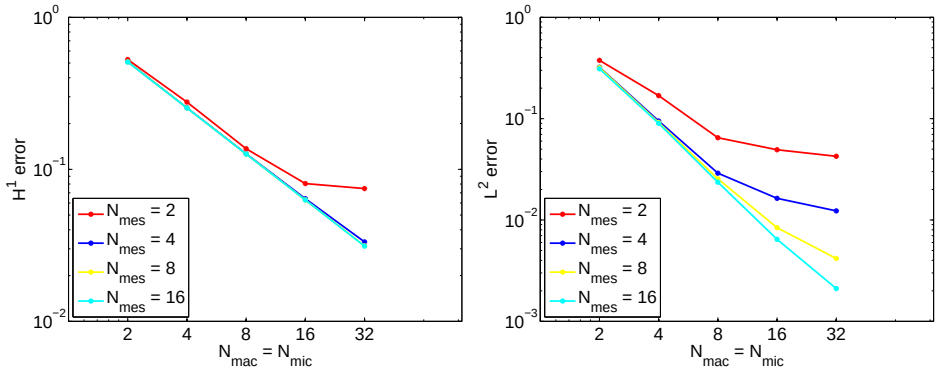


Figure 2: We set $N_{mes} = N_{mic} = N_{mac}^{1/2}$ for computing $\mathscr{H}^1$ error and $N_{mes} = N_{mic} = N_{mac}$ for $\mathscr{L}^2$ error.

2.7 Discussion

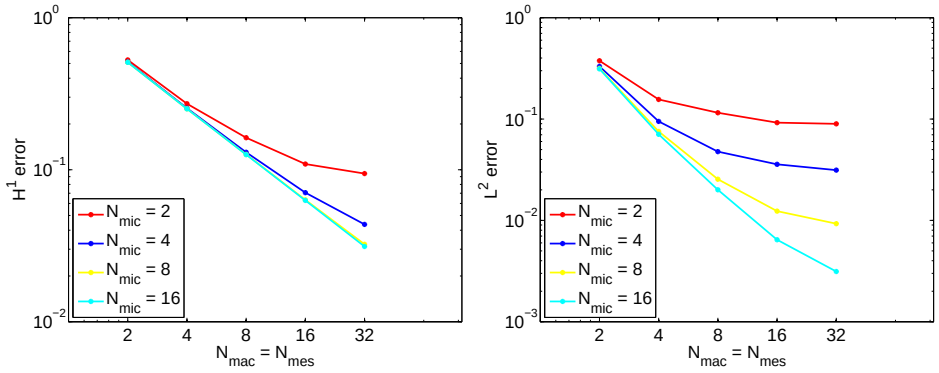
In this chapter, we present the generalized FE-HMM for the problems with more than two scales. Furthermore, we provide a fully discrete a priori error analysis for the $(N+1)$-scale FE-HMM where the error coming from numerical quadrature used in each scale is carefully discussed. We note that this analysis can be applied to the FE-HMM of any order. A complexity analysis is also added for the completeness.

Based on the work in this chapter, there are two aspects that are of interest to explore in the next step. For the first aspect, considering the massive meso and micro cell problems need to be computed, a model reduction technique is required to reduce the cost of the generalized FE-HMM. The second aspect is a concern for engineering applications where physical models in different scales can be different and thus in this case coupling different models in a consistent way should be taken into account.



(a) $\mathscr{H}^1$ Error with fixed N_{mes}

(b) $\mathscr{L}^2$ Error with fixed N_{mes}



(c) $\mathscr{H}^1$ Error with fixed N_{mic}

(d) $\mathscr{L}^2$ Error with fixed N_{mic}

Figure 3: The convergence behavior when N_{mic} or N_{mes} is fixed at different values.

Part II

Reduced basis finite element heterogeneous multiscale method for linear problems

Numerical homogenization methods show large efficiency improvement compared with the classical numerical methods for multiscale problems. In particular, the FE-HMM yields an approximation of an effective problem with complexity independent of ε. One common features shared by several numerical homogenization methods is that the global microscopic problem is split into many local cell problems which are coupled with the macro discretization. According to the complexity analysis of the fully discrete method (see (1.2.17), (1.2.18)), the global macro discretization and the micro discretizations for the cell problems have to be refined simultaneously in order to obtain the optimal convergence rate. For higher order macro methods (where more sampling domains are required) or high dimensional problems, these methods can become computationally very expensive (even though order of magnitude cheaper than a full fine scale approach). For the FE-HMM, attempts to reduce the computational cost have been pursued in [15], where fast micro solvers have been coupled with standard FEM. By selecting a special quadrature formula with integration points on the interfaces of the macro partition, one can also in some situations reduce the computational cost (this does however only reduce the constant in front of the computational cost for the FE-HMM, e.g., a reduction factor of one half is reported in [57] for two dimensional problems with first or second order macro solvers).

Reduced basis (RB) techniques for model reduction, pioneered in [65, 64, 86], have seen recently a renewed interest thanks to the development of new sampling techniques and rigorous a posteriori error bounds for outputs of interest [93] (see also [92, 96] for additional references on the recent literature). In the context of numerical homogenization, the use of RB was first proposed in [45, 46] emphasizing on parametrizing various configurations of cell problems (e.g. inclusion with various shapes, etc.), here, building on [45, 46], we focus on integrating the RB methodology in a micro macro FEM such as the FE-HMM and providing fully discrete error analysis for our new approach.

Outline of Part II.

- Chapter 3: Gives an introduction on the RB method for coercive parametrized linear elliptic problems as well as some implementation details which are used in the later chapters.
- Chapter 4: Proposes the RB-FE-HMM for linear multiscale problems and shows several numerical experiments. This chapter is taken from [7, Section 3 - Section 5].
- Chapter 5: Presents the further exploration of the RB-FE-HMM for adaptive procedures. The chapter is taken from [8, Section 3 - Section 6] with small modifications.

3 Reduced basis method

Many engineering problems can be modeled by PDEs with input parameters, called the parametrized PDEs. The input parameters can be the shape deformation of the modeling objects or physical elements of the environment. The solutions of the parametrized PDEs vary with respect to different input parameters but share some similarities. In the aspect of numerical computation, the parametrized PDE needs to be re-solved when the input parameters are changed. For problems that have many parameters of interest, the numerical computation can be very costly (both time and storage) and real time results cannot be achieved for complex 3D problems or time dependent problems. Motivated by the "real time" and "many query" contexts, the reduced basis (RB) method is developed that aims at reducing the computational cost by a specially constructed lower dimensional space such that real time simulation can be achieved. In this chapter, we introduce the general reduced basis (RB) method for parametrized linear elliptic coercive PDEs [96]. This idea can be similarly developed for parabolic problems [67] and nonlinear problems [76].

3.1 The reduced basis method for linear elliptic coercive PDEs

To be consistent with the literature of general RB method and to distinguish from our multiscale problems, we use the following weak form for the general elliptic problem: Given the input parameter $\mu \in \mathscr{D} \subset \mathbb{R}^d$, there exists a unique $u^e(\mu) \in V^e(\Omega)$ that satisfies

$$a(u^e(\mu), v; \mu) = f(v) \quad \forall v \in V^e(\Omega), \tag{3.1.1}$$

where $a(\cdot,\cdot;\mu)$ is a bilinear form, $f(\cdot) : V^e \to R$ is a parameter independent linear functional and $V^e(\Omega)$ is a Hilbert space which the exact solution $u^e(\mu)$ lies in, equipped with inner product $(\cdot,\cdot)_V$ and norm $\|\cdot\|_V$.

We are interested in the quantity

$$s^e(\mu) = L(u^e(\mu)), \tag{3.1.2}$$

where $L(\cdot): V^e \to \mathbb{R}$ denotes a linear output functional. The choice $L(u^e(\mu)) = f(u^e(\mu))$ is often considered in RB literature which will also be considered. We assume that the bilinear form $a(\cdot,\cdot,\mu)$ is symmetric, bounded and coercive for any $\mu \in \mathscr{D}$. The boundedness is defined as: There exists a constant $\gamma_0 > 0$ such that

$$\gamma_0 \geq \gamma(\mu) := \sup_{w \in V^e} \sup_{v \in V^e} \frac{a(w,v;\mu)}{\|v\|_V \|w\|_V}. \tag{3.1.3}$$

and the coercivity is defined as: There exists a constant $\alpha_0 > 0$ such that

$$0 < \alpha_0 \leq \alpha(\mu) := \inf_{w \in V^e} \frac{a(w,w;\mu)}{\|w\|_V^2}. \tag{3.1.4}$$

Furthermore, we assume the bilinear form $a(\cdot,\cdot)$ can be written in the following affine form, i.e.

$$a(w,v;\mu) = \sum_{q=1}^{Q} \Theta^q(\mu) a^q(w,v), \quad \forall v, w \in V^e(\Omega),\ \mu \in \mathscr{D}, \tag{3.1.5}$$

where $a^q(w,v)$ is a parameter independent bilinear form and $\Theta^q(\mu)$ is a function on $\mathscr{D}$ with variable μ. This assumption is critical to achieve efficiency for the RB method. For the cases that the affine representation is not available, we can apply the so-called empirical interpolation method (EIM) (to be introduced in Section 3.3) to construct for the original bilinear form an approximation in the affine form .

3.1.1 Offline-online strategy

Assume that $V^{\mathcal{N}}(\Omega)$ is the finite element (FE) approximation space of $V^e(\Omega)$ with number of degrees of freedom (DOF) $\mathcal{N}$. Solving (3.1.1) in $V^{\mathcal{N}}$ for any given μ reads: Find $u^h \in V^{\mathcal{N}}$ such that

$$a(u^{\mathcal{N}}(\mu), v^{\mathcal{N}}; \mu) = f(v^{\mathcal{N}}) \quad \forall v^{\mathcal{N}} \in V^{\mathcal{N}}(\Omega). \tag{3.1.6}$$

The RB method aims at building a small dimensional space (the RB space) $V_N^{\mathcal{N}} \subset V^{\mathcal{N}}$ in an offline stage so that (3.1.1) can be solved in $V_N^{\mathcal{N}}$ in an online stage where real time results can be obtained. The notation $V_N^{\mathcal{N}}$ represents that the dimension of this space is N and its basis functions belong to $V^{\mathcal{N}}$.

Algorithm 3.1.1. *(RB offline stage)*

1. *Sample a large training set* $\Xi_{train} = \{\mu_i\}_{i=1}^{N_{train}} \subset \mathscr{D}$ *and compute* $u^{\mathcal{N}}(\mu_1)$ *of* (3.1.6). *Let*

$\xi_1^{\mathcal{N}} = \frac{u^{\mathcal{N}}(\mu_1)}{\|u^{\mathcal{N}}(\mu_1)\|_V}$, $V_1^{\mathcal{N}} = span\{\xi_1^{\mathcal{N}}\}$ *and* $l = 1$.

2. *Compute the a posteriori error estimator* $\Delta_l(\mu_i)$, $i = 1, \cdots, N_{train}$. *If* $max_{i=1,\cdots,N_{train}} \Delta_l(\mu_i) > tol$, *select* $\mu_{l+1} = argmax_{i=1,\cdots,N_{train}} \Delta_l(\mu_i)$ *and go to next step, otherwise offline ends.*

3. *Compute* $u^{\mathcal{N}}(\mu_l)$ *of* (3.1.6) *and denote by* $\xi_l^{\mathcal{N}}$ *the orthogonalization of* $u^{\mathcal{N}}(\mu_l)$ *to* $V_l^{\mathcal{N}}$. *Let* $V_{l+1}^{\mathcal{N}} = V_l^{\mathcal{N}} \bigoplus span\{\xi_l^{\mathcal{N}}\}$ *and* $l = l + 1$. *Go to step 2.*

We store the following offline outputs for the online stage:

$$A_{ij}^q := a^q(\xi_i^{\mathcal{N}}, \xi_j^{\mathcal{N}}), \ F_i := f(\xi_i^{\mathcal{N}}), \ i, j = 1, \cdots, N, \ q = 1, \cdots, Q \tag{3.1.7}$$

where A^q is the $N \times N$ affine stiffness matrix and F is the $N \times 1$ affine vector (i.e. A^q and F are parameter independent). Generally, the dimension of $V_N^{\mathcal{N}}$ is rather small such that the memory requirement to store the offline outputs is low. The issue left for the offline process is how to design the a posteriori error estimator. This point will be discussed in Section 3.1.2.

In the online stage, we solve for any given parameter μ: Find $u_N^{\mathcal{N}}(\mu) \in V_N^{\mathcal{N}}$ such that

$$\sum_{q=1}^{Q} \Theta^q(\mu) a^q(u_N^{\mathcal{N}}(\mu), v_N^{\mathcal{N}}) = f(v_N^{\mathcal{N}}), \ \forall v_N^{\mathcal{N}} \in V_N^{\mathcal{N}}, \tag{3.1.8}$$

where $u_N^{\mathcal{N}}(\mu)$ is presented by the linear combination of the basis functions of $V_N^{\mathcal{N}}$ i.e. $u_N^{\mathcal{N}}(\mu) = \sum_{l=1}^{N} \beta_l(\mu) \xi_l^{\mathcal{N}}$ and $\beta(\mu) = \big(\beta_1(\mu), \cdots, \beta_N(\mu)\big)^T \in \mathbb{R}^N$. Therefore the corresponding linear system of (3.1.8) can be written as

$$(\sum_{q=1}^{Q} \Theta^q(\mu) A^q)\beta(\mu) = F, \tag{3.1.9}$$

and the quantity of interest (3.1.2) is simply $L(u_N^{\mathcal{N}}(\mu)) = F^T \beta(\mu)$. The affine assumption of the bilinear form makes the offline outputs easy to use for the online computation. We emphasize that the linear system 3.1.9 has small DOF which allows to obtain the real-time result.

Remark 3.1.2. *In order to guarantee the precision of the online solutions, it is often required to use a fine FE space* $V^{\mathcal{N}}$ *in the offline stage i.e.* $\mathcal{N}$ *is large and this makes the offline stage costly. However since the offline outputs* A^q *and* F *are parameter independent, they can be repeatedly used in the online stage for different given parameters. Therefore the offline stage is only needed once.*

3.1.2 A posteriori error estimator

We denote that $u_l^{\mathcal{N}}(\mu)$ is the solution of (3.1.1) computed in the RB space $V_l^{\mathcal{N}}$. We would like to control the error $\|\hat{e}_l(\mu)\|_V$ where $\hat{e}_l(\mu) := u^{\mathcal{N}}(\mu) - u_l^{\mathcal{N}}(\mu)$. First we have the following equation

$$a(\hat{e}_l(\mu), v^{\mathcal{N}}; \mu) = f(v^{\mathcal{N}}) - a(u_l^{\mathcal{N}}(\mu), v^{\mathcal{N}}; \mu), \ \ \forall v^{\mathcal{N}} \in V^{\mathcal{N}}.$$

Using Riesz representation theory, there exists a unique $\bar{e}_l(\mu) \in V^{\mathcal{N}}$ such that

$$(\bar{e}_l(\mu), v^{\mathcal{N}})_V = f(v^{\mathcal{N}}) - a(u_l^{\mathcal{N}}(\mu), v^{\mathcal{N}}; \mu). \tag{3.1.10}$$

We define the a posteriori error estimator as

$$\Delta_l(\mu) = \frac{\|\bar{e}_l(\mu)\|_V}{\alpha_{LB}(\mu)},$$

where $\alpha_{LB}(\mu)$ is an approximation of the coercivity factor $\alpha(\mu)$ that satisfies $\alpha(\mu) \geq \alpha_{LB}(\mu) > 0$. We define the energy norm $\|v\|_{\mathcal{E},\mu} = \sqrt{a(v, v; \mu)}$ and the following result is shown in [96]

$$1 \leq \frac{\Delta_l(\mu)}{\|\hat{e}_l(\mu)\|_{\mathcal{E},\mu}} \leq \sqrt{\frac{\gamma(\mu)}{\alpha_{LB}(\mu)}}, \quad \forall \mu \in \mathcal{D}$$

where $\gamma(\mu)$ is defined in (3.1.3). This result indicates that the error $\|\hat{e}_l(\mu)\|_{\mathcal{E},\mu}$ can be bounded on both sides by the a posteriori estimator such that the offline outputs are "certified".

To obtain the a posteriori error, one needs to solve (3.1.10) in the fine FE space $V^{\mathcal{N}}$ (as mentioned in Remark 3.1.2) for all the samples from Ξ_{train} in each loop of enlarging the RB space, which could make the offline stage prohibitively costly. Thanks to the affine bilinear form and the linearity of equation (3.1.10), we can decompose (3.1.10) the following problems: Find $c^{\mathcal{N}}, L_l^{\mathcal{N},q} \in V_N^{\mathcal{N}}$ such that

$$(c^{\mathcal{N}}, v^{\mathcal{N}}) = f(v^{\mathcal{N}}), \ \ \forall v^{\mathcal{N}} \in V_N^{\mathcal{N}} \tag{3.1.11}$$

$$(L_i^{\mathcal{N},q}, v^{\mathcal{N}}) = a^q(\xi_i^{\mathcal{N}}, v^{\mathcal{N}}) \ \ \forall v^{\mathcal{N}} \in V_N^{\mathcal{N}}, \ i = 1, \cdots, l, \ q = 1, \cdots, Q. \tag{3.1.12}$$

Therefore we obtain $\bar{e}_l(\mu) = c^{\mathcal{N}} + \sum_{q=1}^{Q} \Theta^q(\mu) \sum_{i=1}^{l} \beta_i(\mu) L_i^{\mathcal{N}}$. Therefore, we have

$$\|\bar{e}_l(\mu)\|_V^2 \quad = \quad (c^{\mathcal{N}}, c^{\mathcal{N}})_V + 2 \sum_{q=1}^{Q} \Theta^q(\mu)(c^{\mathcal{N}}, L_i^{\mathcal{N},q})_V$$

$$+ \sum_{p,q=1}^{Q} \Theta^q(\mu)\Theta^p(\mu) \sum_{i,j=1}^{l} \beta_i(\mu)\beta_j(\mu)(L_i^{\mathcal{N},q}, L_j^{\mathcal{N},p})_V. \tag{3.1.13}$$

We note that (3.1.11) and (3.1.12) are parameter independent that can be pre-computed only once in the offline stage. Thus the computation of $\|\bar{e}_l(\mu)\|_V^2$ becomes very efficient.

3.2 Coercivity factor approximation $\alpha_{LB}(\mu)$

Now we introduce how to compute the coercivity factor $\alpha_{LB}(\mu)$ to complete the issues of the RB offline stage. The direct way is to solve for each $\mu \in \Xi_{train}$

$$\alpha^{\mathcal{N}}(\mu) = \inf_{v^{\mathcal{N}} \in V^{\mathcal{N}}} \frac{a(v^{\mathcal{N}}, v^{\mathcal{N}}; \mu)}{\|v^{\mathcal{N}}\|_V}, \tag{3.2.14}$$

which is equivalent to solve an eigenvalue problem in $V^{\mathcal{N}}$. Again due to the requirement of the fine FE space $V^{\mathcal{N}}$ and a large initial training set of parameters in the offline stage, directly solving the eigenvalue problem has huge computational cost. Therefore, we need to have more efficient methods to obtain a lower bound α_{LB} of $\alpha^{\mathcal{N}}$ such that $0 < \alpha_{LB}(\mu) \leq \alpha^{\mathcal{N}}(\mu),\ \forall \mu \in \mathscr{D}$. The most popular methods coupled with the RB method are the "min Θ" method [92] and the successive constraint method (SCM) [72, 96] (the SCM for non-symmetric and non-coercive problems are also discussed in these references). The basic assumptions for both of the methods is that $a(\cdot,\cdot;\mu)$ has affine representation (3.1.5). In what follows, we will briefly introduce the two methods and for simplicity we still assume the coercivity and symmetry of $a(\cdot,\cdot;\mu)$.

3.2.1 The "min Θ" method

For the min Θ method, we need to have one more critical assumption: The bilinear form $a(\cdot,\cdot;\mu)$ is affine coercive, i.e.

$$\Theta^q(\mu) > 0, \ \ \forall \mu \in \mathscr{D} \text{ and } a^q(v,v) \geq 0, \ \ \forall v \in V^e(\Omega). \tag{3.2.15}$$

Then we define the lower bound of $\alpha^{\mathcal{N}}(\mu)$ as

$$\alpha_{LB}(\mu) = \min_{q=1,\cdots,Q} \frac{\Theta^q(\mu)}{\Theta^q(\mu^*)} \alpha^{\mathcal{N}}(\mu^*), \tag{3.2.16}$$

where μ^* is any fixed parameter in $\mathscr{D}$.

We show here that α_{LB} satisfies the inequality in (3.1.4): For $\forall v^{\mathcal{N}} \in V^{\mathcal{N}}, \mu \in \mathcal{D}$

$$\begin{aligned} a(v^{\mathcal{N}}, v^{\mathcal{N}}; \mu) &= \sum_{q=1}^{Q} \Theta^q(\mu) a^q(v^{\mathcal{N}}, v^{\mathcal{N}}) = \sum_{q=1}^{Q} \frac{\Theta^q(\mu)}{\Theta^q(\mu^*)} \Theta^q(\mu^*) a^q(v, v) \\ &\geq \min_{q=1,\cdots,Q} \frac{\Theta^q(\mu)}{\Theta^q(\mu^*)} \sum_{q=1}^{Q} \Theta^q(\mu*) a^q(v^{\mathcal{N}}, v^{\mathcal{N}}) \\ &\geq \min_{q=1,\cdots,Q} \frac{\Theta^q(\mu)}{\Theta^q(\mu^*)} \alpha^{\mathcal{N}}(\mu^*) \| v^{\mathcal{N}} \|_V. \end{aligned}$$

Therefore, we have $\alpha_{LB}(\mu) \leq \alpha^{\mathcal{N}}(\mu) = \inf_{v^{\mathcal{N}} \in V^{\mathcal{N}}} \frac{a(v^{\mathcal{N}}, v^{\mathcal{N}}; \mu)}{\| v^{\mathcal{N}} \|_V}$.

However the strict assumption (3.2.15) makes this method prohibitive for general problems. In contrast, the SCM has great generality.

3.2.2 The successive constraint method

The SCM consists of an offline stage launched before the RB offline stage and an online stage launched during the RB offline stage to compute $\alpha_{LB}(\mu)$ for each $\mu \in \Xi_{train}$.

We first introduce a set $\boldsymbol{Z} \subset \mathbb{R}^Q$ that

$$\boldsymbol{Z} := \left\{ z \in \mathbb{R}^Q | \; z = \{z_1, \cdots, z_Q\}^T, \; \exists v_z^{\mathcal{N}} \in V^{\mathcal{N}} \text{ s.t. } z_q = \frac{a^q(v_z^{\mathcal{N}}, v_z^{\mathcal{N}})}{\| v_z^{\mathcal{N}} \|_V}, q = 1, \cdots, Q \right\}.$$

It is easy to see that $\alpha^{\mathcal{N}}(\mu) = \inf_{z \in \boldsymbol{Z}} \sum_{q=1}^{Q} \Theta^q(\mu) z_q$. The idea of the SCM is to find a subset of $\boldsymbol{Z}$ such that one can efficiently obtain an lower bound of $a^{\mathcal{N}}(\mu)$ by solving a minimization problem in this subset.

Next we introduce the so-called "coercivity constraint box"

$$\boldsymbol{B} = \prod_{q=1}^{Q} \Big[\inf_{v^{\mathcal{N}} \in V^{\mathcal{N}}} \frac{a^q(v^{\mathcal{N}}, v^{\mathcal{N}})}{\| v^{\mathcal{N}} \|_V}, \sup_{v^{\mathcal{N}} \in V^{\mathcal{N}}} \frac{a^q(v^{\mathcal{N}}, v^{\mathcal{N}})}{\| v^{\mathcal{N}} \|_V} \Big]. \tag{3.2.17}$$

We define the following two sets that will be used in the SCM algorithm:

$$\boldsymbol{C}_J = \{\mu_1^S, \cdots, \mu_J^S\}, \quad (\text{ define } \boldsymbol{C}_0 = \emptyset)$$

where $\mu_j^S \in \mathcal{D}$ is selected by the greedy algorithm in the SCM offline stage. Define $\boldsymbol{C}_J^M(\mu)$ a set that consists of M elements in $\boldsymbol{C}_J$ which are closest to the given μ and we have $\boldsymbol{C}_J^M(\mu) \subset \boldsymbol{C}_J$ (when $J \leq M$, $\boldsymbol{C}_J^M(\mu) = \boldsymbol{C}_J$). Practically, we choose $M = 2^d$ where d is the dimension of Ω.

Given K_{max} the max size of set $\boldsymbol{C}_J$ generated in the SCM offline stage and the tolerance tol_{SCM} as the stopping criteria, the offline process is stated as follows.

Algorithm 3.2.1. *(the SCM offline stage)*

1. Select a large training set $\Xi^S_{train} = \{\mu_i\}_{i=1}^{N^S_{train}}$, $\mu_i \in \mathscr{D}$. *Construct the "continuity constraint" box* $\boldsymbol{B}$ *defined in* (3.2.17).

2. While $J < K_{max}$*:*

a. Set $\boldsymbol{C}_J = \boldsymbol{C}_{J-1} \cup \{\mu^S_J\}$ *and when* $J = 1$, μ^S_1 *is randomly chosen from* Ξ^S_{train}.

b. Compute $\alpha^{\mathscr{N}}(\mu^S_J)$ *and the corresponding eigen function* $v^{S,\mathscr{N}}_J$ *by solving* (3.2.14). *Next compute* $z^J \in \boldsymbol{Z}$ *where* $z^J_q = \frac{a^q(v^{S,\mathscr{N}}_J, v^{S,\mathscr{N}}_J)}{\|v^{S,\mathscr{N}}_J\|_v}, q = 1, \cdots, Q$. *Let set* $\boldsymbol{E}_J = \boldsymbol{E}_{J-1} \cup \{\alpha^{\mathscr{N}}(\mu^S_J)\}$ *and* $\boldsymbol{Z}^{UB}_J = \boldsymbol{Z}^{UB}_{J-1} \cup \{z^J\}$ *(define* $\boldsymbol{E}_0 = \emptyset$ *and* $\boldsymbol{Z}^{UB}_0 = \emptyset$*).*

c. For $\mu_k \in \Xi^S_{train}$, $k = 1, \cdots, N^S_{train}$*:*

(i). We solve

$$\alpha_{LB}(\mu_k) = \min_{z \in \boldsymbol{Z}^{LB}_{M,J}(\mu_k)} \sum_{q=1}^{Q} \Theta^q(\mu_k) z_q,$$

where $\boldsymbol{Z}^{LB}_{M,J}(\mu_k) := \{z \in \mathbb{R}^Q | z \in \boldsymbol{B}$ *and* $\sum_{q=1}^{Q} \Theta^q(\mu_k) z_q \geq \alpha^{\mathscr{N}}(\mu), \forall \mu \in \boldsymbol{C}^M_J(\mu_k)\}$. *We emphasize that all the* $\alpha^{\mathscr{N}}(\mu)$ *where* $\mu \in \boldsymbol{C}^M_J(\mu_k)$ *are stored in set* $\boldsymbol{E}_J$.

(ii). Solve

$$\alpha_{UB}(\mu_k) = \min_{z \in \boldsymbol{Z}^{UB}_{M,J}(\mu_k)} \sum_{q=1}^{Q} \Theta^q(\mu_k) z_q,$$

where $\boldsymbol{Z}^{UB}_{M,J}(\mu_k)$ *is a subset of* $\boldsymbol{Z}^{UB}_J$*, containing the elements in* $\boldsymbol{Z}^{UB}_J$ *that are indexed by the parameters in* $\boldsymbol{C}^M_J(\mu_k)$. *Since* $\boldsymbol{Z}^{UB}_{M,J}(\mu_k) \subset \boldsymbol{Z}$*, therefore* $\alpha_{UB}(\mu_k) \geq \alpha(\mu_k)$.

(iii). Calculate $\epsilon_J(\mu_k) = \frac{\alpha_{UB}(\mu_k) - \alpha_{LB}(\mu_k)}{\alpha_{UB}(\mu_k)}$.

d. If $\max_{k=1,\cdots,N^S_{train}} \epsilon_J(\mu_k) < tol_{SCM}$*, the offline process ends. Otherwise let* $J = J+1$ *and go to a.*

The outputs of the SCM offline stage are $\boldsymbol{B}, \boldsymbol{C}_J$ and $\boldsymbol{E}_J$. In the SCM online stage, for any given parameter $\mu \in \mathscr{D}$, we just need to solve the following linear optimization problem

$$\alpha_{LB}(\mu) = \min_{z \in \boldsymbol{Z}^{LB}_{M,J}(\mu)} \sum_{q=1}^{Q} \Theta^q(\mu) z_q.$$

It is proved in [96] that $\frac{\alpha_{\mathscr{N}}(\mu)}{\alpha_{LB}(\mu)} \leq \frac{1}{1-\epsilon_{SCM}}$, $\forall \mu \in \mathscr{D}$, where $\epsilon_{SCM} = \max_{\mu \in \Xi^S_{train}} \epsilon_J(\mu)$.

3.3 The empirical interpolation method

The empirical interpolation method (EIM) is proposed in [37] in order to obtain the affine representation for general parametrized functions in form of $g(\mu,x)$. We assume that $g(\mu,x)$ is uniformly continuous on $\mathscr{D}$ and belongs to $\mathscr{L}^\infty(\Omega)$. The goal of the EIM is to construct a few basis functions $\{q_1,\cdots,q_M\}$ that span a low dimensional space and based on this space one can efficiently obtain an affine approximation for $g(\mu,x)$ expressed as $g_M(\mu,x)=\sum_{m=1}^{M}\phi_m(\mu)q_m(x)$ for any given μ. The EIM is also an offline-online strategy based on the greedy algorithm. The EIM offline process consists of two main ingredients: The first ingredient is to select the representative parameters μ from $\mathscr{D}$ and obtain the corresponding functions $g(\mu,\cdot)$ and the second one is to find the interpolation points x in Ω in order to discretize the function $g(\cdot,x)$. The detailed offline process is stated as what follows.

Algorithm 3.3.1. *(The EIM offline stage)*
1. Choose a large training set $\Xi^E_{train}=\{\mu_i\}_{i=1}^{N^E_{train}}\subset\mathscr{D}$ *(we use the superscript E to distinguish from the RB offline notations) and select* $\mu_1^g=argmax_{\mu\in\Xi^E_{train}}\|g(\mu,x)\|_{\mathscr{L}^\infty(\Omega)}$. *Set* $\eta_1(x)=g(\mu_1^g,x)$ *and* $x_1=argsup_{x\in\Omega}|\eta_1(x)|$ [1] *and* $q_1(x)=\frac{\eta_1(x)}{|\eta_1(x_1)|}$. *Let* $m=1$ *and define* $S_1^{E,g}=span\{q_1\}$ *and* $B^1=q_1(x_1)$.

2.Compute for each $\mu\in\Xi^E_{train}$ *that* $\delta_m(\mu)=\|g(\mu,x)-g_m(\mu,x)\|_{\mathscr{L}^\infty(\Omega)}$ *where* $g_m(\mu,x)\in S_m^{E,g}$. *If* $max_{\mu\in\Xi^E_{train}}\delta_m(\mu)<tol^E$, *let* $M=m$ *and the EIM offline process stops; Otherwise:*

i. Select $\mu_{m+1}=argmax_{\mu\in\Xi^E_{train}}\delta_m(\mu)$ *and let* $\eta_{m+1}(x)=g(\mu_{m+1},x)$;

ii. Solve $\sum_{j=1}^{m}\sigma_j^m q_j(x_i)=\eta_{m+1}(x_i), i=1,\cdots,m$;

iii. Let $r_{m+1}(x)=\eta_{m+1}(x)-\sum_{j=1}^{m}\sigma_j^m q_j(x)$;

iv. Denote $x_{m+1}=argsup_{x\in\Omega}|r_{m+1}(x)|$, $q_{m+1}(x)=\frac{r_{m+1}(x)}{|r_{m+1}(x_{m+1})|}$ *and* $B_{ij}^{m+1}=q_j(x_i), i,j=1,\cdots,m+1$, *where* B^{m+1} *is a matrix with dimension* $(m+1)\times(m+1)$.

v. We have $S_{m+1}^{E,g}=span\{q_1,\cdots,q_{m+1}\}$. *Let* $m=m+1$ *and go back to Step 2.*

The outputs of the EIM offline stage are the basis functions $\{q_1,\cdots,q_M\}$, the matrix B^M with dimension $M\times M$ as well as the set of the interpolation points $\{x_1,\cdots,x_M\}$.

In the online stage of the EIM, we construct the affine approximation $g_M(\mu,x)=\sum_{m=1}^{M}\phi_m(\mu)q_m(x)$ of $g(\mu,x)$ for any given $\mu\in\mathscr{D}$. To obtain the coefficients $\phi_m(\mu),\ m=1,\cdots,M$, one just needs to solve a small linear system: $\sum_{j=1}^{M}B_{ij}^M\phi_j(\mu)=g(\mu,x_i),\ \ i=1,\cdots,M$, which has very low time cost. The upper bound of the error introduced by the EIM is analyzed in [63] and numerical experiments show that the EIM is an efficient and reliable strategy.

[1] To be more precise, here the $\sup_{x\in\Omega}|\eta_1(x)|$ should be understood as essential supremum.

4 Reduced Basis finite element heterogeneous multiscale method for linear problems

In this chapter and next, we consider the linear elliptic problems discussed in Chapter 1 in a bounded polyhedron domain $\Omega \subset \mathbb{R}^d$, $d \leq 3$,

$$\begin{aligned} -\nabla \cdot (a^\varepsilon(x)\nabla u^\varepsilon(x)) &= f \quad \text{in } \Omega, \\ u^\varepsilon(x) &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{4.0.1}$$

where we assume that $a^\varepsilon(x) = a(x, \frac{x}{\varepsilon})$ (with macro-micro scale separation) is symmetric uniformly elliptic and bounded (see in (1.0.1)).

As can be seen from the discussion in Chapter 1, the main cost of the FE-HMM comes from the computation of the cell problems, whose number and DOF increase as we refine the macro mesh for an appropriate approximation of the homogenized solution. In this section we explain how RB can be coupled to the FE-HMM to drastically reduce the cost of solving repeatedly a large number of cell problems.

4.1 Parametrized micro problems and numerical homogenized tensor

In what follows, it will be convenient to change several notations defined in Chapter 1. We denote the micro FE space by $S^q(K_{\delta_j}, \mathcal{N})$ instead of $S^q(K_{\delta_j}, \mathcal{T}_h)$ to emphasize on the degrees of freedom (DOF) $\mathcal{N}$ of the micro FE space. Likewise, the micro function $v^h_{K_j}$, solution of problem (1.2.16) will be denoted by $v_{\mathcal{N},K_j}$. We first notice that $v_{\mathcal{N},K_j}$ can be decomposed as

$$v_{\mathcal{N},K_j}(x) = v^H_{lin,j}(x) + \sum_{i=1}^{d} \chi^i_{\mathcal{N},K_j}(x) \frac{\partial v^H_{lin,j}}{\partial x_i}. \tag{4.1.2}$$

where $\chi^i_{\mathcal{N},K_j}(x),\ i=1,\dots,d$ are solutions of

$$\int_{K_{\delta_j}} a^\varepsilon(x)\nabla\chi^i_{\mathcal{N},K_j}(x)\cdot\nabla z_{\mathcal{N}}(x)dx = -\int_{K_{\delta_j}} a^\varepsilon(x)\mathbf{e}_i\cdot\nabla z_{\mathcal{N}}(x)dx \quad \forall z_{\mathcal{N}}\in S^q(K_{\delta_j},\mathcal{N}). \tag{4.1.3}$$

We now map a sampling domain K_{δ_j} in the reference domain Y through $x = G_{x_{K_j}}(y) = x_{K_j}+\delta y$ and consider $\hat{\chi}^i_{\mathcal{N},K_j}$ the solution of

$$\begin{aligned} b(\hat{\chi}^i_{\mathcal{N},K_j},\hat{z}_{\mathcal{N}}) &:= \int_Y a_{x_{K_j}}(y)\nabla\hat{\chi}^i_{\mathcal{N},K_j}(y)\cdot\nabla\hat{z}_{\mathcal{N}}(y)dy \\ &= -\int_Y a_{x_{K_j}}(y)\mathbf{e}_i\cdot\nabla\hat{z}_{\mathcal{N}}(y)dy =: l_i(\hat{z}_{\mathcal{N}})\ \ \forall\hat{z}_{\mathcal{N}}\in S^q(Y,\mathcal{N}), \end{aligned} \tag{4.1.4}$$

where we note that $a^\varepsilon(G_{x_{K_j}}(y))$ can be parametrized by $x_{K_j}\in\Omega$ and we therefore use the notation $a_{x_{K_j}}(y) := a^\varepsilon(G_{x_{K_j}}(y))$. The FE space $S^q(Y,\mathcal{N})$ has a triangulation $\mathcal{T}_{\hat{h}}$ with $\mathcal{N} = \mathcal{O}(\hat{h}^{-d})$ denoting its degrees of freedom. Functions in $S^q(Y,\mathcal{N})$ will have a subscript $\mathcal{N}$ (e.g., $\hat{z}_{\mathcal{N}}$). It is easily seen that

$$v_{\mathcal{N},K_j} = v^H_{lin,j}(x) + \delta\sum_{i=1}^d \hat{\chi}^i_{\mathcal{N},K_j}(G^{-1}_{x_{K_j}}(x))\frac{\partial v^H_{lin,j}}{\partial x_i}. \tag{4.1.5}$$

The following reformulation of the FE-HMM makes a link between the micro problems and the effective tensor obtained by the above micro-macro procedure. We have [5, Lemma 5.4],[4]

$$\frac{1}{|K_{\delta_j}|}\int_{K_{\delta_j}} a^\varepsilon(x)\nabla v_{\mathcal{N},K_j}(x)\cdot\nabla w_{\mathcal{N},K_j}(x)dx = a^0_{\mathcal{N}}(x_{K_j})\nabla v^H(x_{K_j})\cdot\nabla w^H(x_{K_j}). \tag{4.1.6}$$

Inserting (4.1.6) in (1.2.15) we obtain

$$B_H(v^H,w^H) := \sum_{K\in\mathcal{T}_H}\sum_{j=1}^J \omega_{K_j}a^0_{\mathcal{N}}(x_{K_j})\nabla v^H_{lin,j}(x_{K_j})\cdot\nabla w^H(x_{K_j}), \tag{4.1.7}$$

where

$$(a^0_{\mathcal{N}}(x_{K_j}))_{ik} = \int_Y a_{x_{K_j}}(y)\left(\nabla\hat{\chi}^i_{\mathcal{N},K_j}(y)+\mathbf{e}_i\right)\cdot\left(\nabla\hat{\chi}^k_{\mathcal{N},K_j}(y)+\mathbf{e}_k\right)dy. \tag{4.1.8}$$

and $\hat{\chi}^i_{\mathcal{N},K_j},\hat{\chi}^k_{\mathcal{N},K_j}$ are the solutions of (4.1.4).

Remark 4.1.1. *Similarly to* (4.1.7)*, we have the following bilinear form*

$$B_H(v^H,w^H) := \sum_{K\in\mathcal{T}_H}\sum_{j=1}^J \omega_{K_j}\bar{a}^0(x_{K_j})\nabla v^H(x_{K_j})\cdot\nabla w^H(x_{K_j}), \tag{4.1.9}$$

where $\bar{a}^0$ *is obtained from* (4.1.8)*, assuming the solutions of problem* (4.1.3) *are computed exactly. For the analysis in Section 4.5, we also need the auxiliary problem corresponding to the FE discretization with numerical quadrature of the homogenized problem* (1.1.3)*, i.e., the solution* $u^{0,H}$ *of* (1.2.10) *in Chapter 1.*

4.2 Model reduction

Inspired by the parametrization of the solutions of the micro problems (4.1.4) in the reference domain, we now describe a model reduction strategy for micro functions used in the FE-HMM. The overall idea is the following. Instead of computing micro functions in each macro elements at the quadrature points, we identify a small number N of carefully precomputed micro functions (to construct the RB space), whose supports can be chosen in the whole computational domain as sketched in Fig.1 (offline stage).

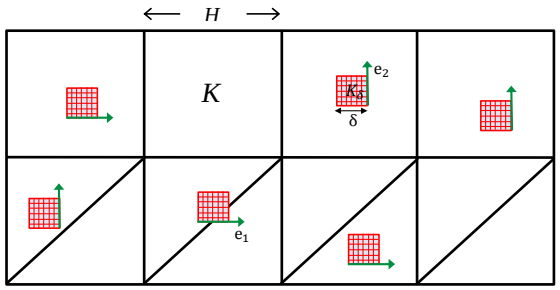


Figure 1: The supports of the RB functions.

In the online stage, the solution of cell problems at the given quadrature points of the macro elements are then computed in the RB space. No mesh, neither stiffness matrix assembly is needed for these later problems which require only the solution of small linear systems of size $N \times N$. Let $T_\delta = x_\tau + (-\delta/2, \delta/2)^d$ be a sampling domain centered at $x_\tau \in \Omega$, chosen such that $T_\delta \subset \Omega$. For $\{(T_\delta, \mathbf{e}_\eta); T_\delta \subset \Omega, \eta = 1, \ldots, d\}$, we introduce the space of "cell solutions",

$$\mathcal{M}^{\mathcal{N}}(Y) := \{\hat{\xi}^\eta_{\mathcal{N},T_\delta}; T_\delta \subset \Omega,\ \eta = 1, \ldots, d\}, \tag{4.2.10}$$

where $\hat{\xi}^\eta_{\mathcal{N},T_\delta}(\cdot) : Y \to \mathbb{R}$ are the solutions of (4.1.4) associated with the mapping G_{x_τ}, i.e., with a tensor $a_{x_\tau}(y) = a^\varepsilon(G_{x_\tau}(y))$ and with right-hand side $l_\eta(\cdot)$. The functions $\hat{\xi}^\eta_{\mathcal{N},T_\delta}$ are computed very accurately. The DOF $\mathcal{N}$ of the FE space $S^q(Y,\mathcal{N})$ is thus assumed to be large.

Affine representation of the tensor. A suitable representation of the tensor $a_{x_\tau}(y)$ is crucial for the efficiency of the RB method (more precisely, we look for an affine parametrization). The

simplest case is when $a_{x_\tau}(y)$ is directly available in an affine form

$$a_{x_\tau}(y) = \sum_{q=1}^{Q} \Theta_q(x_\tau) a_q(y), \ \forall y \in Y. \tag{4.2.11}$$

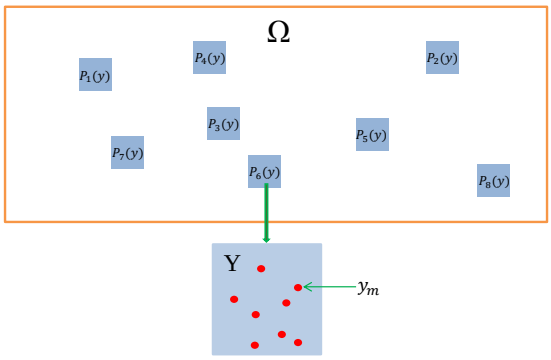


Figure 2: The EIM basis functions $\{P_q(y)\}_{q=1}^{8}$ and the interpolation points $\{y_m\}_{m=1}^{8}$ on the reference sampling domain Y.

If a representation as (4.2.11) is not available, the empirical interpolation method (EIM) [37, 77], can be applied to obtain an affine approximation of $a_{x_\tau}(y)$ in the form

$$a_{x_\tau}^{M}(y) = \sum_{q=1}^{M} \varphi_q(x_\tau) p_q(y). \tag{4.2.12}$$

The idea is to approximate the function $a_{x_\tau}^{M}(y)$ by linear combination of "snapshot" $\{p_{x_{\tau_1}}(y), \ldots, p_{x_{\tau_M}}(y)\}$. For an arbitrary x_τ the linear combination to approximate $a_{x_\tau}^{M}(y)$ will be based on interpolation points $y_1, \ldots, y_M$ in Y. The space of snapshots, called $S_M^{EIM} = span\{p_q(y), \ q = 1, \cdots, M\}$ and the interpolation points $\{y_m\}_{m=1}^{M}$ are computed in an offline stage with the help of a greedy algorithm controlled by available a posteriori error estimates. In the online stage for a given $a_{x_\tau}(y)$ compute (4.2.12) as follows

- evaluate $a_{x_\tau}(y_m)$ at the interpolation points $\{y_m\}_{m=1}^{M}$;
- solve the interpolation problem (a $M \times M$ linear system)

$$\sum_{q=1}^{M} p_q(y_m) \varphi_q(x_\tau) = a_{x_\tau}(y_m), \ m = 1, \ldots, M, \tag{4.2.13}$$

  to find $\varphi_q(x_\tau)_{q=1}^{M}$.

We refer to Section 4.6 for numerical computations with affine and nonaffine multiscale tensors.

A posteriori error estimator. Crucial for the selection of the reduced basis functions is an appropriate a posteriori error estimator. For a given sampling domain $T_\delta \subset \Omega$ let $\xi^i_{\mathcal{N},T_\delta}, \xi^k_{\mathcal{N},T_\delta} \in S^q(Y,\mathcal{N})$ be the solutions of (4.1.4) with right-hand side $l_i(\cdot), l_k(\cdot)$, respectively, as described above. Assume next that $S_l(Y)$ is an l-dimensional linear subspace of $S^q(Y,\mathcal{N})$ and consider $\xi^i_{l,T_\delta}, \xi^k_{l,T_\delta}$ the solution of (4.1.4) in $S_l(Y)$ with right-hand side $l_i(\cdot), l_k(\cdot)$, respectively. Define the following two numerical homogenized tensors

$$(a^0_{\mathcal{N},T_\delta}(x_\tau))_{ik} = \int_Y a_{x_\tau}(y)\left(\nabla\hat{\xi}^i_{\mathcal{N},T_\delta}(y)+\mathbf{e}_i\right)\cdot\left(\nabla\hat{\xi}^k_{\mathcal{N},T_\delta}(y)+\mathbf{e}_k\right)dy, \tag{4.2.14}$$

$$(a^0_{l,T_\delta}(x_\tau))_{ik} = \int_Y a_{x_\tau}(y)\left(\nabla\hat{\xi}^i_{l,T_\delta}(y)+\mathbf{e}_i\right)\cdot\left(\nabla\hat{\xi}^k_{l,T_\delta}(y)+\mathbf{e}_k\right)dy. \tag{4.2.15}$$

We first need the following lemma (see [1, Lemma 3.3]).

Lemma 4.2.1. *Consider the tensors $a^0_{\mathcal{N},T_\delta}(x_\tau), a^0_{l,T_\delta}(x_\tau)$ defined in* (4.2.14), (4.2.15), *respectively. Then*

$$\begin{aligned}&|(a^0_{\mathcal{N},T_\delta}(x_\tau))_{ik}-(a^0_{l,T_\delta}(x_\tau))_{ik}| = \\ &|\int_Y a_{x_\tau}(y)\left(\nabla\hat{\xi}^i_{l,T_\delta}(y)-\nabla\hat{\xi}^i_{\mathcal{N},T_\delta}(y)\right)\cdot\left(\nabla\hat{\xi}^k_{l,T_\delta}(y)-\nabla\hat{\xi}^k_{\mathcal{N},T_\delta}(y))\right)dy|.\end{aligned} \tag{4.2.16}$$

Proof. The proof follows the line of [1, Lemma 3.3]. We sketch it for completeness. As $S_l(Y) \subset S^q(Y,\mathcal{N})$ we have

$$\begin{aligned}&|(a^0_{\mathcal{N},T_\delta}(x_\tau))_{ik}-(a^0_{l,T_\delta}(x_\tau))_{ik}| \\ &= |\int_Y a_{x_\tau}(y)\left(\nabla\hat{\xi}^i_{l,T_\delta}(y)-\nabla\hat{\xi}^i_{\mathcal{N},T_\delta}(y)\right)\cdot\mathbf{e}_k dy| \\ &= |\int_Y a_{x_\tau}(y)\left(\nabla\hat{\xi}^i_{l,T_\delta}(y)-\nabla\hat{\xi}^i_{\mathcal{N},T_\delta}(y)\right)\cdot\left(\mathbf{e}_k+\nabla\hat{\xi}^k_{l,T_\delta}(y)-\nabla\hat{\xi}^k_{\mathcal{N},T_\delta}(y)\right)dy| \\ &= |\int_Y a_{x_\tau}(y)\left(\nabla\hat{\xi}^i_{l,T_\delta}(y)-\nabla\hat{\xi}^i_{\mathcal{N},T_\delta}(y)\right)\cdot\left(-\nabla\hat{\xi}^k_{\mathcal{N},T_\delta}(y)\right)dy|\end{aligned}$$

and the proof follows easily by further adding and subtracting the quantity $\nabla\hat{\xi}^k_{l,T_\delta}(y)$. □

Notice that we have used the symmetry in the above proof. This proof is however valid without symmetry (by using the solution of a dual problem corresponding to (4.1.4), see in [57] or [26, Lemma 4.6],[23]). Next we derive an a posteriori estimator, which allows to control the accuracy of our output of interest (the numerically homogenized tensor). The procedure, which follows standard residual based estimates, is crucial for RB and has been extensively discussed (see [93]

and [45] for a discussion in the homogenization context). Define

$$\hat{e}^i_{l,T_\delta} = \hat{\xi}^i_{l,T_\delta} - \hat{\xi}^i_{\mathcal{N},T_\delta}. \tag{4.2.17}$$

Using (4.1.4) we see that

$$b(\hat{e}^i_{l,T_\delta}, \hat{z}_{\mathcal{N}}) = b(\hat{\xi}^i_{l,T_\delta}, \hat{z}_{\mathcal{N}}) - l_i(\hat{z}_{\mathcal{N}}),\ \forall \hat{z}_{\mathcal{N}} \in S^q(Y,\mathcal{N}). \tag{4.2.18}$$

The right-hand side defines a linear form on $S^q(Y,\mathcal{N})$. Hence by the Riesz theorem, there exists a unique $\bar{e}^i_{l,T_\delta} \in S^q(Y,\mathcal{N})$ such that

$$b(\hat{e}^i_{l,T_\delta}, \hat{z}_{\mathcal{N}}) = (\bar{e}^i_{l,T_\delta}, \hat{z}_{\mathcal{N}})_{\mathcal{W}}, \tag{4.2.19}$$

where $(\cdot,\cdot)_{\mathcal{W}}$, defined as $(v,w)_{\mathcal{W}} = \int_Y \nabla v \cdot \nabla w dy$, denotes the scalar product in the space $\mathcal{W}(Y)$ defined in (1.2.12) or (1.2.13). We notice that $\bar{e}^i_{l,T_\delta}$ can be computed numerically in an efficient way thanks to the affine representation of the tensor $a_{x_\tau}(y)$. This leads to define the residual of the a posteriori error estimator as

$$\Delta^i_{l,T_\delta} := \frac{\|\bar{e}^i_{l,T_\delta}\|_{\mathcal{W}}}{\sqrt{\lambda_{LB}}}. \tag{4.2.20}$$

Here λ_{LB} is an approximation of the coercivity constant λ described in (1.0.1). To compute Δ^i_{l,T_δ}, one needs to solve (4.2.19), which is parameter dependent. Thanks to the affine representation of the tensor, (4.2.19) can be decomposed into several parameter independent FE problems that can be precomputed. Hence, the residual (4.2.20) is cheap to compute which is crucial to get the efficiency of the a posteriori control in the greedy algorithm (see [92, 96] for details). The next lemma gives the bound of the error in quantities of interest (e.g., the numerical homogenized tensors or the cell solutions) in terms of the residual (4.2.20).

Lemma 4.2.2. *Let $\hat{\xi}^i_{l,T_\delta}, \hat{\xi}^i_{\mathcal{N},T_\delta}$ be the solutions of problem* (4.1.4) *in $S_l(Y)$ and $S^q(Y,\mathcal{N})$, respectively, and $\bar{e}^i_{l,T_\delta}, \Delta^i_{l,T_\delta}, a^0_{\mathcal{N},T_\delta}(x_\tau), a^0_{l,T_\delta}(x_\tau)$ be the quantities defined above. Assume that the approximation λ_{LB} of the coercivity constant* (1.0.1) *satisfies $0 < \lambda_{LB} \leq \lambda$. Then we have*

$$\|\hat{\xi}^i_{l,T_\delta} - \hat{\xi}^i_{\mathcal{N},T_\delta}\|_{\mathcal{E},K_j} \leq \Delta^i_{l,T_\delta}, \tag{4.2.21}$$

$$\|\hat{\xi}^i_{l,T_\delta} - \hat{\xi}^i_{\mathcal{N},T_\delta}\|_{\mathcal{W}} \leq \frac{\Delta^i_{l,T_\delta}}{\sqrt{\lambda_{LB}}}, \tag{4.2.22}$$

$$\left(\frac{\lambda_{LB}}{\Lambda}\Delta^i_{l,T_\delta}\right)^2 \leq |(a^0_{\mathcal{N},T_\delta}(x_\tau))_{ii} - (a^0_{l,T_\delta}(x_\tau))_{ii}| \leq (\Delta^i_{l,T_\delta})^2, \tag{4.2.23}$$

where Λ *is the continuity constant defined in* (1.0.1) *and* $\|\cdot\|_{\mathscr{E},T_\delta}$ *is the energy norm defined by*

$$\|v\|_{\mathscr{E},T_\delta} = (b(v,v))^{1/2} := \left(\int_Y a_{x_\tau}(y)\nabla v(y)\cdot\nabla v(y)dy\right)^{1/2}. \tag{4.2.24}$$

Proof. The proof is standard. Plugging $\hat{z}_{\mathscr{N}} = \hat{e}^i_{l,T_\delta}$ in (4.2.19) we get $\|\hat{e}^i_{l,T_\delta}\|^2_{\mathscr{E},K_j} = (\bar{e}^i_{l,T_\delta}, \hat{e}^i_{l,T_\delta})_{\mathscr{W}}$. On one hand, from the coercivity of (4.2.19) and the assumption on λ_{LB} we get

$$\|\hat{e}^i_{l,T_\delta}\|_{\mathscr{W}} \leq \frac{\|\bar{e}^i_{l,T_\delta}\|_{\mathscr{W}}}{\lambda_{LB}},$$

hence (4.2.21). On the other hand, using the inequality $\sqrt{\lambda_{LB}}\|\hat{e}^i_{l,T_\delta}\|_{\mathscr{W}} \leq \|\hat{e}^i_{l,T_\delta}\|_{\mathscr{E},T_\delta}$ easily leads to (4.2.22). The inequality (4.2.23) then follows from (4.2.16), (4.2.21) and the inequality $\|\bar{e}^i_{l,T_\delta}\|_{\mathscr{W}} \leq \Lambda\|\hat{e}^i_{l,T_\delta}\|_{\mathscr{W}}$. □

Remark 4.2.3. *From* (4.2.23)*, we can see that the stability factor* λ_{LB} *plays an important role in the efficiency of the a posteriori estimator. There are two efficient methods proposed in [72, 96, 92] to compute* λ_{LB}*. We simply mention them here (notice that for both methods, the affine representation* (4.2.11) *of the tensor is required). The simplest method is the "*$\min\Theta$*" method, where* $\lambda_{LB}(x_\tau)$ *is estimated by*

$$\lambda_{LB}(x_\tau) = \Big(\min_{q\in\{1,\cdots,Q\}} \frac{\Theta_q(x_\tau)}{\Theta_q(\bar{x}_\tau)}\Big)\lambda(\bar{x}_\tau), \tag{4.2.25}$$

where $\bar{x}_\tau$ *is a randomly chosen point in* Ω*. The "*$\min\Theta$*" method, however, requires that the tensor* (4.2.11) *satisfies the following properties: (i)* $\Theta_q(x) > 0, q = 1,\cdots,Q, \forall x \in \Omega$*; (ii)* $a_q(y)\xi\cdot\xi \geq 0, \forall \xi \in \mathbb{R}^d, y \in Y$*. The above conditions might be restrictive for some applications. A more general but more involved method is the successive constraint method (SCM). This method is based on an offline-online strategy. The SCM offline stage relies on a greedy procedure and is costly, but the online procedure is very efficient. The advantage of this method is that it is a robust and general method which works for all kinds of affine tensors (we refer to [72, 96] for details).*

Offline stage. We select by a greedy algorithm N couple (T_{δ_n}, η_n), where $T_{\delta_n} \subset \Omega$ is a sampling domain and η_n corresponds to the unit vector $\mathbf{e}_{\eta_n}$ belonging to the set canonical basis of $\mathbb{R}^d$, where $\eta_n \in \{1,\ldots,d\}$. Corresponding to the N couple (T_{δ_n}, η_n), we compute $\hat{\xi}^{\eta_n}_{\mathscr{N},T_{\delta_n}}(\cdot)$, the solution of (4.1.4) with a tensor given by $a_{x_{\tau_n}}(y)$ (x_{τ_n} is the barycenter of T_{δ_n}) and a right-hand side given by $l_{\eta_n}(\cdot)$. The following greedy algorithm to determine successively $\{(T_{\delta_n}, \eta_n, \hat{\xi}^{\eta_n}_{\mathscr{N},T_{\delta_n}}),\ n = 1,\ldots,N\}$ is based on the usual procedure of the RB methodology (see [93, 96]).

Algorithm 4.2.4 (Greedy procedure)**.** *Denote by* $\|\cdot\|_{\mathscr{W}}$ *the norm associated to the space* $\mathscr{W}(Y)$ *(defined by* (1.2.13) *or* (1.2.12)*). Given two parameters,* N_{RB} *the maximum basis number, and* tol_{RB} *a stopping tolerance:*

1. *Choose randomly (by a Monte Carlo method)* N_{train} *sampling domains* T_{δ_n} *in such a way that*

$T_{\delta_n} \subset \Omega$. *Define the "training set"* $\Xi_{RB} = \{(T_{\delta_n}, \eta_n); 1 \leq \eta_n \leq d, 1 \leq n \leq N_{train}\}$ [1].

2. *Select randomly* $(T_{\delta_1}, \eta_1) \in \Xi_{RB}$ *and compute* $\hat{\xi}^{\eta_1}_{\mathcal{N},T_{\delta_1}}$, *the solution of* (4.1.4) *with right-hand side* $l_{\eta_1}(\cdot)$ *in* $S^q(Y,\mathcal{N})$, *corresponding to the selected parameters* (T_{δ_1}, η_1). *Set* $l = 1$ *and define* $\hat{\xi}_{1,\mathcal{N}}(y) = \frac{\hat{\xi}^{\eta_1}_{\mathcal{N},T_{\delta_1}}}{\|\hat{\xi}^{\eta_1}_{\mathcal{N},T_{\delta_1}}\|_{\mathcal{W}}}$, *and the corresponding RB space* $S_1(Y) = span\{\hat{\xi}_{1,\mathcal{N}}\}$.

3. *For* $l = 2, \ldots, N_{RB}$

   a. *Compute for each* $(T_\delta, \eta) \in \Xi_{RB}$ *the solution* $\hat{\xi}^{\eta}_{l-1,T_\delta}$ *of* (4.1.4) *with right-hand side* $l_\eta(\cdot)$ *in* $S_{l-1}(Y)$ *and the residual* $\Delta^{\eta}_{l-1,T_\delta}$ *defined in* (4.2.20).

   b. *Select the next reduced basis by choosing*

   $$(T_{\delta_l}, \eta_l) = argmax_{(T_\delta,\eta)\in\Xi_{RB}} \Delta^{\eta}_{l-1,T_\delta},$$

   provided that $\max_{(T_\delta,\eta)\in\Xi_{RB}} (\Delta^{\eta}_{l-1,T_\delta})^2 > tol_{RB}$ [2], *otherwise the algorithm ends.*

   c. *Compute* $\hat{\xi}^{\eta_l}_{\mathcal{N},T_{\delta_l}}$ *the solution of* (4.1.4) *in* $S^q(Y,\mathcal{N})$ *corresponding to the selected parameters* (T_{δ_l}, η_l). *Set* $\hat{\xi}_{l,\mathcal{N}}(y) = \frac{R_l(y)}{\|R_l(y)\|_{\mathcal{W}}}$ *the* $l-th$ *RB basis function, where*

   $$R_l(y) = \hat{\xi}^{\eta_l}_{\mathcal{N},T_{\delta_l}}(y) - \sum_{m=1}^{l-1} (\hat{\xi}^{\eta_l}_{\mathcal{N},T_{\delta_l}}, \hat{\xi}_{m,\mathcal{N}})\hat{\xi}_{m,\mathcal{N}}$$

   Define the RB space $S_l(Y) = span\{\hat{\xi}_{1,\mathcal{N}}, \ldots, \hat{\xi}_{l,\mathcal{N}}\}$. *Set* $l = l+1$ *and go back to a.*

We emphasize that while constructing $S_l(Y)$, with $\hat{\xi}_{l,\mathcal{N}}(y)$ being a linear combination of the solutions of (4.1.4), our output of interest is (4.1.8) that can be computed using the RB (see Lemma 4.2.1). From Lemma 4.2.2, we know that the the square of the residual $\Delta^{\eta}_{l,T_\delta}$ gives an a posteriori error estimate for the output of interest, hence, $(\Delta^{\eta}_{l,T_\delta})^2$ is the quantity that needs to be controlled in the above algorithm. We note that even though (4.1.4) has to be solved for each parameter in Ξ_{RB} in the step 3.a., this procedure is moderately expensive as (4.1.4) is solved in the RB space $S_{l-1}(Y)$ of small dimension $l-1 \leq N$. A similar remark holds for the residuals $\Delta^{\eta}_{l,T_\delta}$ that need to be computed for each parameter in Ξ_{RB}, but only rely for these computations on precomputed quantities (computed once for the whole offline procedure) and small linear problems involving the current RB space $S_{l-1}(Y)$.

Output of the offline procedure. The output of the above procedure is the RB space

$$S_N(Y) = \text{span}\{\hat{\xi}_{n,\mathcal{N}}(y),\ n = 1, .., N\}. \tag{4.2.26}$$

[1] N_{train} should be large enough to ensure that the results of the greedy algorithm are stable with respect to other choices of training sets.

[2] Notice that the error of the outputs of interest scale like the square of the error of the cell functions as can be seen in Lemma 4.2.1 and 4.2.2.

Rather than storing the reduced basis functions, using the affine representation (4.2.11) (or (4.2.12)) described above, the output consists of the following matrices and vectors

$$(A_q)_{nm} := \int_Y a_q(y)\nabla\hat{\xi}_{n,\mathcal{N}}(y)\cdot\nabla\hat{\xi}_{m,\mathcal{N}}dy, \quad (F_q^i)_m := \int_Y a_q(y)\mathbf{e}_i\cdot\nabla\hat{\xi}_{m,\mathcal{N}}(y)dy. \tag{4.2.27}$$

4.3 Online procedure and the RB-FE-HMM

We define a macro method similar to the FE-HMM, with micro functions computed in the RB space. The method reads: find $u^{H,RB}\in S_0^\ell(\Omega,\mathcal{T}_H)$ such that

$$B_{H,RB}(u^{H,RB},v^H)=\int_\Omega fv^H dx,\ \forall v^H\in S_0^\ell(\Omega,\mathcal{T}_H), \tag{4.3.28}$$

with a bilinear form given by

$$B_{H,RB}(v^H,w^H) := \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\frac{\omega_{K_j}}{|K_{\delta_j}|}\int_{K_{\delta_j}} a^\varepsilon(x)\nabla v_{N,K_j}(x)\cdot\nabla w_{N,K_j}(x)dx, \tag{4.3.29}$$

where $v_{N,K_j}(x)$ (respectively $w_{N,K_j}(x)$) is such that $v_{N,K_j}-v^H_{lin,j}(x)\in S_N(K_{\delta_j})$ and

$$\int_{K_{\delta_j}} a^\varepsilon(x)\nabla v_{N,K_j}(x)\cdot\nabla z_N(x)dx=0, \quad \forall z_N\in S_N(K_{\delta_j}). \tag{4.3.30}$$

The space $S_N(K_{\delta_j})$ is defined through the mapping $G_{x_{K_j}}: Y\to K_{\delta_j}$ as

$$S_N(K_{\delta_j}) = \text{span}\{\delta\hat{\xi}_{n,\mathcal{N}}(G^{-1}_{x_{K_j}}(x)) =: \xi_{n,K_j}(x),\ n=1,..,N\}. \tag{4.3.31}$$

The well-posedness of problem (4.3.28) is proved in the following lemma.

Lemma 4.3.1. *Assume that* (1.0.1) *and that* (**Q1**) *hold. Then problem* (4.3.28) *has a unique solution.*

Proof. Similarly as in [1],[3, Section 3.3.1] we can show that

$$\|\nabla v^H_{lin}\|_{\mathscr{L}^2(K_\delta)} \le \|\nabla v_{N,K_j}\|_{\mathscr{L}^2(K_\delta)} \le \sqrt{\frac{\Lambda}{\lambda}}\|\nabla v^H_{lin}\|_{\mathscr{L}^2(K_\delta)}. \tag{4.3.32}$$

Using (1.0.1) and (**Q1**) we then obtain

$$C_1\|\nabla v^H\|^2_{\mathscr{L}^2(\Omega)} \le B_{H,RB}(v^H,v^H), \quad B_{H,RB}(v^H,w^H)\le C_2\|\nabla v^H\|_{\mathscr{L}^2(\Omega)}\|\nabla w^H\|_{\mathscr{L}^2(\Omega)}.$$

The Poincaré inequality and the Lax-Milgram theorem give the stated result (see again [1],[3,

Section 3.3.1] for details). □

Fast solution of micro-problems. Owing to the affine form (4.2.11) of the tensor a^ε, the problem (4.3.30) amounts to solving an $N \times N$ linear system (recall N is small). Indeed, we observe that by writing $v_{N,K_j} - v^H_{lin,j}(x) = \sum_{n=1}^N \alpha_n \xi_{n,K_j}(x)$ (4.3.30) reads

$$\sum_{n=1}^N \alpha_n \int_{K_{\delta_j}} a^\varepsilon(x)\nabla\xi_{n,K_j}(x)\cdot\nabla\xi_{m,K_j}(x)dx = -\sum_{i=1}^d \int_{K_{\delta_j}} a^\varepsilon(x)\mathbf{e}_i\cdot\nabla\xi_{m,K_j}(x)dx\frac{\partial v^H_{lin,j}}{\partial x_i}, \quad (4.3.33)$$

for all $m = 1,\dots N$. Next, again thanks to the affine representation of the tensor (here we are assuming the representation (4.2.11) for simplicity), (4.3.33) can be written as

$$\begin{aligned}&\sum_{n=1}^N \alpha_n \sum_{q=1}^Q \Theta_q(x_{K_j}) \int_Y a_q(y)\nabla\hat{\xi}_{n,\mathcal{N}}(y)\cdot\nabla\hat{\xi}_{m,\mathcal{N}}(y)dy\\ &= -\sum_{i=1}^d \sum_{q=1}^Q \Theta_q(x_{K_j}) \int_Y a_q(y)\mathbf{e}_i\cdot\nabla\hat{\xi}_{m,\mathcal{N}}(y)dy\frac{\partial v^H_{lin,j}}{\partial x_i},\end{aligned} \quad (4.3.34)$$

or equivalently

$$\left(\sum_{q=1}^Q \Theta_q(x_{K_j})A_q\right)\boldsymbol{\alpha} = -\sum_{i=1}^d\left(\sum_{q=1}^Q \Theta_q(x_{K_j})F^i_q\right)\frac{\partial v^H_{lin,j}}{\partial x_i}, \quad (4.3.35)$$

where the $N\times N$ matrices $A_q,\ q=1,\dots,Q$ and the vectors $F^i_q \in \mathbb{R}^N,\ q=1,\dots,Q,\ i=1,\dots,d$ are defined by (4.2.27).

We emphasize that the matrices A_q and the vectors F^i_q are assembled and stored in the offline stage, thus (4.3.35) amounts just in building the linear combination by evaluating $\Theta_q(\cdot)$ at the desired integration points x_{K_j} (or computing the interpolation problem (4.2.12) when we rely on the approximation (4.2.12) for the tensor $a_{x_\tau}(y)$) and solving the $N\times N$ system (4.3.35) for each micro function at the quadrature points needed to assemble (1.2.15).

Reformulation of the RB-FE-HMM. Similar to the reformulation (4.1.7) for the FE-HMM, we have

$$B_{H,RB}(v^H, w^H) := \sum_{K\in\mathcal{T}_H}\sum_{j=1}^J \omega_{K_j} a^0_N(x_{K_j})\nabla v^H_{lin,j}(x_{K_j})\cdot\nabla w^H_{lin,j}(x_{K_j}), \quad (4.3.36)$$

where

$$(a^0_N(x_{K_j}))_{ik} = \int_Y a_{x_{K_j}}(y)\left(\nabla\hat{\chi}^i_{N,K_j}(y)+\mathbf{e}_i\right)\cdot\left(\nabla\hat{\chi}^k_{N,K_j}(y)+\mathbf{e}_k\right)dy. \quad (4.3.37)$$

which is easily seen by noting that $v_{N,K_j}(x)$, the solution of (4.3.30), can be written as

$$v_{N,K_j}(x) = v^H_{lin,j}(x) + \delta \sum_{i=1}^{d} \hat{\chi}^i_{N,K_j}(G^{-1}_{x_{K_j}}(x)) \frac{\partial v^H_{lin,j}}{\partial x_i}, \tag{4.3.38}$$

where $\hat{\chi}^i_{N,K_j}(y)$ is the solution of (4.1.4) in the RB space (4.2.26).

4.4 Reconstruction of the micro solution

We briefly explain a procedure to obtain an approximation to the fine scale solution u^ε of problem (4.0.1). While an error estimates $\|u^\varepsilon - u^{H,RB}\|$ can be obtained in the $\mathscr{L}^2$ norm for locally periodic tensor (see Section 4.5), one cannot expect convergence between ∇u^ε and $\nabla u^{H,RB}$, as $u^{H,RB}$ does not capture the fine scale oscillations of the solution (we do not have strong convergence of ∇u^ε towards ∇u^0 in general). Inspired by the correctors in homogenization theory [74], a numerical corrector for the FE-HMM has been introduced in [1, 60]. A numerical corrector, computed in a post-processing step can also be defined for the RB-FE-HMM. Here we present two approaches for such reconstruction.

For the first approach, which follows the procedure for the FE-HMM, we assume piecewise linear macro solver. For any $K \in \mathscr{T}_H$, we consider the function $u_{N,K} - u^{H,RB}$ known in the sampling domain K_ε (see (4.3.30)). We then consider its periodic extension in K denoted by $(u_{N,K} - u^{H,RB})|^P_K$ and define a corrector in every macro element K as

$$u_p(x)|_K = u^{H,RB} + (u_{N,K} - u^{H,RB})|^P_K. \tag{4.4.39}$$

An error estimate for this procedure is available for the FE-HMM in [1, Thm. 3.11] Also simple, this procedure requires to store the micro solutions in the sampling domains, and the periodic extension might be cumbersome, specially for three dimensional problems with simplicial elements. However, thanks to the precomputed RB space, the computational cost to solve a micro cell problem in the present framework is largely reduced. This allows to consider a second approach for the construction of numerical correctors. For a sampling domain of size ε, given $x \in K \in \mathscr{T}_H$, we can evaluate the reconstructed solution at this particular point by using

$$u^\varepsilon_{p,RB}(x)|_K = u^{H,RB}(x) + \varepsilon \sum_{i=1}^{d} \hat{\chi}^i_{N,T_\varepsilon}(G^{-1}_x(x)) \frac{\partial u^{H,RB}_{lin}(x)}{\partial x_i}, \tag{4.4.40}$$

where $T_\varepsilon = x + (-\varepsilon/2, \varepsilon/2)^d$ and $\hat{\chi}^i_{N,T_\varepsilon}(G^{-1}_x(x))$ can be computed by solving (4.1.4) in the reduced basis space $S_N(T_\varepsilon)$. We note that the second reconstruction procedure (4.4.40) allows to use higher order macro FEMs, whereas for the first procedure (4.4.39), it would require an interpolation procedure.

4.5 A priori error analysis

In this section, we derive an a priori error estimate for the RB-FE-HMM. While a component of the error (relying on the approximation property of the greedy algorithm) relies on assumptions difficult to check in practice, by providing the analysis proposed here, describing the various contributions to the global error is nevertheless of interest. Following [3], an error estimate similar to (1.2.17) can be derived for the RB-FE-HMM.

Theorem 4.5.1. *Let u^0, $u^{H,RB}$ be the solutions of* (1.1.3) *and* (4.3.28)*, respectively, and that assume that $u^0 \in \mathscr{H}^{\ell+1}(\Omega)$. Assume further that (***Q1***), (***Q2***) and* (1.0.1) *hold and that the tensor $a^0(x)$ appearing in* (1.1.3) *is sufficiently regular. Then*

$$\|u^0 - u^{H,RB}\|_{\mathscr{H}^1(\Omega)} \leq C(H^\ell + r_{HMM}), \tag{4.5.41}$$

$$\|u^0 - u^{H,RB}\|_{\mathscr{L}^2(\Omega)} \leq C(H^{\ell+1} + r_{HMM}), \tag{4.5.42}$$

where

$$r_{HMM} = \sup_{K\in\mathscr{T}_H} \sup_{x_{K_j}\in K} \|a^0(x_{K_j}) - a^0_N(x_{K_j})\|_{\mathscr{F}}, \tag{4.5.43}$$

and where the tensor a^0_N is defined in (4.3.37) *and $a^0(x_{K_j})$ is the tensor of the homogenized problem* (1.1.3) *evaluated at the quadrature point x_{K_j}. The constant C is independent of H, N, $\mathscr{N}$ or ε.*

Proof. Decompose the error into ($\|\cdot\|$ stands for the $\mathscr{L}^2$ or $\mathscr{H}^1$ norm)

$$\|u^0 - u^{H,RB}\| \leq \|u^0 - u^{0,H}\| + \|u^{0,H} - u^{H,RB}\|,$$

where $u^{0,H}$ is the solution of (1.2.10).

Following [54], we obtain $\|u^0 - u^{0,H}\|_{\mathscr{H}^1(\Omega)} \leq CH^\ell$ and $\|u^0 - u^{0,H}\|_{\mathscr{L}^2(\Omega)} \leq CH^{l+1}$ for sufficiently regular tensor a^0. We then have

$$\|u^{0,H} - u^{H,RB}\|_{\mathscr{H}^1(\Omega)} \leq C \sup_{w^H\in S^\ell_0(\Omega,\mathscr{T}_H)} \frac{|B_{0,H}(u^{0,H}, w^H) - B_{H,RB}(u^{H,RB}, w^H)|}{\|w^H\|_{\mathscr{H}^1(\Omega)}}.$$

Using the expressions (1.2.11) and (4.3.36) for $B_{0,H}$ and $B_{H,RB}$, respectively, together with (**Q2**), we can bound the right-hand side of the above inequality by $C\sup_{K\in\mathscr{T}_H}\sup_{x_{K_j}\in K}\|a^0(x_{K_j}) - a^0_N(x_{K_j})\|_{\mathscr{F}}\|u^{0,H}\|_{\mathscr{H}^1(\Omega)}$. Using the a priori bound $\|u^{0,H}\|_{\mathscr{H}^1(\Omega)} \leq C\|f\|_{\mathscr{H}^{-1}(\Omega)}$ completes the proof. □

We further decompose

$$r_{HMM} \leq r_{MOD} + r_{MIC} + r_{RB},$$

with

$$\begin{aligned} r_{MOD} &:= \sup_{K\in\mathcal{T}_H}\sup_{x_{K_j}\in K} \|a^0(x_{K_j}) - \bar{a}^0(x_{K_j})\|_{\mathcal{F}}, \\ r_{MIC} &:= \sup_{K\in\mathcal{T}_H}\sup_{x_{K_j}\in K} \|\bar{a}^0(x_{K_j}) - a^0_{\mathcal{N}}(x_{K_j})\|_{\mathcal{F}}, \\ r_{RB} &:= \sup_{K\in\mathcal{T}_H}\sup_{x_{K_j}\in K} \|a^0_{\mathcal{N}}(x_{K_j}) - a^0_N(x_{K_j})\|_{\mathcal{F}}, \end{aligned}$$

where $a^0_{\mathcal{N}}(x_{K_j})$ is defined in (4.1.8) and $\bar{a}^0(x_{K_j})$ is the tensor appearing in (4.1.9). Error bounds for the micro error r_{MIC} were first presented in [1] for linear elliptic problems and generalized to high order in [3] (see also [4, Lemma 6]). Error bounds for the modeling error were first presented in [60] (see also [22] for a situation where r_{MOD} vanishes). The aforementioned error estimates can directly be used for the RB-FE-HMM. It remains to estimate r_{RB}. Consider the space $\mathcal{M}^{\mathcal{N}}(Y)$ as defined in (4.2.10). We want to quantify how well $\mathcal{M}^{\mathcal{N}}(Y)$ can be approximated by the linear space $S_N(Y)$ of dimension N. Such a quantification relies on the notion Kolmogorov N-width.

Definition 4.5.2. *Let F be a subset of $\mathcal{W}(Y)$. We denote the distance of F to any generic $N-$dimensional subspace $W_N(Y)$ of W by*

$$E(F;W_N) = \sup_{x\in F}\inf_{y\in W_N}\|x-y\|_{\mathcal{W}}.$$

The minimal error $E(F;W_N(Y))$ is given by the Kolmogorov N-width of F in W

$$d_N(F,\mathcal{W}(Y)) \quad = \quad \inf\{E(F;W_N(Y)) : W_N(Y) \text{ a } N-\text{dimensional subspace of } \mathcal{W}(Y)\}.$$

It is difficult in general to quantify the Kolmogorov N-width of a given subset of $\mathcal{W}(Y)$. Invoking regularity of the set $\mathcal{M}^{\mathcal{N}}(Y)$ with respect to the parameters one expects usually a fast (e.g. exponential) decay of d_N. Assuming such a decay, it is not obvious that the particular $N-dimensional$ subspace of $\mathcal{W}(Y)$ constructed with the greedy algorithm enjoys such an approximation property. This has been proved in [43, 48]. More precisely the application of [48, Corollary 3] shows the following result. Assume that the parametrized cell solution space $\mathcal{M}^{\mathcal{N}}$ has an exponentially small Kolmogorov N-width $d_N(\mathcal{M}^{\mathcal{N}},W) \le ce^{-rN}$, with r satisfying

$$r > \log(1+(\Lambda/\lambda_{LB})\sqrt{\Lambda/\lambda}), \tag{4.5.44}$$

where λ,Λ are the coercivity and continuity bounds (1.0.1) and λ_{LB} is the approximation of the coercivity constant used in the greedy algorithm (see Section 3.2). Then the reduced basis

method converges exponentially in the sense that there exists a constant $s > 0$ such that

$$\|\hat{\chi}^k_{N,K_j}(y) - \hat{\chi}^k_{\mathcal{N},K_j}(y)\|_{\mathcal{W}} \leq Ce^{-sN}, \tag{4.5.45}$$

for all $K \in \mathcal{T}_H$ and all $x_{K_j} \in K$.

Theorem 4.5.3. *In addition to the assumption of the Theorem 4.5.1, assume that the parametrized cell solution space $\mathcal{M}^{\mathcal{N}}$ has an exponentially small Kolmogorov N-width $d_N(\mathcal{M}^{\mathcal{N}}, W) \leq ce^{-rN}$, where r satisfies* (4.5.44)*. Then,*

$$\|u^0 - u^{H,RB}\|_{\mathcal{H}^1(\Omega)} \leq C(H^\ell + e^{-2sN} + r_{MIC} + r_{MOD}), \tag{4.5.46}$$
$$\|u^0 - u^{H,RB}\|_{\mathcal{L}^2(\Omega)} \leq C(H^{\ell+1} + e^{-2sN} + r_{MIC} + r_{MOD}). \tag{4.5.47}$$

If in addition χ^i, $i = 1,\ldots,d$, the solutions (4.1.3) *in $\mathcal{W}(Y)$ (see* (1.2.13) (1.2.12)*) satisfy*

$$|\chi^i|_{\mathcal{H}^{q+1}(K_{\delta_j})} \leq C\varepsilon^{-q}\sqrt{|K_{\delta_j}|}, \tag{4.5.48}$$

with C independent of ε, the quadrature point x_{K_j}, then

$$\|u^0 - u^{H,RB}\|_{\mathcal{H}^1(\Omega)} \leq C(H^\ell + e^{-2sN} + \left(\frac{h}{\varepsilon}\right)^{2q} + r_{MOD}), \tag{4.5.49}$$
$$\|u^0 - u^{H,RB}\|_{\mathcal{L}^2(\Omega)} \leq C(H^{\ell+1} + e^{-2sN} + \left(\frac{h}{\varepsilon}\right)^{2q} + r_{MOD}), \tag{4.5.50}$$

where $\hat{h} = \frac{h}{\varepsilon}$ is the microscopic meshsize for the micro problems of the offline stage, i.e., $\hat{h} = \mathcal{O}(\mathcal{N}^{-1/d})$.

If in addition $a^\varepsilon(x) = a^\varepsilon(x, x/\varepsilon) = a(x,y)$ is Y-periodic in y, and $a_{ij}(x,y) \in \mathcal{C}\left(\bar{\Omega}; \mathcal{W}^{1,\infty}_{per}(Y)\right)$ for all $i,j = 1,\ldots,d$, then

$$r_{MOD} = 0 \qquad \text{if } \mathcal{W}(K_{\delta_j}) = \mathcal{W}^1_{per}(K_{\delta_j}) \text{ and } \delta/\varepsilon \in \mathbb{N}, \tag{4.5.51}$$
$$r_{MOD} \leq C(\delta + \frac{\varepsilon}{\delta}) \qquad \text{if } \mathcal{W}(K_{\delta_j}) = \mathcal{H}^1_0(K_{\delta_j}) \ (\delta > \varepsilon). \tag{4.5.52}$$

Proof. The estimates (4.5.46),(4.5.47) follow from Theorem 4.5.1, (4.5.45) and Lemma 4.2.1. The estimates (4.5.49),(4.5.50) follow from Theorem 4.5.1, the estimates for the fully discrete error in [1] [5, Corollary 10] (see also [3, Lemma 10] and [4]). The estimate (4.5.52) has been proved in [60] and the estimate (4.5.51) in [22]. □

We notice that using the estimate $\|u^\varepsilon - u^0\|_{\mathcal{L}^2(\Omega)} \leq C\varepsilon$ valid for locally periodic tensor we can

obtain an error estimate

$$\|u^{\varepsilon} - u^{H,RB}\|_{\mathscr{L}^2(\Omega)} \leq C(H^{\ell+1} + e^{-2sN} + \left(\frac{h}{\varepsilon}\right)^{2q} + \varepsilon + r_{MOD}),$$

measuring the approximation of $u^{H,RB}$ to the fine scale solution in the $\mathscr{L}^2$ norm. In Theorem 4.3, various aspects of the convergence behavior of the RB-FE-HMM are described. The overall goal of the RB-FE-HMM is to obtain an approximation of the effective solution u^0. If the effective tensor would be available, then, rates of convergence such as H^{ℓ} or $H^{\ell+1}$ for the $\mathscr{H}^1$ or $\mathscr{L}^2$ norm could be obtained, provided adequate regularity of the effective solution u^0. For effective solution with singularities, appropriate adaptive mesh refinement techniques could be used. In both case, the rate of convergence is described by classical FE analysis. To obtain similar convergence rates in terms of the macroscopic meshsize with the RB-FE-HMM, we need to control the errors arising from the multiscale methodology and the use of micro solvers to recover the unknown effective data. These various errors are listed below

1. The microscopic error $\left(\frac{h}{\varepsilon}\right)^{2q}$, where $\hat{h} = \frac{h}{\varepsilon} = \mathscr{O}(\mathscr{N}^{-1/d})$ is the microscopic meshsize and q the order of the FEM used for the micro problems in the offline stage. In the RB framework, we use a very accurate meshsize for this offline problems, i.e., micro solvers with large value $\mathscr{N}\,of\,DOF$. This effort is compensated by the fact that we solve micro problems only for a *fixed* (usually *small number*) of sampling domains distributed in the macroscopic physical computational, compared to the classical FE-HMM, where micro problems with DOF proportional to the macro DOF have to be solved in *every* macro elements (notice that in this case, not only the microscopic DOF but also the number of sampling domains increase while refining the macro mesh);

2. The a priori RB error e^{-2sN} , which quantifies how well the infinite dimensional manifold of solutions of micro problems can be approximated by a low dimensional linear subspace, explaining why a small number of micro problems usually suffice in the offline stage. In applications, however, the RB a priori estimate is not used (and usually not known), but this a priori error can be bounded (on both sides) by the RB a posteriori estimate which is computed during the RB procedure;

3. The modeling error, which indicates the influence of the micro solutions on the sampling domain sizes and and micro boundary conditions. In general, this is a delicate question already for the FE-HMM (without RB) and still a subject of investigation (see the discussion and references in the reviews [3, 4]). In particular cases, for example for locally periodic tensors with period ε, if we set the cell domain size ε and choose the center of the cells at the quadrature points, then modeling error vanishes, i.e., $r_{MOD} = 0$.

4.6 Numerical examples

In this section we apply the RB-FE-HMM to four test problems. The first three examples are 2D problems with an affine tensor, a discontinuous affine tensor and a nonaffine tensor, respectively. The fourth example is a 3D problem, representing heat transfer in a microchip first described in [19] for the FE-HMM. In the offline stage, the micro functions are computed in the reference

domain Y using a uniform mesh. We also use a uniform macroscopic mesh for the online procedure to compute the RB-FE-HMM solution. Notice that nonuniform meshes could be similarly used. Higher order implementation for 2D is inspired by [39] and for the 3D problem, we use the software CUBIT version 11.1 [98] to generate the macroscopic tetrahedral mesh for the discretization of the considered domain.

Numerical evaluation of the errors. Let u^H be the numerical solution and u^{ref} be a reference solution (for the problem (1.1.3)) computed on a fine triangulation $\mathcal{T}_h$. The error $u^{ref} - u^H$ in the $\mathscr{H}^1$ and $\mathscr{L}^2$ norms are estimated by

$$\begin{aligned} e_{\mathscr{L}^2} &:= \|u^{ref}\|^{-1}_{\mathscr{L}^2(\Omega)} \Big(\sum_{K\in\mathcal{T}_h} \sum_{j=1}^{J} \rho_{K_j} |u^H(z_{K_j}) - u^{ref}(z_{K_j})|^2 \Big)^{1/2}, \\ e_{H^1} &:= \|u^{ref}\|^{-1}_{\mathscr{H}^1(\Omega)} \Big(\sum_{K\in\mathcal{T}_h} \sum_{j=1}^{J} \rho_{K_j} |\nabla u^H(z_{K_j}) - \nabla u^{ref}(z_{K_j})|^2 \Big)^{1/2}, \end{aligned}$$

where we will use $\|u^{ref}\|_{\mathscr{H}^1(\Omega)} \sim (\sum_{K\in\mathcal{T}_H} \|\nabla u^{ref}\|^2_{\mathscr{L}^2(K)})^{1/2}$. Here $\{z_{K_j}, \rho_{K_j}\}$ are quadrature points on the fine triangulation $\mathcal{T}_h$ chosen such that the quadrature formula is exact for the degree of the piecewise polynomials used to compute u^{ref}.

Stability factor computation for the a posteriori error estimates. As explained in Remark (4.2.3), an estimation of the stability factor λ_{LB} is crucial to control the accuracy of the outputs of the greedy algorithm. In the numerical experiments below, we will use the "$\min\Theta$" method when it can be applied (see Remark (4.2.3)), namely for the 2D and 3D affine examples, and the SCM otherwise (for the 2D discontinuous and nonaffine examples).

4.6.1 2-D problems.

Let $\Omega = [0,1]^2$. We consider the following problem

$$\begin{aligned} -\nabla\cdot(a^\varepsilon(x)\nabla u^\varepsilon(x)) &= 1, \quad \text{in } \Omega, \\ u^\varepsilon(x) &= 0, \quad \text{on } \partial\Omega_D, \\ n\cdot(a^\varepsilon(x)\nabla u^\varepsilon(x)) &= 0, \quad \text{on } \partial\Omega_N, \end{aligned} \tag{4.6.53}$$

where for $x = (x_1, x_2)$, $\partial\Omega_D = \{x_1 = 0\} \cup \{x_1 = 1\}$ and $\partial\Omega_N = \{x_2 = 0\} \cup \{x_2 = 1\}$. We will choose various oscillating tensors a^ε for the above problem. The tensors are chosen so that the homogenized tensors can be easily computed to be able to perform careful numerical test on the behavior of the experimental convergence of the RB-FE-HMM. We emphasize that our numerical method can be applied to more general problems, when an explicit form of the homogenized tensor is not available (e.g., non-periodic or random tensors).

2D affine multiscale tensor. We consider

$$a\Big(x,\frac{x}{\varepsilon}\Big)=\begin{pmatrix} x_1^2+0.2+(x_2+1)(\sin(2\pi\frac{x_1}{\varepsilon})+2) & 0 \\ 0 & x_2^2+0.05+(x_1x_2+1)(\sin(2\pi\frac{x_2}{\varepsilon})+2)\end{pmatrix},$$

with a corresponding diagonal homogenized tensor given by

$$\begin{aligned} a^0_{11}(x)&=\int_0^1 \frac{1}{x_1^2+0.2+(x_2+1)(\sin(2\pi y_1)+2)}dy_1)^{-1}\\ a^0_{22}(x)&=(\int_0^1 \frac{1}{x_2^2+0.05+(x_1x_2+1)(\sin(2\pi y_2)+2)}dy_2)^{-1} \end{aligned} \tag{4.6.54}$$

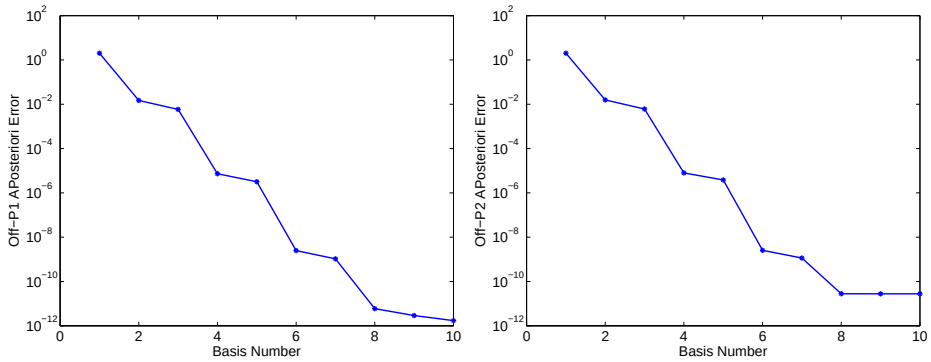
In the offline stage, we use the P1-FEM and the P2-FEM respectively, to compute the RB for the micro problems, that is, we take the FE space $S^q(Y,\mathcal{N})$ with $q=1,2$, with a large number of DOF (as usual in the RB methodology). Periodic coupling is used (i.e., $S^q(Y,\mathcal{N})$ is chosen to be a subspace of (1.2.12)). We also choose sampling domains that match the length of the period of a^ε. To make a fair comparison, we take the same initial sample set Ξ_{RB} for the computations with the P1 and the P2 FEM. In both cases, the tolerance, set to tol_{RB} = 5e-11, is reached by the a

Table 4.1: Parameters for the offline stage (affine tensor).

Offline stage	P1-FEM	P2-FEM
Meshsize for the micro reference domain Y	1500×1500	1200×1200
Initial sample points number $\Xi_{RB}\subset\Omega$	800	800
Tolerance for the offline stage tol_{RB}	5e-11	5e-11
Stability factor method	min Θ	min Θ
RB number	10	10
Final a posteriori error	1.7059e-12	2.7682e-11
Offline CPU time(s)	5330	42724

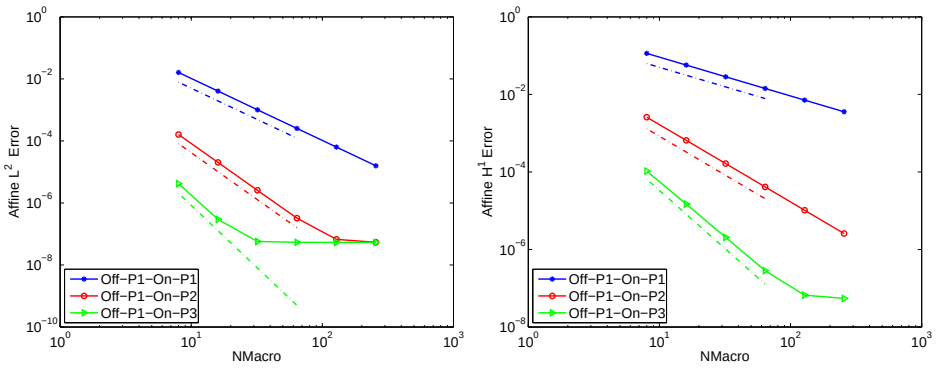
posteriori estimator after the selection of 10 reduced basis. The CPU time in second is reported for our MATLAB computations. We see that the P2 FEM is approximately 8 times more expensive than the P1 FEM for the offline stage. In Fig. 3, we report the decay of the a posteriori error $(\Delta^\eta_{l,T_\delta})^2$ that is fast, as expected.

With these precomputed RB spaces of micro solutions (obtained with P1 and P2 FEM), we now perform online computation and compute the macro solution $u^{H,RB}$ with the RB-FE-HMM. We use P1, P2 and P3 macro FEMs to compute $u^{H,RB}$. The macro meshsize is chosen as 2^{-n}, $n=3,\dots,8$. A reference solution u^{ref} is computed by solving (1.1.3) with a tensor given by (4.6.54) using a FEM with a 1024×1024 mesh and piecewise polynomials of total degree 3. The behavior of the error of the RB-FE-HMM is shown in Fig. 4. In view of the estimates of Theorem 4.5.3, we can make the following observations. First, we notice that $r_{MOD}=0$ as we have periodic boundary conditions for the RB-FE-HMM and sampling domains of size ε. According to the a priori estimates (4.5.50), (4.5.49), the micro error r_{MIC} should be of the order of $10^{-7},10^{-13}$ for the P1, P2 offline computation, respectively, while the macro error should be of the order

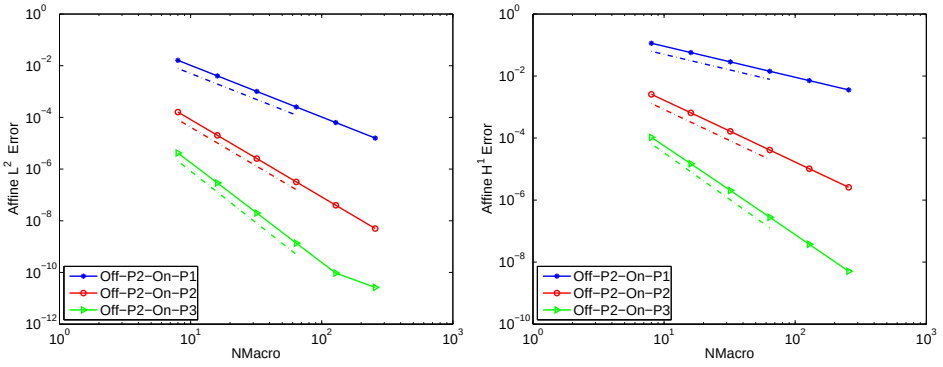


(a) P1-Offline A posteriori Error (b) P2-Offline A posteriori Error

Figure 3: Affine tensor a posteriori error $\max_{(T_\delta,\eta)\in\Xi_{RB}}(\Delta^{\eta}_{l,T_\delta})^2$.



(a) Offline P1-FEM $\mathscr{L}^2$ Error (b) Offline P1-FEM $\mathscr{H}^1$ Error



(c) Offline P2-FEM $\mathscr{L}^2$ Error (d) Offline P2-FEM $\mathscr{H}^1$ Error

Figure 4: Affine tensor example $\|u^{H,RB}-u^{ref}\|$. The offline and RB space parameters are in Table 4.1. Online solver is P1-FEM, P2-FEM, P3-FEM respectively. The dashed lines are the reference lines with slopes 2,3,4 for (a)-(c) and slopes 1,2,3 for (b)-(d).

$2^{-n\cdot(P+1)}$, $2^{-n\cdot P}$ for the $\mathscr{L}^2$ and $\mathscr{H}^1$ errors, respectively, where P is the order of the macro FEM and $n=3,\ldots 8$. For the computations with P1 polynomials in the offline stage, the results are

in accordance with the theoretical results. Whenever the macro error reaches 10^{-7}, we lose the expected convergence rates which may indicate that micro errors are of the same order of accuracy (this can be seen in Fig. 4 (a)-(b)).

For the computation with P2 polynomials in the offline stage, one would expect the effect of micro errors appearing for macro errors around 10^{-13}, but we see in Fig. 4 (c) that the expected convergence rates are lost for errors around 10^{-11}. Here we think that what is seen is the r_{RB} error term, i.e., the accuracy of the reduced basis procedure. Indeed, according to the a posteriori error estimator (controlling this latter error), the accuracy reached for P2 polynomials in the offline procedure with the given mesh is around 10^{-11} (see Table (4.1)).

How does the RB-FE-HMM compare with the FE-HMM? We next show some comparisons with $u^{H,RB}$ computed with P1 offline and online polynomials.

Table 4.2: Comparison between RB-FE-HMM (P1-FEM as offline and online solver) and FE-HMM (P1-FEM as micro and macro solver) for the $\mathscr{L}^2$ error.

Mesh	RB-FE-HMM offline mesh 350×350 $\mathscr{L}^2$ Error	RB-FE-HMM offline mesh 500×500 $\mathscr{L}^2$ Error	FE-HMM $\mathscr{L}^2$ Error
8×8	0.0161	0.0161	0.0176
16×16	0.0040	0.0040	0.0044
32×32	0.0010	0.0010	0.0011
64×64	2.5347e-04	2.5306e-04	2.7702e-04
128×128	6.3969e-05	6.3561e-05	6.9259e-05
256×256	1.6599e-05	1.6184e-05	1.7315e-05

Table 4.3: Comparison of CPU time between the RB-FE-HMM (P1-FEM as the offline and online solver) and the FE-HMM (P1-FEM as the micro and macro solver). The offline CPU time is $193s$ with meshsize 350×350 and $424s$ with meshsize 500×500.

Mesh	RB-FE-HMM Online CPU Time (s) with 1 processor	FE-HMM CPU Time (s) with 1 processor
8×8	0.03	0.14
16×16	0.10	0.98
32×32	0.28	109
64×64	1.21	1760
128×128	4.92	27504
256×256	20.33	332410

We choose two different offline meshes, namely 350×350 and 500×500 and obtain 10 RB (as previously). For the FE-HMM solution u^H, we use P1-FEM for both the macro and micro solvers, where simultaneous refinement is needed according to estimates (1.2.17), (1.2.18)rmic. The errors in the $\mathscr{L}^2$ norm and the computation time are reported in Tables 4.2 (error) and 4.3 (computation time). In Table 4.2, we see that the offline mesh of size 350×350 is fine enough to get the optimal (quadratic) convergence rate. The simultaneous refinement for the FE-HMM,

i.e., $H \simeq (h/\varepsilon) \simeq 2^{-n},\ n = 3,\dots 8$, also gives the optimal (quadratic) convergence rate. The errors for both methods are similar. Now we compare in Table 4.3 the computation time. Taking into account the offline stage, we see that the total cost is an order of magnitude smaller for the RB-FE-HMM except for very coarse macro meshes, where overhead given by the cost of the offline computation for the RB-FE-HMM dominates the cost for the FE-HMM. As can be seen from these computations, for errors smaller than 10^{-4}, the RB-FE-HMM is always more efficient than the FE-HMM.

Notice that in our comparisons between the FE-HMM and the RB-FE-HMM we only used P1 online macro FEMs. Because of the increasing number of cell problems that need to be solved for the FE-HMM when using higher order macro polynomials (due to the increasing number of quadrature points and related sampling domains), this method becomes very expensive. In contrast, only the macro assembly is affected in the RB-FE-HMM (similarly as for standard FEM) when using higher order macro solver. Thus for the RB-FE-HMM the cost of increasing the degree of the macro polynomials is proportional to the macro DOFs only and is *similar* to the cost of increasing the polynomials degree in standard FEM.

2D affine multiscale discontinuous tensor. In this example, we test the RB-FE-HMM on a problem with an oscillating tensor discontinuous on the sampling domains. Such tensors prevent the use of fast microsolvers (e.g., based on pseudo-spectral methods as proposed in [15]). This is why we distinguish in our experiments continuous versus discontinuous affine tensors. We assume that the reference domain is divided into three subdomains $Y = Y_A \cup Y_B \cup Y_C$ with different tensors in the different domains, discontinuous at the interfaces. The tensor is defined by

$$a(x, \frac{x}{\varepsilon}) = a_A(x, \frac{x}{\varepsilon}) I_A + a_B(x, \frac{x}{\varepsilon}) I_B + a_C(x, \frac{x}{\varepsilon}) I_C,$$

where I_A, I_B, I_C are the indicator functions of the domains Y_A, Y_B, Y_C and $a_A(x, \frac{x}{\varepsilon})$, $a_B(x, \frac{x}{\varepsilon})$, $a_C(x, \frac{x}{\varepsilon})$ are diagonal tensors with entries given by

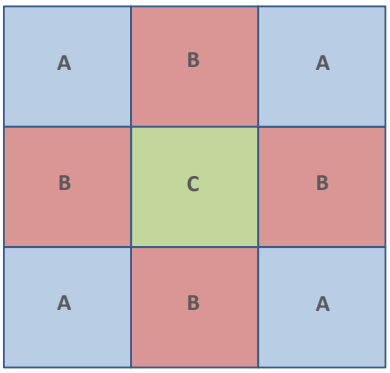
$$a_{A,ii}(x, \frac{x}{\varepsilon}) = \begin{cases} x_1^2 + 0.2 + (x_2+1)(\sin(2\pi \frac{x_1}{\varepsilon}) + 2), & i = 1 \\ x_2^2 + 0.05 + (x_1 x_2 + 1)(\sin(2\pi \frac{x_2}{\varepsilon}) + 2), & i = 2 \end{cases}$$

$$a_{B,ii}(x, \frac{x}{\varepsilon}) = \begin{cases} 3x_1 + x_2^2 + 0.5, & i = 1 \\ e^{(-x_1 - x_2)}(\cos(2\pi \frac{x_1}{\varepsilon}) + 2), & i = 2 \end{cases}$$

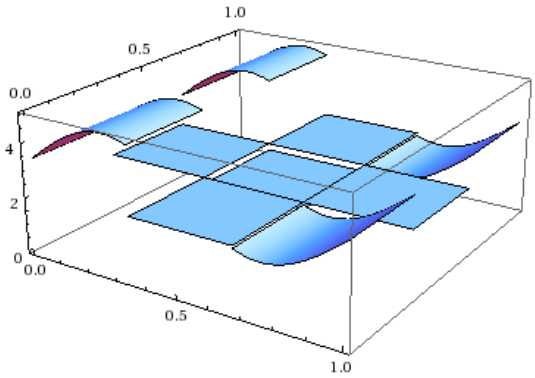
$$a_{C,ii}(x, \frac{x}{\varepsilon}) = x_1 + x_2 + 1,\ i = 1, 2.$$

Notice that the above tensor could model a material having different phases with different conductivity properties in each phase. The discontinuities over the phases are illustrated in Fig. 5 (b)-(c). Table 4.4 provides information of the offline stage.

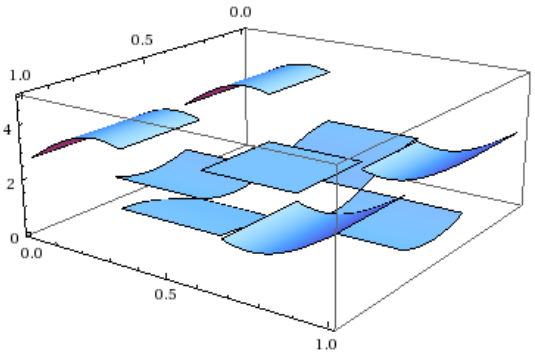
For this example, we use the FE-HMM solution with fine micro and macro meshes to compute a reference solution (we choose a grid of 512×512 for the micro and macro meshes, respectively). We display in Fig. 6 (a) the RB-FE-HMM solution with macro mesh 128×128. In Fig. 6 (b) we report the $\mathscr{L}^2$ and $\mathscr{H}^1$ convergence rates (as we refine the macro mesh for the RB-FE-HMM). As



(a) The substructure of the micro reference domain.



(b) Upper diagonal term of $a(x, \frac{x}{\varepsilon}) = a(x, y)$ with $(x_1, x_2) = (0.5, 0.5)$.



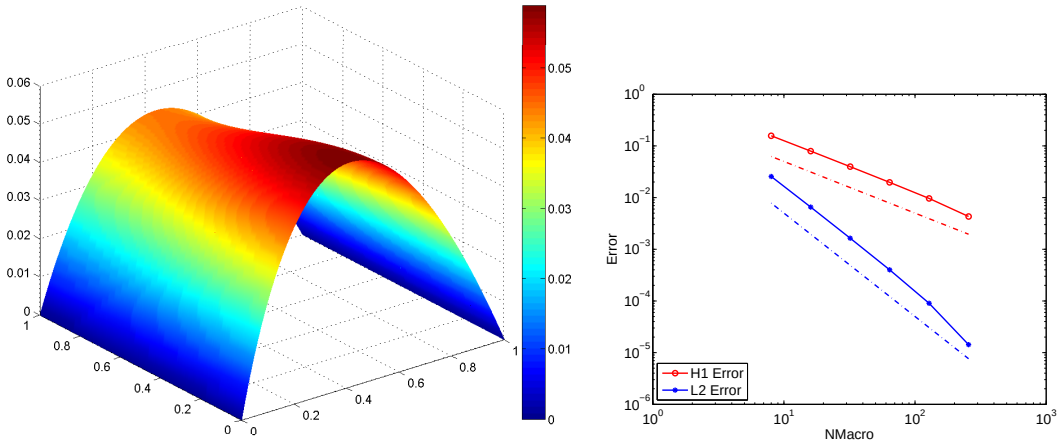
(c) Lower diagonal term of $a(x, \frac{x}{\varepsilon}) = a(x, y)$ with $(x_1, x_2) = (0.5, 0.5)$

Figure 5: Domain and tensor.

Table 4.4: Parameters for RB-FE-HMM offline stage (discontinuous tensor).

Meshsize for the micro reference domain Y	901×901
Initial sample points number $\Xi_{RB} \subset \Omega$	800
Tolerance of the offline stage tol_{RB}	1e-07
Offline solver	P1-FEM
Method used to compute the stability factor	SCM
RB number	40

can be seen, we obtain optimal convergence rates for this example. This shows the efficiency of the RB strategy in situations (discontinuous tensors) that prevent the use of fast micro solvers taking advantage of the smoothness of the micro solution.



(a) $u^{H,RB}$ (b) $\|u^{H,RB} - u^{ref}\|$

Figure 6: RB-FE-HMM for discontinuous tensor. (a). RB-FE-HMM solution computed with a 128×128 online macro mesh. (b). The error $\|u^{H,RB} - u^{ref}\|$ is displayed, where 40 reduced bases are used for $u^{H,RB}$. The online solver is a P1-FEM. The reference solution u^{ref} is computed by the FE-HMM with 512×512 micro and macro meshes. The dashed lines are the reference lines with the slope 1, 2, respectively.

2D nonaffine multiscale tensor. For the last 2-dimensional example, we consider a tensor that is not in affine form. As mentioned in Section 4.2, we have to apply the EIM to obtain an affine representation of the tensor to implement the RB methodology. We take a tensor of the form

$$a(x,\frac{x}{\varepsilon})_{11} = \Big(\sqrt{(x_1^2+\sin(2\pi\frac{x_1}{\varepsilon})+1.2)(x_1x_2+\sin(4\pi\frac{x_1}{\varepsilon})+1.5)}\Big)^{-1}$$
$$a(x,\frac{x}{\varepsilon})_{22} = \big((x_1x_2+\sin(5\pi\frac{x_2}{\varepsilon})+1.2)(x_2^2\cos(2\pi\frac{x_2}{\varepsilon})+x_1+1.5)\big)^{-1}$$
$$a(x,\frac{x}{\varepsilon})_{12} = a(x,\frac{x}{\varepsilon})_{21} = 0$$

chosen in such a way that the homogenized tensor can be computed easily for numerical comparison purpose. It is given by

$$a^0_{11} = \Big(\int_0^1 \sqrt{(x_1^2+\sin(2\pi y_1)+1.2)(x_1x_2+sin(4\pi y_1)+1.5)}\,dy_1\Big)^{-1}$$
$$a^0_{22} = \Big(\int_0^1 (x_1x_2+\sin(5\pi y_2)+1.2)(x_2^2\cos(2\pi y_2)+x_1+1.5)\,dy_2\Big)^{-1}$$
$$a^0_{12} = a^0_{21} = 0$$

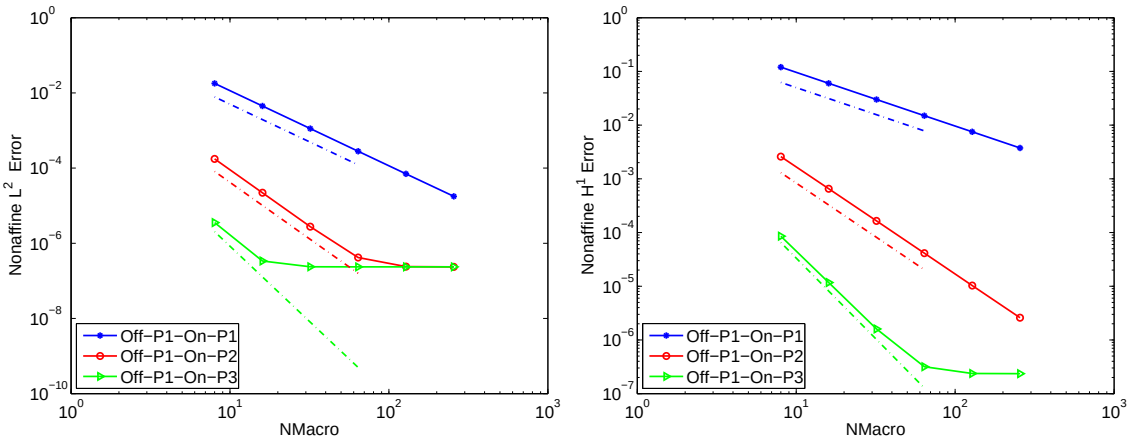
In Table 4.5 we report the parameters of the EIM offline stage. We obtained 19 affine terms for the first diagonal entry of $a(x, \frac{x}{\epsilon})$, and 26 terms for the second diagonal entry.

Table 4.5: Parameters for EIM offline stage.

Initial sample points number $\Xi_{EIM} \subset \Omega$	600
Tolerance of the EIM tol_{EIM}	1e-06
Number affine terms for $(a(x, \frac{x}{\epsilon}))_{11}$	19
Number affine terms for $(a(x, \frac{x}{\epsilon}))_{22}$	26
EIM CPU time(s)	5461

Table 4.6: Parameters for the RB-FE-HMM offline stage.

Meshsize for the micro reference domain Y	1200×1200
Initial sample points number $\Xi_{RB} \subset \Omega$	800
Tolerance of the offline stage tol_{RB}	1e-08
Offline solver	P1 FEM
Method used to compute the stability factor	SCM
RB number	13
Offline CPU time (s)	57354



(a) $\mathscr{L}^2$ Error (b) $\mathscr{H}^1$ Error

Figure 7: Problem with nonaffine tensor. The error $\|u^{H,RB} - u^{ref}\|$ is displayed, where 13 reduced bases are used for computing $u^{H,RB}$. The offline parameters are reported in Table 4.6. For the online macro solver P1-FEM, P2-FEM, P3-FEM are used. The dashed lines are the reference lines with slope equal to 2,3,4 in Fig. (a) and 1,2,3 in Fig. (b). The homogenized FE solution u^{ref} is computed with a P3-FEM on a fine mesh of size 1024×1024.

Next, we report the parameters of the RB-FE-HMM offline stage in Table 4.6. Since the error from the EIM process also influences the output accuracy of the offline stage, we choose a lower accuracy requirement for the offline stage and fix tol_{RB} to be 10^{-8}. This tolerance is met by the the a posteriori error estimator of the greedy procedure after the selection of 13 bases. As we fix the tolerance at 10^{-8}, we do not expect to get optimal convergence rates when the error is smaller

than this threshold. Observe that here, an additional error term should appear in the r_{HMM} error described in Theorem 4.5.1, namely the approximation error due to the EIM. Experimentally we observe a plateau when the error reaches 10^{-7}. Until this threshold, we observe in Fig. 7 optimal convergence for the RB-FE-HMM (with P1, P2 or P3 macro FEMs). Let us mention that the tensor chosen here is continuous and that the performance of the EIM may decrease when the coefficients vary discontinuously within a sampling domain (see [46, Chapter 5]).

4.6.2 3-D test problem.

The FE-HMM, as any numerical homogenization methods, can be costly for three-dimensional problems, due to the repeated computations of micro problems on sampling domains, each of them involving an increasing number of DOF as the mesh on the computational domain is refined (remember that this requires the micro mesh on the sampling domain to be refined simultaneously to the macro mesh).

The problem considered here is the heat transfer in a microchip (see Fig.8), as described in [19]. The volume of the smallest box containing the microchip is $12.2 \times 12.2 \times 1$ mm^3. The macro

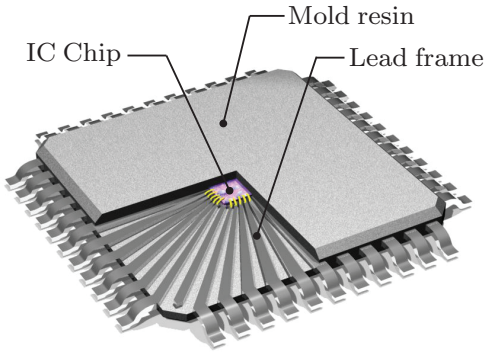


Figure 8: 3D Macro domain structure [19].

domain Ω is composed of three parts, $\Omega = \Omega_{chip} \cup \Omega_{leadframe} \cup \Omega_{resin}$, the domains of the chip, the leadframe and the package, respectively (see Fig. 8). The model equation is given by

$$\begin{aligned} -\nabla \cdot (a^\varepsilon \nabla u^\varepsilon) &= f \quad \text{in } \Omega, \\ n \cdot (a^\varepsilon u^\varepsilon) + \alpha u^\varepsilon &= g_R \quad \text{on } \partial\Omega, \end{aligned} \tag{4.6.55}$$

where $g_R = 5863[\frac{W}{m^2}]$ and $\alpha = 20$ and

$$f = \begin{cases} 1.87 \times 10^8 [\frac{W}{m^3}] & x \in \Omega_{chip}, \\ 0 & otherwise. \end{cases}$$

Here g_R represents the heat flux entering the domain and the heat exchange with the ambient temperature and f is a heat source representing the power of the chip. We take different conductivity tensors for each component. All the tensors are diagonal and given by

$$
\begin{aligned}
a^{\varepsilon}_{ii,chip} &= 140, \\
a^{\varepsilon}_{ii,leadframe} &= 400e^{20(x_1^2+x_2^2)^{\frac{1}{2}}} + 400(\cos(x_3\pi)+1.5)(\cos(2\pi x_i/\varepsilon)+1.1), \\
a^{\varepsilon}_{ii,resin} &= \begin{cases} \frac{1}{\sin(6\pi x_2)+1.6} + 3(\cos(\pi x_3)+1.5x_1^2+1.1)(\sin(4\pi x_1/\varepsilon)+1.1) & i=1, \\ \frac{1}{\sin(6\pi x_2)+1.6} + 3(\cos(\pi x_3)+1.5x_1^2+1.1)(\cos(\pi x_2/\varepsilon)+1.2) & i=2, \\ \frac{2}{\sin(6\pi x_2)+1.6} + 3(\cos(\pi x_3)+1.5x_1^2+1.1)(\cos(6\pi x_3/\varepsilon)+1.2) & i=3. \end{cases}
\end{aligned}
$$

We notice that a^{ε}_{chip} is constant and we do not need to solve any cell problem on the domain Ω_{chip}. We thus apply the RB-FE-HMM strategy on $\Omega_{leadframe}$ and Ω_{resin} respectively. Table 4.7 displays the RB offline parameters. We see that 20 reduced bases are needed for this problem.

Table 4.7: 3D RB-FE-HMM offline parameters.

Domain	Initial sample points number	Offline mesh	tol_{RB}	RB number
$\Omega_{leadframe}$	400	$220 \times 220 \times 220$	1e-10	11
Ω_{resin}	600	$220 \times 220 \times 220$	1e-10	9

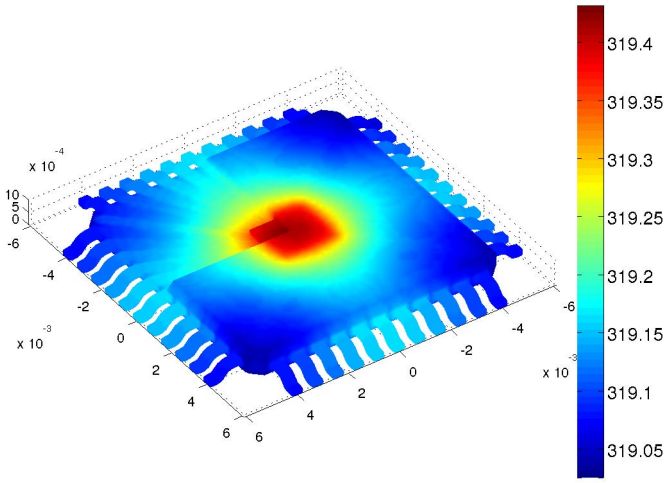


Figure 9: 3D RB-FE-HMM solution $u^{H,RB}$. The offline parameters are reported in Table 4.7. Online DOF: 37011.

A solution of problem (4.6.55) computed with the RB-FE-HMM is shown in Fig. 9. For this computation a macro mesh with 37011 DOF is used. In Table 4.8, we show error estimates when comparing the RB-FE-HMM solution $u^{H,RB}$ with a numerically computed reference solution u^{ref} for the homogenized problem. For 3D problems, it is not a trivial task to compute an

accurate reference solution. The reference solution is computed as follows. As $a^{\varepsilon}_{leadframe}$ and a^{ε}_{resin} are diagonal tensors, the corresponding homogenized tensors $a^{0}_{ii,leadframe}$ and $a^{0}_{ii,resin}$ are the harmonic averages of the fine scale tensors. These harmonic averages (involving 1D integrals) are further evaluated by using an accurate numerical integration scheme.

Because of the difficulty to obtain an accurate reference solution, we only display the accuracy of the RB-FE-HMM for one refinement step. A refinement step (corresponding roughly to a meshsize divided by two), starting with the initial mesh with 37 011 DOF (corresponding to 190 081 tetrahedra), leads to 278 123 DOF (corresponding to 1 520 648 tetrahedra). The reference solution is computed with a mesh with 2 108 977 DOF (corresponding to 12 165 184 tetrahedra). We display in Table 4.8 the error in various norms, when we refine the macro mesh. We see that the $\mathscr{H}^1$ and the $\mathscr{L}^2$ errors have the expected decay rate.

Table 4.8: Error estimates for problem (4.6.55). The solution u^{ref} is computed with 2108977 DOF, $\|u^{ref}\|_A = 24.2616$.

Online mesh DOF	$\|u^{H,RB}\|_A$	$\|u^{H,RB}-u^{ref}\|_{\mathscr{L}^2(\Omega)}$	$\|u^{H,RB}-u^{ref}\|_{\mathscr{H}^1(\Omega)}$	$\frac{\|u^{H,RB}-u^{ref}\|_{\mathscr{L}^\infty(\Omega)}}{\|u^{ref}\|_{\mathscr{L}^\infty(\Omega)}}$
37011	24.2465	0.0001665	0.0214941	0.0002173
278123	24.2604	5.6277e-05	0.0081573	7.9756e-05

We emphasize that once the reduced 20 bases are computed in the offline stage, the online stage is quite inexpensive. In contrast, computations with the FE-HMM would require to solve a very large number of cell problems on sampling domains (about 1.5 million for the mesh with 278 123 DOF), with a meshsize adapted to the rate of decay of the macro meshsize. For a macro mesh that is not overly coarse (this is already required to properly represent the geometry of the microchip), a computation with the FE-HMM is much more costly than with the RB-FE-HMM, even for a single computation when taking into account the offline and the online costs for this latter method. For such 3D problems it is thus very advantageous to use the RB-FE-HMM even for computation with low order (piecewise linear) macro FEMs.

4.6.3 Discussion

We have presented in this chapter an efficient FEM for high order discretizations of elliptic homogenization problems based on micro-macro solvers combined with a RB strategy. In our new method, the RB-FE-HMM, repeated FEM computations of micro problems (at quadrature points of a macro mesh) are avoided. These repeated micro computations are the main computational overhead of a numerical homogenization method such as the FE-HMM, when accurate macro solutions need to be computed. In turn, the RB-FE-HMM is not only more efficient for high order macroscopic disretizations, but also for three-dimensional problems, already for low order macroscopic discretizations when even a single micro problem *in each* element of the macroscopic mesh is expensive to compute. Using interpolation techniques following the reduced basis methodology, we showed that an efficient numerical method can be designed, relying only on a small number of accurately computed micro solutions. An a posteriori error estimate for the selection of representative micro solutions has been discussed. We have derived

an a priori error analysis which allows to describe the decay rate of the various discretization errors involved in our numerical approach. The efficiency of the RB-FE-HMM strategy has been illustrated by several numerical examples and comparisons with the classical FE-HMM have shown significant improvements.

5 Adaptive reduced basis finite element heterogeneous multiscale method

Following the discussion of the previous chapter, we present in this chapter a posteriori error control and mesh refinement strategy for the RB-FE-HMM which can be widely applied to many practical engineering problems. Being able to refine the computational mesh adaptively, based on a given actual numerical solution is of prime importance. Adaptive methods not only provide a criterion that indicates whether a certain prescribed accuracy is met, but also estimate local errors that allow to drive a mesh refinement that equi-distributes the approximation error (e.g., refining in the region where singularity in the solution or in the domain occur). The general adaptive strategy for the adaptive FEM is based on the following cycle

$$\text{Solve} \rightarrow \text{Estimate} \rightarrow \text{Mark} \rightarrow \text{Refine}.$$

The procedure can be understood as follows. One first solves the partial differential equation numerically on the current mesh and computes an a posteriori error by defining suitable error estimators that give estimations of the actual error between the computed and the true solution (in a certain norm). Then, according to the distribution of the error estimator in each element, one marks the elements which have the largest contribution to the estimated error, refines the marked elements and goes back to the step "Solve". The essential step for adaptive method is the construction of reliable error estimator. While, for single scale problems, there is a large literature for adaptive methods (see for example [27, 105, 87, 41, 28, 38] and the references therein), the literature for multiscale FE methods is more scarce, we mention however the work of Ghosh et al. [78, 79], where an adaptive micro-macro method (based on Voroni cell finite elements) was proposed for elasticity problems. While a posteriori indicators have been derived, upper and lower bounds in term of these indicators that take into account macro and micro meshes were not derived. We note also that the algorithm has been extended in [79] to account for model adaptivity, i.e., correction of the homogenized solution where the model is not accurate enough.

One difficulty is that error estimators (that first need to be suitably modeled) depend on multiple scales and the reliability of such error estimators depends on the accuracy of the resolution of the fine scale. We briefly discuss an approach that is complementary to the approach described here, namely that of the hierarchical and goal-oriented adaptivity for multiscale problems proposed

and developed in [107, 89, 87, 88]. The goal there is the following: assuming that an effective homogenization solution u^0 and an effective problem are at hand, the task is then to improve this coarse model in order to better approximate the fine scale solution in a given quantity of interest. Assuming that $Q(u^\varepsilon)$ is a quantity of interest (Q will be either the energy norm or a bounded linear functional in this chapter), the general procedure is based on the decomposition of the error into modeling error $Q(u^\varepsilon - u^0)$ and numerical error $Q(u^0 - u^{H,h})$ [1]. The focus is then on reducing the error $Q(u^\varepsilon - u^0)$ by an adaptive strategy that enriches the solution u^0 with fine scale features (obtained by solving locally the fine scale models) based on suitable a posteriori indicators. The term $Q(u^0 - u^{H,h})$ is not analyzed in this work. It is mentioned in [89] that "the choice of homogenization techniques has significant impact on the performance of error estimators and the success of the (goal oriented) method". This is the focus of this chapter, to propose an adaptive algorithm for the computation of $Q(u^0 - u^{H,h})$. As mentioned in Chapter 1 when the fine scale problem is not globally periodic, then the homogenized problem relies on an infinite number of micro problems and they have thus to be properly localized and solved with an accuracy that is proportional to the macroscopic mesh. In turn, variational crimes are made which prevent the use of classical adaptivity techniques for FEM. Furthermore, as the data of the homogenized problems depend on the accuracy of micro problems, this discretization error has also to be taken into account while deriving a posteriori upper and lower bound for the error. The accuracy $Q(u^0 - u^{H,h})$ then depends on a macro mesh, and micro meshes (used to solve the micro problems) and a modeling error (originating from the constrain between micro and macro solvers). In the present work our adaptive reduced basis method is shown to be an efficient and reliable method to compute adaptively $Q(u^0 - u^{H,h})$ and in this sense both approaches focusing on $Q(u^\varepsilon - u^0)$ or $Q(u^0 - u^{H,h})$ are complementary and need to be carried out for successful adaptive multiscale computations.

A posteriori error analysis for the FE-HMM has first been given in [90] based on a two-scale analysis [81]. A posteriori error analysis for the FE-HMM in the physical energy-norm has been derived in [20, 85] and the extension of the FE-HMM to goal oriented adaptivity (in quantities of interest) has been discussed in [21]. For this latter adaptive method, the solution of dual problems (in a higher macro FE space) are required. Both adaptive methods [20, 21] have proved to be useful and more efficient than the FE-HMM with uniform refinement when the macroscopic domain or the macroscopic solution exhibit singularities. But as mentioned above, the simultaneous mesh refinement still represents a significant computational overhead that prevents to solve efficiently three-dimensional problems adaptively, or to use adaptive higher order macro FEM. In this chapter we discuss a new adaptive method based on the RB-FE-HMM introduced in Chapter 4. This method, that allows to drive a macroscopic mesh refinement using the same precomputed set of RB micro functions to estimate the effective data, improves significantly the adaptive FE-HMM and provides a highly efficient adaptive multiscale computational strategy. In particular this methods allows to use *the same precomputed set* of RB micro functions to:

[1] Here we put a reference to the size of a macroscopic and a microscopic meshes, H and h, respectively to emphasize that an approximation of the homogenized solution usually depends on at least two meshes. The macroscopic mesh that meshes the physical domain and a microscopic mesh that meshes micro problems to retrieve locally the parameters (tensors) of the homogenized problem.

- drive a macroscopic mesh refinement and estimate the effective data;
- use higher order macro FEM in an adaptive procedure;
- compute primal and dual numerical homogenization problems for an adaptive mesh refinement in quantities of interest;
- compute adaptively several online problems with different source terms.

We mention that the adaptive RB-FE-HMM could also be used for h-p refinement strategy. Indeed, the generalization and the computational efficiency of the RB-FE-HMM for higher order macro FEs make this method a good candidate for such simultaneous refinement methods that remain up to now challenging for numerical homogenization methods such as the FE-HMM.

The chapter is organized as follows. In Section 5.1, we discuss the residual based adaptive RB-FE-HMM and the corresponding a posteriori estimates for the upper and lower bound in the energy norm. The goal-oriented adaptive RB-FE-HMM is presented in Section 5.2, where an exact representation of the error in a quantity of interest is derived. Numerical examples for the proposed method in two and three dimensions and comparison with the adaptive FE-HMM are presented in Section 5.3.

5.1 The energy norm based adaptive RB-FE-HMM

In this section, we introduce the adaptive RB-FE-HMM. For simplicity, we only consider simplicial macro FEs in what follows (notice that for such elements **(Q2)** implies **(Q1)**, see Section 1.2.1). For the RB-FE-HMM offline procedure we refer to Section 4.2 and here we start by discussing for the online stage a posteriori error estimates in the energy norm. Such results have first been obtained for the FE-HMM in [20] for linear macro FEs. To obtain a more efficient algorithm, we apply the RB strategy for the adaptive FE-HMM and in turn avoid having to refine the micro mesh simultaneously to the adaptive macro mesh refinement. We also prove the upper and lower a posteriori error bound for arbitrary order of the macro FE space.

Reduced basis multiscale jump and flux. Based on the reformulation of the RB-FE-HMM discussed in Section 4.3, one can write (4.3.37) in the following form (see [4]),

$$a_N^0(x_{K_j}) = \int_Y a_{x_{K_j}}(y)\left(I + J^T_{\hat{\chi}^i_{N,K_j}}(y)\right)dy, \tag{5.1.1}$$

where $J_{\hat{\chi}^i_{N,K_j}}$ is a $d\times d$ tensor, defined as $\left(J_{\hat{\chi}^i_{N,K_j}}(y)\right)_{ik} = (\partial\hat{\chi}^i_{N,K_j})/(\partial y_k)$. Therefore one can show that for $\forall\, v^H \in S^\ell_0(\Omega,\mathcal{T}_H)$ the relation below holds (see [5])

$$\frac{1}{|K_{\delta_j}|}\int_{K_{\delta_j}} a_{x_{K_j}}(G^{-1}_{x_{K_j}}(x))\nabla v_{N,K_j}(x)dx = a_N^0(x_{K_j})\nabla v^H(x_{K_j}). \tag{5.1.2}$$

We introduce in each macro element K an interpolation polynomial $\Pi^K_{a\nabla v_N}(x) \in \mathscr{P}^{l-1}(K)$ based on the J quadrature points used for the RB-FE-HMM (see [5, 20]). Assuming in addition to (**Q2**) that $J = \frac{1}{2}\ell(\ell+1),\ d=2$ or $J = \frac{1}{6}\ell(\ell+1)(\ell+2),\ d=3$, then we define

$$\Pi^K_{a\nabla v_N}(x_{K_j}) = \frac{1}{|K_{\delta_j}|}\int_{K_{\delta_j}} a_{x_{K_j}}(G^{-1}_{x_{K_j}}(x))\nabla v_{N,K_j}(x)dx,\ j=1,\cdots,J. \tag{5.1.3}$$

The polynomial $\Pi^K_{a\nabla v_N}$ is called “multiscale flux". Combing (5.1.2) with (5.1.3), we have

$$\Pi^K_{a\nabla v_N}(x_{K_j}) = a^0_N(x_{K_j})\nabla v^H(x_{K_j}). \tag{5.1.4}$$

Remark 5.1.1. *The interpolation polynomial is uniquely determined by condition* (5.1.3)*, provided* **(Q2)** *holds and*

$$\begin{aligned} &J = \frac{1}{2}\ell(\ell+1) \quad d=2,\\ &J = \frac{1}{6}\ell(\ell+1)(\ell+2) \quad d=3. \end{aligned} \tag{5.1.5}$$

With such a choice for J, one can deduce that $\Pi^K_{a\nabla v_N}(x) \in \mathscr{P}^{\ell-1}(K)$ (see [20, 21]). We notice that quadrature formulas for which (5.1.5) *holds indeed exist (see [106, 101]).*

We denote by K^+ and K^- two macro elements with a non empty interface given by $e = K^+ \cap K^-$. Now we define the RB multiscale jump on the interior interface e as

$$[\![\overline{\Pi^K_{a\nabla v_N}}]\!]_e(s) := \begin{cases} (\Pi^{K^+}_{a\nabla v_N}(s) - \Pi^{K^-}_{a\nabla v_N}(s))\cdot n_e & \text{for } e \not\subset \partial\Omega,\\ 0 & \text{for } e \subset \partial\Omega, \end{cases} \tag{5.1.6}$$

where n_e is the outward normal vector of e on ∂K^+.

Remark 5.1.2. *When the macro FE space is $S^1_0(\Omega,\mathscr{T}_H)$ with simplicial element, $J=1$ and we have only one quadrature point per macro element. Then* (5.1.6) *can be simply written as*

$$[\![\overline{a^\varepsilon\nabla v_{N,K}}]\!]_e := \begin{cases} \left(\frac{1}{|K^+_\delta|}\int_{K^+_\delta} a_{x_{K^+}}(G^{-1}_{x_{K^+}}(x))\nabla v_{N,K^+}(x)dx - \right.\\ \left.\frac{1}{|K^-_\delta|}\int_{K^-_\delta} a_{x_{K^-}}(G^{-1}_{x_{K^-}}(x))\nabla v_{N,K^-}(x)dx\right)\cdot n_e & \text{for } e\not\subset\partial\Omega,\\ 0 & \text{for } e\subset\partial\Omega. \end{cases} \tag{5.1.7}$$

The energy norm based a posteriori error estimates. Following [21], we introduce the local refinement indicator and the data approximation error.

Definition 5.1.3. *The local refinement indicator $\eta_H(K)$ is defined by*

$$\eta_H(K)^2 := H_K^2 \|f^H + \nabla \cdot \Pi^K_{a\nabla u_N}\|^2_{\mathscr{L}^2(K)} + \frac{1}{2}\sum_{e\subset\partial K} H_e \|[\![\overline{\Pi^K_{a\nabla u_N}}]\!]_e\|^2_{\mathscr{L}^2(e)}, \tag{5.1.8}$$

and the local data approximation error is defined by

$$\xi_H(K)^2 := H_K^2 \|f^H - f\|^2_{\mathscr{L}^2(K)} + \|\Pi^K_{a\nabla u_N} - a^0\nabla u^{H,RB}\|^2_{\mathscr{L}^2(K)}, \tag{5.1.9}$$

where a^0 is the tensor of the homogenized problem (1.1.3) *and f^H is an approximation of f in the space $\{g \in \mathscr{L}^2(\Omega), g|_K \in \mathscr{P}^m(K), \forall K \in \mathscr{T}_H\}$.*

Remark 5.1.4. *If the macro FE space is $S^1_0(\Omega,\mathscr{T}_H)$, the local refinement indicator and local data approximation can be written respectively, in the following way,*

$$\eta_H(K)^2 := H_K^2 \|f^H\|^2_{\mathscr{L}^2(K)} + \frac{1}{2}\sum_{e\subset\partial K} H_e \|[\![\overline{a^\varepsilon \nabla u_{N,K}}]\!]_e\|^2_{\mathscr{L}^2(e)}, \tag{5.1.10}$$

$$\xi_H(K)^2 := H_K^2 \|f^H - f\|^2_{\mathscr{L}^2(K)} + \|(a^0_N - a^0)\nabla u^{H,RB}\|^2_{\mathscr{L}^2(\Omega)}. \tag{5.1.11}$$

We now explain how we obtain the expressions of the local indicator (5.1.8) and data approximation (5.1.9).

Lemma 5.1.5. *Define $e^H = u^0 - u^{H,RB}$. For $\forall v \in H^1_0(\Omega)$, the following representation formula holds*

$$\begin{aligned} B_0(e^H, v) &= \int_\Omega f^H v dx - \sum_{K\in\mathscr{T}_H}\Big(\sum_{e\subset\partial K}\int_e [\![\overline{\Pi^K_{a\nabla u_N}}]\!]_e(s) v ds + \int_K \nabla\cdot\Pi^K_{a\nabla u_N} v dx \\ &+ \int_K (\Pi^K_{a\nabla u_N} - a^0\nabla u^{H,RB})\cdot\nabla v dx + \int_K (f - f^H) v dx\Big), \end{aligned} \tag{5.1.12}$$

where $B_0(\cdot,\cdot)$ is the bilinear form of the homogenized equation (1.1.3).

Expression (5.1.12) is instrumental to derive upper and lower a posteriori error bounds.

Proof. The proof follows [20, Lemma 9] and we sketch the idea. First we write

$$B_0(e^H, v) = \int_\Omega f v dx - \sum_{K\in\mathscr{T}_H}\int_K a^0\nabla u^{H,RB}\nabla v dx.$$

We then add and subtract the term $\sum_{K\in\mathscr{T}_H}\int_K \Pi^K_{a\nabla u_N}\cdot\nabla v + f^H v dx$ in the above expression, use integration by part for the expression $\sum_{K\in\mathscr{T}_H}\int_K a^0\nabla u^{H,RB}\nabla v dx$, to finally obtain (5.1.12). □

The following two theorems give upper and lower bounds of the error $\|u^0 - u^{H,RB}\|_{\mathscr{H}^1(\Omega)}$ in terms of the local refinement indicator $\eta_H(K)$ and local data approximation error $\xi_H(K)$. As the proofs of the theorems are very similar to those in [20], we only sketch them here for completeness.

Theorem 5.1.6. *(A posteriori upper bound) Let* $\eta_H(\Omega)^2 = \sum_{K\in T_H}\eta_H(K)^2$, $\xi_H(\Omega)^2 = \sum_{K\in T_H}\xi_H(K)^2$. *There exists a constant* $C > 0$, *such that*

$$\|u^0 - u^{H,RB}\|^2_{\mathscr{H}^1(\Omega)} \leq C(\eta_H(\Omega)^2 + \xi_H(\Omega)^2).$$

Sketch of the proof: Due to the low regularity of the exact solution (only assumed to be in $\mathscr{H}^1(\Omega)$), one considers the Clément interpolation operator $I^H e^H$ of $e^H = u^0 - u^{H,RB}$, where $I^H e^H \in S^1_0(\Omega, \mathscr{T}_H)$ (see [55]). By adding the equation $B_{H,RB}(u^{H,RB}, I^H e^H) - \sum_{K\in\mathscr{T}_H}\int_K f I^H e^H dx = 0$ to (5.1.12) and writing $\psi^H = e^H - I^H e^H$, we obtain

$$\begin{aligned} B_0(e^H, e^H) &= \int_\Omega f^H \psi^H dx + \sum_{K\in\mathscr{T}_H}\Big(\int_K \nabla\cdot\Pi^K_{a\nabla u_N}\psi^H dx + \sum_{e\in\partial K}\int_e [\![\overline{\Pi^K_{a\nabla u_N}}]\!]_e \psi^H ds\Big) \\ &+ \int_\Omega (f - f^H)\psi^H dx + \sum_{K\in\mathscr{T}_H}\int_K (\Pi^K_{a\nabla u_N} - a^0\nabla u^{H,RB})\cdot\nabla e^H dx. \end{aligned}$$

The upper bound can then be obtained by using the Cauchy-Schwarz inequality and the property $\|\psi^H\|_{\mathscr{L}^2(K)} \leq CH\|\nabla e^H\|_{\mathscr{L}^2(K)}$ of the Clement interpolation operator to the above expression.

Theorem 5.1.7. *(A posteriori lower bound) There exists a constant C such that*

$$\eta_H(K)^2 \leq C(\|u^0 - u^{H,RB}\|^2_{\mathscr{H}^1(\omega_K)} + \xi_H(\omega_K)^2),$$

where ω_K *is the union of all the elements sharing an interface with K.*

Sketch of the proof: The first step is to estimate the term $H_K^2\|f^H + \nabla\cdot\Pi^K_{a\nabla u_N}(x)\|^2_{\mathscr{L}^2(K)}$ in (5.1.8) which is the so called interior residual. For that, one needs to consider an interior bubble function ψ_K in an FE space defined over a refinement $\tilde{\mathscr{T}}_H$ of $\mathscr{T}_H$ so that every $K \in \mathscr{T}_H$ has an interior node $\tilde{x}_K$ in $\tilde{\mathscr{T}}_H$ (likewise every edge e of $\tilde{\mathscr{T}}_H$ not on the boundary $\partial\Omega$ mush have an interior node in $\tilde{\mathscr{T}}_H$). For any $K \in \mathscr{T}_H$, ψ_K has the properties that $0 \leq \psi_K \leq 1$, $\psi_K(\tilde{x}_K) = 1$ and $\psi_K = 0 \in \Omega\setminus K$. We choose the test function $v = \psi_K(f^H + \nabla\cdot\Pi^K_{a\nabla u_N})$, insert it into (5.1.12). By applying the Cauchy-Schwarz inequality and the inverse inequality we obtain

$$H^2\|f^H + \nabla\cdot\Pi^K_{a\nabla u_N}\|^2_{\mathscr{L}^2(K)} \leq C(\|\nabla e^H\|^2_{\mathscr{L}^2(K)} + \xi_H(K)^2).$$

The second step of the proof is to estimate the jump residual which corresponds to the second term in (5.1.8). Now we need to introduce the edge bubble function ψ_e. Assume w_e is the

common edge shared by two elements K_1 and K_2 and that $x_{w_e} \in w_e$ is an interior node. We introduce the bubble function ψ_e which satisfies $\psi_e(x_{w_e}) = 1$, $\psi_e|_{\partial w_e} = 0$, $0 \leq \psi_e(s) \leq 1$, $\forall s \in w_e$ and $\psi_e \equiv 0$ on $\Omega \setminus (K_1 \cup K_2)$. Then we choose the test function $v(x)$ such that

$$v(x)|_{w_e} = \overline{[\![\Pi^K_{a\nabla u_N}]\!]_e}(s)\psi_e(s), \text{ and } v(x) = 0,\ x \in \Omega \setminus (K_1 \cup K_2).$$

Inserting this test function into (5.1.12) and using the Cauchy-Schwarz inequality and the estimate $\|\nabla v\|_{\mathscr{L}^2(K_i)} \leq CH^{-1/2}\|\overline{[\![\Pi^K_{a\nabla u_N}]\!]_e}\psi_e\|_{\mathscr{L}^2(w_e)}$, $i = 1,2$, we can show that

$$H\|\overline{[\![\Pi^K_{a\nabla u_N}]\!]_e}\|^2_{\mathscr{L}^2(e)} \leq C(\|\nabla e^H\|^2_{\mathscr{L}^2(\omega_K)} + \xi_H(\omega_K)^2).$$

Combining the estimates for the interior and the jump residual leads to the claimed lower bound.

Remark 5.1.8. *Let us have a closer look at the data approximation error. The contribution to the data approximation error given by the term $H_K^2\|f - f^H\|_{\mathscr{L}^2(K)}$ depends on the accuracy of the approximation f^H of f. This term also arises in single-scale energy norm adaptive FEM. For the second term $\|\Pi^K_{a\nabla u_N} - a^0\nabla u^{H,RB}\|^2_{\mathscr{L}^2(\Omega)}$, if one assumes that $a^\varepsilon(x) = a(x, x/\varepsilon) = a(x,y)$ is Y-periodic in y, $a_{ij}(\cdot,y)|_K$ is constant and $a_{ij}(x,\cdot) \in \mathscr{W}^{1,\infty}_{per}(Y)$, for all $i,j = 1,\cdots,d$, one can prove that $\|\Pi^K_{a\nabla u_N} - a^0\nabla u^{H,RB}\|_{\mathscr{L}^2(\Omega)} \leq Cr_{HMM}$, where $r_{HMM} = r_{MIC} + r_{RB} + r_{MOD}$ (see Section 5.2 for details). We note that r_{MIC} and r_{RB} can be estimated similarly as for the RB-FE-HMM. In particular, r_{MIC} will usually be small due to the requirement of having a very accurate FE computation in the offline stage. Likewise r_{RB} will be small (even exponentially decaying with respect to the RB number N) if appropriate smoothness in the macroscopic variation of the macro tensor holds. For locally periodic problems, when the slow variable of the tensor is collocated with quadrature points in the bilinear form* (4.3.29), *$r_{MOD} = 0$.*

Algorithm. For the adaptive RB-FE-HMM, one needs to have the offline outputs (see (4.2.27)) which can be repeatedly used for the online adaptive procedure. The adaptive online strategy is quite similar to the adaptive FE-HMM. The algorithm is however much faster due to the precomputed RB functions. We state here the complete algorithm.

Algorithm 5.1.9. *The adaptive RB-FE-HMM*

- *Offline stage: Construct the RB space $S_N(Y)$ following Algorithm 4.2.4 in Section 4.2 and store the output* (4.2.27).
- *Online stage:*

1.Solve *Compute the macro solution $u^{H,RB}$ on the current macro mesh by* (4.3.28)-(4.3.30) *and the quantity $\frac{1}{|K_{\delta_j}|}\int_{K_{\delta_j}} a_{x_{K_j}}(G^{-1}_{x_{K_j}}(x))\nabla u_{N,K_j}(x)dx$ at the sampling domains centered at each quadrature point. Store the data for the following step.*

2.Estimate *Construct the RB multiscale flux $\Pi^K_{a\nabla u_N}$ and compute the RB mutliscale jump $\overline{[\![\Pi^K_{a\nabla u_N}]\!]}$. Further, compute the error indicator $\eta_H(K)$ for all $K \in \mathscr{T}_H$. If*

$\sum_{K\in\mathcal{T}_H}\eta_H(K)^2 < tol$, *the process ends otherwise moves to the next step, where* tol *is given as a stopping criterion.*

3.Mark *Identify the macro elements marked for refinement following a suitable marking strategy, based on the refinement indicator* η_H.

4.Refine *Refine the marked the elements by applying the newest vertex refinement strategy [49] which keeps the conformity of the refined mesh. Go back to Step 1 with the refined macro mesh.*

Several marking strategies have been proposed in the literature, for example, [94, 100, 105]. Here we follow the marking strategy E in [105] which consists in finding a minimal subset $\tilde{\mathcal{T}}_H$ of $\mathcal{T}_H$ such that $\sum_{K\in\tilde{\mathcal{T}}_H}\eta_H(K)^2 \geq r^2\eta_H(\Omega)^2$, where $r\in(0,1)$ is a pre-defined parameter.

Complexity comparison with the adaptive FE-HMM. Recall that for the adaptive FE-HMM we need to refine the micro mesh simultaneously to the refinement of the macromesh. As a result, the micro problems of the adaptive FE-HMM

- have to be recomputed in each refined macro element;
- have increasing number of DOF at each iteration of the macro refinement procedure.

In contrast with the RB-FE-HMM, the micro problems are solved in the RB space whose dimension is fixed, usually small and computed once for all in the offline stage The efficiency improvement of the adaptive RB-FE-HMM compared with the adaptive FE-HMM is illustrated numerically in Section 5.3.

5.2 Goal Oriented Reduced Basis Adaptive FE-HMM

In this section, we apply the RB technique to another multiscale adaptive method, the DWR FE-HMM [21], which is based on the framework of the dual-weighted residual method. Here we want to know the error in a certain quantity of interest, e.g., the value of the macro solution at a certain point, directional point-wise derivative of the macro solution or the average of the solution on a subdomain etc. Thus, the error estimators are designed to quantify the accuracy in the quantities of interest. In turn, the mesh refinement is constructed in order to improve the accuracy of the computed quantity of interest. Generally, we define a linear bounded functional $\mathcal{J}: H_0^1(\Omega)\to\mathbb{R}$ to represent the quantity of interest. The main concern is the macroscopic error $e^H := u^0 - u^{H,RB}$ in the form of quantity of interest, i.e.

$$\mathcal{J}(e^H) = \mathcal{J}(u^0) - \mathcal{J}(u^{H,RB}).$$

The construction of the error estimators relies on a primal problem and a dual problem that are described below. In both primal and dual problems, the use of RB to compute suitable micro problems can significantly improve the efficiency of the DWR FE-HMM.

Primal and dual problems. The primal problem is the homogenized problem (1.1.3) that reads in weak form

$$B_0(u^0, v) = \int_\Omega f v dx, \ \forall v \in H_0^1(\Omega). \tag{5.2.13}$$

We use $u^{H,RB}$ to approximate u^0 and consider

$$B_{H,RB}(u^{H,RB}, v^H) = \int_\Omega f v^H dx, \ \forall v^H \in S_0^\ell(\Omega, \mathcal{T}_H). \tag{5.2.14}$$

The dual problem of (5.2.13) consists in finding $z^0 \in H_0^1(\Omega)$ which satisfies

$$B_0(\phi, z^0) = \mathcal{J}(\phi), \ \forall \phi \in H_0^1(\Omega). \tag{5.2.15}$$

Thus we can deduce that

$$\mathcal{J}(u^0 - u^{H,RB}) = B_0(u^0 - u^{H,RB}, z^0). \tag{5.2.16}$$

We emphasize that when we say in the sequel that we compute a RB-FE-HMM approximation of (5.2.13) or (5.2.15), it should be understood that we compute an effective solution based on the fine scale model for (5.2.13) or (5.2.15). Of course we do not assume (5.2.13) and (5.2.15) to be available.

We next consider a numerical approximation of (5.2.15) that will allow to estimate the right-hand side of (5.2.16). We observe that for u^H (the FE solution without numerical quadrature of the primal problem (5.2.13)), we have $\mathcal{J}(u^0 - u^H) = 0$ if taking an approximation z^H of z^0 from the same FE space as u^H because of the Galerkin orthogonality. It has thus been suggested in [36, 84] to use an FE approximation of the dual problem (5.2.15) in a higher order polynomial space. The same strategy is used for the RB-FE-HMM. We thus consider the RB-FE-HMM solution $z^{\mathcal{H},\mathcal{RB}}$ of (5.2.15) in the FE space $S^{\hat{\ell}}(\Omega, \mathcal{T}_H)$, $\hat{\ell} > \ell$, i.e. find $z^{\mathcal{H},\mathcal{RB}}$ s.t.

$$B_{\mathcal{H},\mathcal{RB}}(\phi^{\mathcal{H}}, z^{\mathcal{H},\mathcal{RB}}) = \mathcal{J}(\phi^{\mathcal{H}}), \ \ \forall \phi^{\mathcal{H}} \in S^{\hat{\ell}}(\Omega, \mathcal{T}_H), \tag{5.2.17}$$

where $B_{\mathcal{H},\mathcal{RB}}$ is the RB-FE-HMM bilinear form (4.3.29) with an appropriate quadrature scheme (i.e. satisfying **(Q2)** for the higher order polynomial space).

Remark 5.2.1. *We emphasize that $z^{\mathcal{H},\mathcal{RB}}$ is computed using the same RB space as $u^{H,RB}$ (for the micro functions). Thus the error r_{HMM} is the same for both $z^{\mathcal{H},\mathcal{RB}}$ and $u^{H,RB}$ and fixed after the offline stage. This also means that the cost to solve the micro problems in the online stage does*

not increase when using higher order macro FE space, since the DOF of the RB space remain unchanged in the online stage.

The error indicators. Before we define the error indicators, we need to mention the definition of the interior residual $R_{I,H}$ and jump residual $R_{J,H}$. These quantities are given by,

$$R_{I,H}(x)|_K = f^H(x) + \nabla \cdot (\Pi^K_{a\nabla u_N}(x)),$$
$$R_{J,H}(s)|_e = -\frac{1}{2}[\![\overline{\Pi^K_{a\nabla u_N}}]\!]_e(s),$$

where $\Pi^K_{a\nabla u_N}$ and $[\![\overline{\Pi^K_{a\nabla u_N}}]\!]_e(s)$ are defined in (5.1.3) and (5.1.6), respectively, and f^H is defined in Definition 5.1.3.

Next, we give the definition of the local error indicator $\eta_{H,G}(K)$ and data approximation $\xi_{H,G}(K)$. To distinguish them from the residual based indicators and data approximation, we will add a subscript "G" (goal oriented) for these quantities.

Definition 5.2.2. *The local error indicator is defined as*

$$\eta_{H,G}(K) := \int_K R_{I,H} z^{\mathcal{H},\mathcal{RB}} dx + \int_{\partial K} R_{J,H} z^{\mathcal{H},\mathcal{RB}} ds. \tag{5.2.18}$$

To guide the mesh refinement, we actually use the unsigned local refinement indicator defined as $\bar{\eta}_{H,G}(K) = |\eta_{H,G}(K)|$ *(see [21]).*

Remark 5.2.3. *The unsigned local refinement indicator* $\bar{\eta}_{H,G}(K)$ *leads to a good mesh refinement but not optimal since it is always positive while the global (signed) error indicator (the summation of the element contribution of* (5.2.18)*) allows cancellation between elements. This cancellation cannot be expressed by* $\bar{\eta}_{H,G}(K)$ *(we notice that finding the optimal mesh is already an issue for single scale DWR). The mesh refinement based on* $\bar{\eta}_{H,G}(K)$ *does nevertheless lead to good convergence rates. One issue is that the decay of the error can have oscillation (see the numerical experiments in Section 5.3.3).*

In order to analyse the variational crimes introduced for example by using numerical quadrature and computing $z^{\mathcal{H},\mathcal{RB}}$ in the higher order FE space $S^{\hat{\ell}}(\Omega, \mathcal{T}_H)$ etc, we need to introduce a suitable data approximation error.

Definition 5.2.4. *The data approximation error* $\xi_{H,G}(K)$ *is given by*

$$\begin{aligned}\xi_{H,G}(K) &= \int_K (\Pi^K_{a\nabla u_N} - a^0 \nabla u^{H,RB}) \cdot \nabla z^{\mathcal{H},\mathcal{RB}} dx \\ &+ B_{0,K}(u^0 - u^{H,RB}, z^0 - z^{\mathcal{H},\mathcal{RB}}) \\ &- \int_K (f^H - f) z^{\mathcal{H},\mathcal{RB}} dx,\end{aligned} \tag{5.2.19}$$

where $B_{0,K}$ is the bilinear form B_0 restricted to the element K.

The definition of data approximation error is motivated by the following exact DWR RB-FE-HMM error representation.

Theorem 5.2.5. *The exact representation of the error $e^H = u^0 - u^{H,RB}$ in the quantity of interest is given by*

$$\mathscr{J}(e^H) = \sum_{K\in\mathscr{T}_H} \eta_{H,G}(K) + \xi_{H,G}(K). \tag{5.2.20}$$

The proof follows [21, Theorem 4]. We sketch the main steps. Considering (5.2.15) and taking $v = u^0 - u^{H,RB}$, we can deduce

$$\mathscr{J}(u^0 - u^{H,RB}) \quad = \quad B_0(u^0 - u^{H,RB}, z^{\mathcal{H},\mathcal{RB}}) + B_0(u^0 - u^{H,RB}, z^0 - z^{\mathcal{H},\mathcal{RB}}), \tag{5.2.21}$$

where u^0 is the solution of (5.2.13), z^0 is the solution of (5.2.15), $u^{H,RB}$ is the solution of (5.2.14), and $z^{\mathcal{H},\mathcal{RB}}$ is the solution of (5.2.17). The last two terms on the right-hand side of (5.2.21) are part of the data approximation error. The first term can be further written as

$$B_0(u^0 - u^{H,RB}, z^{\mathcal{H},\mathcal{RB}}) = \int_\Omega f^H z^{\mathcal{H},\mathcal{RB}} dx - \int_\Omega (f^H - f) z^{\mathcal{H},\mathcal{RB}} dx + B_0(u^{H,RB}, z^{\mathcal{H},\mathcal{RB}}). \tag{5.2.22}$$

By the definition of the multiscale flux and integration by parts, one can deduce that

$$\begin{aligned} B_0(u^{H,RB}, z^{\mathcal{H},\mathcal{RB}}) \quad &= \quad -\sum_{K\in\mathscr{T}_H} \int_K \nabla\cdot(\Pi^K_{a\nabla u_N}) z^{\mathcal{H},\mathcal{RB}} dx \\ &+ \quad \frac{1}{2}\sum_{K\in\mathscr{T}_H}\sum_{e\subset\partial K}\int_e [\![\overline{\Pi^K_{a\nabla u_N}}]\!]_e z^{\mathcal{H},\mathcal{RB}} ds \\ &- \quad \sum_{K\in\mathscr{T}_H}\int_K (\Pi^K_{a\nabla u_N} - a^0\nabla u^{H,RB})\cdot\nabla z^{\mathcal{H},\mathcal{RB}} dx. \end{aligned} \tag{5.2.23}$$

Combining (5.2.21) (5.2.22) and (5.2.23) proves the theorem.

The following result follows immediately from Theorem.5.2.5.

Corollary 5.2.6. *The a posteriori upper bound is given by*

$$|\mathscr{J}(u^0 - u^{H,RB})| \le \sum_{K\in\mathscr{T}_H} \bar{\eta}_{H,G}(K) + |\xi_{H,G}(K)|.$$

According to the expression (5.2.20), whether $\sum_{K\in\mathcal{T}_H}\eta_{H,G}(K)$ can provide a good approximation to the exact error in the quantity of interest $\mathscr{J}(e^H)$ depends on the quality of the data approximation error. The following theorem gives an upper bound for the data approximation error.

Theorem 5.2.7. *Assume that the triangulation is regular, that the homogenization tensor $a^0(x)$ is smooth enough and that assumption* **(H1)** *and* (1.2.19) *hold. Assume the condition (***Q2***) holds with $\sigma = 2\ell - 2$ (for primal solution) and $\sigma = 2\hat{\ell} - 2$ (for dual solution). In addition, assume that* (5.1.5) *holds for the QF for the primal solution and that the multiscale tensor $a_{ij}(\cdot, y)$ is locally constant for each $K \in \mathcal{T}_H$. Then we have for $\xi_{H,G}(\Omega) = \sum_{K\in\mathcal{T}_H}\xi_{H,G}(K)$ the following upper bound*

$$|\xi_{H,G}(\Omega)| \quad \leq \quad C\big(H^{\ell+\hat{\ell}} + C\|f - f^H\|_{\mathscr{L}^2(\Omega)} + (\frac{h}{\varepsilon})^{2q} + r_{MOD} + r_{RB}\big). \tag{5.2.24}$$

Proof. We sketch the proof for this theorem that follows the steps of the proof of [21, Theorem 6]. For the first term of (5.2.19), we apply the decomposition,

$$\begin{aligned} &\int_K (\Pi^K_{a\nabla u_N} - a^0\nabla u^{H,RB})\cdot\nabla z^{\mathcal{H},\mathcal{RB}} dx \\ = \quad &\underbrace{\int_K (\Pi^K_{a\nabla u_N} - \Pi^K_{a^0\nabla u^{H,RB}})\cdot\nabla z^{\mathcal{H},\mathcal{RB}} dx}_{I} + \underbrace{\int_K (\Pi^K_{a^0\nabla u^{H,RB}} - a^0\nabla u^{H,RB})\cdot\nabla z^{\mathcal{H},\mathcal{RB}} dx}_{II}, \end{aligned} \tag{5.2.25}$$

where $\Pi^K_{a^0\nabla u^{H,RB}}(x)$ is the interpolation polynomial of $a^0\nabla u^{H,RB}$ on the element K using the same interpolation points as for $\Pi^K_{a\nabla u_N}$. By [21, Lemma 11], we have

$$|I| \leq C(\frac{h}{\varepsilon})^{2q} + r_{MOD} + r_{RB}. \tag{5.2.26}$$

Next, II vanishes by our assumption on the multiscale tensor.

The second term of the data approximation error defined in (5.2.19) can be estimated as follows,

$$\begin{aligned} &\sum_{K\in\mathcal{T}_H} |B_{0,K}(u^0 - u^{H,RB}, z^0 - z^{\mathcal{H},\mathcal{RB}})| \\ &\qquad \leq C(H^{\ell} + (\frac{h}{\varepsilon})^{2q} + r_{MOD} + r_{RB})(H^{\hat{\ell}} + (\frac{h}{\varepsilon})^{2q} + r_{MOD} + r_{RB}). \end{aligned} \tag{5.2.27}$$

As noticed in Remark 5.2.1, the dual problem has the same micro and modeling error $(\frac{h}{\varepsilon})^{2q} + r_{MOD} + r_{RB}$ as the primal problem.

The third terms of the data approximation can be bounded as

$$\sum_{K\in\mathcal{T}_H} |\int_K (f^H - f) z^{\mathcal{H},\mathcal{RB}} dx| \leq C\|f^H - f\|_{\mathscr{L}^2(\Omega)}. \tag{5.2.28}$$

Combining (5.2.26),(5.2.27) and (5.2.28) gives the theorem. □

Remark 5.2.8. *If the FE space for the primal problem is $S_0^1(\Omega,\mathcal{T}_H)$, the assumption on the multiscale tensor in Theorem 5.2.7 that $a_{ij}(\cdot,y)|_K$ is constant can be removed. In turn, the term II is no longer zero but can be bounded by CH^1 (see [21]).*

Observe from Theorem 5.2.7 that if r_{RB} is small (e.g., a fast decaying Kolmogorov N-width holds), then the data approximation error for the DWR RB-FE-HMM is smaller than the corresponding data approximation error for the FE-HMM. Indeed, the bound for the micro error $r_{MIC} \leq C(\frac{h}{\varepsilon})^{2q} = \mathcal{O}(\mathcal{N}^{-\frac{2q}{d}})$ for the RB-FE-HMM, depends on the fine mesh used in the offline stage that is usually smaller than the micro mesh used for the FE-HMM (where a mesh proportional to the macro-mesh is used for efficiency, see the complexity discussion in Section 1.2.4).

We next summarize our algorithm.

Algorithm 5.2.9. *(DWR RB-FE-HMM)*

- *Offline stage: Construct the RB space $S_N(Y)$ following Algorithm 4.2.4 and store the output* (4.2.27).
- *Online stage:*

1.Solve *On the current macro mesh, first compute the macro solution $u^{H,RB}$ and $\frac{1}{|K_{\delta_j}|}\int_{K_{\delta_j}} a_{x_{K_j}}(G_{x_{K_j}}^{-1}(x))\nabla u_{N,K_j}(x)dx$ at the sampling domains K_{δ_j} of each quadrature point. Then compute the dual solution $z^{\mathcal{H},\mathcal{RB}} \in S^{\hat{\ell}}(\Omega,\mathcal{T}_H)$ also on the current macro mesh. Store the data for the next step.*

2.Estimate *Construct the RB multiscale flux $\Pi^K_{a\nabla u_N}$ and compute the RB mutliscale jump $[\![\overline{\Pi^K_{a\nabla u_N}}]\!]$. Further compute the error indicator $\eta_{H,G}(K)$ for all $K \in \mathcal{T}_H$ as an approximation to $\mathcal{J}(e^H)$. If $\sum_{K\in\mathcal{T}_H}|\eta_{H,G}(K)| < tol$, the process ends otherwise moves to next step, where tol is given as a stopping criterion.*

3.Mark *Mark the macro elements based on $\bar{\eta}_{H,G}(K)$ by a given marking strategy (e.g. maximum marking strategy in [94, 100]).*

4.Refine *Refine the marked the elements e.g., by applying the newest vertex refinement strategy (briefly explained in Section 5.3) which keeps the conformity of the refined mesh. Go back to Step 1 (solve).*

Remark 5.2.10. *For the DWR FE-HMM, the introduction of the dual problem in each element and its solution using higher order FEM leads to high computational cost (the number of micro problems increases due to the need of using a higher order QF likewise the DOF increase in order to match the accuracy of the increasingly refined macro FE mesh). With the proposed RB strategy, the computation of the micro problems is decoupled from the macro adaptive procedure. Accurate RB functions for the micro problems are computed once and can be used for the primal and dual FE-HMM problems while macro meshes are refined. This brings huge computational saving as can be seen in the numerical experiments.*

5.3 Numerical Experiments

In this section, we present numerical experiments for the energy norm based adaptive RB-FE-HMM for a 2-D problem on a crack domain in Section 5.3.1 and a 3-D problem on an L-shape domain in Section 5.3.2. We also present numerical experiments for the DWR RB-FE-HMM for two types of quantities of interest in Section 5.3.3. In the whole section, we consider the model problem defined in (4.0.1), with possibly non-homogeneous boundary conditions,

$$\begin{aligned} -\nabla\cdot(a^\varepsilon(x)\nabla u^\varepsilon(x)) &= f \quad \text{in } \Omega, \\ u^\varepsilon(x) &= g_d \quad \text{on } \partial\Omega, \end{aligned} \tag{5.3.29}$$

where $\varepsilon = 4\times 10^{-5}$. We choose for Section 5.3.1 and Section 5.3.3 tensors $a^\varepsilon = a(x,y)$ Y-periodic with respect to y for which explicit expressions of the homogenized tensors are available in order to be able to accurately quantify the errors. We emphasize that the algorithms proposed in this chapter are valid for problems with general tensors. All the numerical simulations are implemented in Matlab R2010b. The core code for the FE-HMM follows [19]. The refinement strategy and FEM are partly based on AFEM@[49] which we shortly explain in the following.

Newest vertex bisection. In the later numerical experiments, we only consider this refinement strategy. Let $\mathcal{T}_H$ be a shape regular triangulation of Ω. For each element $K\in\mathcal{T}_H$, we assign the "newest vertex" to one of the vertices for the initial setting and call the opposing edge as the refinement edge. Based on this setting, the newest vertex bisection can be implemented as follows.
1. Bisect an element K into two children elements by connecting the newest vertex to the midpoint of the refinement edge.
2. Assign the the midpoint of the refinement edge of the parent element as the newest vertex for both children elements and their refinement edges correspondingly.

In order to keep the conformity of $\mathcal{T}_H$, some of the neighbouring elements need to be bisected besides the marked elements in the adaptivity steps. We note that the nodes of the newly generated mesh are nested with the old ones and the triangulation keeps shape regular and conforming.

Numerical evaluation of the errors. Let u^H be the numerical solution and u^{ref} be a reference solution (for the effective problem (1.1.3)) computed on a fine triangulation $\mathcal{T}_h$. The error $u^{ref}-u^H$ in the $\mathcal{H}^1$ and $\mathcal{L}^2$ norms are estimated by

$$\begin{aligned} e_{\mathcal{L}^2} &:= \|u^{ref}\|^{-1}_{\mathcal{L}^2(\Omega)}\Big(\sum_{K\in\mathcal{T}_h}\sum_{j=1}^{J}\rho_{K_j}|u^H(z_{K_j})-u^{ref}(z_{K_j})|^2\Big)^{1/2}, \\ e_{\mathcal{H}^1} &:= \|u^{ref}\|^{-1}_{\mathcal{H}^1(\Omega)}\Big(\sum_{K\in\mathcal{T}_h}\sum_{j=1}^{J}\rho_{K_j}|\nabla u^H(z_{K_j})-\nabla u^{ref}(z_{K_j})|^2\Big)^{1/2}, \end{aligned}$$

where we will use $\|u^{ref}\|_{\mathcal{H}^1(\Omega)} \sim (\sum_{K\in\mathcal{T}_H} \|\nabla u^{ref}\|^2_{\mathcal{L}^2(K)})^{1/2}$. Here $\{z_{K_j}, \rho_{K_j}\}$ is the quadrature formula on the fine triangulation $\mathcal{T}_h$.

Macro FEM and QF used in the examples. In the following examples, when using P1 triangular (tetrahedral) elements for the macro problems, we choose the barycenter of the element as the quadrature point and the weight $\hat{\omega} = |\hat{K}|$. When we use P2 triangular elements for the macro problems, we choose the Gauss three points quadrature formula with barycentric coordinates $(1/6, 1/6, 2/3)$ and weights $\hat{\omega}_i = |\hat{K}|/3,\ i = 1,2,3$.

Computing environment. We note that all the numerical experiments were made with MATLAB in a serial computing mode on a workstation with 16 CPU processors (8 cores). The RB-FE-HMM could of course benefit from parallel computing environment. Finally, we will report computing time (and comparison with the FE-HMM) for all the 2D experiments for which the linear algebra (using the MATLAB sparse LU solver) is done reasonably well. For 3D, however, the unoptimized linear algebra solver that we used does not allow to give meaningful computing time and we just report the convergence rates.

5.3.1 Energy norm based adaptive RB-FE-HMM applied to crack problem

Let $\Omega \subset \mathbb{R}^2$ be a square domain with a crack (see Fig. 1) and let $f = 1$, $g_d = 0$. The diagonal entries of the multiscale tensor are

$$\begin{aligned} a(x, \frac{x}{\varepsilon})_{11} &= x_1^2 + 0.2 + (x_2 + 1.2)(\sin(2\pi\frac{x_1}{\varepsilon}) + 2), \\ a(x, \frac{x}{\varepsilon})_{22} &= x_2^2 + 0.05 + (x_1 x_2 + 1.5)(\sin(2\pi\frac{x_2}{\varepsilon}) + 2). \end{aligned} \tag{5.3.30}$$

The corresponding homogenized tensor is also diagonal and can be computed as (see [42, 74])

$$\begin{aligned} a^0(x)_{11} &= (\int_0^1 \frac{1}{x_1^2 + 0.2 + (x_2 + 1.2)(\sin(2\pi y_1) + 2)} dy_1)^{-1}, \\ a^0(x)_{22} &= (\int_0^1 \frac{1}{x_2^2 + 0.5 + (x_1 x_2 + 1.5)(\sin(2\pi y_2) + 2)} dy_2)^{-1}. \end{aligned} \tag{5.3.31}$$

Offline stage. We use FEM with piecewise linear basis functions (called P1-FEM) as the offline solver to compute the reduced basis functions and set the offline meshsize to 1600×1600 so that the a priori error bound given in Theorem 4.5.3 reads $r_{MIC} = \mathcal{O}(10^{-7})$. As discussed in [92], the error for the output of interest for the RB method, i.e., the numerical homogenization tensor (4.2.23), can be bounded by the a posteriori error estimator $max_{(T_\delta,\eta)\in\Xi_{RB}} (\Delta^{\eta}_{N,T_\delta})^2$. In our experiment, this error reads $\mathcal{O}(10^{-11})$ for the particular tolerance chosen (see Table 5.1) since $max_{(T_\delta,\eta)\in\Xi_{RB}} (\Delta^{\eta}_{N,T_\delta})^2 \leq tol_{RB} = \mathcal{O}(10^{-11})$. Furthermore we choose sampling domain sizes of the same length as ε and periodic boundary condition on the cell problems so that $r_{MOD} = 0$ (see (1.2.20)). As a result of the offline stage, we obtain 10 RB functions (see Table 5.1 for the

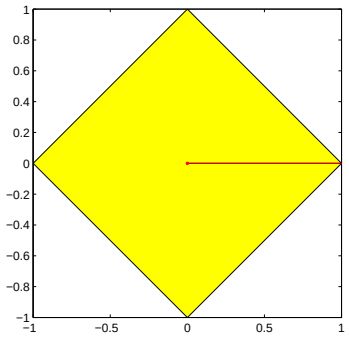


Figure 1: The domain Ω for the crack problem is in yellow. The red line describes the crack.

parameters related to the offline stage).

Table 5.1: Parameters for the RB-FE-HMM offline stage.

Solver	P1-FEM
Mesh	1600×1600
Basis number	10
tol_{RB}	5e-11
CPU time(s)	5100

Adaptive P1 RB-FE-HMM online stage. Now we start the adaptive online procedure. Here we present two tests with different online solvers. In the first test we use P1-FEM as the online solver. The corresponding a posteriori error estimator is defined in (5.1.10). In Fig. 2, we display the macro solution $u^{H,RB}$ after 17 macroscopic iterations of the adaptive RB-FE-HMM.

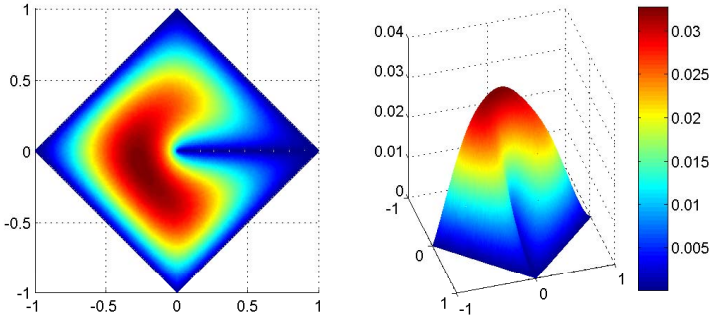


Figure 2: The adaptive RB-FE-HMM solution computed by P1-FEM on the crack domain.

We recall that the a priori error estimates for RB-FE-HMM derived in Section 4.5 read $\|u^{H,RB} - u^0\|_{\mathscr{H}^1(\Omega)} \leq C(H^1 + r_{HMM})$ and $\|u^{H,RB} - u^0\|_{\mathscr{L}^2(\Omega)} \leq C(H^2 + r_{HMM})$, where $u^{H,RB} \in S^1_0(\Omega, \mathscr{T}_H)$. We denote by M_{mac} the macro DOF $M_{mac} = \mathscr{O}(H^{-1/2})$ and replace H with $\frac{1}{\sqrt{M_{mac}}}$ in those estimates

and obtain

$$\|u^{H,RB}-u^0\|_{\mathscr{H}^1(\Omega)} \leq C(\frac{1}{\sqrt{M_{mac}}}+r_{HMM}),\ \ \|u^{H,RB}-u^0\|_{\mathscr{L}^2(\Omega)} \leq C(\frac{1}{M_{mac}}+r_{HMM}).$$

We see in Fig. 3 that when the macro mesh is refined, the $\mathscr{H}^1$ and $\mathscr{L}^2$ error decay with the optimal rates $\mathscr{O}(M_{mac}^{-1/2})$ and $\mathscr{O}(M_{mac}^{-1})$, respectively. The reference solution u^{ref} is computed with an adaptive FEM with piecewise quadratic polynomials called P2-FEM up to 25 iterations with the initial mesh obtained by uniformly refining 2 times the final online mesh (final mesh for $u^{H,RB}$). The number of DOF of u^{ref} is given by 1 757 821. We can also observe in Fig. 3 that the profile of the error indicator is parallel to the $\mathscr{H}^1$ error, which corroborates the estimates of Theorem 5.1.7.

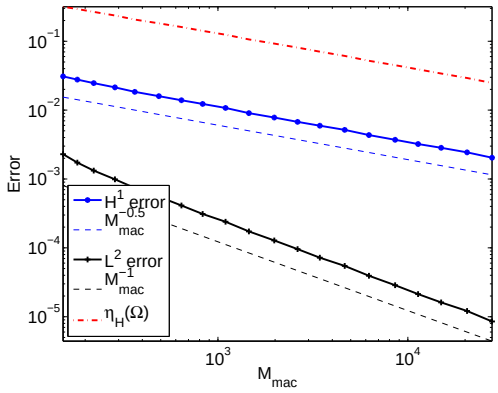


Figure 3: The refinement indicator $\eta_H(\Omega)$ the $\mathscr{H}^1$ and $\mathscr{L}^2$ errors of the macro solution (crack Problem) of the adaptive P1 RB-FE-HMM for 20 iterations.

Adaptive P2 RB-FE-HMM online stage. In this test, we apply the adaptive P2-FEM in the online stage. As we explained in Section 5.1, an interpolation polynomial $\Pi^K_{a\nabla u_N}$ (see (5.1.3)) is introduced to define the high order a posteriori error estimator. In this test, $\Pi^K_{a\nabla u_N} \in \mathscr{P}^1(K)$ is interpolated on three quadrature points of the simplicial element. For the P2 RB-FE-HMM, the a priori error estimate yields $\|u^{ref}-u^{H,RB}\|_{\mathscr{H}^1(\Omega)}$ proportional to $\mathscr{O}(M_{mac}^{-1})$ and $\|u^{ref}-u^{H,RB}\|_{\mathscr{L}^2(\Omega)}$ proportional to $\mathscr{O}(M_{mac}^{-3/2})$. We can observe in Fig. 4 that the errors for the adaptive P2 RB-FE-HMM indeed decay as $\mathscr{O}(M_{mac}^{-1})$ and $\mathscr{O}(M_{mac}^{-3/2})$ in the $\mathscr{H}^1$ and $\mathscr{L}^2$ norms, respectively. We can also observe for the adaptive P2 RB-FE-HMM, that the decay of the error indicator has the optimal $\mathscr{H}^1$ error decay. The reference solution u^{ref} is also computed by an adaptive P2-FEM similarly as described in the first test (but with 20 iterations). The number of DOF is 9 721 276.

Comparisons. Here we first compare the efficiency and accuracy of the adaptive RB-FE-HMM with P1-FEM and P2-FEM as macro solvers. We then compare the performance of the adaptive RB-FE-HMM with the adaptive FE-HMM.

- **Adaptive P1 RB-FE-HMM and adaptive P2 RB-FE-HMM.** We first present the online mesh refinements in Fig. 5. The refinements of both P1 and P2 RB-FE-HMM can detect the crack

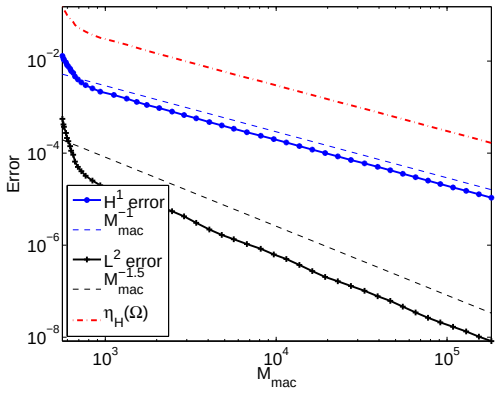
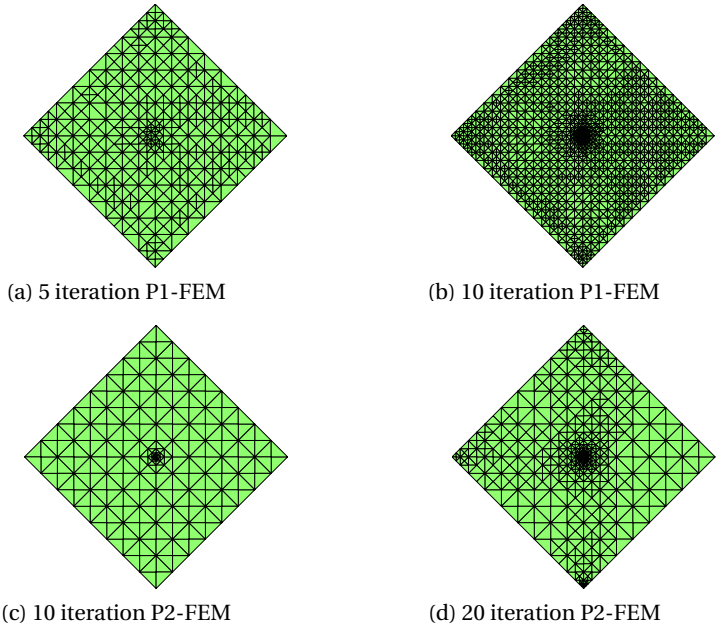


Figure 4: The refinement indicator $\eta_H(\Omega)$, the $\mathscr{H}^1$ and $\mathscr{L}^2$ errors of the macro solution (crack Problem) of the adaptive P2 RB-FE-HMM for 46 iterations. Online CPU time 276s.



(a) 5 iteration P1-FEM

(b) 10 iteration P1-FEM

(c) 10 iteration P2-FEM

(d) 20 iteration P2-FEM

Figure 5: Online refinement. (a) $M_{mac} = 365$. (b) $M_{mac} = 1456$. (c) $M_{mac} = 180$. (d) $M_{mac} = 516$.

and the P2 refinement is more concentrated at the vicinity of the singularities than the P1 refinement. This higher order macro FEM provides better convergence rates away from the singularities and can therefore use coarser meshes. Now we compare the performance

Table 5.2: Comparison of the adaptive P1 RB-HMM-FEM and P2 RB-HMM-FEM.

	P1 RB-FE-HMM	P2 RB-FE-HMM
# iter	20	19
$e_{\mathscr{H}^1}$	0.0020	0.0011
$e_{\mathscr{L}^2}$	8.5412e-06	8.1887e-06
M_{mac}	27725	1750
Online CPU time	25.18 s	3.92 s

of the two online solvers in Table 5.2. We can see that for similar errors, the number of the macro DOF M_{mac} for the P2 RB-HMM-FEM is only 6.3% of the M_{mac} for the P1 RB-HMM-FEM and the time cost for the P2 RB-HMM-FEM is 15.6% of the CPU time for the P1 RB-HMM-FEM. This shows the potential advantage of using higher order macro solver for adaptivity.

- **Adaptive P1 RB-FE-HMM and adaptive P1 FE-HMM.** In this numerical experiment, we compare the accuracy and time cost of the adaptive RB-FE-HMM and the adaptive FE-HMM. In Table 5.3 we presents the $\mathscr{H}^1$ error of both methods for each iteration and the corresponding effectivity index which is defined as Eff $:= \frac{\eta_H}{e_{\mathscr{H}^1}}$. For this test, the macro error dominates so that in Table 5.3 the RB-FE-HMM and FE-HMM give almost the same error $e_{\mathscr{H}^1}$ and the same effectivity index Eff as the standard adaptive FEM. However, as we see in Table 5.4, the CPU time comparison immediately shows the advantage of using the adaptive RB-FE-HMM which yields the same accuracy as the adaptive FE-HMM with only 0.14% of its computing time. Even if taking account the offline overhead, the adaptive RB-FE-HMM takes only about 2.8% of the time used for the adaptive FE-HMM.

Control of the offline parameters. We next perform the offline algorithm with a 200×200 offline mesh instead of the 1600×1600 used before. We nevertheless keep the offline tolerance as $tol = 5e{-}11$ (similarly as for Table.5.1). The a priori error estimate for the RB functions computed with this mesh indicates an error of $r_{MIC} = O(10^{-5})$. We note that the algorithm nevertheless terminates when the a posteriori estimator for r_{RB} indicates an error of $O(10^{-10})$. We note that this quantity is controlled by the a posteriori error estimator in the offline stage according to the tolerance set by the user. We note that a failure to meet the offline tolerance could indicate a poor decay of the Kolmogorov N-width originating for example from a lack of self-similarity of the microscopic data.

We then apply the P2 adaptive RB-FE-HMM for the crack problem as the online solver using the new offline outputs. As can be seen in Fig. 6, the $\mathscr{L}^2$ error becomes constant after a certain number of iteration steps indicating that the error r_{MIC} dominates r_{MAC}. This experiment also illustrates that by monitoring the online macroscopic error obtained by using different sets of RB with different accuracy can be used to assess the r_{MIC} error in an actual computational

Table 5.3: Effectivity index and $\mathscr{H}^1$ error for the adaptive RB-FE-HMM, FE-HMM and FEM for the crack problem.

		RB-FE-HMM		FE-HMM		FEM
Iter no.	M_{mac}	$e_{\mathscr{H}^1}$	Eff	$e_{\mathscr{H}^1}$	Eff	Eff
1	153	0.0309	10.2577	0.0309	10.2588	10.2577
2	182	0.0278	10.5756	0.0278	10.5766	10.5756
3	222	0.0247	10.8022	0.0247	10.8031	10.8022
4	287	0.0214	11.0437	0.0214	11.0444	11.0437
5	365	0.0184	11.2158	0.0184	11.2162	11.2158
6	489	0.0159	11.4192	0.0159	11.4196	11.4192
7	643	0.0139	11.5564	0.0139	11.5567	11.5564
8	831	0.0123	11.5224	0.0123	11.5526	11.5224
9	1097	0.0107	11.5932	0.0107	11.5934	11.5932
10	1456	0.0090	11.7484	0.0090	11.7485	11.7484
11	1981	0.0078	11.8570	0.0078	11.8571	11.8570
12	2626	0.0068	11.9613	0.0068	11.9614	11.9613
13	3438	0.0059	11.8667	0.0059	11.8668	11.8667
14	4648	0.0052	11.8540	0.0052	11.8540	11.8540
15	6237	0.0043	11.9844	0.0043	11.9844	11.9844
16	8566	0.0037	12.1085	0.0037	12.1085	12.1085
17	11320	0.0032	12.1627	0.0032	12.1628	12.1627
18	14960	0.0028	11.9973	0.0028	11.9973	11.9973
19	20503	0.0024	11.9997	0.0024	11.9997	11.9997
20	27725	0.0020	12.1779	0.0020	12.1779	12.1779

Table 5.4: CPU time comparison between the adaptive P1 RB-FE-HMM and the adaptive P1 FE-HMM for crack problem.

	RB-FE-HMM	FE-HMM	$t_{RB}/t_{FE-HMM}(\%)$
Total online CPU time for 20 iterations	26.13 s	186110 s	0.14%
Total offline-online CPU time	5126.13 s	186110 s	2.8%

procedure.

5.3.2 The energy norm based adaptive RB-FE-HMM applied to a 3-D problem on an L-shape domain

In this example we investigate the performance of the adaptive RB-FE-HMM on a three dimensional problem. We consider an exponential stationary model for the infiltration of a fluid in unsaturated porous media. We choose here an L-shape computational domain (similar test problems have been considered in [50, 61] for regular domains). The multiscale tensor is defined as

$$a^{\varepsilon}(x)_{ii} = 10\alpha^{\varepsilon}(x)e^{\beta^{\varepsilon}(x)} + x_1^2 + (x_2 - x_3)^2 + 0.5 \;\; i = 1,2,3, \tag{5.3.32}$$

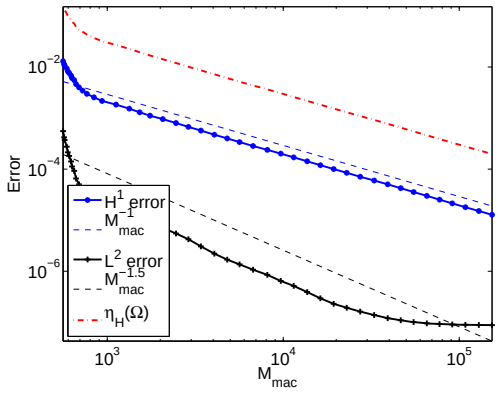
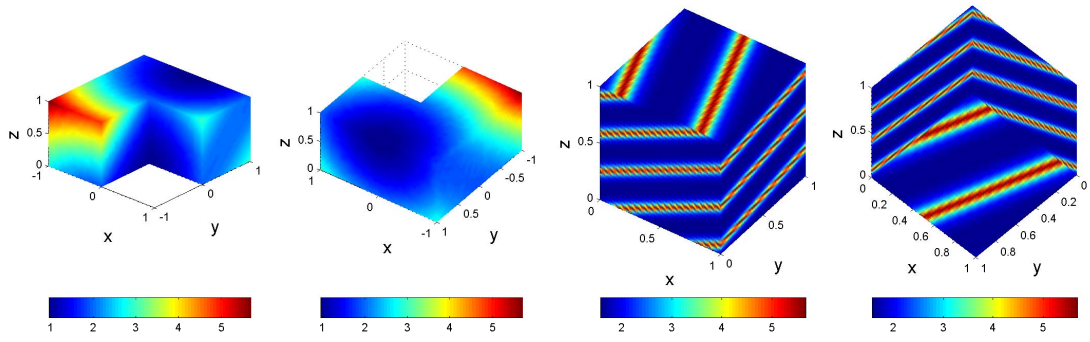


Figure 6: Refinement indicator $\eta_H(\Omega)$, $\mathscr{H}^1$ and $\mathscr{L}^2$ errors of the macro solution (crack Problem) for the adaptive P2 RB-FE-HMM for 45 iterations with inaccurate offline outputs.

where

$$\begin{aligned}\beta^\varepsilon(x) &= -\alpha^\varepsilon(x)\big((x_1-c_1\mu_1(x))^2+(x_2-c_1\mu_2(x))^2+(x_3-c_1\mu_3(x))^2\big)\cdot\\ &\quad\big((x_1-0.8-c_2\mu_1(x))^2+(x_2-0.5-c_2\mu_2(x))^2+(x_3-0.7-c_2\mu_3(x))^2\big),\\ \alpha^\varepsilon(x) &= \frac{1}{20+18\sin(6\pi\frac{x_3}{\varepsilon}-\pi(3\frac{x_1}{\varepsilon}+\frac{x_2}{\varepsilon}))},\end{aligned}$$

and where $\mu(x)=(\mu_1(x),\mu_2(x),\mu_3(x))$ is a uniform random mapping from Ω to $[0,1]^3$ and c_1,c_2 are constant parameters chosen to be $c_1=0.2,\ c_2=0.1$. We set $f=10$ and $g_d=0.5$. In general infiltration models, a^ε represents the microscopic permeability of the porous media. As shown



(a) Permeability $a^\varepsilon(x)$ on Ω (b) $a_{x^*}(y)$ on the reference sampling domain

Figure 7: (a) Permeability $a^\varepsilon(x)$ on the L-shape domain Ω. (b) $a_{x^*}(y)$ on the reference sample domain $Y=[0,1]^3$ with fixed macro variable $x^*=(0.750,0.625,0.125)$ and $\mu(x^*)=(0.799,0.499,0.643)$.

in Fig. 7 (a), we can observe that the media has lowest permeability around points $(0,0,0)$ and $(0.8,0.5,0.7)$. When we fix the macro variable to be x^* and map $a^\varepsilon(x)$ to the reference sampling

domain $Y = [0,1]^3$ by the mapping G_{x^*} defined in Section 4.2, we obtain the parametrized tensor $a_{x^*}(y)$ with respect to parameter x^* which displays locally periodic pattern shown in Fig. 7 (b). Since the tensor $a^\varepsilon(x)$ in (5.3.32) is not in an affine form, we need to apply the empirical

Table 5.5: Parameters for the RB-FE-HMM offline stage (3D problem).

Reference sampling domain Y	$[0,1]^3$
Mesh	200^3
tol_{EIM}	1e-5
tol_{RB}	3e-7
Affine terms	11
Basis number	21

interpolation method (see [7]) to obtain the affine approximation of the tensor in the offline stage. The offline parameters and output information are presented in Table 5.5. After the offline stage 21 basis functions are obtained. Constrained by the computation environment, the finest offline mesh generated is 200^3. Considering the macro meshes we test in the online stage (the macroscopic DOF M_{mac} varies in the range of 325 to 63749 $\approx 40^3$), the offline mesh is sufficiently fine according to the estimates of Theorem 4.5.3.

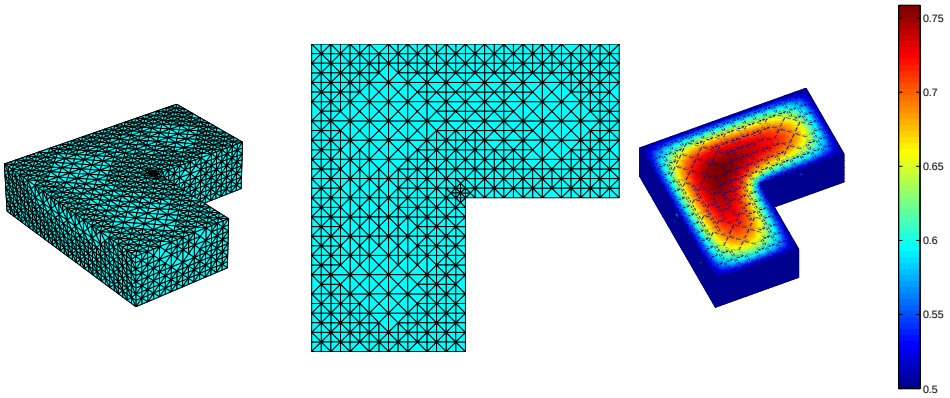
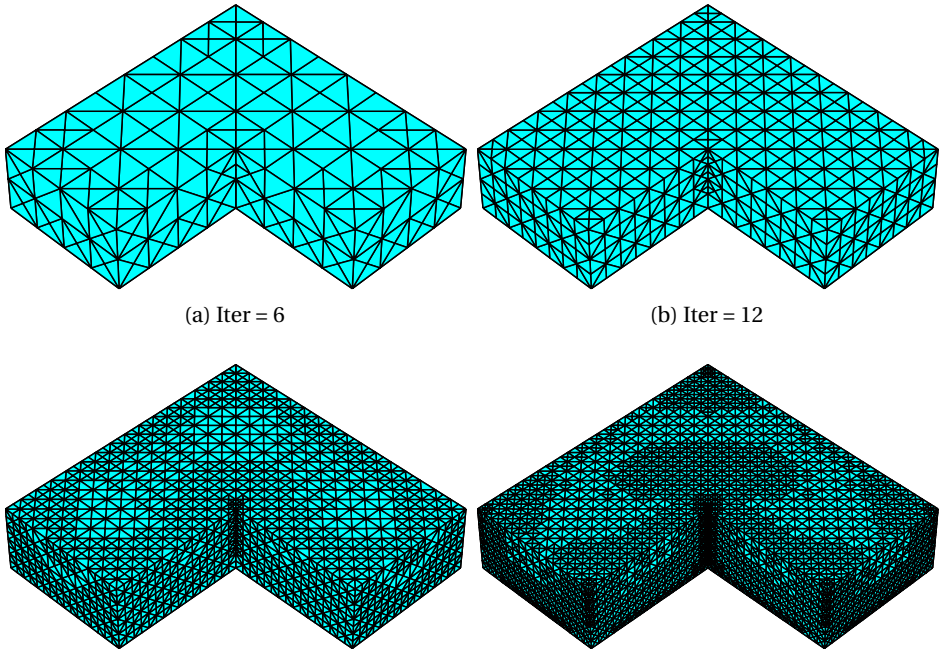


Figure 8: The macro mesh and $u^{H,RB}$ after 20 iterations.

In the online stage, we set the initial macro mesh to be uniform tetrahedron with $M_{mac} = 325$. We perform 24 iterations of the adaptive RB-FE-HMM. The solution $u^{H,RB}$ of the 20th iteration and corresponding macro meshes on the surface of the segment obtained by a cut through the plan $z = 0.5$ are displayed in Fig. 8. As shown in Fig. 9, the mesh refinement mostly takes place in the vicinity of the singularities (the corners of the domain) which also illustrates that the refinement indicators provide effective information for the macro mesh refinement. The error indicator $\eta_H(\Omega)$ for 24 iterations of the adaptive RB-FE-HMM is shown in Fig.10. We observe that the error indicator decays with an optimal convergence rate $\mathcal{O}(M_{mac}^{-1/3})$. This again corroborates



(a) Iter = 6 (b) Iter = 12

(c) Iter = 18 (d) Iter = 24

Figure 9: The macro mesh refinements.

the estimates of Theorem 5.1.6 and Theorem 5.1.7.

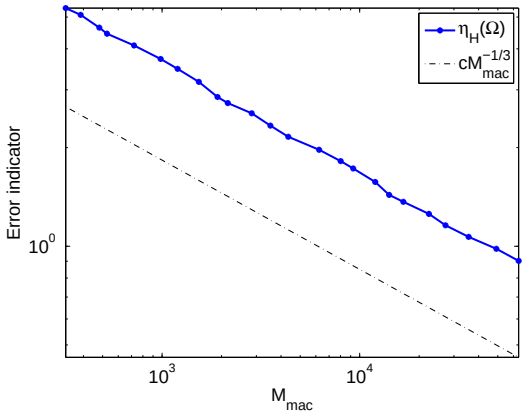


Figure 10: The a posteriori error decay for adaptivity on the 3-D L-shape domain for 24 iterations.

5.3.3 DWR RB-FE-HMM.

In this subsection, we present numerical experiment for goal oriented adaptive computations with the DWR RB-FE-HMM. We consider the model equation (5.3.29) with $f = 1000$, a domain

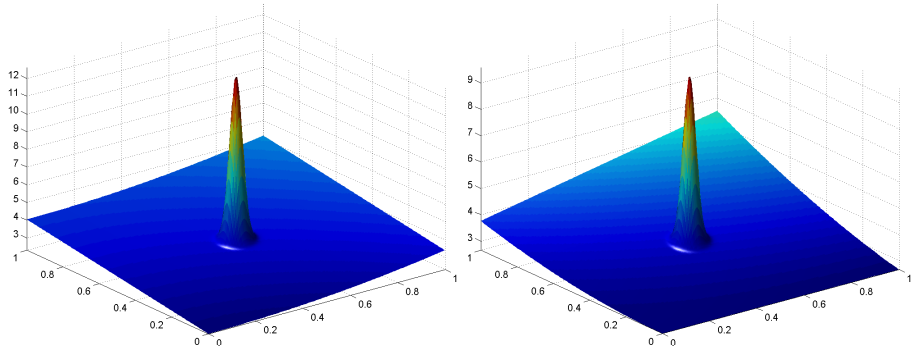
$\Omega = [0,1]^2$ and the diagonal multiscale tensor given by

$$
\begin{aligned}
a(x, \frac{x}{\varepsilon})_{11} &= (x_1^2 + 0.2)E(x) + (x_2 + 1.2)(\sin(2\pi\frac{x_1}{\varepsilon}) + 2), \\
a(x, \frac{x}{\varepsilon})_{22} &= (x_2^2 + 0.05)E(x) + (x_1 x_2 + 1.5)(\sin(2\pi\frac{x_2}{\varepsilon}) + 2),
\end{aligned} \tag{5.3.33}
$$

where

$$E(x) = 1 + 20e^{-1000((x_1-0.5)^2+(x_2-0.5)^2)}.$$

The corresponding homogenized tensor is



(a) $a^0(x)_{11}$ (b) $a^0(x)_{22}$

Figure 11: The homogenized tensor (5.3.34).

$$
\begin{aligned}
a^0(x)_{11} &= (\int_0^1 \frac{1}{(x_1^2 + 0.2)E + (x_2 + 1.2)(\sin(2\pi y_1) + 2)} dy_1)^{-1}, \\
a^0(x)_{22} &= (\int_0^1 \frac{1}{(x_2^2 + 0.5)E + (x_1 x_2 + 1.5)(\sin(2\pi y_2) + 2)} dy_2)^{-1},
\end{aligned} \tag{5.3.34}
$$

which is shown in Fig. 11.

The offline parameters for this example are presented in Table 5.6. For the online stage, we propose two different types of quantities of interest respectively.

Table 5.6: Parameters for the DWR RB-FE-HMM offline stage

Solver	P1-FEM
Mesh	1600×1600
Basis number	10
tol_{RB}	5e-11
CPU time(s)	6772

Quantity of interest 1. We use a local average of the macro solution as the quantity of interest

defined as

$$\mathscr{J}(u^0) = \frac{1}{|\Omega_s|}\int_{\Omega_s} u^0 dx, \tag{5.3.35}$$

where $\Omega_s \subset \Omega$ is the domain of interest. We choose $\Omega_s = [0.25, 0.5]^2$ in this test. We show in

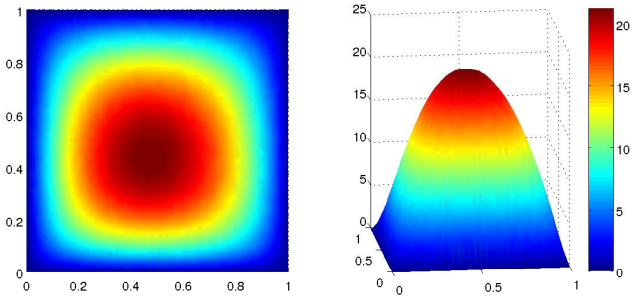
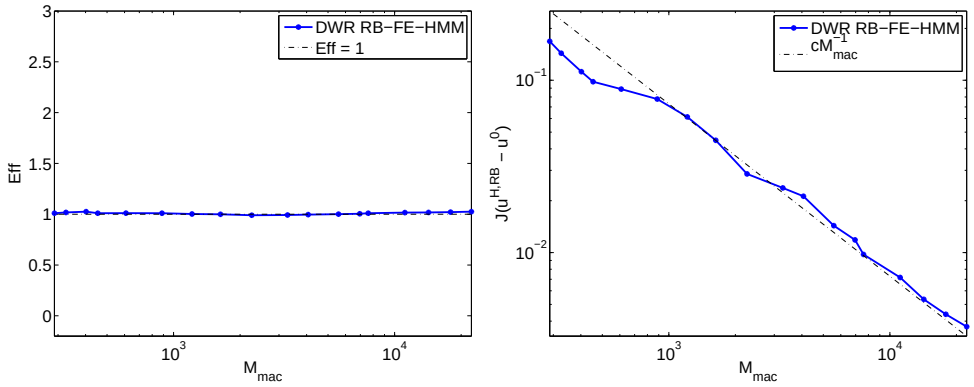


Figure 12: $u^{H,RB}$ after 11 iterations.

Fig. 12 the solution $u^{H,RB}$ after 10 iterations with 3279 DOF. In Fig.13 (a), we observe that the $\text{Eff} := \frac{|\eta_H(\Omega)|}{\mathscr{J}(u^{H,RB}-u^{ref})} \approx 1$ where $\eta_H(\Omega) = \sum_K \eta_H(K)$, which illustrates that the refinement indicator can accurately estimate the error in the quantity of interest and that the data approximation error ξ_H is much smaller than the refinement indicator. In Fig. 13 (b) we observe that the error in quantity of interest converges with a rate of $\mathscr{O}(M_{mac}^{-1})$.



(a) Effectivity index

(b) Error in quantity of interest

Figure 13: The effectivity index and error of the DWR RB-FE-HMM for the quantity of interest (5.3.35)

Now we compare the performance of the DWR RB-FE-HMM and DWR FE-HMM. The comparison starts with an initial non-uniform mesh with 289 DOF. For the FE-HMM, we apply the optimal $\mathscr{H}^1$ refinement to the micro problems that is, $h_K = \sqrt{H_K}$ for the P1 micro FEM as the solver for primal problem (5.2.14) and $h_K = H_K^{2/3}$ for the P2 micro FEM for the dual problem (5.2.17).

According to Fig. 13 (a), $|\eta_{H,G}(\Omega)|$ gives a good estimate of the error in the quantity of interest. We can observe in Table 5.7 that the two methods yield similar refinements and accuracy while Table 5.8 shows that the total cost for the DWR RB-FE-HMM is only 2.1% of the time for the DWR FE-HMM. We emphasize that while the iteration 15 is obtained with the DWR RB-FE-HMM in about 3 hours on our workstation, the iteration 15 with the FE-HMM costs about 4 days on the same workstation !

Table 5.7: Refinement indicators of the DWR RB-FE-HMM and the DWR FE-HMM (quantity of interest 1).

	DWR RB-FE-HMM		DWR FE-HMM	
Iter no.	M_{mac}	$\|\eta_{H,G}(\Omega)\|$	M_{mac}	$\|\eta_{H,G}(\Omega)\|$
1	289	0.1698	289	0.1698
2	326	0.1460	326	0.1438
3	402	0.1150	424	0.1065
4	454	0.0991	598	0.0917
5	609	0.0898	764	0.0856
6	886	0.0785	1163	0.0671
7	1212	0.0614	1447	0.0528
8	1630	0.0448	2228	0.0297
9	2259	0.0283	2949	0.0248
10	3279	0.0235	4370	0.0193
11	4066	0.0211	5731	0.0147
12	5574	0.0143	7006	0.0120
13	6957	0.0119	8430	0.0091
14	7598	0.0098	10513	0.0071
15	11133	0.0073	14671	0.0050

Table 5.8: CPU time comparison between DWR RB-FE-HMM and DWR FE-HMM (quantity of interest 1).

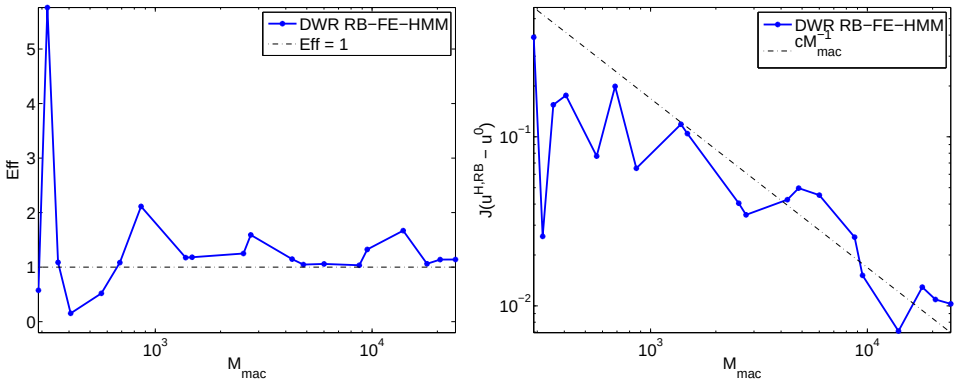
	RB-FE-HMM	FE-HMM	$t_{RB}/t_{FE-HMM}(\%)$
M_{mac} of last iteration	17921	14671	
Total online CPU time	325 s	343130 s	0.1%
Total offline-online CPU time	7097 s	343130 s	2.1%

Quantity of interest 2. In this test we consider a quantity of interest given by a (regularized) pointwise directional derivative (in the direction of the unit vector $\bar{s}$) at $x^* \in \Omega$

$$\mathscr{J}\left(u^0\right) = \frac{1}{|S^r|}\int_{S^r} \nabla u^0 \cdot \bar{s}\, dx, \tag{5.3.36}$$

where S^r is a ball of (small) radius r around the point $x^\star$. We choose $x^* = (0.75, 0.75)$ and $\bar{s} = (\sqrt{2}/2, \sqrt{2}/2)$. We observe in Fig.14 (a) that the effectivity index varies between 0.15 and 5.8 and when $M_{mac} > 1200$ the effectivity index only varies in a small range from 1 to 1.7 which is close to the optimal value 1. In Fig. 14 (b), we sketch the rate of the error $\mathscr{J}(u^{H,RB} - u^0)$ which

approximately decays with an order $\mathcal{O}(M_{mac}^{-1})$ but with large oscillations when $M_{mac} \leq 1200$. The factor which causes the initial oscillations in Fig. 14 might be the strong cancellation effect of the local error indicator $\eta_{H,G}(K)$ for coarse macro meshes.



(a) Effectivity index

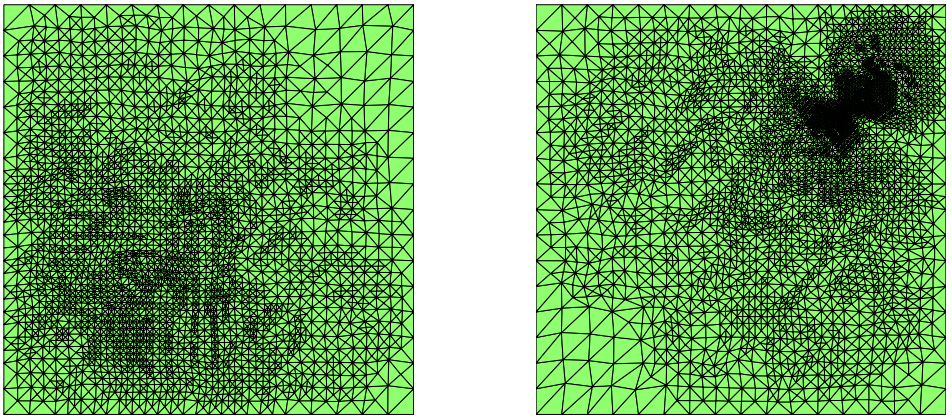
(b) Quantity of interest

Figure 14: The effectivity index and the error in the quantity of interest for the DWR RB-FE-HMM for 20 iterations (quantity of interest of type 2).

In Fig. 15, we plot the mesh refinement for the two different quantities of interest. We use an initial non-uniform mesh for the tests of the two quantities of interest with meshsize 17×17. After several iterations, the two online meshes show different patterns according to the specific quantity of interest. We can see that in Fig. 15 (a), that the mesh is mostly refined in the target domain Ω_s where the quantity of interest is defined while in Fig. 15 (b), the mesh is more refined around the target point $(0.75, 0.75)$ along the direction $(\sqrt{2}/2, \sqrt{2}/2)$ where the directional derivative is defined. For both quantities of interest, the macro mesh refinements also take place in the vicinity of the singularity where the homogenized tensor has a big jump (i.e., in the center of the domain Ω see Fig. 11).

5.4 Discussion

We have presented an efficient adaptive FEM, the RB-FE-HMM, for elliptic homogenization problems based on micro-macro solvers combined with a RB strategy. We have shown that repeated FEM computations of micro problems are avoided during the macro mesh refinement, in contrast to the adaptive FE-HMM, as the micro solutions are computed in a finite dimensional space spanned by a small number of accurately computed representative micro solutions (the reduced basis) obtained by a greedy algorithm in an offline stage. This methodology allows to bypass the derivation of micro a posterior error estimates to calibrate the micro mesh during the macro mesh adaptive cycles. We have presented an a posteriori error analysis for a residual based adaptive RB-FE-HMM for arbitrary order of the macro FE space. Error estimation for the RB-FE-HMM in quantities of interest has also been presented. The efficiency and the sharpness of the derived error bounds have been illustrated by several numerical examples in two and three



(a) Mesh refinement of quantity of interest 1 after 10 iterations, $M_{mac} = 3279$

(b) Mesh refinement of quantity of interest 2 after 12 iterations, $M_{mac} = 4274$

Figure 15: The macro meshes for the two quantities of interest.

dimensions. These examples show the significant advantage of the adaptive RB-FE-HMM over the adaptive FE-HMM.

We close this discussion by two remarks. First as explained in the beginning of this chapter, the RB-FE-HMM would be an excellent solver to be used for goal oriented multiscale computation as discussed in [107, 89, 87, 79, 88] to provide an accurate homogenized solution that could be refined locally to approximate the fine scale solution. The generalization and the computational efficiency of the RB-FE-HMM for higher order macro FEs make this method also a good candidate for an h-p implementation, that remains up to now challenging for a numerical homogenization method.

Part III

Reduced basis finite element heterogeneous multiscale method for nonlinear problems

Quasilinear elliptic equations enter the modeling of numerous problems such as phase transitions, flow in porous media, or reaction and diffusion in electrolysis to mention a few examples [32]. Numerical approximations of such problems have been analyzed by many authors. We mention the works of Douglas and Dupont [73], and Nitsche [83], where the first a priori error analysis was given for the finite element method (FEM). Much recently and relevant for the present work, we mention the analysis obtained in [25] for a FEM with numerical quadrature, i.e., when the continuous variational form originating from the nonlinear problem is approximated by a quadrature formula. Here we are interested in quasilinear elliptic problems with highly oscillatory data of the form

$$-\nabla\cdot\left(a^\varepsilon(x,u^\varepsilon(x))\nabla u^\varepsilon(x)\right)=f(x)\ \text{ in }\Omega, \tag{6.0.1}$$

in a domain $\Omega\subset\mathbb{R}^d$, $d\leq 3$, where $a^\varepsilon(x,u)=(a^\varepsilon_{mn}(u,s))_{1\leq m,n\leq d}$ is a $d\times d$ tensor, associated to $\varepsilon>0$, a sequence of positive real numbers going to zero and $f\in\mathscr{H}^{-1}(\Omega)$. For simplicity we assume homogeneous Dirichlet boundary conditions $u^\varepsilon=0$ on $\partial\Omega$ but we emphasize that more general boundary conditions could be considered.

Such problems arise for example in infiltration of water in an unsaturated porous media modeled by the (stationary) Richards equation [40] or (stationary) heat conduction in a composite material [75]. For efficient numerical computations, an appropriate upscaling of Equation (6.0.1) is needed. Such coarse graining procedures are rigorously described by the mathematical homogenization theory [42, 74] and are studied for the class of problems (6.0.1) in [33, 44, 66]. These analyses show that the solution u^ε of (6.0.1) converges in a weak sense to u^0 as $\varepsilon\to 0$, where the homogenized function u^0 is the solution of an effective (homogenized) equation that is of the same quasilinear type as the original equation with an effective homogenized tensor $a^0(x,u^0(x))$ that depends nonlinearly on u^0. Numerical homogenization methods for problems of the type (6.0.1) are derived in [51] for the multiscale finite element method (MsFEM) and in [60, 26] Here we focus on the FE-HMM proposed in [60, 26] for quasilinear problems. The practical implementation relies on a Newton method for the macroscopic nonlinear FEM. Since the value of the corresponding macroscopic solution is updated at each Newton iteration, the microscopic problems in each element of the macroscopic mesh need to be recomputed. Although the micro problems can be solved independently in parallel, the cost of the procedure mentioned above can be prohibitive, especially for high dimensional problems. In this part, we show how the use of the reduced basis (RB) method (see [93, 92, 96] and references therein) for computing the micro problems permits to considerably improve the efficiency of the standard nonlinear FE-HMM.

Outline of Part III.

- Chapter 6: Presents the RB-FE-HMM for quasilinear problems as well as the fully discrete error analysis and the convergence analysis of the Newton method.
- Chapter 7: Explains the details of the implementation and provides extensive numerical examples.

6 The RB-FE-HMM for quasilinear problems

The algorithm proposed in this chapter for nonlinear problems relies also on the online and offline procedures. However, there are substantial differences from Part II, as here the micro problems are parametrized by both the location of the cell problems in the domain Ω and the macroscopic solution at this location. The greedy algorithm allows to choose an appropriate basis of micro functions (computed with high accuracy) for selected values of the parameters. For the online stage, a Newton method for the RB-FE-HMM implementation is proposed with microscopic solutions computed in the reduced basis space, which amounts to solve small dimensional linear systems in each element of the macroscopic mesh. The overall computational cost of the online macroscopic Newton method is similar to the cost of single scale nonlinear problems. One difficulty is the design of an a posteriori error estimator in the offline stage that is both efficient and also guarantees that the online Newton method converges. We propose in this chapter a new a posteriori error estimators and prove the convergence of the online Newton method and the uniqueness of the numerical solution. Furthermore, a fully discrete error analysis of the quasilinear RB-FE-HMM is derived.

This chapter is taken from [11] and organized as follows. In Section 6.1, we briefly recall the framework of homogenization theory in our context of quasilinear elliptic problems of non-monotone type. We then present in Section 6.2 the new nonlinear RB-FE-HMM with its offline and online procedures, and analyze its convergence in Section 6.3.

6.1 Homogenization of quasilinear elliptic problems

We assume that the tensor $a^\varepsilon(x,s)$ in (6.0.1) is uniformly elliptic and bounded with respect to s and ε, i.e., there exist $\lambda, \Lambda_1 > 0$ such that

$$\lambda|\xi|^2 \le a^\varepsilon(x,s)\xi\cdot\xi, \quad |a^\varepsilon(x,s)\xi| \le \Lambda_1|\xi|, \quad \forall \xi \in \mathbb{R}^d, \forall s \in \mathbb{R}, \text{ a.e. } x \in \Omega, \tag{6.1.1}$$

and that the functions $a^\varepsilon_{mn}(x,s),\ m,n = 1,\dots,d$ are continuous, bounded and uniformly Lipschitz continuous with respect to s.

Then, for all fixed $\varepsilon > 0$, the weak form of (6.0.1) has a unique solution $u^\varepsilon \in \mathscr{H}^1_0(\Omega)$ (we refer

for example to [52, Theorem 11.6] for a proof). The solution, for each ε, satisfies the a priori bound $\|u^\varepsilon\|_{\mathscr{H}^1(\Omega)} \leq C\|f\|_{\mathscr{H}^{-1}(\Omega)}$, hence one can apply standard compactness arguments to the sequence of solution u^ε that ensure the existence of a subsequence of $\{u^\varepsilon\}$ converging weakly in $\mathscr{H}^1(\Omega)$. The homogenization result is shown in [44, Theorem 3.6] (see also [66]) and reads as follows: there exists a subsequence of $\{a^\varepsilon(\cdot, s)\}$ (again indexed by ε) such that the corresponding sequence of solutions $\{u^\varepsilon\}$ converges weakly to u^0 in $\mathscr{H}^1(\Omega)$. The limit function u^0 is the solution of the homogenized problem

$$-\nabla\cdot\left(a^0(x,u^0(x))\nabla u^0(x)\right) = f(x) \text{ in } \Omega, \qquad u^0(x) = 0 \text{ on } \partial\Omega. \tag{6.1.2}$$

The tensor $a^0(x,s)$, the homogenized tensor, can be shown to be Lipschitz continuous with respect to s, uniformly elliptic, and bounded [44, Prop. 3.5], i.e., there exists $\Lambda_2 > 0$ such that

$$\|a^0(x,s_1) - a^0(x,s_2)\|_{\mathscr{F}} \leq \Lambda_2|s_1 - s_2|, \text{ a.e. } x\in\Omega, \forall s_1, s_2 \in \mathbb{R}, \tag{6.1.3}$$

and there exist $\lambda, \Lambda_1 > 0$ such that a^0 satisfies (6.1.1) (possibly with different constants). Under these assumptions, the homogenized problem (6.1.2) has also a unique solution $u^0 \in \mathscr{H}^1_0(\Omega)$.

We mention that for a locally periodic tensor of the form $a^\varepsilon(x,s) = a(x, x/\varepsilon, s)$ where $a(x,y,s)$ is Y periodic with respect to y, the weak convergence of u^ε to the solution of (6.1.2) holds for the whole sequence $\{u^\varepsilon\}$ and the homogenized tensor can be characterized in the following way [33]:

$$a^0(x,s) = \int_Y a(x,y,s)(I + J^T_{\chi(x,y,s)})\,dy, \qquad \text{for } x\in\Omega, s\in\mathbb{R}, \tag{6.1.4}$$

where $J_{\chi(x,y,s)}$ is a $d\times d$ matrix with entries $J_{\chi(x,y,s)\,ij} = (\partial\chi^i)/(\partial y_j)$ and $\chi^i(x,\cdot,s)$, $i = 1,\dots,d$ are the unique solutions in $\mathscr{W}^1_{per}(Y)$ of the linear cell problems with parameters $x\in\Omega, s\in\mathbb{R}$

$$\int_Y a(x,y,s)\nabla_y\chi^i(x,y,s)\cdot\nabla w(y)\,dy = -\int_Y a(x,y,s)\mathbf{e}_i\cdot\nabla w(y)\,dy, \quad \forall w\in\mathscr{W}^1_{per}(Y). \tag{6.1.5}$$

Remark 6.1.1. *We sometimes refer to the problems* (6.1.2) *or* (6.0.1) *as "non monotone problems". This stems from the following fact: writing for example* (6.1.2) *in weak form*

$$B(u^0;u^0,v) = \int_\Omega a^0(x,u^0(x))\nabla u^0(x)\nabla v(x)\,dx = (f,v), \quad \forall v\in\mathscr{H}^1_0(\Omega),$$

we observe that the monotonicity property $B(u^0;u^0,u^0-v) - B(v;v,u^0-v) \geq C\|u^0-v\|^2_{\mathscr{H}^1(\Omega)}$ *with* $C\geq 0$ *does not hold in general for the quasilinear problem* (6.1.2) *(or* (6.0.1)*). This lack of monotonicity makes the numerical analysis for FEM a nontrivial task, in particular when quadrature formula are used [25].*

For our analysis, we will further assume that the tensor a^ε is symmetric (and thus also a^0) and that the homogenized tensor is continuous,

$$a^0_{mn} \in \mathscr{C}^0(\overline{\Omega}\times\mathbb{R}), \qquad \forall m,n = 1,\dots,d. \tag{6.1.6}$$

6.2 Reduced basis FE-HMM for quasilinear problems

As the homogenized tensor a^0 in (6.1.2) is in general unknown, the task in numerical homogenization is to design an algorithm capable of computing an approximation of the homogenized solution u^0 without knowing a^0, relying on a finite number of localized micro problems, i.e. cell problems, chosen in such a way that the overall computation is both efficient and reliable. Here, we generalize the RB-FE-HMM introduced in Chapter 4 for linear elliptic problems to quasilinear elliptic problems.

6.2.1 Preliminaries

We still use the notations defined in Section 1.2.1 for the macro FEM space. Following the discussion in Section 4.1, we will again denote the micro FE space by $S^q(K_{\delta_j},\mathcal{N})$ instead of $S^q(K_{\delta_j},\mathcal{T}_h)$ (defined in Section1.2.2) to emphasize on the dimension $\mathcal{N}$ of the micro FE space which in RB strategy is required to be large. Analogously, the functions in $S^q(Y,\mathcal{N})$ are denoted using the subscript $\mathcal{N}$ (e.g., $\hat{z}_{\mathcal{N}}$). We then recall the RB space defined in (4.2.26), which is a subspace of $S^q(Y,\mathcal{N})$ with a low dimension denoted

$$S_N(Y) = \text{span}\{\hat{\xi}_{n,\mathcal{N}}(y),\ n = 1,..,N\} \subset S^q(Y,\mathcal{N}), \tag{6.2.7}$$

where $\hat{\xi}_{n,\mathcal{N}}(y),\ n = 1,..,N$ denotes the reduced basis. Notice here that for the analysis of the RB-FE-HMM, we shall also consider a new RB space of the form

$$\overline{S}_N(Y) = \text{span}\{(\hat{\xi}_{n,\mathcal{N}},\hat{\zeta}_{n,\mathcal{N}}),\ n = 1,..,N\} \subset S^q(Y,\mathcal{N})^2,$$

which is a subspace of dimension N of $(S^q(Y,\mathcal{N}))^2$ involving the same functions $\hat{\xi}_{n,\mathcal{N}}$ as in $S_N(Y)$ and where $\hat{\zeta}_{n,\mathcal{N}} \in S^q(Y,\mathcal{N}),\ n = 1,..,N$. The construction of the RB spaces $S_N(Y)$ and $\overline{S}_N(Y)$ is discussed in Section 6.2.4 below.

For each macro element $K \in \mathcal{T}_H$ and each quadrature point $x_{K_j} \in K,\ j = 1,\ldots,J$, we have the sampling domains $K_{\delta_j} = x_{K_j} + (-\delta/2,\delta/2)^d$, $(\delta \geq \varepsilon)$. We recall that each sampling domain K_{δ_j} is in correspondence with Y through the affine transformation as defined in Section 4.1

$$y \in Y \mapsto G_{x_{K_j}}(y) = x_{K_j} + \delta y \in K_{\delta_j} \tag{6.2.8}$$

This transformation applied to the RB space (6.2.7) permits to define the RB space $S_N(K_{\delta_j})$ associated to each sampling domain K_{δ_j} as

$$S_N(K_{\delta_j}) = \text{span}\{\delta\hat{\xi}_{n,\mathcal{N}}(G^{-1}_{x_{K_j}}(x)) =: \xi_{n,K_j}(x),\ n = 1,..,N\}. \tag{6.2.9}$$

6.2.2 Online procedure: the RB-FE-HMM

Assuming that the RB space has been pre-constructed in the offline stage described in the next section, we introduce a macro method similar to the FE-HMM with the micro problems solved in the RB space.

The nonlinear RB-FE-HMM for (6.0.1) is defined as follows: find $u^{H,RB} \in S_0^\ell(\Omega, \mathcal{T}_H)$ such that

$$B_{H,RB}(u^{H,RB}; u^{H,RB}, v^H) = \int_\Omega f v^H dx, \ \forall v^H \in S_0^\ell(\Omega, \mathcal{T}_H), \tag{6.2.10}$$

with a bilinear form defined for all $u^H, v^H, w^H \in S_0^\ell(\Omega, \mathcal{T}_H)$ by

$$B_{H,RB}(u^H; v^H, w^H) := \sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J} \frac{\omega_{K_j}}{|K_{\delta_j}|} \int_{K_{\delta_j}} a^\varepsilon(x, u^H(x_{K_j})) \nabla v_{N,K_j}^{u^H(x_{K_j})}(x) \cdot \nabla w_{N,K_j}^{u^H(x_{K_j})}(x) dx, \tag{6.2.11}$$

where for the scalar parameter $s = u^H(x_{K_j})$, the function v_{N,K_j}^s solves $v_{N,K_j}^s - v_{lin,j}^H \in S_N(K_{\delta_j})$ and

$$\int_{K_{\delta_j}} a^\varepsilon(x,s) \nabla v_{N,K_j}^s(x) \cdot \nabla z_N(x) dx = 0, \quad \forall z_N \in S_N(K_{\delta_j}) \tag{6.2.12}$$

and similarly for $w_{N,K_j}^s(x)$. The problem (6.2.12) requires the solution of an $N \times N$ linear system, where the details of the offline output and the online implementation are discussed in Chapter 7. The efficiency of the RB procedure relies in the fact that the dimension N of the RB space is usually small. Furthermore, in contrast to the standard FE-HMM, the number of degrees of freedom (DOF) of the micro (RB) space remains fixed during the online procedure and does not increase as the macroscopic DOF increase. This is in sharp contrast with the FE-HMM for which the simultaneous refinement of the macro and micro DOF is a major computational issue [1].

6.2.3 Solution of the macro quasilinear problem and Newton method

While the cell problems (6.2.12) are linear, the macroscopic problem (6.2.11) is nonlinear and is usually solved by a Newton method.

The following reformulation of the bilinear form of the RB-FE-HMM will be useful to define the Newton method used in practice to compute a numerical solution $u^{H,RB}$ of (6.2.10). The bilinear form (6.2.11) can be rewritten as

$$B_{H,RB}(u^H; v^H, w^H) = \sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J} \omega_{K_j} a_{N,K_j}^0(u^H(x_{K_j})) \nabla v^H(x_{K_j}) \cdot \nabla w^H(x_{K_j}), \tag{6.2.13}$$

where we define the numerical homogenized tensor as

$$(a_{N,K_j}^0(x_{K_j}, s))_{ik} = \int_Y a_{x_{K_j},s}(y) \left(\nabla \hat{\chi}_{N,K_j}^{i,s}(y) + \mathbf{e}_i \right) \cdot \left(\nabla \hat{\chi}_{N,K_j}^{k,s}(y) + \mathbf{e}_k \right) dy. \tag{6.2.14}$$

where $\hat{\chi}^{i,s}_{N,K_j} \in S_N(Y),\ i = 1,\dots,d$ is the solution of a cell problem (see (6.2.25) below) corresponding to the sampling domain K_{δ_j}.

Inspired by [73, 26], we explain here how to solve the nonlinear problem (6.2.10) with the Newton method. For given $z^H, v^H, w^H \in S^\ell_0(\Omega, \mathcal{T}_H)$ we first define the Fréchet derivative ∂B_H obtained by differentiating the nonlinear quantity $B_H(z^H, z^H, w^H)$ with respect to z^H

$$\partial B_{H,RB}(z^H; v^H, w^H) := B_{H,RB}(z^H; v^H, w^H) + B'_{H,RB}(z^H; v^H, w^H), \tag{6.2.15}$$

where by the reformulation of the RB-FE-HMM bilinear form (6.2.13) we derive

$$B'_H(z^H; v^H, w^H) = \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \omega_{K_j} \frac{d}{ds} a^0_{N,K_j}(s)|_{s=z^H(x_{K_j})} v^H(x_{K_j}) \nabla z^H(x_{K_j}) \cdot \nabla w^H(x_{K_j}). \tag{6.2.16}$$

The Newton method for approximating a solution u^H of the nonlinear RB-FE-HMM (6.2.10) by a sequence $\{u^H_k\}$ reads in weak form

$$\partial B_H(u^H_k; u^H_{k+1} - u^H_k, w^H) = F_H(w^H) - B_H(u^H_k; u^H_k, w^H), \quad \forall\, w^H \in S^\ell_0(\Omega, \mathcal{T}_H). \tag{6.2.17}$$

The fact that the Newton method is well defined and convergence is discussed in Section 6.3.2 while an efficient implementation is detailed in next chapter.

6.2.4 Offline procedure: RB for quasilinear problems

This section describes the offline stage of the RB algorithm in our context of quasilinear elliptic problems. The task is to construct a low dimensional RB space $S_N(Y)$ spanned by a small number $N \ll \mathcal{N}$ of representative solutions of the cell problems (6.2.22) below (depending on the quadrature node x_{K_j} and the nonlinear parameter s). Here again, $\mathcal{N}$ denotes the (large) DOF of the FE space used to obtain a highly resolved solution of (6.2.22).

The main novelty is that the proposed RB algorithm permits to compute efficiently with a reliable a posteriori error control not only the solutions of the cell problems (6.2.22) but also their derivatives with respect to the nonlinear parameter s. This is an essential ingredient to prove in Section 6.2.3 the uniqueness of the RB-FE-HMM macro solution and the convergence of the Newton method.

Considering an affine representation of the tensor, we first describe a suitable formulation of the cell problems before presenting the parametrized cell solution space itself. We then introduce a new a posteriori error estimator and analyze its efficiency and reliability. This is the key ingredient of the Greedy algorithm for the construction of the RB space that concludes this section.

Affine representation of the tensor. A suitable representation of the tensor

$$a_{x_{K_j},s}(y) := a^\varepsilon(G_{x_{K_j}}(y), s), \tag{6.2.18}$$

where we use the transformation (6.2.8) that is crucial for the RB methodology, i.e., an affine representation of the form

$$a_{x_\tau,s}(y) = \sum_{p=1}^{P} \Theta_p(x_\tau, s) a_p(y), \ \forall y \in Y. \tag{6.2.19}$$

We remark that if such direct affine representation is unavailable, the EIM can also be used to approximate a nonaffine tensor by an affine one of the form (6.2.19) as we discussed in Section 3.3.

Cell problems. The micro problems in the FE-HMM are based on the FE approximation of the cell functions $\chi^{i,s}_{K_j} \in \mathcal{W}(K_{\delta_j})$, solving the linear problem

$$\int_{K_{\delta_j}} a^\varepsilon(x,s)\nabla\chi^{i,s}_{K_j}(x)\cdot\nabla z(x)dx = -\int_{K_{\delta_j}} a^\varepsilon(x,s)\mathbf{e}_i\cdot\nabla z(x)dx, \ \forall z \in \mathcal{W}(K_{\delta_j}). \tag{6.2.20}$$

which has a unique solution using (6.1.1). For the design of the RB method, is more convenient to work in the space $\mathcal{W}(Y)$ (defined in either (1.2.12) or (1.2.13)) rather than the quadrature node dependent space $\mathcal{W}(K_{\delta_j})$. We thus consider the transformation (6.2.8) and using the notations

$$\begin{aligned} b(\hat{v},\hat{z}) &:= \int_Y a_{x_{K_j},s}(y)\nabla\hat{v}(y)\cdot\nabla\hat{z}(y)dy \ \ \forall \hat{v},z \in \mathcal{W}(Y), \\ l_i(\hat{z}) &:= -\int_Y a_{x_{K_j},s}(y)\mathbf{e}_i\cdot\nabla\hat{z}(y)dy \ \ \forall \hat{z} \in \mathcal{W}(Y), \end{aligned} \tag{6.2.21}$$

the problem (6.2.20) with $\hat{\chi}^{i,s}_{K_j}(y) = \chi^{i,s}_{K_j}(G_{x_{K_j}}(y))$ can be transformed into

$$b(\hat{\chi}^{i,s}_{K_j},\hat{z}) = l_i(\hat{z}), \ \forall z \in \mathcal{W}(Y). \tag{6.2.22}$$

On $\mathcal{W}(Y)$ we consider the scalar product $(v,w)_{\mathcal{W}} = \int_Y \nabla v\cdot\nabla w dy$ and associated norm $\|v\|_{\mathcal{W}} = ((v,v)_{\mathcal{W}})^{1/2}$ and for $(T_\delta, s) \in \mathcal{D}$ the energy norm

$$\|v\|_{\mathcal{E},T_\delta,s} := (b(v,v))^{1/2} = \left(\int_Y a_{x_\tau,s}(y)\nabla v(y)\cdot\nabla v(y)dy\right)^{1/2}, \tag{6.2.23}$$

and notice that from the ellipticity of the tensor it holds

$$\|v\|_{\mathcal{W}} \leq \frac{1}{\sqrt{\lambda}}\|v\|_{\mathcal{E},T_\delta,s}. \tag{6.2.24}$$

Consider $\hat{\chi}^{i,s}_{\mathcal{N},K_j} \in S^q(Y,\mathcal{N})$ the solution of the linear problem

$$b(\hat{\chi}^{i,s}_{\mathcal{N},K_j}, \hat{z}_{\mathcal{N}}) = l_i(\hat{z}_{\mathcal{N}}) \ \ \forall \hat{z}_{\mathcal{N}} \in S^q(Y,\mathcal{N}), \tag{6.2.25}$$

We notice using (6.1.1) that problem (6.2.25) has a unique solution.

For the convergence of the Newton method explained in Section 6.2.3 we will also need to control the derivatives with respect to the parameter s of the cell functions $\hat{\chi}^{i,s}_{K_j}$. We assume[1]

$$\begin{aligned} &s \in \mathscr{R} \mapsto a^\varepsilon(\cdot,s) \in (\mathscr{L}^\infty(\Omega))^{d\times d} \text{ is of class } C^1, \\ &|\partial_s a^\varepsilon(x,s)\xi| \le \Lambda_2|\xi|, \ \forall s \in \mathbb{R}, \text{ a.e. } x \in \Omega, \forall \xi \in \mathbb{R}^d. \end{aligned} \tag{6.2.26}$$

Lemma 6.2.1. *Assume that* (6.1.1) *and* (6.2.26) *hold. Consider the solution* $\hat{\chi}^{i,s}_{\mathcal{N},K_j}$ *of* (6.2.25). *Then, the map* $s \mapsto \hat{\chi}^{i,s}_{\mathcal{N},K_j} \in \mathscr{H}^1(T_\delta)$ *is of class* C^1 *and satisfies*

$$\frac{\partial}{\partial s}\hat{\chi}^{i,s}_{\mathcal{N},K_j} = \hat{\phi}^{i,s}_{\mathcal{N},K_j}, \qquad \frac{\partial}{\partial s}\nabla\hat{\chi}^{i,s}_{\mathcal{N},K_j} = \nabla\hat{\phi}^{i,s}_{\mathcal{N},K_j}, \tag{6.2.27}$$

where for all $\hat{\zeta}_{\mathcal{N}} \in S^q(Y,\mathcal{N})$,

$$\int_Y a_{x_\tau,s}(y)\nabla\hat{\phi}^{i,s}_{\mathcal{N},K_j}(y)\cdot\nabla\hat{\zeta}_{\mathcal{N}}(y)dy = -\int_Y \partial_s a_{x_\tau,s}(\nabla\hat{\chi}^{i,s}_{\mathcal{N},K_j}(y)+\mathbf{e}_i)\cdot\nabla\hat{\zeta}_{\mathcal{N}}(y)dx. \tag{6.2.28}$$

Proof. This is a standard result for FEM problems depending smoothly on a parameter (see e.g. Lemma 6.1 in [26] for details). □

Parametrized cell solution space. We consider a compact subspace $\mathscr{D}$ of $\Omega\times\mathbb{R}$. For any randomly chosen parameter[2] $(x_\tau,s)\in\mathscr{D}$, we have the map G_{x_τ} from the physical sampling domain $T_\delta = x_\tau + (-\delta/2,\delta/2)^d$ centered at x_τ to the reference domain Y and consider (6.2.25),(6.2.28) with the tensor $a_{x_\tau,s}(y)$. Next indexed by $\{(T_\delta,s,\mathbf{e}_\eta);(T_\delta,s)\in\mathscr{D} \text{ and } \eta=1,\cdots,d\}$, we define the parametrized cell solution space $\mathscr{M}^{\mathcal{N}}(Y)\subset\mathscr{W}(Y)^2$ given by

$$\mathscr{M}^{\mathcal{N}}(Y) := \big\{(\hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta}, \partial_s\hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta}); (T_\delta,s)\in\mathscr{D} \text{ and } \eta=1,\cdots,d\big\}, \tag{6.2.29}$$

where $\hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta} \in S^q(Y,\mathcal{N})$, $\partial_s\hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta} := \frac{\partial}{\partial s}\hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta} \in S^q(Y,\mathcal{N})$ are the solutions of (6.2.25),(6.2.28) associated with the mapping G_{x_τ} and the Hilbert space $\mathscr{W}(Y)$ is defined in either (1.2.12) or (1.2.13). On the Hilbert product space $\mathscr{W}(Y)^2$ we define the norms

$$\|(u,v)\|_{\mathscr{W}\times\mathscr{W}} := (\|u\|^2_{\mathscr{W}}+\|v\|^2_{\mathscr{W}})^{1/2} \quad \text{and} \quad \|(u,v)\|_{\mathscr{E}\times\mathscr{E},T_\delta,s} := (\|u\|^2_{\mathscr{E},T_\delta,s}+\|v\|^2_{\mathscr{E},T_\delta,s})^{1/2}. \tag{6.2.30}$$

[1] It is shown in [44, Rem. 3.3, Prop. 3.5] that the best constant Λ_2 in (6.1.3) may differ from the one in (6.2.26).
[2] $\mathscr{D}$ should be chosen such that $T_\delta\subset\Omega$, for all $(x_\tau,s)\in\mathscr{D}$.

The goal of the Greedy procedure described below is to find an $N-$dimensional subspace of $\mathscr{M}^{\mathscr{N}}(Y)$, called $\overline{S}_N(Y)$, that minimizes the projection error of functions in $\mathscr{M}^{\mathscr{N}}(Y)$ over other choices of $N-$dimensional subspaces. We emphasize that the derivative functions $\partial_s\hat{\xi}^{\eta,s}_{\mathscr{N},T_\delta}$ involved in the definition (6.2.29) of $\mathscr{M}^{\mathscr{N}}(Y)$ are considered only for the analysis, but should not be computed explicitly in the implementation. Hence the solution of the online cell problem (6.2.12) will involve the reduced basis space $S_N(Y)$, defined as the first component of each couple of functions in $\overline{S}_N(Y)$.

A posteriori error estimator. The procedure of selecting the representative cell solutions is conducted by an a posteriori error estimator which allows to control the accuracy of our output of interest (the numerically homogenized tensor) [93, 45].

Assume that the RB space of dimension l, denoted by $\overline{S}_l(Y)$, is available (its construction will be detailed in Algorithm 6.2.4). Given the parameters (x_τ, s, i), consider $(\hat{\xi}^{i,s}_{\mathscr{N},T_\delta}, \partial_s\hat{\xi}^{i,s}_{\mathscr{N},T_\delta})$, $(\hat{\xi}^{i,s}_{l,T_\delta}, \partial_s\hat{\xi}^{i,s}_{l,T_\delta})$ the solutions of (6.2.25),(6.2.28) in $S^q(Y,\mathscr{N})^2$ and $\overline{S}_l(Y)$, respectively (i.e. with test functions $(z_\mathscr{N}, \zeta_\mathscr{N})$ in $S^q(Y,\mathscr{N})^2$ and $\overline{S}_l(Y)$, respectively). We then consider

$$\hat{e}^{i,s}_{l,T_\delta} = \hat{\xi}^{i,s}_{l,T_\delta} - \hat{\xi}^{i,s}_{\mathscr{N},T_\delta}, \tag{6.2.31}$$

$$\partial_s\hat{e}^{i,s}_{l,T_\delta} = \partial_s\hat{\xi}^{i,s}_{l,T_\delta} - \partial_s\hat{\xi}^{i,s}_{\mathscr{N},T_\delta}. \tag{6.2.32}$$

We derive an a posteriori estimator for both $\hat{e}^{i,s}_{l,T_\delta}$ and $\partial_s\hat{e}^{i,s}_{l,T_\delta}$ will be analyzed in Lemma 6.2.3. We have that

$$b(\hat{e}^{i,s}_{l,T_\delta}, \hat{z}_\mathscr{N}) = b(\hat{\xi}^{i,s}_{l,T_\delta}, \hat{z}_\mathscr{N}) - l_i(\hat{z}_\mathscr{N}),\ \forall \hat{z}_\mathscr{N} \in S^q(Y,\mathscr{N}), \tag{6.2.33}$$

where the right-hand side defines a linear form on $S^q(Y,\mathscr{N})$. Hence, by the Riesz theorem, there exists a unique $\bar{e}^{i}_{l,T_\delta} \in S^q(Y,\mathscr{N})$ such that

$$b(\hat{e}^{i,s}_{l,T_\delta}, \hat{z}_\mathscr{N}) = (\bar{e}^{i,s}_{l,T_\delta}, \hat{z}_\mathscr{N})_\mathscr{W},\ \forall \hat{z}_\mathscr{N} \in S^q(Y,\mathscr{N}). \tag{6.2.34}$$

We then define the residual of the a posteriori error estimator as

$$\Delta^{i,s}_{l,T_\delta} := \frac{\|\bar{e}^{i,s}_{l,T_\delta}\|_\mathscr{W}}{\sqrt{\lambda_{LB}}} + \frac{\|\partial_s\bar{e}^{i,s}_{l,T_\delta}\|_\mathscr{W}}{\sqrt{\lambda_{LB}}}, \tag{6.2.35}$$

where λ_{LB} is an approximation of the coercivity constant λ defined in (6.1.1). We notice that the first term in (6.2.35) is the standard residual used for linear problems as shown in Chapter 4. The second term arises from the nonlinearity of our problem and its control is needed to ensure the uniqueness of the nonlinear RB-FE-HMM and the convergence of the Newton method used in the implementation.

Remark 6.2.2. *To compute the residual* $\bar{e}^{i,s}_{l,T_\delta}$ *in* (6.2.35)*, we first observe that we need to solve* (6.2.34)*, which depends on the parameter* s*. As discussed in Chapter 4, thanks to the affine representation of the tensor,* $\|\bar{e}^{i,s}_{l,T_\delta}\|_{\mathscr{W}}$ *is efficient to compute. Second, for evaluating* $\partial_s \bar{e}^{i,s}_{l,T_\delta}$ *one can simply consider the finite difference approximation*

$$\partial_s \bar{e}^{i,s}_{l,T_\delta} \approx \frac{\bar{e}^{i,s+\sqrt{eps}}_{l,T_\delta} - \bar{e}^{i,s}_{l,T_\delta}}{\sqrt{eps}},$$

where eps *is the machine precision. This can be done by solving* (6.2.34) *twice with parameters* s *and* $s+\sqrt{eps}$*, respectively. In the analysis, we shall neglect the error of the above finite difference.*

The next lemma gives a bound for the a posteriori error in output of interest in terms of the norms (6.2.30). It is a generalization of the result [7, Lemma 3.3] in the context of linear elliptic problems. These results are needed in our nonlinear context to control the microscopic error in the macroscopic (nonlinear) solver.

Consider $\bar{e}^{i,s}_{l,T_\delta}$ defined in (6.2.34) and the residual $\Delta^{i,s}_{l,T_\delta}$ defined in (6.2.35). Define

$$(a^0_{\mathscr{N},T_\delta}(x_\tau,s))_{ij} = \int_Y a_{x_\tau,s}(y)\left(\nabla\hat{\xi}^{i,s}_{\mathscr{N},T_\delta}(y)+\mathbf{e}_i\right)\cdot\left(\nabla\hat{\xi}^{j,s}_{\mathscr{N},T_\delta}(y)+\mathbf{e}_j\right)dy, \tag{6.2.36}$$

$$(a^0_{l,T_\delta}(x_\tau,s))_{ij} = \int_Y a_{x_\tau,s}(y)\left(\nabla\hat{\xi}^{i,s}_{l,T_\delta}(y)+\mathbf{e}_i\right)\cdot\left(\nabla\hat{\xi}^{j,s}_{l,T_\delta}(y)+\mathbf{e}_j\right)dy. \tag{6.2.37}$$

Lemma 6.2.3. *Assume* (6.1.1) *and* (6.2.26)*. Let* $(\hat{\xi}^{i,s}_{\mathscr{N},T_\delta}, \partial_s\hat{\xi}^{i,s}_{\mathscr{N},T_\delta})$ *and* $(\hat{\xi}^{i,s}_{l,T_\delta}, \partial_s\hat{\xi}^{i,s}_{l,T_\delta})$ *be the solution of problem* (6.2.25)-(6.2.28) *in* $S^q(Y,\mathscr{N})^2$ *and* $\overline{S}_l(Y)$*, with test functions* $(z_{\mathscr{N}},\zeta_N)$ *in* $S^q(Y,\mathscr{N})^2$ *and* $\overline{S}_l(Y)$*, respectively. Assume that the approximation* λ_{LB} *of the coercivity constant satisfies* $0<\lambda_{LB}\le\lambda$*. Consider the quantities* $\hat{e}^{i,s}_{l,T_\delta}$ *and* $\partial_s\hat{e}^{i,s}_{l,T_\delta}$ *defined in* (6.2.31)*. Then*

$$\|(\hat{e}^{i,s}_{l,T_\delta}, \partial_s\hat{e}^{i,s}_{l,T_\delta})\|_{\mathscr{E}\times\mathscr{E},T_\delta,s} \le (2+\frac{\Lambda_2}{\lambda_{LB}})\Delta^{i,s}_{l,T_\delta}, \tag{6.2.38}$$

$$(2\Lambda_1+\Lambda_2)^{-1}\lambda_{LB}^{1/2}\Delta^i_{l,T_\delta} \le \|(\hat{e}^{i,s}_{l,T_\delta}, \partial_s\hat{e}^{i,s}_{l,T_\delta})\|_{\mathscr{W}\times\mathscr{W}} \le (2\lambda_{LB}^{-1/2}+\Lambda_2\lambda_{LB}^{-3/2})\Delta^i_{l,T_\delta}, \tag{6.2.39}$$

$$|(a^0_{\mathscr{N},T_\delta}(s))_{ij}-(a^0_{l,T_\delta}(s))_{ij}| + |\partial_s(a^0_{\mathscr{N},T_\delta}(s))_{ij}-\partial_s(a^0_{l,T_\delta}(s))_{ij}| \le 3\Big(1+\frac{\Lambda_2}{\lambda_{LB}}\Big)\Delta^{i,s}_{l,T_\delta}\Delta^{j,s}_{l,T_\delta}, \tag{6.2.40}$$

where Λ_1,Λ_2 *are the constants in* (6.2.26),(6.1.1) *and* $\|\cdot\|_{\mathscr{E},T_\delta,s}$ *is the energy norm defined in* (6.2.23).

Proof of Lemma 6.2.3. Taking $\hat{z}_{\mathscr{N}} = \hat{e}^{i,s}_{l,T_\delta}$ in (6.2.34) and using (6.2.24) yields successively,

$$\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathscr{E},T_\delta,s} \le \Delta^{i,s}_{l,T_\delta}, \tag{6.2.41}$$

$$\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathscr{W}} \le \frac{\Delta^{i,s}_{l,T_\delta}}{\sqrt{\lambda_{LB}}}. \tag{6.2.42}$$

A consequence of (6.2.36),(6.2.37), and the symmetry of the tensor, is the identity

$$\begin{aligned}&(a^0_{l,T_\delta}(x_\tau,s))_{ij}-(a^0_{\mathcal{N},T_\delta}(x_\tau,s))_{ij}\\&=\int_Y a_{x_\tau,s}(y)\left(\nabla\hat{\xi}^{i,s}_{\mathcal{N},T_\delta}(y)-\nabla\hat{\xi}^{i,s}_{l,T_\delta}(y)\right)\cdot\left(\nabla\hat{\xi}^{j,s}_{\mathcal{N},T_\delta}(y)-\nabla\hat{\xi}^{j,s}_{l,T_\delta}(y)\right)dy.\end{aligned}\tag{6.2.43}$$

We deduce from the Cauchy-Schwarz inequality and (6.2.41),

$$|(a^0_{\mathcal{N},T_\delta}(s))_{ij}-(a^0_{l,T_\delta}(s))_{ij}|\le\Delta^{i,s}_{l,T_\delta}\Delta^{j,s}_{l,T_\delta}.\tag{6.2.44}$$

Using Lemma 6.2.1, we obtain after differentiation of (6.2.34) with respect to the parameter s,

$$(\partial_s\bar{e}^{i,s}_{l,T_\delta},\hat{z}_{\mathcal{N}})_{\mathcal{W}}=\int_Y\partial_s a_{x_\tau,s}\nabla\hat{e}^{i,s}_{l,T_\delta}\cdot\nabla\hat{z}_{\mathcal{N}}dy+\int_Y a_{x_\tau,s}\nabla(\partial_s\hat{e}^{i,s}_{l,T_\delta})\cdot\nabla\hat{z}_{\mathcal{N}}dy,\ \forall\hat{z}_{\mathcal{N}}\in S^q(Y,\mathcal{N}).\tag{6.2.45}$$

We take $\hat{z}_{\mathcal{N}}=\partial_s\hat{e}^{i,s}_{l,T_\delta}$ in (6.2.45) and we write

$$\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|^2_{\mathcal{E},T_\delta,s}=(\partial_s\bar{e}^{i,s}_{l,T_\delta},\partial_s\hat{e}^{i,s}_{l,T_\delta})_{\mathcal{W}}-\int_Y\partial_s a_{x_\tau,s}(y)\nabla\hat{e}^{i,s}_{l,T_\delta}(y)\cdot\nabla(\partial_s\hat{e}^{i,s}_{l,T_\delta}(y))dy.$$

We deduce from the Cauchy-Schwarz inequality and (6.2.24),(6.2.26),

$$\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|^2_{\mathcal{E},T_\delta,s}\le\lambda_{LB}^{-1/2}\|\partial_s\bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{E},T_\delta,s}+\Lambda_2\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}$$

which gives, using (6.2.42),(6.2.24),

$$\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{E},T_\delta,s}\le(1+\Lambda_2/\lambda_{LB})\Delta^{i,s}_{l,T_\delta}.\tag{6.2.46}$$

The estimates (6.2.41) and (6.2.46) yield (6.2.38), and using in addition (6.2.24) proves the upper bound in (6.2.39). Next, taking $\hat{z}_{\mathcal{N}}=\bar{e}^{i,s}_{l,T_\delta}$ in (6.2.34) using the Cauchy-Schwarz inequality yields $\|\bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\le\Lambda_1\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}$, while taking $\hat{z}_{\mathcal{N}}=\partial_s\bar{e}^{i,s}_{l,T_\delta}$ in (6.2.45) yields $\|\partial_s\bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\le\Lambda_2\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}+\Lambda_1\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}$. We obtain $\Delta^i_{l,T_\delta}\le\lambda_{LB}^{-1/2}(\Lambda_1+\Lambda_2)\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}+\lambda_{LB}^{-1/2}\Lambda_1\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}$ which yields the lower bound in (6.2.39). We finally prove (6.2.40). Differentiating the equality (6.2.43) and using (6.2.45) with $\hat{z}_{\mathcal{N}}=\hat{e}^{j,s}_{l,T_\delta}$ we obtain (using the Cauchy-Schwarz inequality)

$$\begin{aligned}&|(\partial_s a^0_{\mathcal{N},T_\delta}(x_\tau,s))_{ij}-(\partial_s a^0_{l,T_\delta}(x_\tau,s))_{ij}|\\&\le3\Lambda_2\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\|\hat{e}^{j,s}_{l,T_\delta}\|_{\mathcal{W}}+\|\partial_s\bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\|\hat{e}^{j,s}_{l,T_\delta}\|_{\mathcal{W}}+\|\partial_s\bar{e}^{j,s}_{l,T_\delta}\|_{\mathcal{W}}\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\\&\le\left(3\frac{\Lambda_2}{\lambda_{LB}}+2\right)\Delta^{i,s}_{l,T_\delta}\Delta^{j,s}_{l,T_\delta}\end{aligned}$$

where we used (6.2.42) and the definition (6.2.35) in the last inequality. Finally, using (6.2.44) concludes the proof. □

Offline algorithm. In the offline stage, we select by a greedy algorithm N triples of the form

$(T_{\delta_n}, s, \eta_n)$, where (T_{δ_n}, s) belongs to a given compact $\mathscr{D} \subset \Omega \times \mathbb{R}$ (since the range of the parameter s can only be obtained when the macro solution $u^{H,RB}$ is computed, we propose in next chapter an *ad hoc* method to find an a priori range of s) and η_n corresponds to the unit vector $\mathbf{e}_{\eta_n}$ belonging to the canonical basis of $\mathbb{R}^d$. Corresponding to the N couples of $(T_{\delta_n}, s, \eta_n)$, we compute $\hat{\xi}^{\eta_n,s}_{\mathscr{N},T_{\delta_n}}$, the solution of (6.2.25) with a tensor given by $a_{x_{\tau_n},s}(y)$ (x_{τ_n} is the barycenter of T_{δ_n}) and a right-hand side given by $l_{\eta_n}(\cdot)$. The complete offline algorithm stated below is also based on the usual procedure of the RB methodology (see [93, 96]) and in Chapter 7, we will discuss about the details of implementation.

Notice that in the case of a linear elliptic problem (i.e. $a^\varepsilon(x,s)$ independent of s), it coincides with the Greedy procedure proposed in Chapter 4.

Algorithm 6.2.4 (Greedy procedure)**.** *Given the maximum basis number N_{RB} and a stopping tolerance tol_{RB}:*

1. *Choose randomly (by a Monte Carlo method) N_{train} parameters $(T_{\delta_n}, s_n) \in \mathscr{D}$ (N_{train} large). Define the "training set" $\Xi_{RB} = (T_{\delta_n}, s_n, \eta_n); 1 \le \eta_n \le d, 1 \le n \le N_{train}\}$.*

2. *Select randomly $(T_{\delta_1}, s_1, \eta_1) \in \Xi_{RB}$ and compute $\hat{\xi}^{\eta_1,s_1}_{\mathscr{N},T_{\delta_1}}$, the solution of* (6.2.25) *with right-hand side $l_{\eta_1}(\cdot)$ in $S^q(Y,\mathscr{N})$, corresponding to the selected parameter $(T_{\delta_1}, s_1, \eta_1)$. Set $l = 1$ and define $\hat{\xi}_{1,\mathscr{N}}(y) = \frac{\hat{\xi}^{\eta_1,s_1}_{\mathscr{N},T_{\delta_1}}(y)}{\|\hat{\xi}^{\eta_1,s_1}_{\mathscr{N},T_{\delta_1}}\|_W}$, and the corresponding RB space $\overline{S}_1(Y) = span\{(\hat{\xi}^{\eta_1,s_1}_{\mathscr{N},T_{\delta_1}}, \partial_s \hat{\xi}^{\eta_1,s_1}_{\mathscr{N},T_{\delta_1}})\}$.*

3. *For $l = 2, \dots, N_{RB}$*

   a. *Compute for each $(T_\delta, s, \eta) \in \Xi_{RB}$ the residual $\Delta^{\eta,s}_{l-1,T_\delta}$ defined in* (6.2.35) *and select the next reduced basis by choosing*

   $$(T_{\delta_l}, s_l, \eta_l) = argmax_{(T_\delta,s,\eta)\in\Xi_{RB}} \Delta^{\eta,s}_{l-1,T_\delta},$$

   provided that[3] $\max_{(T_\delta,s,\eta)\in\Xi_{RB}} (\Delta^{\eta,s}_{l-1,T_\delta})^2 > tol_{RB}$*, otherwise the algorithm ends.*

   b. *Compute $\hat{\xi}^{\eta_l,s_l}_{\mathscr{N},T_{\delta_l}}$ the solution of* (6.2.25) *in $S^q(Y,\mathscr{N})$ corresponding to the selected parameters $(T_{\delta_l}, s_l, \eta_l)$. Enlarge the RB space: $\overline{S}_l(Y) = \overline{S}_{l-1}(Y) \oplus span\{(\hat{\xi}^{\eta_l,s_l}_{\mathscr{N},T_{\delta_l}}, \partial_s \hat{\xi}^{\eta_l,s_l}_{\mathscr{N},T_{\delta_l}})\}$. Set $l = l+1$ and go back to a.*

We emphasize once again that the derivative functions $\partial_s \hat{\xi}^{\eta_l,s_l}_{\mathscr{N},T_{\delta_l}}$ involved in Algorithm 6.2.4 do not need to be computed in the implementation and shall be considered only in the analysis. Thanks to Remark 6.2.2 for the a posteriori error estimator evaluation, as output of the Greedy algorithm, it is sufficient to compute only the list of functions $\hat{\xi}_{l,\mathscr{N}}, l = 1, \dots, N$ that span the space $S_N(Y) := span\{\hat{\xi}_{1,\mathscr{N}}, \dots, \hat{\xi}_{N,\mathscr{N}}\}$. These RB functions are obtained by orthogonalizing in

[3] Notice that the error of the outputs of interest scale like the square of the error of the cell functions (6.2.40).

$\mathscr{W}(Y)$ the functions $\hat{\xi}^{\eta_l,s_l}_{\mathscr{N},T_{\delta_l}}$, $l=1,\dots,N$ and they are defined as

$$\hat{\xi}_{l,\mathscr{N}}(y) := \frac{R_l(y)}{\|R_l\|_{\mathscr{W}}}, \quad \text{where} \quad R_l(y) := \hat{\xi}^{\eta_l,s_l}_{\mathscr{N},T_{\delta_l}}(y) - \sum_{m=1}^{l-1} (\hat{\xi}^{\eta_l,s_l}_{\mathscr{N},T_{\delta_l}}, \hat{\xi}_{m,\mathscr{N}})_{\mathscr{W}} \hat{\xi}_{m,\mathscr{N}}.$$

Remark 6.2.5. *Notice that we choose to orthogonalize the RB in $\mathscr{W}(Y)$ with respect to the scalar product $(\cdot,\cdot)_{\mathscr{W}}$ rather than the RB in $\mathscr{W}(Y)^2$ with respect to the scalar product $(\cdot,\cdot)_{\mathscr{W}^2}$ associated to the norm $\|\cdot\|_{\mathscr{W}^2}$ in* (6.2.30) *(as normally expected in the usual RB methodology) because this is more convenient in the implementation (avoiding the computation of $\partial_s \hat{\xi}^{\eta_l,s_l}_{\mathscr{N},T_{\delta_l}}$). Since $\max_{(T_\delta,s,\eta)\in\Xi_{RB}}(\Delta^{\eta,s}_{l,T_\delta})$ decays exponentially as l increases (under the assumptions of Theorem 6.3.2, a slight variation of the result in [48, Corollary 4.1] and [43]), we have $\hat{\xi}^{\eta_l,s_l}_{\mathscr{N},T_{\delta_l}} \notin S_{l-1}$ and thus $\|R_l\|_{\mathscr{W}} \neq 0$ and the above orthogonalization procedure succeeds.*

6.3 Analysis of the RB-FE-HMM

In this section we first derive a priori error estimates for the RB-FE-HMM. We then show the uniqueness of the numerical approximation which is based on the convergence of the Newton method.

6.3.1 A priori error analysis

We introduce the following quantity to measure the error between the tensor a^0 of the homogenized problem (6.1.2) and the numerical homogenized tensor $a^0_{\mathscr{N},K_j}$ in (6.2.14).

$$r_{HMM} := \sup_{K\in\mathscr{T}_H, x_{K_j}\in K} \|a^0(x_{K_j}, u^{H,RB}(x_{K_j})) - a^0_{\mathscr{N},K_j}(u^{H,RB}(x_{K_j}))\|_{\mathscr{F}}. \tag{6.3.47}$$

Theorem 6.3.1. *Consider u^0 the solution of problem* (6.1.2)*. Let $\ell \geq 1$ and $\mu = 0$ or 1. Consider a quasi-uniform macro mesh satisfying (***Q1***), (***Q2***) (introduced in Section 1.2.1). Assume*

$$u^0 \in \mathscr{H}^{\ell+1}(\Omega)\cap\mathscr{W}^{1,\infty}(\Omega), \qquad a^0_{mn} \in \mathscr{W}^{\ell+\mu,\infty}(\Omega\times\mathbb{R}), \quad \forall m,n=1\dots d.$$

Assume further that (6.1.1),(6.1.3),(6.1.6) *hold and that $\partial_u a^0_{mn} \in \mathscr{W}^{1,\infty}(\Omega\times\mathbb{R})$, and that the coefficients $a^0_{mn}(x,s)$ are twice differentiable with respect to s, with the first and second order derivatives continuous and bounded on $\overline{\Omega}\times\mathbb{R}$, for all $m,n=1\dots d$. Then, there exist $r_0>0$ and $H_0>0$ such that, provided*

$$H \leq H_0 \quad \textit{and} \quad r_{HMM} \leq r_0, \tag{6.3.48}$$

any solution $u^{H,RB}$ of (6.2.10) *satisfies*

$$\|u^0 - u^{H,RB}\|_{\mathscr{H}^1(\Omega)} \quad \leq \quad C(H^\ell + r_{HMM}) \qquad \textit{if } \mu = 0,1$$

$$\|u^0 - u^{H,RB}\|_{\mathscr{L}^2(\Omega)} \quad \leq \quad C(H^{\ell+1} + r_{HMM}) \ \ \textit{if}\, \mu = 1,$$

where C is independent of $H, h, N, \mathscr{N}, \varepsilon$.

Proof. We apply Lemma 6.4.1 (a result from [26] stated in the Section 6.4 Appendix) with $\tilde{a}(x_{K_j}, s) = a^0_{\mathscr{N},K_j}(s)$ and $\tilde{u}^H = u^{H,RB}$. □

We next have to quantify the error r_{HMM} defined in (6.3.47) which can be decomposed as

$r_{HMM} \leq r_{MOD} + r_{MIC} + r_{RB}$, i.e. with the modeling, micro, and RB errors, respectively. These quantities are defined by

$$r_{MOD} := \sup_{K\in\mathscr{T}_H, x_{K_j}\in K} \|a^0(x_{K_j}, u^{H,RB}(x_{K_j})) - \overline{a}^0(x_{K_j}, u^{H,RB}(x_{K_j}))\|_{\mathscr{F}}, \tag{6.3.49}$$

$$r_{MIC} := \sup_{K\in\mathscr{T}_H, x_{K_j}\in K} \|\overline{a}^0(x_{K_j}, u^{H,RB}(x_{K_j})) - a^0_{N,K_j}(u^{H,RB}(x_{K_j}))\|_{\mathscr{F}}, \tag{6.3.50}$$

$$r_{RB} := \sup_{K\in\mathscr{T}_H, x_{K_j}\in K} \|a^0_{N,K_j}(u^{H,RB}(x_{K_j})) - a^0_{\mathscr{N},K_j}(u^{H,RB}(x_{K_j}))\|_{\mathscr{F}}. \tag{6.3.51}$$

To estimate the quantities r_{MIC} and r_{MOD}, we make the following smoothness and structure assumptions on the tensor:

(**HN1**) Given the degree q of the micro FE space $S^q(K_{\delta_j}, \mathscr{T}_h)$, the cell functions $\chi^{i,s}_{K_j}$ solution of (6.2.20) satisfy the bound $|\chi^{i,s}_{K_j}|_{\mathscr{H}^{q+1}(K_{\delta_j})} \leq C\varepsilon^{-q}\sqrt{|K_{\delta_j}|}$, with C independent of ε, the quadrature point x_{K_j}, the domain K_{δ_j}, and the parameter s for all $i = 1 \dots d$.

(**HN2**) for all $m, n = 1, \dots, d$, we assume $a^\varepsilon_{mn}(x, s) = a_{mn}(x, x/\varepsilon, s)$, where $a_{mn}(x, y, s)$ is y-periodic in Y, and the map $(x, s) \mapsto a_{mn}(x, \cdot, s)$ is Lipschitz continuous and bounded from $\overline{\Omega} \times \mathbb{R}$ into $\mathscr{W}^{1,\infty}_{per}(Y)$.

We next discuss the reduced basis error. Consider the space $\mathscr{M}^{\mathscr{N}}(Y)$ as defined in (6.2.29). We want to quantify how well $\mathscr{M}^{\mathscr{N}}(Y)$ can be approximated by the linear space $\overline{S}_N(Y)$ of dimension N. Such a quantification relies on the Kolmogorov N-width, see Definition 4.5.2.

In [48, Corollary 4.1] and [43], it is shown, for a class of symmetric linear uniformly coercive elliptic problems with continuity bound Λ and coercivity constant λ, that if the parametrized space of the RB algorithm has an exponentially small $N-$width, then the RB algorithm converges exponentially fast with respect to the dimension N of the RB space.

Notice that problem (6.2.25)-(6.2.28) with solution $(\hat{\chi}^{i,s}_{\mathscr{N},K_j}, \partial_s\hat{\chi}^{i,s}_{\mathscr{N},K_j}) \in \mathscr{W}(Y)^2$ is not coercive due to the nonlinearity of the tensor. Nevertheless, problem (6.2.25)-(6.2.28) still satisfies the following Céa inequality (see Lemma 6.4.3 in Section 6.4) in the Hilbert space $\mathscr{W}(Y)^2$ with norm

defined in (6.2.30),

$$\|(\hat{\chi}^{i,s}_{\mathcal{N},K_j},\partial_s\hat{\chi}^{i,s}_{\mathcal{N},K_j})-(\hat{\chi}^{i,s}_{l,K_j},\partial_s\hat{\chi}^{i,s}_{l,K_j})\|_{\mathcal{W}\times\mathcal{W}}\leq C_0\inf_{z\in\overline{S}_l(Y)}\|(\hat{\chi}^{i,s}_{\mathcal{N},K_j},\partial_s\hat{\chi}^{i,s}_{\mathcal{N},K_j})-z\|_{\mathcal{W}\times\mathcal{W}} \tag{6.3.52}$$

where we consider the solution $(\hat{\chi}^{i,s}_{l,K_j},\partial_s\hat{\chi}^{i,s}_{l,K_j})$ of (6.2.25)-(6.2.28) in the space $\overline{S}_l(Y)$ (i.e. taking test functions $(\hat{z}_l,\hat{\zeta}_l)\in\overline{S}_l(Y)$). The above constant is given by

$$C_0=\sqrt{\frac{\Lambda_1}{\lambda}\Big(3+\frac{8\Lambda_2^2}{\lambda^2}\Big)} \tag{6.3.53}$$

where $\lambda,\Lambda_1,\Lambda_2$ are the coercivity and continuity bounds (6.1.1),(6.2.26). In addition, recall the a posteriori estimate (6.2.39) of the form $C_{low}\Delta^i_{l,T_\delta}\leq\|(\hat{e}^{i,s}_{l,T_\delta},\partial_s\hat{e}^{i,s}_{l,T_\delta})\|_{\mathcal{W}\times\mathcal{W}}\leq C_{up}\Delta^i_{l,T_\delta}$, with

$$C_{low}=(2\Lambda_1+\Lambda_2)^{-1}\lambda_{LB}^{1/2},\qquad C_{up}=2\lambda_{LB}^{-1/2}+\Lambda_2\lambda_{LB}^{-3/2}. \tag{6.3.54}$$

We obtain the following result which states that the reduced basis method converges exponentially. This is a slight adaptation of the result in [48, Corollary 4.1] and [43].

Theorem 6.3.2. *In addition to* (6.1.1) *and* (6.2.26)*, assume that the parametrized cell solution space* $\mathcal{M}^{\mathcal{N}}$ *in* (6.2.29) *has an exponentially small Kolmogorov N-width,*

$$d_N(\mathcal{M}^{\mathcal{N}},\mathcal{W}(Y)^2)\leq Ce^{-rN},\quad \textit{with } r>\log((1+C_{up}/C_{low})C_0), \tag{6.3.55}$$

with constants in (6.3.52),(6.3.54)*. Then, there exists constants* $c,\kappa>0$ *independent of N such that*

$$\|\hat{\chi}^{i,s}_{N,K_j}-\hat{\chi}^{i,s}_{\mathcal{N},K_j}\|_{\mathcal{W}}\leq ce^{-\kappa N},\qquad \|\partial_s\hat{\chi}^{i,s}_{N,K_j}-\partial_s\hat{\chi}^{i,s}_{\mathcal{N},K_j}\|_{\mathcal{W}}\leq ce^{-\kappa N} \tag{6.3.56}$$

for all $K\in\mathcal{T}_H$ *and all* $x_{K_j}\in K$*, where* $\hat{\chi}^{i,s}_{\mathcal{N},K_j}$ *and* $\hat{\chi}^{i,s}_{N,K_j}$ *are the solutions of the cell problem* (6.2.25) *in* $S^q(Y,\mathcal{N})$ *and* $S_N(Y)$*, respectively, with corresponding test functions* $\hat{z}_{\mathcal{N}}\in S^q(Y,\mathcal{N})$ *and* $\hat{z}_N\in S_N(Y)$*.*

Proof. Inspecting the proof of [48, Corollary 4.1] reveals that the coercivity of the problem is not needed and the Céa inequality (6.3.52) is sufficient to obtain the exponential convergence (6.3.56) of the RB algorithm using the RB space $\overline{S}_N(Y)$ constructed in the Greedy algorithm 6.2.4. □

Theorem 6.3.3. *Consider* u^0 *the solution of problem* (6.1.2)*, and* $u^{H,RB}$ *the solution of* (6.2.10)*. In addition to the assumptions of Theorem 6.3.1, assume that* $u^{H,RB}(x_{K_j})\in(u^{0,low},u^{0,up})$*, for all* $K\in\mathcal{T}_H, x_{K_j}\in K$*. Assume further (***HN1***), (***HN2***), and* (6.3.55)*. Then, there exist* $H_0>0$ *and* $r_0>0$

such at if $H \le H_0$, $h/\varepsilon \le r_0$, *and* $ce^{-\kappa N} \le r_0$ *then for* $\mu = 0, 1$,

$$\|u^0 - u^{H,RB}\|_{\mathscr{H}^{1-\mu}(\Omega)} \le \begin{cases} C(H^{\ell+\mu} + (\frac{h}{\varepsilon})^{2q} + \delta) + r_{RB}, & \text{if } \mathscr{W} = \mathscr{W}^1_{per} \text{ and } \delta/\varepsilon \in \mathbb{N}, \\ C(H^{\ell+\mu} + (\frac{h}{\varepsilon})^{2q}) + r_{RB}, & \text{if } \mathscr{W} = \mathscr{W}^1_{per},\ \delta/\varepsilon \in \mathbb{N}, \text{ and } a^\varepsilon(x,s) \text{ is replaced by } a(x_{K_j}, \frac{x}{\varepsilon}, s) \text{ in (6.2.18), (6.2.11), (6.2.12), (6.2.20),} \\ C(H^{\ell+\mu} + (\frac{h}{\varepsilon})^{2q} + \delta + \frac{\varepsilon}{\delta}) + r_{RB}, & \text{if } \mathscr{W} = \mathscr{H}^1_0\ (\delta > \varepsilon), \end{cases}$$

where $r_{RB} \le \Lambda_1(ce^{-\kappa N})^2$ *with* Λ_1 *given in* (6.1.1). *We also assume* $\delta \le r_0$ *or* $\delta + \varepsilon/\delta \le r_0$ *in the first and third cases, respectively. We use the notation* $\mathscr{H}^0(\Omega) = \mathscr{L}^2(\Omega)$. *The constants* C *are independent of* $H, h, \varepsilon, \delta, N, \mathscr{N}$.

Proof. In view of Theorem 6.3.1, we estimate r_{HMM}. Using (**HN1**), the estimate $r_{MIC} \le C(h/\varepsilon)^{2q}$ follows from [1] (see also [4]). Using (**HN2**), The estimates $r_{MOD} \le C\delta$, $r_{MOD} = 0$, $r_{MOD} \le C(\delta + \varepsilon/\delta)$ follows from [22, 60]. Finally, the estimate $r_{RB} \le \Lambda_1(ce^{-\kappa N})^2$ follows from (6.3.56) and the identity (6.2.43). □

6.3.2 Uniqueness of the RB-FE-HMM solution and the Newton method

Consider the derivatives with respect to s of the exact and numerical homogenized tensors in (6.1.2) and (6.2.14). We define

$$r'_{HMM} := \sup_{K \in \mathscr{T}_H, x_{K_j} \in K} \left\| \partial_s a^0(x_{K_j}, u^{H,RB}(x_{K_j})) - \partial_s a^0_{\mathscr{N},K_j}(u^{H,RB}(x_{K_j})) \right\|_{\mathscr{F}}. \tag{6.3.57}$$

The proof of the uniqueness of the RB-FE-HMM solution relies on the following result which is an adaptation of Lemma 4.11 in [26].

Theorem 6.3.4. *Assume that the hypotheses of Theorem 6.3.1 and* (6.3.55) *hold. Then, there exist positive constants* H_0, r_0 *such that if*

$$H \le H_0 \qquad \text{and} \qquad H^{-1/2} r_{HMM} + r'_{HMM} \le r_0 \tag{6.3.58}$$

then the solution $u^{H,RB}$ *of* (6.2.10) *is unique.*

The proof of Theorem 6.3.4 relies on the convergence of the Newton method stated in the following lemma.

Lemma 6.3.5. *Assume that the hypotheses of Theorem 6.3.4 hold. Let* $u^{H,RB}$ *be a solution of* (6.2.10). *Then, there exists* $H_0, r_0, \nu > 0$, *such that provided a smallness assumption of the form* (6.3.58), *for all* $u_0^H \in S_0^\ell(\Omega, \mathscr{T}_H)$ *satisfying*

$$\sigma_H \|u_0^H - u^{H,RB}\|_{\mathscr{H}^1(\Omega)} \le \nu, \tag{6.3.59}$$

the sequence $\{u_k^H\}$ *of the Newton method* (6.2.17) *with initial value* u_0^H *is well defined and* $\|u_{k+1}^H - u^{H,RB}\|_{\mathscr{H}^1(\Omega)} \leq C\sigma_H \|u_k^H - u^{H,RB}\|^2_{\mathscr{H}^1(\Omega)}$, *where* C *is a constant independent of* $H, h, k, \mathscr{N}, N, \varepsilon$.

Proof. We apply Lemma 6.4.2 (a result from [26] stated in Section 6.4 Appendix) with $\tilde{a}(x_{K_j}, s) = a^0_{\mathscr{N},K_j}(s)$ and $\tilde{u}^H = u^{H,RB}$. □

Proof of Theorem 6.3.4. The proof is an immediate consequence of Lemma 6.3.5, where given two numerical solutions $u^{H,RB}, \tilde{u}^{H,RB}$ of (6.2.10), we apply the Newton method with the initial guess $u_0^H := \tilde{u}^{H,RB}$. The smallness assumption (6.3.58) together with the $\mathscr{H}^1$ a priori error estimate of Theorem 6.3.1 permits to satisfy the condition (6.3.59). □

For the estimation of r'_{HMM} in (6.3.57), we consider the decomposition

$$r'_{HMM} \leq r'_{MOD} + r'_{MIC} + r'_{RB}$$

where r'_{MOD}, r'_{MIC}, r'_{RB} are defined similarly to (6.3.49),(6.3.50),(6.3.51), respectively, with the exception that all the tensors are differentiated with respect to the s parameter. Hence hypothesis (**HN1'**) and (**HN2'**), similar to (**HN1**) and (**HN2**) but for $\partial_s \chi^{i,s}_{K_j}$ are needed. Following [23], the quantities r'_{MOD}, r'_{MIC} satisfy analogous estimates to those of r_{MOD}, r_{MIC}. It remains to estimate

$$r'_{RB} := \sup_{K\in\mathscr{T}_H, x_{K_j}\in K} \left\| \partial_s a^0_{N,K_j}(u^{H,RB}(x_{K_j})) - \partial_s a^0_{\mathscr{N},K_j}(u^{H,RB}(x_{K_j})) \right\|_{\mathscr{F}}.$$

This is done in the following lemma.

Lemma 6.3.6. *Assume that the hypotheses of Theorem 6.3.3 hold with a periodic coupling with* $\delta = \varepsilon$. *Assume further* (6.2.26), *and* (6.3.55). *Then, there exist constants* $c, \kappa > 0$ *such that*

$$r'_{RB} \leq (2\Lambda_1 + \Lambda_2)(ce^{-\kappa N})^2.$$

Proof. Differentiating (6.2.43) with respect to s and using Theorem 6.3.2 conclude the proof. □

Remark 6.3.7. *Notice that a similar a posteriori estimator as used here in the offline stage to control* r_{RB} *and* r'_{RB} *could be used to define a RB-FE-HMM for linear parabolic multiscale problems with a time-dependent tensor of the form* $\frac{\partial u^\varepsilon}{\partial t}(x,t) = \nabla\cdot(a^\varepsilon(x,t)\nabla u^\varepsilon(x,t)) + f(x,t)$ *as analysed in [24]. In this case, the micro problems would be parametrized by the location of the cell problem in the domain* Ω *and the time variable of the tensor* a^ε.

6.4 Some technical lemmas

The proof of Theorem 6.3.1 relies on the following lemma taken from [26] and based on the analysis for standard FEM with numerical quadrature from [25]. It is a reformulation of the statement of Theorem 3.1 in [26], its proof is thus omitted.

Lemma 6.4.1. *Consider u^0 the solution of problem* (6.1.2)*. Assume the assumptions of Theorem 6.3.1. Then, there exist $r_0 > 0$ and $H_0 > 0$ such that, for all tensor $\tilde{a}(x,s)$ satisfying* (6.1.3)*,* (6.1.1) *and continuous on $\overline{\Omega} \times \mathbb{R}$, for all $H \leq H_0$, and for all solution $\tilde{u}^H$ of the nonlinear FEM problem*

$$\sum_{K \in \mathcal{T}_H} \sum_{j=1}^{J} \omega_{K,j} \tilde{a}(x_{K_j}, \tilde{u}^H(x_{K_j})) \nabla \tilde{u}^H(x_{K_j}) \cdot \nabla w^H(x_{K_j}) = \int_\Omega f(x) w^H(x) dx, \quad \forall w^H \in S_0^\ell(\Omega, \mathcal{T}_H), \tag{6.4.60}$$

provided

$$Q_H := \sup_{K \in \mathcal{T}_H, x_{K_j} \in K} \left\| \tilde{a}(x_{K_j}, u^H(x_{K_j})) - a^0(x_{K_j}, u^H(x_{K_j})) \right\|_{\mathscr{F}} \leq r_0,$$

we have the $\mathscr{H}^1$ and $\mathscr{L}^2$ error estimates

$$\begin{aligned} \| u^0 - \tilde{u}^H \|_{\mathscr{H}^1(\Omega)} &\leq C(H^\ell + Q_H) \quad \text{if } \mu = 0, 1 \\ \| u^0 - \tilde{u}^H \|_{\mathscr{L}^2(\Omega)} &\leq C(H^{\ell+1} + Q_H) \quad \text{if } \mu = 1, \end{aligned}$$

where C is independent of H, Q_H and the tensor $\tilde{a}$.

The proof of Theorem 6.3.4 relies on the following result which is a reformulation of Lemma 4.11 in [26].

Lemma 6.4.2. *Assume that the hypotheses of Lemma 6.4.1 hold. Assume further that $\tilde{a}(x,s)$ is twice continuously differentiable with respect to s with derivatives continuous and bounded on $\overline{\Omega} \times \mathbb{R}$. Then, there exists $H_0, R_0, \nu > 0$, such that for*

$$Q_H \leq H \leq H_0, \quad Q'_H := \sup_{K \in \mathcal{T}_H, x_{K_j} \in K} \left\| \partial_s \tilde{a}(x_{K_j}, u^H(x_{K_j})) - \partial_s a^0(x_{K_j}, u^H(x_{K_j})) \right\|_{\mathscr{F}} \leq R_0,$$

for all $\tilde{u}^H$ solution of 6.4.1 and for all for all $u_0^H \in S_0^\ell(\Omega, \mathcal{T}_H)$ satisfying

$$\sigma_H \| u_0^H - \tilde{u}^H \|_{\mathscr{H}^1(\Omega)} \leq \nu,$$

the sequence $\{u_k^H\}$ of the Newton method (6.2.17) *applied to the problem* (6.4.60) *with initial value u_0^H is well defined and $\| u_{k+1}^H - \tilde{u}^H \|_{\mathscr{H}^1(\Omega)} \leq C\sigma_H \| u_k^H - \tilde{u}^H \|^2_{\mathscr{H}^1(\Omega)}$, where C is a constant independent of H, h, k.*

Finally, for the proof of Theorem 6.3.2, we need to prove the Céa inequality (6.3.52) for the problem (6.2.25)-(6.2.28).

Lemma 6.4.3. *Assume* (6.1.1) *and* (6.2.26)*. Then* (6.3.52) *holds with constant C_0 in* (6.3.53)*.*

Proof. We denote $\hat{e} = \hat{\chi}^{i,s}_{\mathcal{N},K_j} - \hat{\chi}^{i,s}_{l,K_j}$ and $\partial_s \hat{e} = \partial_s \hat{\chi}^{i,s}_{\mathcal{N},K_j} - \partial_s \hat{\chi}^{i,s}_{l,K_j}$. Considering the problem (6.2.25)-(6.2.28) with test functions in $S^q(Y,\mathcal{N})^2$ and $\overline{S}_l(Y)$, respectively, and subtracting, we deduce for

all $(z_l, \zeta_l) \in \overline{S}_l$,

$$b(\partial_s \hat{e}, \hat{\zeta}_l) = -\int_Y \partial_s a_{x_{K_j},s}(y) \nabla \hat{e}(y) \cdot \nabla \hat{\zeta}_l dy, \tag{6.4.61}$$

where the symmetric bilinear form $b(\cdot,\cdot)$ is defined in (6.2.21). Using

$$b(\partial_s \hat{e}, \partial_s \hat{e}) = b(\partial_s \hat{e}, \partial_s \hat{\chi}^{i,s}_{\mathcal{N},K_j} - \zeta_l) + b(\partial_s \hat{e}, \zeta_l - \partial_s \hat{\chi}^{i,s}_{l,K_j}),$$

we obtain from the Cauchy-Schwarz inequality and (6.4.61),(6.2.26),

$$b(\partial_s \hat{e}, \partial_s \hat{e}) \leq b(\partial_s \hat{e}, \partial_s \hat{e})^{1/2} b(\partial_s \hat{e} - \zeta_l, \partial_s \hat{e} - \zeta_l)^{1/2} + \Lambda_2 \|\hat{e}\|_{\mathcal{W}} (\|\partial_s \hat{\chi}^{i,s}_{\mathcal{N},K_j} - \zeta_l\|_{\mathcal{W}} + \|\partial_s \hat{e}\|_{\mathcal{W}}).$$

We deduce from the Young inequality and (6.1.1),

$$\lambda \|\partial_s \hat{e}\|^2_{\mathcal{W}} \leq b(\partial_s \hat{e}, \partial_s \hat{e}) \leq \Lambda_1 \|\partial_s \hat{\chi}^{i,s}_{\mathcal{N},K_j} - \zeta_l\|^2_{\mathcal{W}} + 2\Lambda_2 \|\hat{e}\|_{\mathcal{W}} \|\partial_s \hat{\chi}^{i,s}_{\mathcal{N},K_j} - \zeta_l\|_{\mathcal{W}} + 2\Lambda_2 \|\hat{e}\|_{\mathcal{W}} \|\partial_s \hat{e}\|_{\mathcal{W}}$$

Using again the Young inequality yields

$$\|(\hat{e}, \partial_s \hat{e})\|^2_{\mathcal{W}\times\mathcal{W}} \leq (1 + \frac{2\Lambda_1}{\lambda}) \|\partial_s \hat{\chi}^{i,s}_{\mathcal{N},K_j} - \zeta_l\|^2_{\mathcal{W}} + (1 + \frac{8\Lambda_2^2}{\lambda^2}) \|\hat{e}\|^2_{\mathcal{W}}.$$

Finally, the application of the Céa lemma to (6.2.25) yields $\|\hat{e}\|_{\mathcal{W}} \leq \sqrt{\frac{\Lambda_1}{\lambda}} \inf_{z_l \in S_l(Y)} \|\hat{\chi}^{i,s}_{\mathcal{N},K_j} - z_l\|_{\mathcal{W}}$ which permits to conclude the proof. □

7 Implentation issues for the quasilinear RB-FE-HMM

This chapter is taken from [10]. Here we discuss implementation issue for the RB-FE-HMM applied to quasilinear problems, We also discuss an extension of the method to parabolic problems. Finally we explain the construction of corrector function that can be used to obtain energy approximation of the fine scale problem. We consider quasilinear elliptic problems of the form

$$-\nabla\cdot\big(a^\varepsilon(x,u^\varepsilon(x))\nabla u^\varepsilon(x)\big)=f(x)\ \text{ in }\Omega,\qquad u^\varepsilon=0\ \text{ on }\partial\Omega, \tag{7.0.1}$$

or quasilinear parabolic problems of the form

$$\frac{\partial u^\varepsilon(x,t)}{\partial t}-\nabla\cdot\big(a^\varepsilon(x,u^\varepsilon(x,t))\nabla u^\varepsilon(x,t)\big)=f(x,t)\ \text{ in }\Omega\times[0,T],\quad u^\varepsilon=0\ \text{ on }\partial\Omega\times[0,T], \tag{7.0.2}$$

where a^ε is a nonlinear tensor that oscillates rapidly in space at the scale ε and a^ε satisfies the assumptions (6.1.1) and (6.2.26). As discussed in last chapter, there exists a homogenized equation for (7.0.1) or (7.0.2), i.e.

$$-\nabla\cdot\big(a^0\big(x,u^0(x)\big)\nabla u^0(x)\big)=f(x)\ \text{ in }\Omega,\qquad u^0=0\ \text{ on }\partial\Omega, \tag{7.0.3}$$

or

$$\frac{\partial u^0(x,t)}{\partial t}-\nabla\cdot\big(a^0\big(x,u^0(x,t)\big)\nabla u^0(x,t)\big)=f(x)\ \text{ in }\Omega,\qquad u^0=0\ \text{ on }\partial\Omega. \tag{7.0.4}$$

For the completeness of the algorithm description, we restate the following macro FEM for the quasilinear problem. Considering a FEM space $S_0^\ell(\Omega,\mathcal{T}_H)$ with mesh grid size H, such method approximates a solution u^H of

$$B_H(u^H;u^H,w^H)=F_H(w^H),\quad \forall\, w^H\in S_0^\ell(\Omega,\mathcal{T}_H), \tag{7.0.5}$$

by a sequence $u_k^H, k=0,1,2,3,\ldots$ and reads in weak form

$$\partial B_H(u_k^H;u_{k+1}^H-u_k^H,w^H)=F_H(w^H)-B_H(u_k^H;u_k^H,w^H),\quad \forall\, w^H\in S_0^\ell(\Omega,\mathcal{T}_H), \tag{7.0.6}$$

where $\partial B_H(z^H; v^H, w^H) := B_H(z^H; v^H, w^H) + B'_H(z^H; v^H, w^H)$. We used the notations

$$B_H(z^H; v^H, w^H) := \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \omega_{K_j} a^0(x_{K_j}, z^H(x_{K_j}))\, v^H(x_{K_j}) \nabla z^H(x_{K_j}) \cdot \nabla w^H(x_{K_j}), \tag{7.0.7}$$

$$B'_H(z^H; v^H, w^H) := \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \omega_{K_j} \partial_s a^0(x_{K_j}, z^H(x_{K_j}) v^H(x_{K_j}) \nabla z^H(x_{K_j}) \cdot \nabla w^H(x_{K_j}), \tag{7.0.8}$$

and $F_H(w^H)$ is an approximation of $\int_\Omega f w^H dx$.

This chapter is organized as follows. In Section 7.1, we present the offline procedure of the RB-FE-HMM and related implementation issues. Then, we describe in Section 7.2 how an online procedure, with a computational cost analogous to that of solving a one scale problem, permits to compute the homogenized solution u^0. In Section 7.3, we describe a reconstruction procedure for approximating the oscillatory solution u^ε at a negligible overcost. In Section 7.4, the performance of the algorithm is illustrated by several numerical examples.

7.1 Offline procedure

This section describes the offline stage of the RB-FE-HMM which permits to precompute the nonlinear homogenized tensor a^0 that is needed in the online stage to compute the homogenized solution u^0 itself.

Based on the mapping defined in (6.2.8), we can write oscillatory tensor as

$$a_{x_{K_j},s}(y) := a^\varepsilon(G_{x_{K_j}}(y), s). \tag{7.1.9}$$

Notice that if the tensor a^ε is locally periodic, i.e. has the form $a^\varepsilon(x,s) = a(x, x/\varepsilon, s)$ with periodicity with respect to the second argument x/ε, then we shall consider instead

$$a_{x_{K_j},s}(y) := a(x_{K_j}, G_{x_{K_j}}(y), s),$$

and if in addition the oscillatory period ε is known, we shall take $\delta = \varepsilon$ for the size of the sampling domains K_{δ_j}. We need to assume the following affine representation for simplicity (if not available, then the EIM introduced in Section 3.3 can be applied):

$$a_{x,s}(y) = \sum_{p=1}^{P} \Theta_p(x,s) a_p(y), \ \forall y \in Y.$$

For all the point x in Ω and all parameter s, we recall the linear cell problem: Find $\hat{\chi}^{i,s}_{\mathcal{N},x} \in$

$S^q(Y,\mathcal{N})$ [1] such that

$$b(\hat{\chi}^{i,s}_{\mathcal{N},x},\hat{z}_{\mathcal{N}}) = -\int_Y a_{x,s}(y)\mathbf{e}_i\cdot\nabla\hat{z}_{\mathcal{N}}(y)dy \ \ \forall \hat{z}_{\mathcal{N}}\in S^q(Y,\mathcal{N}), \tag{7.1.10}$$

where

$$b(\hat{v},\hat{z}) := \int_Y a_{x,s}(y)\nabla\hat{v}(y)\cdot\nabla\hat{z}(y)dy \ \ \forall \hat{v},\hat{z}\in\mathcal{W}^1_{per}(Y).$$

We consider the Sobolev space $\mathcal{W}^1_{per}(Y)$ [2], and still use notation $S^q(Y,\mathcal{N})$ for the micro FE space. Our aim is to compute a small set of representative basis functions $\hat{\xi}_{l,\mathcal{N}}, l=1,\dots,N$ that span the RB space

$$S_N(Y) := span\{\hat{\xi}_{1,\mathcal{N}},\dots,\hat{\xi}_{N,\mathcal{N}}\}. \tag{7.1.11}$$

The above *hatmap* notation is used to recall that these functions are defined over the reference domain Y.

A posteriori error estimator. We recall that the a posteriori error estimator for the parameter (x_τ,s,i) is defined as

$$\Delta^{i,s}_{l,T_\delta} := \frac{\|\bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}+\|\partial_s\bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}}{\sqrt{\lambda_{LB}}}, \tag{7.1.12}$$

where λ_{LB} with $\lambda_{LB}\le\lambda$ is an approximation of the coercivity constant λ defined in (6.1.1) (the detailed computation is introduced in Section 3.2). The residual $\partial_s\bar{e}^{i,s}_{l,T_\delta}$ in (7.1.12) is defined by differentiating with respect to s the FE problem (4.2.19). As discussed in last chapter, this a posteriori error estimator is designed to guarantee not only the convergence of the RB-FE-HMM, but also the Newton method convergence and the uniqueness of the numerical solution, as analyzed in [25] using the a posteriori error estimates for $i,j=1,\dots,d$, $l=1,\dots,N$,

$$C_1\Delta^i_{l,T_\delta}\le\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}+\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\le C_2\Delta^i_{l,T_\delta} \tag{7.1.13}$$

where C_1,C_2 depend on $\lambda,\Lambda_1,\Lambda_2,\lambda_{LB}$ in (6.1.1), (6.1.3). The above estimator can be computed as follows. In view of (4.2.19), considering the decomposition $\hat{\xi}^{i,s}_{l-1,T_\delta}=\sum_{j=1}^{l-1}\alpha_j(s)\hat{\xi}_{j,\mathcal{N}}$ in the RB space S_{l-1}, we compute following the implementation in [92, Section 4.4]

$$\bar{e}^{i,s}_{l,T_\delta} = \sum_{p=1}^{P}\Theta_p(x,s)\left(\sum_{j=1}^{l-1}\alpha_j(s)L_{p,j}-M_{p,i}\right) \tag{7.1.14}$$

[1] For the simplicity of presentation, we write $\hat{\chi}^{i,s}_{\mathcal{N},x}$ as the cell solution instead of $\hat{\chi}^{i,s}_{\mathcal{N},T_\delta}$ used in Chapter 6.

[2] Notice that $\mathcal{H}^1_0(Y)$ could also be considered for a coupling with Dirichlet boundary conditions and sampling domain size $\delta>\varepsilon$, see the review [14].

where $L_{p,j}, M_{p,i} \in S^q(Y,\mathcal{N}), p = 1,\dots,P, i = 1,\dots,d, j = 1,\dots,N$ are obtained by solving the FE projection problems

$$(L_{p,j},\hat{z}_{\mathcal{N}})_{\mathcal{W}} = \int_Y a_p(y)\nabla\hat{\xi}_{j,\mathcal{N}}(y)\cdot\nabla\hat{z}_{\mathcal{N}}(y)dy, \quad (M_{p,i},\hat{z}_{\mathcal{N}})_{\mathcal{W}} = \int_Y a_p(y)\nabla\mathbf{e}_i\cdot\nabla\hat{z}_{\mathcal{N}}(y)dy, \ \forall\hat{z}_{\mathcal{N}} \in S^q(Y,\mathcal{N}).$$

Notice that the functions $L_{p,j}, M_{p,i}$ are independent of the parameter x, s and thus computed only once. We refer to Remark 6.2.2 for the computation of $\partial_s\bar{e}^{i,s}_{l,T_\delta}$.

RB construction by the Greedy algorithm We now describe the construction procedure of the RB in (7.1.11) based on the a posteriori error estimator (7.1.12).

A first step is the guess of the range of nonlinear parameters $s = u^H(x_{K_j})$ involved in the nonlinear tensor $a(x,s)$. This is done using the following empirical algorithm motivated by the Voigt-Reiss inequality [74, Chapter 1.6],

$$\Big(\int_Y a_{x,s}(y)^{-1}dy\Big)^{-1} \le a^0(x,s) \le \int_Y a_{x,s}(y)dy, \tag{7.1.15}$$

for parameters x, s, and where $a^0(x,s)$ is the homogenized tensor. We then apply the following procedure to derive the training set $\mathscr{D}$, a compact of $\Omega\times\mathbb{R}$.

Algorithm 7.1.1 (Construction of the training set)**.**

1. *Using a standard one-scale FEM method, solve* (6.1.2) *where the homogenized tensor* $a^0(x,s)$ *is replaced alternatively with the tensors of the left and right-hand sides of* (7.1.15)*.*
2. *Set a maximum range* $(u^{0,low}, u^{0,up})$ *by taking the minimum and maximum values of the two solutions using theses easy to evaluate tensors.*
3. *Define the training set as* $\mathscr{D} = \Omega_\delta \times [u^{0,low} - \alpha, u^{0,up} + \alpha]$ *where* Ω_δ *is a compact subset of* Ω *such that for any* $x \in \Omega_\delta$ *the corresponding sampling domain centered at* x *is inscribed in* Ω*, and the range of parameter* s *is enlarged by* $\pm\alpha$ *(a safety factor of about* 10%*).*

Based on the training set $\mathscr{D}$, we can launch the greedy procedure to construct the basis functions. This procedure has been described in Algorithm 6.2.4 of Section 6.2.4. We just mention here that the following offline outputs are required for the online stage: the $N\times N$ stiffness matrices A_p and the length N vectors $F_{p,i}$, with $p = 1,\dots, P, i = 1,\dots,d$, defined as

$$(A_p)_{mn} = \int_Y a_p(y)\nabla\hat{\xi}_{n,\mathcal{N}}(y)\cdot\nabla\hat{\xi}_{m,\mathcal{N}}(y)dy, \qquad F_{p,i} = \int_Y a_p(y)\mathbf{e}_i\cdot\nabla\hat{\xi}_{m,\mathcal{N}}(y)dy.$$

We shall also need the matrix of averages

$$(G_p)_{mn} = \int_Y a_p(y)\mathbf{e}_m\cdot\mathbf{e}_n dy = \int_Y (a_p(y))_{mn}dy.$$

The above integrals on Y can be computed using a sufficiently accurate quadrature formula on the micro mesh $\mathscr{T}_{\hat{h}}$ of the micro FEM space $S^q(Y,\mathcal{N})$.

7.2 Online procedure

The online version of the RB-FE-HMM is identical to a standard FEM with numerical quadrature applied to a one-scale problem (6.1.2) using a Newton method. The only difference is that one has to approximate the homogenized tensor $a^0(x_{K_j}, s)$ for a given quadrature node x_{K_j} and a nonlinear parameter s. We explain how this can be performed at a negligible overcost using the outputs of the offline procedure.

In the case of a locally periodic tensor a^ε, the homogenized tensor coefficients given by the homogenization theory [44] can be written as

$$(a^0(x,s))_{mn} = \int_Y a_{x_{K_j},s}(y)\left(\hat{\chi}_x^{m,s}(y) + \mathbf{e}_m\right)\cdot \mathbf{e}_n dy, \quad m,n = 1,\dots,d,$$

where $\hat{\chi}_x^{m,s} \in \mathcal{W}^1_{per}(Y)$ is the solution of the cell problem (7.1.10) with test functions in $\mathcal{W}^1_{per}(Y)$. Analogously, we define the numerical homogenized tensor a_N^0 as

$$(a_N^0(x,s))_{mn} = \int_Y a_{x,s}(y)\left(\nabla \hat{\chi}_{N,x}^{m,s}(y) + \mathbf{e}_m\right)\cdot \mathbf{e}_n dy, \quad m,n = 1,\dots,d, \tag{7.2.16}$$

where $\hat{\chi}_{N,x}^{m,s} \in S_N(Y),\ m = 1,\dots,d$ is the solution of (7.1.10) in the RB space $S_N(Y)$, i.e. with test functions in $S_N(Y)$. This function can be decomposed in the RB space $S_N(Y)$ as

$$\hat{\chi}_{N,x}^{m,s}(y) = \sum_{j=1}^{N} \alpha_j^{(x,s,m)} \hat{\xi}_{j,\mathcal{N}}, \tag{7.2.17}$$

where $\boldsymbol{\alpha}^{(x,s,m)} = (\alpha_1^{(x,s,m)},\dots,\alpha_N^{(x,s,m)})^T$ is the vector of coordinates in this basis.

Elliptic case. Inspired by (7.2.16) and the decomposition (7.2.17), we are now in position to state the algorithm for the online procedure to compute the homogenized solution u^0 of the multiscale quasilinear problem (7.0.1).

Algorithm 7.2.1 (Online algorithm for elliptic problems (7.0.1))**.** *Given an initial guess $u_0^H \in S_0^\ell(\Omega, \mathcal{T}_H)$, compute the Newton method sequence $u_k^H, k = 1,2,3,\dots$ defined in* (7.0.6)*, where B_H and ∂B_H are defined similarly to* (7.0.7), (7.0.8)*, with the unknown homogenized tensor $a^0(x,s)$ replaced by the numerical tensor $a_N^0(x,s)$ and computed as*

$$(a_N^0(x,s))_{mn} = \sum_{p=1}^{P} \Theta_q(x,s)\left(\boldsymbol{\alpha}^{(x,s,m)}\cdot F_{p,n} + (G_p)_{mn}\right), \quad m,n = 1,\dots,d,$$

where the vectors $\boldsymbol{\alpha}^{(x,s,i)} \in \mathbb{R}^N$ with $i = 1,\dots,d$ are obtained by solving the $N \times N$ linear system

$$\left(\sum_{p=1}^{P} \Theta_p(x,s) A_p\right)\boldsymbol{\alpha}^{(x,s,i)} = -\sum_{p=1}^{P} \Theta_p(x,s) F_{p,i}. \tag{7.2.18}$$

The derivative $\partial_s a_N^0(x,s)$ of the numerical tensor involved in ∂B_H can be computed using a finite difference formula (analogously to $\partial_s \bar{e}_{l,T_\delta}^{i,s}$ in Remark 6.2.2).

It is shown in Chapter 6 that under suitable smoothness assumptions, the Newton method in the online algorithm 7.2.1 is well defined and converges for sufficiently fine mesh parameters H, h to the approximation u^H of u^0. This numerical solution u^H is uniquely defined. We observe that the a posteriori estimator (7.1.12) involving the derivative of the Riesz projection of the residual (4.2.17) is instrumental for establishing the convergence of the Newton method.

Parabolic case. We next explain the extension of the algorithm for solving parabolic problems (7.0.2). First, we note that for quasilinear parabolic problems, the effective tensor can be computed by solving time-independent (quasilinear elliptic) problems. Second, we observe that Algorithm 7.1.1 can be applied to obtain the training set by solving standard parabolic FE problems using the bounds in (7.1.15). Finally, we can use the Algorithm 6.2.4 for the Greedy procedure.

Notice that various time integrators can be used in the online stage of RB-FE-HMM, see e.g. [24] in the context of linear multiscale parabolic problems. Here, we use the linearized backward Euler (see [82] for details) and consider a constant time stepsize. Its advantage over the standard backward Euler scheme is that it permits to avoid Newton iterations.

Algorithm 7.2.2 (Online algorithm for parabolic problems (7.0.2))**.** *Given the time step Δt and $u_0^H \in S_0^\ell(\Omega, \mathcal{T}_H)$ associated to the initial condition, compute u_n^H, $n = 1,2,3,\ldots$ the solution at time $t = n\Delta t$, of the linear system*

$$\int_\Omega \frac{u_n^H - u_{n-1}^H}{\Delta t} v^H dx + B_H(u_{n-1}^H; u_n^H, v^H) = F_H(v^H),\ \forall v^H \in S_0^\ell(\Omega, \mathcal{T}_H), \tag{7.2.19}$$

where the form B_H is computed identically as is Algorithm 7.2.1.

7.3 Construction of a corrector

The online algorithm 7.2.1 permits to approximate the homogenized solution u^0 of the multiscale problem (7.0.1) as $\varepsilon \to 0$. To reconstruct the oscillatory solution u^ε of (7.0.1) for a fixed value of ε, and in turn obtaining an approximation of u^ε in the energy norm (i.e. $u_{corr}^\varepsilon \simeq u^0 + \varepsilon u^1$), a suitable corrector procedure is needed [42, 74] and can be obtained as follows. Consider the linear multiscale problem

$$-\nabla \cdot \left(a^\varepsilon(x, u^0)\nabla v^\varepsilon(x)\right) = f(x) \text{ in } \Omega, \qquad v^\varepsilon = 0 \text{ on } \partial\Omega,$$

where compared to (7.0.1), the tensor is evaluated at u^0 instead of u^ε. It is shown in [44, Section 3.4.2] that a corrector u_{corr} for the above linear problem is also a corrector for the solution u^ε of the nonlinear problem (7.0.1). This permits to define the following numerical corrector for the approximation of u^ε similarly to the linear case.

Nonlinear corrector. Given a quadrature point $x_{K_j} \in K \in \mathcal{T}_H$, we consider the inverse map $\widetilde{G}_{x_{K_j}} := G_{x_{K_j}}^{-1} : T_{\delta_j} \to Y$ where $G_{x_{K_j}}$ is defined in (6.2.8). We next extend this map periodically on the macro element K by setting $\widetilde{G}_{x_{K_j}}(x+\delta \boldsymbol{k}) = \widetilde{G}_{x_{K_j}}(x)$ for all $\boldsymbol{k} \in \mathbb{Z}^d$, $x \in T_{\delta_j}$. We next define for all $x \in K$,

$$u^{H,\varepsilon}_{corr}(x) := u^H(x) + \varepsilon \hat{\chi}^{m,s}_{N,x_0}\big(\widetilde{G}_{x_{K_j}}(x)\big) \simeq u^\varepsilon(x),$$

where $u^H \in S^\ell_0(\Omega, \mathcal{T}_H)$ is the output of the online algorithm 7.2.1 and the functions $\hat{\chi}^{m,s}_{N,x_0} \in S_N(Y)$ can be evaluated using the RB decomposition (7.2.17) with the RB produced by the Greedy algorithm 6.2.4. Notice that the coefficients $\alpha^{(x,s,m)}_j$ in (7.2.17) are already computed (7.2.18) and need not to be recomputed.

7.4 Numerical experiments

We illustrate the performances of the proposed implementation of RB-FE-HMM on various elliptic and parabolic multiscale problems in 2D and 3D. We compare the algorithm with the standard nonlinear FE-HMM code [26] designed without the reduced basis technique, and based on the implementation in [19] in the context of linear problems. We consider here (non-parallel[3]) implementations in Matlab on a desktop computer.

7.4.1 A simple illustrative example

We consider the model problem (7.0.1) in $\Omega = [0,1]^2$ with $f = 50e^{(x_1-0.2)^2+(x_2-0.3)^2}$, and the following mixed boundary conditions,

$$\begin{aligned} &u^\varepsilon(x) = 2x_1^2(x_1-1)^2 + 3x_2^2(x_2-1)^2 + 1 \quad \text{on } \{x_1=0\} \cup \{x_1=1\}, \\ &n \cdot (a^\varepsilon(x,u^\varepsilon(x))\nabla u^\varepsilon(x)) = 0 \quad \text{on } \{x_2=0\} \cup \{x_2=1\}. \end{aligned} \tag{7.4.20}$$

Consider a diagonal multiscale tensor with the following affine expression [4]

$$\begin{aligned} a^\varepsilon(x,s)_{11} &= (x_1^2+0.2) + (x_2\sin(s\pi)+2)(\sin(2\pi\frac{x_1}{\varepsilon})+2), \\ a^\varepsilon(x,s)_{22} &= (\frac{1}{s+1}e^{x_2}+0.05) + (x_1x_2+1)(\sin(2\pi\frac{x_2}{\varepsilon})+2). \end{aligned} \tag{7.4.21}$$

Using the homogenization theory [74], the corresponding homogenized tensor is also diagonal with entries given by the harmonic averages

$$a^0_{ii} = \Big(\int_Y a(x,y;s)^{-1}dy\Big)^{-1},\ i=1,2. \tag{7.4.22}$$

[3] Notice that the linear systems (7.2.18) involving different parameters x, s are independent and can be solved in parallel.

[4] Recall that since the RB-FE-HMM computes the solution of the effective problem as $\varepsilon \to 0$, the actual value of ε is not needed in the algorithm.

Offline stage. In the offline stage, we set the parameter space to be $\mathscr{D} = \Omega_i \times U$, where Ω_i is a closed subset of Ω such that $\bar{T}_\delta = x_\tau + [-\delta/2, \delta/2]^d \subset \bar{\Omega}$ for all $\tau \in \Omega_i$, and U is a closed bounded interval of $\mathbb{R}$ (an estimation of the range for u^0). In order to obtain U, motivated by (7.1.15), we first solve (7.0.1) on a coarse 8×8 macro mesh by replacing the homogenized tensor respectively with the arithmetic and harmonic averages of the multiscale tensor. The ranges of the corresponding solutions are shown in Table 7.1 and we choose $U = [0.9, 3.66]$ adding a safety correction.

Table 7.1: A priori estimate for the solution range. Mesh size = 8×8.

tensor type	solution range
$\int_Y a(x,y;s)dy$	[1, 3.14]
$(\int_Y a(x,y;s)^{-1}dy)^{-1}$	[1, 3.56]

For the RB offline stage, we propose in Section 6.2.4 a new a posteriori error estimator (7.1.12) in order to guarantee the convergence of the Newton method. We will also check the computational overhead of this new estimator compared to the (standard) a posteriori estimator

$$\tilde{\Delta}^{i,s}_{l,T_\delta} := \frac{\|\bar{e}^{i,s}_{l,T_\delta}\|_{\mathscr{W}}}{\sqrt{\lambda_{LB}}} \tag{7.4.23}$$

used for linear problems in Chapter 4. The offline parameters are collected in Table 7.2 and the comparison of the two estimators are shown in Table 7.3.

We observe in this test that the offline outputs using Δ_{l,T_δ} have only one additional basis function.

Online stage: convergence rates for the P1 and P2 RB-FE-HMM. Using the computed offline outputs (obtained by via Δ_{l,T_δ} as the offline estimator), we consider a P1 FEM and a P2 FEM for the online stage. The reference solution $u^{ref} \approx u^0$ is obtained using the P2 FEM with a 1024×1024 uniform mesh.

By the a priori estimates of Theorems 6.3.1-6.3.3 and the mesh size used for the offline computation, we have the bound $\Delta^2_{l,T_\delta} = \mathscr{O}(tol_{RB}) \sim 10^{-10}$ for r_{RB} and $r_{MIC} \sim 10^{-7}$. As we choose sampling domains with size $\delta = \varepsilon$ with periodic boundary condition we have $r_{MOD} = 0$ and we expect $r_{HMM} \sim 10^{-7}$. We observe in Fig. 1 the expected convergence rates with respect to the macro mesh.

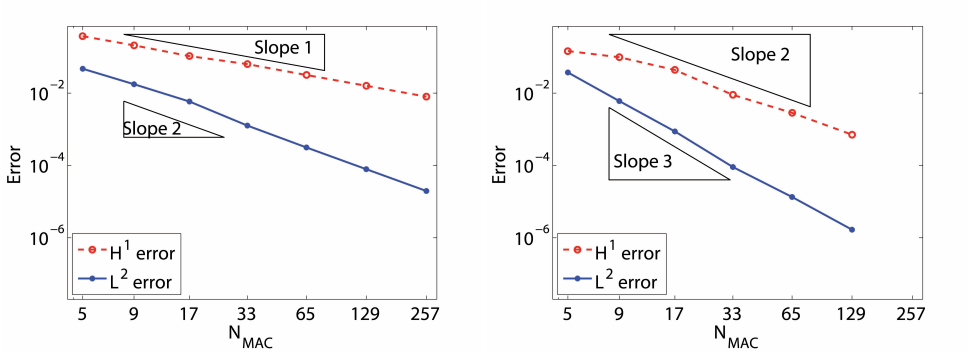
RB-FE-HMM v.s FE-HMM. In this test, we compare the efficiency and accuracy between the

Table 7.2: Offline parameters

Parameter space	$[0,1]^2 \times [0.9, 3.66]$
Training set size	4400
Solver	P1 FEM
Mesh	1500×1500
tol_{RB}	1e-10

Table 7.3: A posteriori estimators

	Δ_{l,T_δ}	$\tilde{\Delta}_{l,T_\delta}$
Basis number	9	8
Offline CPU time	1300	1100



(a) P1 RB-FE-HMM (b) P2 RB-FE-HMM

Figure 1: Test problem (7.0.1)-(7.4.20)-(7.4.21). The errors $\|u^{H,RB} - u^{ref}\|_{\mathscr{L}^2(\Omega)}$ and $\|u^{H,RB} - u^{ref}\|_{\mathscr{H}^1(\Omega)}$ versus $N_{MAC} = 1/H$ for the P1 RB-FE-HMM and the P2 RB-FE-HMM, respectively.

P1 RB-FE-HMM and the P1 FE-HMM. For the P1 FE-HMM, we set $N_{MIC} = N_{MAC}$ for each refinement step ($\mathscr{L}^2$ refinement strategy), where N_{MAC} and N_{MIC} are the numbers of macro and micro DOF in one space direction, respectively. We can see in Table 7.4 that the $\mathscr{H}^1$ and $\mathscr{L}^2$ errors for the two methods decay with the same rates which is consistent with the a priori estimates. However, the RB-FE-HMM has a considerably reduced computational cost for fine meshes (up to two orders of magnitude in this example). Next we present in Table 7.5 a comparison that takes

Table 7.4: Comparison between the RB-FE-HMM and the FE-HMM.

	RE-FE-HMM			FE-HMM		
DOF	$\mathscr{H}^1$ error	$\mathscr{L}^2$ error	online time (s)	$\mathscr{H}^1$ error	$\mathscr{L}^2$ error	time cost (s)
5×5	0.3727	0.0471	0.08	0.3724	0.0481	0.26
9×9	0.2086	0.0176	0.23	0.2082	0.0167	1.59
17×17	0.1053	0.0058	0.90	0.1052	0.0056	11.49
33×33	0.0632	0.0013	3.82	0.0631	0.0012	160.20
65×65	0.0316	3.15e-04	19.83	0.0316	3.03e-04	2802.68
129×129	0.0159	7.89e-05	146.75	0.0159	7.61e-05	49260.89

into account the computational overhead from the offline stage for the RB-FE-HMM. Here we have $t_{RB}^{offline} = 1300$s and we denote the total offline and online time by t_{RB}^{total}. We see that the FE-HMM is still significantly more expensive except for coarse macroscopic meshes. This indicates that even for one computation, the RB-FE-HMM can provide an important computational speed-up.

Table 7.5: CPU comparison between the RB-FE-HMM and the FE-HMM.

DOF	$t_{RB}^{online}/t_{FE-HMM}$	$t_{RB}^{total}/t_{FE-HMM}$
33×33	2.38%	813.87%
65×65	0.69%	47%
129×129	0.3%	2.9%

Table 7.6: Offline settings and outputs for the time dependent problem (7.0.2) with boundary conditions (7.4.24).

Parameter space	$[0,1]^2 \times [0.9, 3.93]$
Train set size	4400
Mesh	1500×1500
tol_{RB}	1e-10
RB Basis number	8

7.4.2 A 2D time dependent problem

We consider a time dependent test problem of the form (7.0.2). We choose an example where a reference solution can be easily computed to check the accuracy of the method, and for simplicity we still take the affine tensor (7.4.21) defined on domain $\Omega = [0,1]^2$ and we choose the boundary conditions

$$\begin{aligned} u^\varepsilon(x,t) &= 1, \quad \text{on } \{x_1 = 0\} \cup \{x_1 = 1\}, \\ n \cdot (a^\varepsilon(x, u^\varepsilon(x,t)) \nabla u^\varepsilon(x,t)) &= 0, \quad \text{on } \{x_2 = 0\} \cup \{x_2 = 1\}. \end{aligned} \tag{7.4.24}$$

We set $f(x) = 50e^{(x_1-0.2)^2+(x_2-0.3)^2}$ for the right hand side function, $u^\varepsilon(x,0) = 1$ as the initial condition, and consider the time interval $[0,T] = [0,0.5]$.

We apply the offline stage as described in Algorithm 6.2.4 and obtain the RB with 8 functions. The offline settings and outputs are presented in Table 7.6.

In this computation, we set the time step $\Delta t = 0.001$ (so that the corresponding $\mathcal{O}(\Delta t)$ error is negligible compared to the spatial discretization). A reference solution u^{ref} is computed by using the standard FEM with piecewise linear FEs applied to the homogenized problem with tensor a^0, with a mesh grid 257×257 and the linearized backward Euler scheme as time integrator. We can see from Fig. 2 that we obtain the usual convergence rates for the FEM applied to parabolic problems. Indeed, $\|u^H - u^{ref}\|_{\mathscr{L}^2([0,T];\mathscr{H}^1(\Omega))}$ decreases with rate $\mathcal{O}(H)$ and $\|u^H - u^{ref}\|_{\mathscr{L}^\infty([0,T];\mathscr{L}^2(\Omega))}$ decreases with rate $\mathcal{O}(H^2)$.

We list the online CPU times in Table 7.7 for various macro meshes of the square domain Ω.

We tested that the CPU time cost for the FE-HMM with $N_{MAC} = N_{MIC} = 65$ for one time step is 280s. As we have $T/\Delta t = 500$ time steps, the total CPU time for the FE-HMM without reduced basis would be about 39 hours. In contrast, the RB-FE-HMM CPU time is only 446s, i.e. 300 times less.

The total CPU time of the RB-FE-HMM (including the offline procedure) is only 1.3% of the FE-HMM cost.

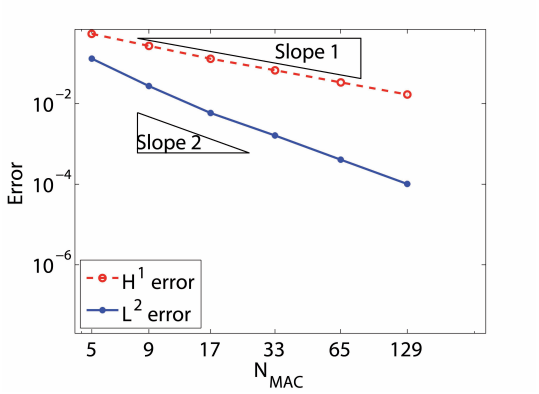


Figure 2: Time dependent problem (7.0.2), (7.4.24) with $T = 0.5$. We plot the errors $\|u^H - u^{ref}\|_{\mathscr{L}^\infty([0,T];\mathscr{L}^2(\Omega))}$ and $\|u^H - u^{ref}\|_{\mathscr{L}^2([0,T];\mathscr{H}^1(\Omega))}$ versus $N_{MAC} = 1/H$ for the RB-FE-HMM.

Table 7.7: Online CPU times for the parabolic test problem (7.0.2), (7.4.24), $T = 0.5, \Delta t = 0.001$.

Macro mesh	5×5	9×9	17×17	33×33	65×65
Online CPU time	14.87	17.53	19.49	26.46	445.93

7.4.3 Stationary Richards problem

We consider the stationary Richards equation for describing the fluid pressure in an unsaturated porous media

$$-\nabla\cdot(K^\varepsilon(x,u^\varepsilon(x))\nabla(u^\varepsilon(x)-x_2)) = f(x) \text{ in } [0,1]^2. \tag{7.4.25}$$

with a nonlinear permeability tensor $K^\varepsilon(s)$ similar to the one in [50, Section 5.1] written as

$$K^\varepsilon(x,s) = (200\alpha^\varepsilon e^{-(s-2-x_2)^2\alpha^\varepsilon(x)} + (x_1-0.3)^2 + x_2^2 + 2)I,\ \alpha^\varepsilon(x) = \frac{0.005}{2+1.8\sin(2\pi\frac{x_2}{\varepsilon}-6\pi\frac{x_1}{\varepsilon})}.$$

Notice that this problem can be cast in the form (7.0.1) by using the change of variable $v^\varepsilon(x) = u^\varepsilon(x) - x_2$. We set $f = 1$ and consider Dirichlet conditions on the top boundary of the domain and Neumann conditions on the rest of boundaries, that is

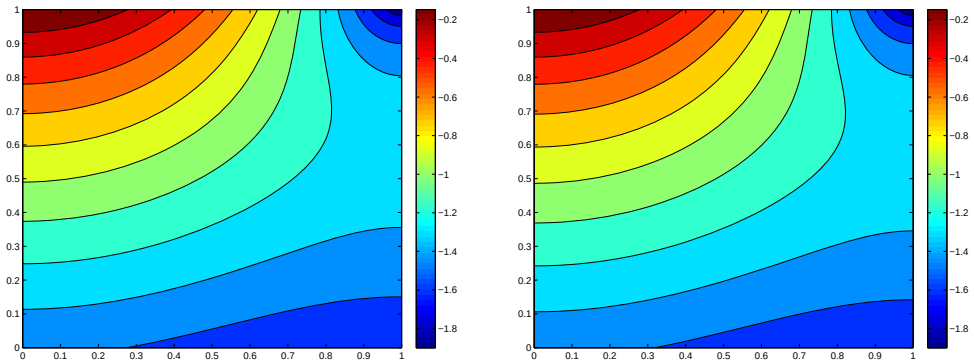
$$\begin{aligned} &u^\varepsilon(x) = 1 - 1.9x_1^2, \text{ on } [0,1]\times\{1\},\\ &n\cdot\left(K^\varepsilon(x,u^\varepsilon(x))\nabla(u^\varepsilon(x)-x_2)\right) = 0, \text{ on } [0,1]\times\{0\}\cup\{0,1\}\times[0,1]. \end{aligned}$$

Offline stage. The parameters are given in Table 7.8. We determine an a priori range for the homogenized solution $U = [0.9, 3.93]$. As the permeability tensor K^ε does not have an affine representation (6.2.19), we need to apply the EIM, which introduces another error term r_{EIM} in r_{HMM}, as discussed in [7] (this term can be controlled by the prescribed tolerance tol_{EIM} [37]).

Table 7.8: RB offline settings and outputs for the nonaffine test problem (7.4.25).

Parameter space	$[0,1]^2 \times [-3.1,-0.8]$
Training set size	4400
Mesh	1000×1000
tol_{EIM}	1e-6
EIM basis number	5
EIM CPU time	550.19s
tol_{RB}	1e-9
Solver	P1-FEM
RB Basis number	4
Offline CPU time	912.73s

Online stage. We plot in Fig. 3 the online solution $u^{H,RB}$ on a uniform 65×65 macro mesh for the P1 RB-FE-HMM (left picture) and the FE-HMM (right picture). We observe that the two solutions are very similar as expected.



(a) P1 RB-FE-HMM (b) P1 FE-HMM

Figure 3: Richards stationary problem (7.4.25). The RB-FE-HMM solution and the FE-HMM solution on a 65×65 macro mesh.

We notice that the range of $u^{H,RB}$ is $[-2.9,-1]$ which safely lies in our a priori range $[-3.1,-0.8]$. Therefore, the offline outputs can be successfully used for the online computation. In Table 7.9, we present the errors of the Newton method iterations and the corresponding CPU times. Due to the quadratic convergence rate of the Newton method, only four iterations are needed in all considered cases to reach the machine precision.

7.4.4 Heat transfer in a 3D rotor

To illustrate the performances of the RB-FE-HMM, we consider in this section a simplified model in 3D inspired from [102] of a car brake rotor in a stationary regime (see left picture of Fig. 4). We focus on the heat propagation in a part of the rotor as the break system is activated. The geometry of this part is represented in Fig. 4 (right picture). Its external and internal radius

Table 7.9: Richards stationary problem (7.4.25). Online CPU times and Newton iteration errors for the RB-FE-HMM.

DOF	CPU time	Iteration 1 err	Iteration 2 err	Iteration 3 err	Iteration 4 err
5×5	0.15	2.78	0.0053	4.87e-8	1.78e-15
9×9	0.2	2.80	0.0047	5.76e-8	5.77e-15
17×17	0.8	2.81	0.0046	5.78-8	1.46e-14
33×33	3.13	2.82	0.0045	5.71e-8	1.42e-14
65×65	14.72	2.82	0.0045	5.70e-8	3.02e-14
129×129	55.56	2.82	0.0045	5.70e-8	4.39e-14

are given respectively by $R = 15$cm and $r = 8$cm and its thickness is 5cm. We assume that the

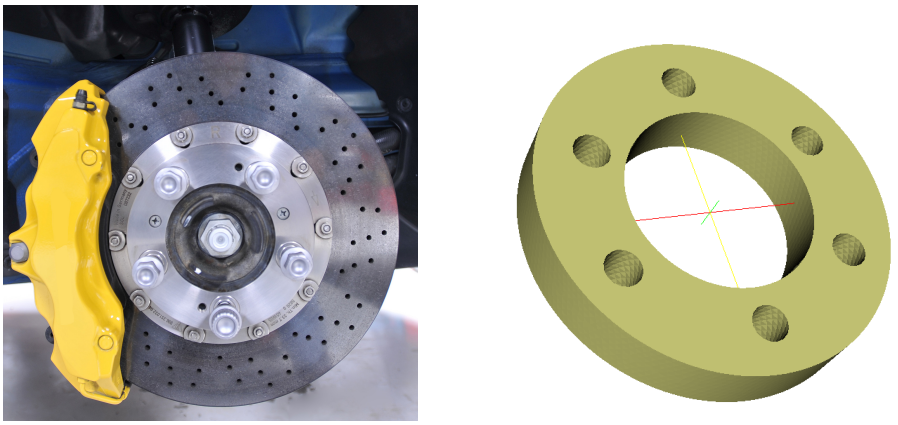


Figure 4: Left picture: example of a rotor break system in a car. Right picture: the geometry of the brake rotor part considered as the computational domain Ω.

material has a composite structure with a microscopic characteristic length ε. We consider a nonlinear heat propagation modeled by problem (7.0.1) with heat source $f = 0$ where $u^\varepsilon(x)$ denotes the temperature (in Kelvin degrees) as a function of x in the 3D computational domain. The nonlinear multiscale conductivity tensor is assumed diagonal with entries given by

$$\begin{aligned} a^\varepsilon_{11}(x,s) &= \alpha(s)\big(\cos(4\pi\frac{x_3}{\varepsilon})+2\big)+\beta(x)\big(\sin(2\pi\frac{x_1}{\varepsilon})+2\big), \\ a^\varepsilon_{22}(x,s) &= \alpha(s)\big(\cos(4\pi\frac{x_1}{\varepsilon})+2\big)+\beta(x)\big(\sin(2\pi\frac{x_2}{\varepsilon})+2\big), \\ a^\varepsilon_{33}(x,s) &= \alpha(s)\big(\cos(4\pi\frac{x_2}{\varepsilon})+2\big)+\beta(x)\big(\sin(2\pi\frac{x_3}{\varepsilon})+2\big), \end{aligned}$$

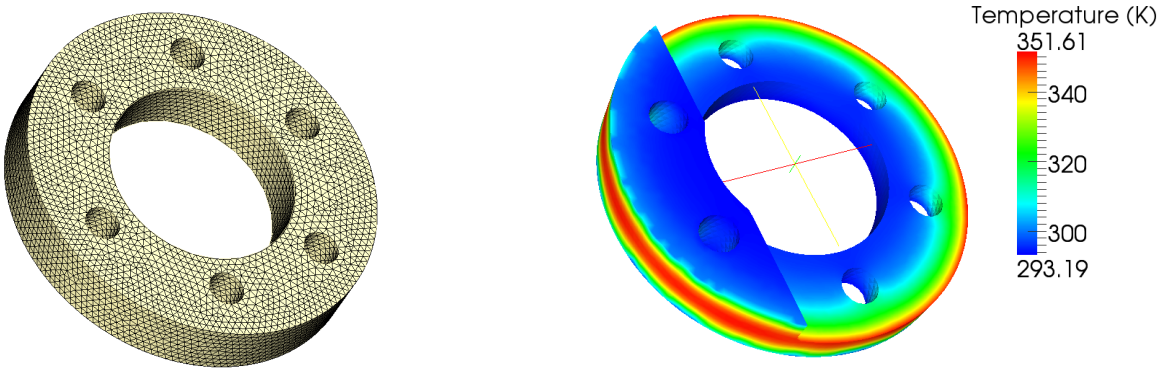
where $\alpha(s) = 0.01\big(e^{-(s-300)^2}+0.5\big)$, $\beta(x) = \big(0.001(x_1^2+x_2^2+x_3^2)\big)^{1/2}+0.005$.

Our aim here is to show the capability of our nonlinear algorithm on a realistic 3D geometry and we therefore choose to model the microstructure by ad-hoc oscillatory tensors. We note that for realistic composite materials these conductivity tensors could be obtained via imaging techniques. We next describe the boundary conditions that are considered in the model. The

external circular boundary surface (diameter R) of the considered rotor part is attached to a brake plate (see left picture in Fig. 4). It permits to slow down and stop the disk rotation by the friction due to the brake pads pushing against this brake plate (yellow part). The heat flux arising from the braking is modeled by a Neumann boundary condition on this external cylindrical boundary

$$-\nabla\cdot\left(a^{\varepsilon}(x,u^{\varepsilon}(x))\nabla u^{\varepsilon}(x)\right)\cdot n = g_B \tag{7.4.26}$$

where n is the external unit vector normal to the boundary and we set the surface power $g_B = 75 W m^{-2}$.



(a) (b)

Figure 5: RB-FE-HMM macro solution u^H of the 3D brake rotor elliptic problem.

The six small holes in the computational domain (see Fig. 4) as well as the main hole in the center are normally filled with screws or other components (not considered here) and we choose for simplicity homogeneous Neumann boundary conditions at these interfaces (we thus use (7.4.26) with g_B replaced by zero). The rest of the boundary is in contact with air at temperature $u^{ext} = 293.15\,K$. The corresponding convective heat transfer is modeled by a Robin boundary condition and we use (7.4.26) with g_B replaced by $-\alpha(u^{\varepsilon}(x) - u^{ext})$ and set $\alpha = 10 W m^{-2} K^{-1}$.

Table 7.10: RB offline pre-process for the 3D rotor problem. $DOF = 17804$.

Tensor type	$\min \bar{u}^0_H$	$\max \bar{u}^0_H$
$< a(x,y;s) >_Y$	293.2	348.7
$< a(x,y;s)^{-1} >_Y^{-1}$	293.2	353.7

We apply the RB-FE-HMM to this 3D problem and we collect the offline parameters and in Table 7.11. For the online stage, we consider a macro mesh of the computational domain Ω with about

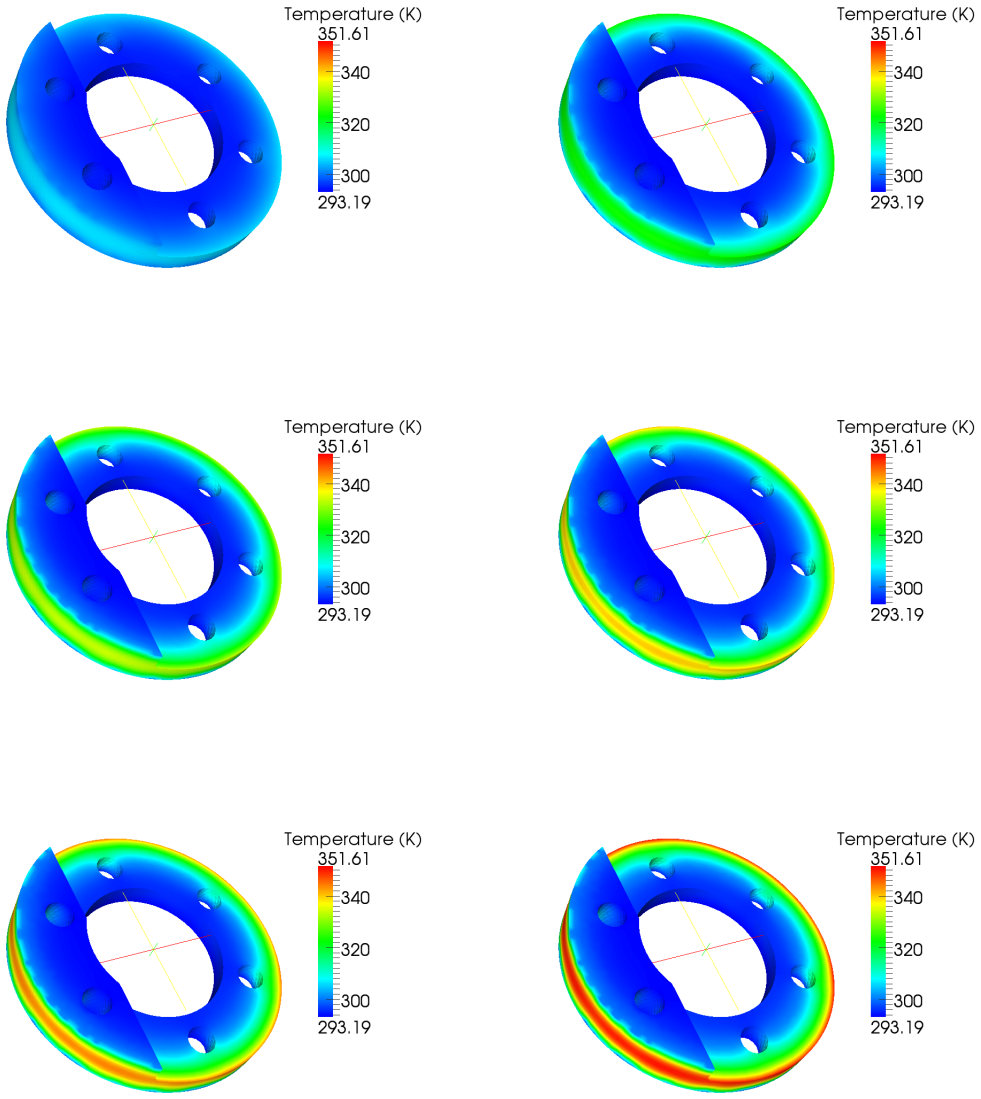


Figure 6: RB-FE-HMM macro solution of the 3D brake rotor parabolic model at times $t = 0.6, 2.4, 4.2, 6.0, 7.8, 50$ (respectively from left to right and top to bottom).

Table 7.11: Offline parameters and outputs for the 3D rotor problem.

Parameter space	$\Omega \times [288, 360]$
Training set size	4000
Number of tetrahedra	29478000
tol_{RB}	1e-8
RB Basis number	6

90000 tetrahedra generated by Cubit13.2 [98], see Fig. 5 (a). The CPU time for the online stage of the nonlinear elliptic RB-FE-HMM is 96.4s for 6 Newton iterations. In Fig. 5 (b), we plot the computed temperature distribution on the rotor (a cut parallel to the top surface is displayed to see the internal heat distribution). As we can see, the range of the computed temperature is $[293.85, 349.2]$ which lies in our training set based on the data in Table 7.10.

We finally consider the parabolic model of the form (7.0.2) where we use the same boundary conditions as above and we consider the initial temperature $u^\varepsilon(x,0) = u^{ext} = 293.15\,K$ and consider the time interval $[0,T] = [0,50]$. The RB produced by the offline procedure of the elliptic problem can be reused for this time evolutionary problem because the range of temperature in Table 7.10 remains valid. The CPU computational time for the parabolic online stage is $8\,s$ per time step, using a constant time step $\Delta t = 0.2$. We plot in Figure 6 the RB-FE-HMM solutions obtained by the parabolic online procedure. We observe that the heat propagates from the exterior to the interior of the rotor and approaches the steady solution (Figure 5).

7.5 Discussion

We have presented an efficient implementation of the RB-FE-HMM algorithm for the resolution of elliptic or parabolic nonlinear multiscale problems at the cost of single-scale nonlinear problems. This proposed algorithm combines a numerical homogenization strategy (the FE-HMM) with a model reduction strategy (the RB method). The accuracy of the model reduction strategy and the outputs of interest are controlled by an appropriate a posteriori error estimator for nonlinear problems. A remarkable feature of the offline-online strategy is that the computational cost of the method is independent of the smallness of the oscillatory parameter ε and the oscillatory solution u^ε can be reconstructed from the homogenized one u^0 with negligible overhead. The efficiency of the proposed algorithm has been illustrated on two and three dimensional time dependent nonlinear problems on a non-trivial computational domain.

8 Conclusion and Outlook

8.1 Conclusion

In this thesis we have first provided a detailed analysis for the generalized FE-HMM applied to elliptic multiscale problems with a finite number of well separated scales. Then, we have proposed a new multiscale method that combines the FE-HMM with reduced order modeling techniques such as the reduced basis. This method has been shown to be significantly more efficient than the FE-HMM, in particular for high-dimensional problems or high order simulations.

As previous a priori error analysis for the FE-HMM was restricted to two-scale problems, we proposed and analyzed a generalized FE-HMM for an arbitrary number of well-separated scales. In our analysis, both the FE error and the quadrature error are considered and detailed estimates for various quadrature formulas are presented. Due to the hierarchic structure of the algorithm and the cascade of inter-related cell problems, the cost of this method can be prohibitive for complex problems. Nevertheless, the a priori error analysis provides a theoretical foundation for future research in combining with the generalized FE-HMM with reduced order modeling techniques.

For two-scale FE-HMM, the efficiency issue arising from the large number of repeated micro computations has been addressed by our new method, based on an offline-online strategy. In the offline stage, several representative micro cell problems are selected by an a posteriori estimator that controls the reliability of the offline output to form the microscopic reduced basis. In the online stage, this microscopic reduced basis is then used to assemble the FE-HMM. The online computational time cost can be compared to the single scale FEM for the macro triangulation.

Thanks to the fact that the construction of the RB space is independent of the macro solver, the same RB space can be coupled with macro solvers of different orders or iterative macro solvers. This has been exploited to extend the RB-FE-HMM in two directions. In the first direction, we designed the adaptive RB-FE-HMM that can be used to deal with complex and real engineering problems. We considered two different adaptive a posteriori error estimators: the energy norm based estimator and the goal oriented estimator. In the second direction, we extended the RB-FE-HMM for quasilinear multiscale problems. For such problems, the computation of new

micro problems at each iteration of the macro solver (needed for the solution of the nonlinear equation) needed with classical numerical homogenization methods, can be avoided with the RB-FE-HMM. Therefore the use of RB-FE-HMM largely reduces the computing time cost of the FE-HMM. Of special interest is the design of a new a posteriori error estimator for the RB offline stage which does not only controls the precision of the offline output but also guarantees the convergence of the Newton method used in the online stage.

In each part of this thesis, we have presented extensive numerical examples to corroborate our numerical analysis and illustrate large efficiency improvement compared to the FE-HMM.

8.2 Outlook

For the future work, we would like to extend the RB-FE-HMM for different problems. Based on the numerical methods proposed in this thesis, various research topics could be explored that we summarize in the following.

N+1 scale problems. As mentioned in the Chapter 2, the hierarchic structure of the cell problems in the FE-HMM makes the method costly. Combing this method with RB technique is certainly of high interest.

The adaptive RB-FE-HMM for modeling fluid in porous media. In [12] a multiscale method, where the effective permeability of Darcy problem recovered from Stokes problem is proposed. Optimal convergence rates for the multiscale method can only be obtained when the adaptive algorithms are applied for both micro problems and macro problem.Therefore it would be of great interest to use RB techniques in this methodology.

The RB-FE-HMM for monotone nonlinear problems. In this thesis we have proposed the RB-FE-HMM for quasilinear problems of nonmonotone type. An obvious question is the applicability of RB techniques for (nonlinear) monotone problems. Based on the structure of the monotone problem studied in [17], for standard FE-HMM, the Newton method is applied to solve both macro and micro problems. However the design of the RB-FE-HMM is not straight forward because of the nonlinearity in the micro problems. Furthermore, the analysis for the FE-HMM for this type of problems is already very involved, and thus for the extension to the RB-FE-HMM could be quite challenging.

The RB-FE-HMM for stochastic problems. In this thesis, all the problems discussed and numerical examples presented were based on deterministic PDEs. It would be of interest to study if and how our techniques could account for stochastic coefficients.

In the end, we hope that the RB-FE-HMM with efficient performance and certified error control could be considered for real world engineering applications.

Bibliography

[1] A. Abdulle. On a priori error analysis of fully discrete heterogeneous multiscale FEM. *SIAM, Multiscale Model. Simul.*, 4(2):447–459, 2005.

[2] A. Abdulle. Analysis of a heterogeneous multiscale FEM for problems in elasticity. *Math. Models Methods Appl. Sci.*, 16(4):615–635, 2006.

[3] A. Abdulle. The finite element heterogeneous multiscale method: a computational strategy for multiscale PDEs. *GAKUTO Int. Ser. Math. Sci. Appl.*, 31:135–184, 2009.

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

[5] A. Abdulle. Discontinuous Galerkin finite element heterogeneous multiscale method for elliptic problems with multiple scales. *Math. Comp.*, 81(278):687–713, 2012.

[6] A. Abdulle. Numerical homogenization methods. *to appear in Springer, Encyclopedia of Applied and Computational Mathematics*, 2013.

[7] A. Abdulle and Y. Bai. Reduced basis finite element heterogeneous multiscale method for high-order discretizations of elliptic homogenization problems. *J. Comput. Phys.*, 231(21):7014–7036, 2012.

[8] A. Abdulle and Y. Bai. Adaptive reduced basis finite element heterogeneous multiscale method. *Comput. Methods Appl. Mech. Engrg.*, 257:201–220, 2013.

[9] A. Abdulle and Y. Bai. Fully discrete analysis of the heterogeneous multiscale method for elliptic problems with multiple scales. *submitted to publication*, 2013.

[10] A. Abdulle, Y. Bai, and G. Vilmart. An offline-online homogenization strategy to solve quasilinear two-scale problems at the cost of one-scale problems. *submitted to publication*, http://mathicse.epfl.ch/page-68906-en.html, 2013.

[11] A. Abdulle, Y. Bai, and G. Vilmart. Reduced basis finite element heterogeneous multiscale method for quasilinear elliptic homogenization problems. *submitted to publication*, http://mathicse.epfl.ch/page-68906-en.html, 2013.

[12] A. Abdulle and O. Budác. An adaptive finite element heterogeneous multiscale method for the stokes problem in porous media. *Preprint*, 2013.

[13] A. Abdulle and W. E. Finite difference heterogeneous multi-scale method for homogenization problems. *J. Comput. Phys.*, 191(1):18–39, 2003.

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

[15] A. Abdulle and B. Engquist. Finite element heterogeneous multiscale methods with near optimal computational complexity. *SIAM, Multiscale Model. Simul.*, 6(4):1059–1084, 2007.

[16] A. Abdulle and M. E. Huber. Discontinuous galerkin finite element heterogeneous multiscale method for advection-diffusion problems with multiple scales. *Numer. Math., in press*, 2013.

[17] A. Abdulle, M. E. Huber, and G. Vilmart. Fully-discrete space-time analysis for parabolic nonlinear monotone single scale and multiscale problems. *Preprint*, 2013.

[18] A. Abdulle and A. 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.

[19] A. Abdulle and A. Nonnenmacher. A short and versatile finite element multiscale code for homogenization problem. *Comput. Methods Appl. Mech. Engrg*, 198(37-40):2839–2859, 2009.

[20] A. Abdulle and A. Nonnenmacher. Adaptive finite element heterogeneous multiscale method for homogenization problems. *Comput. Methods Appl. Mech. Engrg.*, 200(37-40):2710–2726, 2011.

[21] A. Abdulle and A. Nonnenmacher. A posteriori error estimate in quantities of interest for the finite element heterogeneous multiscale method. *Numerical Methods for Partial Differential Equations, in press*, 2013.

[22] A. Abdulle and C. Schwab. Heterogeneous multiscale FEM for diffusion problems on rough surfaces. *SIAM, Multiscale Model. Simul.*, 3(1):195 – 220, 2005.

[23] A. Abdulle and G. Vilmart. The effect of numerical integration in the finite element method for nonmonotone nonlinear elliptic problems with application to numerical homogenization methods. *C. R. Acad. Sci. Paris, Ser. I*, 349(19-20):1041 – 1046, 2011.

[24] A. Abdulle and G. Vilmart. Coupling heterogeneous multiscale FEM with Runge-Kutta methods for parabolic homogenization problems: a fully discrete space-time analysis. *Math. Models Methods Appl. Sci.*, 22(6):1250002/1–1250002/40, 2012.

[25] A. Abdulle and G. Vilmart. A priori error estimates for finite element methods with numerical quadrature for nonmonotone nonlinear elliptic problems. *Numer. Math.*, 121(3):397–431, 2012.

[26] A. Abdulle and G. Vilmart. Analysis of the finite element heterogeneous multiscale method for nonmonotone elliptic homogenization problems. *Math. Camp., in press*, 2013.

[27] M. Ainsworth and J. Oden. *A posteriori error estimation in finite element analysis*. John Wiley & Sons, New york, 2000.

[28] M. Ainsworth and R. Rankin. Guaranteed computable bounds on quantities of interest in finite element computations. *International Journal of Numerical Methods in Engineering*, 89(13):1605–1634, 2012.

[29] G. Allaire. Homogenization and two-scale convergence. *SIAM J. Math. Anal.*, 23(6):1482–1518, 1992.

[30] G. Allaire and M. Briane. Multiscale convergence and reiterated homogenisation. *Proc. Roy. Soc. Edinburgh Sect. A*, 126(2):297–342, 1996.

[31] G. Allaire and R. Brizzi. A multiscale finite element method for numerical homogenization. *Multiscale Model. Simul.*, 4(3):790–812 (electronic), 2005.

[32] H. Amann. Nonhomogeneous linear and quasilinear elliptic and parabolic boundary value problems. In *Function spaces, differential operators and nonlinear analysis (Friedrichroda, 1992)*, volume 133 of *Teubner-Texte Math.*, pages 9–126. Teubner, Stuttgart, 1993.

[33] M. Artola and G. Duvaut. Un résultat d'homogénéisation pour une classe de problèmes de diffusion non linéaires stationnaires. *Ann. Fac. Sci. Toulouse Math.*, 4(5):1–28, 1982.

[34] I. Babŭska and J. Osborn. Homogenization and its application. mathematical and computational problems. *Numerical solution of partial differential equations, III (Proc. Third Sympos. (SYNSPADE), Univ. Maryland, College Park, Md., 1975)*, pages 89–116, 1976.

[35] I. Babŭska and J. Osborn. Generalized finite element methods: their performance and their relation to mixed methods. *SIAM J. Numer. Anal.*, 20:510–536, 1983.

[36] W. Bangerth and R. Rannacher. *Adaptive finite element methods for differential equations*. Birkhäuser Verlag, Basel, 2003.

[37] M. Barrault, Y. Maday, N. Nguyen, and A. Patera. An 'empirical interpolation method': Application to efficient reduced-basis discretization of partial differential equations. *C. R. Acad. Sci. Paris*, Ser.I 339:667–672, 2004.

[38] S. Bartels and C. Carstensen. Each averaging technique yields reliable a posteriori error control in FEM on unstructured grids. II. Higher order FEM. *Math. Comp.*, 71(239):971–994, 2002.

[39] S. Bartels, C. Carstensen, and A. Hecht. P2Q2Iso2D=2D isoparametric FEM in Matlab. *J. Comput. Appl. Math.*, 192(2):219–250, 2006.

[40] J. Bear and Y. Bachmat. *Introduction to modelling of transport phenomena in porous media*. Kluwer Academic, Dordrecht, The Netherlands, 1991.

[41] R. Becker and R. Rannacher. An optimal control approach to a posteriori error estimation in finite element methods. *Acta numerica*, 10:1–102, 2001.

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

[43] P. Binev, A. Cohen, W. Dahmen, R. Devore, G. Petrova, and P. Wojtaszczyk. Convergence rates for greedy algorithms in reduced basis methods. *SIAM J. Math. Anal.*, 43:1457–1472, 2011.

[44] L. Boccardo and F. Murat. Homogénéisation de problèmes quasi-linéaires. *Publ. IRMA, Lille*, 3(7):13–51, 1981.

[45] S. Boyaval. Reduced-basis approach for homogenization beyond the periodic setting. *Multiscale Model. Simul.*, 7(1):466–494, 2008.

[46] S. Boyaval. *Mathematical modeling and simulation for material science*. PhD thesis, University Paris Est, 2009.

[47] F. Brezzi and R. Russo. Choosing bubbles for advection-diffusion problems. *Math. Models Methods Appl. Sci.*, 4(4):571– 587, 1994.

[48] A. Buffa, Y. Maday, A. T. Patera, C. Prud'homme, and G. Turinici. A priori convergence of the greedy algorithm for the parametrized reduced basis. *ESAIM: M2AN*, 46:595–603, 2012.

[49] L. Chen and C. Zhang. *AFEM@matlab: a matlab package of adaptative finite element methods*. technical report. 2006.

[50] Z. Chen, W. Deng, and H. Ye. Upscaling of a class of nonlinear parabolic equations for the flow transport in heterogeneous porous media. *Commun. Math. Sci.*, 3(4):493–515, 2005.

[51] Z. Chen and T. Y. Savchuk. Analysis of the multiscale finite element method for nonlinear and random homogenization problems. *SIAM J. Numer. Anal.*, 46(1):260–279, 2007/08.

[52] M. Chipot. Elliptic equations: an introductory course. *Birkhäuser Advanced Texts: Basler Lehrbücher. Birkhäuser Verlag, Basel*, 2009.

[53] P. Ciarlet. *Basic error estimates for Elliptic problems*, volume 2 of *Handbook of numerical analysis*. Elsevier science publishers B.V., North-Holland, 1991.

[54] P. Ciarlet and P. Raviart. The combined effect of curved boundaries and numerical integration in isoparametric finite element methods. *Math. Foundation of the FEM with Applications to PDE*, pages 409–474, 1972.

[55] P. Clément. Approximation by finite element functions using local regularization. *Revue française d?automatique, informatique, recherche opéra-tionnelle. Analyse numérique*, 9(R2):77–84, 1975.

[56] E. De Giorgi and S. Spagnolo. Sulla convergenza degli integrali dell'energia per operatori ellittici del secondo ordine. *Boll. Un. Mat. Ital.*, 4(8):391–411, 1973.

[57] R. Du and P. Ming. Heterogeneous multiscale finite element method with novel numerical integration schemes. *Commm. Math. Sci.*, 8(4):863–885, 2010.

[58] W. E. *Principles of multiscale modeling*. Cambridge University Press, 2011.

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

[60] W. E, P. Ming, and P. Zhang. Analysis of the heterogeneous multiscale method for elliptic homogenization problems. *J. Amer. Math. Soc.*, 18(1):121–156, 2005.

[61] Y. Efendiev and T. Y. Hou. *Multiscale finite element methods. Theory and applications*, volume 4 of *Surveys and Tutorials in the Applied Mathematical Sciences*. Springer, New york, 2009.

[62] Y. Efendiev and X. Wu. Multiscale finite element for problems with highly oscillatory coefficients. *Numer. Math.*, 90:459–486, 2002.

[63] J. Eftang, M. Grepl, and A. Patera. A posteriori error bounds for the empirical interpolation method. *C. R. Acad. Sci. Paris*, Ser.I 348(9-10):575–579, 2004.

[64] J. P. Fink and W. C. Rheinboldt. On the error behavior of the reduced basis technique for nonlinear finite element approximations. *Z. Angew. Math. Mech.*, 63:21–28, 1983.

[65] R. L. Fox and H. Miura. An approximate analysis technique for design calculations. *AIAA Journal*, 9:177–179, 1971.

[66] N. Fusco and G. Moscariello. On the homogenization of quasilinear divergence structure operators. *Ann. Mat. Pura Appl.*, 146(4):1–13, 1987.

[67] M. Grepl and A. Patera. A posteriori error bounds for reduced-basis approximations of parametrized parabolic partial differential equations. *ESAIM: M2AN*, 39(1):157–181, 2005.

[68] V. H. Hoang and christoph Schwab. High-dimensional finite elements for elliptic problems with multiple scales. *Multiscale Model. Simul*, 3(1):168–194, 2005.

[69] T. Hou, X. Wu, and Z. Cai. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. *Math. Comp.*, 68(227):913–943, 1999.

[70] T. Hughes. Multiscale phenomena: Green's functions, the dirichlet-to- neumann formulation, subgrid scale models, bubbles and the origin of stablized methods. *Comput. Methods Appl. Mech. Engrg.*, 128:387–401, 1995.

[71] T. Hughes, G. Feijóo, L. Mazzei, and J.-B. Quincy. The variational multiscale method - a paradigm for computational mechanics. *Comput. Methods Appl. Mech. Engrg.*, 166:3–24, 1998.

[72] D. Huynh, G. Rozza, S. Sen, and A. Patera. A successive constraint linear optimaization method for lower bounds of parametric coercivity and inf-sup stability constants. *C. R. Math. Acad. Sci. Paris, Ser.I*, 345:473–478, 2007.

[73] J. J. Douglas and T. Dupont. A Galerkin method for a nonlinear Dirichlet problem. *Math. Comp.*, 29(131):689–696, 1975.

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

[75] A. Karageorghis and D. Lesnic. Steady-state nonlinear heat conduction in composite materials using the method of fundamental solutions. *Comput. Methods Appl. Mech. Engrg.*, 197(33–40):3122–3137, 2008.

[76] K.Veroy, C. Prud'homme, D. Rovas, and A. Patera. A posteriori error bounds for reduced-basis approximation of parametrized noncoercive and nonlinear elliptic partial differential equations. *Proceedings of the 16th AIAA Computational Fluid Dynamics Conference*, 2003.

[77] Y. Maday, N. C. Nguyen, A. T. Patera, and G. Pau. A general multipurpose interpolation procedure: the magic points. *Commun. Pure Appl. Anal.*, 8(1):383–404, 2009.

[78] S. Moorthy and S. Ghosh. Adaptivity and convergence in the voronoi cell Âőnite element model for analyzing heterogeneous materials. *Comput. Methods Appl. Mech. Engrg.*, 185(185):37–74, 2000.

[79] S. Moorthy and S. Ghosh. Concurrent multi-scale analysis of elastic composites by a multi-level computational model. *Comput. Methods Appl. Mech. Engrg.*, 193:497–538, 2004.

[80] F. Murat and L. Tartar. H-convergence, topics in the mathematical modeling of composite materials. *Progr. Nonlinear Differential Equations Appl.*, 31:21–43, 1997.

[81] G. Nguetseng. A general convergence result for a functional related to the theory of homogenization. *SIAM J. Math. Anal.*, 20(3):608–623, 1989.

[82] Y.-Y. Nie and V. Thomée. A lupmped mass finite-element method with quadrature for a non-linear parabolic problem. *IMA Journal of Numerical Analysis*, 5:371–396, 1985.

[83] J. A. Nitsche. On L_∞-convergence of finite element approximations to the solution of a nonlinear boundary value problem. In *Topics in numerical analysis, III (Proc. Roy. Irish Acad. Conf., Trinity Coll., Dublin, 1976)*, pages 317–325. Academic Press, London, 1977.

[84] R. Nochetto, A. Veeser, and M. Verani. A safeguarded dual weighted residual method. *IMA journal of Numerical Analysis*, 29(1):126–140, 2009.

[85] A. Nonnenmacher. *Adaptive finite element methods for multiscale partial differential equations*. PhD thesis, École polytechnique fédérale de Lausanne, 2011.

[86] A. K. Noor. Recent advances in reduction methods for nonlinear problems. *Comput. Struct.*, 13:31–44, 1981.

[87] J. Oden and S. Prudhomme. Goal-oriented error estimation and adaptivity for the finite element method. *Computers & Mathematics with Applications*, 41(5-6):735 – 756, 2001.

[88] J. Oden, S. Prudhomme, A. Romkes, and P. Bauman. Multiscale modeling of physical phenomena: adaptive control of models. *SIAM J. Sci. Comput.*, 28(6):2359–2389, 2006.

[89] J. T. Oden and K. 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.

[90] M. Ohlberger. A posteriori error estimates for the heterogeneous multiscale finite element method for elliptic homogenization problems. *Multiscale Model. Simul*, 4(1):88–114, 2005.

[91] H. Owhadi and L. Zhang. Localized bases for finite-dimensional homogenization approximations with nonseparated sacles and high contrast. *Multiscale Model. Simul.*, 9(4):1373–1398, 2011.

[92] A. T. Patera and G. Rozza. *Reduced Basis Approximation and A Posteriori Error Estimation for Parametrized Partial Differential Equations*. to appear in (tentative rubric) MIT Pappalardo Graduate Monographs in Mechanical Engineering, 2007.

[93] C. Prud'homme, D. V. Rovas, K. Veroy, L. Machiels, Y. Maday, A. T. Patera, and G. Turinici. Reliable real-time solution of parametrized partial differential equations: Reduced-basis output bounds methods. *J. Fluids Eng.*, 124:70–80, 2002.

[94] S. Prudhomme and J. Oden. On goal-oriented error estimation for elliptic problems: application to the control of pointwise errors. *Computer Methods in Applied Mechanics and Engineering*, 176(1-4):313–331, 1999.

[95] A. Quarteroni. *Numerical Models for Differential Problems*. Springer, 2009.

[96] G. Rozza, D. Huynh, and A. T. Patera. Reduced basis approximation and a posteriori error estimation for affinely parametrized elliptic coercive partial differential equations. *Arch. Comput. Methods. Eng.*, 15:229–275, 2008.

[97] E. Sanchez-Palencia. *Nonhomogeneous Media and Vibration Theory*. Lecture Notes in Phys. 127. Springer, Berlin, 1980.

[98] Sandia National Lab. *http://cubit.sandia.gov*. Sandia Corporation, 1997-2010.

[99] G. Sangalli. Capturing small scales in elliptic problems using a residual-free bubbles finite element method. *Multiscale Model. Simul.*, 1(15):485–503, 2003.

[100] A. Schmidt and K. Siebert. *Design of adaptive finite element software: The finite element toolbox ALBERTA*, volume 42 of *Lecture Notes in Computational Science and Engineering*. Springer-Verlag, Berlin, 2005.

[101] H. Schmidt. On cubature formulae with a minimal number of knots. *Numer. Math.*, 31(3):281–297, 1978.

[102] F. Talati and S. Jalalifar. Analysis of heat conduction in a disk brake system. *Heat Mass Transfer*, 45:1047–1059, 2009.

[103] L. Tartar. *Estimations des coefficients homogénéisés*. Springer-Verlag, Berlin, 1977.

[104] L. Tartar. *Incompressible fluid flow through porous media convergence of the homogenization process.* in appendix of [97]. Springer, Berlin, 1980.

[105] R. Verfürth. *A review of a posteriori error estimation and adaptive mesh-refinement techniques.* Wiley-Teubner, New-York, 1996.

[106] O. Zienkiewicz. *The finie element method.* Elsevier/Butterworth-Heinemann, Amsterdam, 2005.

[107] T. I. Zohdi, J. T. Oden, and G. J. Rodin. Hierarchical modeling of heterogeneous bodies. *Comput. Methods Appl. Mech. Engrg.*, 138(1-4):273–298, 1996.

YUN BAI

Date of birth: 09. 03. 1984. **Nationality:** China.

EDUCATION

Ph.D. in Applied Mathematics 08. 2009 - 09. 2013
École Polytechnique Fédérale de Lausanne, Switzerland.
Thesis advisor: Professor Assyr Abdulle.

M.S. in Applied Mathematics 09. 2006 - 07. 2009
University of Science and Technology of China.
Thesis advisor: Professor Xingye Yue.

B.S. in Mathematics 09. 2002 - 07. 2006
Special Class for Gifted Young, University of Science and Technology of China.

PUBLICATIONS

[1] A. Abdulle, Y. Bai, G. Vilmart, *Reduced basis finite element heterogeneous multiscale method for quasilinear homogenization problems.* Submitted to publication, 2013.

[2] A. Abdulle, Y. Bai, G. Vilmart, *An offline-online homogenization strategy for quasilinear problems.* Submitted to publication, 2013.

[3] A. Abdulle, Y. Bai, *Fully discrete analysis of the heterogeneous multiscale method for elliptic problems with multiple scales*. To be published in IMAJNA, 2013.

[4] A. Abdulle, Y. Bai, *Adaptive reduced basis finite element heterogeneous multiscale method.* Comput. Methods Appl. Mech. Engrg., 2013.

[5] A. Abdulle, Y. Bai, *Reduced basis finite element heterogeneous multiscale method for high-order discretizations of elliptic homogenization problems*. J. Comput. Phys., 2012

[6] Y. Bai, X. Yue, Q. Zeng, *Multil-scale Modeling and Numerical Simulation for CVI Process.* Commun. Comput. Phys., 2010.

PRESENTATIONS AT INTERNATIONAL CONFERENCES

"Reduced basis finite element heterogeneous multiscale method for quasilinear problems", MAFELAP 2013, Uxbridge, UK, 2013.

"Reduced basis finite element heterogeneous multiscale method and its applications" , AIMS Conference, Florida, USA, 2012.

"Reduced basis finite element heterogeneous multiscale method",
Swiss Numerical Colloquium, Bern, Switzerland, 2012.

"Multiscale modeling for CVI process",
International Ceramic Composite Material Seminar, Xi'an, China, 2008.