MULTISCALE MODEL. SIMUL.
Vol. 12, No. 3, pp. 1078–1107

AN ADAPTIVE MULTISCALE FINITE ELEMENT METHOD*

PATRICK HENNING†, MARIO OHLBERGER‡, AND BEN SCHWEIZER§

Abstract. This work is devoted to an adaptive multiscale finite element method (MsFEM) for solving elliptic problems with rapidly oscillating coefficients. Starting from a general version of the MsFEM with oversampling, we derive an a posteriori estimate for the H^1-error between the exact solution of the problem and a corresponding MsFEM approximation. Our estimate holds without any assumptions on scale separation or on the type of the heterogeneity. The estimator splits into different contributions which account for the coarse grid error, the fine grid error, and the oversampling error. Based on the error estimate, we construct an adaptive algorithm that is validated in numerical experiments.

Key words. adaptivity, MsFEM, multiscale, oversampling

AMS subject classifications. 35J15, 65G99, 65N30

DOI. 10.1137/120886856

1. Introduction.

We consider the following elliptic problem: find u with

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

Here, $\Omega \subset \mathbb{R}^n$ is a bounded domain, f is a source term, and A is a rapidly oscillating, highly heterogeneous matrix, i.e., a rough coefficient function. Problems of this type have various applications in hydrology and industrial engineering. Examples are Darcy flow in porous media or heat transport in composite materials. Solving problems of the above type numerically is typically rather problematic. Standard methods require a computational grid that resolves the fine scale. This leads to an enormous computational demand which easily exceeds the capabilities of available computers. To overcome these difficulties, various approaches have been proposed and discussed within the last two decades. Just to name some of them, there is the variational multiscale method (VMM) initially proposed by Hughes et al. [34, 36], the heterogeneous multiscale finite element method (HMM) by E and Engquist [11, 12, 13], the two-scale finite element method by Matache and Schwab [42, 47], and the mortar multiscale methods by Arbogast, Peszyńska, and coauthors [45, 44, 4]. The method that we focus on is the multiscale finite element method (MsFEM) that was introduced by Hou and Wu [31]. The idea of MsFEM is to construct a set of conforming multiscale basis functions. These basis functions are determined by adding fine-scale information to the original (low-dimensional) set of Lagrange basis functions of a finite element space. The MsFEM approximation is then defined as the Galerkin approximation of the original problem in the discrete space that is spanned by the multiscale basis. Once the

*Received by the editors August 3, 2012; accepted for publication (in revised form) April 25, 2014; published electronically July 17, 2014. This work was supported by the Deutsche Forschungsgemeinschaft under contracts OH 98/4-2 and SCHW 639/3-1.
http://www.siam.org/journals/mms/12-3/88685.html

†ANMC, Section de Mathématiques, École polytechnique fédérale de Lausanne, 1015 Lausanne, Switzerland (patrick.henning@epfl.ch).

‡Institut für Numerische und Angewandte Mathematik, Westfälische Wilhelms-Universität Münster, Einsteinstr. 62, 48149 Münster, Germany (mario.ohlberger@uni-muenster.de).

§Technische Universität Dortmund, Fakultät für Mathematik, Vogelpothsweg 87, D-44227 Dortmund, Germany (ben.schweizer@tu-dortmund.de).

multiscale basis is computed, the remaining problem is low dimensional and therefore cheap to solve. The MsFEM approach can be combined with various oversampling strategies (to reduce the effect of possible boundary layers). In such approaches the multiscale basis functions are constructed in a sampling domain that is larger than the support of the original basis function. The information close to the boundary of the sampling domain is ignored and the remaining relevant "inner" information is used to construct a new set of conforming basis functions (cf. [31, 16, 15]). In our contribution we do not specify the explicit realization of the oversampling strategy; instead we work with a general conforming projection operator that might be specialized to realize specific strategies. The MsFEM has been successfully applied to a large variety of applications. For linear elliptic equations, we refer the reader to the works of Hou et al. [31, 32, 15] or to [30] for an overview; for nonlinear elliptic equations we refer the reader to [16]; two-phase flow in porous media was treated in [14], stochastic porous media flow in [1], and problems of uncertainty quantification were approached in [10]. Convergence of MsFEM approximations in general homogenization settings were, e.g., treated in [17, 18, 19, 25, 49], and explicit a priori error estimates in periodic or stochastic scenarios were obtained in [31, 32, 20, 16, 7]. However, so far we do not know of any a posteriori error estimate for MsFEM approximations. In this paper we wish to close this gap.

A posteriori error estimates for other types of multiscale methods are already available. For the HMM, see [3, 2, 27, 26, 43] for L^2-, H^1-, and energy-norm estimates. However, the results in these contributions only hold up to a remaining modeling error which vanishes in periodic or stochastic scenarios. Furthermore, the estimates do not include oversampling control. Adaptive versions of the VMM have been extensively studied by Larson and Målqvist. In [39, 41] they introduce an adaptive VMM based on energy-norm estimates, and the method in [38, 40] makes use of a duality-based a posteriori error representation formula. These approaches are based on a partition of unity by means of a coarse grid (finite element) Lagrange basis. This partition is used to split the global fine-scale problem into localized (decoupled) fine-scale problems. For each basis function, one local problem is obtained where the corresponding solution has fast decay to 0 outside of the support of the basis function. As this strategy is only possible if the considered problem is linear, there is no straightforward possibility to generalize it to nonlinear problems.

Our basic ansatz to error estimation is similar to the one proposed in [39] for the VMM, but with several new ingredients. For instance, as already mentioned, for the VMM in [39] every solution of a localized fine-scale problem rapidly decays outside of the support of the associated coarse grid basis function (i.e., partition of unity function). The local problems for the MsFEM do not have such a decay, since the corresponding right-hand sides have global support in the whole domain Ω. Larson and Målqvist use conservative fluxes to determine the computational domain for the local problems of the VMM. When the conservative flux over the boundary of the local domain is small enough, this computational domain seems to be large enough. For the MsFEM, the fluxes need to be first split into macro and micro contributions. Then the jump in the flux of the micro contributions is used as an indicator for the size of the computational domain. An additional advantage of our approach is the straightforward extension to nonlinear problems, which is not possible in the VMM setting. We show in section 5 how a nonlinear problem can be treated.

Outline. In section 2 we motivate and propose a general version of the MsFEM. In section 3 we discuss the role of conservative fluxes and present our final a posteriori error estimate and a corresponding adaptive algorithm. A proof of the main a pos-

teriori error estimate is given in section 4. In section 5 we sketch how we transfer our results to a nonlinear scenario, and in section 6 we present detailed numerical experiments to validate the method, the error estimate, and the adaptive algorithm.

2. A multiscale finite element method. In this and the next two sections, we consider a linear multiscale diffusion equation with homogeneous Dirichlet boundary condition. We choose this model problem to develop our method, but we emphasize that all the subsequent considerations can be generalized to other types of elliptic multiscale problems. We develop a similar a posteriori error estimate for a strongly monotone nonlinear problem in section 5. The linear model problem is the following:

$$\text{find} \quad u \in \mathring{H}^1(\Omega): \quad \int_\Omega A\nabla u \cdot \nabla \Phi = \int_\Omega f\Phi \quad \forall \Phi \in \mathring{H}^1(\Omega). \tag{2.1}$$

Here, $\Omega \subset \mathbb{R}^n$, $n \in \mathbb{N}_{>0}$ denotes a domain with a polygonal boundary, and we define $\mathring{H}^1(\Omega) := \overline{C_c^\infty(\Omega)}^{\|\cdot\|_{H^1(\Omega)}}$. Furthermore, we assume that $A \in (L^\infty(\Omega))^{n\times n}$ and $f \in L^2(\Omega)$. For A, we also suppose ellipticity, i.e., there exists some $\alpha \in \mathbb{R}_{>0}$ with

$$A(x)\xi \cdot \xi \geq \alpha|\xi|^2 \quad \forall \xi \in \mathbb{R}^n \text{ and for a.e. } x \in \Omega.$$

We recall that we are interested in the case that A (and hence also u) exhibits microscopic features.

2.1. Motivation. In order to motivate the MsFEM that we propose in Definition 2.2, we start to rewrite problem (2.1) in the style of a MsFEM. This helps to understand the final formulation of the method and to identify the relevant parameters, which we need to "fine-tune" to improve the approximation. The subsequent considerations are typically made in order to motivate the variational multiscale method (VMM; cf. [39]). Even though the MsFEM exhibits crucial differences to the VMM, both methods are still related and can be formulated in a common framework (cf. [28]). We therefore make the following considerations. Let $\mathcal{V}_H^c$ denote a finite-dimensional "coarse-scale" subspace of $\mathring{H}^1(\Omega)$. Let $I^c : \mathring{H}^1(\Omega) \to \mathcal{V}_H^c$ denote a projection operator. We define the continuous fine-scale space $\mathcal{V}^f$ by

$$\mathcal{V}^f := \{\phi \in \mathring{H}^1(\Omega) | I^c(\phi) = 0\}.$$

This space is a Hilbert space with the H^1-scalar product, and we can accordingly decompose the full space with some complementary space $\mathcal{V}_H^c$ as

$$\mathring{H}^1(\Omega) = \mathcal{V}^f \oplus \mathcal{V}_H^c.$$

Our next tool is a reconstruction operator $R : \mathcal{V}_H^c \to \mathring{H}^1(\Omega)$. We define this operator with the help of a corrector operator $Q : \mathcal{V}_H^c \to \mathcal{V}^f \subset \mathring{H}^1(\Omega)$ as $R = \mathrm{id} + Q$. For each basis function $\Phi_j \in \mathcal{V}_H^c$ we define the corrector $Q(\Phi_j) \in \mathcal{V}^f$ as the solution of

$$\int_\Omega A\left(\nabla\Phi_j + \nabla Q(\Phi_j)\right) \cdot \nabla\phi = \int_\Omega f\phi \quad \forall \phi \in \mathcal{V}^f. \tag{2.2}$$

Equation (2.2) is solvable by the Lax–Milgram theorem. With this construction, the $J := \dim \mathcal{V}_H^c$ functions $\{R(\Phi_j) | j \leq J\}$ form a multiscale basis for the multiscale space $V^M := \mathrm{span}\{R(\Phi_j) | j \leq J\}$. In this space, we consider the following MsFEM-like problem:

$$\text{find} \quad \bar{u} \in V^M: \quad \int_\Omega A\nabla\bar{u} \cdot \nabla\Phi_H = \int_\Omega f\Phi_H \quad \forall \Phi_H \in \mathcal{V}_H^c\,. \tag{2.3}$$

We emphasize that, at least formally, problem (2.3) is a low-dimensional problem. Nevertheless, we claim that (2.3) is actually equivalent to the original problem, i.e., that we obtain the original solution $\bar{u} = u$. To verify this claim, we first use the definition of V^M and write $\bar{u} \in V^M$ as $\bar{u} = R(u^c)$ for some $u^c \in \mathcal{V}_H^c$. Now let $\Phi \in \mathring{H}^1(\Omega)$ be an arbitrary test function. According to the decomposition of $\mathring{H}^1(\Omega)$ we can write this function as $\Phi = \Phi_H + \phi$ with $\Phi_H \in \mathcal{V}_H^c$ and $\phi \in \mathcal{V}^f$. We have

$$\int_\Omega A\nabla\bar{u} \cdot \nabla\Phi_H = \int_\Omega f\Phi_H$$

by (2.3) and

$$\int_\Omega A\nabla R(u^c) \cdot \nabla\phi = \int_\Omega f\phi$$

by the definition of Q in (2.2). Adding the two equations yields that $\bar{u} = R(u^c)$ solves

$$\int_\Omega A\nabla\bar{u} \cdot \nabla\Phi = \int_\Omega f\Phi,$$

the original problem. Since the solution is unique, there holds $\bar{u} = u$.

Our result is that an MsFEM can be formulated as follows:

$$\text{find} \quad u^c \in \mathcal{V}_H^c : \quad \int_\Omega A\nabla R(u^c) \cdot \nabla\Phi_H = \int_\Omega f\Phi_H \quad \forall\Phi_H \in \mathcal{V}_H^c\,.$$

The reconstruction is defined as $R(\Phi_H) := \Phi_H + Q(\Phi_H)$, where $Q(\Phi_H) \in \mathcal{V}^f$ solves

$$\int_\Omega A\left(\nabla\Phi_H + \nabla Q(\Phi_H)\right) \cdot \nabla\phi = \int_\Omega f\phi \quad \forall\phi \in \mathcal{V}^f. \tag{2.4}$$

In the above setting, the reconstruction $R(u^c)$ provides the true solution u. In many methods, the right-hand side f in the corrector problem (2.4) is replaced by 0 to make the operator Q linear and source independent. This can be motivated with the fact that relevant test functions ϕ in (2.4) should be rapidly oscillating to capture the variations of A. This means that such a test function exhibits a very large gradient, whereas the function itself (respectively, its L^2-norm) can be neglected in comparison to this large gradient. The left- and right-hand sides of the corrector problem are of different orders of magnitude, which prompts us to replace f by 0. This heuristic consideration can be quantified in the above setting: if $\phi \in \mathcal{V}^f$, then $I^c(\phi) = 0$ and therefore

$$\left|\int_\Omega f\phi\right| \leq \|f\|_{L^2(\Omega)}\|\phi - I^c\phi\|_{L^2(\Omega)},$$

where $\|\phi - I^c\phi\|_{L^2(\Omega)}$ behaves like $C(H)\|\nabla\phi\|_{L^2(\Omega)}$ for proper choices of I^c.

Starting from the above formulation of the problem, we can construct a fully discrete MsFEM: we replace $\mathcal{V}^f$ by a discrete space $\mathcal{W}_h$. We replace f by 0 on the right-hand side of the fine-scale problem (2.4). We localize the fine-scale problem by replacing Ω by smaller domains. These three steps are performed in the next subsection. Note that it is not clear a priori that such simplifications still lead to a reasonable method. However, we will subsequently cite two major convergence results that justify the method as it will be proposed in Definition 2.2.

2.2. Formulation of the method. In this section and in the subsequent ones, we let $\mathcal{T}_H$ and $\mathcal{T}_h$ be shape-regular triangulations of Ω. Furthermore, we assume that $\mathcal{T}_h$ is obtained from a regular refinement of $\mathcal{T}_H$. We subsequently call $\mathcal{T}_h$ the *fine grid* and $\mathcal{T}_H$ the *coarse grid*. The local coarse mesh size is defined by $H_T := |T|^{1/d}$ for $T \in \mathcal{T}_H$, and H denotes the maximum of these values. h_T and h are defined in the same way. By $\mathcal{V}_H$ and $\mathcal{W}_h$ we denote the corresponding $p1$-Lagrange finite element spaces, i.e.,

$$\mathcal{V}_H := \{\Phi_H \in \mathring{H}^1(\Omega) \cap C^0(\Omega) \;\; |\Phi_H|_T \in \mathbb{P}^1(T) \;\; \forall T \in \mathcal{T}_H\} \text{ and}$$
$$\mathcal{W}_h := \{\phi_h \in \mathring{H}^1(\Omega) \cap C^0(\Omega) \;\; |\phi_h|_S \in \mathbb{P}^1(S) \;\; \forall S \in \mathcal{T}_h\}.$$

Hence, we have the inclusion $\mathcal{V}_H \subset \mathcal{W}_h \subset \mathring{H}^1(\Omega)$.

If the structure of A can be identified (or if we have any other a priori knowledge), we should chose $\mathcal{W}_h$ to be sufficiently accurate to capture the oscillations of the data, i.e., to fulfill a condition $\inf_{v_h \in \mathcal{W}_h} \|u - v_h\|_{H^1(\Omega)} \leq \text{TOL}$. Such a condition can, for instance, be checked by using standard a priori error estimates for finite element functions. In order to impose boundary conditions on subdomains, we define $\mathring{\mathcal{W}}_h(\omega) := \mathcal{W}_h \cap \mathring{H}^1(\omega)$ for $\omega \subset \Omega$. By A_h we denote an approximation of A, which is piecewise constant with respect to the fine grid $\mathcal{T}_h$.

DEFINITION 2.1 (admissible environment). *For $T \in \mathcal{T}_H$, we call $U(T)$ an admissible environment of T if it is connected, if $T \subset U(T) \subset \Omega$, and if it is the union of elements of $\mathcal{T}_h$, i.e.,*

$$U(T) = \bigcup_{S \in \mathcal{T}_h^*} S, \quad \textit{where } \mathcal{T}_h^* \subset \mathcal{T}_h.$$

Note that the extreme choices $U(T) = T$ and $U(T) = \Omega$ provide admissible environments. Intermediate choices are useful for oversampling.

We can now introduce the MsFEM in a Petrov–Galerkin formulation with oversampling. The typical construction of an explicit multiscale finite element basis is already incorporated into the method.

DEFINITION 2.2. *Let $\mathcal{V}_H, \mathcal{W}_h \subset \mathring{H}^1(\Omega)$ be the discrete spaces defined above, and let $\mathcal{U}_H = \{U(T)| \;\; T \in \mathcal{T}_H\}$ be a family of admissible environments of the elements of $\mathcal{T}_H$. For each simplex $T \in \mathcal{T}_H$ and $\Phi_H \in \mathcal{V}_H$ we define the local corrector $Q_{h,T}(\Phi_H) \in \mathring{\mathcal{W}}_h(U(T))$ as the solution of*

$$\int_{U(T)} A_h \left(\nabla \Phi_H|_T + \nabla Q_{h,T}(\Phi_H)\right) \cdot \nabla \phi_h = 0 \quad \forall \phi_h \in \mathring{\mathcal{W}}_h(U(T)). \tag{2.5}$$

Note that by $\nabla \Phi_H|_T$, we mean the constant value of $\nabla \Phi_H$ on T extended to $U(T)$, i.e., $\nabla \Phi_H|_T := \chi_{U(T)} \fint_T \nabla \Phi_H$, where $\chi_{U(T)}$ is the indicator function of $U(T)$. The local correctors $Q_{h,T}$ are stringed together to a global corrector $\tilde{Q}_h$ by using a conforming projection $P_{H,h}$ which maps piecewise continuous functions on $\mathcal{T}_H$ to elements of $\mathcal{W}_h$,

$$P_{H,h} : \{\phi_h \in L^2(\Omega)| \;\; \phi_h \in \mathcal{W}_h(T) \;\; \forall T \in \mathcal{T}_H\} \longrightarrow \mathcal{W}_h \,.$$

With such a projection, we set the global corrector to be

$$\tilde{Q}_h(\Phi_H) := P_{H,h} \left(\sum_{T \in \mathcal{T}_H} \chi_T \;\; Q_{h,T}(\Phi_H) \right),$$

where χ_T is the indicator function of the simplex T. For $\Phi_H \in \mathcal{V}_H$, the total global reconstruction $R_h(\Phi_H)$ is finally defined as $R_h(\Phi_H) := \Phi_H + \tilde{Q}_h(\Phi_H)$. We call $R_h(u_H) \in \mathcal{W}_h$ the MsFEM approximation of u if $u_H \in \mathcal{V}_H$ solves

$$\sum_{T\in\mathcal{T}_H} \int_T A_h \nabla R_h(u_H) \cdot \nabla \Phi_H = \int_\Omega f \Phi_H \quad \forall \Phi_H \in \mathcal{V}_H. \tag{2.6}$$

Remark 2.3. The projection $P_{H,h}$ can be constructed by using a local average on the edges of T. For instance, in the case of piecewise linear functions, let $\mathcal{N}_h$ denote the set of nodes for the fine mesh $\mathcal{T}_h$ and let $\mathcal{T}_H(x_h) := \{T \in \mathcal{T}_H |\ x_h \in \overline{T}\}$ denote the set of coarse elements that share the node $x_h \in \mathcal{N}_h$. Then the projection $\tilde{Q}_h(\Phi_H) \in \mathcal{W}_h$ is uniquely defined by the following values of $\tilde{Q}_h(\Phi_H)$ in these nodes:

$$\begin{aligned} \tilde{Q}_h(\Phi_H)(x_h) &:= P_{H,h}\left(\sum_{T\in\mathcal{T}_H} \chi_T Q_{h,T}(\Phi_H)\right)(x_h) \\ &:= \frac{1}{\sharp\mathcal{T}_H(x_h)} \sum_{T\in\mathcal{T}_H(x_h)} Q_{h,T}(\Phi_H)(x_h) \qquad \forall x_h \in \mathcal{N}_h. \end{aligned} \tag{2.7}$$

A more sophisticated way of defining $P_{H,h}$ is to formulate new micro problems on an environment $U(E)$ of each macro edge $E = T_1 \cap T_2$. Boundary conditions for these problems are obtained by using the values of Q_{h,T_1} and Q_{h,T_2}. However, the standard numerical experiments in section 6 indicate that this is not necessary. The averaging operator is extremely cheap and reaches a high accuracy. Also note that we might exchange the projection of the nonconforming part and the solving of the discrete macro problem (2.6); i.e., first we solve (2.6) with the (nonconforming) operator $R_h(\Phi_H) := \Phi_H + \sum_{T\in\mathcal{T}_H} \chi_T Q_{h,T}(\Phi_H)$ and then use the conforming function $u_H + P_{H,h}(\sum_{T\in\mathcal{T}_H} \chi_T Q_{h,T}(u_H))$ as your final approximation of u. In particular, in nonlinear scenarios (or if the projection operator is nonlinear) this might be a reasonable strategy. Furthermore, such a procedure does not change the final a posteriori error estimate in Theorem 3.6 below. We still get the same error contributions, where only the term $\|A_h\nabla(Q_{h,T} - P_{H,h}(Q_{h,T}))(u_H)\|_{L^2(T)}$ must be replaced by the slightly different term $\|A\nabla(Q_{h,T} - P_{H,h}(Q_{h,T}))(u_H)\|_{L^2(T)}$. This is an easy observation when looking at the proof of Theorem 3.6 in section 4.

Remark 2.4. For the choice $U(T) = T$, Definition 2.2 provides the typical formulation of the MsFEM without oversampling. In this case we also get $R_h(u_H) \in \mathcal{W}_h$ without any projection.

Let us make a heuristic consideration. If $h \to 0$, if $U(T) \to \Omega$, and if we replace the right-hand side in the local problems (2.5) by $\int_\Omega f\phi$ and $\nabla\Phi_H|_T$ by $\nabla\Phi_H(x)$, we obtain the exact method of section 2.1, and therefore $R_h(u_H) = u$. This implies that we can tune three parameters: h (accuracy with which we solve the micro-scale equations), $U(T)$ (the computational domains for the micro-scale equations), and H (since we ignore the influence of f for the micro-scale equations). In section 3 we present an a posteriori error estimate which can be used to decide how to fine-tune these three quantities. However, let us first present the central analytical results that justify the usage of the MsFEM defined in 2.2.

2.3. Convergence results. Let $\mathcal{V}_H$ and $\mathcal{W}_h$ be as stated at the beginning of section 2.2. In this subsection we cite two convergence results for the MsFEM that justify the specific formulation in Definition 2.2. One result is abstract and does not rely on structural assumptions on the coefficient A, and the other result is more

classical, relying on structural assumptions on A but giving more quantitative answers about the speed of convergence. To state the convergence results properly, let us introduce an additional definition.

DEFINITION 2.5 (patch radius). *Let $U(T)$ be an admissible patch; then we call the set $U(T) \setminus T$ the oversampling layer of T. Furthermore, let $x_{U(T)} \in U(T)$ denote the barycenter of the patch $U(T)$; then we call $d_{U(T)}$ the patch radius if*

$$|x_{U(T)} - \bar{x}| \geq d_{U(T)} \quad \forall \bar{x} \in \partial U(T) \setminus \partial\Omega$$

and if there exists at least one point $\bar{x} \in \partial U(T) \setminus \partial\Omega$ with $|x_{U(T)} - \bar{x}| = d_{U(T)}$.

The above definition measures the extension of $U(T)$ in directions where it does not touch the boundary of Ω, i.e., where we do not know the correct boundary condition. Where $U(T)$ touches the boundary of Ω, the correct boundary condition is known. To state the first convergence result, we recall the notion of H-convergence (cf. [23]).

DEFINITION 2.6 (H-convergence). *Let $\Omega \subset \mathbb{R}^d$ denote a bounded domain, and let $(A^\epsilon)_{\epsilon>0} \subset [L^\infty(\Omega)]^{d\times d}$ denote a sequence of matrices that are uniformly bounded and coercive; i.e., there exist $\alpha, \beta \in \mathbb{R}_{>0}$ such that for a.e. $x \in \Omega$ and for all $\xi \in \mathbb{R}^d$ it holds that $\alpha|\xi|^2 \leq A^\epsilon(x)\xi \cdot \xi \leq \beta|\xi|^2$. Then we say $(A^\epsilon)_{\epsilon>0}$ is G-convergent to $A^0 \in [L^\infty(\Omega)]^{d\times d}$ if for any $F \in H^{-1}(\Omega)$ the solutions $v^\epsilon \in H_0^1(\Omega)$ of $(A^\epsilon \nabla v^\epsilon, \nabla\Phi)_{L^2(\Omega)} = F(\Phi)$ for all $\Phi \in \mathring{H}^1(\Omega)$ satisfy the relations $v^\epsilon \rightharpoonup v^0$ in $\mathring{H}^1(\Omega)$ and $A^\epsilon \nabla v^\epsilon \rightharpoonup A^0 \nabla v^0$ in $[L^2(\Omega)]^d$, where $v^0 \in \mathring{H}^1(\Omega)$ is the solution of $(A^0 \nabla v^0, \nabla\Phi)_{L^2(\Omega)} = F(\Phi)$ for all $\Phi \in \mathring{H}^1(\Omega)$.*

The abstract notion of H-convergence is very general and particularly covers the typical quasi-periodic and stochastic settings. However, it allows predictions without further structural assumptions.

THEOREM 2.7. *Let $f \in L^2(\Omega)$, and let $(A^\epsilon)_{\epsilon>0} \subset (L^\infty(\Omega))^{n\times n}$ be a sequence of matrices that is uniformly bounded and coercive; i.e., there exist $\alpha, \beta \in \mathbb{R}_{>0}$ with*

$$\alpha|\xi|^2 \leq A^\epsilon(x)\xi \cdot \xi \leq \beta|\xi|^2 \quad \forall \xi \in \mathbb{R}^n, \text{ for a.e. } x \in \Omega, \text{ and } \forall \epsilon > 0.$$

Furthermore, we assume that A^ϵ is H-convergent and let $u_H^\epsilon \in \mathcal{V}_H$ denote the corresponding nonconforming (semidiscrete) MsFEM approximation that solves

$$\sum_{T\in\mathcal{T}_H} \int_T A^\epsilon \nabla \left(u_H^\epsilon + Q_T(u_H^\epsilon)\right) \cdot \nabla\Phi_H = \int_\Omega f\Phi_H \quad \forall \Phi_H \in \mathcal{V}_H,$$

where for $\Phi_H \in \mathcal{V}_H$ the corrector $Q_T(\Phi_H) \in \mathring{H}^1(U(T))$ solves

$$\int_{U(T)} A\left(\nabla\Phi_H(x_T) + \nabla Q_T(\Phi_H)\right) \cdot \nabla\phi = 0 \quad \forall \phi \in \mathring{H}^1(U(T)).$$

If

$$\frac{\operatorname{diam}(U(T)) - \operatorname{diam}(T)}{\operatorname{diam}(T)} \to 0 \text{ for } H \to 0, \tag{2.8}$$

then we have

$$\lim_{H\to 0} \lim_{\varepsilon\to 0} \left(\sum_{T\in\mathcal{T}_H} \|u^\epsilon - (u_H^\epsilon + Q_T(u_H^\epsilon))\|^2_{H^1(T)} \right)^{\frac{1}{2}} = 0. \tag{2.9}$$

In the spirit of Theorem 2.7, the general diffusion matrix A in (2.1) can be seen as being equal to some element A^{ϵ_0} of the sequence $(A^\epsilon)_{\epsilon>0}$ for a fixed sufficiently small ϵ_0. If this is the case, Theorem 2.7 predicts a small H^1-error. The theorem was proved by Gloria for general nonsymmetric coefficients A^ϵ in [23, Theorem 5]. A generalization to nonlinear problems is given in [21, Theorem 6 and Remark 7]. At first sight, assumption (2.8) appears counterintuitive, since the oversampling converges to zero instead of being increased. However, observe that the first limit in (2.9) is ϵ, which lets the relative thickness of the oversampling layer grow to infinity. Thus, for fixed ϵ the local patches $U(T)$ should be increased more and more to get better results and ϵ should be small compared to the thickness of the oversampling layer. We end up in a typical homogenization setting, where optimal corrector problems are typically equations formulated on whole $\mathbb{R}^d$ (cf. [37]).

A more quantitative result was obtained by Hou, Wu, and Zhang in the periodic setting [33].

THEOREM 2.8. *Suppose that $d = 2$, $f \in L^2(\Omega)$, and $A^\epsilon(x) = A_p(\frac{x}{\epsilon})$ for a bounded, elliptic, symmetric, and 1-periodic matrix $A_p \in (C^3([0,1]^d))^{d\times d}$. Let $u_H^\epsilon \in \mathcal{V}_H$ denote the corresponding nonconforming MsFEM approximation that solves*

$$\sum_{T\in\mathcal{T}_H} \int_T A^\epsilon \nabla \left(u_H^\epsilon + Q_{h,T}(u_H^\epsilon)\right) \cdot \nabla \Phi_H = \int_\Omega f\Phi_H \quad \forall \Phi_H \in \mathcal{V}_H,$$

where $Q_{h,T}$ is given by (2.5). *Then it holds that*

$$\left(\sum_{T\in\mathcal{T}_H} \|u^\epsilon - (u_H^\epsilon + Q_{h,T}(u_H^\epsilon))\|^2_{H^1(T)}\right)^{\frac{1}{2}} \le C\left(\frac{\epsilon}{d_H} + H + \frac{h}{\epsilon} + \epsilon^{\frac{1}{2}}\right), \tag{2.10}$$

where d_H denotes the minimum patch radius according to Definition 2.5.

A proof of this theorem is given in [33], where the assumption $d_H \simeq H$ is made (and does hence not appear in (2.10)). However, it is possible to work out the role of d_H by following their proofs closely so that estimate (2.10) is obtained as above. Generalizations and improvements of this estimate by small modifications of the local problems are presented in [22, Theorem 3.1] and [23, sections 5.3 and 5.4].

In addition to the above estimates let us recall that if $H < \epsilon$ (i.e., if the coarse scale resolves the micro structure), then the MsFEM solution yields the same approximation order as standard FEMs on the full fine scale, i.e., $\mathrm{O}(H)$ for H^2-solutions. This result does not require structural assumptions on A (cf. [25, Proposition 1]).

Theorems 2.7 and 2.8 now give us a clear motivation for the necessity of a rigorous a posteriori error estimate that does not rely on structural assumptions. We can make out the following sources that can pollute an MsFEM approximation.

1. If h is not small enough, $\frac{h}{\epsilon}$ can dominate the error.
2. If H is not small enough, H can dominate the error.
3. If $d_H \simeq H \simeq \epsilon$ (even only locally), $\frac{\epsilon}{d_H}$ will dominate the error strongly.

Hence, an adequate adaptive algorithm must be able to detect regions where the MsFEM approximation is not sufficiently accurate (for instance, because A^ϵ yields no structure, ϵ cannot be identified, or ϵ is too coarse) to refine the coarse grid locally to enter the regime of standard FEMs. In short, we want to construct an estimate that tells us the subdomains where we can use a simplified model (i.e., use MsFEM with $H \gg \epsilon$) and the subdomains where we have to refine $\mathcal{T}_H$ locally to resolve the fine scale because scale separation and structures are missing (i.e., use MsFEM with $H < \epsilon$, which behaves like standard FEMs).

3. The a posteriori error estimate. In this section, we consider again the case that the partitions are simplicial and that the discrete spaces consist of piecewise linear functions, i.e., the case as specified at the beginning of section 2.2. The estimate that we present in this section can also be generalized to other cases. First, we require some additional definitions for formulating the a posteriori error estimate properly.

DEFINITION 3.1. *Let $\mathcal{T}_H$ and $\mathcal{T}_h$ be defined as in section* 2.2. *For an element $T \in \mathcal{T}_H$ we subsequently denote the barycenter of T by x_T. The set of the inner (coarse) faces is defined by*

$$\Gamma(\mathcal{T}_H) := \{E |\, E = T \cap \tilde{T},\, T, \tilde{T} \in \mathcal{T}_H \ \text{and}\ \operatorname{codim}(E) = 1\},$$

and $\Gamma(\mathcal{T}_h)$ is defined analogously. For any subdomain $\omega \subset \Omega$, the local restriction of $\mathcal{W}_h$ to ω is given by $\mathcal{W}_h(\omega) := \{(\phi_h)_{|\omega} |\, \phi_h \in \mathcal{W}_h\}$. A_h denotes an approximation of A, which is constant on every $S \in \mathcal{T}_h$.

To analyze the method of Definition 2.2 further, we next introduce a *local corrector basis* $(w^i_{h,T})_{T,i}$. The definition of these functions is analogous to the construction of the cell problem basis functions in homogenization theory. The basis functions $w^i_{h,T}$ will later provide an indicator for the oversampling error.

DEFINITION 3.2 (local corrector basis). *Let $i \in \mathbb{N}$ be such that $1 \le i \le n$, and let $v_i(x) := x_i$ denote the linear function that maps $x = (x_1, \dots, x_n)$ to its ith component. For that we define $w^i_{h,T} := Q_{h,T}(v_i)$. We note that the corrector functions $\{w^i_{h,T} | i = 1, \dots, n\}$ allow us to write*

$$Q_{h,T}(\Phi_H)(x) = \sum_{i=1}^{n} \partial_{x_i}\Phi_H(x_T)\ w^i_{h,T}(x). \tag{3.1}$$

Equation (3.1) can be verified as follows: let us fix T and $U(T)$, and let $e_i \in \mathbb{R}^n$ denote the ith uni vector (i.e., $(e_i)_j = 0$ for $i \neq j$ and $(e_i)_i = 1$), and therefore $\nabla v_i = e_i$. According to $w^i_{h,T} = Q_{h,T}(v_i)$ and because of the definition of $Q_{h,T}$, we obtain

$$\int_{U(T)} A_h \left(e_i + \nabla w^i_{h,T}\right) \cdot \nabla \phi_h = 0 \quad \forall \phi_h \in \mathring{\mathcal{W}}_h(U(T)).$$

Now we multiply both sides by $\partial_{x_i}\Phi_H(x_T)$ and form the sum over i, which gives us

$$\begin{aligned} 0 &= \sum_{i=1}^{n} \int_{U(T)} A_h \left(\partial_{x_i}\Phi_H(x_T) e_i + \partial_{x_i}\Phi_H(x_T)\nabla w^i_{h,T}\right) \cdot \nabla \phi_h \\ &= \int_{U(T)} A_h \left(\nabla \Phi_H(x_T) + \nabla \left(\sum_{i=1}^{n} \partial_{x_i}\Phi_H(x_T) w^i_{h,T}\right)\right) \cdot \nabla \phi_h. \end{aligned}$$

Since this equation holds for all $\phi_h \in \mathring{\mathcal{W}}_h(U(T))$ and since $\left(\sum_{i=1}^{n} \partial_{x_i}\Phi_H(x_T) w^i_{h,T}\right) \in \mathring{\mathcal{W}}_h(U(T))$ solves the same equation as $Q_{h,T}(\Phi_H) \in \mathring{\mathcal{W}}_h(U(T))$, namely,

$$\int_{U(T)} A_h \left(\nabla \Phi_H(x_T) + \nabla Q_{h,T}(\Phi_H)\right) \cdot \nabla \phi_h = 0 \quad \forall \phi_h \in \mathring{\mathcal{W}}_h(U(T)),$$

we conclude from the uniqueness of the solution (due to Lax–Milgram) that

$$Q_{h,T}(\Phi_H)(x) = \sum_{i=1}^{n} \partial_{x_i}\Phi_H(x_T) w^i_{h,T}(x).$$

Remark 3.3. Working with the functions $(w^i_{h,T})_{T,i}$ instead of using $Q_{h,T}(\Phi_H)$ also has advantages in the implementation and run time. To illustrate this, let us consider the two-dimensional case. In the original formulation, $Q_{h,T}(\Phi_H)$ has to be computed for each macroscopic basis function Φ_H with support on T. In general, these are three local problems to solve for each triangle T. Using the local corrector basis, we only need to solve two problems for each T (one for v_1 and one for v_2). Additionally, the a posteriori error estimate below uses expressions that depend on $w^i_{h,T}$.

3.1. Conservative corrector flux. In the discrete scheme, the local correctors $Q_{h,T}(\Phi_H)$ are glued together in order to obtain a global corrector $Q_h(\Phi_H)$. This will not provide the "exact corrector" $Q(\Phi_H)$ of section 2.1. Therefore, one of the errors of the discrete scheme can be determined by checking how well two local correctors $Q_{h,T_1}(\Phi_H)$ and $Q_{h,T_2}(\Phi_H)$ can be glued together. We hence study the jump in the flux over interfaces E.

When doing this, we must consider two aspects. The first aspect is that we should not look at the jump $[A_h \nabla Q_{h,T}(\Phi_H)]_E$, because this jump contains a macroscopic contribution due to the jump of $\nabla \Phi_H$ over E. This jump cannot be reduced by increasing the admissible environments $U(T_1)$ and $U(T_2)$. The right strategy is to first remove the influence of the jump of $\nabla \Phi_H$ over E and then concentrate on the jump of the corrector flux. With this purpose in mind, we introduced the local corrector basis $\{w^i_{h,T} | 1 \leq i \leq n\}$. We will evaluate $[A_h \nabla w^i_{h,T}]_E$ instead of $[A_h \nabla Q_{h,T}(\Phi_H)]$.

The second aspect is the choice of an adequate flux measure. We will use a *conservative corrector flux*. This flux is defined according to the results obtained in the work of Hughes et al. [35]. In the mentioned work, the main finding is that the continuous Galerkin method is locally conservative with respect to subdomains consisting of a union of grid elements. The conservative flux is a variational approximation of the numerical flux over the boundary of such a subdomain. This suggests that the conservative flux is a good indicator for oversampling in the following sense: the closer the sum of the conservative fluxes over a macro edge E comes to zero, the more the method behaves like a global FEM on the fine grid. In this case we get the best possible approximations. Note that we should not use the numerical flux as an indicator (i.e., the gradient jumps of the numerical approximations over the individual micro edges that contribute to E), since this is not mass conservative and since it typically does not accurately approximate the real flux. For further details and experiments on the differences between the numerical flux and the conservative flux, we refer the reader to the work of Hughes et al. [35].

Given the basis functions $w^i_{h,T}$, we define flux functions $q_{h,T,i}$ as follows. Regarding the name we note that the property of mass conservation can be verified by choosing $\phi_h = 1$ in the definition of $q_{h,T,i}$.

DEFINITION 3.4 (conservative corrector flux). *We define the conservative corrector flux $q_{h,T,i} \in \mathcal{W}_h(\partial T)$ of $w^i_{h,T}$ as the unique solution (continuous and piecewise linear on ∂T) of*

$$(-q_{h,T,i}, \phi_h)_{L^2(\partial T)} = \int_T A_h \left(e_i + \nabla w^i_{h,T}\right) \cdot \nabla \phi_h \quad \forall \phi_h \in \mathcal{W}_h(T). \tag{3.2}$$

The a posteriori error estimate will depend on the flux functions $q_{h,T,i}$. Using the error estimate in an adaptive numerical scheme leads to the following strategy: If the jump of the conservative corrector flux over a face $E = T_1 \cap T_2$ is too large, we increase the admissible environments $U(T_1)$ and $U(T_2)$. The jump is measured by the

quantities of the next definition.

DEFINITION 3.5 (flux and flux jump measures). *For any simplex $T \in \mathcal{T}_H$, we denote the outer normal function by $\nu_T : \partial T \to \mathbb{R}^n$. For two simplices T_1 and T_2, with $E := \bar{T}_1 \cap \bar{T}_2$ and for a function $g \in (L^\infty(T_1 \cup T_2))^n$ with $g_{|T_j} \in \left(C^0(T_j)\right)^n$, $j = 1, 2$, we define the jump $[g]_E : E \to \mathbb{R}$ of g over E by*

$$[g]_E(x) := \lim_{\delta \to 0} g(x - \delta \nu_{T_1}) \cdot \nu_{T_1}(x) + \lim_{\delta \to 0} g(x - \delta \nu_{T_2}) \cdot \nu_{T_2}(x).$$

With the conservative corrector fluxes $q_{h,T,i}$ of Definition 3.4 we set

$$[q_{h,i}]_E(x) := q_{h,T_1,i}(x) + q_{h,T_2,i}(x).$$

For $\Phi_H \in \mathcal{V}_H$ (which has a piecewise constant gradient), we define

$$\begin{aligned}[q_E(\Phi_H)](x) := &\left| \sum_{i=1}^n \lim_{\delta \to 0} \left(\partial_{x_i} \Phi_H(x - \delta \nu_{T_1})\right) [q_{h,i}]_E(x) \right| \\ &+ \left| \sum_{i=1}^n \lim_{\delta \to 0} \left(\partial_{x_i} \Phi_H(x - \delta \nu_{T_2}) \right) [q_{h,i}]_E(x) \right|\end{aligned}$$

and

$$\begin{aligned}[\gamma_E(\Phi_H)](x) := &\left| \sum_{i=1}^n \lim_{\delta \to 0} \left(\partial_{x_i} \Phi_H(x - \delta \nu_{T_1}) - \partial_{x_i} \Phi_H(x - \delta \nu_{T_2})\right) q_{h,T_1,i}(x) \right| \\ &+ \left| \sum_{i=1}^n \lim_{\delta \to 0} \left(\partial_{x_i} \Phi_H(x - \delta \nu_{T_1}) - \partial_{x_i} \Phi_H(x - \delta \nu_{T_2})\right) q_{h,T_2,i}(x) \right|.\end{aligned}$$

3.2. Main result and adaptive algorithm. In this section we present the a posteriori error estimate and the associated adaptive algorithm. The proof is given in section 4. The notation $f \precsim g$ is used if $f \leq Cg$, where $C > 0$ does not depend on the discretization.

THEOREM 3.6 (a posteriori error estimate). *Let u be the solution of the original problem (2.1), and let $R_h(u_H)$ be the solution of the discrete scheme of Definition 2.2. We want to investigate the error $e := u - R_h(u_H) \in \mathring{H}^1(\Omega)$. For the error, we have the following a posteriori error estimate:*

$$\begin{aligned}\|e\|_{H^1(\Omega)} \precsim &\left(\sum_{T \in \mathcal{T}_H} H_T^2 \|f\|^2_{L^2(T)} \right)^{\frac{1}{2}} + \left(\sum_{S \in \mathcal{T}_h} \|(A - A_h) \nabla R_h(u_H)\|^2_{L^2(S)} \right)^{\frac{1}{2}} \\ &+ \left(\sum_{E_S \in \Gamma(\mathcal{T}_h)} h_{E_S} \|[A_h \nabla R_h(u_H)]_{E_S}\|^2_{L^2(E_S)} \right)^{\frac{1}{2}} \\ &+ \left(\sum_{T \in \mathcal{T}_H} \|A_h \nabla (Q_{h,T} - P_{H,h}(Q_{h,T}))(u_H)\|^2_{L^2(T)} \right)^{\frac{1}{2}} \\ &+ \left(\sum_{E \in \Gamma(\mathcal{T}_H)} H_E \|[q_E(u_H)]\|^2_{L^2(E)} \right)^{\frac{1}{2}} + \left(\sum_{E \in \Gamma(\mathcal{T}_H)} H_E \|[\gamma_E(u_H)]\|^2_{L^2(E)} \right)^{\frac{1}{2}}.\end{aligned}$$

Let us consider the six terms on the right-hand side of the error estimate. The first and the last term account for the discretization error in the macro grid. The second term is due to the approximation of the diffusion matrix A. The third term is a measure for the discretization error in the local problems, and the fourth term for the enforcement of global continuity. The fifth term accounts for the error that occurs if $U(T)$ is not large enough, i.e., the error contribution which can be decreased with larger oversampling.

Remark 3.7 (effectivity of the estimate). It currently seems out of reach to derive a general effectivity result for the above estimate including the projection and the oversampling errors. If we increase the size of each environment $U(T)$ to converge to the whole domain Ω, the projection part vanishes as well and we are in a known scenario. However, we cannot forecast the interaction of projection and oversampling in a general manner so that these parts can be bounded by the local H^1-error itself. Besides these, all the other error contributions (i.e., the macro and micro grid residuals) are effective in the typical sense and can be treated in the standard way using localization with bubble functions (cf. [50]).

Remark 3.8 (effectivity of the conservative fluxes). Let w_T^i denote the exact solution of a certain local problem (i.e., $w_{h,T}^i \to w_T^i$ for $h \to 0$), and let $q_{T,i}$ denote the corresponding exact flux. Then effectivity of the conservative flux was shown by Brezzi, Hughes, and Süli [6] in simple cases (Poisson problem and certain geometrical assumptions on Ω). Here, efficiency is in the sense that the accuracy of $q_{h,T,i}$ approximating $q_{T,i}$ is of the same order as $\nabla\xi_h$ approximating $\nabla\xi$ in L^2. More general results are not available, even though they seem to hold true due to [6].

Based on the a posteriori result of Theorem 3.6 we formulate an adaptive algorithm. The contributions from fine-grid residual and approximation error are used for locally refining $\mathcal{T}_h$, the contributions from the coarse-grid residual are used for an adaptive refinement of $\mathcal{T}_H$, and contributions that depend on the conservative fluxes are used to determine the sizes of the admissible environments $U(T)$. In order to formulate the algorithm properly, we need some additional definitions.

DEFINITION 3.9. *The local error indicators for macro, micro, approximation, projection, and oversampling errors are given by*

$$\eta_T^{macro} := H_T\|f\|_{L^2(T)} + \frac{1}{\sqrt{2}} \sum_{\substack{E\in\Gamma(\mathcal{T}_H)\\ E\subset T}} \sqrt{H_E}\|[\gamma_E(u_H)]\|_{L^2(E)},$$

$$\eta_T^{micro} := \sum_{\substack{E_S\in\Gamma(\mathcal{T}_h)\\ E_S\subset T}} \sqrt{h_{E_S}}\|[A_h\nabla R_h(u_H)]_{E_S}\|_{L^2(E_S)},$$

$$\eta_T^{approx} := \sum_{\substack{S\in\mathcal{T}_h\\ S\subset T}} \|(A-A_h)\nabla R_h(u_H)\|_{L^2(S)},$$

$$\eta_T^{proje} := \|A\nabla(Q_{h,T} - P_{H,h}(Q_{h,T}))(u_H)\|_{L^2(T)},$$

$$\eta_T^{overs} := \frac{1}{\sqrt{2}} \sum_{\substack{E\in\Gamma(\mathcal{T}_H)\\ E\subset T}} \sqrt{H_E}\|[q_E(u_H)]\|_{L^2(E)}.$$

Accordingly we define the global indicators:

$$\eta^* := \left(\sum_{T\in\mathcal{T}_H} (\eta_T^*)^2\right)^{\frac{1}{2}},$$

where "" stands for either "macro," "micro," "approx," "proje," or "overs." The total estimated error η^{total} is defined as the sum of all the global indicators. The set of admissible environments is denoted by $\mathcal{U} := \{U(T)|T \in \mathcal{T}_H\}$. By "increasing $U(T) \in \mathcal{U}$" we mean that we add another layer of elements of $\mathcal{T}_h$ to the corresponding subgrid.*

ALGORITHM: adaptiveRefine(TOL, $\mathcal{T}_H$, $\mathcal{T}_h$, $\mathcal{U}$, σ, k).

Let $c_i \in (0,1)$, $1 \le i \le 4$ with $\sum_{i=1}^{4} c_i = 1$.

while $\eta^{total} > \mathsf{TOL}$ **do**
  Compute $R_h(u_H)$ with $\mathcal{T}_H$, $\mathcal{T}_h$ and $\mathcal{U}$.
  Compute η^{total}.
  if $\eta^{total} < \mathsf{TOL}$ **then**
    break.
  end
  Compute η^{macro}, η^{micro}, η^{approx}, η^{overs}.
  if $\eta^{micro} > c_1\,\eta^{total}$ **or** $\eta^{approx} > c_2\,\eta^{total}$ **then**
    foreach $T \in \mathcal{T}_H$ **do**
      if $\eta_T^{micro} \ge \frac{1}{|\mathcal{T}_H|}\eta^{micro}$ **then**
        mark all elements of $\mathcal{T}_h \cap T$ for one refinement.
      end
    end
    Refine the grid $\mathcal{T}_h$.
  end
  if $\eta^{overs} > c_3\,\eta^{total}$ **then**
    foreach $T \in \mathcal{T}_H$ **do**
      if $\eta_T^{overs} \ge \frac{1}{|\mathcal{T}_H|}\eta^{overs}$ **then**
        increase $U(T) \in \mathcal{U}$ by k layers.
      end
    end
  end
  if $\eta^{macro} > c_4\,\eta^{total}$ **then**
    foreach $T \in \mathcal{T}_H$ **do**
      if $\eta_T^{macro} \ge \sigma\frac{1}{|\mathcal{T}_H|}\eta^{macro}$ **then**
        mark T for one refinement.
      else
        do nothing.
      end
    end
    Refine the grid $\mathcal{T}_H$.
  end
end

The input for the algorithm is a coarse-grid triangulation $\mathcal{T}_H$, a fine-grid triangulation $\mathcal{T}_h$, a set of admissible environments $\mathcal{U}$, a positive number σ describing the permissible deviation from a given average, and a positive integer k describing how many layers of fine-grid elements are added to a certain environment. Adding one layer to T means that we add all fine-grid elements to $U(T)$ that share at least one

node with a fine grid element in T. We put this into a definition.

DEFINITION 3.10. *Let $K \subset \Omega$ denote a set that consists of elements of $\mathcal{T}_h$, i.e.,*

$$K = \bigcup_{S \in \mathcal{T}_h^*} S, \quad \text{where } \mathcal{T}_h^* \subset \mathcal{T}_h.$$

We say $U(K)$ is created by enriching K by one layer if

$$U(K) = \bigcup \{S \in \mathcal{T}_h | \ \overline{S} \cap \overline{K} \neq \emptyset\}.$$

We denote $U(K)$ to be an oversampling region with k layers if this procedure is iteratively applied k times.

In the algorithm, we use phrases such as "mark T for one refinement." By "mark an element for 1 refinement" we mean that we mark this element to be bisected at the end of the current loop of the adaptive algorithm. The algorithm also includes the step "mark all elements of $\mathcal{T}_h \cap T$ for one refinement," which means all elements of $\mathcal{T}_h$ that are contained in T should be bisected at the end of the loop. This might also influence elements of $\mathcal{T}_h$ that are not contained in T in order to avoid hanging nodes. Let us note that during the run of the algorithm, $\mathcal{T}_h$ needs to remain a refinement of $\mathcal{T}_H$. This means that changing $\mathcal{T}_H$ might also mean to add some elements to $\mathcal{T}_h$. The algorithm is validated in the numerical experiments in section 6.

4. Proof of the a posteriori error estimate. In this section we prove the a posteriori error estimate in Theorem 3.6. We note that the strategy of the proof could also be generalized to other multiscale methods such as the heterogeneous multiscale method (HMM). However, in that case, one would need to define a suitable extension operator instead of a projection operator in order to glue the local solutions $Q_{h,T}$ together. One possibility would be to extend the local parts by periodicity and then use a projection to again glue them together. In comparison to the MsFEM case it seems to be more reasonable that such a projection should involve additional local problems around the interfaces where the local solutions are connected.

Proof of Theorem 3.6. Let $I_H : \mathring{H}^1(\Omega) \to \mathcal{V}_H$ and $I_h : H^1(\Omega) \to \mathcal{W}_h$ denote Clément interpolation operators. For these operators we have the following estimates (cf. [8]):

$$\|\phi - I_H(\phi)\|_{L^2(T)} \leq C_1 H_T |\phi|_{H^1(\omega_T)} \quad \forall \phi \in H^1(\Omega), \ \forall T \in \mathcal{T}_H, \tag{4.1}$$

$$\|\phi - I_H(\phi)\|_{L^2(E)} \leq C_2 H_E^{\frac{1}{2}} |\phi|_{H^1(\omega_E)} \quad \forall \phi \in H^1(\Omega), \ \forall E \in \Gamma(\mathcal{T}_H), \tag{4.2}$$

$$\|\phi - I_h(\phi)\|_{L^2(S)} \leq C_3 h_S |\phi|_{H^1(\omega_S)} \quad \forall \phi \in H^1(\Omega), \ \forall S \in \mathcal{T}_h, \tag{4.3}$$

$$\|\phi - I_h(\phi)\|_{L^2(E_S)} \leq C_4 h_{E_S}^{\frac{1}{2}} |\phi|_{H^1(\omega_{E_S})} \quad \forall \phi \in H^1(\Omega), \ \forall E_S \in \Gamma(\mathcal{T}_h). \tag{4.4}$$

In these estimates we used the notation

$$\omega_T := \bigcup_{K \in \mathcal{T}_H, \overline{K} \cap \overline{T} \neq \emptyset} \overline{K}, \quad \text{and} \quad \omega_E := \overline{T_1} \cup \overline{T_2}, \ \text{where } \overline{T_1} \cap \overline{T_2} = E,$$

and ω_S and ω_{E_S} accordingly for the fine grid.

The estimate is based on an L^2-approach, exploiting a testing procedure and Galerkin orthogonality. We use the positive constant $c_P > 0$ from the Poincaré

inequality to calculate

$$
\begin{aligned}
c_P\alpha\|e\|^2_{H^1(\Omega)} &\le \int_\Omega A\nabla e\cdot\nabla e\\
&= \int_\Omega A\nabla u\cdot\nabla e - \int_\Omega A\nabla R_h(u_H)\cdot\nabla e\\
&\overset{(2.1)}{=} \int_\Omega fe - \int_\Omega (A-A_h)\nabla R_h(u_H)\cdot\nabla e - \int_\Omega A_h\nabla R_h(u_H)\cdot\nabla e\\
&\overset{(2.6)}{=} \int_\Omega f(e-I_H(e)) - \sum_{T\in\mathcal{T}_H}\int_T A_h\nabla R_h(u_H)\cdot\nabla(e-I_H(e))\\
&\quad - \int_\Omega (A-A_h)\nabla R_h(u_H)\cdot\nabla e\\
&= \int_\Omega f(e-I_H(e)) - \int_\Omega (A-A_h)\nabla R_h(u_H)\cdot\nabla e\\
&\quad - \sum_{S\in\mathcal{T}_h}\int_S A_h\nabla R_h(u_H)\cdot\nabla\left((e-I_H(e)) - I_h(e-I_H(e))\right)\\
&\quad - \sum_{T\in\mathcal{T}_H}\int_T A_h\nabla R_h(u_H)\cdot\nabla\left(I_h(e-I_H(e))\right)\\
&= \underbrace{\int_\Omega f(e-I_H(e))}_{=:\mathrm{I}} - \underbrace{\int_\Omega (A-A_h)\nabla R_h(u_H)\cdot\nabla e}_{=:\mathrm{II}}\\
&\quad \underbrace{- \sum_{S\in\mathcal{T}_h}\int_{\partial S}(A_h\nabla R_h(u_H)\cdot\nu_S)\left((e-I_H(e)) - I_h(e-I_H(e))\right)}_{=:\mathrm{III}}\\
&\quad \underbrace{+ \sum_{T\in\mathcal{T}_H}\int_T A_h\nabla(u_H+Q_{h,T}(u_H)-R_h(u_H))\cdot\nabla\left(I_h(e-I_H(e))\right)}_{=:\mathrm{IV}}\\
&\quad \underbrace{- \sum_{T\in\mathcal{T}_H}\int_T A_h(\nabla u_H+\nabla Q_{h,T}(u_H))\cdot\nabla\left(I_h(e-I_H(e))\right).}_{=:\mathrm{V}}
\end{aligned}
$$

It remains to estimate the terms I to V. For I we obtain with (4.1)

$$
\begin{aligned}
|\mathrm{I}| &\le \sum_{T\in\mathcal{T}_H}\|f\|_{L^2(T)}\|e-I_H(e)\|_{L^2(T)} \le C\sum_{T\in\mathcal{T}_H}\|f\|_{L^2(T)}H_T\|e\|_{H^1(\omega(T))}\\
&\le C\left(\sum_{T\in\mathcal{T}_H}\|f\|^2_{L^2(T)}H_T^2\right)^{\frac12}\|e\|_{H^1(\Omega)}.
\end{aligned}
$$

Term II can be estimated directly by

$$
\begin{aligned}
|\mathrm{II}| &\le \sum_{S\in\mathcal{T}_h}\|(A-A_h)\nabla R_h(u_H)\|_{L^2(S)}\|\nabla e\|_{L^2(S)}\\
&\le \left(\sum_{S\in\mathcal{T}_h}\|(A-A_h)\nabla R_h(u_H)\|^2_{L^2(S)}\right)^{\frac12}\|e\|_{H^1(\Omega)}.
\end{aligned}
$$

For III we use (4.4) and the fact that the Clément interpolation operator is H^1-stable. Furthermore, note that $(e - I_H(e)) - I_h(e - I_H(e))$ vanishes on all faces that intersect the boundary of Ω. We obtain

$$\begin{aligned}|\mathrm{III}| &= \left| \sum_{E_S \in \Gamma(\mathcal{T}_h)} \int_{E_S} [A_h \nabla R_h(u_H)]_{E_S} \left((e - I_H(e)) - I_h(e - I_H(e)) \right) \right| \\ &\leq C \left(\sum_{E_S \in \Gamma(\mathcal{T}_h)} \| [A_h \nabla R_h(u_H)]_{E_S} \|^2_{L^2(E_S)} h_{E_S} \right)^{\frac{1}{2}} \|e\|_{H^1(\Omega)}.\end{aligned}$$

Again, using the H^1-stability of the Clément operator, we get for IV

$$\begin{aligned}|\mathrm{IV}| &= \left| \sum_{T \in \mathcal{T}_H} \int_T A_h \nabla \left(u_H + Q_{h,T}(u_H) - R_h(u_H) \right) \cdot \nabla \left(I_h(e - I_H(e)) \right) \right| \\ &= \left| \sum_{T \in \mathcal{T}_H} \int_T A_h \nabla (Q_{h,T} - \tilde{Q}_h)(u_H) \cdot \nabla \left(I_h(e - I_H(e)) \right) \right| \\ &\leq C \left(\sum_{T \in \mathcal{T}_H} \| A_h \nabla (Q_{h,T} - \tilde{Q}_h)(u_H) \|^2_{L^2(T)} \right)^{\frac{1}{2}} \|e\|_{H^1(\Omega)}.\end{aligned}$$

It remains to estimate V. We abbreviate the second factor as $\phi_h := I_h(e - I_H(e))$. Furthermore, for an edge $E \in \Gamma(\mathcal{T}_H)$, let us denote by $T_{E,1}$ and $T_{E,2}$ the two elements of $\mathcal{T}_H$ that share E, i.e., $E = T_{E,1} \cap T_{E,2}$. With this notation, we get

$$\begin{aligned}\mathrm{V} &= - \sum_{T \in \mathcal{T}_H} \int_T A_h (\nabla u_H + \nabla Q_{h,T}(u_H)) \cdot \nabla \phi_h \\ &\overset{(3.1)}{=} - \sum_{T \in \mathcal{T}_H} \sum_{i=1}^n \partial_{x_i} u_H(x_T) \int_T A_h \left(e_i + \nabla w^i_{h,T} \right) \cdot \nabla \phi_h \\ &\overset{(3.2)}{=} \sum_{T \in \mathcal{T}_H} \sum_{i=1}^n \partial_{x_i} u_H(x_T) (q_{h,T,i}, \phi_h)_{L^2(\partial T)} \\ &= \frac{1}{2} \sum_{E \in \Gamma(\mathcal{T}_H)} \sum_{i=1}^n \left(\partial_{x_i} u_H(x_{T_{E,1}}) \int_E q_{h,T_{E,1},i}\, \phi_h + \partial_{x_i} u_H(x_{T_{E,2}}) \int_E q_{h,T_{E,2},i}\, \phi_h \right) \\ &= \frac{1}{2} \sum_{E \in \Gamma(\mathcal{T}_H)} \left(\sum_{i=1}^n \left(\partial_{x_i} u_H(x_{T_{E,1}}) \int_E (q_{h,T_{E,1},i} + q_{h,T_{E,2},i}) \phi_h \right) \right. \\ &\qquad \left. - \sum_{i=1}^n \left((\partial_{x_i} u_H(x_{T_{E,1}}) - \partial_{x_i} u_H(x_{T_{E,2}})) \int_E q_{h,T_{E,2},i} \phi_h \right) \right).\end{aligned}$$

With Definition 3.5 we can write

$$\begin{aligned}|\mathrm{V}| &\leq \frac{1}{2} \sum_{E \in \Gamma(\mathcal{T}_H)} \int_E [q_E(u_H)] |\phi_h| + \frac{1}{2} \sum_{E \in \Gamma(\mathcal{T}_H)} \int_E [\gamma_E(u_H)] |\phi_h| \\ &\leq \frac{1}{2} \sum_{E \in \Gamma(\mathcal{T}_H)} \int_E [q_E(u_H)] |I_h(e - I_H(e))| + \frac{1}{2} \sum_{E \in \Gamma(\mathcal{T}_H)} \int_E [\gamma_E(u_H)] |I_h(e - I_H(e))|\end{aligned}$$

$$\leq C \left(\sum_{E \in \Gamma(\mathcal{T}_H)} H_E \| [q_E(u_H)] \|^2_{L^2(E)} \right)^{\frac{1}{2}} \|e\|_{H^1(\Omega)}$$
$$+ C \left(\sum_{E \in \Gamma(\mathcal{T}_H)} H_E \| [\gamma_E(u_H)] \|^2_{L^2(E)} \right)^{\frac{1}{2}} \|e\|_{H^1(\Omega)}.$$

Adding the estimates for I to V and dividing by $\|e\|_{H^1(\Omega)}$ yields the desired result. ∎

5. A strictly monotone nonlinear problem. In comparison to the adaptive VMM proposed by Larson and Målqvist [38, 39], our method directly generalizes to nonlinear problems. The VMM in [38, 39] is based on a partition of unity which is used to localize the fine-scale computations. The big advantage of this strategy is that this yields a natural, globally continuous fine-scale approximation of the exact solution. No projection is necessary and the homogeneous Dirichlet boundary condition for the local problems is close to the correct boundary condition. On the other hand, this strategy does not carry over to the nonlinear setting since we can no longer split the problem linearly into local problems by means of a partition of unity. Our method does not have such a restriction. Let us sketch the method and the estimate in the case of monotone operators; i.e., we assume $A(\cdot, \xi) \in (L^\infty(\Omega))^n$ and that there exist two constants $0 < \alpha \leq \beta < \infty$ such that uniformly for a.e. x in Ω:

$$\begin{aligned} (A(x,\xi_1) - A(x,\xi_2), \xi_1 - \xi_2) &\geq \alpha |\xi_1 - \xi_2|^2 \quad \text{(strong monotonicity)}, \\ |A(x,\xi_1) - A(x,\xi_2)| &\leq \beta |\xi_1 - \xi_2| \quad \text{(Lipschitz continuity)}, \\ A(x,0) &= 0. \end{aligned}$$

We then consider the problem of finding $u \in \mathring{H}^1(\Omega)$ with

$$\int_\Omega A(x, \nabla u) \cdot \nabla \Phi(x)\, dx = \int_\Omega f(x) \Phi(x)\, dx \qquad \forall \Phi \in \mathring{H}^1(\Omega). \tag{5.1}$$

There is a unique solution of (5.1) due to the Browder–Minty theorem (cf. [46]). In the following A_h denotes a piecewise x-independent strictly monotone operator (with respect to the fine grid $\mathcal{T}_h$). The formulation of the McFEM is identical to the linear case, i.e., $R_h(u_H) \in \mathcal{W}_h \subset \mathring{H}^1(\Omega)$ is the MsFEM approximation of u, if $u_H \in \mathcal{V}_H$ solves

$$\sum_{T \in \mathcal{T}_H} \int_T A_h(\cdot, \nabla R_h(u_H)) \cdot \nabla \Phi_H = \int_\Omega f \Phi_H \quad \forall \Phi_H \in \mathcal{V}_H.$$

Here, for each simplex $T \in \mathcal{T}_H$ and for $\Phi_H \in \mathcal{V}_H$ the local corrector $Q_{h,T}(\Phi_H) \in \mathring{\mathcal{W}}_h(U(T))$ is the solution of

$$\int_{U(T)} A_h(\cdot, \nabla \Phi_H(x_T) + \nabla Q_{h,T}(\Phi_H)) \cdot \nabla \phi_h = 0 \quad \forall \phi_h \in \mathring{\mathcal{W}}_h(U(T)).$$

We define the reconstruction by $R_h(\Phi_H)_{|T} := \Phi_H + Q_{h,T}(\Phi_H)$. With a suitable projection $P_{H,h}$, the final (continuous) approximation of u is given by $P_{H,h}(R_h(u_H))$.

The arising nonlinear equations can be solved with the Newton scheme framework for multiscale methods as proposed in [28].

There is only one difference in this nonlinear setting if a posteriori error estimates are considered. Due to the nonlinearity, it is not possible to separate the flux into its macro- and micro-scale parts (which we established with a *local corrector basis*). Therefore, the conservative corrector flux needs to be defined in a more general manner so that we determine the flux for the whole correction of u_H: for $\Phi_H \in \mathcal{V}_H$, we define the conservative corrector flux in the nonlinear case $q_{h,T}(\Phi_H) \in \mathcal{W}_h(\partial T)$ of $Q_{h,T}(\Phi_H)$ as the unique solution of

$$(-q_{h,T}(\Phi_H), \phi_h)_{L^2(\partial T)} = \int_T A_h\left(\cdot, \nabla\Phi_H + \nabla Q_{h,T}(\Phi_H)\right) \cdot \nabla\phi_h \quad \forall \phi_h \in \mathcal{W}_h(T).$$

In addition we set

$$[q_E^{nl}(\Phi_H)]_E(x) := q_{h,T_1}(\Phi_H)(x) + q_{h,T_2}(\Phi_H)(x).$$

With this modification, the a posteriori error estimate can be derived analogously to the preceding section.

THEOREM 5.1. *Let u denote the exact solution of problem (5.1) and $R_h(u_H)$ the corresponding MsFEM approximation. Then for the total error $e := u - P_{H,h}(R_h(u_H)) \in \mathring{H}^1(\Omega)$ we obtain*

$$\begin{aligned}
\|e\|_{H^1(\Omega)} \precsim & \left(\sum_{T\in\mathcal{T}_H} H_T^2 \|f\|^2_{L^2(T)}\right)^{\frac{1}{2}} + \left(\sum_{S\in\mathcal{T}_h} \|(A - A_h)(\cdot, \nabla R_h(u_H))\|^2_{L^2(S)}\right)^{\frac{1}{2}} \\
& + \left(\sum_{E_S\in\Gamma(\mathcal{T}_h)} h_{E_S} \|[A_h\left(\cdot, \nabla R_h(u_H)\right)]_{E_S}\|^2_{L^2(E_S)}\right)^{\frac{1}{2}} \\
& + \left(\sum_{T\in\mathcal{T}_H} \|A\left(\cdot, \nabla(Q_{h,T} - P_{H,h}(Q_{h,T}))(u_H)\right)\|^2_{L^2(T)}\right)^{\frac{1}{2}} \\
& + \left(\sum_{E\in\Gamma(\mathcal{T}_H)} H_E \|[q_E^{nl}(u_H)]\|^2_{L^2(E)}\right)^{\frac{1}{2}}.
\end{aligned}$$

With Theorem 5.1 we have an a posteriori error estimate that allows us to deal numerically with strictly monotone nonlinear problems. Regarding limitations of the above results we mention that in relevant applications in hydrology, the nonlinear multiscale problems are degenerate; see, e.g., [29] and [48]. For such a setting, the construction of an adaptive MsFEM has yet to be developed.

6. Numerical experiments. In this section we validate the a posteriori error estimate and the adaptive MsFEM in various numerical experiments. All implementations are made in C++ using the DUNE-FEM module [9] and the DUNE-SUBGRID module [24] of the software toolbox DUNE (cf. [5]). We apply the method to two model problems with increasing complexity. The first model problem involves a periodically oscillating coefficient function. Here, the exact solution is available, and we can state quantitative results such as explicit L^2- and H^1-errors. The second model problem involves a periodic structure that is locally perturbed by stripes of high conductivity within a patch of low conductivity. Here, the exact solution is not available, and we can only evaluate our adaptive MsFEM approximations by comparing them

TABLE 1

Model Problem 1. *The table contains L^2- and H^1-errors between the exact solution and different MsFEM approximations. The MsFEM approximations are obtained for different coarse-grid and fine-grid resolutions H and h. The grids are uniformly refined. Each coarse element $T \in \mathcal{T}_H$ is enriched by* 10 *layers to create the oversampling set $U(T)$. We also list the total estimated error, the coarse-grid error indicator η^{macro}, and fine-grid error indicator η^{micro}.*

H	h	$\|e\|_{L^2(\Omega)}$	$\|e\|_{H^1(\Omega)}$	η^{total}	η^{macro}	η^{micro}	η^{approx}
2^{-2}	2^{-4}	0.1669	2.4887	5.6398	4.2013	0.7787	0.2236
2^{-3}	2^{-5}	0.0810	1.9847	4.8033	3.2779	0.6314	0.1773
2^{-4}	2^{-6}	0.0243	1.0391	2.8838	1.7292	0.3177	0.0947
2^{-5}	2^{-7}	0.0074	0.5629	1.8501	0.8714	0.1689	0.0489

qualitatively with standard finite element computations on a highly resolved grid. To recall what we mean by "k oversampling layers," we refer the reader to Definition 3.10. In this section, the error indicators (obtained from Theorem 3.6 and stated in Definition 3.9) are multiplied by a factor of order $\frac{\beta}{\alpha}$, which is a typical scaling of the stability constants for these types of error estimates. In particular, for Model Problem 1 we use the factor 10.

MODEL PROBLEM 1. *Let $\Omega :=]0,1[^2$ and $\epsilon = 5 \cdot 10^{-2}$. We define*

$$u(x_1, x_2) := \sin(2\pi x_1)\sin(2\pi x_2) + \frac{\epsilon}{2}\cos(2\pi x_1)\sin(2\pi x_2)\sin\left(2\pi\frac{x_1}{\epsilon}\right),$$

which is the exact solution of the problem

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

where A is given by

$$A(x_1, x_2) := \frac{1}{8\pi^2}\begin{pmatrix} 2(2+\cos(2\pi\frac{x_1}{\epsilon}))^{-1} & 0 \\ 0 & 1+\frac{1}{2}\cos(2\pi\frac{x_1}{\epsilon}) \end{pmatrix}$$

and f by

$$f(x) := -\nabla \cdot (A(x)\nabla u(x)) \approx \sin(2\pi x_1)\sin(2\pi x_2).$$

In Table 1 we can see a listing of various L^2-, H^1-, and estimated errors. Here, the number of layers (for oversampling) is fixed at 10, but H and h are coupled by the factor 2^{-2}. Ten layers are sufficient so that the oversampling error takes only a minor role. We ignore the first row of computations (for 2^{-2} and 2^{-4}) since these values are not yet representative. We obtain an average experimental order of convergence (EOC) of 1.72 for the L^2-error and 0.91 for the H^1-error. Note that we cannot expect optimal orders of convergence for the errors, since we do not solve the local problems globally but only in local sampling domains $U(T)$. However, we still observe a very nice reduction of the error. The estimated error shows an average EOC of 0.7. Again, the error contributions of the conservative flux jumps (i.e., the indicator for the size of the oversampling error) prevent an optimal order of 1.0. Still, the reduction of the estimated error is adequate and reasonable. On the other hand, the error indicators for the coarse-grid residual (EOC 0.96), the fine-grid residual (EOC 0.95), and approximation error (EOC 0.93) show, as expected, almost optimal order. Similar results are obtained if we only regard convergence in the macro mesh size H. The

TABLE 2

Model Problem 1. *MsFEM computations for fixed fine-grid resolution with* $h = 2^{-7}$ *and fixed number of* 10 *oversampling layers. We see a listing of the errors, the total estimated error* η^{total}, *and its coarse-grid error contribution* η^{macro}.

H	$\|e\|_{L^2(\Omega)}$	$\|e\|_{H^1(\Omega)}$	η^{total}	η^{macro}
2^{-3}	0.0625	1.7719	4.1246	3.4181
2^{-4}	0.0196	0.9231	2.5679	1.7652
2^{-5}	0.0074	0.5629	1.8501	0.8714

TABLE 3

Model Problem 1. *MsFEM computations for increasing coarse- and fine-grid resolutions and for an increasing number of oversampling layers. We see a listing of the errors and the total estimated error* η^{total} *and corresponding averaged EOCs.*

H	h	Number of layers	$\|e\|_{L^2(\Omega)}$	$\|e\|_{H^1(\Omega)}$	η^{total}
2^{-3}	2^{-5}	6	0.1101	2.1897	5.6399
2^{-4}	2^{-6}	8	0.0259	1.0437	2.8856
2^{-5}	2^{-7}	10	0.0074	0.5629	1.8501
Average EOC			1.95	0.98	0.81

MsFEM computations in Table 2 are for a fixed fine-grid resolution with $h = 2^{-7}$. Here, the effect of a reduced EOC is even more pronounced due to the constant micro-mesh contributions. The EOC for the L^2-error is between 1.67 and 1.41, the EOC for the H^1-error is between 0.94 and 0.71, and the EOC for the estimated H^1-error is between 0.68 and 0.47. However, again the pure contribution of the coarse-grid residual converges with an average order of 0.99. These first experiments indicate that the various contributions of the error (coarse grid, fine grid, and oversampling) interact with each other and that they all have a significant influence on the final approximation. The total convergence of the method can be improved and comes close to the optimal order if we combine the macro and micro refinements with an increasing size of the oversampling region. This is depicted in Table 3, where we can see that the L^2-error converges with order 1.95 and the H^1-error with order 0.98.

Furthermore, the overall accuracy obtained with the method is really high. This becomes clear when we look at Figure 1, where we can see a comparison between the isolines of the exact solution and the MsFEM approximation for $H = 2^{-5}$, $h = 2^{-7}$, and 10 oversampling-layers. Except for a small discrepancy near the maximum and the minimum, the isolines match each other perfectly. Even fine-scale fluctuations are captured nicely. For the same computation, the explicit effect of the correction operator is depicted in Figure 2. Here, we see the smooth coarse part u_H that does not resolve the oscillations of u and the corresponding fine-scale correction $Q(u_H)$ which adds significant information about the oscillations to the coarse part.

The influence of the size of the oversampling set is illustrated in Table 4. In these computations, we fix the coarse grid and the fine grid with $(H, h) = (H = 2^{-4}, 2^{-8})$. The number of layers increased step by step. We see that, solely by using oversampling, the L^2-error is almost halved and the H^1-error is reduced by about 25%. Comparably, the estimated error is reduced by 30.7%, which suggests that the oversampling indicator η^{overs} does a reliable job. The largest reduction of the error appears when we add the first oversampling layer. We see the same effect for the estimated error. After this first step, we still have a permanent reduction, but

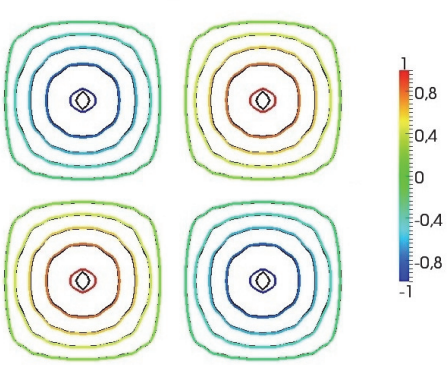
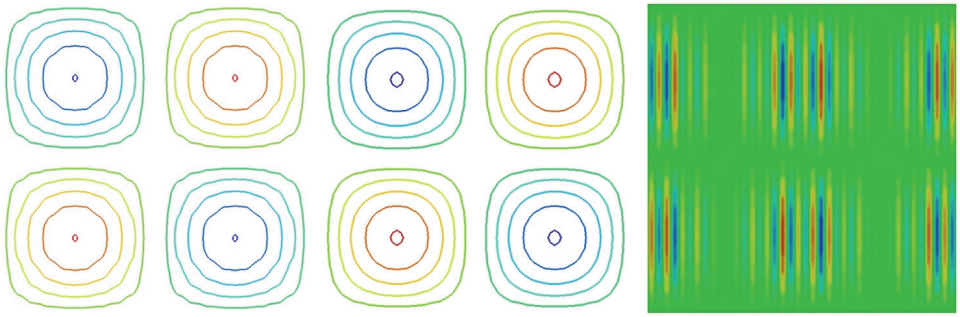


FIG. 1. *Model Problem* 1. *MsFEM computation for* $H = 2^{-5}$, $h = 2^{-7}$, *and* 10 *oversampling layers. The figure shows a comparison of the isolines of the MsFEM approximation with the isolines of the exact solution. The black lines beneath are from the MsFEM solution. (See online version for color images.)*



(a) (b) (c)

FIG. 2. *Model Problem* 1. *MsFEM computation for* $H = 2^{-5}$, $h = 2^{-7}$, *and* 10 *layers (cf. Table* 1*). Panel* (a) *shows the total MsFEM approximation* $R_h(u_H)$, *and panel* (b) *shows the coarse part* u_H *of the MsFEM approximation (isolines, colorshading for both: blue (*-1.0*) to red (*1.0*)). Panel* (c) *shows the fine-scale part* $Q(u_H)$ *of MsFEM approximation (colorshading: blue (*-0.029*) to red (*0.029*)).*

the respective reductions get smaller. Nevertheless, no visible stagnation occurs. The oversampling error indicator shows a similar behavior. All in all, we see that the jump in the conservative fluxes virtually signals whether the used oversampling sets $\mathcal{U}_H$ are really large enough. From Table 4, we also see that the contribution of the projection error can be ignored. We do not have to perform expensive fine-scale computations to improve the quality of the projection operator. Simple averaging as proposed in (2.7) produces very convenient results. It is obvious from the results that the projection error itself is a bad indicator for the size of an oversampling set, since it takes its minimum for 0 layers, then it increases for several steps while the real error decreases.

In Figure 3 we fixed an inner coarse-grid element T. We see a corresponding corrector basis element $w^1_{h,T}$ (see Definition 3.2) with support $U(T)$. $U(T)$ is equal to T enriched by 10 layers. On the left side of the figure we can see the location of $w^1_{h,T}$ in the coarse grid (i.e., we see how much we oversample), and on the right side of the figure we see its location within the fine grid (i.e., we can see the resolution with

TABLE 4

Model Problem 1. *MsFEM computations for fixed coarse grid with* $H = 2^{-4}$ *and fixed fine grid with* $h = 2^{-8}$, *but with different sizes for the oversampling domains* $U(T)$. *If* k *denotes the number of layers, the coarse-grid element* T *has been enriched by* k *layers of fine-grid elements to create* $U(T)$. *We depict the* L^2*-error, the* H^1*-error, the total estimated error* η^{total}, *the oversampling indicator* η^{overs} *(size of corrector flux jumps), and the projection error indicator* η^{proje}.

Number of layers	$\|e\|_{L^2(\Omega)}$	$\|e\|_{H^1(\Omega)}$	η^{total}	η^{overs}	η^{proje}
0	0.0332	1.2085	3.4370	1.3754	0.0
1	0.0296	1.0878	3.0868	0.9920	$5.4 \cdot 10^{-3}$
2	0.0265	1.0652	2.8130	0.6468	$9.9 \cdot 10^{-3}$
3	0.0242	1.0599	2.6324	0.4576	$1.1 \cdot 10^{-2}$
4	0.0223	1.0227	2.5404	0.4069	$9.6 \cdot 10^{-3}$
5	0.0208	0.9720	2.4848	0.4063	$7.1 \cdot 10^{-3}$
10	0.0172	0.9524	2.4320	0.3532	$7.0 \cdot 10^{-3}$
15	0.0186	0.9305	2.4143	0.3369	$5.7 \cdot 10^{-3}$
20	0.0182	0.9161	2.3821	0.3166	$4.2 \cdot 10^{-3}$

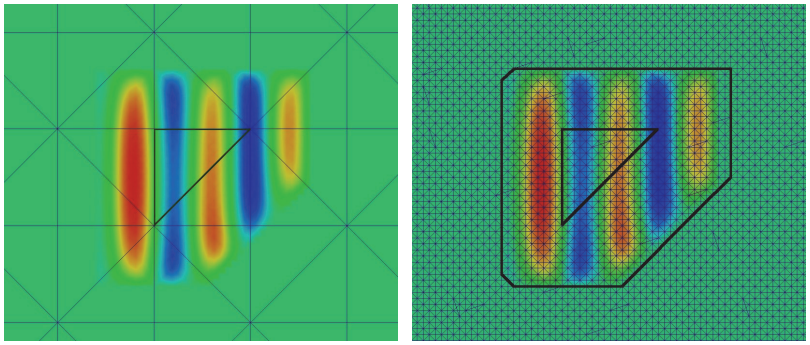


FIG. 3. *Model Problem* 1. *MsFEM computation for* $H = 2^{-4}$, $h = 2^{-8}$, *and* 10 *layers. The figure depicts the corrector basis element* $w^1_{h,T}$ *(recall Definition* 3.2*) for an arbitrary inner coarse grid element* T. *Coarse-grid triangulation (left) and fine-grid triangulation (right) with* T *and* $U(T)$ *indicated by black lines.*

which we solve the local problems). In Figure 4 we can also see the corresponding conservative flux which is almost constant along the edges of T, even though $w^1_{h,T}$ is oscillating. Figure 5 shows a good agreement between the correctors of two adjacent coarse-grid elements. Both correctors might be easily glued together at the common edge of the two coarse elements. This matching becomes even clearer if we have a look at Figure 6. Here, we see the conservative fluxes for the correctors w^1_{h,T_1} and w^1_{h,T_2} of two adjacent coarse elements T_1 and T_2. The flow over the interface from the left (almost constant with value 0.0157587) is about the negative of the flow from the right (almost constant value -0.015759). This implies that the jump in the flux is very close to zero and, therefore, the oversampling sets are chosen large enough. Hence, the usage of conservative fluxes seems to be a good choice.

In Tables 5 and 6 we show two different results of the adaptive algorithm that we suggested in section 3.2 for $c_1 = c_2 = c_3 = c_4 = 0.25$ and $\sigma = 1.1$. In both cases we start in the first cycle with uniformly refined grids with $H = 2^{-2}$ and $h = 2^{-4}$ and corrector problems without oversampling. In the first example, i.e., Table 5, six cycles are required to fall below the tolerance $\eta^{total} < 2.0$. During the algorithm the

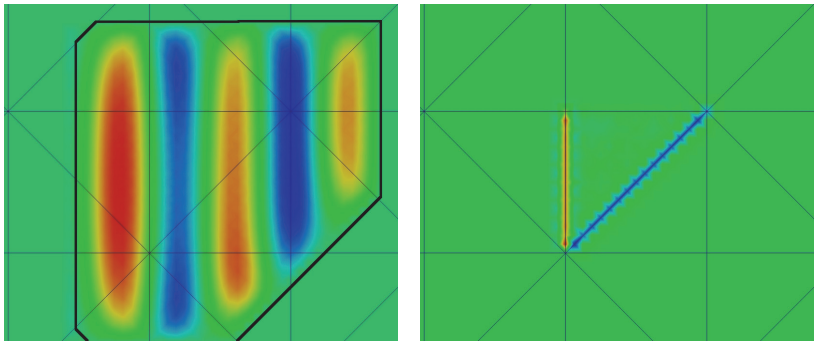


FIG. 4. *Model Problem* 1. *MsFEM computation for* $H = 2^{-4}$, $h = 2^{-8}$, 10 *layers, and for an arbitrary inner coarse-grid element* T. *Left: corrector basis element* $w^1_{h,T} := Q_h(v_1)_{|T}$ *with* $v_1(x) := x_1$ *(recall Definition* 3.2*). Colorshading: min* $-3.58 \cdot 10^{-3}$ *(blue) to max* $6.18 \cdot 10^{-3}$ *(red). Right: corresponding conservative corrector flux* $q_{h,T,1}$ *(recall Definition* 3.4*). Colorshading: min* $-1.09 \cdot 10^{-2}$ *(blue) to max* $1.58 \cdot 10^{-2}$ *(red).*

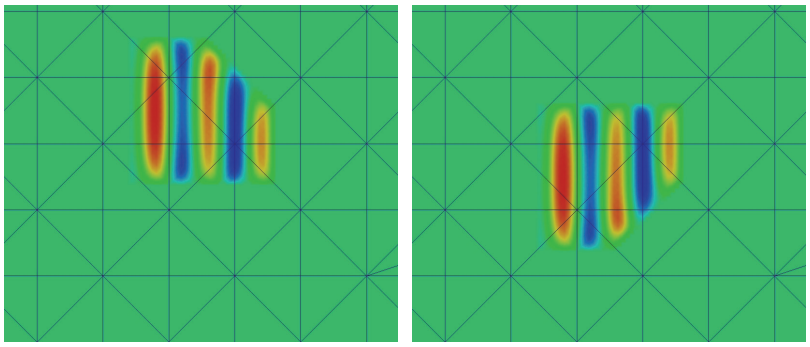


FIG. 5. *Model Problem* 1. *MsFEM computation for* $H = 2^{-4}$, $h = 2^{-8}$, *and* 10 *layers. The figure depicts two corrector basis elements* w^1_{h,T_1} *and* w^1_{h,T_2} *(see Definition* 3.2*) for two adjacent coarse grid elements* T_1 *and* T_2.

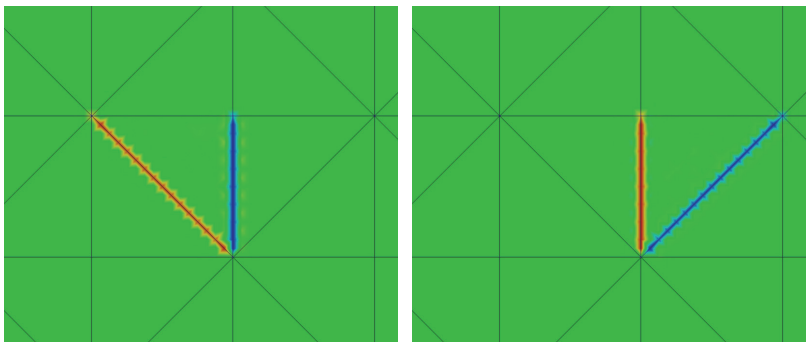


FIG. 6. *Model Problem* 1. *MsFEM computation for* $H = 2^{-4}$, $h = 2^{-8}$, *and* 10 *layers. The figure depicts two conservative corrector fluxes* $q_{h,T_1,1}$ *and* $q_{h,T_2,1}$ *for two adjacent coarse grid elements* T_1 *and* T_2. *Left figure colorshading: min* $-1.5759 \cdot 10^{-2}$ *(blue) to max* $1.09188 \cdot 10^{-2}$ *(red). Right figure colorshading: min* $-1.09188 \cdot 10^{-2}$ *(blue) to max* $1.57587 \cdot 10^{-2}$ *(red).*

TABLE 5

Model Problem 1. *MsFEM computations obtained with the adaptive algorithm proposed in section* 3.2. *We start with a uniformly refined coarse and fine grids with* $H = 2^{-2}$ *and* $h = 2^{-4}$. *We also start with* 0 *layers, i.e.,* $U(T) = T$ *for all* $T \in \mathcal{T}_H$. *During the algorithm, the number of layers is only increased uniformly for all elements at once (with* $k = 5$*). The permissible deviation* σ *is* 1.1, *and we have equal weighting of the components by* $c_1 = c_2 = c_3 = c_4 = \frac{1}{4}$. *In the first column we depict the number of the cycle of the algorithm.*

Run	$\|e\|_{L^2(\Omega)}$	$\|e\|_{H^1(\Omega)}$	η^{total}
1	0.1671	2.4884	5.6112
2	0.1072	2.1508	5.4531
3	0.0576	1.6258	4.9596
4	0.0209	1.0243	2.7894
5	0.0096	0.6095	2.0227
6	0.0054	0.4976	1.6495

TABLE 6

Model Problem 1. *MsFEM computations obtained with the adaptive algorithm proposed in section* 3.2 *with the same configuration as described in Table* 5, *but with the difference that if* η^{macro} *is identified as being dominant, each relevant coarse element is marked for* 4 *bisections instead of* 2. *Again, we use an equal weighting of the components by* $c_1 = c_2 = c_3 = c_4 = \frac{1}{4}$. *In the first column we depict the number of the cycle of the algorithm.*

Run	$\|e\|_{L^2(\Omega)}$	$\|e\|_{H^1(\Omega)}$	η^{total}
1	0.16709	2.4884	5.9112
2	0.04536	1.5138	4.9976
3	0.00954	0.6406	1.8942

number of oversampling layers goes up to 15, and in the last cycle we have a coarse grid that is locally refined up to $H = 2^{-6}$, a fine grid locally refined up to $h = 2^{-8}$. In order to decrease the number of cycles, we slightly modified the algorithm for the results in Table 6. Here, we allow a coarse element to be marked for 4 bisections instead of 2. Now, the algorithm reaches the tolerance $\eta^{total} < 2.0$ already after three cycles, where we get a very accurate approximation of the exact solution. The maximum resolution of the (adaptively refined) coarse grid corresponds with $H = 2^{-4}$, the maximum resolution of the fine grid corresponds with $h = 2^{-7}$, and the coarse elements were enriched by only 5 layers for oversampling. Nevertheless, the errors are already quite close to the errors for a uniform computation with $(H, h) = (2^{-5}, 2^{-7})$ and an enrichment by 10 layers (as depicted in Table 1). The error reduction after each cycle is significant. The algorithm produces very good results. The coarse grid and the final MsFEM approximation after the third cycle is depicted in Figure 7. We observe the strongest refinement close to the maximum values of the solution, where we also have a strong curvature of the solution. The grid remains quite coarse in most regions where the solution tends to be constant.

In order to investigate the influence of the "weighting constants" $c_1, \ldots, c_4$ on the performance of the adaptive algorithm, we repeated the computations with various combinations of these constants. In Table 7 we see the results for the original version of the algorithm (for a fixed number of 5 cycles), whereas Table 8 depicts the results for the case that coarse elements can be marked for 4 bisections (for a fixed number of 3 cycles). It turns out that the constant c_4 (which decides whether the macro error is dominant) plays the key role in the performance of the algorithm and should not be chosen too large (not more than 0.3). However, in an investigated span of

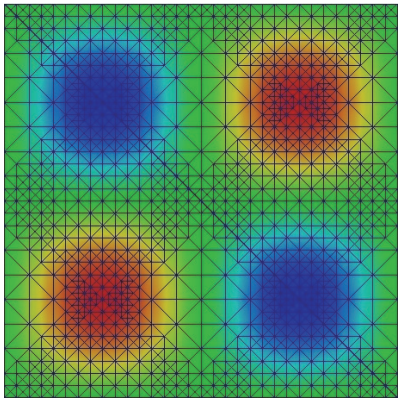


FIG. 7. *Model Problem* 1. *Adaptively refined coarse grid and corresponding MsFEM approximation obtained after the third cycle of the algorithm as indicated in Table* 6.

TABLE 7

Model Problem 1. *MsFEM computations obtained with the adaptive algorithm proposed in section* 3.2 *with the same configuration as described in Table* 5. *The influence of the weighting constants* $c_1, \dots, c_4$ *is tested. For simplification we define* $c_1 := c^{micro}$, $c_2 := c^{approx}$, $c_3 := c^{overs}$, $c_4 := c^{macro}$. *Column* 5 *depicts the number of runs of the algorithm for which the errors are stated.*

c^{micro}	c^{macro}	c^{approx}	c^{overs}	Number of runs	$\|e\|_{L^2(\Omega)}$	$\|e\|_{H^1(\Omega)}$	η^{total}
0.1	0.4	0.1	0.4	5	0.0130	0.7378	2.0672
0.2	0.3	0.3	0.3	5	0.0101	0.6396	2.0477
0.25	0.25	0.25	0.25	5	0.0096	0.6095	2.0227
0.3	0.1	0.3	0.3	5	0.0101	0.6396	2.0477
0.3	0.2	0.3	0.2	5	0.0096	0.6095	2.0227
0.3	0.3	0.3	0.1	5	0.0099	0.6124	2.0237
0.4	0.1	0.4	0.1	5	0.0099	0.6124	2.0237
0.4	0.05	0.4	0.15	5	0.0099	0.6124	2.0237

$0.05 \leq c_4 \leq 0.3$ no noteworthy changes in the performance could be observed. The choice of the other constants (in a span between 0.1 and 0.4) does not seem to have a crucial influence, as long as c_4 is chosen properly. The choice $c_1 = c_2 = c_3 = c_4 = 0.25$ gave the best performance results, even though it is only nuances over similar constellations.

MODEL PROBLEM 2. *Let us define* $\Omega :=]0,1[^2$. *Find* $u \in \mathring{H}^1(\Omega)$ *with*

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

Here, the synthetic scalar coefficient A *is depicted on the left side of Figure* 8. A *is rapidly oscillating in an outer region. In an inner region, the conductivity is very low* $(5 \cdot 10^{-4})$ *but still contains layers of constant high conductivity* $(5 \cdot 10^{-2})$.

For Model Problem 2, we do not have access to the exact solution. However, on the right side of Figure 8 we can see an approximation that was obtained with a finite element computation with a uniformly refined grid with $h = 2^{-8}$. We observe small fine-scale fluctuations outside the inner patch and a dominant oval-shaped region in the middle. Obviously, this middle region must be resolved by the coarse grid, otherwise this part cannot be captured. Due to the absence of a reliable approximation

TABLE 8

Model Problem 1. *MsFEM computations obtained with the adaptive algorithm proposed in section* 3.2 *with the same configuration as described in Table* 6 *(i.e., relevant coarse elements are marked for* 4 *bisections instead of* 2*). We test the influence of the weighting constants* c_1, c_4. *For simplification we define* $c_1 := c^{micro}$, $c_2 := c^{approx}$, $c_3 := c^{overs}$, $c_4 := c^{macro}$.

c^{micro}	c^{macro}	c^{approx}	c^{overs}	Number of runs	$\|e\|_{L^2(\Omega)}$	$\|e\|_{H^1(\Omega)}$	η^{total}
0.1	0.4	0.1	0.4	3	0.0231	0.9675	2.6569
0.2	0.3	0.3	0.3	3	0.0095	0.6406	1.8942
0.25	0.25	0.25	0.25	3	0.0095	0.6406	1.8942
0.3	0.1	0.3	0.3	3	0.0095	0.6406	1.8942
0.3	0.2	0.3	0.2	3	0.0095	0.6406	1.8942
0.3	0.3	0.3	0.1	3	0.0095	0.6406	1.8942
0.4	0.1	0.4	0.1	3	0.0095	0.6406	1.8942
0.4	0.05	0.4	0.15	3	0.0095	0.6406	1.8942

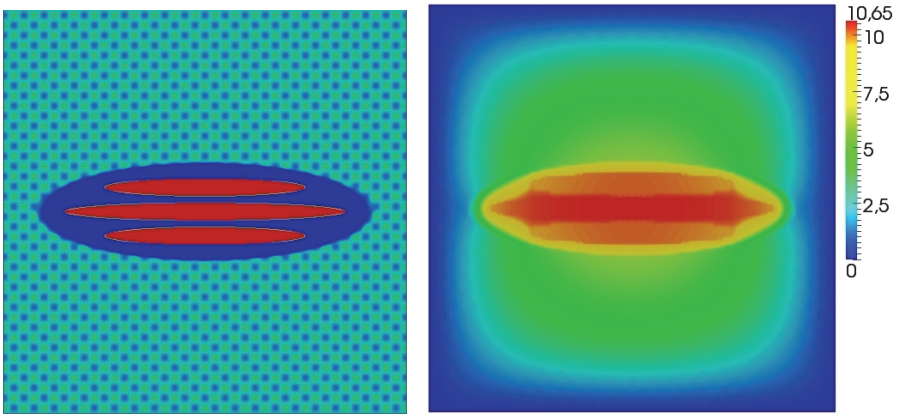


FIG. 8. *Left: plot of the diffusion coefficient A used in Model Problem* 2. *The colorshading is from red* (0.05*) to blue* (0.0005*). The micro structure outside the inner patch is periodic and given by* $(8\pi^2)^{-2}\left(1+2^{-1}\cos(2\pi\frac{x_0}{\epsilon})\sin(2\pi\frac{x_1}{\epsilon})\right)$ *with* $\epsilon = 5\cdot 10^{-2}$. *The transition is smooth. Right: approximation of the exact solution of Model Problem* 2. *Approximation obtained with a standard FEM fine-scale computation with* $h = 2^{-8}$.

of u we only compare the solutions qualitatively.

As for Model Problem 1, we apply the algorithm stated in section 3.2. We start with a uniform coarse grid and a uniform fine grid, where $(H, h) = (2^{-2}, 2^{-4})$. No oversampling is used in the first cycle, and the permissible deviation σ is 1.1. By choosing $c_1 = c_2 = c_3 = c_4 = \frac{1}{4}$ we get an equal weighting of the components. After 3 cycles we obtain an adaptively refined coarse grid (with 3044 elements) and a corresponding fine grid. The coarse grid reaches a maximum resolution of $H = 2^{-4}$, whereas the fine grid resolves up to $h = 2^{-7}$. The number of oversampling layers reaches 10 for every coarse element T. The final approximation and the associated coarse grid can be seen in Figure 9. Indeed, as expected, we observe strong refinements around the problematic middle patch. The grid is also refined around the corners and in the areas where the gradient of the coarse part ∇u_H becomes relatively large. The MsFEM solution shows visible fine-scale fluctuations outside the inner patch just like the detailed FEM approximation in Figure 8 (right side). This becomes even clearer when we have a look at Figure 10, which depicts the fine-scale part of the MsFEM

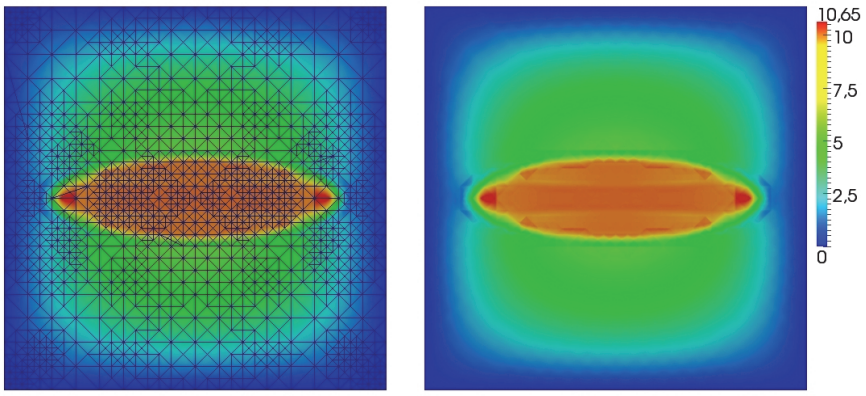


FIG. 9. *Model Problem* 2. *MsFEM computations obtained with the adaptive algorithm proposed in section* 3.2. *We start with a uniformly refined coarse and fine grid with* $H = 2^{-2}$ *and* $h = 2^{-4}$, *and we start with* 0 *layers, i.e.,* $U(T) = T$ *for all* $T \in \mathcal{T}_H$. *The permissible deviation* σ *is* 1.1, *and we have* $c_1 = c_2 = c_3 = c_4 = \frac{1}{4}$ *(equal weighting of the components). The figure depicts the adaptively refined coarse grid (*3044 *elements) and the MsFEM approximation after the third cycle of the algorithm.*

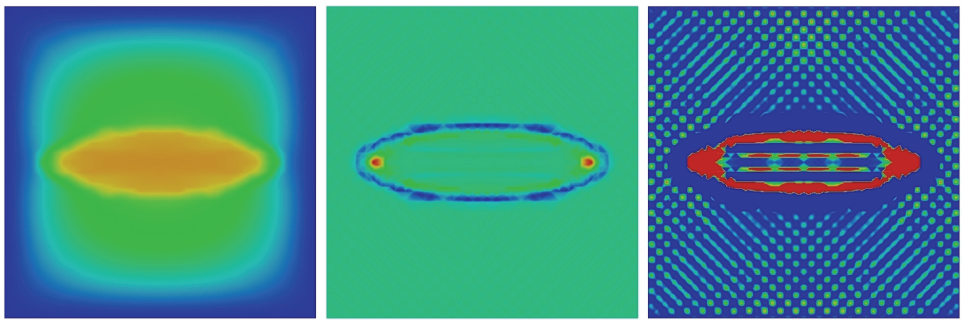


FIG. 10. *Model Problem* 2. *Adaptively computed MsFEM approximation as already depicted in Figure* 9. *The left figure shows the coarse scale part (colorshading identical to the right side of Figure* 8*). The middle figure shows the fine-scale part with a maximum/minimum colorshading from blue (*−1.72*) to red (*3.52*). The right figure also depicts the fine-scale part but (to see the micro structure) with a scaled and clamped colorshading from blue (*0.0*) to red (*0.6*).*

approximation and where the oscillating structure is clearly perceptible. The behavior of the MsFEM solution in the inner patch is also close to the behavior of the detailed FEM approximation, and we conclude that every characterizing feature seems to be captured by the MsFEM solution. Note that the FEM approximation was obtained on a grid with almost 150 000 elements in comparison to the MsFEM coarse grid with only 3044 elements. Finally, in Figure 11 we can see the number of oversampling layers for each coarse-grid element after the second cycle of the algorithm. Due to the heterogeneity, we modified the algorithm in the following way: We define $\omega_T := \eta_T^{overs}(\frac{1}{|\mathcal{T}_H|}\eta^{overs})^{-1}$ (i.e., ω_T describes the deviation of the local oversampling error from the average oversampling error). If $\omega_T \in [\frac{m}{5}, \frac{m+1}{5})$, $m \in \mathbb{N}$, then increase $U(T) \in \mathcal{U}$ by $m+1$ layers. This strategy is more flexible with regard to significantly different oversampling errors depending on the location of the coarse-grid element. We observe that the most layers are added to the elements located outside the inner

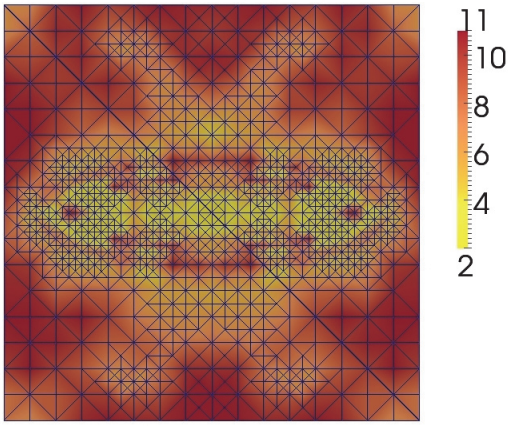


FIG. 11. *Model Problem* 2. *Results for the adaptive MsFEM computation described in Figure* 9 *after the second cycle of the algorithm. The depicted grid is the coarse grid. We see a visualization of the number of layers for each of the coarse-grid elements. The colorshading is from yellow (*2 *layers) to red (*11 *layers).*

patch. The coarse-grid elements located in the inner patch do not receive a lot of layers. This behavior completely resembles what we expect. In the inner patch, the behavior is primarily macroscopic and the microscopic behavior is almost constant. This yields solutions of the local problems which must be close to zero. But the zero solutions are already accurately approximated without a large oversampling set. Therefore, we only need a few layers in this region. One the other hand, outside the inner patch we are dealing with fast fine-scale oscillations. The solutions of the local problems are highly variable and must not be forced to a zero boundary condition. To reduce the effect of this wrong boundary condition, several oversampling layers are required. The algorithm and oversampling error estimator perfectly adapt to this situation. Again, the advantage of conservative fluxes used as oversampling error indicators can be confirmed.

7. Conclusion. In this contribution, we derived the first rigorous a posteriori error estimate for multiscale finite element approximations in a general scenario with no further assumptions on the micro structure. Based on this estimate we were able to derive an algorithm for oversampling control and adaptive mesh refinement. We showed how to transfer our results to the nonlinear case and demonstrated the applicability of the adaptive method in numerical experiments.

Acknowledgment. The authors gratefully acknowledge the helpful and accurate comments of the anonymous referees, which helped to improve this paper considerably.

REFERENCES

[1] J. E. AARNES AND Y. EFENDIEV, *Mixed multiscale finite element methods for stochastic porous media flows*, SIAM J. Sci. Comput., 30 (2008), pp. 2319–2339.

[2] A. ABDULLE AND A. NONNENMACHER, *A posteriori error analysis of the heterogeneous multiscale method for homogenization problems*, C. R. Math. Acad. Sci. Paris, 347 (2009), pp. 1081–1086.

[3] A. Abdulle and A. Nonnenmacher, *Adaptive finite element heterogeneous multiscale method for homogenization problems*, Comput. Methods Appl. Mech. Engrg., 200 (2011), pp. 2710–2726.
[4] T. Arbogast, G. Pencheva, M. F. Wheeler, and I. Yotov, *A multiscale mortar mixed finite element method*, Multiscale Model. Simul., 6 (2007), pp. 319–346.
[5] P. Bastian, M. Blatt, A. Dedner, C. Engwer, R. Klöfkorn, R. Kornhuber, M. Ohlberger, and O. Sander, *A generic grid interface for parallel and adaptive scientific computing.* II. *Implementation and tests in DUNE*, Computing, 82 (2008), pp. 121–138.
[6] F. Brezzi, T. J. R. Hughes, and E. Süli, *Variational approximation of flux in conforming finite element methods for elliptic partial differential equations: A model problem*, Atti Accad. Naz. Lincei Cl. Sci. Fis. Mat. Natur. Rend. Lincei (9) Mat. Appl., 12 (2001), pp. 159–166 (2002).
[7] Z. Chen and T. Y. Savchuk, *Analysis of the multiscale finite element method for nonlinear and random homogenization problems*, SIAM J. Numer. Anal., 46 (2008), pp. 260–279.
[8] P. Clément, *Approximation by finite element functions using local regularization*, Rev. Française Automat. Informat. Recherche Opérationnelle Sér. Rouge Anal. Numér., 9 (1975), pp. 77–84.
[9] A. Dedner, R. Klöfkorn, M. Nolte, and M. Ohlberger, *A generic interface for parallel and adaptive discretization schemes: Abstraction principles and the dune-fem module*, Computing, 90 (2010), pp. 165–196.
[10] P. Dostert, Y. Efendiev, and T. Y. Hou, *Multiscale finite element methods for stochastic porous media flow equations and applications to uncertainty quantification*, Comput. Methods Appl. Mech. Engrg., 197 (2008), pp. 3445–3455.
[11] W. E and B. Engquist, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1 (2003), pp. 87–132.
[12] W. E and B. Engquist, *Multiscale modeling and computation*, Notices Amer. Math. Soc., 50 (2003), pp. 1062–1070.
[13] W. E and B. Engquist, *The heterogeneous multi-scale method for homogenization problems*, in Multiscale Methods in Science and Engineering, Lect. Notes Comput. Sci. Eng. 44, Springer, Berlin, 2005, pp. 89–110.
[14] Y. Efendiev and T. Y. Hou, *Multiscale finite element methods for porous media flows and their applications*, Appl. Numer. Math., 57 (2007), pp. 577–596.
[15] Y. Efendiev and T. Y. Hou, *Multiscale Finite Element Methods. Theory and Applications*, Surv. Tutor. Appl. Math. Sci. 4, Springer, New York, 2009.
[16] Y. Efendiev, T. Y. Hou, and V. Ginting, *Multiscale finite element methods for nonlinear problems and their applications*, Commun. Math. Sci., 2 (2004), pp. 553–589.
[17] Y. Efendiev and A. Pankov, *Numerical homogenization of monotone elliptic operators*, Multiscale Model. Simul., 2 (2003), pp. 62–79.
[18] Y. Efendiev and A. Pankov, *Numerical homogenization and correctors for nonlinear elliptic equations*, SIAM J. Appl. Math., 65 (2004), pp. 43–68.
[19] Y. Efendiev and A. Pankov, *Homogenization of nonlinear random parabolic operators*, Adv. Differential Equations, 10 (2005), pp. 1235–1260.
[20] Y. R. Efendiev, T. Y. Hou, and X.-H. Wu, *Convergence of a nonconforming multiscale finite element method*, SIAM J. Numer. Anal., 37 (2000), pp. 888–910.
[21] A. Gloria, *An analytical framework for numerical homogenization.* Part II: *Windowing and oversampling*, Multiscale Model. Simul., 7 (2008), pp. 274–293.
[22] A. Gloria, *Reduction of the resonance error—Part* 1: *Approximation of homogenized coefficients*, Math. Models Methods Appl. Sci., 21 (2011), pp. 1601–1630.
[23] A. Gloria, *Numerical homogenization: Survey, new results, and perspectives*, ESAIM Proc., 37 (2012), pp. 50–116.
[24] C. Gräser and O. Sander, *The dune-subgrid module and some applications*, Computing, 86 (2009), pp. 269–290.
[25] P. Henning, *Convergence of MsFEM approximations for elliptic, non-periodic homogenization problems*, Netw. Heterog. Media, 7 (2012), pp. 503–524.
[26] P. Henning and M. Ohlberger, *The heterogeneous multiscale finite element method for elliptic homogenization problems in perforated domains*, Numer. Math., 113 (2009), pp. 601–629.
[27] P. Henning and M. Ohlberger, *The heterogeneous multiscale finite element method for advection-diffusion problems with rapidly oscillating coefficients and large expected drift*, Netw. Heterog. Media, 5 (2010), pp. 711–744.
[28] P. Henning and M. Ohlberger, *A Newton-scheme framework for multiscale methods for nonlinear elliptic homogenization problems*, in Proceedings of the ALGORITMY 2012,

19th Conference on Scientific Computing, Vysoke Tatry, Podbanske, 2012, pp. 65–74.
[29] P. Henning, M. Ohlberger, and B. Schweizer, *Homogenization of the degenerate two-phase flow equations*, Math. Models Methods Appl. Sci., 23 (2013), pp. 2323–2352.
[30] P. Henning and D. Peterseim, *Oversampling for the multiscale finite element method*, Multiscale Model. Simul., 11 (2013), pp. 1149–1175.
[31] T. Y. Hou and X.-H. Wu, *A multiscale finite element method for elliptic problems in composite materials and porous media*, J. Comput. Phys., 134 (1997), pp. 169–189.
[32] T. Y. Hou, X.-H. Wu, and Z. Cai, *Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients*, Math. Comp., 68 (1999), pp. 913–943.
[33] T. Y. Hou, X.-H. Wu, and Y. Zhang, *Removing the cell resonance error in the multiscale finite element method via a Petrov-Galerkin formulation*, Commun. Math. Sci., 2 (2004), pp. 185–205.
[34] T. J. R. Hughes, *Multiscale phenomena: Green's functions, the Dirichlet-to-Neumann formulation, subgrid scale models, bubbles and the origins of stabilized methods*, Comput. Methods Appl. Mech. Engrg., 127 (1995), pp. 387–401.
[35] T. J. R. Hughes, G. Engel, L. Mazzei, and M. G. Larson, *The continuous Galerkin method is locally conservative*, J. Comput. Phys., 163 (2000), pp. 467–488.
[36] T. J. R. Hughes, G. R. Feijóo, L. Mazzei, and J.-B. Quincy, *The variational multiscale method: A paradigm for computational mechanics*, Comput. Methods Appl. Mech. Engrg., 166 (1998), pp. 3–24.
[37] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik, *Homogenization of Differential Operators and Integral Functionals*, Springer-Verlag, Berlin, 1994; translated from the Russian by G. A. Yosifian,
[38] M. G. Larson and A. Målqvist, *Adaptive variational multiscale methods based on a posteriori error estimation: Duality techniques for elliptic problems*, in Multiscale Methods in Science and Engineering, Lect. Notes Comput. Sci. Eng. 44, Springer, Berlin, 2005, pp. 181–193.
[39] M. G. Larson and A. Målqvist, *Adaptive variational multiscale methods based on a posteriori error estimation: Energy norm estimates for elliptic problems*, Comput. Methods Appl. Mech. Engrg., 196 (2007), pp. 2313–2324.
[40] M. G. Larson and A. Målqvist, *An adaptive variational multiscale method for convection-diffusion problems*, Comm. Numer. Methods Engrg., 25 (2009), pp. 65–79.
[41] M. G. Larson and A. Målqvist, *A mixed adaptive variational multiscale method with applications in oil reservoir simulation*, Math. Models Methods Appl. Sci., 19 (2009), pp. 1017–1042.
[42] A.-M. Matache and C. Schwab, *Two-scale FEM for homogenization problems*, M2AN Math. Model. Numer. Anal., 36 (2002), pp. 537–572.
[43] M. Ohlberger, *A posteriori error estimates for the heterogeneous multiscale finite element method for elliptic homogenization problems*, Multiscale Model. Simul., 4 (2005), pp. 88–114.
[44] M. Peszyńska, *Mortar adaptivity in mixed methods for flow in porous media*, Int. J. Numer. Anal. Model., 2 (2005), pp. 241–282.
[45] M. Peszyńska, M. F. Wheeler, and I. Yotov, *Mortar upscaling for multiphase flow in porous media*, Comput. Geosci., 6 (2002), pp. 73–100.
[46] M. Růžička, *Nichtlineare Funktionalanalysis*, Springer-Verlag, Berlin, Heidelberg, New York, 2004.
[47] C. Schwab and A.-M. Matache, *Generalized FEM for homogenization problems*, in Multiscale and Multiresolution Methods, Lect. Notes Comput. Sci. Eng. 20, Springer, Berlin, 2002, pp. 197–237.
[48] B. Schweizer, *Homogenization of degenerate two-phase flow equations with oil trapping*, SIAM J. Math. Anal., 39 (2008), pp. 1740–1763.
[49] B. Schweizer and M. Veneroni, *The needle problem approach to non-periodic homogenization*, Netw. Heterog. Media, 6 (2011), pp. 755–781.
[50] R. Verfürth, *A Review of A Posteriori Error Estimation and Adaptive Mesh-Refinement Techniques*, Wiley-Teubner, Stuttgart, 1996.