Reduced basis finite element heterogeneous multiscale method for quasilinear elliptic homogenization problems

Assyr Abdulle[1], Yun Bai[1], and Gilles Vilmart[2]

April 10, 2013

Abstract

The reduced basis finite element heterogeneous multiscale method (RB-FE-HMM) for a class of nonlinear homogenization elliptic problems of nonmonotone type is introduced. In this approach, the solutions of the micro problems needed to estimate the macroscopic data of the homogenized problem are selected by a Greedy algorithm and computed in an offline stage. It is shown that the use of reduced basis (RB) for nonlinear numerical homogenization reduces considerably the computational cost of the finite element heterogeneous multiscale method (FE-HMM). As the precomputed microscopic functions depend nonlinearly on the macroscopic solution, we introduce a new a posteriori error estimator for the Greedy algorithm that guarantees the convergence of the online Newton method. A priori error estimates and uniqueness of the numerical solution are also established. Numerical experiments illustrate the efficiency of the proposed method.

Keywords: nonlinear nonmonotone elliptic problems, numerical homogenization, reduced basis method, a posteriori error estimator, finite element method

AMS subject classification (2010): 65N30,65M60,74D10,74Q05.

1 Introduction

Quasilinear elliptic problems 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 [14]. Numerical approximations of such problems have been analyzed by many authors. We mention the works of Douglas and Dupont [30], and Nitsche [35], 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 [11] for a FEM with numerical quadrature, i.e., when the continuous variational form originating from the nonlinear problem is approximated by a quadrature formula. In this paper we are interested in quasilinear elliptic problems with highly oscillatory data of the form

$$-\nabla \cdot (a^\varepsilon(x, u^\varepsilon(x))\nabla u^\varepsilon(x)) = f(x) \ \text{ in } \Omega, \tag{1}$$

in a domain $\Omega \subset \mathbb{R}^d,\ d \leq 3$, where $a^\varepsilon(x,u)$ is a $d \times d$ tensor, associated to $\varepsilon > 0$, a sequence of positive real numbers going to zero. Seeking the solution u^ε in $H^1(\Omega)$ we assume $f \in H^{-1}(\Omega)$.

[1]ANMC, Section de Mathématiques, École Polytechnique Fédérale de Lausanne, 1015 Lausanne, Switzerland, Assyr.Abdulle@epfl.ch, Yun.Bai@epfl.ch

[2]École Normale Supérieure de Cachan, Antenne de Bretagne, INRIA Rennes, IRMAR, CNRS, UEB, Campus de Ker Lann, F-35170 Bruz, France, Gilles.Vilmart@bretagne.ens-cachan.fr

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 [17] or (stationary) heat conduction in a composite material [32]. For the standard FEM, a finescale resolution is needed for a satisfactory approximation. If the ratio between the scale of interest and the finest scale in the problem is too large, the FEM approximation will have a prohibitive number of degrees of freedom (DOF), leading to an enormous computational cost. For efficient numerical computations, an appropriate upscaling of equation (1) is thus needed. Such coarse graining procedures are rigorously described by the mathematical homogenization theory [18, 31] and are studied for the class of problems (1) in [15, 20, 29]. These analyses show that the solution u^ε of (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 (1) are derived in [24] for the multiscale finite element method (MsFEM) and in [28, 12] for the finite element heterogeneous multiscale method (FE-HMM) [6, 27]. The MsFEM is based on a standard FE space enriched with oscillatory functions, while the FE-HMM is based on a strategy first proposed in [27] that consists in macroscopic FEM on a macroscopic mesh with quadrature formula (QF), with effective data (the homogenized tensor at the quadrature points) recovered on the fly from micro problems. These micro problems, defined on sampling domains centered at the macroscopic quadrature points of the QF, use only the oscillatory data given by the problem (1). We focus on the FE-HMM proposed in [28, 12] 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 paper, we show how the use of the reduced basis (RB) method (see [37, 36, 38] and references therein) for computing the micro problems permits to considerably improve the efficiency of the standard nonlinear FE-HMM. The use of RB for multiscale problems with an emphasis on parametrizing various configurations of micro problems have been proposed in [21, 34, 13]. The application of RB to the FE-HMM is designed and analyzed for a class of linear elliptic problems in [3]. In this paper, we extend the analysis and the method, denoted RB-FE-HMM, for a class of nonlinear problems of nonmonotone type. We focus in this paper on the analysis of the method, while in [5] we detail the implementation of the algorithm for practitioners, explain the numerical construction of a corrector and also provide 3D numerical examples for both static and time dependent nonlinear problems. Following the methodology of RB, the method relies online and offline procedures: in the offline procedure accurate micro solutions for the original problem on sampling domains are selected and computed. These micro problems are parametrized by the location of the cell problem in the domain Ω and the macroscopic solution at this location. A Greedy algorithm allows to choose an optimal basis of micro functions (computed with high accuracy) for selected values of the parameters. In 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 paper such 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 paper is organized as follows. In Sect. 2, we briefly recall the framework of homogenization theory in our context of quasilinear elliptic problems of nonmonotone type. We then present in Sect. 3 the new nonlinear RB-FE-HMM with its offline and online procedures, and analyze its convergence in Sect. 4. We explain some implementation issues in Sect. 5. Finally, numerical experiments in Sect. 6 show how the use of reduced basis considerably improves the efficiency by reducing drastically the number of degrees of freedom for various problems.

2 Homogenization of quasilinear elliptic problems

We assume that the tensor $a^\varepsilon(x,s)$ in (1) is uniformly elliptic and bounded with respect to s and ε, i.e., there exist $\lambda, \Lambda_1 > 0$ such that

$$\lambda|\xi|^2 \le a^\varepsilon(x,s)\xi\cdot\xi, \quad |a^\varepsilon(x,s)\xi| \le \Lambda_1|\xi|, \quad \forall \xi\in\mathbb{R}^d, \forall s\in\mathbb{R}, \text{ a.e. } x\in\Omega, \tag{2}$$

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 (1) has a unique solution $u^\varepsilon\in H^1_0(\Omega)$ (we refer for example to [25, Theorem 11.6] for a proof). The solution, for each ε, satisfies the a priori bound $\|u^\varepsilon\|_{H^1(\Omega)} \le C\|f\|_{H^{-1}(\Omega)}$, 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 $H^1(\Omega)$. The homogenization result is shown in [20, Theorem 3.6] (see also [29]) 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 $H^1(\Omega)$. The limit function u^0 is the solution of the homogenized problem

$$-\nabla\cdot\big(a^0(x,u^0(x))\nabla u^0(x)\big) = f(x) \ \text{ in } \Omega, \qquad u^0(x)=0 \ \text{ on } \partial\Omega. \tag{3}$$

The tensor $a^0(x,s)$, called the homogenized tensor, can be shown to be Lipschitz continuous with respect to s, uniformly elliptic, and bounded [20, Prop. 3.5], i.e., there exists $\Lambda_2>0$ such that[1]

$$\|a^0(x,s_1)-a^0(x,s_2)\|_F \le \Lambda_2|s_1-s_2|, \text{ a.e. } x\in\Omega, \forall s_1,s_2\in\mathbb{R}, \tag{4}$$

and there exist $\lambda,\Lambda_1>0$ such that a^0 satisfies (2) (possibly with different constants). Under these assumptions, the homogenized problem (3) has also a unique solution $u^0\in H^1_0(\Omega)$.

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 (3) holds for the whole sequence $\{u^\varepsilon\}$ and the homogenized tensor can be characterized in the following way [15]:

$$a^0(x,s) = \int_Y a(x,y,s)(I+J^T_{\chi(x,y,s)})dy, \qquad \text{for } x\in\Omega, s\in\mathbb{R}, \tag{5}$$

[1] In this paper, we use the Frobenius norm on matrices defined as $\|M\|_F := \mathrm{trace}(M^TM)$.

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 $W^1_{per}(Y) := \{z \in H^1_{per}(Y);\ \int_Y z dx = 0\}$ of the linear cell problems with parameters $x \in \Omega, s \in \mathbb{R}$

$$\int_Y a(x,y,s)\nabla_y\chi^i(x,y,s)\cdot\nabla w(y)dy = -\int_Y a(x,y,s)\mathbf{e}_i\cdot\nabla w(y)dy, \quad \forall w \in W^1_{per}(Y), \tag{6}$$

where $H^1_{per}(Y) := \{g \in H^1(Y)|\ g \text{ periodic in } Y\}$ and $\mathbf{e}_i$, $i=1,\dots,d$ are the vectors of the canonical basis of $\mathbb{R}^d$.

Remark 2.1. *We sometimes refer to the problems* (3) *or* (1) *as "non monotone problems". This stems from the following fact: writing for example* (3) *in weak form*

$$B(u^0;u^0,v) = \int_\Omega a^0(x,u^0(x))\nabla u^0(x)\nabla v(x)dx = (f,v), \quad \forall v \in H^1_0(\Omega),$$

we observe that the monotonicity property $B(u^0;u^0,u^0-v)-B(v;v,u^0-v) \geq C\|u^0-v\|^2_{H^1(\Omega)}$ *with* $C \geq 0$ *does not hold in general for the quasilinear problem* (3) *(or* (1)*). This lack of monotonicity makes the numerical analysis for FEM a nontrivial task, in particular when quadrature formulas are used [11].*

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,

$$a^0_{mn} \in C^0(\overline{\Omega}\times\mathbb{R}), \qquad \forall m,n=1,\dots,d. \tag{7}$$

3 Reduced basis FE-HMM for quasilinear problems

As the homogenized tensor a^0 in (3) 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 [4] for linear elliptic problems to quasilinear elliptic problems. This method relies on a macroscopic solver with macroscopic data recovered by microscopic simulations (the micro problems) performed on sampling domains located at appropriate quadrature points of the macroscopic mesh. In addition, in order to avoid repeated micro computations, the solution of the micro problem are computed in finite dimensional space of low dimension spanned by a so-called reduced basis obtained in an offline procedure.

3.1 Preliminaries

We describe here the macro and micro finite element spaces needed to define and analyze the RB-FE-HMM.

Macroscopic mesh and quadrature formulas. The RB-FE-HMM is based on a macro finite element (FE) space

$$S^\ell_0(\Omega,\mathcal{T}_H) = \{v^H \in H^1_0(\Omega);\ v^H|_K \in \mathcal{R}^\ell(K),\ \forall K \in \mathcal{T}_H\},$$

where $\mathcal{T}_H$ is a shape-regular family of (macro) partition of Ω in simplicial or quadrilateral elements K of diameter H_K, and $\mathcal{R}^\ell(K)$ is the space $\mathcal{P}^\ell(K)$ of polynomials on K of total degree at most ℓ if K is a simplicial FE, or the space $\mathcal{Q}^\ell(K)$ of polynomials on K of degree at most ℓ in each variable if K is a parallelogram FE. For a given macro partition, we define as usual $H := \max_{K\in\mathcal{T}_H} H_K$. We highlight that H in our discretization is allowed to be much larger than ε.

For each element K of the macro partition we consider an affine transformation F_K such that $K = F_K(\hat{K})$, where $\hat{K}$ is the reference element (simplicial or parallelogram). For a given quadrature formula $\{\hat{x}_j, \hat{\omega}_j\}_{j=1}^J$ on $\hat{K}$, the quadrature weights and integration points on $K \in \mathcal{T}_H$ are then given by $\omega_{K_j} = \hat{\omega}_j |\det(\partial F_K)|,\ x_{K_j} = F_K(\hat{x}_j),\ j = 1, \ldots, J$. We make the following assumptions on the quadrature formulas, which are standard assumptions also for linear elliptic problems [26]:

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

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

Microscopic mesh and RB. We consider a micro FE space $S^q(Y,\mathcal{N}) \subset W(Y)$ with simplicial or quadrilateral FEs and piecewise polynomial of degree q on the domain $Y = (-1/2, 1/2)^d$. We apply a conformal and shape regular family of triangulation $\mathcal{T}_{\hat{h}}$ and use $\mathcal{N}$ to denote the number of degrees of freedom (DOF) of $S^q(Y,\mathcal{N})$. The space $W(Y)$ denotes either the Sobolev space

$$W(Y) = W^1_{per}(Y) \tag{8}$$

for a periodic coupling or

$$W(Y) = H^1_0(Y) \tag{9}$$

for a coupling with Dirichlet boundary conditions.

We then consider the RB space, which is a subspace of $S^q(Y,\mathcal{N})$ with a low dimension denoted

$$S_N(Y) = \text{span}\{\hat{\xi}_{n,\mathcal{N}}(y),\ n = 1, .., N\} \subset S^q(Y,\mathcal{N}). \tag{10}$$

where $\hat{\xi}_{n,\mathcal{N}}(y),\ n = 1, .., N$ denotes the reduced basis. Notice that for the analysis of the RB-FE-HMM, we shall also consider a 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 Sect. 3.4 below.

For each macro element $K \in \mathcal{T}_H$ and each quadrature point $x_{K_j} \in K,\ j = 1, \ldots, J$, we define the sampling domains $K_{\delta_j} = x_{K_j} + (-\delta/2, \delta/2)^d$, $(\delta \geq \varepsilon)$. We observe that each sampling domain K_{δ_j} is in correspondence with Y through the affine transformation

$$y \in Y \mapsto G_{x_{K_j}}(y) = x_{K_j} + \delta y \in K_{\delta_j} \tag{11}$$

This transformation applied to the RB space (10) permits to define the RB space $S_N(K_{\delta_j})$ associated to each sampling domain K_{δ_j} as

$$S_N(K_{\delta_j}) = \text{span}\{\delta\hat{\xi}_{n,\mathcal{N}}(G^{-1}_{x_{K_j}}(x)) =: \xi_{n,K_j}(x),\ n = 1, .., N\}. \tag{12}$$

3.2 Online procedure: the RB-FE-HMM

Assuming that the RB space has been pre-constructed in the offline stage that we will present in Section 3.4, we describe here the online stage which relies on a macro method analogous to the standard FE-HMM with the exception that the micro problems are solved in the RB space, which has a low dimension, rather that the original FE space $S^q(Y,\mathcal{N})$.

The online procedure of the nonlinear RB-FE-HMM for (1) is described as follows: we compute $u^{H,RB} \in S_0^\ell(\Omega,\mathcal{T}_H)$ such that

$$B_{H,RB}(u^{H,RB};u^{H,RB},v^H) = \int_\Omega f v^H dx,\ \forall v^H \in S_0^\ell(\Omega,\mathcal{T}_H), \tag{13}$$

with a bilinear form defined for all $u^H, v^H, w^H \in S_0^\ell(\Omega,\mathcal{T}_H)$ by

$$B_{H,RB}(u^H;v^H,w^H) := \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\frac{\omega_{K_j}}{|K_{\delta_j}|}\int_{K_{\delta_j}} a^\varepsilon(x,u^H(x_{K_j}))\nabla v_{N,K_j}^{u^H(x_{K_j})}(x)\cdot\nabla w_{N,K_j}^{u^H(x_{K_j})}(x)dx, \tag{14}$$

where for the scalar parameter $s = u^H(x_{K_j})$, the function v^s_{N,K_j} (similarly for $w^s_{N,K_j}(x)$) solves $v^s_{N,K_j} - v^H_{lin,j} \in S_N(K_{\delta_j})$ and

$$\int_{K_{\delta_j}} a^\varepsilon(x,s)\nabla v^s_{N,K_j}(x)\cdot\nabla z_N(x)dx = 0, \quad \forall z_N \in S_N(K_{\delta_j}) \tag{15}$$

where we used the notation $v^H_{lin,j}(x) := v^s_{N,K_j}(x_{K_j}) + (x-x_{K_j})\cdot\nabla v^s_{N,K_j}(x_{K_j})$. We highlight that the RB space $S_N(K_{\delta_j})$ with dimension N is computed only once during the offline stage with a precision controlled by an a posteriori estimator described in Section 3.4. It remains fixed during the online procedure of the RB-FE-HMM. The problem (15) requires the solution of an $N\times N$ linear system, where the details of the offline output and the online implementation are discussed in Sect. 5. 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].

3.3 Solution of the macro quasilinear problem and Newton method

While the cell problems (15) are linear, the macroscopic problem (14) 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 (13). The bilinear form (14) can be rewritten as

$$B_{H,RB}(u^H;v^H,w^H) = \sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\omega_{K_j}a^0_{N,K_j}(x_{K_j},u^H(x_{K_j}))\nabla v^H(x_{K_j})\cdot\nabla w^H(x_{K_j}), \tag{16}$$

where we define the numerical homogenized tensor as

$$(a^0_{N,K_j}(x_{K_j},s))_{ik} = \int_Y a_{x_{K_j},s}(y)\left(\nabla\hat{\psi}^{i,s}_{N,K_j}(y)+\mathbf{e}_i\right)\cdot\left(\nabla\hat{\psi}^{k,s}_{N,K_j}(y)+\mathbf{e}_k\right)dy. \tag{17}$$

where $\hat{\psi}^{i,s}_{N,K_j} \in S_N(Y),\ i = 1, \ldots, d$ is the solution of a cell problem (see (28) below) corresponding to the sampling domain K_{δ_j}.

Inspired by [30, 12], we explain here how to solve the nonlinear problem (13) 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 and in the direction v^H,

$$\partial B_{H,RB}(z^H; v^H, w^H) := B_{H,RB}(z^H; v^H, w^H) + B'_{H,RB}(z^H; v^H, w^H), \tag{18}$$

where by the reformulation of the RB-FE-HMM bilinear form (16) we derive

$$B'_H(z^H; v^H, w^H) = \sum_{K\in\mathcal{T}_H} \sum_{j=1}^{J} \omega_{K_j} \frac{d}{ds} a^0_{N,K_j}(s)|_{s=z^H(x_{K_j})} v^H(x_{K_j}) \nabla z^H(x_{K_j}) \cdot \nabla w^H(x_{K_j}). \tag{19}$$

The Newton method for approximating a solution u^H of the nonlinear RB-FE-HMM (13) by a sequence $\{u^H_k\}$ reads in weak form

$$\partial B_H(u^H_k; u^H_{k+1} - u^H_k, w^H) = F_H(w^H) - B_H(u^H_k; u^H_k, w^H), \quad \forall w^H \in S^\ell_0(\Omega, \mathcal{T}_H), \tag{20}$$

where $F_H(\cdot)$ is a linear functional on $S^\ell_0(\Omega, \mathcal{T}_H)$ which is an approximation of $F(w^H) = \int_\Omega f(x) w^H(x) dx$, obtained by using quadrature formulas satisfying (**Q2**). The fact that the Newton method is well defined and the convergence is discussed in Sect. 4.2 while an efficient implementation is detailed in Sect. 5.

3.4 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 (25) below (depending on the quadrature node x_{K_j} and the nonlinear parameter s). In the RB method, the dimension $\mathcal{N}$ is expected to be large to obtain a highly resolved numerical solution of (25).

The main novelty here 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 (25) but also their derivatives with respect to the nonlinear parameter s. This is an essential ingredient to prove in Sect. 4.2 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

$$a_{x_{K_j},s}(y) := a^\varepsilon(G_{x_{K_j}}(y), s), \tag{21}$$

where we use the transformation (11) which is crucial for the RB methodology, i.e., an affine representation of the form

$$a_{x_\tau,s}(y) = \sum_{p=1}^{P} \Theta_p(x_\tau, s) a_p(y),\ \forall y \in Y. \tag{22}$$

We notice however that such direct affine representation is generally unavailable and a Greedy algorithm, called the empirical interpolation method (EIM) can be used to approximate a nonaffine tensor by an affine one of the form (22) (see [16]).

Cell problems. The micro problems in the FE-HMM are based on the FE approximation of the cell functions $\psi_{K_j}^{i,s} \in W(K_{\delta_j})$, solving the linear problem

$$\int_{K_{\delta_j}} a^\varepsilon(x,s)\nabla\psi_{K_j}^{i,s}(x)\cdot\nabla z(x)dx = -\int_{K_{\delta_j}} a^\varepsilon(x,s)\mathbf{e}_i\cdot\nabla z(x)dx,\ \forall z\in W(K_{\delta_j}). \tag{23}$$

which has a unique solution using (2). For the design of the RB method, is more convenient to work in the space $W(Y)$ (defined in either (8) or (9)) rather than the quadrature node dependent space $W(K_{\delta_j})$. We thus consider the transformation (11) and using the notations

$$\begin{aligned} b(\hat v,\hat z) &:= \int_Y a_{x_{K_j},s}(y)\nabla\hat v(y)\cdot\nabla\hat z(y)dy \quad \forall \hat v,\hat z\in W(Y),\\ l_i(\hat z) &:= -\int_Y a_{x_{K_j},s}(y)\mathbf{e}_i\cdot\nabla\hat z(y)dy \quad \forall \hat z\in W(Y), \end{aligned} \tag{24}$$

the problem (23) with $\hat\psi_{K_j}^{i,s}(y) = \psi_{K_j}^{i,s}(G_{x_{K_j}}(y))$ can be transformed into

$$b(\hat\psi_{K_j}^{i,s},\hat z) = l_i(\hat z),\ \forall \hat z\in W(Y). \tag{25}$$

On $W(Y)$ we consider the scalar product $(v,w)_{\mathcal W} = \int_Y \nabla v\cdot\nabla w dy$ and associated norm $\|v\|_{\mathcal W} = ((v,v)_{\mathcal W})^{1/2}$. For the sampling domain T_δ centered at a given point x_τ and given a parameter s, we define the energy norm

$$\|v\|_{\mathcal E,T_\delta,s} := (b(v,v))^{1/2} = \left(\int_Y a_{x_\tau,s}(y)\nabla v(y)\cdot\nabla v(y)dy\right)^{1/2}, \tag{26}$$

and notice that from the ellipticity of the tensor it holds

$$\|v\|_{\mathcal W} \le \frac{1}{\sqrt{\lambda}}\|v\|_{\mathcal E,T_\delta,s}. \tag{27}$$

In what follows, it will be convenient to 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 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., $\hat z_{\mathcal N}$). The FE space $S^q(Y,\mathcal N)$ has a (shape-regular) triangulation $\mathcal T_{\hat h}$ with $\mathcal N = \mathcal O(\hat h^{-d})$ denoting its number of DOF. Consider $\hat\psi_{\mathcal N,K_j}^{i,s}\in S^q(Y,\mathcal N)$ the solution of the linear problem

$$b(\hat\psi_{\mathcal N,K_j}^{i,s},\hat z_{\mathcal N}) = l_i(\hat z_{\mathcal N}) \quad \forall \hat z_{\mathcal N}\in S^q(Y,\mathcal N), \tag{28}$$

We notice using (2) that problem (28) has a unique solution.

For the convergence of the Newton method explained in Section 3.3 we will also need to control the derivatives with respect to the parameter s of the cell functions $\hat\psi_{K_j}^{i,s}$. We assume[2]

$$\begin{aligned} &s\in\mathbb R \mapsto a^\varepsilon(\cdot,s)\in (L^\infty(\Omega))^{d\times d} \text{ is of class } C^1,\\ &|\partial_s a^\varepsilon(x,s)\xi| \le \Lambda_2|\xi|,\ \forall s\in\mathbb R,\ \text{a.e. } x\in\Omega, \forall\xi\in\mathbb R^d. \end{aligned} \tag{29}$$

[2]It is shown in [20, Rem. 3.3, Prop. 3.5] that the best constant Λ_2 in (4) may differ from the one in (29).

Lemma 3.1. *Assume that* (2) *and* (29) *hold. Consider the solution* $\hat{\psi}^{i,s}_{\mathcal{N},K_j}$ *of* (28)*. Then, the map* $s \mapsto \hat{\psi}^{i,s}_{\mathcal{N},K_j} \in H^1(T_\delta)$ *is of class* C^1 *and satisfies*

$$\partial_s \hat{\psi}^{i,s}_{\mathcal{N},K_j} = \hat{\phi}^{i,s}_{\mathcal{N},K_j}, \qquad \partial_s \nabla \hat{\psi}^{i,s}_{\mathcal{N},K_j} = \nabla \hat{\phi}^{i,s}_{\mathcal{N},K_j}, \tag{30}$$

where we use the notation $\partial_s := \frac{\partial}{\partial s}$ *and for all* $\hat{\zeta}_{\mathcal{N}} \in S^q(Y,\mathcal{N})$*, we have*

$$\int_Y a_{x_\tau,s}(y) \nabla \hat{\phi}^{i,s}_{\mathcal{N},K_j}(y) \cdot \nabla \hat{\zeta}_{\mathcal{N}}(y) dy = -\int_Y \partial_s a_{x_\tau,s}(\nabla \hat{\psi}^{i,s}_{\mathcal{N},K_j}(y) + \mathbf{e}_i) \cdot \nabla \hat{\zeta}_{\mathcal{N}}(y) dx. \tag{31}$$

Proof. This is a standard result for FEM problems depending smoothly on a parameter (see e.g. Lemma 6.1 in [12] for details). □

Parametrized cell solution space. We consider a compact subspace $\mathcal{D}$ of $\Omega \times \mathbb{R}$ (the choice of $\mathcal{D}$ is discussed in Sect. 5). For any randomly chosen parameter $(x_\tau, s) \in \mathcal{D}$, we define the map G_{x_τ} from the physical sampling domain $T_{\delta_\tau} = x_\tau + (-\delta/2, \delta/2)^d$ centered at x_τ to the reference domain $Y = (-1/2, 1/2)^d$ and the index τ of T_{δ_τ} will often be omitted in the notation when this causes no ambiguity. We consider (28), (31) with a tensor $a_{x_\tau,s}(y) = a^\varepsilon(G_{x_\tau}(y), s)$. Next indexed by $\{(T_\delta, s, \mathbf{e}_\eta); (T_\delta, s) \in \mathcal{D} \text{ and } \eta = 1, \cdots, d\}$ [3], we define the parametrized cell solution space $\mathcal{M}^{\mathcal{N}}(Y) \subset W(Y)^2$ given by

$$\mathcal{M}^{\mathcal{N}}(Y) := \{(\hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta}, \partial_s \hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta}); (T_\delta, s) \in \mathcal{D} \text{ and } \eta = 1, \cdots, d\}, \tag{32}$$

where $\hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta} \in S^q(Y,\mathcal{N})$, $\partial_s \hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta} \in S^q(Y,\mathcal{N})$ are the solutions of (28),(31) associated with the mapping G_{x_τ} and the Hilbert space $W(Y)$ is defined in either (8) or (9). On the Hilbert product space $W(Y)^2$ we define the norms

$$\|(u,v)\|_{\mathcal{W}\times\mathcal{W}} := (\|u\|^2_{\mathcal{W}} + \|v\|^2_{\mathcal{W}})^{1/2} \quad \text{and} \quad \|(u,v)\|_{\mathcal{E}\times\mathcal{E},T_\delta,s} := (\|u\|^2_{\mathcal{E},T_\delta,s} + \|v\|^2_{\mathcal{E},T_\delta,s})^{1/2}. \tag{33}$$

The goal of the Greedy procedure described below is to find an $N-$dimensional subspace of $\mathcal{M}^{\mathcal{N}}(Y)$, called $\overline{S}_N(Y)$, that minimizes the projection error of functions in $\mathcal{M}^{\mathcal{N}}(Y)$ over other choices of $N-$dimensional subspaces. We emphasize that the derivative functions $\partial_s \hat{\xi}^{\eta,s}_{\mathcal{N},T_\delta}$ involved in the definition (32) of $\mathcal{M}^{\mathcal{N}}(Y)$ are considered only for the analysis, but should not be computed explicitly in the implementation. Hence the solution of the online cell problem (15) 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 in the offline stage is conducted by an a posteriori error estimator which allows to control the accuracy of our output of interest (the numerically homogenized tensor) [37, 21].

Assume that the RB space of dimension l, denoted by $\overline{S}_l(Y)$, is available (its construction will be detailed in Algorithm 3.4). Given the parameters (x_τ, s, i), consider $(\hat{\xi}^{i,s}_{\mathcal{N},T_\delta}, \partial_s \hat{\xi}^{i,s}_{\mathcal{N},T_\delta})$, $(\hat{\xi}^{i,s}_{l,T_\delta}, \partial_s \hat{\xi}^{i,s}_{l,T_\delta})$ the solutions of (28),(31) in $S^q(Y,\mathcal{N})^2$ and $\overline{S}_l(Y)$, respectively (i.e. with test functions $(z_{\mathcal{N}}, \zeta_{\mathcal{N}})$ in $S^q(Y,\mathcal{N})^2$ and $\overline{S}_l(Y)$, respectively). We then consider

$$\hat{e}^{i,s}_{l,T_\delta} = \hat{\xi}^{i,s}_{l,T_\delta} - \hat{\xi}^{i,s}_{\mathcal{N},T_\delta}, \tag{34}$$

$$\partial_s \hat{e}^{i,s}_{l,T_\delta} = \partial_s \hat{\xi}^{i,s}_{l,T_\delta} - \partial_s \hat{\xi}^{i,s}_{\mathcal{N},T_\delta}. \tag{35}$$

[3] $\mathcal{D}$ should be chosen such that $T_\delta \subset \Omega$, for all $(x_\tau, s) \in \mathcal{D}$.

We derive an a posteriori estimator for both $\hat{e}^{i,s}_{l,T_\delta}$ and $\partial_s \hat{e}^{i,s}_{l,T_\delta}$ that will be analyzed in Lemma 3.3. We have that

$$b(\hat{e}^{i,s}_{l,T_\delta}, \hat{z}_{\mathcal{N}}) = b(\hat{\xi}^{i,s}_{l,T_\delta}, \hat{z}_{\mathcal{N}}) - l_i(\hat{z}_{\mathcal{N}}),\ \forall \hat{z}_{\mathcal{N}} \in S^q(Y,\mathcal{N}), \tag{36}$$

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

$$b(\hat{e}^{i,s}_{l,T_\delta}, \hat{z}_{\mathcal{N}}) = (\bar{e}^{i,s}_{l,T_\delta}, \hat{z}_{\mathcal{N}})_{\mathcal{W}},\ \forall \hat{z}_{\mathcal{N}} \in S^q(Y,\mathcal{N}). \tag{37}$$

We then define the residual of the a posteriori error estimator as

$$\Delta^{i,s}_{l,T_\delta} := \frac{\|\bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}}{\sqrt{\lambda_{LB}}} + \frac{\|\partial_s \bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}}{\sqrt{\lambda_{LB}}}, \tag{38}$$

where λ_{LB} is an approximation of the coercivity constant λ defined in (2). We notice that the first term in (38) is the standard residual used for linear problems [36, 38]. 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 3.2. *To compute $\bar{e}^{i,s}_{l,T_\delta}$ in (38), we first observe that we need to solve (37), which depends on the parameters (x_τ, s, i). Thanks to the affine representation of the tensor, (37) can be decomposed into several parameter independent FE problems that can be precomputed before applying the Greedy procedure in the offline stage [36, 38] and hence $\|\bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}$ is cheap to compute. 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 (37) twice with parameters s and $s+\sqrt{eps}$, respectively. In the analysis, we shall neglect the error of the above finite difference. The above finite difference introduces an small error of size $\mathcal{O}(\sqrt{eps})$ in the Jacobian involved in the Newton method. We highlight that this does not affect neither the convergence of the Newton method obtained in practice, nor the RB-FE-HMM solution accuracy which is independent of the Jacobian approximation.

The next lemma gives a bound for the a posteriori error in the output of interest in terms of the norms (33). It is a generalization of the result [3, 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 (37) and the residual $\Delta^{i,s}_{l,T_\delta}$ defined in (38). Define

$$(a^0_{\mathcal{N},T_\delta}(x_\tau,s))_{ij} = \int_Y a_{x_\tau,s}(y)\left(\nabla \hat{\xi}^{i,s}_{\mathcal{N},T_\delta}(y) + \mathbf{e}_i\right)\cdot\left(\nabla \hat{\xi}^{j,s}_{\mathcal{N},T_\delta}(y) + \mathbf{e}_j\right) dy, \tag{39}$$

$$(a^0_{l,T_\delta}(x_\tau,s))_{ij} = \int_Y a_{x_\tau,s}(y)\left(\nabla \hat{\xi}^{i,s}_{l,T_\delta}(y) + \mathbf{e}_i\right)\cdot\left(\nabla \hat{\xi}^{j,s}_{l,T_\delta}(y) + \mathbf{e}_j\right) dy. \tag{40}$$

Lemma 3.3. *Assume* (2) *and* (29). *Let* $(\hat{\xi}^{i,s}_{\mathcal{N},T_\delta}, \partial_s\hat{\xi}^{i,s}_{\mathcal{N},T_\delta})$ *and* $(\hat{\xi}^{i,s}_{l,T_\delta}, \partial_s\hat{\xi}^{i,s}_{l,T_\delta})$ *be the solution of problem* (28)-(31) *in* $S^q(Y,\mathcal{N})^2$ *and* $\overline{S}_l(Y)$, *with test functions* $(z_\mathcal{N}, \zeta_N)$ *in* $S^q(Y,\mathcal{N})^2$ *and* $\overline{S}_l(Y)$, *respectively. Assume that the approximation* λ_{LB} *of the coercivity constant satisfies* $0 < \lambda_{LB} \leq \lambda$. *Consider the quantities* $\hat{e}^{i,s}_{l,T_\delta}$ *and* $\partial_s\hat{e}^{i,s}_{l,T_\delta}$ *defined in* (34). *Then*

$$\|(\hat{e}^{i,s}_{l,T_\delta}, \partial_s\hat{e}^{i,s}_{l,T_\delta})\|_{\mathcal{E}\times\mathcal{E},T_\delta,s} \leq (2 + \frac{\Lambda_2}{\lambda_{LB}})\Delta^{i,s}_{l,T_\delta}, \tag{41}$$

$$(2\Lambda_1 + \Lambda_2)^{-1}\lambda_{LB}^{1/2}\Delta^i_{l,T_\delta} \leq \|(\hat{e}^{i,s}_{l,T_\delta}, \partial_s\hat{e}^{i,s}_{l,T_\delta})\|_{\mathcal{W}\times\mathcal{W}} \leq (2\lambda_{LB}^{-1/2} + \Lambda_2\lambda_{LB}^{-3/2})\Delta^i_{l,T_\delta}, \tag{42}$$

$$|(a^0_{\mathcal{N},T_\delta}(s))_{ij} - (a^0_{l,T_\delta}(s))_{ij}| + |\partial_s(a^0_{\mathcal{N},T_\delta}(s))_{ij} - \partial_s(a^0_{l,T_\delta}(s))_{ij}| \leq 3\Big(1 + \frac{\Lambda_2}{\lambda_{LB}}\Big)\Delta^{i,s}_{l,T_\delta}\Delta^{j,s}_{l,T_\delta}, \tag{43}$$

where Λ_1, Λ_2 *are the constants in* (29),(2) *and* $\|\cdot\|_{\mathcal{E},T_\delta,s}$ *is the energy norm defined in* (26).

Proof of Lemma 3.3. Taking $\hat{z}_\mathcal{N} = \hat{e}^{i,s}_{l,T_\delta}$ in (37) and using (27) yields successively,

$$\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{E},T_\delta,s} \leq \Delta^{i,s}_{l,T_\delta}, \tag{44}$$

$$\|\hat{e}^{i,s}_{l,T_\delta}\|_\mathcal{W} \leq \frac{\Delta^{i,s}_{l,T_\delta}}{\sqrt{\lambda_{LB}}}. \tag{45}$$

A consequence of (39),(40), and the symmetry of the tensor, is the identity

$$\begin{aligned}&(a^0_{l,T_\delta}(x_\tau, s))_{ij} - (a^0_{\mathcal{N},T_\delta}(x_\tau, s))_{ij}\\ &= \int_Y a_{x_\tau,s}(y)\left(\nabla\hat{\xi}^{i,s}_{\mathcal{N},T_\delta}(y) - \nabla\hat{\xi}^{i,s}_{l,T_\delta}(y)\right)\cdot\left(\nabla\hat{\xi}^{j,s}_{\mathcal{N},T_\delta}(y) - \nabla\hat{\xi}^{j,s}_{l,T_\delta}(y)\right)dy.\end{aligned} \tag{46}$$

We deduce from the Cauchy-Schwarz inequality and (44),

$$|(a^0_{\mathcal{N},T_\delta}(s))_{ij} - (a^0_{l,T_\delta}(s))_{ij}| \leq \Delta^{i,s}_{l,T_\delta}\Delta^{j,s}_{l,T_\delta}. \tag{47}$$

Using Lemma 3.1, we obtain after differentiation of (37) with respect to the parameter s,

$$(\partial_s\bar{e}^{i,s}_{l,T_\delta}, \hat{z}_\mathcal{N})_\mathcal{W} = \int_Y \partial_s a_{x_\tau,s}\nabla\hat{e}^{i,s}_{l,T_\delta}\cdot\nabla\hat{z}_\mathcal{N}dy + \int_Y a_{x_\tau,s}\nabla(\partial_s\hat{e}^{i,s}_{l,T_\delta})\cdot\nabla\hat{z}_\mathcal{N}dy,\ \forall\hat{z}_\mathcal{N} \in S^q(Y,\mathcal{N}). \tag{48}$$

We take $\hat{z}_\mathcal{N} = \partial_s\hat{e}^{i,s}_{l,T_\delta}$ in (48) 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 (27),(29),

$$\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|^2_{\mathcal{E},T_\delta,s} \leq \lambda_{LB}^{-1/2}\|\partial_s\bar{e}^{i,s}_{l,T_\delta}\|_\mathcal{W}\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{E},T_\delta,s} + \Lambda_2\|\hat{e}^{i,s}_{l,T_\delta}\|_\mathcal{W}\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|_\mathcal{W}$$

which gives, using (45),(27),

$$\|\partial_s\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{E},T_\delta,s} \leq (1 + \Lambda_2/\lambda_{LB})\Delta^{i,s}_{l,T_\delta}. \tag{49}$$

The estimates (44) and (49) yield (41), and using in addition (27) proves the upper bound in (42). Next, taking $\hat{z}_\mathcal{N} = \bar{e}^{i,s}_{l,T_\delta}$ in (37) using the Cauchy-Schwarz inequality yields $\|\bar{e}^{i,s}_{l,T_\delta}\|_\mathcal{W} \leq$

$\Lambda_1\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}$, while taking $\hat{z}_{\mathcal{N}} = \partial_s \bar{e}^{i,s}_{l,T_\delta}$ in (48) yields $\|\partial_s \bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}} \leq \Lambda_2\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}} + \Lambda_1\|\partial_s \hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}$. We obtain $\Delta^i_{l,T_\delta} \leq \lambda_{LB}^{-1/2}(\Lambda_1+\Lambda_2)\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}} + \lambda_{LB}^{-1/2}\Lambda_1\|\partial_s \hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}$ which yields the lower bound in (42). We finally prove (43). Differentiating the equality (46) and using (48) with $\hat{z}_{\mathcal{N}} = \hat{e}^{j,s}_{l,T_\delta}$ we obtain (using the Cauchy-Schwarz inequality)

$$
\begin{aligned}
&|(\partial_s a^0_{\mathcal{N},T_\delta}(x_\tau, s))_{ij} - (\partial_s a^0_{l,T_\delta}(x_\tau, s))_{ij}| \\
&\leq 3\Lambda_2\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\|\hat{e}^{j,s}_{l,T_\delta}\|_{\mathcal{W}} + \|\partial_s \bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}\|\hat{e}^{j,s}_{l,T_\delta}\|_{\mathcal{W}} + \|\partial_s \bar{e}^{j,s}_{l,T_\delta}\|_{\mathcal{W}}\|\hat{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}} \\
&\leq \left(3\frac{\Lambda_2}{\lambda_{LB}} + 2\right)\Delta^{i,s}_{l,T_\delta}\Delta^{j,s}_{l,T_\delta}
\end{aligned}
$$

where we used (45) and the definition (38) in the last inequality. Finally, using (47) concludes the proof. □

Offline algorithm. We now state step by step the offline stage of the RB 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 $\mathcal{D} \subset \Omega \times \mathbb{R}$ (since the range of the parameter s can only be obtained when the macro solution $u^{H,RB}$ is computed, we propose in Sect. 5 an *ad hoc* method to find an a priori range of s) and η_n corresponds to the unit vector $\mathbf{e}_{\eta_\mathbf{n}}$ belonging to the canonical basis of $\mathbb{R}^d$. Corresponding to the N triples of $(T_{\delta_n}, s, \eta_n)$, we compute $\hat{\xi}^{\eta_n,s}_{\mathcal{N},T_{\delta_n}}$, the solution of (28) 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 based on the usual procedure of the RB methodology (see [37, 38]). 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 [3].

Algorithm 3.4 (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 $(T_{\delta_n}, s_n) \in \mathcal{D}$. Define the "training set" $\Xi_{RB} = (T_{\delta_n}, s_n, \eta_n); 1 \leq \eta_n \leq d, 1 \leq n \leq N_{train}\}$* [4].
2. *Select randomly $(T_{\delta_1}, s_1, \eta_1) \in \Xi_{RB}$ and compute $\hat{\xi}^{\eta_1,s_1}_{\mathcal{N},T_{\delta_1}}$, the solution of (28) with right-hand side $l_{\eta_1}(\cdot)$ in $S^q(Y,\mathcal{N})$, corresponding to the selected parameter $(T_{\delta_1}, s_1, \eta_1)$. Set $l = 1$ and define $\hat{\xi}_{1,\mathcal{N}}(y) = \frac{\hat{\xi}^{\eta_1,s_1}_{\mathcal{N},T_{\delta_1}}(y)}{\|\hat{\xi}^{\eta_1,s_1}_{\mathcal{N},T_{\delta_1}}\|_{\mathcal{W}}}$, and the corresponding RB space $\overline{S}_1(Y) = span\{(\hat{\xi}^{\eta_1,s_1}_{\mathcal{N},T_{\delta_1}}, \partial_s \hat{\xi}^{\eta_1,s_1}_{\mathcal{N},T_{\delta_1}})\}$.*
3. *For $l = 2, \ldots, N_{RB}$*
   a. *Compute for each $(T_\delta, s, \eta) \in \Xi_{RB}$ the residual $\Delta^{\eta,s}_{l-1,T_\delta}$ defined in (38) 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[5] $\max_{(T_\delta,s,\eta)\in\Xi_{RB}}(\Delta^{\eta,s}_{l-1,T_\delta})^2 > tol_{RB}$*, otherwise the algorithm ends.*
   b. *Compute $\hat{\xi}^{\eta_l,s_l}_{\mathcal{N},T_{\delta_l}}$ the solution of (28) in $S^q(Y,\mathcal{N})$ corresponding to the selected parameters $(T_{\delta_l}, s_l, \eta_l)$. Enlarge the RB space: $\overline{S}_l(Y) = \overline{S}_{l-1}(Y) \oplus span\{(\hat{\xi}^{\eta_l,s_l}_{\mathcal{N},T_{\delta_l}}, \partial_s \hat{\xi}^{\eta_l,s_l}_{\mathcal{N},T_{\delta_l}})\}$. Set $l = l+1$ and go back to a.*

[4] 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.

[5] Notice that the error of the outputs of interest scale like the square of the error of the cell functions (43).

We emphasize once again that the derivative functions $\partial_s \hat{\xi}^{\eta_l,s_l}_{\mathcal{N},T_{\delta_l}}$ involved in Algorithm 3.4 do not need to be computed in the implementation and shall be considered only in the analysis. Thanks to Remark 3.2 for the a posteriori error estimator evaluation, as output of the Greedy algorithm, it is sufficient to compute only the list of functions $\hat{\xi}_{l,\mathcal{N}}, l = 1, \ldots, N$ that span the space $S_N(Y) := span\{\hat{\xi}_{1,\mathcal{N}}, \ldots, \hat{\xi}_{N,\mathcal{N}}\}$. These RB functions are obtained by orthogonalizing in $W(Y)$ the functions $\hat{\xi}^{\eta_l,s_l}_{\mathcal{N},T_{\delta_l}}, l = 1, \ldots, N$ and they are defined as

$$\hat{\xi}_{l,\mathcal{N}}(y) := \frac{R_l(y)}{\|R_l\|_{\mathcal{W}}}, \quad \text{where} \quad R_l(y) := \hat{\xi}^{\eta_l,s_l}_{\mathcal{N},T_{\delta_l}}(y) - \sum_{m=1}^{l-1} (\hat{\xi}^{\eta_l,s_l}_{\mathcal{N},T_{\delta_l}}, \hat{\xi}_{m,\mathcal{N}})_{\mathcal{W}} \hat{\xi}_{m,\mathcal{N}}.$$

Remark 3.5. *Notice that we choose to orthogonalize the RB in $W(Y)$ with respect to the scalar product $(\cdot,\cdot)_{\mathcal{W}}$ rather than the RB in $W(Y)^2$ with respect to the scalar product $(\cdot,\cdot)_{\mathcal{W}^2}$ associated to the norm $\|\cdot\|_{\mathcal{W}^2}$ in (33) (as normally expected in the usual RB methodology) because this is more convenient in the implementation (avoiding the computation of $\partial_s \hat{\xi}^{\eta_l,s_l}_{\mathcal{N},T_{\delta_l}}$). 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 4.3, a slight variation of the result in [22, Corollary 4.1] and [19]), we have $\hat{\xi}^{\eta_l,s_l}_{\mathcal{N},T_{\delta_l}} \notin S_{l-1}$ and thus $\|R_l\|_{\mathcal{W}} \neq 0$ and the above orthogonalization procedure succeeds.*

4 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.

4.1 A priori error analysis

We introduce the following quantity to measure the error between the tensor a^0 of the homogenized problem (3) and the numerical homogenized tensor a^0_{N,K_j} in (17).

$$r_{HMM} := \sup_{K\in\mathcal{T}_H, x_{K_j}\in K} \|a^0(x_{K_j}, u^{H,RB}(x_{K_j})) - a^0_{N,K_j}(u^{H,RB}(x_{K_j}))\|_F, \tag{50}$$

where for a given $d \times d$ tenser A, we denote $\|A\|_F = \sqrt{\sum_{m,n} |A_{mn}|^2}$ its Frobenius norm.

Theorem 4.1. *Consider u^0 the solution of problem (3). Let $\ell \geq 1$ and $\mu = 0$ or 1. Consider a quasi-uniform macro mesh and quadrature formulas satisfying (**Q1**), (**Q2**). Assume*

$$u^0 \in H^{\ell+1}(\Omega) \cap W^{1,\infty}(\Omega), \qquad a^0_{mn} \in W^{\ell+\mu,\infty}(\Omega \times \mathbb{R}), \quad \forall m, n = 1 \ldots d.$$

Assume further that (2),(4),(7) hold and that $\partial_u a^0_{mn} \in W^{1,\infty}(\Omega\times\mathbb{R})$, 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 $m, n = 1 \ldots d$. Then, there exist $r_0 > 0$ and $H_0 > 0$ such that, provided

$$H \leq H_0 \quad \text{and} \quad r_{HMM} \leq r_0, \tag{51}$$

any solution $u^{H,RB}$ of (13) satisfies

$$\begin{aligned} \|u^0 - u^{H,RB}\|_{H^1(\Omega)} &\leq C(H^\ell + r_{HMM}) \qquad \text{if } \mu = 0, 1 \\ \|u^0 - u^{H,RB}\|_{L^2(\Omega)} &\leq C(H^{\ell+1} + r_{HMM}) \quad \text{if } \mu = 1, \end{aligned}$$

where C is independent of $H, h, N, \mathcal{N}, \varepsilon$.

Proof. We apply Lemma 7.1 (a result from [12] stated in the Appendix) with $\tilde{a}(x_{K_j}, s) = a^0_{\mathcal{N},K_j}(s)$ and $\tilde{u}^H = u^{H,RB}$. □

We next have to quantify the error r_{HMM} defined in (50) 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

$$r_{MOD} := \sup_{K\in\mathcal{T}_H, x_{K_j}\in K} \|a^0(x_{K_j}, u^{H,RB}(x_{K_j})) - \overline{a}^0(x_{K_j}, u^{H,RB}(x_{K_j}))\|_F, \tag{52}$$

$$r_{MIC} := \sup_{K\in\mathcal{T}_H, x_{K_j}\in K} \|\overline{a}^0(x_{K_j}, u^{H,RB}(x_{K_j})) - a^0_{\mathcal{N},K_j}(u^{H,RB}(x_{K_j}))\|_F, \tag{53}$$

$$r_{RB} := \sup_{K\in\mathcal{T}_H, x_{K_j}\in K} \|a^0_{\mathcal{N},K_j}(u^{H,RB}(x_{K_j})) - a^0_{N,K_j}(u^{H,RB}(x_{K_j}))\|_F, \tag{54}$$

where $\overline{a}^0(x_{K_j}, s)$ is defined as

$$(\overline{a}^0(x_{K_j}, s))_{ik} := \int_{K_{\delta_j}} a^\varepsilon(x_{K_j}, s)(\nabla\bar{\psi}^{i,s}_{K_j}(x) + \mathbf{e}_i)\cdot(\nabla\bar{\psi}^{k,s}_{K_j}(x) + \mathbf{e}_k)dx$$

and $\bar{\psi}^{i,s}_{K_j}$ is the solution of the follow problem

$$\int_{K_{\delta_j}} a^\varepsilon(x_{K_j}, s)\nabla\bar{\psi}^{i,s}_{K_j}(x)\cdot\nabla z(x)dx = -\int_{K_{\delta_j}} a^\varepsilon(x_{K_j}, s)\mathbf{e}_i\cdot\nabla z(x)dx,\ \forall z \in W(K_{\delta_j}).$$

To estimate the quantities r_{MIC} and r_{MOD}, we make the following smoothness and structure assumptions on the tensor:

(**H1**) Given the degree q of the micro FE space $S^q(K_{\delta_j}, \mathcal{T}_h)$, the cell functions $\psi^{i,s}_{K_j}$ solution of (23) satisfy the bound $|\psi^{i,s}_{K_j}|_{H^{q+1}(K_{\delta_j})} \leq C\varepsilon^{-q}\sqrt{|K_{\delta_j}|}$, with C independent of ε, the quadrature point x_{K_j}, the domain K_{δ_j}, and the parameter s for all $i = 1\dots d$.

(**H2**) 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 $W^{1,\infty}_{per}(Y)$.

We next discuss the reduced basis error. Consider the space $\mathcal{M}^{\mathcal{N}}(Y)$ as defined in (32). We want to quantify how well $\mathcal{M}^{\mathcal{N}}(Y)$ can be approximated by the linear space $\overline{S}_N(Y)$ of dimension N. Such a quantification relies on the notion Kolmogorov N-width.

Definition 4.2. *Let F be a subset of a Banach space X. We denote the distance of X to any generic $N-$dimensional subspace X_N of X by*

$$E(F; X_N) = \sup_{x\in F}\inf_{y\in X_N}\|x - y\|_W.$$

The minimal error $E(F; X_N)$ is given by the Kolmogorov N-width of F in X

$$d_N(F, X) = \inf\{E(F; X_N) : X_N \text{ a } N-\text{dimensional subspace of } X\}.$$

We say that F has an exponentially small Kolmogorov N-width if there exists constants $C, r > 0$ independent of N such that

$$d_N(F, X) \leq Ce^{-rN}. \tag{55}$$

In [22, Corollary 4.1] and [19], 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 (55) with constant

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

where $0 < \lambda_{LB} \leq \lambda$ is an estimate of the coercivity bound, then the RB algorithm converges exponentially fast with respect to the dimension N of the RB space. Such assumption (55) is motivated by [33] where it is proved in the special case of one-dimensional parameter linear elliptic problems.

Notice that problem (28)-(31) with solution $(\hat{\psi}^{i,s}_{\mathcal{N},K_j}, \partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}) \in W(Y)^2$ is not coercive due to the nonlinearity of the tensor. Nevertheless, problem (28)-(31) still satisfies the following Céa inequality (see Lemma 7.3 in the Appendix) in the Hilbert space $W(Y)^2$ with norm defined in (33),

$$\|(\hat{\psi}^{i,s}_{\mathcal{N},K_j}, \partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}) - (\hat{\psi}^{i,s}_{l,K_j}, \partial_s\hat{\psi}^{i,s}_{l,K_j})\|_{\mathcal{W}\times\mathcal{W}} \leq C_0 \inf_{z\in\overline{S}_l(Y)} \|(\hat{\psi}^{i,s}_{\mathcal{N},K_j}, \partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}) - z\|_{\mathcal{W}\times\mathcal{W}} \quad (56)$$

where we consider the solution $(\hat{\psi}^{i,s}_{l,K_j}, \partial_s\hat{\psi}^{i,s}_{l,K_j})$ of (28)-(31) 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

$$C_0 = \sqrt{\frac{\Lambda_1}{\lambda}\Big(3 + \frac{8\Lambda_2^2}{\lambda^2}\Big)} \quad (57)$$

where $\lambda, \Lambda_1, \Lambda_2$ are the coercivity and continuity bounds (2),(29). In addition, recall the a posteriori estimate (42) 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

$$C_{low} = (2\Lambda_1 + \Lambda_2)^{-1}\lambda_{LB}^{1/2}, \qquad C_{up} = 2\lambda_{LB}^{-1/2} + \Lambda_2\lambda_{LB}^{-3/2}. \quad (58)$$

We obtain the following result which states that the reduced basis method converges exponentially. This is a slight adaptation of the result in [22, Corollary 4.1] and [19].

Theorem 4.3. *In addition to* (2) *and* (29)*, assume that the parametrized cell solution space* $\mathcal{M}^{\mathcal{N}}$ *in* (32) *has an exponentially small Kolmogorov* N*-width,*

$$d_N(\mathcal{M}^{\mathcal{N}}, W(Y)^2) \leq Ce^{-rN}, \quad \text{with } r > \log((1 + C_{up}/C_{low})C_0), \quad (59)$$

with constants in (56),(58)*. Then, there exists constants* $c, \kappa > 0$ *independent of* N *such that*

$$\|\hat{\psi}^{i,s}_{N,K_j} - \hat{\psi}^{i,s}_{\mathcal{N},K_j}\|_{\mathcal{W}} \leq ce^{-\kappa N}, \qquad \|\partial_s\hat{\psi}^{i,s}_{N,K_j} - \partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}\|_{\mathcal{W}} \leq ce^{-\kappa N} \quad (60)$$

for all $K \in \mathcal{T}_H$ *and all* $x_{K_j} \in K$*, where* $\hat{\psi}^{i,s}_{\mathcal{N},K_j}$ *and* $\hat{\psi}^{i,s}_{N,K_j}$ *are the solutions of the cell problem* (28) *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 [22, Corollary 4.1] reveals that the coercivity of the problem is not needed and the Céa inequality (56) is sufficient to obtain the exponential convergence (60) of the RB algorithm using the RB space $\overline{S}_N(Y)$ constructed in the Greedy algorithm 3.4. In view of Lemma 3.3, we thus deduce Theorem 4.3 from [22, Corollary 4.1] applied to the augmented problem (28)-(31). □

Theorem 4.4. *Consider u^0 the solution of problem* (3)*, and $u^{H,RB}$ the solution of* (13)*. In addition to the assumptions of Theorem 4.1, assume that $u^{H,RB}(x_{K_j})$ is uniformly bounded for all $K \in \mathcal{T}_H, x_{K_j} \in K$. Assume further (***H1***), (***H2***), and* (59)*. Then, there exist $H_0 > 0$ and $r_0 > 0$ such at if $H \leq H_0$, $h/\varepsilon \leq r_0$, and $ce^{-\kappa N} \leq r_0$ then for $\mu = 0, 1$,*

$$\|u^0 - u^{H,RB}\|_{H^{1-\mu}(\Omega)} \leq \begin{cases} C(H^{\ell+\mu} + (\frac{h}{\varepsilon})^{2q} + \delta) + r_{RB}, & \text{if } W = W^1_{per} \text{ and } \delta/\varepsilon \in \mathbb{N}, \\ C(H^{\ell+\mu} + (\frac{h}{\varepsilon})^{2q}) + r_{RB}, & \text{if } W = W^1_{per}, \ \delta/\varepsilon \in \mathbb{N}, \text{ and } a^\varepsilon(x,s) \text{ is replaced by } a(x_{K_j}, \frac{x}{\varepsilon}, s) \text{ in (21),(14),(15),(23),} \\ C(H^{\ell+\mu} + (\frac{h}{\varepsilon})^{2q} + \delta + \frac{\varepsilon}{\delta}) + r_{RB}, & \text{if } W = H^1_0 \ (\delta > \varepsilon), \end{cases}$$

where $r_{RB} \leq \Lambda_1(ce^{-\kappa N})^2$ with Λ_1 given in (2)*. We also assume $\delta \leq r_0$ or $\delta + \varepsilon/\delta \leq r_0$ in the first and third cases, respectively. We use the notation $H^0(\Omega) = L^2(\Omega)$. The constants C are independent of $H, h, \varepsilon, \delta, N, \mathcal{N}$.*

Proof. In view of Theorem 4.1, we estimate r_{HMM}. Using (**H1**), the estimate $r_{MIC} \leq C(h/\varepsilon)^{2q}$ follows from [1] (see also [2]). Using (**H2**), the estimates $r_{MOD} = 0$, and $r_{MOD} \leq C\delta$, $r_{MOD} \leq C(\delta + \varepsilon/\delta)$ follow respectively from [8, 28]. Finally, the estimate $r_{RB} \leq \Lambda_1(ce^{-\kappa N})^2$ follows from (60) and the identity (46). □

4.2 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 (3) and (17). We define

$$r'_{HMM} := \sup_{K \in \mathcal{T}_H, x_{K_j} \in K} \left\| \partial_s a^0(x_{K_j}, u^{H,RB}(x_{K_j})) - \partial_s a^0_{\mathcal{N},K_j}(u^{H,RB}(x_{K_j})) \right\|_F . \tag{61}$$

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 [12].

Theorem 4.5. *Assume that the hypotheses of Theorem 4.1 and* (59) *hold. Then, there exist positive constants H_0, r_0 such that if*

$$H \leq H_0 \qquad \text{and} \qquad H^{-1/2} r_{HMM} + r'_{HMM} \leq r_0 \tag{62}$$

then the solution $u^{H,RB}$ of (13) *is unique.*

The proof of Theorem 4.5 relies on the convergence of the Newton method stated in the following lemma.

Lemma 4.6. *Assume that the hypotheses of Theorem 4.5 hold. Let $u^{H,RB}$ be a solution of* (13)*. Then, there exist $H_0, r_0, \nu > 0$, such that provided a smallness assumption of the form* (62)*, for all $u_0^H \in S_0^\ell(\Omega, \mathcal{T}_H)$ satisfying*

$$\sigma_H \|u_0^H - u^{H,RB}\|_{H^1(\Omega)} \leq \nu, \tag{63}$$

the sequence $\{u_k^H\}$ of the Newton method (20) *with initial value u_0^H is well defined and $\|u_{k+1}^H - u^{H,RB}\|_{H^1(\Omega)} \leq C\sigma_H \|u_k^H - u^{H,RB}\|^2_{H^1(\Omega)}$, where C is a constant independent of $H, h, k, \mathcal{N}, N, \varepsilon$.*

Proof. We apply Lemma 7.2 (a result from [12] stated in the Appendix) with $\tilde{a}(x_{K_j}, s) = a^0_{\mathcal{N},K_j}(s)$ and $\tilde{u}^H = u^{H,RB}$. □

Proof of Theorem 4.5. The proof is an immediate consequence of Lemma 4.6, where given two numerical solutions $u^{H,RB}, \tilde{u}^{H,RB}$ of (13), we apply the Newton method with the initial guess $u^H_0 := \tilde{u}^{H,RB}$. The smallness assumption (62) together with the H^1 a priori error estimate of Theorem 4.1 permits to satisfy the condition (63). □

For the estimation of r'_{HMM} in (61), we consider the decomposition

$$r'_{HMM} \leq r'_{MOD} + r'_{MIC} + r'_{RB}$$

where r'_{MOD}, r'_{MIC}, r'_{RB} are defined similarly to (52),(53),(54), respectively, with the exception that all the tensors are differentiated with respect to the s parameter. Hence hypotheses analogous to (**H1**) and (**H2**) but for $\partial_s \psi^{i,s}_{K_j}$ and $\partial_s a^\varepsilon_{mn}(x,s)$, respectively, are needed. Following [9], the quantities r'_{MOD}, r'_{MIC} satisfy analogous estimates to those of r_{MOD}, r_{MIC}. It remains to estimate

$$r'_{RB} := \sup_{K\in\mathcal{T}_H, x_{K_j}\in K} \left\| \partial_s a^0_{N,K_j}(u^{H,RB}(x_{K_j})) - \partial_s a^0_{\mathcal{N},K_j}(u^{H,RB}(x_{K_j})) \right\|_F .$$

This is done in the following lemma.

Lemma 4.7. *Assume that the hypotheses of Theorem 4.4 hold with a periodic coupling with $\delta = \varepsilon$. Assume further* (29)*, and* (59)*. Then, there exist constants $c, \kappa > 0$ such that*

$$r'_{RB} \leq (2\Lambda_1 + \Lambda_2)(ce^{-\kappa N})^2.$$

Proof. Differentiating (46) with respect to s and using Theorem 4.3 conclude the proof. □

Remark 4.8. *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 [10]. 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.*

5 Implementation issues

We give in this section details on an efficient implementation of the offline and online procedures of the proposed RB-FE-HMM.

5.1 Offline procedure

Output of the offline procedure. The output of the offline procedure is the RB space (10). Rather than storing the reduced basis functions, using the affine representation (22) described above, it is sufficient to compute the following matrices and vectors

$$(A_q)_{nm} := \int_Y a_q(y)\nabla\hat{\xi}_{n,\mathcal{N}}(y)\cdot\nabla\hat{\xi}_{m,\mathcal{N}}(y)dy, \quad (F^i_q)_m := \int_Y a_q(y)\mathbf{e}_i\cdot\nabla\hat{\xi}_{m,\mathcal{N}}(y)dy. \tag{64}$$

We emphasize that the offline stage is only operated once and the output of the offline stage can be repeatedly used for the online stage computations independently of the chosen online solver.

Remark 5.1. *Taking advantage that the a posteriori error estimator* (38) *can be efficiently computed (see Remark 3.2), it could be used in the online stage to refine the reduced basis and treat new input data. The storage of low dimensional precomputed stiffness matrices in addition to* (64) *would be needed in this case.*

Preprocess of the offline stage. In Algorithm 3.4, we assume that the value of the solution $u^H(x_\tau)$ at the quadrature point x_τ lies in a bounded real interval for each τ in the training set. But the range of u^H can only be known once the macro solution is computed due to the nonlinearity of the problem. One possibility to address this issue is to map $\mathbb{R}$ into a bounded interval. However, our numerical tests indicate that this approach fails in general unless an unreasonably large training set is used. Therefore, we propose to use an empirical algorithm to get an a priori guess of the solution range, motivated by the following classical result [31, Chapter 1.6].

Lemma 5.2. *(Voigt-Reiss' inequality) Assume $a_{x_\tau,s}(y)$ to be symmetric, uniformly elliptic and bounded for $\forall x_\tau \in \Omega, \forall s \in \mathbb{R}$ and $\forall y \in Y$. Then we have*

$$\Big(\int_Y a_{x_\tau,s}(y)^{-1}dy\Big)^{-1} \leq a^0(x_\tau,s) \leq \int_Y a_{x_\tau,s}(y)dy, \tag{65}$$

where $a^0(x_\tau,s)$ is the exact homogenized tensor defined in (5).

We then apply the following procedure: we first solve (3) replacing the homogenized tensor $a^0(x_\tau,s)$ alternatively with $\big(\int_Y a_{x_\tau,s}(y)^{-1}dy\big)^{-1}$ and $\int_Y a_{x_\tau,s}(y)dy$. We then set a maximum range $(u^{0,low}, u^{0,up})$ by taking the minimum and maximum values of the two solutions obtained from the first step. In the end, we consider the range of parameter s to be $(u^{0,low}-\alpha, u^{0,up}+\alpha)$ enlarged by $\pm\alpha$ (a safety factor of about 10%).

5.2 Online procedure

We describe here how the online stage of the RB-FE-HMM can be efficiently implemented.

Fast solution of micro problems. Owing to the affine form (22) of the tensor a^ε, the problem (15) amounts to solving an $N \times N$ linear system (recall that N, the number of RB functions, is small). Indeed, writing $v_{N,K_j}^{v^H(x_{K_j})} - v_{lin,j}^H(x) = \sum_{n=1}^N \alpha_n \xi_{n,K_j}(x)$, we observe that (15) reads

$$\begin{aligned} &\sum_{n=1}^N \alpha_n \int_{K_{\delta_j}} a^\varepsilon(x,u^H(x_{K_j}))\nabla\xi_{n,K_j}(x)\cdot\nabla\xi_{m,K_j}(x)dx \\ = \; &-\sum_{i=1}^d \int_{K_{\delta_j}} a^\varepsilon(x,u^H(x_{K_j}))\mathbf{e}_i\cdot\nabla\xi_{m,K_j}(x)dx\ \frac{\partial v_{lin,j}^H}{\partial x_i}, \end{aligned} \tag{66}$$

for all $m = 1,\dots N$. Next, again thanks to the affine representation of the tensor (here we are assuming the representation (22) for simplicity), (66) can be written as

$$\begin{aligned} &\sum_{n=1}^N \alpha_n \sum_{p=1}^P \Theta_p(x_{K_j},u^H(x_{K_j})) \int_Y a_p(y)\nabla\hat{\xi}_{n,\mathcal{N}}(y)\cdot\nabla\hat{\xi}_{m,\mathcal{N}}(y)dy \\ = &-\sum_{i=1}^d \sum_{p=1}^P \Theta_p(x_{K_j},u^H(x_{K_j})) \int_Y a_p(y)\mathbf{e}_i\cdot\nabla\hat{\xi}_{m,\mathcal{N}}(y)dy \frac{\partial v_{lin,j}^H}{\partial x_i}, \end{aligned} \tag{67}$$

or equivalently

$$\big(\sum_{p=1}^{P}\Theta_p(x_{K_j},u^H(x_{K_j}))A_p\big)\boldsymbol{\alpha}=-\sum_{i=1}^{d}\big(\sum_{p=1}^{P}\Theta_p(x_{K_j},u^H(x_{K_j}))F_p^i\big)\frac{\partial v^H_{lin,j}}{\partial x_i}, \tag{68}$$

where the $N\times N$ matrices $A_p,\ p=1,\dots,P$ and the vectors $F_p^i\in\mathbb{R}^N,\ p=1,\dots,P,\ i=1,\dots,d$ are defined by (64).

We emphasize that the matrices A_p and the vectors F_p^i are assembled and stored in the offline stage, thus (68) amounts just in building the linear combination by evaluating $\Theta_p(\cdot,\cdot)$ at the desired parameter $(x_{K_j},u^H(x_{K_j}))$ for the tensor $a_{x_\tau,s}(y)$ and solving the $N\times N$ system (68) for each micro function at the quadrature points needed to assemble (14).

Newton method implementation. We consider a sequence of $\{u_k^H\}$ in $S_0^\ell(\Omega,\mathcal{T}_H)$ and express each function in the FE basis of $S_0^\ell(\Omega,\mathcal{T}_H)$ as $u_k^H=\sum_{i=1}^{M_{macro}}U_k^i\phi_i^H$. We further denote $U_k=(U_k^1,\dots,U_k^{M_{macro}})^T$. The Newton method (20) can be written out in terms of matrices as

$$\big(B(u_k^H)+B'(u_k^H)\big)(U_{k+1}-U_k)=-B(u_k^H)U_k+F, \tag{69}$$

where $B(u_k^H),B'(u_k^H)$ are the stiffness matrices associated to the bilinear forms $B_H(z^H;\cdot,\cdot)$, $B'_H(z^H;\cdot,\cdot)$ defined in (14) and (19), respectively. Here, F is a vector associated to the source term (13), which also contains the boundary data if for instance general Dirichlet or Neumann boundary conditions are considered.

Stiffness matrices. Following the implementation in [7] we consider for each element $K\in\mathcal{T}_H$ the FE basis functions $\{\phi^H_{K,i}\}_{i=1}^{n_K}$ associated with this element and the local contribution $B_K(u_k^H)$ to the stiffness matrix $(B_K(u_k^H))_{p,q=1}^{n_K}=\sum_{j=1}^{J}(B_{K,j}(u_k^H))_{p,q=1}^{n_K}$ with

$$(B_{K,j}(u_k^H))_{p,q=1}^{n_K}=\frac{\omega_{K_j}}{|K_{\delta_j}|}\int_{K_{\delta_j}}a^\varepsilon(x,z_k^H(x_{K_j}))\nabla\varphi_{K_j,p}^{h,z^H(x_{K_j})}(x)\cdot\nabla\varphi_{K_j,q}^{h,z^H(x_{K_j})}(x)dx, \tag{70}$$

where $\varphi_{K_j,p}^{h,z^H(x_{K_j})},\varphi_{K_j,q}^{h,z^H(x_{K_j})}$ are the solutions of (15) constrained by $\phi^H_{K,p},\phi^H_{K,q}$, linearized at x_{K_j}, respectively.

Similarly to the FE-HMM implementation [12], we see by differentiating (70) that the stiffness matrix $B'(U)$ in (69) associated to the non-symmetric form $B'_H(z^H;\cdot,\cdot)$ defined in (19) is given by the sum of J products of $n_K\times n_K$ matrices

$$B'_K(u_k^H)=\sum_{j=1}^{J}\Big(\frac{\partial}{\partial s}(B_{K,j}(s))\Big|_{s=z^H(x_{K_j})}\Big)\Big(U_{K,k}(\phi^H_{K_1}(x_{K_j}),\dots,\phi^H_{K_{n_K}}(x_{K_j}))\Big)$$

where the column vector $U_{K,k}$ of size n_K gives the components of z^H in the basis $\{\phi^H_{K,i}\}_{i=1}^{n_K}$ of the macro element $K\in\mathcal{T}_H$. Here, the derivative with respect to s of the $n_K\times n_K$ matrix $B_{K,j}(s)$ can be simply approximated by the finite difference $\frac{\partial}{\partial s}(B_{K,j}(s))\approx\frac{B_{K,j}(s+\sqrt{eps})-B_{K,j}(s)}{\sqrt{eps}}$, where *eps* is the machine precision. Therefore, the cost of computing the stiffness matrices for both $B(u_k^H)$ and $B'(u_k^H)$ is about twice the cost of computing the stiffness matrix $B(u_k^H)$ alone.

6 Numerical experiments

This section is dedicated to the numerical illustration of the theoretical results in Sect. 4 and the performances of the proposed RB-FE-HMM compared to the FE-HMM. All computations are performed in Matlab on the same machine (single thread computation). We first consider a simple illustrative example and then the stationary Richards problem which has a nonaffine tensor.

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

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

where $\{z_{K_j}, \rho_{K_j}\}$ denotes the quadrature formula on the fine triangulation $\mathcal{T}_h$.

Macro FEM and quadrature formulas 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 single 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 $\hat{\omega}_i = |\hat{K}|/3,\ i = 1, 2, 3$.

CPU time notations. In the following text, we use $t_{RB}^{offline}$ to denote the CPU time of the RB-FE-HMM offline stage that constructs the RB space and t_{RB}^{online} for the CPU time of the RB-FE-HMM online stage that solves the homogenized equation with the unknown data recovered by the cell solutions in the RB space. We denote the total CPU time of the RB-FE-HMM $t_{RB}^{total} := t_{RB}^{offline} + t_{RB}^{online}$. In the comparison test, we use t_{FE-HMM} to represent the FE-HMM CPU time. All CPU times are expressed in seconds.

6.1 A simple illustrative example

We consider the model problem (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,

$$\begin{aligned} &u^\varepsilon(x) = 2x_1^2(x_1-1)^2 + 3x_2^2(x_2-1)^2 + 1 \quad \text{on } \{x_1 = 0\} \cup \{x_1 = 1\}, \\ &n \cdot (a^\varepsilon(x, u^\varepsilon(x))\nabla u^\varepsilon(x)) = 0 \quad \text{on } \{x_2 = 0\} \cup \{x_2 = 1\}. \end{aligned} \tag{71}$$

Consider a diagonal multiscale tensor with the following affine expression [6]

$$\begin{aligned} a^\varepsilon(x, s)_{11} &= (x_1^2 + 0.2) + (x_2 \sin(s\pi) + 2)(\sin(2\pi\frac{x_1}{\varepsilon}) + 2), \\ a^\varepsilon(x, s)_{22} &= (\frac{1}{s+1}e^{x_2} + 0.05) + (x_1x_2 + 1)(\sin(2\pi\frac{x_2}{\varepsilon}) + 2). \end{aligned} \tag{72}$$

[6] 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.

Using the homogenization theory [31], the corresponding homogenized tensor is also diagonal with entries given by the harmonic averages

$$a^0_{ii} = \Big(\int_Y a(x,y;s)^{-1} dy \Big)^{-1},\ i = 1, 2. \tag{73}$$

Offline stage. In the offline stage, we set the parameter space to be $\mathcal{D} = \Omega_i \times U$, 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 Lemma 5.2, we first solve (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 1 and we choose $U = [0.9, 3.66]$ adding a safety correction.

Table 1: A priori estimate for the solution range. Mesh size $= 8 \times 8$.

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 Sect. 3.4 a new a posteriori error estimator (38) 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

$$\tilde{\Delta}^{i,s}_{l,T_\delta} := \frac{\|\bar{e}^{i,s}_{l,T_\delta}\|_{\mathcal{W}}}{\sqrt{\lambda_{LB}}} \tag{74}$$

used for linear problems [3]. The offline parameters are collected in Table 2.

We observe in Table 3 that our new a posteriori error estimator (38) yields a reduced basis with only one additional function compared to the classical error estimator (74) where 8 basis functions are obtained.

Online stage: convergence rates for the P1 and P2 RB-FE-HMM. Using the computed offline output (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.

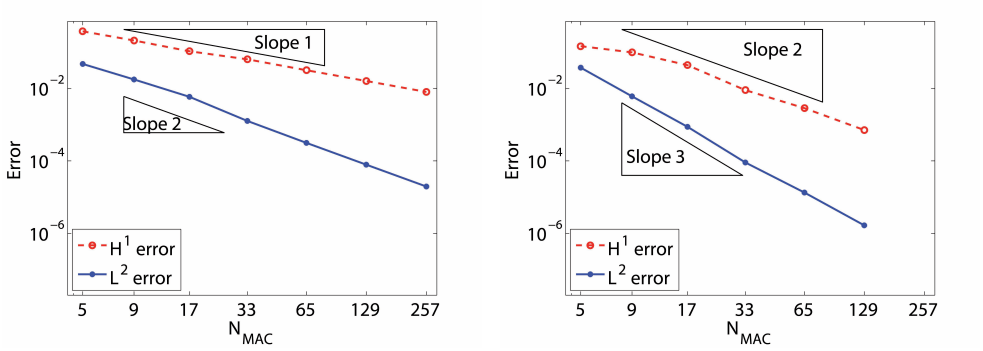
By the a priori estimates of Theorems 4.1-4.4 and the mesh size used for the offline computation, we have the bound $\Delta^2_{l,T_\delta} = \mathcal{O}(tol_{RB}) \sim 10^{-10}$ for r_{RB} and $r_{MIC} \sim 10^{-7}$. As we choose sampling domains with size $\delta = \varepsilon$ with periodic boundary condition we have $r_{MOD} = 0$ and we expect $r_{HMM} \sim 10^{-7}$. We observe in Fig. 1 (a) that for the P1 RB-FE-HMM as the

Table 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 3: A posteriori estimators

	Δ_{l,T_δ}	$\tilde{\Delta}_{l,T_\delta}$
RB space dimension	9	8
$t^{offline}_{RB}$	1300	1100



(a) P1 RB-FE-HMM (b) P2 RB-FE-HMM

Figure 1: Test problem (1)-(71)-(72). The errors $\|u^{H,RB}-u^{ref}\|_{L^2(\Omega)}$ and $\|u^{H,RB}-u^{ref}\|_{H^1(\Omega)}$ versus $N_{MAC} = 1/H$ for the P1 RB-FE-HMM and the P2 RB-FE-HMM, respectively.

macro mesh is refined, $\|u^{H,RB} - u^{ref}\|_{H^1(\Omega)}$ decays as $\mathcal{O}(N_{MAC}^{-1})$ and $\|u^{H,RB} - u^{ref}\|_{L^2(\Omega)}$ decays as $\mathcal{O}(N_{MAC}^{-2})$. For the P2 RB-FE-HMM, the decays become $\mathcal{O}(N_{MAC}^{-2})$ and $\mathcal{O}(N_{MAC}^{-3})$, respectively, which can be seen in Fig. 1 (b). Since $N_{MAC}^{-1} \sim H$ is related to the macro mesh size, this confirms the theoretical a priori error estimate, and shows that we have great flexibility to choose different macro solvers with guaranteed accuracy.

RB-FE-HMM v.s FE-HMM. In this test, we compare the efficiency and accuracy between the P1 RB-FE-HMM and the P1 FE-HMM. For the P1 FE-HMM, we set $N_{MIC} = N_{MAC}$ for each refinement step (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 4 that the H^1 and 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

Table 4: Comparison between the RB-FE-HMM and the FE-HMM.

	RB-FE-HMM			FE-HMM		
DOF	H^1 error	L^2 error	t_{RB}^{online}	H^1 error	L^2 error	t_{FE-HMM}
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

in Table 5 a comparison that takes into account the computational overhead from the offline stage for the RB-FE-HMM (as presented in Table 3 we have $t_{RB}^{offline} = 1300$ seconds). 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 5: CPU time comparison between the RB-FE-HMM (t_{RB}^{online} and t_{RB}^{total}) and the FE-HMM (t_{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%

6.2 Stationary Richards problem

We consider the stationary Richards equation for describing the fluid pressure in an unsaturated porous media

$$-\nabla \cdot (K^\varepsilon(x, u^\varepsilon(x))\nabla(u^\varepsilon(x) - x_2) = f(x) \text{ in } [0,1]^2. \tag{75}$$

with a nonlinear permeability tensor $K^\varepsilon(s)$ similar to the one in [23, Sect. 5.1] written as

$$K^\varepsilon(x,s) = (200\alpha^\varepsilon e^{-(s-2-x_2)^2\alpha^\varepsilon(x)} + (x_1 - 0.3)^2 + x_2^2 + 2)I,\ \alpha^\varepsilon(x) = \frac{0.005}{2 + 1.8\sin(2\pi\frac{x_2}{\varepsilon} - 6\pi\frac{x_1}{\varepsilon})}.$$

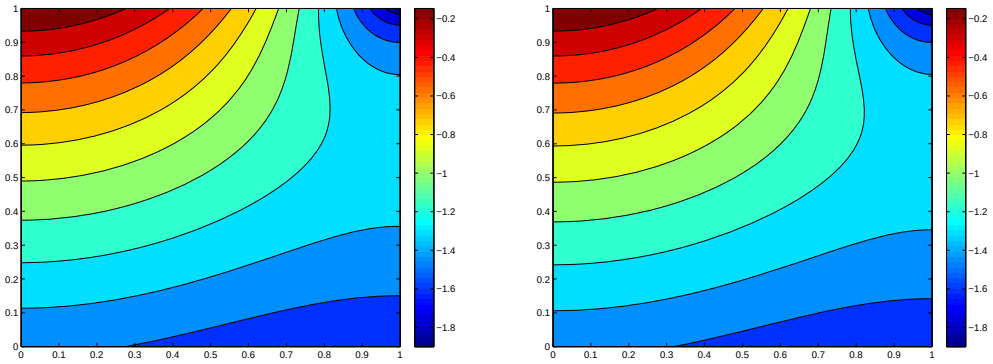
Notice that this problem can be cast in the form (1) by using the change of variable $v^\varepsilon(x) = u^\varepsilon(x) - x_2$. We set $f = 1$ and consider Dirichlet conditions on the top boundary of the domain and Neumann conditions on the rest of boundaries, that is

$$\begin{aligned} &u^\varepsilon(x) = 1 - 1.9x_1^2, \quad \text{on } [0,1] \times \{1\}, \\ &n \cdot \big(K^\varepsilon(x, u^\varepsilon(x))\nabla(u^\varepsilon(x) - x_2)\big) = 0, \quad \text{on } [0,1] \times \{0\} \cup \{0,1\} \times [0,1]. \end{aligned}$$

Offline stage. The parameters are given in Table 6. Similarly as explained for the previous example, 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 (22), we need to apply the EIM, which introduces another error term r_{EIM} in r_{HMM}, as discussed in [3] (this term can be controlled by the prescribed tolerance tol_{EIM} [16]).

Table 6: EIM and RB offline settings and output for the nonaffine test problem (75).

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	550.19
tol_{RB}	1e-9
Solver	P1-FEM
RB Basis number	4
$t_{RB}^{offline}$	912.73



(a) P1 RB-FE-HMM (b) P1 FE-HMM

Figure 2: Richards stationary problem (75). The RB-FE-HMM solution and the FE-HMM solution on a 65×65 macro mesh.

Online stage. We plot in Fig. 2 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.

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 output can be successfully used for the online computation. In Table 7, we present the errors of the Newton method iterations and the corresponding online 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.

Table 7: Richards stationary problem (75). Online CPU times and Newton iteration errors for the RB-FE-HMM.

DOF	t_{RB}^{online}	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

As predicted by our analysis, the convergence of the Newton method is insured by the offline procedure involving the new a posteriori error estimator. The above numerical test shows that the nonlinear RB-FE-HMM can be also efficiently applied to the problems with tensors written in nonaffine forms, which largely broadens the applicable range of this method.

7 Appendix

The proof of Theorem 4.1 relies on the following lemma taken from [12] and based on the analysis for standard FEM with numerical quadrature from [11]. It is a reformulation of the statement of Theorem 3.1 in [12], its proof is thus omitted.

Lemma 7.1. *Consider u^0 the solution of problem (3). Assume the assumptions of Theorem 4.1. Then, there exist $r_0 > 0$ and $H_0 > 0$ such that, for all tensors $\tilde{a}(x, s)$ satisfying (4), (2)*

and continuous on $\overline{\Omega}\times\mathbb{R}$*, for all* $H\le H_0$*, and for all solutions* $\tilde{u}^H$ *of the nonlinear FEM problem*

$$\sum_{K\in\mathcal{T}_H}\sum_{j=1}^{J}\omega_{K,j}\tilde{a}(x_{K_j},\tilde{u}^H(x_{K_j}))\nabla\tilde{u}^H(x_{K_j})\cdot\nabla w^H(x_{K_j})=\int_\Omega f(x)w^H(x)dx,\quad\forall w^H\in S_0^\ell(\Omega,\mathcal{T}_H),\tag{76}$$

provided

$$Q_H:=\sup_{K\in\mathcal{T}_H,x_{K_j}\in K}\left\|\tilde{a}(x_{K_j},u^H(x_{K_j}))-a^0(x_{K_j},u^H(x_{K_j}))\right\|_F\le r_0,$$

we have the H^1 *and* L^2 *error estimates*

$$\begin{aligned}\|u^0-\tilde{u}^H\|_{H^1(\Omega)}&\le C(H^\ell+Q_H)\qquad\text{if }\mu=0,1\\ \|u^0-\tilde{u}^H\|_{L^2(\Omega)}&\le C(H^{\ell+1}+Q_H)\quad\text{if }\mu=1,\end{aligned}$$

where C *is independent of* H,Q_H *and the tensor* $\tilde{a}$*.*

The proof of Theorem 4.5 relies on the following result which is a reformulation of Lemma 4.11 in [12].

Lemma 7.2. *Assume that the hypotheses of Lemma 7.1 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*

$$Q_H\le H\le H_0,\quad Q'_H:=\sup_{K\in\mathcal{T}_H,x_{K_j}\in K}\left\|\partial_s\tilde{a}(x_{K_j},u^H(x_{K_j}))-\partial_s a^0(x_{K_j},u^H(x_{K_j}))\right\|_F\le R_0,$$

for all $\tilde{u}^H$ *solution of 7.1 and for all for all* $u_0^H\in S_0^\ell(\Omega,\mathcal{T}_H)$ *satisfying*

$$\sigma_H\|u_0^H-\tilde{u}^H\|_{H^1(\Omega)}\le\nu,$$

the sequence $\{u_k^H\}$ *of the Newton method* (20) *applied to the problem* (76) *with initial value* u_0^H *is well defined and* $\|u_{k+1}^H-\tilde{u}^H\|_{H^1(\Omega)}\le C\sigma_H\|u_k^H-\tilde{u}^H\|^2_{H^1(\Omega)}$*, where* C *is a constant independent of* H,h,k*.*

Finally, for the proof of Theorem 4.3, we need to prove the Céa inequality (56) for the problem (28)-(31).

Lemma 7.3. *Assume* (2) *and* (29)*. Then* (56) *holds with constant* C_0 *in* (57)*.*

Proof. We denote $\hat{e}=\hat{\psi}^{i,s}_{\mathcal{N},K_j}-\hat{\psi}^{i,s}_{l,K_j}$ and $\partial_s\hat{e}=\partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}-\partial_s\hat{\psi}^{i,s}_{l,K_j}$. Considering the problem (28)-(31) with test functions in $S^q(Y,\mathcal{N})^2$ and $\overline{S}_l(Y)$, respectively, and subtracting, we deduce for all $(z_l,\zeta_l)\in\overline{S}_l$,

$$b(\partial_s\hat{e},\hat{\zeta}_l)=-\int_Y\ \partial_s a_{x_{K_j},s}(y)\nabla\hat{e}(y)\cdot\nabla\hat{\zeta}_l dy,\tag{77}$$

where the symmetric bilinear form $b(\cdot,\cdot)$ is defined in (24). Using

$$b(\partial_s\hat{e},\partial_s\hat{e})=b(\partial_s\hat{e},\partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}-\zeta_l)+b(\partial_s\hat{e},\zeta_l-\partial_s\hat{\psi}^{i,s}_{l,K_j}),$$

we obtain from the Cauchy-Schwarz inequality and (77),(29),

$$b(\partial_s\hat{e},\partial_s\hat{e})\le b(\partial_s\hat{e},\partial_s\hat{e})^{1/2}b(\partial_s\hat{e}-\zeta_l,\partial_s\hat{e}-\zeta_l)^{1/2}+\Lambda_2\|\hat{e}\|_{\mathcal{W}}(\|\partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}-\zeta_l\|_{\mathcal{W}}+\|\partial_s\hat{e}\|_{\mathcal{W}}).$$

We deduce from the Young inequality and (2),

$$\lambda\|\partial_s\hat{e}\|^2_{\mathcal{W}} \leq b(\partial_s\hat{e},\partial_s\hat{e}) \leq \Lambda_1\|\partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}-\zeta_l\|^2_{\mathcal{W}}+2\Lambda_2\|\hat{e}\|_{\mathcal{W}}\|\partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}-\zeta_l\|_{\mathcal{W}}+2\Lambda_2\|\hat{e}\|_{\mathcal{W}}\|\partial_s\hat{e}\|_{\mathcal{W}}$$

Using again the Young inequality yields

$$\|(\hat{e},\partial_s\hat{e})\|^2_{\mathcal{W}\times\mathcal{W}} \leq (1+\frac{2\Lambda_1}{\lambda})\|\partial_s\hat{\psi}^{i,s}_{\mathcal{N},K_j}-\zeta_l\|^2_{\mathcal{W}}+(1+\frac{8\Lambda_2^2}{\lambda^2})\|\hat{e}\|^2_{\mathcal{W}}.$$

Finally, the application of the Céa lemma to (28) yields $\|\hat{e}\|_{\mathcal{W}} \leq \sqrt{\frac{\Lambda_1}{\lambda}}\inf_{z_l\in S_l(Y)}\|\hat{\psi}^{i,s}_{\mathcal{N},K_j}-z_l\|_{\mathcal{W}}$ which permits to conclude the proof. □

8 Conclusion

We have extended and analyzed a reduced basis finite element multiscale method for a class of quasilinear elliptic problems of nonmonotone type. Our analysis relies on a new a posteriori estimator for the offline stage that allows to control the error coming from the greedy algorithm in a nonlinear context. Fully discrete a priori error estimates are derived in both the L^2 and the H^1 norm, taking into account the error originating from the macroscopic and microscopic meshes, from the reduced basis algorithm, and from the possible mismatch with the true micro problems originating from the microscopic boundary conditions and the sampling domain sizes. We show that the use of reduced order modeling techniques for nonlinear multiscale problems allow to considerably improve the efficiency of the nonlinear FE-HMM previously derived in [12] while maintaining its accuracy and convergence rates. In particular the cost of the Newton iterations in the online stage is essentially similar to the cost of solving single scale nonlinear problems. In a companion paper [5] we provide details on the practical implementation of the algorithm for practitioners, explain the numerical construction of a corrector and also provide 3D numerical examples for both static and time dependent nonlinear problems.

Acknowledgements. The research of A. A. and Y. B. is partially supported by the Swiss National Foundation Grant 200021_134716.

References

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

[3] 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.

[4] A. Abdulle and Y. Bai. Adaptive reduced basis finite element heterogeneous multiscale method. *to appear in Comput. Methods Appl. Mech. Engrg.*, 2013.

[5] A. Abdulle, Y. Bai, and G. Vilmart. An offline-online homogenization strategy to solve quasilinear two-scale problems at the cost of one-scale problems. *Preprint submitted for publication*, 2013.

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

[7] 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.

[8] A. Abdulle and C. Schwab. Heterogeneous multiscale fem for diffusion problems on rough surfaces. *SIAM, Multiscale Model. Simul.*, 3(1):195–220, 2005.

[9] 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.

[10] 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.

[11] 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.

[12] A. Abdulle and G. Vilmart. Analysis of the finite element heterogeneous multiscale method for quasilinear elliptic homogenization problems. *To appear in Math. Comp.*, 2013.

[13] F. Albrecht, B. Haasdonk, S. Kaulmann, and M. Ohlberger. The localized reduced basis multiscale method. *Proc. of ALGORITMY*, 2012.

[14] 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.

[15] 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.

[16] 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.

[17] J. Bear and Y. Bachmat. *Introduction to modelling of transport phenomena in porous media.* Kluwer Academic, Dordrecht, The Netherlands, 1991.

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

[19] 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.

[20] L. Boccardo and F. Murat. Homogénéisation de problèmes quasi-linéaires. *Publ. IRMA, Lille*, 3(7):13–51, 1981.

[21] S. Boyaval. Reduced-basis approach for homogenization beyond the periodic setting. *Multiscale Model. Simul.*, 7(1):466–494, 2008.

[22] 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.

[23] 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.

[24] 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.

[25] M. Chipot. Elliptic equations: an introductory course. *Birkhäuser Advanced Texts: Basler Lehrbücher. Birkhäuser Verlag, Basel*, 2009.

[26] 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.

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

[28] 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.

[29] N. Fusco and G. Moscariello. On the homogenization of quasilinear divergence structure operators. *Ann. Mat. Pura Appl.*, 146(4):1–13, 1987.

[30] J. J. Douglas and T. Dupont. A Galerkin method for a nonlinear Dirichlet problem. *Math. Comp.*, 29(131):689–696, 1975.

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

[32] 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.

[33] Y. Maday, A. Patera, and G. Turinici. A priori convergence theory for reduce-basis approximations of single-parameter elliptic partial differential equations. *J. Sci. Comput.*, 17(1-4):437–446, 2002.

[34] N. C. Nguyen. A multiscale reduced-basis method for parametrized elliptic partial differential equations with multiple scales. *J. Comp. Phys.*, 227:9807–9822, 2008.

[35] 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.

[36] 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.

[37] 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.

[38] 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.