# OVERSAMPLING FOR THE MULTISCALE FINITE ELEMENT METHOD*

PATRICK HENNING† AND DANIEL PETERSEIM‡

**Abstract.** This paper reviews standard oversampling strategies as performed in the multiscale finite element method (MsFEM). Common to those approaches is that the oversampling is performed in the full space restricted to a patch including coarse finite element functions. We suggest, by contrast, performing local computations with the additional constraint that trial and test functions be linear independent from coarse finite element functions. This approach reinterprets the variational multiscale method in the context of computational homogenization. This connection gives rise to a general fully discrete error analysis for the proposed multiscale method with constrained oversampling without any resonance effects. In particular, we are able to give the first rigorous proof of convergence for an MsFEM with oversampling.

**Key words.** a priori error estimate, finite element method, multiscale method, MsFEM, oversampling

**AMS subject classifications.** 35J15, 65N12, 65N30

**DOI.** 10.1137/120900332

**1. Introduction.** The numerical treatment of partial differential equations with rapidly varying and strongly heterogeneous coefficient functions is still a challenging area of present research, especially with regard to applications such as porous media flow or the transport of solutes in groundwater. In such problems, the occurring permeabilities and hydraulic conductivities have rapidly changing features due to different types of soil, microscopic inclusions in the bottom, or porous subsurface rock formations. Any meaningful numerical simulation of relevant physical effects has to account for these highly heterogeneous fine scale structures in the whole computational domain. This means that the underlying computational mesh has to be sufficiently fine to resolve microscopic details. If pore scale effects become relevant or if domains spread over kilometers, then the computational load becomes extremely large and in several applications even too large to treat the problem with standard finite element or finite volume methods. This is just one instance of a so-called multiscale problem as it arises in hydrology, physics, or industrial engineering.

In recent years, many numerical methods have been designed to deal with these computational issues that come along with multiscale problems. Most of them aim to decouple the global fine scale problem into localized subproblems which can be treated independently from each other plus some global coarse problem. The list of proposed multiscale methods, meanwhile, is long. Amongst the most popular methods are the finite element heterogeneous multiscale method (HMM), initially introduced by E and Engquist [9] (see also [10, 11, 1]), the variational multiscale method (VMM) by Hughes [29] and Hughes et al. [30] (see also [32, 33, 34]), the approaches by Owhadi and Zhang [35, 36], or that of Babuska and Lipton [2].

In this paper, we deal with another popular method: the multiscale finite element method (MsFEM) proposed by Hou and Wu [26] and further investigated in several contributions [27, 16, 15, 28]. There is an ongoing development of the method to apply it to various fields and equations. For instance, an MsFEM for nonlinear elliptic problems is proposed in [13], a formulation for two phase flow problems in porous media is presented in [12], advection diffusion problems are treated in [8], and an application to elliptic interface problems with high contrast coefficients is presented in [7]. A survey on the method is given in the book by Efendiev and Hou [14]. There is a vast literature devoted to the method, but there are still open questions of strong interest. The most relevant issue is a rigorous error analysis of the method, in particular in the case of nonperiodic microstructures.

The MsFEM is related to some common finite element space with an underlying coarse grid. The essential idea is to modify the corresponding basis functions in such a way that fine scale variations on finer scales are sufficiently well captured. More specifically, local fine scale computations are performed to determine so-called *corrector functions*. These corrector functions can be added as local perturbations to the original set of basis functions of the coarse finite element space.

However, it is well known that the classical MsFEM suffers from so-called resonance errors, which are typically of order $O(\frac{\varepsilon}{H})$, where $\varepsilon$ denotes a characteristic size of the small scale and where $H$ denotes the mesh size of the coarse grid (cf. [27, 28]). This implies that the numerical error becomes large in regions where the coarse grid size is close to the characteristic length scale of the microscopic oscillations. There are two different explanations for this error. The first one is a mismatch between the boundary conditions imposed for the local fine scale problems and global behavior of the oscillatory exact solution (cf. [16]). The second explanation is due to the size and geometry of the sampling patch (cf. [28]). The averaged behavior in such a patch should be "representative" so that we can speak about a perfect sample size. If this is not the case, the final approximation might be distorted. In the periodic setting, for instance, the sampling domain should be some multiple of the periodic cell. On triangular patches with cathetuses of the length of a period, this patch is only half a periodic cell (i.e., the patch has bad size and geometry) and lacks essential information. This yields a completely wrong approximation (cf. [21]). In the periodic setting considerable improvements were obtained by Gloria [19, 20], who proposed a regularization of the local (patch) problems by adding a zero-order term. With this strategy, both sources of the oversampling error could be significantly reduced (cf. [19, Theorem 3.1] and [20, sections 5.3 and 5.4]).

In a lot of applications, such as oil reservoir simulations or the transport of solutes in groundwater, a characteristic microscopic length scale $\varepsilon$ is unknown, cannot be identified, or does not exist at all. In scenarios without a clear scale separation it is often impossible to predict whether or not we are in the regime of resonance errors. It is very likely to actually hit the problematic regime. Hence, the quality of the final approximation cannot be determined unless resonance errors are eliminated. For this purpose, different *oversampling strategies* have been proposed. The fundamental idea of each of these techniques is to extend the local problems to larger patches and perform the computation on these *oversampling domains* but feed the coarse scale equation only with the information obtained within the original smaller patches. This reduces the effect of wrong boundary conditions and bad sampling sizes. In this paper, we present the two major strategies for oversampling and discuss their advantages and disadvantages. On the basis of these considerations we propose a new strategy that overcomes the issues of the existing strategies. The new approach is

closely related to the VMM-type method presented in [34]. We prove quantitative error estimates for the corresponding multiscale approximations under very general assumptions on the diffusion coefficient.

This contribution is structured as follows: In section 2 we recall the classical formulation of the MsFEM without oversampling. The most popular approaches for oversampling are discussed in section 3. In section 4 we propose a new strategy for which we present a quantitative error analysis. Numerical experiments are presented in section 5. The paper closes with a short conclusion.

**2. The multiscale finite element method.** In this section, we state the setting of this paper and we establish the required notation. We recall the classical multiscale finite element method (MsFEM) as initially proposed by Hou and Wu [26].

**2.1. Setting and notation.** Consider a bounded Lipschitz domain $\Omega \subset \mathbb{R}^d$ with a piecewise flat boundary and some matrix-valued coefficient $A \in L^\infty(\Omega, \mathbb{R}^{d \times d}_{sym})$ with uniform spectral bounds $\gamma_{\min} > 0$ and $\gamma_{\max} \geq \gamma_{\min}$,

$$(2.1) \qquad \sigma(A(x)) \subset [\gamma_{\min}, \gamma_{\max}] \quad \text{for almost all } x \in \Omega.$$

Given $f \in L^2(\Omega)$, we seek the weak solution of

$$-\nabla \cdot A\nabla u = f \quad \text{in } \Omega,$$
$$u = 0 \quad \text{on } \partial\Omega;$$

i.e., we seek $u \in H_0^1(\Omega) := \{v \in H^1(\Omega) \mid v|_{\partial\Omega} = 0 \text{ in the sense of traces}\}$ that satisfies

$$(2.2) \qquad a(u,v) := \int_\Omega A\nabla u \cdot \nabla v = \int_\Omega fv =: F(v) \quad \text{for all } v \in H_0^1(\Omega).$$

We consider two discretization scales $H \geq h > 0$. The coarse scale $H$ is arbitrary, whereas the small scale parameter $h$ may be constrained by the problem. Typically, it is assumed to be smaller than the characteristic length scales of the variations of the diffusion coefficient $A$.

Let $\mathcal{T}_H$, $\mathcal{T}_h$ denote corresponding subdivisions of $\Omega$ into (closed) triangles (for $d = 2$) and tetrahedra (for $d = 3$), i.e., $\bar{\Omega} = \bigcup_{t \in \mathcal{T}_h} t = \bigcup_{T \in \mathcal{T}_H} T$. We assume that $\mathcal{T}_H$, $\mathcal{T}_h$ are regular in the sense that any two elements are either disjoint or share exactly one face or share exactly one edge or share exactly one vertex. For simplicity we assume that $\mathcal{T}_h$ is derived from $\mathcal{T}_H$ by some regular, possibly nonuniform, mesh refinement.

For $\mathcal{T} = \mathcal{T}_H, \mathcal{T}_h$, let

$$P_1(\mathcal{T}) = \{v \in C^0(\Omega) \mid \forall T \in \mathcal{T}, v|_T \text{ is a polynomial of total degree} \leq 1\}$$

denote the set of continuous and piecewise affine functions.

Accordingly, $V_h := P_1(\mathcal{T}_h) \cap H_0^1(\Omega)$ denotes the "high resolution" finite element space and the "coarse space" is given by $V_H := P_1(\mathcal{T}_H) \cap H_0^1(\Omega) \subset V_h$. For any given subset $\omega \subset \Omega$ we define the restriction of $V_h$ to $\omega$ with a zero boundary condition by $\mathring{V}_h(\omega) := V_h \cap H_0^1(\omega)$. The nonconforming fine space $V_{h,\mathcal{T}_H}$ is defined by

$$V_{h,\mathcal{T}_H} := \{v_h \mid \forall T \in \mathcal{T}_H, (v_h)|_T \in V_h \cap H^1(T)\}.$$

A general function in $v \in V_{h,\mathcal{T}_H}$ may jump across edges of the coarse mesh $\mathcal{T}_H$ and, hence, does not belong to $H_0^1(\Omega)$. However, the $\mathcal{T}_H$-piecewise gradient $\nabla_H v$, with

$(\nabla_H v)|_T = \nabla(v|_T)$ for all $T \in \mathcal{T}_H$, exists. Typically, MsFEM approximations obtained with oversampling are nonconforming approximations of the exact solution in the sense that they do not belong to $H_0^1(\Omega)$.

In the following $x_T \in T$ denotes an arbitrary point, for instance the barycenter of $T$. For $\Phi_H \in V_H$ and $T \in \mathcal{T}_H$, the affine extension operator $E_T : V_H \to P_1(\Omega)$ is given by

$$E_T(\Phi_H)(x) := (x - x_T) \cdot \nabla\Phi_H(x_T) + \Phi_H(x_T).$$

Finally, by $\chi_T$ we denote the characteristic (or indicator) function with $\chi_T(x) = 1$ for $x \in T$ and $\chi_T(x) = 0$ elsewhere.

For the sake of simplicity, all fine scale computations are performed in subspaces of the fine scale finite element space $V_h$. The Galerkin solution $u_h \in V_h$ which satisfies

$$(2.3) \qquad a_h(u_h, v) = F(v) \quad \text{for all } v \in V_h$$

may, hence, be considered as a reference approximation. Note that we never solve this large scale equation. The function $u_h$ serves as a reference solution to compare our multiscale approximations with. The underlying assumption is that the mesh $\mathcal{T}_h$ is chosen sufficiently fine so that $u_h$ is sufficiently accurate.

Throughout this paper, standard notation on Lebesgue and Sobolev spaces is employed and $a \lesssim b$ abbreviates an inequality $a \leq C b$ with some generic constant $0 \leq C < \infty$ that may depend on the shape regularity of finite element meshes and the contrast $\gamma_{\max}/\gamma_{\min}$ but *not* on the mesh sizes $H$, $h$ and the regularity or the variations of the diffusion matrix $A$; $a \approx b$ abbreviates $a \lesssim b \lesssim a$.

**2.2. The classical MsFEM and reformulation.** We first present the classical MsFEM without oversampling as originally stated by Hou and Wu [26], and similarly by Brezzi et al. [3]. They proposed the strategy to enrich the set of standard finite element basis functions by fine scale information. The information is determined by solving local problems on the fine scale. We briefly recall the method and reformulate it in terms of a correction operator $Q_h$ and a corresponding corrector basis.

Let $N$ denote the dimension of the coarse space $V_H$, and let $\{\Phi_i \mid 1 \leq i \leq N\}$ denote the usual nodal basis of $V_H$. Given some basis function $\Phi_i$, the corresponding MsFEM basis function $\Phi_i^{\mathrm{MsFEM}} \in V_h$ is uniquely determined by the condition that for all $T \in \mathcal{T}_H$ and for all $\phi_h \in \mathring{V}_h(T)$ it holds that

$$(2.4) \qquad \int_T A(x)\nabla\Phi_i^{\mathrm{MsFEM}}(x) \cdot \nabla\phi_h(x)\, dx = 0 \quad \text{and} \quad \Phi_i^{\mathrm{MsFEM}} = \Phi_i \text{ on } \partial T.$$

The span of these MsFEM functions is called the MsFEM solution space

$$V_H^{\mathrm{MsFEM}} := \mathrm{span}\{\Phi_i^{\mathrm{MsFEM}} \mid 1 \leq i \leq N\}.$$

This space is conforming in the sense of $V_H^{\mathrm{MsFEM}} \subset V_h \subset H_0^1(\Omega)$, because the set $\{\Phi_i^{\mathrm{MsFEM}} \mid 1 \leq i \leq N\}$ defines a conforming set of basis functions. The classical MsFEM in the Petrov–Galerkin (PG) formulation due to [28] reads as follows.

DEFINITION 2.1 (MsFEM without oversampling). *The MsFEM approximation $u_H^{\mathrm{MsFEM}} \in V_H^{\mathrm{MsFEM}}$ is defined as the solution of*

$$(2.5) \qquad \int_\Omega A(x)\nabla u_H^{\mathrm{MsFEM}}(x) \cdot \nabla\Phi_H(x)\, dx = \int_\Omega f(x)\Phi_H(x)\, dx \quad \text{for all } \Phi_H \in V_H.$$

In [26], the MsFEM was originally proposed in the Galerkin formulation; i.e., the test functions $\Phi_H \in V_H$ in (2.5) are replaced by test functions $\Phi_H^{\mathrm{MsFEM}} \in V_H^{\mathrm{MsFEM}}$. Observe that due to the orthogonality property (2.4) both formulations are almost identical (in the absence of oversampling). For structural reasons we used the PG version to introduce the MsFEM.

With regard to the general framework for oversampling that we present in the subsequent sections, we note that the MsFEM can be rewritten in the following way.

*Remark* 2.2. If $u_H^{\mathrm{MsFEM}} \in V_H^{\mathrm{MsFEM}}$ denotes the MsFEM approximation stated in Definition 2.1, then we have $u_H^{\mathrm{MsFEM}} = u_H + Q_h(u_H)$, where $u_H \in V_H$ solves

$$(2.6.a) \qquad \int_\Omega A\left(\nabla u_H + \nabla Q_h(u_H)\right) \cdot \nabla \Phi_H = \int_\Omega f\Phi_H \quad \text{for all } \Phi_H \in V_H,$$

with

$$(2.6.b) \qquad Q_h(\Phi_H)(x) := \sum_{T \in \mathcal{T}_H} \sum_{i=1}^{d} \partial_{x_i}\Phi_H(x_T) w_{T,i}(x),$$

and $w_{T,i} \in \mathring{V}_h(T)$ is the unique solution of

$$(2.6.c) \qquad \int_T A\nabla w_{T,i} \cdot \nabla \phi_h = -\int_T Ae_i \cdot \nabla \phi_h \quad \text{for all } \phi_h \in \mathring{V}_h(T).$$

The set of all functions $w_{T,i}$ is what we are going to call a *local corrector basis*. From the computational point of view, it seems at first glance to be cheaper to compute the corrector basis given by (2.6.c) instead of directly computing the set of multiscale basis functions given by (2.4). The latter formally involves more problems to solve. For instance, if $d = 2$, the assembling of the corrector basis $\{w_{T,i} \mid T \in \mathcal{T}_H, i = 1, 2\}$ requires the solution of $2 \cdot |\mathcal{T}_H|$ local problems, whereas the solutions of $3 \cdot |\mathcal{T}_H|$ local problems are required to assemble $\{\Phi_i^{\mathrm{MsFEM}} \mid 1 \le i \le N\}$ by using the gradients of coarse basis functions (for which we have 3 per coarse element). Still, it is possible to use the partition of unity property of the basis functions to equally decrease the costs of the original version of the MsFEM from $d \cdot |\mathcal{T}_H|$ to $(d-1) \cdot |\mathcal{T}_H|$. In particular, restricted to $T$, the gradients of $(d-1)$ basis functions associated with $(d-1)$ corners of the element $T$ span the gradient of the missing $d$th basis function on $T$.

The equivalence between the formulations (2.5) and (2.6) can be easily verified by the relation $\Phi_i^{\mathrm{MsFEM}} = \Phi_i + Q_h(\Phi_i)$. Observe that for every $i$, for every $T \in \mathcal{T}_H$, and for every $\phi_h \in \mathring{V}_h(T)$,

$$\int_T A(x) \left(\nabla\Phi_i(x) + \nabla Q_h(\Phi_i)(x)\right) \cdot \nabla\phi(x)\, dx$$

$$= \sum_{i=1}^{d} \partial_{x_i}\Phi_H(x_T) \int_T A(x) \left(e_i + \nabla w_T^i(x)\right) \cdot \nabla\phi(x)\, dx = 0$$

and $\Phi_i + Q_h(\Phi_i) = \Phi_i$ on $\partial T$, which is the definition of $\Phi_i^{\mathrm{MsFEM}}$.

A symmetric formulation of (2.6.a) is given by the following: find $u_H \in V_H$ with

$$(2.7) \quad \int_\Omega A\left(\nabla u_H + \nabla Q_h(u_H)\right) \cdot \left(\nabla\Phi_H + \nabla Q_h(\Phi_H)\right) = \int_\Omega f\Phi_H \quad \text{for all } \Phi_H \in V_H.$$

Note that (2.6.a) and (2.7) are identical, because

$$\int_T A\left(\nabla u_H + \nabla Q_h(u_H)\right) \cdot \nabla\phi_h = 0 \quad \text{for all } \phi_h \in \mathring{V}_h(T).$$

**3. Oversampling strategies.** As already discussed in the introduction, the classical MsFEM in Definition 2.1 can be strongly affected or even dominated by resonance errors (cf. [14]). In the absence of scale separation or any knowledge about a suitable sample size for the local problems, the classical MsFEM needs a modification. *Oversampling* is considered to be a remedy to this issue. Oversampling means that the local problems (2.6.a) are solved on larger domains, but only the interior information (i.e., we restrict the gained fine scale information to $T$) is communicated to the coarse scale equation (2.6.a).

There is no unique way of extending the local problems (2.6.a) to larger patches. Different extensions lead to different oversampling strategies. In this section, we present the two common approaches for oversampling. We rephrase both approaches so that they fit into a common framework. We discuss the advantages and disadvantages of the methods, and then we propose our new oversampling strategy. Note that each of the subsequent strategies is a generalization of the case without oversampling.

We shall introduce some additional notation.

DEFINITION 3.1 (admissible patch). *For $T \in \mathcal{T}_H$, we call $U(T)$ an admissible patch of $T$ if it is nonempty, open, and connected, if $T \subset U(T) \subset \Omega$, and if it is the union of elements of $\mathcal{T}_h$, i.e.,*

$$U(T) = \text{int} \bigcup_{\tau \in \mathcal{T}_h^*} \tau, \quad \text{where} \;\; \mathcal{T}_h^* \subset \mathcal{T}_h.$$

A given set of admissible patches is given by $\mathcal{U}$, i.e.,

$$\mathcal{U} := \{U(T) \mid T \in \mathcal{T}_H \;\; \text{and} \;\; U(T) \;\; \text{is an admissible patch}\},$$

where $\mathcal{U}$ contains one and only one patch $U(T)$ for each $T \in \mathcal{T}_H$. The set $U(T) \setminus T$ is called an *oversampling layer*. The thickness of the oversampling layer is denoted by $d_{\mathcal{U},T} := \text{dist}(T, \partial U(T))$. Furthermore, we define

$$d_{\mathcal{U}}^{\min} := \min_{T \in \mathcal{T}_H} d_{\mathcal{U},T} \quad \text{and} \quad d_{\mathcal{U}}^{\max} := \max_{T \in \mathcal{T}_H} d_{\mathcal{U},T}$$

as the minimum and maximum thickness.

In the spirit of (2.6.a) and (2.7), we now define the coarse scale equation for an arbitrary MsFEM with a chosen oversampling strategy. As we will see later on, all MsFEM realizations differ only in the correction operator $Q_h$ that determines the oversampling strategy.

DEFINITION 3.2 (framework for oversampling strategies). *Let $\alpha = 1, 2, 3$ denote the index of the oversampling strategy to be specified later on, and let*

$$\{w_{h,T,i}^{\mathcal{U},\alpha} \mid 1 \le i \le d, \, T \in \mathcal{T}_H\}$$

*denote a given local corrector basis that depends on the chosen strategy $\alpha$ (see (2.6.a)–(2.6.c) for the trivial case of such a basis). Then, a (not necessarily conforming) correction operator $Q_h^{\mathcal{U},\alpha} : V_H \to V_{h,\mathcal{T}_H}$ is defined by*

$$(3.1) \qquad Q_h^{\mathcal{U},\alpha}(\Phi_H)(x) := \sum_{T \in \mathcal{T}_H} \chi_T(x) \sum_{i=1}^{d} \partial_{x_i} \Phi_H(x_T) \, w_{h,T,i}^{\mathcal{U},\alpha}(x) \quad \text{for} \;\; \Phi_H \in V_H.$$

*The MsFEM approximation $u_H^\alpha + Q_h^{\mathcal{U},\alpha}(u_H^\alpha)$ obtained with strategy $\alpha$ in the PG formulation reads as follows: find $u_H^\alpha \in V_H$ such that*

$$(3.2) \qquad \sum_{T \in \mathcal{T}_H} \int_T A \left( \nabla u_H^\alpha + \nabla Q_h^{\mathcal{U},\alpha}(u_H^\alpha) \right) \cdot \nabla \Phi_H = \int_\Omega f \Phi_H \quad \text{for all} \;\; \Phi_H \in V_H.$$

*The MsFEM approximation $u_H^{\alpha,\mathrm{sym}} + Q_h^{\mathcal{U},\alpha}(u_H^{\alpha,\mathrm{sym}})$ obtained with strategy $\alpha$ and a (not necessarily equivalent) symmetric formulation is given by the following: find $u_H^{\alpha,\mathrm{sym}} \in V_H$ with*

$$(3.3) \quad \sum_{T \in \mathcal{T}_H} \int_T A\left(\nabla u_H^{\alpha,\mathrm{sym}} + \nabla Q_h^{\mathcal{U},\alpha}(u_H^{\alpha,\mathrm{sym}})\right) \cdot \left(\nabla \Phi_H + \nabla Q_h^{\mathcal{U},\alpha}(\Phi_H)\right)$$

$$= \int_\Omega f(\Phi_H + Q_h^{\mathcal{U},\alpha}(\Phi_H))$$

*for all $\Phi_H \in V_H$. Observe that strategies can differ only in the choice of the corrector basis. The remaining structure of the methods is always the same.*

In the subsequent sections, we demonstrate how existing oversampling strategies fit into the framework presented in Definition 3.2.

**3.1. Classical strategy initially introduced by Hou and Wu.** The classical oversampling strategy was proposed by Hou and Wu [26] and further used and investigated in several works (cf. [13, 6, 14]).

Let $T \in \mathcal{T}_H$ be fixed, and let $\{\Phi_1^T, \Phi_2^T, \ldots, \Phi_{d+1}^T\} \subset V_H$ denote the basis functions that belong to the $d+1$ nodal points in $T$. Hou and Wu [26] proposed the following oversampling strategy: solve for $\tilde{\Phi}_j^T \in V_h(U(T))$ with

$$(3.4) \quad \int_{U(T)} A\nabla\tilde{\Phi}_j^T \cdot \nabla\phi_h = 0 \quad \text{for all } \phi_h \in \mathring{V}_h(U(T))$$

and the boundary condition $\tilde{\Phi}_j^T = E_T(\Phi_j^T)$ on $\partial U(T)$, where $E_T(\Phi_j^T)$ denotes the affine extension of $(\Phi_j^T)|_T$. Then, for a given coarse function $\Phi_H \in V_H$, $\Phi_H^{\mathrm{MsFEM}}$ is defined by

$$\Phi_H^{\mathrm{MsFEM}} = \sum_{j=1}^{d+1} c_j \tilde{\Phi}_j^T,$$

where the $c_j$ are such that $\Phi_H^{\mathrm{MsFEM}}(x_j) = \Phi_H(x_j)$ for all $d+1$ coarse nodes $x_j$ of $T$. The final coarse scale equation in the PG formulation reads as follows: find $u_H^{\mathbf{1},\mathrm{MsFEM}} \in V_{h,\mathcal{T}_H}$ with

$$(3.5) \quad \int_\Omega A\nabla_H u_H^{\mathbf{1},\mathrm{MsFEM}} \cdot \nabla\Phi_H = \int_\Omega f\Phi_H$$

for all $\Phi_H \in V_H$. Observe that $u_H^{\mathrm{MsFEM}}$ is a nonconforming approximation of $u$. In [28, 16], a slightly different condition is used to define the coefficients $c_i$. However, it turns out that this modified condition leads to nothing but Oversampling Strategy 2 below.

We shall rephrase this multiscale method with oversampling strategy in the framework of Definition 3.2. Let $Q_T(\Phi_H) := \Phi_H^{\mathrm{MsFEM}} - E_T(\Phi_H)$ define the local corrector, i.e., an operator that communicates fine scale information to the coarse scale equation. The corresponding reduced fine scale space $\mathring{V}_h^r(U(T))$ is given by

$$(3.6) \quad \mathring{V}_h^r(U(T)) := \mathring{V}_h(U(T)) \setminus \mathrm{span}\{\Phi_1^T, \Phi_2^T, \ldots, \Phi_{d+1}^T\}$$

with nodal basis functions $\{\Phi_1^T, \Phi_2^T, \ldots, \Phi_{d+1}^T\} \subset V_H$. Since

$$Q_T(\Phi_H)(x_i) = \Phi_H^{\mathrm{MsFEM}}(x_i) - \Phi_H(x_i) = 0 \quad \text{for all nodes } x_i \text{ in } T,$$

$Q_T(\Phi_H) \in \mathring{V}_h^r(U(T))$. Moreover, by the definition of $\Phi_H^{\mathrm{MsFEM}}$, $Q_T(\Phi_H) \in \mathring{V}_h^r(U(T))$ satisfies

$$\int_{U(T)} A\left(\nabla\Phi_H(x_T) + \nabla Q_T(\Phi_H)\right) \cdot \nabla\phi_h$$

$$= \int_{U(T)} A\left(\nabla E_T(\Phi_H) + \nabla Q_T(\Phi_H)\right) \cdot \nabla\phi_h$$

$$= \sum_{i=1}^{d} c_i \int_{U(T)} A\nabla\tilde{\Phi}_i^T \cdot \nabla\phi_h = 0$$

for all $\phi_h \in \mathring{V}_h^r(U(T))$. Since $\nabla\Phi_H(x_T)$ is a constant in $U(T)$, we may rewrite $Q_T(\Phi_H)$ in terms of a corrector basis. This gives us the first definition of oversampling within our framework.

OVERSAMPLING STRATEGY 1. *Let $\mathring{V}_h^r(U(T))$ denote the reduced fine scale space given by (3.6), and let $w_{h,T,i}^{\mathcal{U},\mathbf{1}} \in \mathring{V}_h^r(U(T))$ (for $i \in \{1,2,\ldots,d\}$) denote the solution of*

$$(3.7) \qquad \int_{U(T)} A\nabla w_{h,T,i}^{\mathcal{U},\mathbf{1}} \cdot \nabla\phi_h = -\int_{U(T)} Ae_i \cdot \nabla\phi_h \quad \text{for all } \phi_h \in \mathring{V}_h^r(U(T)).$$

*For $\Phi_H \in V_H$ we define the corrector $Q_h^{\mathcal{U},\mathbf{1}}(\Phi_H) \in V_{h,\mathcal{T}_H}$ by*

$$Q_h^{\mathcal{U},\mathbf{1}}(\Phi_H) := \sum_{T \in \mathcal{T}_H} \chi_T(x) \sum_{i=1}^{d} \partial_{x_i}\Phi_H(x_T) \, w_{h,T,i}^{\mathcal{U},\mathbf{1}}(x).$$

*Let $u_H^{\mathbf{1}} \in V_H$ be the solution of (3.2), i.e.,*

$$\sum_{T \in \mathcal{T}_H} \int_T A\left(\nabla u_H^{\mathbf{1}} + \nabla Q_h^{\mathcal{U},\alpha}(u_H^{\mathbf{1}})\right) \cdot \nabla\Phi_H = \int_\Omega f\Phi_H \quad \text{for all } \Phi_H \in V_H.$$

*Then, $u_H^{\mathbf{1},\mathrm{MsFEM}} := u_H^{\mathbf{1}} + Q_h^{\mathcal{U},\mathbf{1}}(u_H^{\mathbf{1}})$ defines the MsFEM approximation obtained with Oversampling Strategy 1. Obviously, $u_H^{\mathbf{1},\mathrm{MsFEM}}$ solves (3.5).*

*Remark* 3.3. The explicit boundary condition for the local problems (3.4) is often missing in the literature (cf. [26, 14]). However, it seems that these computations were performed for the case described above; i.e., the solution $\tilde{\Phi}_i^T$ of (3.4) takes the values of an affine function on $\partial U(T)$ (cf. [16, 28], which also refer to the numerical experiments in [26]). In some works (cf. [13]) the local problems (3.4) are formulated with the boundary condition $\tilde{\Phi}_i^T = \Phi_i^T$ on $\partial U(T)$. This seems to be a mistake, because the new basis functions will be equal to zero whenever $U(T)$ is larger than the support of the original basis functions.

**3.2. Oversampling motivated from homogenization theory.** The second type of oversampling is motivated from numerical homogenization theory. Assume that we regard

$$\text{find } u^\varepsilon \in H_0^1(\Omega) \text{ with } \quad \int_\Omega A^\varepsilon \nabla u^\varepsilon \cdot \nabla\Phi = \int_\Omega f\Phi \quad \text{for all } \Phi \in H_0^1(\Omega),$$

and assume that $A^\varepsilon$ is uniformly bounded and coercive in $\varepsilon$, that $A^\varepsilon$ is $H$-convergent to some matrix $A^0$, and that $u^\varepsilon \rightharpoonup u^0$ in $H^1(\Omega)$, where $u^0 \in H_0^1(\Omega)$ is called the homogenized solution.

Then, a numerical approximation of the homogenized solution $u^0$ can be obtained by discretizing a more convenient equation (see (3.8) below). For this purpose, let $B(x, \eta)$ denote an open ball centered at $x \in \Omega$ with radius $\eta > 0$, and let $N(x, \eta)$ denote an open neighborhood of $x \in \Omega$ with a Lipschitz boundary. It is assumed that there exist $0 < c \leq C$ so that for all $\eta > 0$ and all $x \in \Omega$ there holds $c|B(x, \eta)| \leq |N(x, \eta)| \leq C|B(x, \eta)|$. We seek $u^{\varepsilon, \eta, \zeta} \in H_0^1(\Omega)$ that solves

$$(3.8) \quad \int_\Omega |N(x, \eta)|^{-1} \int_{N(x, \eta)} A^\varepsilon(y) \left( \nabla u^{\varepsilon, \eta, \zeta}(x) + \nabla_y Q(u^{\varepsilon, \eta, \zeta})(x, y) \right) \cdot \nabla \Phi(x) \, dy \, dx$$

$$= \int_\Omega f(x) \Phi(x) \, dx$$

for $\Phi \in H_0^1(\Omega)$, where for given $\Psi \in H_0^1(\Omega)$ the corrector $Q(\Psi)(x, \cdot) \in H_0^1(N(x, \eta + \zeta))$ is determined by

$$\int_{N(x, \eta+\zeta)} A^\varepsilon(y) \left( \nabla \Psi(x) + \nabla_y Q(\Psi)(x, y) \right) \cdot \nabla \phi(y) \, dy = 0 \text{ for all } \psi \in H_0^1(N(x, \eta + \zeta)).$$

If $\zeta = \zeta(\eta)$ and $\lim_{\eta \to 0} \frac{\zeta(\eta)}{\eta} = 0$, then it holds that

$$\lim_{\eta, \zeta \to 0} \lim_{\varepsilon \to 0} \|u^0 - u^{\varepsilon, \eta, \zeta}\|_{H^1(\Omega)} = 0.$$

As a consequence thereof, we get that

$$\lim_{\eta, \zeta \to 0} \lim_{\varepsilon \to 0} \|u^\varepsilon - u^{\varepsilon, \eta, \zeta}\|_{L^2(\Omega)} = 0.$$

This result was shown by Gloria [17, 18] in a general nonlinear setting. Since $u^{\varepsilon, \eta, \zeta}$ yields a good approximation of $u^\varepsilon$, the result suggests looking at discretizations of (3.8). This was exploited, for instance, in [23]. A general numerical framework that can be seen as a discretization of (3.8) was proposed in [24]. Particularly, the HMM and the MsFEM are recovered from the framework, which leads to a straightforward oversampling strategy. This strategy can be formulated as follows (cf. [18, 24, 25]).

OVERSAMPLING STRATEGY 2. *For* $i \in \{1, 2, \ldots, d\}$, *let* $w_{h,T,i}^{\mathcal{U}, \mathbf{2}} \in \mathring{V}_h(U(T))$ *solve*

$$(3.9) \quad \int_{U(T)} A \nabla w_{h,T,i}^{\mathcal{U}, \mathbf{2}} \cdot \nabla \phi_h = - \int_{U(T)} A e_i \cdot \nabla \phi_h \quad \text{for all } \phi_h \in \mathring{V}_h(U(T)),$$

*and for* $\Phi_H \in V_H$, *let* $Q_h^{\mathcal{U}, \mathbf{2}}(\Phi_H) \in V_{h, \mathcal{T}_H}$ *denote the corrector given by* (3.1). *If* $u_H^{\mathbf{2}} \in V_H$ *is the solution of* (3.2), *then* $u_H^{\mathbf{2}} + Q_h^{\mathcal{U}, \mathbf{2}}(u_H^{\mathbf{2}})$ *defines the MsFEM approximation obtained with Oversampling Strategy* 2. *We therefore denote*

$$u_H^{\mathbf{2}, \text{MsFEM}} := u_H^{\mathbf{2}} + Q_h^{\mathcal{U}, \mathbf{2}}(u_H^{\mathbf{2}}).$$

We immediately see that Oversampling Strategies 1 and 2 differ only in the fine scale trial space for the local problems and that they are identical for $U(T) = T$, even though Oversampling Strategy 2 was formulated independently of Oversampling Strategy 1. In [28, 16], Oversampling Strategy 2 is written in terms of an asymptotic expansion in the periodic case. Also note that this second approach is closely related to the finite element HMM, where the same type of oversampling is used (cf. [9, 10, 11, 22]). Notably, the HMM and the MsFEM can be reinterpreted in a common homogenization framework (cf. [17, 18]) and in a common numerical framework (cf. [24]).

**3.3. Discussion of the strategies.** As we just discussed, there are two widely used strategies for oversampling for the MsFEM. However, the difference between both approaches is only minor and the behavior of the resulting approximations appears to be qualitatively the same. The small difference in the local trial spaces does not seem to have a significant impact. At least, the error estimates available for Oversampling Strategies 1 and 2 are very similar. The literature does not even distinguish between these strategies. For instance, [6] (using Oversampling Strategy 1) claims to generalize the results of [16] (using Oversampling Strategy 2). Such a mixture of strategies can be observed in several works on this topic. To the best of our knowledge, even though both approaches seem to behave identically, a rigorous proof of this conjecture is still missing. Oversampling Strategy 1 suggests fixing the corrector $Q_T^1(\Phi)$ in the corners of the coarse grid element $T$ (forcing it to zero), whereas the corrector proposed by Oversampling Strategy 2 does not have such a restriction leaving it completely free in these corners.

*Remark* 3.4. As already mentioned, the MsFEM might also be considered in a symmetric formulation (cf. [16]); i.e., the coarse scale equation reads as follows: find $u_H \in V_H$ with

$$\sum_{T \in \mathcal{T}_H} \int_T A \left( \nabla u_H^\alpha + \nabla Q_h^{\mathcal{U},\alpha}(u_H^\alpha) \right) \cdot \left( \nabla \Phi_H + \nabla Q_h^{\mathcal{U},\alpha}(\Phi_H) \right)$$

$$= \int_\Omega f(\Phi_H + Q_h^{\mathcal{U},\alpha}(\Phi_H))$$

for all $\Phi_H \in V_H$ and where $Q_h^{\mathcal{U},\alpha}$ is defined either with Oversampling Strategy 1 or 2. However, the theoretical and numerical results in [28] show that this version of the method still suffers from resonance errors. One explanation was suggested by Gloria [18], who proposed a simple computation:

$$\int_T A \left( \nabla u_H^\alpha + \nabla Q_h^{\mathcal{U},\alpha}(u_H^\alpha) \right) \cdot \left( \nabla \Phi_H + \nabla Q_h^{\mathcal{U},\alpha}(\Phi_H) \right)$$

$$= \int_T A \left( \nabla u_H^\alpha + \nabla Q_h^{\mathcal{U},\alpha}(u_H^\alpha) \right) \cdot \nabla \Phi_H$$

$$+ \int_T A \left( \nabla u_H^\alpha + \nabla Q_h^{\mathcal{U},\alpha}(u_H^\alpha) \right) \cdot \nabla Q_h^{\mathcal{U},\alpha}(\Phi_H)$$

$$= \int_T A \left( \nabla u_H^\alpha + \nabla Q_h^{\mathcal{U},\alpha}(u_H^\alpha) \right) \cdot \nabla \Phi_H$$

$$+ \int_{U(T)\backslash T} A \left( \nabla u_H^\alpha + \nabla Q_h^{\mathcal{U},\alpha}(u_H^\alpha) \right) \cdot \nabla Q_h^{\mathcal{U},\alpha}(\Phi_H).$$

This means that the effective MsFEM bilinear forms in the PG and non-PG formulations differ in the term

$$\sum_{T \in \mathcal{T}_H} \int_{U(T)\backslash T} A \left( \nabla u_H^\alpha + \nabla Q_h^{\mathcal{U},\alpha}(u_H) \right) \cdot \nabla Q_h^{\mathcal{U},\alpha}(\Phi_H),$$

which still seems to contain the problematic boundary layers that we tried to get rid of. Observe that we integrate over the layer $U(T)\backslash T$. This is exactly the region where we encounter unpleasant boundary effects of the correctors $Q_h^{\mathcal{U},\alpha}(u_H)$ and $Q_h^{\mathcal{U},\alpha}(\Phi_H)$. This might imply that preference should be given to the PG formulation. Note,

however, that uniqueness and existence of discrete solutions have not been proved for general oversampling so far.

Let us review the two essential results concerning the convergence of MsFEM approximations with oversampling. The first result is due to Gloria and is the most general result currently available for Oversampling Strategy 2.

THEOREM 3.5. *Let $f \in L^2(\Omega)$ and $A^\varepsilon \in L^\infty(\Omega, \mathbb{R}^{d \times d})$ be a sequence of (possibly nonsymmetric) matrices with uniform spectral bounds $\gamma_{\min} > 0$ and $\gamma_{\max} \geq \gamma_{\min}$,*

$$(3.10) \qquad \sigma(A^\varepsilon(x)) \subset [\gamma_{\min}, \gamma_{\max}] \quad \text{for almost all } x \in \Omega \text{ and for all } \varepsilon > 0,$$

*and assume that $A^\varepsilon$ is H-convergent. Furthermore, let $u_H^\varepsilon \in V_H$ denote the corresponding MsFEM approximation obtained with Oversampling Strategy 2, and let*

$$\frac{\text{diam}(U(T)) - \text{diam}(T)}{\text{diam}(T)} \to 0 \quad \text{for } H \to 0.$$

*Then we have*

$$\lim_{H \to 0} \lim_{\varepsilon \to 0} \|u^\varepsilon - u_H^\varepsilon\|_{L^2(\Omega)} = 0.$$

The proof for general nonsymmetric coefficients is given in [20, Theorem 5], and the case of nonlinear problems is presented in [18, Theorem 6 and Remark 7]. At first glance the result appears counterintuitive in the sense that it suggests letting the oversampling converge to zero. However, the first limit is in $\varepsilon$, which makes the relative thickness $\frac{\varepsilon}{d_{\mathcal{U}}^{\min}}$ of the oversampling layer grow to infinity. Hence, the correct interpretation is that for fixed $\varepsilon$ the computational domains should blow up to infinity. In this case, the optimal corrector problem is an equation formulated on the whole $\mathbb{R}^d$. These corrector problems are exactly the cell problems known from periodic and stochastic homogenization theory. In the periodic setting the classical cell problems can be extended to the $\mathbb{R}^d$ by periodicity, and in the stochastic setting they are directly formulated in $\mathbb{R}^d$ to obtain the correct stochastic average (cf. [31]).

Theorem 3.5 gives a clear message in the case of extremely small microscopic variations. If $\varepsilon$ (the characteristic length scale of the fine scale oscillations) is (globally) sufficiently small, then the resulting MsFEM approximation yields very good approximations. This is a very important result, but it is purely qualitative. For example, it does not answer the question of how (thick) to choose an oversampling patch. We cannot predict how the method behaves if there is a large spectrum of oscillations without a scale separation. For instance, we might encounter variations, where it is hard to tell which of them are macroscopic and which are microscopic (i.e., "$\varepsilon$-dependent"). In practice, we do not construct an artificial sequence in $\varepsilon$; we have only a given scenario and a given set of data.

The next theorem due to Hou, Wu, and Zhang is much more restrictive, but it gives a more quantitative answer than Theorem 3.5.

THEOREM 3.6. *Assume that $d = 2$, $f \in L^2(\Omega)$, and $A$ is a bounded, elliptic, symmetric, and $\varepsilon$-periodic $C^3$-matrix, i.e., $A(x) = A_p(\frac{x}{\varepsilon})$, with $A_p \in C^3([0,1]^d, \mathbb{R}^{d \times d}_{sym})$ being periodic. Let $u_H^\varepsilon \in V_{h, \mathcal{T}_H}$ denote the MsFEM approximation obtained with Oversampling Strategy 2. Then*

$$\|u_H^\varepsilon - u^\varepsilon\|_{L^2(\Omega)} \leq C \left( \frac{\varepsilon}{d_{\mathcal{U}}^{\min}} + H + \varepsilon(\log H)^{\frac{1}{2}} \right),$$

$$\left( \sum_{T \in \mathcal{T}_H} \|\nabla u_H^\varepsilon - \nabla u^\varepsilon\|_{L^2(\Omega)}^2 \right)^{\frac{1}{2}} \leq C \left( \frac{\varepsilon}{d_{\mathcal{U}}^{\min}} + H + \varepsilon^{\frac{1}{2}} \right).$$

A proof of this theorem is given in [28]. The assumption $d = 2$ seems to be essential for their strategy. Note that in [28] the theorem is formulated without the $\frac{1}{d_{\mathcal{U}}^{\min}}$ contribution. Instead, the authors make the assumption that the oversampling layer is sufficiently large. Following their proofs one can easily see that the generalized estimate reads as above (cf. [14] for the case $d_{\mathcal{U}}^{\min} = CH$). In particular, the $\frac{\varepsilon}{d_{\mathcal{U}}^{\min}}$-term describes the decay of the error between the exact corrector and the corrector with wrong boundary conditions in a coarse element $T$. The decay turns out to be inversely proportional to the thickness of the layer. Because of the $\varepsilon$ scaling of the solution, the effective term becomes $\frac{\varepsilon}{d_{\mathcal{U}}^{\min}}$. This seems to be a sharp estimate for the decay due to the findings in [16, 5]. A proof of Theorem 3.6 for Oversampling Strategy 1 can be achieved in the same fashion as in [28]. Theorem 3.6 predicts the following: if locally O($H$)=O($\varepsilon$), the patch size of the local problems must not be of order O($H$) to preserve convergence. Still, the theorem gives only an answer of how to choose the oversampling patches $\mathcal{U}$ if $\varepsilon$ is a known parameter.

If the thickness of the oversampling layer is of order O($h$), both estimates in Theorem 3.6 receive an order O($\frac{\varepsilon}{h}$) term and the right-hand sides remain large. In general, the thickness $d_{\mathcal{U}}^{\min}$ must be large in comparison to $\varepsilon$. Analytically, this implies that O($H$)-oversampling might be not enough in regions where we deal with resonance errors due to O($H$)=O($\varepsilon$). This seems to show up in the numerical experiments in [28], where the authors observe a stagnation in the convergence of the $H^1$-error for $H$ entering the region with O($H$)=O($\varepsilon$). The effect on the $L^2$-error is less strong. However, the value of $d_{\mathcal{U}}^{\min}$ is missing in the experiments in [28], so we can assume only that $d_{\mathcal{U}}^{\min}$ is of order $H$. Otherwise the computation of the MsFEM basis functions becomes quite expensive. However, we note that there also exists a modification of Oversampling Strategy 2 proposed by Gloria (cf. [19, Theorem 3.1] and [20, paragraphs 5.3 and 5.4]) where the local problems are regularized by adding the term $\kappa^{-1}(w_{h,T,i}^{\mathcal{U},\mathbf{2}}, \phi_h)_{L^2(U(T))}$ (for large $\kappa > 0$) to the left-hand side of problem (3.9). Using this modified strategy, the Theorem 3.6–type estimates can be improved enormously, even without restrictions on space dimensions and much weaker assumptions on the regularity of $A^\varepsilon$.

*Remark* 3.7. In [16], the symmetric version (3.3) of the MsFEM is considered. Here, the derived $L^2$-estimate reads as

$$\|u_H^\varepsilon - u^\varepsilon\|_{L^2(\Omega)} \leq C\left(\varepsilon + H^2 + \varepsilon(\log H) + \frac{\varepsilon}{d_{\mathcal{U}}^{\min}} + C_r\left(\frac{\varepsilon}{H}\right)^2\right),$$

where $u_H^\varepsilon$ denotes the MsFEM approximation of the symmetric problem (3.3) determined with Oversampling Strategy 1. Due to the numerical experiments in [16], $C_r$ seems to have a considerable size so that this term is dominating the estimate if locally O($\varepsilon$)=O($H$). We notice that the estimate is worse than the $L^2$-estimate for the PG version of the method, because the last term cannot be reduced even for large $d_{\mathcal{U}}^{\min}$. These observations are consistent with Remark 3.4. However, this leads to an additional problem of the MsFEM with Oversampling Strategy 1 or 2. On the one hand, the PG version should be preferred over the symmetric version (see the estimates). On the other hand, the existence and uniqueness of the corresponding MsFEM approximations have not been established so far, not to mention the stability. For the symmetric version, we can simply exploit the ellipticity of $A$ to conclude that the method is well posed and stable. For the PG version there is no such argument. The only result is a perturbation result due to Gloria [18], saying that if the oversampling size is small enough (i.e., if the difference between the PG formulation and

symmetric formulation is small enough), then we still have existence and uniqueness. The lack of knowledge regarding the general well-posedness of the PG MsFEM with Oversampling Strategies 1 and 2 is a big issue of these approaches.

The $\varepsilon$-terms in the estimates that cannot be reduced with $H \gtrsim \varepsilon$ should be seen as fixed modeling errors. They describe the error between exact solution and homogenized solution. In a general nonperiodic nonstochastic scenario they cannot be quantified.

In conclusion we have two findings. First, in general, both approaches do not show clear asymptotics for a convergence to the exact solution (for $H \gtrsim \varepsilon$). There is always a remainder of order $\varepsilon$, even if $U(T) = \Omega$. In particular, this is a problem if $\varepsilon$ is unknown or if the microstructure is heterogeneous. Second, if the modeling error of order $\varepsilon$ is negligible, Theorem 3.6 still suggests that linear convergence (with respect to $H$) can be achieved only if the oversampling thickness scales with $O(1)$, which makes the local problems prohibitively expensive. Since the estimate for the decay rate $\frac{\varepsilon}{d_{\mathcal{U}}^{\min}}$ of the corrector error is sharp (in the periodic setting), we cannot hope for much improvement of the final error estimates stated in Theorem 3.6.

We may summarize the following issues, which we address to solve with our new oversampling strategy to be proposed in the next section:

(a) elimination of resonance errors of any kind,
(b) clear prediction for the size of oversampling patches without explicit knowledge about the microstructure or scale separation,
(c) construction of a conforming approximation in $H_0^1(\Omega)$,
(d) a quantitative error analysis in $H$ without restrictive regularity assumptions on the coefficients and for all space dimensions,
(e) a priori error estimation in the fully discrete setting (previous results were obtained under the assumption that the local problems are solved exactly),
(f) formulation of a stable approach for which we can guarantee existence and uniqueness of the resulting MsFEM approximation, and
(g) prevention of unstable splittings due to point evaluations as, e.g., required to implement the constraint in Oversampling Strategy 1 (cf. the definition of $\mathring{V}_h^r(U(T))$ in (3.6)).

Note that points (d) and (e) could be done in the periodic setting for Oversampling Strategies 1 and 2 by using, e.g., the techniques presented in [19, 20].

**4. Constrained oversampling.** In this section we introduce a third oversampling strategy for which we derive a quantitative a priori error estimate. The results are presented in subsection 4.1, and a corresponding proof is given in subsection 4.2. All the results require solely the assumptions stated in section 2.1 to be satisfied, i.e., $A \in L^\infty(\mathbb{R}_{\mathrm{sym}}^{d \times d})$ uniformly positive definite and $f \in L^2(\Omega)$.

**4.1. New strategy and quantitative error estimates.** In the following, let $\mathcal{N}_H$ denote the set of interior vertices of the coarse grid $\mathcal{T}_H$. For a given node $z \in \mathcal{N}_H$, $\Phi_z \in V_H$ denotes the corresponding nodal basis function as before.

Our new approach is based on some multiscale decomposition of the space $V_h$,

$$(4.1) \qquad\qquad\qquad V_h = V_H \oplus W_h,$$

where the space $W_h$ contains the "fine scale" functions of $V_h$, i.e., functions that are not captured by $V_H$. More precisely, we choose $W_h$ to be the kernel of some Clément-type quasi-interpolation operator $I_H : H_0^1(\Omega) \to V_H$,

$$(4.2) \qquad\qquad\qquad W_h := \{v \in V_h \mid I_H(v) = 0\}.$$

Several choices for $I_H$ are possible. We refer the reader to [34] for an axiomatic characterization. In this paper, for the sake of simplicity, we choose the particular operator introduced in [4]. Given $v \in H_0^1(\Omega)$, $I_H v := \sum_{z \in \mathcal{N}_H} (I_H v)(z) \Phi_z$ is determined by the nodal values

$$(4.3) \qquad (I_H v)(z) := \frac{\int_\Omega v \Phi_z}{\int_\Omega \Phi_z} \quad \text{for } z \in \mathcal{N}_H.$$

The nodal values are weighted averages of the function over nodal patches $\omega_z := \operatorname{supp} \Phi_z$. The operator is linear, surjective, bounded, and invertible on the finite element space $V_H$. Hence, the decomposition (4.1) exists and is stable; it is even orthogonal in $L^2(\Omega)$.

Recall the (local) approximation and stability properties of the interpolation operators $I_H$ [4]: There exists a generic constant $C$ such that for all $v \in H_0^1(\Omega)$ and for all $K \in \mathcal{T}_H$ it holds that

$$(4.4) \qquad H_T^{-1} \|v - I_H v\|_{L^2(K)} + \|\nabla(v - I_H v)\|_{L^2(K)} \leq C \|\nabla v\|_{L^2(\omega_K)},$$

where $\omega_K := \cup \{K' \in \mathcal{T}_H \mid K' \cap K \neq \emptyset\}$. The constant $C$ depends on the shape regularity of the finite element mesh $\mathcal{T}_H$ but not on the local mesh size $H_T := \operatorname{diam}(T)$.

*Remark* 4.1 (nodal interpolation). Since we consider a fully discrete setting, where corrector problems are solved in the fine scale finite element space $V_h$, we could have chosen nodal interpolation instead of Clément-type interpolation. The subsequent definitions and results will be almost verbatim the same. However, nodal interpolation does not satisfy the estimate (4.4) with an $h$-independent constant if $d > 1$. The best constant $C = C_d(h)$ reads as $C_2(h) = \log(H/h)$ and $C_3(h) = (H/h)^{-1}$ depending on the spatial dimension $d$ (cf. [38]). Since this constant enters basically all error estimates below, we would end up with an $h$-dependence of the multiplicative constants in the final error estimates. In two dimensions this can still be acceptable, because the dependence on $h$ is only logarithmic.

With the decomposition (4.1) we do not search the local correctors in the full fine scale space $V_h$ but only in the constrained space $W_h$. The advantage is the following: as stated in the previous section for Oversampling Strategies 1 and 2, the standard decay for the difference between the local correctors and the global "exact" corrector is of order $\frac{1}{d_{\mathcal{U}}^{\min}}$ (see Theorem 3.6), but in the constrained space $W_h$ we can achieve an exponential-type decay (cf. Lemma 4.9 below).

We now propose our new oversampling strategy.

OVERSAMPLING STRATEGY 3 (constrained oversampling). *Let $W_h$ denote the space given by (4.2), and define*

$$(4.5) \qquad \mathring{W}_h(U(T)) := \{v_h \in W_h \mid v_h|_{\Omega \setminus U(T)} = 0\}.$$

*The local correctors $w_{h,T,i}^{\mathcal{U},\mathbf{3}} \in \mathring{W}_h(U(T))$ (for $i \in \{1, 2, \ldots, d\}$) are defined as the (unique) solutions of*

$$(4.6) \qquad \int_{U(T)} A \nabla w_{h,T,i}^{\mathcal{U},\mathbf{3}} \cdot \nabla \phi_h = - \int_T A e_i \cdot \nabla \phi_h \quad \text{for all } \phi_h \in \mathring{W}_h(U(T)).$$

*For general $\Phi_H \in V_H$ we define the correction operator $Q_h^{\mathcal{U},\mathbf{3}} : V_H \to V_h$ by*

$$Q_h^{\mathcal{U},\mathbf{3}}(\Phi_H)(x) := \sum_{T \in \mathcal{T}_H} \sum_{i=1}^d \partial_{x_i} \Phi_H(x_T) w_{h,T,i}^{\mathcal{U},\mathbf{3}}(x).$$

*The global coarse scale approximation $u_H^{\mathbf{3}} \in V_H$ is the solution of (3.3); i.e., it solves*

$$\mathcal{A}^{\mathbf{3}}(u_H^{\mathbf{3}}, \Phi_H) := \int_\Omega A\left(\nabla u_H^{\mathbf{3}} + \nabla Q_h^{\mathcal{U},\mathbf{3}}(u_H^{\mathbf{3}})\right) \cdot \left(\nabla \Phi_H + \nabla Q_h^{\mathcal{U},\mathbf{3}}(\Phi_H)\right)$$

(4.7)
$$= \int_\Omega f(\Phi_H + Q_h^{\mathcal{U},\mathbf{3}}(\Phi_H)) \quad \text{for all } \Phi_H \in V_H.$$

*The corresponding MsFEM approximation is given by*

$$u_H^{\mathbf{3},\text{MsFEM}} := u_H^{\mathbf{3}} + Q_h^{\mathcal{U},\mathbf{3}}(u_H^{\mathbf{3}}).$$

Using the above definition of the localized space $\mathring{W}_h(U(T))$ does not assure that our new method boils down to the classical MsFEM in the case without oversampling. Nevertheless, this can be achieved by introducing a localized interpolation operator. Given some element $T \in \mathcal{T}_H$ and an admissible patch $U(T)$, we can define $I_H^{U(T)}$ to be the Clément-type quasi-interpolation operator with respect to the domain $U(T)$ (with extension by zero in $\Omega \setminus U(T)$). Then, the localized space $\mathring{W}_h(U(T))$ can be defined in analogy to $W_h$ with $I_H^{U(T)}$ replacing $I_H$. With this modification we obtain the classical MsFEM for $U(T) = T$. This is only a subtle detail, and all results still remain valid for these modified local spaces; however, this version would generate some technicalities in the proofs later on, which is why we decided to work with the definition (4.5).

*Remark* 4.2. The crucial differences between the classical Oversampling Strategies 1 and 2 and Oversampling Strategy 3 are the following:

(a) The variational problem for the local corrector in Oversampling Strategy 3 is posed in the constrained space $\mathring{W}_h(U(T))$, whereas the classical corrector problem seeks the local corrector in the full space $\mathring{V}_h(U(T))$ restricted to the patch.

(b) The support of the integrals on the right-hand sides in (3.7) and (3.9) is $U(T)$. In our new version we use only the element $T$. This allows us to exploit nice summation properties of the local projectors, without using indicator functions $\chi_T$ that lead to discontinuities.

(c) In the classical setting, the local correctors are restricted to the corresponding elements to derive the global corrector. For Oversampling Strategy 3, we simply sum up (weighted by the coefficients of the finite element function) the local contributions to get the global corrector. Note that our global corrector is conforming in the sense that its image is a subset of $V_h \subset H_0^1(\Omega)$, whereas the classical setup leads to a nonconforming corrector.

(d) In Oversampling Strategy 3, we do not use a PG formulation for the global problem (4.7). Since $A$ is assumed symmetric, a symmetric discretization appears more natural. Furthermore, we immediately inherit coercivity for the global bilinear form $\mathcal{A}^{\mathbf{3}}$. This gives us the existence and uniqueness of $u_H^{\mathbf{3}}$, and the arising MsFEM approximation is well posed and the method stable. The typical disadvantage of the symmetric version, which still suffers from resonance errors (which is why the PG formulation is typically preferred), does not remain for our strategy.

(e) In contrast to Oversampling Strategies 1 and 2, the corrector $Q_h^{\mathcal{U},\mathbf{3}}(\Phi_H)$ does not preserve the support of $\Phi_H$. In other words, the set of multiscale basis functions $\Phi_z + Q_h^{\mathcal{U},\mathbf{3}}(\Phi_z)$ with $z \in \mathcal{N}_H$ has an extended support. This results in a loss of sparsity in the stiffness matrix that corresponds with the global

problem (4.7). In order to still assemble the stiffness matrix in an efficient way, one might store the intersection domain for each two given oversampling patches (in storage types with low memory requirements). This can be easily done at the same time the grids for the local patches are being generated. Once all intersection domains are available, the matrix can be assembled efficiently. A quadrature rule that resolves the microstructure is needed for each of the strategies.

*Remark* 4.3 (perturbation of the right-hand side). We might also replace the right-hand side of (4.7) by the term $\int_\Omega f\Phi_H$. This introduces only a perturbation of order $\|Hf\|_{L^2(\Omega)}$ in the $H^1$-error.

*Remark* 4.4 (nonsymmetric formulation). As for the classical strategies, one might also consider the nonsymmetric PG formulations: find $u_H^{\mathbf{3}} \in V_H$ such that

$$\int_\Omega A\left(\nabla(u_H^{\mathbf{3}} + Q_h^{\mathcal{U},\mathbf{3}}(u_H^{\mathbf{3}}))\right)\cdot\nabla\Phi_H = \int_\Omega f\Phi_H \quad \text{for all } \Phi_H \in V_H$$

or

$$\int_\Omega A\left(\nabla u_H^{\mathbf{3}}\right)\cdot\nabla(\Phi_H + Q_h^{\mathcal{U},\mathbf{3}}(\Phi_H)) = \int_\Omega f\Phi_H \quad \text{for all } \Phi_H \in V_H.$$

In the spirit of homogenization theory, one might pose the question of whether Theorem 3.5 still holds for MsFEM approximations obtained with Oversampling Strategy 3. At least, this seems to be likely. The reason is that Theorem 3.5 in particular covers the case without oversampling (see also [17]), and the proof given in [18] goes back to the arguments used for the case without oversampling. But for $\mathcal{U} = \mathcal{T}_H$ (no oversampling), Oversampling Strategies 1 and 2 are identical, and Oversampling Strategy 3 is at least close to the classical approach. Especially concerning Oversampling Strategy 3, if the thickness of the oversampling layer decreases faster than the coarse mesh size, we are almost in the case of Oversampling Strategy 1, up to a small perturbation of the source term that is of order $\max_{T\in\mathcal{T}_H}\frac{|U(T)\setminus T|}{|T|}$ and that converges to zero under the assumptions of Theorem 3.5. However, such arguments still need a detailed investigation. In this sense, one might carefully study whether Oversampling Strategy 3 also covers the homogenization setting established by Gloria, with $u_H^{\mathbf{3}}$ converging to the homogenized solution as in Theorem 3.5. This might be an interesting result to ensure that Oversampling Strategy 3 is not worse than the classical strategies with respect to a homogenization setting.

Besides the advantages of our new strategy mentioned previously, e.g., its conformity, stability, and unique solvability, we formulate the main error estimate, which is proved in subsection 4.2.

THEOREM 4.5 (quantitative a priori error estimates). *Assume that we have* $A \in L^\infty(\Omega, \mathbb{R}_{sym}^{d\times d})$ *and* $f \in L^2(\Omega)$ *as in the general assumptions in section* 2.1. *Let* $\mathcal{T}_H$ *be a given coarse triangulation, and let* $\mathcal{U}$ *denote a corresponding set of admissible patches, with the property* $d_{\mathcal{U}}^{\min} \gtrsim H\log(H^{-1})$. *By* $\mathcal{T}_h$ *we denote a sufficiently accurate fine triangulation of* $\Omega$ *and by* $u_h$ *the associated finite element solution of* (2.3). *If* $u_H^{\mathbf{3},\text{MsFEM}}$ *is the MsFEM approximation determined with Oversampling Strategy* 3 *and if* $u_H^{\mathbf{3}}$ *denotes the corresponding coarse part, then the following a priori error estimates holds true for arbitrary mesh sizes* $H \geq h$:

$$\|\nabla u_h - \nabla u_H^{\mathbf{3},\text{MsFEM}}\|_{L^2(\Omega)} \leq CH,$$
$$\|u_h - u_H^{\mathbf{3},\text{MsFEM}}\|_{L^2(\Omega)} \leq CH^2,$$
$$\|u_h - u_H^{\mathbf{3}}\|_{L^2(\Omega)} \leq CH.$$

*Here, $C$ denotes generic constants that depend on $f$, $\gamma_{\min}$, and $\gamma_{\max}$ but not on $H$, $h$, the regularity of the exact solution, or the variations of $A$. Details on the constants are given in Theorems 4.13 and 4.15.*

**4.2. Proof of the main result.** Before we prove the error estimates for the MsFEM with the correctors presented in Oversampling Strategy 3, we introduce some simplifying notation for this subsection.

DEFINITION 4.6 (notation for Oversampling Strategy 3). *Let $w_{h,T,i}^{\mathcal{U},\mathbf{3}} \in \mathring{W}_h(U(T))$ denote the local corrector basis given by (4.6), let $Q_h^{\mathcal{U},\mathbf{3}}$ denote the corresponding corrector operator from Oversampling Strategy 3, and let $u_H^{\mathbf{3}}$ denote the arising (coarse) MsFEM approximation. In the following, we skip the redundant indices and use the following notation:*

$$w_T^i := w_{h,T,i}^{\mathcal{U},\mathbf{3}}, \quad Q_h := Q_h^{\mathcal{U},\mathbf{3}}, \quad u_H := u_H^{\mathbf{3}}, \quad and \quad u^{\textit{MsFEM}} := u_H + Q_h(u_H).$$

The first lemma treats the (unpractical) case of maximal oversampling.

LEMMA 4.7 (error estimate for maximal oversampling). *Let $U(T) = \Omega$ for all $T \in \mathcal{T}_H$. Then the multiscale approximation $u_H$ that solves (4.7) satisfies the error estimate*

$$\|\nabla u_h - \nabla(u_H + Q_h(u_H))\|_{L^2(\Omega)} \lesssim \gamma_{\min}^{-1}\|Hf\|_{L^2(\Omega)},$$

*where $u_h$ solves the reference problem (2.3).*

*If, moreover, $(f, w_h)_{L^2(\Omega)} = 0$ for all fine scale functions $w_h \in W_h$, then $u_H + Q_h(u_H) = u_h$.*

*Proof.* For $U(T) = \Omega$, $Q_h$ maps onto the fine scale space $W_h$. Given $\Phi_H \in V_H$, it is easily checked that $Q_h(\Phi_H) = \sum_{T \in \mathcal{T}_H} \sum_{i=1}^d \partial_{x_i} \Phi_H(x_T) w_T^i$ satisfies

$$\begin{aligned}
a(Q_h(\Phi_H), \phi_h) &= \int_\Omega A\nabla\left(\sum_{T\in\mathcal{T}_H}\sum_{i=1}^d \partial_{x_i}\Phi_H(x_T)w_T^i(x)\right)\cdot\nabla\phi_h \\
&= \int_\Omega A\left(\sum_{T\in\mathcal{T}_H}\sum_{i=1}^d \partial_{x_i}\Phi_H(x_T)\nabla w_T^i(x)\right)\cdot\nabla\phi_h \\
&= \sum_{T\in\mathcal{T}_H}\sum_{i=1}^d \partial_{x_i}\Phi_H(x_T)\int_\Omega A\nabla w_T^i(x)\cdot\nabla\phi_h \\
&= -\sum_{T\in\mathcal{T}_H}\sum_{i=1}^d \partial_{x_i}\Phi_H(x_T)\int_T Ae_i\cdot\nabla\phi_h \\
&= -\sum_{T\in\mathcal{T}_H}\int_T A\nabla\Phi_H\cdot\nabla\phi_h \\
&= -a(\Phi_H, \phi_h)
\end{aligned}$$

for all $\phi_h \in W_h$. This means that $Q_h$ is the orthogonal projection of $\Phi_H$ onto the fine scale space $W_h$ with respect to the scalar product $a(\cdot, \cdot)$. This yields the orthogonal decomposition

(4.8) $$V_h = \tilde{V}_H \oplus_{\perp_a} W_h, \quad \text{where } \tilde{V}_H := \{\Phi_H + Q_h(\Phi_H) \mid \Phi_H \in V_H\}.$$

Moreover, Galerkin orthogonality holds; i.e., for $e_h := u_h - (u_H + Q_h(u_H))$ and for arbitrary $\Phi_H + Q_h(\Phi_H) \in \tilde{V}_H$,

$$a(e_h, \Phi_H + Q_h(\Phi_H)) = a(u_h, \Phi_H + Q_h(\Phi_H)) - a(u_H + Q_h(u_H), \Phi_H + Q_h(\Phi_H))$$

(4.9) $$\stackrel{(4.7)}{=} 0.$$

The combination of (4.8) and (4.9) shows that $e_h \in W_h$, and therefore $I_H(e_h) = 0$. We obtain

$$\gamma_{\min}\|\nabla e_h\|_{L^2(\Omega)}^2 \leq a(e_h, e_h) = a(u_h, e_h) = \int_\Omega f e_h = \int_\Omega f(e_h - I_H(e_h)).$$

The application of the Cauchy–Schwarz inequality on the element level and the estimate (4.4) for the interpolation error yield the assertion. $\quad\square$

COROLLARY 4.8. *The new MsFEM is exact (up to the discretization error on the fine scale and oscillations* $\|Hf\|_{L^2(\Omega)}$ *of the right-hand side $f$) in the limit of maximal oversampling. This results holds true independent of the upper spectral bound $\gamma_{\max}$ and the variations of A. This is the next difference from the previous Oversampling Strategies 1 and 2.*

Although the error estimate in Lemma 4.7 is encouraging, maximal oversampling is not feasible. We shall study the decay of the correctors away from element they are associated with. For all $T \in \mathcal{T}_H$, define element patches in the coarse mesh $\mathcal{T}_H$ by

(4.10)
$$\begin{aligned}
&U_0(T) := T, \\
&U_k(T) := \cup\{T' \in \mathcal{T}_H \mid T' \cap U_{k-1}(T) \neq \emptyset\} \quad k = 1, 2, \ldots.
\end{aligned}$$

LEMMA 4.9 (decay of the ideal correctors). *Let $U(T) = \Omega$ for all $T \in \mathcal{T}_H$ in Oversampling Strategy 3, and let $w_T^i$ denote the corresponding local correctors defined in Definition 4.6 (and (4.6)). Then, for all $T \in \mathcal{T}_H$ and all $k \in \mathbb{N}$,*

$$\|A^{1/2}\nabla w_T^i\|_{L^2(\Omega\setminus U_k(T))} \lesssim e^{-rk}\|A^{1/2}\nabla w_T^i\|_{L^2(\Omega)},$$

*where $r$ is a positive constant that depends on the square root of the contrast but not on the mesh size or the variations of A.*

The proof of Lemma 4.9 requires the definition of cutoff functions and an additional lemma. For $T \in \mathcal{T}_H$ and $\ell, k \in \mathbb{N}$ with $k > \ell$, define $\eta_{T,k,\ell} \in P_1(\mathcal{T}_H)$ with nodal values

(4.11)
$$\begin{aligned}
&\eta_{T,k,\ell}(z) = 0 \quad \text{for all } z \in \mathcal{N}_H \cap U_{k-\ell}(T), \\
&\eta_{T,k,\ell}(z) = 1 \quad \text{for all } z \in \mathcal{N}_H \cap (\Omega \setminus U_k(T)), \text{ and} \\
&\eta_{T,k,\ell}(z) = \frac{m}{\ell} \quad \text{for all } x \in \mathcal{N}_H \cap \partial U_{k-\ell+m}(T), \; m = 0, 1, 2, \ldots, \ell.
\end{aligned}$$

For a sketch in one dimension, see Figure 4.1.

Given some $w \in W_h$, the product $\eta_{T,k,\ell}w$ is not in $W_h$ in general. However, the distance of $\eta_{T,k,\ell}w$ and $W_h$ is small in the following sense.

LEMMA 4.10. *Given $w \in W_h$ and some cutoff function $\eta_{T,k,\ell} \in P_1(\mathcal{T}_H)$ as in (4.11), there exists some $\tilde{w} \in \mathring{W}_h(\Omega \setminus U_{k-\ell-1}(T)) \subset W_h$ such that*

$$\|\nabla(\eta_{T,k,\ell}w - \tilde{w})\|_{L^2(\Omega)} \lesssim \ell^{-1}\|\nabla w\|_{L^2(U_{k+2}(T)\setminus U_{k-\ell-2}(T))}.$$

*Proof.* Fix some $T \in \mathcal{T}_H$ and $k \in \mathbb{N}$, and let $\eta_\ell := \eta_{T,k,\ell}$. The operator $I_h : H_0^1(\Omega) \cap C(\bar{\Omega}) \to V^h$ denotes the nodal interpolant with respect to the mesh $\mathcal{T}_h$.
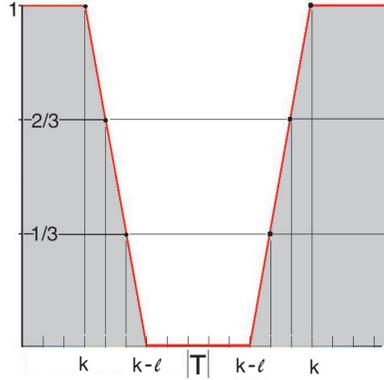
FIG. 4.1. *Sketch of $\eta_{T,k,\ell}$ (red curve) in one dimension for $k = 5$ and $\ell = 3$. $\eta_{T,k,\ell}$ is equal to zero on $T$ and also on the first 2 $(= k - \ell)$ coarse grid layers around $T$. Then it grows linearly on the layers $\ell = 3$ until $k = 5$. On the remaining layers $\eta_{T,k,\ell}$ is constantly equal to 1.*

Recall that for all quadratic polynomials $p$ and all $t \in \mathcal{T}_h$, $I_h$ fulfills the (local) approximation and stability estimates

$$(4.12) \qquad \|\nabla(p - I_h p)\|_{L^2(t)} \lesssim h_t \|\nabla^2 p\|_{L^2(t)} \quad \text{and} \quad \|\nabla(I_h p)\|_{L^2(t)} \lesssim \|\nabla p\|_{L^2(t)}.$$

We will use this estimate for the $\mathcal{T}_h$-piecewise quadratic function $p = \eta_\ell w$. Since $\nabla^2 \eta_\ell = \nabla^2 w = 0$ in every $t \in \mathcal{T}_h$, we have that $\nabla_h^2 \eta_\ell w = \nabla \eta_\ell \cdot \nabla w$ in $t$.

According to [34, Lemma 1], there exists some $v \in V^h$ such that
$$(4.13)$$
$$I_H v = I_H I_h(\eta_\ell w), \quad \|\nabla v\|_{L^2(\Omega)} \lesssim \|\nabla I_H I_h(\eta_\ell w)\|_{L^2(\Omega)}, \quad \text{and} \quad \text{supp}(v) \subset \Omega \setminus U_{k-\ell-1}(T).$$

Hence, $\tilde{w} := I_h(\eta_\ell w) - v \in \mathring{W}_h(\Omega \setminus U_{k-\ell-1}(T))$. Since $I_H I_h(cw) = c I_H w = 0$ for any $c \in \mathbb{R}$, we set $c_K^\ell := |\omega_K|^{-1} \int_{\omega_K} \eta_\ell$ for $K \in \mathcal{T}_H$ and get

$$(4.14)$$
$$\|\nabla I_H I_h(\eta_\ell w)\|_{L^2(\Omega)}^2$$

$$\overset{(4.11)}{=} \sum_{\substack{K \in \mathcal{T}_H: \\ K \subset \overline{U_{k+1}(T)} \setminus U_{k-\ell-1}(T)}} \left\| \nabla I_H I_h \left( (\eta_\ell - c_K^\ell) w \right) \right\|_{L^2(K)}^2$$

$$\overset{(4.12),(4.4)}{\lesssim} \sum_{\substack{K \in \mathcal{T}_H: \\ K \subset \overline{U_{k+1}(T)} \setminus U_{k-\ell-1}(T)}} \left\| \nabla \left( (\eta_\ell - c_K^\ell) w \right) \right\|_{L^2(\omega_K)}^2$$

$$\overset{(4.2)}{\lesssim} \sum_{\substack{K \in \mathcal{T}_H: \\ K \subset \overline{U_{k+1}(T)} \setminus U_{k-\ell-1}(T)}} \|(\nabla \eta_\ell)(w - I_H w)\|_{L^2(\omega_K)}^2 + \left\| (\eta_\ell - c_K^\ell) \nabla w \right\|_{L^2(\omega_K)}^2$$

$$\overset{(4.11)}{\lesssim} \sum_{\substack{K \in \mathcal{T}_H: \\ K \subset \overline{U_k(T)} \setminus U_{k-\ell}(T)}} \|(\nabla \eta_\ell)(w - I_H w)\|_{L^2(K)}^2 \quad + \sum_{\substack{K \in \mathcal{T}_H: \\ K \subset \overline{U_{k+1}(T)} \setminus U_{k-\ell-1}(T)}} \left\| (\eta_\ell - c_K^\ell) \nabla w \right\|_{L^2(\omega_K)}^2$$

$$\lesssim \quad \|H\nabla\eta_\ell\|_{L^\infty(\Omega)}^2\|\nabla w\|_{L^2(U_{k+1}(T)\setminus U_{k-\ell-1}(T))}^2 + \sum_{\substack{K\in\mathcal{T}_H: \\ K\subset\overline{U_{k+1}(T)}\setminus U_{k-\ell-1}(T)}} \left\|\left(\eta_\ell - c_K^\ell\right)\nabla w\right\|_{L^2(\omega_K)}^2$$

$$\lesssim \quad \|H\nabla\eta_\ell\|_{L^\infty(\Omega)}^2\|\nabla w\|_{L^2(U_{k+2}(T)\setminus U_{k-\ell-2}(T))}^2.$$

In the last step, we used the Lipschitz bound

$$\|\eta_\ell - c_K^\ell\|_{L^\infty(\omega_K)}^2 \lesssim H^2\|\nabla\eta_\ell\|_{L^\infty(\omega_K)}^2.$$

In summary we get with the previous computations

$$\|\nabla(\eta_\ell w - \tilde{w})\|_{L^2(\Omega)}^2 \overset{(4.13)}{\lesssim} \|\nabla(\eta_\ell w - I_h(\eta_\ell w))\|_{L^2(\Omega)}^2 + \|\nabla I_H I_h(\eta_\ell w)\|_{L^2(\Omega)}^2$$

$$\overset{(4.12),(4.14)}{\lesssim} \|h\nabla\eta_\ell\cdot\nabla w\|_{L^2(\Omega)}^2 + \|H\nabla\eta_\ell\|_{L^\infty(\Omega)}^2\|\nabla w\|_{L^2(U_{k+2}(T)\setminus U_{k-\ell-2}(T))}^2$$

$$\overset{(4.11)}{\lesssim} \left(\|h\nabla\eta_\ell\|_{L^\infty(\Omega)}^2 + \|H\nabla\eta_\ell\|_{L^\infty(\Omega)}^2\right)\|\nabla w\|_{L^2(U_{k+2}(T)\setminus U_{k-\ell-2}(T))}^2$$

$$\overset{(4.11)}{\lesssim} \ell^{-2}\|\nabla w\|_{L^2(U_{k+2}(T)\setminus U_{k-\ell-2}(T))}^2.$$

This proves the assertion. $\qquad\square$

*Proof of Lemma* 4.9. The proof exploits some recursive Caccioppoli argument as in [34]. We fix some $T\in\mathcal{T}_H$ and $k\in\mathbb{N}$. Given $\ell\in\mathbb{N}$ with $\ell < k-1$, let $\eta_\ell := \eta_{T,k-2,\ell-4}\in V_H$ be some cutoff function as in (4.11). Lemma 4.10 shows that there exists some $\tilde{w}_T^i\in W_h$ such that $\|\nabla(\eta_\ell w_T^i - \tilde{w}_T^i)\|_{L^2(\Omega)}\lesssim \ell^{-1}\|\nabla w_T^i\|_{L^2(U_k(T)\setminus U_{k-\ell}(T))}$. Since $\tilde{w}_T^i\in\mathring{W}_h(\Omega\setminus U_{k-\ell+1}(T))$ and, hence, $\tilde{w}_T^i|_T = 0$, it holds that

$$(4.15)\qquad \int_{\Omega\setminus U_{k-\ell}(T)} A\nabla w_T^i\cdot\nabla\tilde{w}_T^i = \int_\Omega A\nabla w_T^i\cdot\nabla\tilde{w}_T^i = -\int_T Ae_i\cdot\nabla\tilde{w}_T^i = 0.$$

The definition of $\eta_\ell$, the product rule, (4.6), and (4.2) yield

$$\int_{\Omega\setminus U_k(T)} A\nabla w_T^i\cdot\nabla w_T^i \leq \int_{\Omega\setminus U_{k-\ell}(T)} \eta_\ell A\nabla w_T^i\cdot\nabla w_T^i$$

$$= \int_{\Omega\setminus U_{k-\ell}(T)} A\nabla w_T^i\cdot\left(\nabla(\eta_\ell w_T^i) - w_T^i\nabla\eta_\ell\right)$$

$$\overset{(4.15)}{=} \int_{\Omega\setminus U_{k-\ell}(T)} A\nabla w_T^i\cdot\left(\nabla(\eta_\ell w_T^i - \tilde{w}_T^i) - (w_T^i - \underbrace{I_H(w_T^i)}_{=0})\nabla\eta_\ell\right).$$

Observe that, by (4.11), $\|\nabla\eta_\ell\|_{L^\infty(K)} = |\nabla\eta_\ell(x_K)| \lesssim \ell^{-1}H_K^{-1}$ for all $K\in\mathcal{T}_H$. This and the estimate (4.4) for the interpolation error show that

$$\|(w_T^i - I_H(w_T^i))\nabla\eta_\ell\|_{L^2(K)}^2 \lesssim \|\nabla\eta_\ell\|_{L^\infty(K)}^2\|w_T^i - I_H(w_T^i)\|_{L^2(K)}^2$$

$$\lesssim H_K^2\|\nabla\eta_\ell\|_{L^\infty(K)}^2\|\nabla w_T^i\|_{L^2(\omega_K)}^2$$

$$\lesssim \ell^{-2}\|\nabla w_T^i\|_{L^2(\omega_K)}^2$$

for any $K\in\mathcal{T}_H$. The combination of the previous estimates and Cauchy–Schwarz inequalities proves that there is some constant $C_1 > 0$ independent of $T$, $\ell$, $k$, and the oscillations of $A$ such that

$$(4.16)\qquad \|A^{1/2}\nabla w_T^i\|_{L^2(\Omega\setminus U_k(T))} \leq C_1\ell^{-1/2}\|A^{1/2}\nabla w_T^i\|_{L^2(\Omega\setminus U_{k-\ell-1}(T))}.$$

The choice $\ell := \lceil C_1 e \rceil$ and the recursive application of (4.16) readily yield the assertion. $\square$

This exponential decay justifies the approximation of the correctors on local patches $U_k(T)$ as proposed in (4.10). We denote by $Q_h^k$ the corrector that corresponds to the choice $U(T) = U_k(T)$ in Oversampling Strategy 3 and by $Q_h^\Omega$ the one for $U(T) = \Omega$.

COROLLARY 4.11 (truncation/localization error). *Let $U(T) = \Omega$ for all $T \in \mathcal{T}_H$ in Oversampling Strategy 3. Then, for all $T \in \mathcal{T}_H$ and all $k \in \mathbb{N}$,*

$$\|A^{1/2}\nabla(w_T^i - w_T^{i,k})\|_{L^2(\Omega \setminus U_k(T))} \lesssim e^{-r \cdot k}\|A^{1/2}e_i\|_{L^2(T)},$$

*where $r > 0$ is as in Lemma 4.9 (independent of the variations of $A$ or the mesh size).*

*Proof.* Galerkin orthogonality yields

$$\|A^{1/2}\nabla(w_T^i - w_T^{i,k})\|_{L^2(\Omega)}^2 \le \|A^{1/2}\nabla(w_T^i - \tilde{w})\|_{L^2(U_{k-1}(T))}^2 + \|A^{1/2}\nabla w_T^i\|_{L^2(\Omega \setminus U_{k-1}(T))}^2,$$

where $\tilde{w} \in W_h$ is the fine scale function that corresponds to $(1-\eta_{T,k-1,1})w_T^i$ and which is constructed in the same way as $\tilde{w}$ in the proof of Lemma 4.10. Here, $\eta_{T,k-1,1}$ is some cutoff function as in (4.11). Since $\operatorname{supp}(\tilde{w}) \subset \operatorname{supp}((1-\eta_{T,k-1,1})w_T^i) \subset U_{k-1}(T)$, we have that $\tilde{w} \in \mathring{W}_h(U_k(T))$ and the use of Galerkin orthogonality is justified. Proceeding as in Lemma 4.10 shows that

$$\|A^{1/2}\nabla(w_T^i - w_T^{i,k})\|_{L^2(\Omega)}^2 \lesssim \|A^{1/2}\nabla w_T^i\|_{L^2(\Omega \setminus U_{k-2}(T))}^2,$$

and the application of Lemma 4.9 yields the assertion. $\square$

The proof of the main theorem requires one technical result.

LEMMA 4.12. *Let $k \in \mathbb{N}_{>0}$, and let $\Phi_H \in V_H$; then*

(4.17)

$$\left\|A^{1/2}\nabla(Q_h^\Omega - Q_h^k)\Phi_H\right\|_{L^2(\Omega)}^2 \lesssim k^d \sum_{T \in \mathcal{T}_H} \sum_{i=1}^d |\partial_{x_i}\Phi_H(x_T)|^2 \left\|A^{1/2}\nabla(w_T^i - w_T^{i,k})\right\|_{L^2(\Omega)}^2.$$

*Proof.* Let $\eta_{T,k,1}$ be as in (4.11), and define $z := (Q_h^\Omega - Q_h^k)\Phi_H \in W_h$. We decompose the error as follows:

$$\left\|A^{1/2}\nabla(Q_h^\Omega - Q_h^k)\Phi_H\right\|_{L^2(\Omega)}^2 = a(z,z)$$

$$= \underbrace{\sum_{T \in \mathcal{T}_H} \sum_{i=1}^d \partial_{x_i}\Phi_H(x_T)a(w_T^i - w_T^{i,k}, z(1 - \eta_{T,k,1}))}_{=:\mathrm{I}}$$

$$+ \underbrace{\sum_{T \in \mathcal{T}_H} \sum_{i=1}^d \partial_{x_i}\Phi_H(x_T)a(w_T^i - w_T^{i,k}, z\eta_{T,k,1})}_{=:\mathrm{II}}.$$

For the first term we get

$$|\mathrm{I}| \lesssim \sum_{T \in \mathcal{T}_H} \sum_{i=1}^d |\partial_{x_i}\Phi_H(x_T)|\|A^{\frac{1}{2}}\nabla(w_T^i - w_T^{i,k})\|_{L^2(\Omega)}\|\nabla(z(1 - \eta_{T,k,1}))\|_{L^2(U_{k+1}(T))},$$

where with $I_H(z) = 0$

$$\|\nabla\left(z(1-\eta_{T,k,1})\right)\|_{L^2(U_{k+1}(T))} \leq \|\nabla z\|_{L^2(U_{k+1}(T))} + \|z\nabla\left(1-\eta_{T,k,1}\right)\|_{L^2(U_{k+1}(T)\backslash U_k(T))}$$

$$\lesssim \|\nabla z\|_{L^2(U_{k+1}(T))} + \frac{1}{H}\|z - I_H(z)\|_{L^2(U_{k+1}(T)\backslash U_k(T))}$$

$$\lesssim \|\nabla z\|_{L^2(U_{k+2}(T))},$$

and therefore

$$|\mathrm{I}| \lesssim \sum_{T\in\mathcal{T}_H}\sum_{i=1}^d |\partial_{x_i}\Phi_H(x_T)|\|A^{\frac{1}{2}}\nabla(w_T^i - w_T^{i,k})\|_{L^2(\Omega)}\|\nabla z\|_{L^2(U_{k+2}(T))}$$

$$\lesssim k^{\frac{d}{2}}\left(\sum_{T\in\mathcal{T}_H}\sum_{i=1}^d |\partial_{x_i}\Phi_H(x_T)|^2\|A^{\frac{1}{2}}\nabla(w_T^i - w_T^{i,k})\|_{L^2(\Omega)}^2\right)^{\frac{1}{2}}\|\nabla z\|_{H^1(\Omega)}.$$

To estimate the second term, we use Lemma 4.10, which gives us the existence of some $\tilde{z} \in \mathring{W}_h(\Omega \setminus U_{k-2}(T))$ with $a(w_T^i - w_T^{i,k}, \tilde{z}) = 0$ (as in (4.15)) and the property $\|\nabla(z\eta_{T,k,1} - \tilde{z})\|_{L^2(\Omega)} \lesssim \|\nabla z\|_{L^2(U_{k+2}(T))}$. This yields

$$|\mathrm{II}| = \left|\sum_{T\in\mathcal{T}_H}\sum_{i=1}^d \partial_{x_i}\Phi_H(x_T)a(w_T^i - w_T^{i,k}, z\eta_{T,k,1} - \tilde{z})\right|$$

$$\leq \sum_{T\in\mathcal{T}_H}\sum_{i=1}^d |\partial_{x_i}\Phi_H(x_T)|\|A^{\frac{1}{2}}\nabla(w_T^i - w_T^{i,k})\|_{L^2(\Omega)}\|\nabla z\|_{L^2(U_{k+2}(T))}$$

$$\lesssim k^{\frac{d}{2}}\left(\sum_{T\in\mathcal{T}_H}\sum_{i=1}^d |\partial_{x_i}\Phi_H(x_T)|^2\|A^{\frac{1}{2}}\nabla(w_T^i - w_T^{i,k})\|_{L^2(\Omega)}^2\right)^{\frac{1}{2}}\|\nabla z\|_{H^1(\Omega)}.$$

Combining the estimates for I and II and dividing by $\|\nabla z\|_{H^1(\Omega)} \lesssim a(z,z)^{\frac{1}{2}}$ yields the assertion.    $\square$

THEOREM 4.13 ($H^1$–error estimate).  *Given $k \in \mathbb{N}$, let $U(T) = U_k(T)$ for all $T \in \mathcal{T}_H$ in Oversampling Strategy 3. Then the multiscale approximation $u_H^k$ that solves (4.7) satisfies the error estimate*

$$\|\nabla u_h - \nabla(u_H^k + Q_h^k(u_H^k))\| \lesssim \gamma_{\min}^{-1}\|Hf\|_{L^2(\Omega)} + k^{\frac{d}{2}}e^{-rk}\|f\|_{H^{-1}(\Omega)},$$

*where $u_h$ is the reference solution from (2.3) and $r > 0$ as in Lemma 4.9.*

*Remark* 4.14 (relation to the results in [34]).  In the case of maximal oversampling, the new MsFEM with constrained oversampling coincides with the ideal version (without localization) of the VMM presented in [34]. The localized versions are different and allow similar, but not identical, error estimates. The upper bound obtained in [34] reads (up to some multiplicative constant) as

$$\|Hf\|_{L^2(\Omega)} + H^{-1}e^{-rk}\|f\|_{H^{-1}(\Omega)}.$$

Our new localization strategy allows for an improved estimate in the sense that the unpleasant factor $H^{-1}$ does not appear. Note that the proof of the error estimate in Theorem 4.13 does not generalize to the localization strategy used in [34] and must therefore be seen independently. The reason is that the structure of the local problems

(4.6) gives us a nice summation property which we were able to exploit but which is not available in [34]. This observation indicates that better numerical approximations for equal sizes of oversampling patches are possible with our new approach. This will be investigated in future works.

*Proof of Theorem* 4.13. Using the fact that Galerkin approximation minimizes the error in the energy norm, we obtain with the definitions of $u_H^k$ and $u_h$ that for all $\Phi_H \in V_H$

(4.18)
$$\|A^{1/2}(\nabla u_h - \nabla(u_H^k + Q_h^k(u_H^k)))\|_{L^2(\Omega)} \leq \|A^{1/2}(\nabla u_h - \nabla(\Phi_H + Q_h^k(\Phi_H)))\|_{L^2(\Omega)}.$$

Let $u_H$ be the solution of (4.7) with the ideal corrector $Q_h = Q_h^\Omega$. Then

$$\|A^{1/2}(\nabla u_h - \nabla(u_H^k + Q_h^k(u_H^k)))\|_{L^2(\Omega)}$$

$$\overset{(4.18)}{\leq} \|A^{1/2}(\nabla u_h - \nabla(u_H + Q_h^k(u_H)))\|_{L^2(\Omega)}$$

$$\leq \|A^{1/2}(\nabla u_h - \nabla(u_H + Q_h^\Omega(u_H)))\|_{L^2(\Omega)}$$

$$\quad + \|A^{1/2}(\nabla(u_H + Q_h^\Omega(u_H)) - \nabla(u_H + Q_h^k(u_H)))\|_{L^2(\Omega)}$$

$$\lesssim \gamma_{\min}^{-1/2}\|Hf\|_{L^2(\Omega)} + \|A^{1/2}\nabla((Q_h^\Omega - Q_h^k)(u_H))\|_{L^2(\Omega)}.$$

By Corollary 4.11, we get

$$\|A^{1/2}\nabla((Q_h^\Omega - Q_h^k)(u_H))\|_{L^2(\Omega)}^2$$

$$= \left\|\sum_{T \in \mathcal{T}_H}\sum_{i=1}^d \partial_{x_i}u_H(x_T)A^{\frac{1}{2}}\nabla(w_T^i - w_T^{i,k})\right\|_{L^2(\Omega)}^2$$

$$\overset{(4.17)}{\lesssim} k^d \sum_{T \in \mathcal{T}_H}\sum_{i=1}^d |\partial_{x_i}u_H(x_T)|^2 \left\|A^{\frac{1}{2}}\nabla(w_T^i - w_T^{i,k})\right\|_{L^2(\Omega)}^2$$

$$\lesssim k^d e^{-2r\cdot k}\sum_{T \in \mathcal{T}_H}\sum_{i=1}^d |\partial_{x_i}u_H(x_T)|^2 \|A^{1/2}e_i\|_{L^2(T)}^2$$

$$\lesssim k^d e^{-2r\cdot k}\sum_{T \in \mathcal{T}_H}\sum_{i=1}^d \|A^{1/2}\nabla u_H\|_{L^2(T)}^2$$

$$\lesssim k^d e^{-2rk}\|f\|_{H^{-1}(\Omega)}^2.$$

In the last step we have used that $u_H = I_H(u_H + Q_h^\Omega(u_H))$, the stability of $I_H$, and the energy estimate $\|A^{1/2}\nabla(u_H + Q_h^\Omega(u_H)\|_{L^2(T)} \lesssim \gamma_{\min}^{-1/2}\|f\|_{H^{-1}(\Omega)}$. □

THEOREM 4.15 ($L^2$-*estimates*). *Given* $k \in \mathbb{N}$, *let* $U(T) = U_k(T)$ *for all* $T \in \mathcal{T}_H$ *in Oversampling Strategy* 3. *Then the multiscale approximation* $u_H^k$ *that solves* (4.7) *satisfies the error estimates*

$$\|u_h - (u_H^k + Q_h^k(u_H^k))\|_{L^2(\Omega)} \lesssim (\gamma_{\min}^{-1}\|H\|_{L^\infty(\Omega)} + k^{d/2}e^{-rk})^2\|f\|_{L^2(\Omega)}$$

*and*

$$\|u_h - u_H^k\|_{L^2(\Omega)} \lesssim \min_{v_H \in V_H}\|u_h - v_H\|_{L^2(\Omega)} + (\gamma_{\min}^{-1}\|H\|_{L^\infty(\Omega)} + k^{d/2}e^{-rk})^2\|f\|_{L^2(\Omega)},$$

*where* $u_h$ *is the reference solution from* (2.3) *and* $r$ *is a positive constant.*

*Proof.* A standard Aubin–Nitsche duality argument yields the first estimate. The second estimate follows from the first one and the quasi optimality and stability of the interpolation $I_H$ in $L^2(\Omega)$.    □

*Remark* 4.16 (smooth coefficient with known smallest scale $\varepsilon$). Let $\Omega$ be convex, and let $f \in L^2(\Omega)$ with $\|f\|_{L^2(\Omega)} \lesssim 1$, $A \in W^{1,\infty}(\Omega)$ with $\|\nabla A\|_{L^\infty(\Omega)} \lesssim \varepsilon^{-1}$ with some small scale parameter $\varepsilon > 0$. Choose uniform meshes $\mathcal{T}_H$ and $\mathcal{T}_h$ with $H \gtrsim \varepsilon \gtrsim h$. Under these assumptions, the error of the reference solution $u_h \in V_h$ is bounded as follows:

$$\|\nabla(u - u_h)\| \lesssim h\varepsilon^{-1}.$$

We refer the reader to [37] for details. If $k \gtrsim \log(H^{-1})$, Theorems 4.13 and 4.15 yield the error bounds

$$\|\nabla(u - u_H^k - Q_h^k(u_H^k))\|_{L^2(\Omega)} \lesssim H + \tfrac{h}{\varepsilon},$$
$$\|u - u_H^k - Q_h^k(u_H^k)\|_{L^2(\Omega)} \lesssim H^2 + \left(\tfrac{h}{\varepsilon}\right)^2,$$
$$\|u - u_H^k\|_{L^2(\Omega)} \lesssim H + \left(\tfrac{h}{\varepsilon}\right)^2.$$

**5. Numerical experiments.** In this section we present numerical experiments to confirm the derived error estimates and to compare the numerical accuracies of Oversampling Strategies 1, 2, and 3. Here we use Oversampling Strategies 1 and 2 in the PG formulation (due to the findings in [28]) and Oversampling Strategy 3 in the symmetric formulation. We consider the following model problem.

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

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

which is the exact solution of the problem

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

where $A$ is given by

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

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 5.1 we depict the results for $h = 2^{-6}$ and various combinations of $H$ with different numbers of oversampling layers. For a better illustration we state the number of fine grid layers and the number of coarse grid layers ($k$) that corresponds with that. The results in Table 5.1 match nicely with the analytically predicted behavior. In Table 5.2 we state a comparison between the $L^2$- and $H^1$-errors for the three oversampling strategies obtained for identical values of $H$, $h$, and $\mathcal{U}$. We observe that our oversampling strategy, in contrast to the classical ones, does not suffer from a loss in accuracy when $H$ is close to the microscopic parameter $\varepsilon$. Moreover, the accuracy obtained for Oversampling Strategy 3 is very promising in general.

TABLE 5.1
*Computations made for $h = 2^{-6}$. $k$ denotes the number of coarse layers. $u_h$ denotes the fine scale reference given by (2.3), and $u_H^{3,\text{MsFEM}}$ denotes the MsFEM approximation obtained with Oversampling Strategy 3. The table depicts various errors between $u_h$ and $u_H^{3,\text{MsFEM}}$.*

| $H$ | Fine layers | k | $\|u_h - u_H^{3,\text{MsFEM}}\|_{L^2(\Omega)}$ | $\|u_h - u_H^{3,\text{MsFEM}}\|_{H^1(\Omega)}$ |
|---|---|---|---|---|
| $2^{-1}$ | 16 | 0.5 | 0.490063 | 4.49575 |
| $2^{-2}$ | 8 | 0.5 | 0.09491 | 1.66315 |
| $2^{-2}$ | 24 | 1.5 | 0.06376 | 1.08960 |
| $2^{-3}$ | 4 | 0.5 | 0.033691 | 1.017150 |
| $2^{-3}$ | 8 | 1 | 0.007125 | 0.406317 |
| $2^{-3}$ | 12 | 1.5 | 0.007115 | 0.331458 |
| $2^{-3}$ | 16 | 2 | 0.003241 | 0.165703 |
| $2^{-4}$ | 2 | 0.5 | 0.012808 | 0.655269 |
| $2^{-4}$ | 4 | 1 | 0.004164 | 0.348814 |
| $2^{-4}$ | 6 | 1.5 | 0.004029 | 0.329306 |
| $2^{-4}$ | 8 | 2 | 0.001451 | 0.162747 |
| $2^{-4}$ | 12 | 2.5 | 0.000850 | 0.114040 |
| $2^{-4}$ | 16 | 3 | 0.000696 | 0.096378 |

TABLE 5.2
*Computations made for $h = 2^{-6}$. $k$ denotes the number of coarse layers. $u_h$ denotes the fine scale reference given by (2.3), and $u_H^{i,\text{MsFEM}}$ denotes the MsFEM approximation obtained with Oversampling Strategy 1. The error is denoted by $e_i := u_h - u_H^{i,\text{MsFEM}}$. The second column depicts the number of fine grid layers.*

| $H$ | $k$ | Oversampling Strategy 1 | | Oversampling Strategy 2 | | Oversampling Strategy 3 | |
|---|---|---|---|---|---|---|---|
| | | $\|e_1\|_{L^2}$ | $\|e_1\|_{H^1}$ | $\|e_2\|_{L^2}$ | $\|e_2\|_{H^1}$ | $\|e_3\|_{L^2}$ | $\|e_3\|_{H^1}$ |
| $2^{-2}$ | 1 | 0.1399 | 1.9812 | 0.1399 | 1.9812 | 0.0638 | 1.0896 |
| $2^{-3}$ | 1 | 0.0594 | 1.6250 | 0.0594 | 1.6250 | 0.0071 | 0.4063 |
| $2^{-3}$ | 2 | 0.0593 | 1.6250 | 0.0593 | 1.6250 | 0.0032 | 0.1657 |
| $2^{-4}$ | 1 | 0.0166 | 0.8067 | 0.0172 | 0.8048 | 0.0042 | 0.3488 |
| $2^{-4}$ | 2 | 0.0160 | 0.8057 | 0.0168 | 0.7955 | 0.0015 | 0.1628 |
| $2^{-4}$ | 3 | 0.0153 | 0.8016 | 0.0152 | 0.7937 | 0.0007 | 0.0964 |

**6. Conclusion.** In this work, we proposed a new oversampling strategy for the MsFEM, which generalizes the original method without oversampling. The new strategy is based on an additional constraint for the solution spaces of the local problems. The error analysis shows that oversampling layers of thickness $H \log(H^{-1})$ suffice to preserve the common convergence rates with respect to $H$ without any preasymptotic effects. Moreover, this choice prevents resonance errors even for general $L^\infty$ coefficients without any assumptions on the geometry of the microstructure or the regularity of $A$. In this respect, the method is reliable. The method is also efficient in the sense that structural knowledge about the coefficient, e.g., (local) periodicity or scale separation, may be exploited to reduce the number of corrector problems considerably. Whether the oversampling can be reduced to very small layers in the case of, e.g., periodicity should be investigated numerically and/or analytically in future works.

## REFERENCES

[1] A. ABDULLE, *The finite element heterogeneous multiscale method: A computational strategy for multiscale PDEs*, in Multiple Scales Problems in Biomathematics, Mechanics, Physics and Numerics, GAKUTO Internat. Ser. Math. Sci. Appl. 31, Gakkōtosho, Tokyo, 2009, pp. 133–181.

[2] I. BABUSKA AND R. LIPTON, *Optimal local approximation spaces for generalized finite element methods with application to multiscale problems*, Multiscale Model. Simul., 9 (2011), pp. 373–406.

[3] F. BREZZI, L. P. FRANCA, T. J. R. HUGHES, AND A. RUSSO, $b = \int g$, Comput. Methods Appl. Mech. Engrg., 145 (1997), pp. 329–339.

[4] C. CARSTENSEN AND R. VERFÜRTH, *Edge residuals dominate a posteriori error estimates for low order finite element methods*, SIAM J. Numer. Anal., 36 (1999), pp. 1571–1587.

[5] Z. CHEN AND T. Y. HOU, *A mixed multiscale finite element method for elliptic problems with oscillating coefficients*, Math. Comp., 72 (2003), pp. 541–576.

[6] 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.

[7] C.-C. CHU, I. G. GRAHAM, AND T.-Y. HOU, *A new multiscale finite element method for high-contrast elliptic interface problems*, Math. Comp., 79 (2010), pp. 1915–1955.

[8] W. DENG, X. YUN, AND C. XIE, *Convergence analysis of the multiscale method for a class of convection-diffusion equations with highly oscillating coefficients*, Appl. Numer. Math., 59 (2009), pp. 1549–1567.

[9] W. E AND B. ENGQUIST, *The heterogeneous multiscale methods*, Commun. Math. Sci., 1 (2003), pp. 87–132.

[10] W. E AND B. ENGQUIST, *Multiscale modeling and computation*, Notices Amer. Math. Soc., 50 (2003), pp. 1062–1070.

[11] 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.

[12] Y. EFENDIEV AND T. HOU, *Multiscale finite element methods for porous media flows and their applications*, Appl. Numer. Math., 57 (2007), pp. 577–596.

[13] Y. EFENDIEV, T. HOU, AND V. GINTING, *Multiscale finite element methods for nonlinear problems and their applications*, Commun. Math. Sci., 2 (2004), pp. 553–589.

[14] Y. EFENDIEV AND T. Y. HOU, *Multiscale Finite Element Methods: Theory and Applications*, Surv. Tutor. Appl. Math. Sci. 4, Springer, New York, 2009.

[15] Y. EFENDIEV AND A. PANKOV, *Numerical homogenization of monotone elliptic operators*, Multiscale Model. Simul., 2 (2003), pp. 62–79.

[16] 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.

[17] A. GLORIA, *An analytical framework for the numerical homogenization of monotone elliptic operators and quasiconvex energies*, Multiscale Model. Simul., 5 (2006), pp. 996–1043.

[18] A. GLORIA, *An analytical framework for numerical homogenization. Part II: Windowing and oversampling*, Multiscale Model. Simul., 7 (2008), pp. 274–293.

[19] A. GLORIA, *Reduction of the resonance error—Part 1: Approximation of homogenized coefficients*, Math. Models Methods Appl. Sci., 21 (2011), pp. 1601–1630.

[20] A. GLORIA, *Numerical homogenization: Survey, new results, and perspectives*, in Mathematical and Numerical Approaches for Multiscale Problem, ESAIM Proc. 37, EDP Sci., Les Ulis, 2012, pp. 50–116.

[21] P. HENNING, *Convergence of MsFEM approximations for elliptic, non-periodic homogenization problems*, Netw. Heterog. Media, 7 (2012), pp. 503–524.

[22] 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.

[23] P. HENNING AND M. OHLBERGER, *Error control and adaptivity for heterogeneous multiscale*

*approximations of nonlinear monotone problems*, Discrete Contin. Dyn. Syst. Ser. S, to appear.

[24] 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.

[25] P. Henning, M. Ohlberger, and B. Schweizer, *An Adaptive Multiscale Finite Element Method*, Preprint 05/12 - N, University of Münster, Münster, Germany, 2012.

[26] 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.

[27] 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.

[28] 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.

[29] 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.

[30] 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.

[31] V. V. Jikov, S. M. Kozlov, and O. A. Oleĭnik, *Homogenization of Differential Operators and Integral Functionals*, Springer-Verlag, Berlin, 1994.

[32] 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.

[33] A. Målqvist, *Multiscale methods for elliptic problems*, Multiscale Model. Simul., 9 (2011), pp. 1064–1086.

[34] A. Målqvist and D. Peterseim, *Localization of elliptic multiscale problems*, Math. Comp., to appear.

[35] H. Owhadi and L. Zhang, *Metric-based upscaling*, Comm. Pure Appl. Math., 60 (2007), pp. 675–723.

[36] H. Owhadi and L. Zhang, *Localized bases for finite-dimensional homogenization approximations with nonseparated scales and high contrast*, Multiscale Model. Simul., 9 (2011), pp. 1373–1398.

[37] D. Peterseim and S. Sauter, *Finite elements for elliptic problems with highly varying, nonperiodic diffusion matrix*, Multiscale Model. Simul., 10 (2012), pp. 665–695.

[38] H. Yserentant, *On the multilevel splitting of finite element spaces*, Numer. Math., 49 (1986), pp. 379–412.