AN EFFICIENT HIGH ORDER HETEROGENEOUS MULTISCALE METHOD FOR ELLIPTIC PROBLEMS

RUO LI, PINGBING MING, AND FENGYANG TANG

Abstract. We propose an efficient heterogeneous multiscale finite element method based on a local least-squares reconstruction of the effective matrix using the data retrieved from the solution of cell problems posed on the vertices of the triangulation. The method achieves high order accuracy for high order macroscopic solver with essentially the same cost as the linear macroscopic solver. Optimal error bounds are proved for the elliptic problem. Numerical results demonstrate that the new method significantly reduces the cost without loss of accuracy.

1. Introduction

We consider the classical elliptic problem

$$\begin{cases} -\operatorname{div}\left(a^\varepsilon(\boldsymbol{x})\nabla u^\varepsilon(\boldsymbol{x})\right) = f(\boldsymbol{x}), & \boldsymbol{x}\in D\subset\mathbb{R}^d,\\ u^\varepsilon(\boldsymbol{x}) = 0, & \boldsymbol{x}\in\partial D, \end{cases} \tag{1.1}$$

where ε is a small parameter that signifies explicitly the multiscale nature of the coefficient a^ε, which is not necessarily symmetric. We assume $a^\varepsilon\in\mathcal{M}(\alpha,\beta,D)$ that is defined as

$$\mathcal{M}(\alpha,\beta,D) = \Big\{\mathcal{B}\in[L^\infty(D)]^{d^2} \;\Big|\; (\mathcal{B}(\boldsymbol{x})\boldsymbol{\xi},\boldsymbol{\xi})\geq\alpha|\boldsymbol{\xi}|^2, |\mathcal{B}(\boldsymbol{x})\boldsymbol{\xi}|\leq\beta|\boldsymbol{\xi}| \\ \text{for any } \boldsymbol{\xi}\in\mathbb{R}^d \text{ and } a.e., \boldsymbol{x} \text{ in } D\Big\},$$

where $(\cdot,\cdot)$ denotes the inner product on $\mathbb{R}^d$, while $|\cdot|$ is the corresponding norm, and D is a bounded domain in $\mathbb{R}^d$.

Date: November 8, 2011.

2000 *Mathematics Subject Classification.* 65N12, 65N30, 74Q05, 74Q15.

Key words and phrases. Heterogeneous multiscale method, H-convergence, Least-squares reconstruction.

This work of Li was partially supported by Fok Ying Tong Education Foundation and NCET in China. The work of Ming was partially supported by National Natural Science Foundation of China grants 10871197, 10932011, and by the funds from Creative Research Groups of China through grant 11021101, and by the support of CAS National Center for Mathematics and Interdisciplinary Sciences.

We are very grateful to the referees for many thoughtful suggestions which help to improve the paper.

On the analytic side, the following fact is known about (1.1). In the sense of H-convergence (see [36]), for every $a^\varepsilon \in \mathcal{M}(\alpha, \beta, D)$ and $f \in H^{-1}(D)$, the sequence $\{u^\varepsilon\}$ of the solutions of (1.1) satisfies, in the sense of substraction of a subsequence,

$$\begin{aligned} u^\varepsilon &\rightharpoonup U_0 && \text{weakly in} \quad H_0^1(D), \\ a^\varepsilon \nabla u^\varepsilon &\rightharpoonup \mathcal{A}\nabla U_0 && \text{weakly in} \quad [L^2(D)]^d, \end{aligned}$$

where U_0 is the solution of

$$\begin{cases} -\operatorname{div}(\mathcal{A}(\boldsymbol{x})\nabla U_0(\boldsymbol{x})) = f(\boldsymbol{x}), & \boldsymbol{x} \in D, \\ U_0(\boldsymbol{x}) = 0, & \boldsymbol{x} \in \partial D, \end{cases} \tag{1.2}$$

and $\mathcal{A} \in \mathcal{M}(\alpha, \beta^2/\alpha, D)$. Here $H_0^1(D), L^2(D)$ and $H^{-1}(D)$ are standard Sobolev spaces [5], and we denote the $L^2(D)$ inner product by $(\cdot, \cdot)$.

The heterogeneous multiscale method (HMM for short) introduced by E AND ENGQUIST [16] is a general methodology for efficient computation of multiscale problems. It consists of two components: selection of a macroscopic solver, and estimating the missing macroscale data by solving the microscale problem locally. The choice of the macroscopic solver depends on the nature of the problem. Finite element method is often used as the macroscopic solver for Problem (1.1) due to its variational structure (HMM-FEM for short).

The missing data in HMM-FEM is the effective matrix evaluated at the quadrature nodes, which is obtained through solving the cell problems posed on the quadrature nodes. The cost of HMM-FEM mainly comes from solving the cell problems, and the cost increases dramatically when one were to employ the higher-order macroscopic solver since the number of cell problems grows rapidly. DU AND MING [15] proposed a new quadrature rule for the linear element and the quadratic element that preferably makes use of element vertices or element edge centers as the quadrature nodes, which seemingly contradicts with the criterion of *a good quadrature formula* [33] that uses the points lying within the element as the quadrature nodes. However, compared to the original HMM-FEM, the method based on such nonconventional quadrature rule has smaller cost without loss of accuracy, because the effective matrix evaluated at the element vertices and element edge centers can be shared by more than one element, and therefore, less cell problems need to be solved. This idea has been extended to solve three-dimensional problem by WANG [39] and the numerical tests confirm the efficiency of such idea for quadratic and cubic macroscopic solvers. Unfortunately, the gain of such method for even higher-order macroscopic solver is less pronounced because the quadrature nodes for higher-order element get to accumulate inside the elements [38]. The following question arises: Can we design a better high order HMM-FEM? The new method should retain high order accuracy with relatively low cost.

In this paper, we introduce a new method that employs a recovering operation to retrieve the effective matrix from suitable sampling points. Given a macroscopic

solver of degree k, we fit a polynomial of degree m to values of the entry of the effective matrix at some sampling points by a local discrete least-squares method. Here m is determined by k that will be made clear in §4. For a typical mesh, solving the cell problems on the vertices is often the most economical one among all possible alternatives. Therefore, only the cell problems posed at the vertices need to be solved regardless of the order of the macroscopic solver. Both theoretical results and the numerical tests show that the method converges with optimal order while the cost is essentially the same as HMM-FEM with a linear macroscopic solver. The underlying assumption of this method is that the effective matrix is smooth to certain degree, which may not be true for general problems, however, this assumption may be valid for certain practical cases, e.g., we concern with the macroscopic response of the composite materials. It is interesting to note that the prescribed sampling points are not necessarily related to the triangulation in use, it may be any scattering points on the whole domain, namely, it can be meshless. Such idea will be elaborated in our future work.

The idea of our method is related to Zienkiewicz-Zhu (ZZ) gradient patch recovery [43], and polynomial patch recovery by Zhang and Naga [42] in which the superconvergence properties of the gradient information is the main concern. A similar idea has also been used in ENO/WENO method, in which a high order polynomial is reconstructed from the cell mean of the stencil; see [37].

There are also other types high order multiscale methods. We just name a few. Based on the idea of multiscale finite element method [25], Allaire and Brizzi [6] proposed a high order numerical homogenization method; a high order residual-free bubbles is introduced in [10]; and some high order generalized finite element methods have been reviewed in [7].

The convergence behavior of HMM-FEM applied to Problem (1.1) is by now well understood (see [17, 1, 2, 15, 14]). Discretization error of the cell problems are mainly limited to a special periodic boundary condition and the Dirichlet boundary condition, while leaving the cell problem with other boundary conditions open [24, 41]. In this paper, we shall give a unified analysis for discretization error of the microscopic problems supplemented with the Dirichlet, the Neumann or the periodic boundary condition when a^ε is a locally periodic matrix. Our result holds true under realistic regularity assumption on the solutions of the cell problems. The proof relies on an interpolant arising from the homogenization theory [8].

The rest of the paper is organized as follows. In the next section, we introduce a new method that is based on a local discrete least-squares reconstruction. The accuracy of the proposed method is analyzed in §3. Numerical examples are reported in §4. We draw the conclusion in the last section.

2. Algorithms

The macroscopic solver is chosen as the standard $\mathbb{P}_k$ Lagrange element, which is defined as the set of polynomials with degree less than k for the sum of all variables. The finite element space is denoted by $\mathcal{V}_H$ corresponding to the triangulation $\mathcal{T}_H$ with mesh size H that is the maximum of the element size H_K for all elements $K \in \mathcal{T}_H$. Here H_K is the diameter of K. The mesh is assumed to be shape-regular in the sense of Ciarlet-Raviart [11]. The HMM solution $U_H \in \mathcal{V}_H$ satisfies

$$a_H(U_H, V) = (f, V) \quad \text{for all } V \in \mathcal{V}_H, \tag{2.1}$$

where the bilinear form a_H is defined for any $V, W \in \mathcal{V}_H$ by

$$a_H(V, W) = \sum_{K \in \mathcal{T}_H} \int_K \nabla W(\boldsymbol{x}) \cdot \mathcal{A}_H(\boldsymbol{x}) \nabla V(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x},$$

where $\mathcal{A}_H$ is reconstructed on each element K as follows. For $i, j = 1, \ldots, d$, the entry $(\mathcal{A}_H)_{ij}$ is defined as the solution of a discrete least-squares problem:

$$(\mathcal{A}_H)_{ij} = \arg \min_{p \in \mathbb{P}_m\left(S(K)\right)} \sum_{\boldsymbol{x}_\ell \in \mathcal{I}_K} \left| \left(\widetilde{\mathcal{A}}_H(\boldsymbol{x}_\ell)\right)_{ij} - p(\boldsymbol{x}_\ell) \right|^2, \tag{2.2}$$

where $\mathcal{I}_K$ is the nodal set of all elements that belong to $S(K)$, and $S(K)$ is a convex patch of elements around K (including K). Its precise definition will be given in the next section. We refer to Fig. 1 for an example of $S(K)$. At each vertex $\boldsymbol{x}_\ell$, the effective matrix $\widetilde{\mathcal{A}}_H(\boldsymbol{x}_\ell)$ is defined by

$$\widetilde{\mathcal{A}}_H(\boldsymbol{x}_\ell) \left\langle \nabla v_h^\varepsilon \right\rangle_{I_\delta} \equiv \left\langle a^\varepsilon \nabla v_h^\varepsilon \right\rangle_{I_\delta}, \tag{2.3}$$

where the cell $I_\delta(\boldsymbol{x}_\ell) \equiv \boldsymbol{x}_\ell + \delta Y$ with $Y \equiv (-1/2, 1/2)^d$, and δ is the cell size. We use $\langle \cdot \rangle_{I_\delta}$ to denote the integral mean over $I_\delta(\boldsymbol{x}_\ell)$. Here $v_h^\varepsilon - V_\ell \in \mathcal{V}_h$ satisfies

$$\left(a^\varepsilon \nabla v_h^\varepsilon, \nabla \varphi\right)_{L^2(I_\delta)} = 0 \quad \text{for all } \varphi \in \mathcal{V}_h, \tag{2.4}$$

where $V_\ell \equiv V(\boldsymbol{x}_\ell) + (\boldsymbol{x} - \boldsymbol{x}_\ell) \cdot \nabla V(\boldsymbol{x}_\ell)$ is the linear approximation of V at $\boldsymbol{x}_\ell$. We call (2.4) the Dirichlet cell problem if

$$\mathcal{V}_h = \mathcal{V}_{D,h} \equiv \left\{ v \in H_0^1\left(I_\delta(\boldsymbol{x}_\ell)\right) \mid v|_K \in \mathbb{P}_{k'}(K), \quad K \in \mathcal{T}_h \right\}.$$

We call (2.4) the Neumann cell problem if

$$\mathcal{V}_h = \mathcal{V}_{N,h} \equiv \left\{ v \in H^1\left(I_\delta(\boldsymbol{x}_\ell)\right) \mid v|_K \in \mathbb{P}_{k'}(K), \langle \nabla v \rangle_{I_\delta} = 0 \;\; K \in \mathcal{T}_h \right\}.$$

We call (2.4) the periodic cell problem if

$$\mathcal{V}_h = \mathcal{V}_{P,h} \equiv \left\{ v \in H_\sharp^1\left(I_\delta(\boldsymbol{x}_\ell)\right) \mid v|_K \in \mathbb{P}_{k'}(K), \langle v \rangle_{I_\delta} = 0 \;\; K \in \mathcal{T}_h \right\},$$

where $\mathcal{T}_h$ is the triangulation of $I_\delta(\boldsymbol{x}_\ell)$ with mesh size h and $k' \in \mathbb{N}$. $H_\sharp^1\left(I_\delta(\boldsymbol{x}_\ell)\right)$ is the closure of $C_\sharp^\infty$ for the H^1 norm, and $C_\sharp^\infty$ is the subset of $C^\infty(I_\delta(\boldsymbol{x}_\ell))$ of $I_\delta-$periodic function. We shall deal with all these three cell problems and refer to [41] for the implementation details.

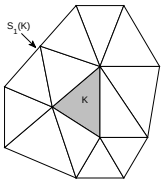


FIG. 1. Example of the element patch $S(K)$ and the nodal set $\mathcal{I}_K$.

An alternative definition of the effective matrix is to solve the following constrained discrete least-squares problem: for any $m \in \mathbb{N}$,

$$(\mathcal{A}_H)_{ij} = \arg \min_{p \in \mathbb{P}_m(S(K))} \sum_{\boldsymbol{x}_\ell \in \mathcal{I}_K} \left| \left(\widetilde{\mathcal{A}}_H(\boldsymbol{x}_\ell)\right)_{ij} - p(\boldsymbol{x}_\ell) \right|^2 \tag{2.5}$$

subject to the constraints

$$p(\boldsymbol{x}_\ell) = \left(\widetilde{\mathcal{A}}_H\right)_{ij}(\boldsymbol{x}_\ell) \quad \text{for all the vertices } \boldsymbol{x}_\ell \text{ of } K.$$

It will be shown in §4 that the method based on the constrained reconstruction has the same convergence order with the one without constraints, while it is numerically more accurate.

Remark. We can also define $\mathcal{A}_H$ as the solution of a weighted least-squares problem with a suitable weight, which may be more efficient in certain case. We shall leave it for further study.

3. Convergence

In this section, we analyze the proposed method with the least-squares reconstruction (2.2) or (2.5). In what follows, we assume that $\mathcal{A}$ is smooth and the domain D is a convex polytope, and define

$$e(\text{HMM}) = \max_{\substack{\boldsymbol{x} \in K \\ K \in \mathcal{T}_H}} \left\| (\mathcal{A} - \mathcal{A}_H)(\boldsymbol{x}) \right\|_F,$$

where $\left\| \cdot \right\|_F$ is the Euclidean norm.

Lemma 3.1. *If $e(HMM) < \alpha$, then for all $V, W \in \mathcal{V}_H$, there holds*

$$\begin{aligned} a_H(V,V) &\geq \big(\alpha - e(HMM)\big) \left\| \nabla V \right\|^2_{L^2(D)}, \\ |a_H(V,W)| &\leq (\beta^2/\alpha + \alpha) \left\| \nabla V \right\|_{L^2(D)} \left\| \nabla W \right\|_{L^2(D)}. \end{aligned} \tag{3.1}$$

Proof. By the ellipticity of the effective matrix $\mathcal{A}$ and the definition of $e(\mathrm{HMM})$, we obtain

$$\begin{aligned} a_H(V,V) &= \int_D \nabla V\cdot\mathcal{A}(\boldsymbol{x})\nabla V\,\mathrm{d}\boldsymbol{x} + \sum_{K\in\mathcal{T}_H}\int_K \nabla V\cdot(\mathcal{A}_H-\mathcal{A})(\boldsymbol{x})\nabla V\,\mathrm{d}\boldsymbol{x}\\ &\geq \big(\alpha - e(\mathrm{HMM})\big)\big\|\nabla V\big\|^2_{L^2(D)}. \end{aligned}$$

This gives the lower bound $(3.1)_1$. The upper bound can be obtained similarly by noting $\mathcal{A}\in\mathcal{M}(\alpha,\beta^2/\alpha,D)$ and the condition $e(\mathrm{HMM})<\alpha$. □

The above lemma gives the existence and uniqueness of the HMM-FEM solution (2.1). The following error estimate is based on Lemma 3.1 and a theorem of Berger, Scott and Strang [9, Theorem I, equation (11)] and, except for the explicit constants in the estimate (3.2), can be found in [17, Theorem 1.1].

Lemma 3.2. *Let U_0 and U_H be the solutions of* (1.2) *and* (2.1)*, respectively. If $e(HMM)<\alpha/2$, then,*

$$\big\|\nabla(U_0-U_H)\big\|_{L^2(D)} \leq \frac{\beta}{\alpha}\inf_{V\in\mathcal{V}_H}\big\|\nabla(U_0-V)\big\|_{L^2(D)} + \frac{2c_p}{\alpha^2}\|f\|_{H^{-1}(D)}e(HMM), \tag{3.2}$$

where c_p is the constant in the following discrete Poincaré's inequality:

$$\|V\|_{H^1(D)}\leq c_p\big\|\nabla V\big\|_{L^2(D)} \qquad \text{for all } V\in\mathcal{V}_H.$$

If in additional $f\in L^2(D)$ so that $U_0\in H^2(D)$, then there exists C such that

$$\|U_0-U_H\|_{L^2(D)}\leq C\Big(H\inf_{V\in\mathcal{V}_H}\big\|\nabla(U_0-V)\big\|_{L^2(D)} + e(HMM)\Big). \tag{3.3}$$

3.1. **Properties of the reconstruction.** It remains to estimate $e(\mathrm{HMM})$, which obviously relates to the reconstruction. For any $t\in\mathbb{N}$, we define the element patch $S_t(K)$ in a recursive manner as

$$S_0(K)=K, \qquad S_t(K)=\bigcup_{\substack{\widetilde{K}\in\mathcal{T}_H,\ \overline{\widetilde{K}}\cap\overline{K'}\neq\emptyset\\ K'\subset S_{t-1}(K)}}\widetilde{K}, \tag{3.4}$$

where $\overline{\widetilde{K}}$ denotes the closure of $\widetilde{K}$. To highlight the dependence on t, for the following discussion we denote $\mathcal{I}_K$ by $\mathcal{I}_t(K)$.

We assume that $\mathcal{T}_H$ satisfies *the inverse assumption*, i.e., there exists a constant $\nu>0$ such that for any $K\in\mathcal{T}_H$,

$$\frac{H}{\rho_K}\leq\nu,$$

where ρ_K is the diameter of the largest ball inscribed into K. Though the above definition is not the same with the standard definition of the inverse assumption as in [11, p. 140], they are equivalent for any shape regular mesh. We use this definition for easy of exposition.

In the sequel we make the following assumption on the nodal set $\mathcal{I}_t(K)$.

Assumption A. For any $K \in \mathcal{T}_H$ and $g \in \mathbb{P}_m(S_t(K))$,

$$g|_{\mathcal{I}_t(K)} = 0 \quad \text{implies} \quad g|_{S_t(K)} \equiv 0.$$

For any $K \in \mathcal{T}_H$ and $g \in C^0\big(S_t(K)\big)$, we define a normalized discrete ℓ_2-norm as

$$\|g\|_{\ell_2} = \left(\frac{1}{\#\mathcal{I}_t(K)} \sum_{\boldsymbol{x}\in\mathcal{I}_t(K)} g^2(\boldsymbol{x})\right)^{1/2},$$

where $\#\mathcal{I}_t(K)$ is the cardinality of $\mathcal{I}_t(K)$. Next we define

$$\Lambda\big(m, \mathcal{I}_t(K)\big) = \max_{g\in\mathbb{P}_m(S_t(K))} \frac{\max_{\boldsymbol{x}\in S_t(K)}|g(\boldsymbol{x})|}{\max_{\boldsymbol{x}\in\mathcal{I}_t(K)}|g(\boldsymbol{x})|}. \tag{3.5}$$

By the equivalence of the norms over finite dimensional space $\mathbb{P}_m\big(S_t(K)\big)$, we have

$$\Lambda\big(m, \mathcal{I}_t(K)\big) < \infty. \tag{3.6}$$

This inequality can be viewed as a quantitative version of Assumption A.

The reconstruction procedure satisfies the following properties.

Theorem 3.3. *If Assumption A holds, then there exists a unique solution of* (2.2) *or* (2.5). *The unique solution will be denoted by* $\mathcal{R}_m\big(\widetilde{A}_H\big)_{ij}$ *for* $i, j = 1, \ldots, d$.

Moreover, $\mathcal{R}_m$ *satisfies*

$$\mathcal{R}_m g = g \quad \textit{for all } g \in \mathbb{P}_m(S_t(K)). \tag{3.7}$$

The stability property holds true for any $K \in \mathcal{T}_H$ *and* $g \in C^0\big(S_t(K)\big)$ *as*

$$\big\|\mathcal{R}_m g\big\|_{L^\infty(K)} \le \Lambda\big(m, \mathcal{I}_t(K)\big)\sqrt{\#\mathcal{I}_t(K)} \max_{\boldsymbol{x}\in\mathcal{I}_t(K)} |g(\boldsymbol{x})|, \tag{3.8}$$

and

$$\begin{aligned}\big\|g - \mathcal{R}_m g\big\|_{L^\infty(K)} \le{}& \Big(1 + \Lambda\big(m, \mathcal{I}_t(K)\big)\sqrt{\#\mathcal{I}_t(K)}\Big) \\ &\times \inf_{p\,\in\mathbb{P}_m\left(S_t(K)\right)} \big\|g - p\big\|_{L^\infty\left(S_t(K)\right)}.\end{aligned} \tag{3.9}$$

The above result is well known; see e.g., [35, Theorem 2.1]. The novelty of Theorem 3.3 lies in the fact that the constants in the estimates are explicitly characterized, which is crucial for the stability and accuracy of our algorithm.

Proof. For any real-valued functions g, h defined on $\mathcal{I}_t(K)$, we define a bilinear form as

$$\langle g, h\rangle_{\mathcal{I}_t(K)} = \frac{1}{\#\mathcal{I}_t(K)} \sum_{\boldsymbol{x}\in\mathcal{I}_t(K)} g(\boldsymbol{x})h(\boldsymbol{x}).$$

This defines an inner product over $\mathcal{I}_t(K)$ with the corresponding norm $\big\|g\big\|_{\ell_2}$ by Assumption A. Therefore, we may write the discrete least-squares problem (2.2) as a minimization problem of the ℓ_2-distance between $(\widetilde{\mathcal{A}}_H)_{ij}$ and $\mathbb{P}_m(S_t(K))$. While the constrained problem (2.5) can be viewed as a minimization problem of the ℓ_2-distance between $(\widetilde{\mathcal{A}}_H)_{ij}$ and a subspace of $\mathbb{P}_m(S_t(K))$. Therefore, the

existence and the uniqueness of the discrete least-squares problem is a special case of a projection theorem on a finite dimensional space [28].

The identity (3.7) is clear by regarding $\mathcal{R}_m$ as a projection operator from $C^0\big(S_t(K)\big)$ to $\mathbb{P}_m\big(S_t(K)\big)$ with respect to the discrete ℓ_2 norm.

By (3.6), we have

$$\big\|\mathcal{R}_m g\big\|_{L^\infty(K)} \le \big\|\mathcal{R}_m g\big\|_{L^\infty\left(S_t(K)\right)} \le \Lambda\big(m,\mathcal{I}_t(K)\big) \max_{\boldsymbol{x}\in\mathcal{I}_t(K)} |R_m g(\boldsymbol{x})|.$$

By the projection property of $\mathcal{R}_m$, we get

$$\big\|\mathcal{R}_m g\big\|_{\ell_2} \le \big\|g\big\|_{\ell_2}.$$

We get (3.8) by combining the above two inequalities.

Next we choose $p_0 \in \mathbb{P}_m\big(S_t(K)\big)$ such that

$$\big\|g - p_0\big\|_{L^\infty\left(S_t(K)\right)} = \inf_{p\in\mathbb{P}_m\left(S_t(K)\right)} \big\|g - p\big\|_{L^\infty\left(S_t(K)\right)}.$$

We apply (3.8) with g replaced by $g - p_0$, and use (3.7) to get

$$\begin{aligned}\big\|\mathcal{R}_m g - p_0\big\|_{L^\infty(K)} &= \big\|\mathcal{R}_m(g - p_0)\big\|_{L^\infty(K)} \\ &\le \Lambda\big(m,\mathcal{I}_t(K)\big)\sqrt{\#\mathcal{I}_t(K)} \max_{\boldsymbol{x}\in\mathcal{I}_t(K)} |(g-p_0)(\boldsymbol{x})| \\ &\le \Lambda\big(m,\mathcal{I}_t(K)\big)\sqrt{\#\mathcal{I}_t(K)}\big\|g - p_0\big\|_{L^\infty\left(S(K)\right)}.\end{aligned}$$

Therefore,

$$\begin{aligned}\big\|g - \mathcal{R}_m g\big\|_{L^\infty(K)} &\le \big\|g - p_0\big\|_{L^\infty(K)} + \big\|\mathcal{R}_m g - p_0\big\|_{L^\infty(K)} \\ &\le \Big(1 + \Lambda\big(m,\mathcal{I}_t(K)\big)\sqrt{\#\mathcal{I}_t(K)}\Big) \inf_{p\in\mathbb{P}_m\left(S_t(K)\right)} \big\|g - p\big\|_{L^\infty\left(S_t(K)\right)}.\end{aligned}$$

This gives (3.9). □

By (3.9), we conclude that the discrete least-squares approximation $\mathcal{R}_m g$ is a nearly optimal uniform approximation polynomial to g if $\Lambda\big(m,\mathcal{I}_t(K)\big)\sqrt{\#\mathcal{I}_t(K)}$ can be controlled. For a special case when $\#\mathcal{I}_t(K) = \binom{m+d}{d} = \dim\mathbb{P}_m(\mathbb{R}^d)$, we may replace $\Lambda\big(m,\mathcal{I}_t(K)\big)\sqrt{\#\mathcal{I}_t(K)}$ by the Lebesgue constant $\mathcal{L}\big(S_t(K)\big)$ [32, p. 24] that is defined by

$$\mathcal{L}\big(S(K)\big) = \max_{\boldsymbol{x}\in S(K)} \sum_{\boldsymbol{x}_\ell\in\mathcal{I}_t(K)} |l_{\boldsymbol{x}_\ell}(\boldsymbol{x})|,$$

where $l_{\boldsymbol{x}_\ell}$ is the Lagrange fundamental polynomial associated with $\boldsymbol{x}_\ell$. In this case, we only need modify the proof of (3.8) as follows. By Assumption A and $\sharp\mathcal{I}_t(K) = \binom{m+d}{d}$, we conclude that $\mathcal{R}_m g$ is the Lagrange interpolation of g for any $g \in C^0\big(S(K)\big)$. We have, for any $K \in \mathcal{T}_H$,

$$\big\|\mathcal{R}_m g\big\|_{L^\infty(K)} \le \mathcal{L}\big(S_t(K)\big) \max_{\boldsymbol{x}\in\mathcal{I}_t(K)} |g(\boldsymbol{x})|.$$

Unfortunately, we have little knowledge of Lebesgue constant in high dimension.

The next two lemmas seek conditions to bound such quantities. In next lemma, we give an upper bound for $\#\mathcal{I}_t(K)$.

Lemma 3.4. *If $\mathcal{T}_H$ satisfies the inverse assumption and $S_t(K)$ is convex, then*

$$\#\mathcal{I}_t(K) \leq \pi(t^2+3t)\nu + 3, \quad d=2, \tag{3.10}$$

$$\#\mathcal{I}_t(K) \leq \frac{2}{3}(2t^3+9t^2+13t)\nu^3 + 2t + 4, \quad d=3. \tag{3.11}$$

Proof. We firstly prove the case for $d=2$. For any $K \in \mathcal{T}_H$, $S_t(K)$ is covered by a circle centered at one vertex of K with radius $(t+1)H$. Notice that $S_t(K)$ is convex, we have

$$p_t(K) \leq 2\pi(t+1)H,$$

where $p_t(K)$ is the perimeter of $S_t(K)$. Using the fact that

$$p_t(K) \geq \#v_t(K) \min_{K\in S_t(K)} H_K \geq \#v_t(K) \min_{K\in S_t(K)} \rho_K,$$

where $\#v_t(K)$ is the number of vertices at the boundary of $S_t(K)$. Combining the above two inequalities and the inverse assumption gives

$$\#v_t(K) \leq 2\pi(t+1)\nu.$$

Using $\#v_t(K) = \#I_t(K) - \#I_{t-1}(K)$, we get the following recursive relation

$$\#I_t(K) - \#I_{t-1}(K) \leq 2\pi(t+1)\nu.$$

Solving the above recursive equation we obtain (3.10).

As to $d=3$, we firstly find a lower bound for the area of any face F of an element K. Denote by m_d the altitude from the vertex to the face F. Using the fact

$$\text{mes}K = \frac{m_d}{3}\text{mes}F \leq \frac{H_K}{3}\text{mes}F,$$

and

$$\text{mes}K \geq \frac{4\pi}{3}\left(\frac{\rho_K}{2}\right)^3,$$

we obtain

$$\text{mes}F \geq \frac{\pi}{2\nu}\rho_K^2.$$

Denote by $\#f_t(K)$ the number of faces at the boundary of $S_t(K)$. Note the area of the outer surface of $S_t(K)$ is less than $4\pi(t+1)^2H^2$ since $S_t(K)$ is convex. This fact together with the above inequality leads to

$$\#f_t(K) \min_{K\in S_t(K)} \frac{\pi}{2\nu}\rho_K^2 \leq 4\pi(t+1)^2H^2,$$

which gives the upper bound of $\#f_t(K)$.

$$\#f_t(K) \leq 8\nu^3(t+1)^2. \tag{3.12}$$

Next by *Euler's formula*,

$$\#v_t(K) - \#e + \#f_t(K) = 2,$$

where $\#e$ is the total number of the edges on the outer faces of $S_t(K)$, respectively. Using the fact that every edge belongs to two faces, we have

$$\#e = \frac{3}{2}\#f_t(K).$$

Combining the above two identities, we get

$$\#v_t(K) = \frac{1}{2}\#f_t(K) + 2.$$

We write the above equation as

$$\#I_t(K) - \#I_{t-1}(K) = \frac{1}{2}\#f_t(K) + 2.$$

Substituting the inequality (3.12) into the above equation, and solving this recursive relation, we get (3.11). □

It is not easy to find an explicit upper bound for $\Lambda\big(m, \mathcal{I}_t(K)\big)$ in general. If $d = 1$ and the nodes are equally spaced, then Coppersmitht and Rivlin [12] proved

$$\Lambda(m, \mathcal{I}_t(K)) \simeq \exp\left(\frac{cm^2}{\#\mathcal{I}_t(K)}\right) \tag{3.13}$$

with c a universe constant. The sharpness of this estimate and an interesting discussion on $\Lambda(m, \mathcal{I}_t(K))$ can be found in [34].

In next lemma, we state a condition under which $\Lambda(m, \mathcal{I}_t(K))$ is uniformly bounded by 2.

Lemma 3.5. *If $\mathcal{T}_H$ satisfies the inverse assumption and $S_t(K)$ is convex, and if in addition*

$$\begin{aligned} m &< \left(\frac{\pi\#S_t(K)}{16\#v_t(K)}\right)^{1/2}\nu^{-1}, \qquad d = 2, \\ m &< \left(\frac{\pi\#S_t(K)}{12\#f_t(K)}\right)^{1/2}\nu^{-3/2}, \qquad d = 3, \end{aligned} \tag{3.14}$$

then

$$\Lambda\big(m, \mathcal{I}_t(K)\big) \le 2, \tag{3.15}$$

where $\#S_t(K)$ denotes the number of elements belong to $S_t(K)$.

For a shape regular mesh, we have $\#v_t(K) \simeq \sqrt{\#\mathcal{I}_t(K)}$ in case of $d = 2$. As to $d = 3$, we have $\#f_t(K) \simeq \big(\#S_t(K)\big)^{2/3}$ and $\#S_t(K) \simeq \#\mathcal{I}_t(K)$. Hence,

$$\#f_t(K) \simeq \big(\#\mathcal{I}_t(K)\big)^{2/3}.$$

Therefore, the number of the sampling points is required to be

$$\#\mathcal{I}_t(K) \simeq m^{2d}$$

for the validity of the assumption (3.14), which is consistent with the one dimensional estimate (3.13). We note that the number of the sampling points is much

larger than $\dim\mathbb{P}_m(\mathbb{R}^d)$. This is just the price we have to pay for the uniform bound of $\Lambda\big(m,\mathcal{I}_t(K)\big)$.

Proof. We firstly prove the two-dimensional case. For any polynomial p of degree m, let $\widetilde{\boldsymbol{x}} \in S_t(K)$ such that $|p(\widetilde{\boldsymbol{x}})| = \max_{\boldsymbol{x}\in S_t(K)}|p(\boldsymbol{x})|$. Denote by $\widetilde{\boldsymbol{x}}_\ell = \arg\min_{\boldsymbol{y}\in\mathcal{I}_t(K)}|\widetilde{\boldsymbol{x}}-\boldsymbol{y}|$. By Taylor expansion, we have

$$p(\widetilde{\boldsymbol{x}}_\ell) = p(\widetilde{\boldsymbol{x}}) + (\widetilde{\boldsymbol{x}}_\ell - \widetilde{\boldsymbol{x}})\cdot\nabla p(\xi_{\boldsymbol{x}})$$

with $\xi_{\boldsymbol{x}}$ a point on the line with end points $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{x}}_\ell$. This gives

$$|p(\widetilde{\boldsymbol{x}})| \le |p(\widetilde{\boldsymbol{x}}_\ell)| + H_K \max_{\boldsymbol{x}\in S_t(K)}|\nabla p(\boldsymbol{x})|,$$

where

$$|\nabla p(\boldsymbol{x})| = \left(\sum_{i=1}^d \left|\frac{\partial p}{\partial x_i}(\boldsymbol{x})\right|^2\right)^{1/2}.$$

By Markov inequality [40], we have

$$\max_{\boldsymbol{x}\in S_t(K)}|\nabla p(\boldsymbol{x})| \le \frac{4m^2}{w(K)}\max_{\boldsymbol{x}\in S_t(K)}|p(\boldsymbol{x})|, \tag{3.16}$$

where $w(K)$ is the width of the convex polygon $S_t(K)$.

Next we look for a lower bound of $w(K)$. By the following inequality for the plane convex set [27],

$$2\,\mathrm{mes}S_t(K) \le w(K)p_t(K). \tag{3.17}$$

It is easy to see

$$p_t(K) \le \#v_t(K)H$$

and

$$\mathrm{mes}S_t(K) \ge \#S_t(K)\min_{K\in\mathcal{T}_H}\frac{\pi}{4}\rho_K^2.$$

Substituting the above two inequalities into (3.17), we obtain a lower bound for the width $w(K)$:

$$w(K) \ge \frac{\pi}{2}\min_{K\in S_t(K)}\frac{\rho_K^2}{H}\frac{\#S_t(K)}{\#v_t(K)}. \tag{3.18}$$

Substituting the above inequality into (3.16), using $(3.14)_1$, we get

$$\max_{\boldsymbol{x}\in S_t(K)}|p(\boldsymbol{x})| \le |p(\widetilde{\boldsymbol{x}}_\ell)| + \frac{1}{2}\max_{\boldsymbol{x}\in S_t(K)}|p(\boldsymbol{x})|,$$

which gives (3.15) for $d=2$.

The proof of (3.15) for $d=3$ is essentially the same since the inequalities (3.16) and (3.17) are valid for polyhedron. □

Remark. The condition (3.14) is less explicit. In the appendix, we shall give a condition that only concerns with m,ν and t when $d=2$, under which (3.15) remains true.

Remark. The convexity of $S_t(K)$ is assumed in Lemma 3.4 and Lemma 3.5. Nevertheless, the method can be applied to a nonconvex element patch $S_t(K)$; see § 4.1.

Remark. The inverse assumption for $\mathcal{T}_H$ actually may be removed at the cost of ν in the inequalities (3.10), (3.11) and (3.14) replaced by ν^k with a power k that depends on t. In this aspect, we refer to [13] for a discussion.

Define

$$\Lambda_m = \max_{K\in\mathcal{T}_H} \Lambda\big(m, \mathcal{I}_t(K)\big)\sqrt{\#\mathcal{I}_t(K)}.$$

If the condition (3.14) is valid, then Λ_m is uniformly bounded by Lemma 3.4 and Lemma 3.5, and the upper bound depends only on t and ν.

3.2. **Discretization error.** Without taking into account the discretization error of the cell problem, at each vertex $\boldsymbol{x}_\ell$, we define the effective matrix $\widehat{\mathcal{A}}_H(\boldsymbol{x}_\ell)$ as

$$\widehat{\mathcal{A}}_H(\boldsymbol{x}_\ell)\,\langle\nabla v^\varepsilon\rangle_{I_\delta} \equiv \langle a^\varepsilon \nabla v^\varepsilon\rangle_{I_\delta}\,,$$

where $v^\varepsilon - V_\ell \in \mathcal{V}$ satisfies

$$(a^\varepsilon\nabla v^\varepsilon, \nabla\varphi)_{L^2(I_\delta)} = 0 \quad \text{for all } \varphi\in\mathcal{V}. \tag{3.19}$$

Here $\mathcal{V}$ may be $\mathcal{V}_D, \mathcal{V}_N$ or $\mathcal{V}_P$ that is defined by

$$\begin{aligned}
\mathcal{V}_D &\equiv H_0^1\big(I_\delta(\boldsymbol{x}_\ell)\big),\\
\mathcal{V}_N &\equiv \big\{\, v\in H^1\big(I_\delta(\boldsymbol{x}_\ell)\big) \;\big|\; \langle\nabla v\rangle_{I_\delta} = 0\,\big\},\\
\mathcal{V}_P &\equiv \big\{\, v\in H^1_\sharp\big(I_\delta(\boldsymbol{x}_\ell)\big) \;\big|\; \langle v\rangle_{I_\delta} = 0\,\big\}.
\end{aligned}$$

Next lemma characterizes the discretization error of the cell problems. The key argument is hidden in [1, Lemma 3.1] for the case when a^ε is symmetric and in [15, Theorem 3.3, equation (3.23)] for the general case.

Lemma 3.6. *At each vertex $\boldsymbol{x}_\ell$, we have*

$$\begin{aligned}
\big\|(\widetilde{\mathcal{A}}_H - \widehat{\mathcal{A}}_H)(\boldsymbol{x}_\ell)\big\|_F \le \frac{\beta^3}{\alpha^2}&\left(\sum_{i=1}^d \inf_{v\in\mathcal{V}_h}\big\|\nabla(v_i^\varepsilon - x_i - v)\big\|^2_{L^2(I_\delta(\boldsymbol{x}_\ell))}\right)^{1/2}\\
&\times\left(\sum_{i=1}^d \inf_{v\in\mathcal{V}_h}\big\|\nabla(\widetilde{v}_i^\varepsilon - x_i - v)\big\|^2_{L^2(I_\delta(\boldsymbol{x}_\ell))}\right)^{1/2},
\end{aligned} \tag{3.20}$$

where v_i^ε is the solution of (3.19) *with $V_\ell = x_i$, while $\widetilde{v}_i^\varepsilon$ is the solution of* (3.19) *with $V_\ell = x_i$ and a^ε replaced by its transpose $(a^\varepsilon)^t$.*

Proof. Using the definition of $\widehat{\mathcal{A}}_H$, we write, at each vertex $\boldsymbol{x}_\ell$,

$$\begin{aligned}
\widehat{\mathcal{A}}_H(\boldsymbol{x}_\ell)_{ij} &= \big\langle\nabla v_i^\varepsilon\big\rangle_{I_\delta(\boldsymbol{x}_\ell)}\cdot\widehat{\mathcal{A}}_H(\boldsymbol{x}_\ell)\big\langle\nabla v_j^\varepsilon\big\rangle_{I_\delta(\boldsymbol{x}_\ell)} = \nabla x_i\cdot\big\langle a^\varepsilon\nabla v_j^\varepsilon\big\rangle_{I_\delta(\boldsymbol{x}_\ell)}\\
&= \big\langle\nabla x_i\cdot a^\varepsilon\nabla v_j^\varepsilon\big\rangle_{I_\delta(\boldsymbol{x}_\ell)} = \big\langle\nabla\widetilde{v}_i^\varepsilon\cdot a^\varepsilon\nabla v_j^\varepsilon\big\rangle_{I_\delta(\boldsymbol{x}_\ell)},
\end{aligned}$$

where we have used (3.19) in the last step of the above equation. Proceeding in the same manner, we obtain

$$\widetilde{\mathcal{A}}_H(\boldsymbol{x}_\ell)_{ij} = \left\langle \nabla \widetilde{v}^\varepsilon_{i,h} \cdot a^\varepsilon \nabla v^\varepsilon_{j,h} \right\rangle_{I_\delta(\boldsymbol{x}_\ell)},$$

where $\widetilde{v}^\varepsilon_{i,h}$ is the solution of (2.3) with a^ε replaced by its transpose. Combining the above two equations, we have

$$\left(\widehat{\mathcal{A}}_H - \widetilde{\mathcal{A}}_H\right)_{ij}(\boldsymbol{x}_\ell) = \left\langle \nabla(\widetilde{v}^\varepsilon_i - \widetilde{v}^\varepsilon_{i,h}) \cdot a^\varepsilon \nabla v^\varepsilon_j \right\rangle_{I_\delta(\boldsymbol{x}_\ell)} + \left\langle \nabla \widetilde{v}^\varepsilon_{i,h} \cdot a^\varepsilon \nabla (v^\varepsilon_j - v^\varepsilon_{j,h}) \right\rangle_{I_\delta(\boldsymbol{x}_\ell)}.$$

The first term in the right-hand side of the above equation is zero since $\widetilde{v}^\varepsilon_i - \widetilde{v}^\varepsilon_{i,h} \in \mathcal{V}$. Applying the same argument to the second term, we get

$$\begin{aligned}
\left(\widehat{\mathcal{A}}_H - \widetilde{\mathcal{A}}_H\right)_{ij}(\boldsymbol{x}_\ell) &= \left\langle \nabla \widetilde{v}^\varepsilon_{i,h} \cdot a^\varepsilon \nabla (v^\varepsilon_j - v^\varepsilon_{j,h}) \right\rangle_{I_\delta(\boldsymbol{x}_\ell)} \\
&= \left\langle \nabla (v^\varepsilon_j - v^\varepsilon_{j,h}) \cdot (a^\varepsilon)^t \nabla \widetilde{v}^\varepsilon_{i,h} \right\rangle_{I_\delta(\boldsymbol{x}_\ell)} \\
&= \left\langle \nabla (v^\varepsilon_j - v^\varepsilon_{j,h}) \cdot (a^\varepsilon)^t \nabla (\widetilde{v}^\varepsilon_{i,h} - \widetilde{v}^\varepsilon_i) \right\rangle_{I_\delta(\boldsymbol{x}_\ell)},
\end{aligned}$$

which gives (3.20) by combining the following standard estimates

$$\left\|\nabla(v^\varepsilon_i - v^\varepsilon_{i,h})\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le \frac{\beta}{\alpha} \inf_{v \in \mathcal{V}_h} \left\|\nabla(v^\varepsilon_i - x_i - v)\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}, \tag{3.21}$$

and

$$\left\|\nabla(\widetilde{v}^\varepsilon_j - \widetilde{v}^\varepsilon_{j,h})\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le \frac{\beta}{\alpha} \inf_{v \in \mathcal{V}_h} \left\|\nabla(\widetilde{v}^\varepsilon_j - x_j - v)\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}.$$

□

If a^ε is a symmetric matrix, then (3.20) changes to [1]

$$\left\|(\widetilde{\mathcal{A}}_H - \widehat{\mathcal{A}}_H)(\boldsymbol{x}_\ell)\right\|_F \le \frac{\beta^3}{\alpha^2} \sum_{i=1}^d \inf_{v \in \mathcal{V}_h} \left\|\nabla(v^\varepsilon_i - x_i - v)\right\|^2_{L^2(I_\delta(\boldsymbol{x}_\ell))}.$$

So far we make no assumption on the form of the coefficient except that $a^\varepsilon \in \mathcal{M}(\alpha, \beta, D)$. For $i = 1, \ldots, d$, let Πv^ε_i be the standard Lagrange interpolant of v^ε_i, which is well-defined because v^ε_i is Hölder continuous in $I_\delta(\boldsymbol{x}_\ell)$ [20]. Taking $v = \Pi(v^\varepsilon_i - x_i)$ in (3.21), we get

$$v^\varepsilon_i - x_i - v = v^\varepsilon_i - x_i - \Pi(v^\varepsilon_i - x_i) = v^\varepsilon_i - \Pi v^\varepsilon_i.$$

Therefore,

$$\left\|\nabla(v^\varepsilon_i - x_i - v)\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} = \left\|\nabla(v^\varepsilon_i - \Pi v^\varepsilon_i)\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le C h^{k'} \left\|v^\varepsilon_i\right\|_{H^{k'+1}(I_\delta(\boldsymbol{x}_\ell))}$$

provided that $\left\|v^\varepsilon_i\right\|_{H^{k'+1}(I_\delta(\boldsymbol{x}_\ell))}$ is bounded. However, this regularity result may not be true for $k' > 1$ [26]. Moreover, even this is true, we have to clarify the dependence of $\left\|v^\varepsilon_i\right\|_{H^{k'+1}(I_\delta(\boldsymbol{x}_\ell))}$ on the parameters ε and δ, which is not easy if it is not possible at all.

[1] If a^ε is symmetric, then we may replace β^3/α^2 in (3.20) with β by employing the energy norm $\|v\|_a \equiv \left(\int_D \nabla v \cdot a^\varepsilon \nabla v \, \mathrm{d}\boldsymbol{x}\right)^{1/2}$; see cf., [11, Remark 2.4.1].

When $k'=1$, under the assumption

$$|\nabla a^\varepsilon(x)| \le C/\varepsilon \quad a.e.,\ x\in D,$$

Du and Ming [15] proved

$$\left\|\nabla v_i^\varepsilon\right\|_{H^1(I_\delta(\boldsymbol{x}_\ell))} \le C/\varepsilon,$$

where C is independent of ε and δ. This immediately implies

$$\inf_{v\in\mathcal{V}_h}\left\|\nabla(v_i^\varepsilon - x_i - v)\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le Ch/\varepsilon, \tag{3.22}$$

where C is independent of ε,δ and h.

If a^ε is a locally periodic matrix, i.e., $a^\varepsilon(\boldsymbol{x}) = a(\boldsymbol{x},\boldsymbol{x}/\varepsilon)$ and $a(\boldsymbol{x},\boldsymbol{y})$ is periodic in $\boldsymbol{y}$ with period Y, then the situation is slightly different. When $k'=1$ and the periodic cell problem is used with $\delta=\varepsilon$, Abdulle [1] proved (3.22) under the assumption that $\|\boldsymbol{\chi}\|_{W^{2,\infty}(Y)}$ is bounded, where $\boldsymbol{\chi}$ is the solution of certain auxiliary problem, whose definition can be found in (3.23) below. For the periodic cell problems with $\delta/\varepsilon\in\mathbb{N}$, we may use the method in [29, Chapter 3] to get

$$\left\|v_i^\varepsilon\right\|_{H^{k'+1}(I_\delta(\boldsymbol{x}_\ell))} \le C\varepsilon^{-k'}.$$

We would have, for periodic cell problem with $\delta/\varepsilon\in\mathbb{N}$ and $k'>1$,

$$\inf_{v\in\mathcal{V}_h}\left\|\nabla(v_i^\varepsilon - x_i - v)\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le C\left(h/\varepsilon\right)^{k'}.$$

The same estimate can be found in [15, Corollary 3.10] and [2, Remark 9], which is based on a different argument. However, it is unclear whether the above regularity result holds true when $\delta/\varepsilon\notin\mathbb{N}$. A related discussion on the discretization error for the Dirichlet and the periodic cell problems can be found in [2].

In what follows, we estimate the discretization error for the Dirichlet, the Neumann, and the periodic cell problems when a^ε is a locally periodic matrix. Instead of using Lagrange interpolant in (3.21), we construct a special interpolant that is motivated by the following result

Lemma 3.7. [14, Lemma 3.2] *Let v^ε be the solution of* (3.19)*, and define*

$$\widehat{V}^\varepsilon \equiv V_\ell + \varepsilon(\boldsymbol{\chi}\cdot\nabla)V_\ell,$$

where $\boldsymbol{\chi}(\boldsymbol{x},\boldsymbol{y}) = \{\chi^j(\boldsymbol{x},\boldsymbol{y})\}_{j=1}^d$ is periodic in $\boldsymbol{y}$ with period Y and it satisfies

$$-\frac{\partial}{\partial y_i}\left(a_{ik}\frac{\partial\chi^j}{\partial y_k}\right)(\boldsymbol{x},\boldsymbol{y}) = \left(\frac{\partial}{\partial y_i}a_{ij}\right)(\boldsymbol{x},\boldsymbol{y}) \text{ in } Y, \quad \int_Y \chi^j(\boldsymbol{x},\boldsymbol{y})\,\mathrm{d}\boldsymbol{y} = 0. \tag{3.23}$$

Here V_ℓ is the linear approximation of V at $\boldsymbol{x}_\ell$. If $a^\varepsilon = a(\boldsymbol{x},\boldsymbol{x}/\varepsilon)$ with $a(\boldsymbol{x},\boldsymbol{y})\in C^{0,1}(D;L^\infty(Y))$, and $a(\boldsymbol{x},\boldsymbol{y})$ is periodic in $\boldsymbol{y}$ with period Y, then there holds

$$\left\|\nabla(v^\varepsilon - \widehat{V}^\varepsilon)\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le C\frac{\beta^2}{\alpha^2}\left(\frac{\varepsilon}{\delta}\right)^{1/2}\left\|\nabla V_\ell\right\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}. \tag{3.24}$$

A direct consequence of the above result is

Corollary 3.8. *Define*

$$\widetilde{V}^\varepsilon = V_\ell + \varepsilon(\varrho^\varepsilon \boldsymbol{\chi}\cdot\nabla)V_\ell,$$

where $\varrho^\varepsilon \in C_0^\infty(I_\delta)$ *is a cut-off function that satisfies* $|\nabla\varrho^\varepsilon| \le C/\varepsilon$, *and*

$$\varrho^\varepsilon(\boldsymbol{x}) = \begin{cases} 1 & \text{if } dist(\boldsymbol{x}, \partial I_\delta(\boldsymbol{x}_\ell)) \ge 2\varepsilon, \\ 0 & \text{if } dist(\boldsymbol{x}, \partial I_\delta(\boldsymbol{x}_\ell)) \le \varepsilon. \end{cases}$$

If $a^\varepsilon = a(\boldsymbol{x}, \boldsymbol{x}/\varepsilon)$ *with* $a(\boldsymbol{x},\boldsymbol{y}) \in C^{0,1}(D; L^\infty(Y))$, *and* $a(\boldsymbol{x},\boldsymbol{y})$ *is periodic in* $\boldsymbol{y}$ *with period* Y, *then there holds*

$$\big\|\nabla(v^\varepsilon - \widetilde{V}^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le C\frac{\beta^2}{\alpha^2}\left(\frac{\varepsilon}{\delta}\right)^{1/2}\big\|\nabla V_\ell\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}. \tag{3.25}$$

Proof. Note that $\widehat{V}^\varepsilon - \widetilde{V}^\varepsilon = (\widehat{V}^\varepsilon - V_\ell)(1-\varrho^\varepsilon)$, by [14, Lemma 3.1], we obtain

$$\big\|\nabla[(\widehat{V}^\varepsilon - V)(1-\varrho^\varepsilon)]\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le C\frac{\beta}{\alpha}\left(\frac{\varepsilon}{\delta}\right)^{1/2}\big\|\nabla V_\ell\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))},$$

which combines with (3.24) yields

$$\begin{aligned}\big\|\nabla(v^\varepsilon - \widetilde{V}^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} &\le \big\|\nabla(v^\varepsilon - \widehat{V}^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} + \big\|\nabla(\widehat{V}^\varepsilon - \widetilde{V}^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \\ &\le C\frac{\beta^2}{\alpha^2}\left(\frac{\varepsilon}{\delta}\right)^{1/2}\big\|\nabla V_\ell\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}.\end{aligned}$$

□

Based on the above result, we estimate the error between v^ε and v_h^ε.

Lemma 3.9. *Let* v^ε *and* v_h^ε *be the solutions of Problems* (2.4) *and* (3.19), *respectively. If*

$$\big\|\boldsymbol{\chi}\big\|_{H^{k'+1}(Y)} < \infty, \tag{3.26}$$

and if in addition $a^\varepsilon = a(\boldsymbol{x}, \boldsymbol{x}/\varepsilon)$ *with* $a(\boldsymbol{x},\boldsymbol{y}) \in C^{0,1}(D; L^\infty(Y))$, *and* $a(\boldsymbol{x},\boldsymbol{y})$ *is periodic in* $\boldsymbol{y}$ *with period* Y, *then*

$$\big\|\nabla(v^\varepsilon - v_h^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le C\left(\left(\frac{\varepsilon}{\delta}\right)^{1/2} + \frac{h^{k'}}{\varepsilon^{k'}}\right)\big\|\nabla V_\ell\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}. \tag{3.27}$$

Proof. Similar to (3.21), we have

$$\big\|\nabla(v^\varepsilon - v_h^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le \frac{\beta}{\alpha}\inf_{v\in\mathcal{V}_h}\big\|\nabla(v^\varepsilon - V_\ell - v)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}. \tag{3.28}$$

For the Dirichlet cell problem, we have $\Pi\widetilde{V}^\varepsilon = V_\ell + \Pi(\widetilde{V}^\varepsilon - V_\ell)$. If follows from $\widetilde{V}^\varepsilon - V_\ell \in H_0^1(I_\delta(\boldsymbol{x}_\ell))$ that $\Pi(\widetilde{V}^\varepsilon - V_\ell) \in H_0^1(I_\delta(\boldsymbol{x}_\ell))$. This yields $\Pi\widetilde{V}^\varepsilon - V_\ell \in \mathcal{V}_{D,h}$.

For the Neumann cell problem, an integration by parts gives

$$\big\langle\nabla(\Pi\widetilde{V}^\varepsilon - V_\ell)\big\rangle_{I_\delta(\boldsymbol{x}_\ell)} = \big\langle\nabla[\Pi(\widetilde{V}^\varepsilon - V_\ell)]\big\rangle_{I_\delta(\boldsymbol{x}_\ell)} = \boldsymbol{0},$$

which leads to $\Pi\widetilde{V}^\varepsilon - V_\ell \in \mathcal{V}_{N,h}$.

For the periodic cell problem, we replace $\Pi\widetilde{V}^\varepsilon - V_\ell$ by $\Pi\widetilde{V}^\varepsilon - V_\ell + c$, where c is a suitable constant such that $\big\langle\Pi\widetilde{V}^\varepsilon - V_\ell + c\big\rangle_{I_\delta(\boldsymbol{x}_\ell)} = 0$. Therefore, $\Pi\widetilde{V}^\varepsilon - V_\ell + c \in \mathcal{V}_{P,h}$.

Taking $v = \Pi\widetilde{V}^\varepsilon - V_\ell$ or $\Pi\widetilde{V} - V_\ell + c$ in (3.28), and notice

$$\nabla(v^\varepsilon - V_\ell - v) = \nabla(v^\varepsilon - V_\ell - \Pi\widetilde{V}^\varepsilon + V_\ell) = \nabla(v^\varepsilon - \Pi\widetilde{V}^\varepsilon),$$

and

$$\nabla(v^\varepsilon - V_\ell - v) = \nabla(v^\varepsilon - V_\ell - \Pi\widetilde{V}^\varepsilon + V_\ell - c) = \nabla(v^\varepsilon - \Pi\widetilde{V}^\varepsilon),$$

we get

$$\begin{aligned}\big\|\nabla(v^\varepsilon - v_h^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} &\le \frac{\beta}{\alpha}\big\|\nabla(v^\varepsilon - \Pi\widetilde{V}^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \\ &\le \frac{\beta}{\alpha}\Big(\big\|\nabla(v^\varepsilon - \widetilde{V}^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} + \big\|\nabla(\widetilde{V}^\varepsilon - \Pi\widetilde{V}^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}\Big).\end{aligned}$$

By the standard interpolation estimate, we have

$$\big\|\nabla(\widetilde{V}^\varepsilon - \Pi\widetilde{V}^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le Ch^{k'}\big\|\nabla^{k'+1}\widetilde{V}^\varepsilon\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}.$$

A direct calculation gives

$$\begin{aligned}\big\|\nabla^{k'+1}\widetilde{V}^\varepsilon\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} &\le C\varepsilon^{-k'}\delta^{d/2}\big\|\nabla_{\boldsymbol{y}}^{k'+1}\boldsymbol{\chi}\big\|_{L^2(Y)}|\nabla V_\ell| \\ &= C\varepsilon^{-k'}\big\|\nabla_{\boldsymbol{y}}^{k'+1}\boldsymbol{\chi}\big\|_{L^2(Y)}\big\|\nabla V_\ell\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))}.\end{aligned}$$

Combining the above two inequalities gives

$$\big\|\nabla(\widetilde{V}^\varepsilon - \Pi\widetilde{V}^\varepsilon)\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))} \le C(h/\varepsilon)^{k'}\big\|\nabla_{\boldsymbol{y}}^{k'+1}\boldsymbol{\chi}\big\|_{L^2(Y)}\big\|\nabla V_\ell\big\|_{L^2(I_\delta(\boldsymbol{x}_\ell))},$$

which together with (3.25) and (3.26) leads to (3.27). □

Combining (3.27) and (3.20), we get

Lemma 3.10. *Under the same condition of Lemma 3.9, we have*

$$\big\|(\widetilde{\mathcal{A}}_H - \widehat{\mathcal{A}}_H)(\boldsymbol{x}_\ell)\big\|_F \le C\Big(\frac{\varepsilon}{\delta} + \frac{h^{2k'}}{\varepsilon^{2k'}}\Big). \tag{3.29}$$

Before proving the main theorem, we need an auxiliary result that quantifies the error between $\mathcal{A}$ and $\widehat{\mathcal{A}}_H$ at each vertex.

Lemma 3.11. [14, Theorem 3.4] *If $a^\varepsilon = a(\boldsymbol{x}, \boldsymbol{x}/\varepsilon)$ with $a(\boldsymbol{x}, \boldsymbol{y}) \in C^{0,1}(D; L^\infty(Y))$, and $a(\boldsymbol{x}, \boldsymbol{y})$ is periodic in $\boldsymbol{y}$ with period Y, then, at each vertex $\boldsymbol{x}_\ell$,*

$$\big\|(\mathcal{A} - \widehat{\mathcal{A}}_H)(\boldsymbol{x}_\ell)\big\|_F \le C\frac{\beta^4}{\alpha^3}\Big(\delta + \frac{\varepsilon}{\delta}\Big). \tag{3.30}$$

We are ready to prove the main theorem of the paper.

Theorem 3.12. *For $i, j = 1, \ldots, d$, and let $a_{ij}(\boldsymbol{x}, \boldsymbol{y})$ be a periodic function in y with period Y. If $a_{ij}(\boldsymbol{x}, \boldsymbol{y})$ is smooth in both $\boldsymbol{x}$ and $\boldsymbol{y}$, and mth order reconstruction is used. Moreover, if Assumptions A, the conditions* (3.14) *and* (3.26) *hold, then*

$$e(HMM) \le C\Big(H^{m+1} + \delta + \frac{\varepsilon}{\delta} + \frac{h^{2k'}}{\varepsilon^{2k'}}\Big). \tag{3.31}$$

Proof. For any $K \in \mathcal{T}_H$ and $\boldsymbol{x} \in K$, using (3.9), we have

$$\big\|(\mathcal{A} - \mathcal{R}_m\mathcal{A})(\boldsymbol{x})\big\|_F \le C\,(1+\Lambda_m)\,H^{m+1},$$

where C depends on $\big\|\mathcal{A}\big\|_{W^{m+1,\infty}\left(S_t(K)\right)}$. Next, using (3.8), we have

$$\begin{aligned} e(\mathrm{HMM}) &\le \max_{\substack{\boldsymbol{x}\in K\\ K\in\mathcal{T}_H}} \Big(\big\|(\mathcal{A} - \mathcal{R}_m\mathcal{A})(\boldsymbol{x})\big\|_F + \big\|\mathcal{R}_m(\mathcal{A} - \widetilde{\mathcal{A}}_H)(\boldsymbol{x})\big\|_F\Big) \\ &\le C\,(1+\Lambda_m)\,H^{m+1} + \Lambda_m \max_{\substack{\boldsymbol{x}_\ell\in\mathcal{I}_t(K)\\ K\in\mathcal{T}_H}} \big\|(\mathcal{A} - \widetilde{\mathcal{A}}_H)(\boldsymbol{x}_\ell)\big\|_F. \end{aligned}$$

Combining the above estimate with (3.30) and (3.29), we get (3.31). □

Combining Lemma 3.2 and the above theorem, we obtain the error estimate of the proposed method.

Corollary 3.13. *Under the same condition of Theorem 3.12, we have*

$$\begin{aligned} \big\|\nabla(U_0 - U_H)\big\|_{L^2(D)} &\le C\bigg(H^k + H^{m+1} + \delta + \frac{\varepsilon}{\delta} + \frac{h^{2k'}}{\varepsilon^{2k'}}\bigg), \\ \big\|U_0 - U_H\big\|_{L^2(D)} &\le C\bigg(H^{k+1} + H^{m+1} + \delta + \frac{\varepsilon}{\delta} + \frac{h^{2k'}}{\varepsilon^{2k'}}\bigg). \end{aligned} \tag{3.32}$$

4. Numerical Results

To demonstrate the efficiency of the method, we report numerical results of Problem (1.1) with the following data:

$$\begin{cases} a(\boldsymbol{x}, \boldsymbol{x}/\epsilon) = \dfrac{(R_1 + R_2\sin(2\pi x_1))(R_1 + R_2\cos(2\pi x_2))}{(R_1 + R_2\sin(2\pi x_1/\varepsilon))(R_1 + R_2\sin(2\pi x_2/\varepsilon))} I, \\ \qquad f(\boldsymbol{x}) = 1, \\ \qquad u(\boldsymbol{x}) = 0 \quad \text{on } \partial D, \end{cases} \tag{4.1}$$

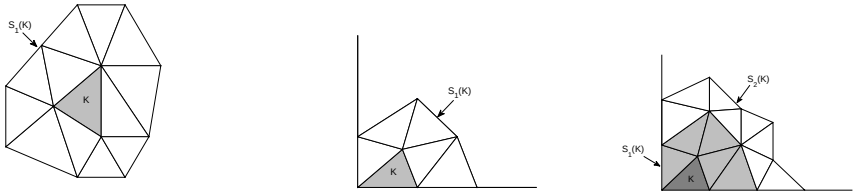
where $\varepsilon = 10^{-6}$, $D = (0,1)\times(0,1)$ and I is the 2 by 2 identity matrix. This problem has been studied in [15]. All of our computation are carried out on an IBM laptop with core speed 2.50 GHz.

A direct calculation gives the explicit formula of the effective matrix as

$$\mathcal{A}(x_1, x_2) = \frac{(R_1 + R_2\sin(2\pi x_1))(R_1 + R_2\cos(2\pi x_2))}{R_1\sqrt{R_1^2 - R_2^2}} I. \tag{4.2}$$

We take $R_1 = 2.5$ and $R_2 = 1.5$ in the simulation. The standard Gauss quadrature rule is used to compute the stiffness matrix in (2.1). At each quadrature node, $\mathcal{A}_H$ is calculated by (2.2) or (2.5). To compute each entry of $\widetilde{\mathcal{A}}_H$, we use (2.3) and take the boundary data V_ℓ in the cell problem (2.4) as $\boldsymbol{e}_i \cdot \boldsymbol{x}$, where $\{\boldsymbol{e}_i\}_{i=1}^2$ are the canonical basis. The domain D is triangulated by *EasyMesh*[2] with $H = 1/N$, and

[2]see *http://www-dinma.univ.trieste.it/nirftc/research/easymesh/*



(a) K is an interior element (b) K is on the boundary (c) $S_2(K)$

FIG. 2. Examples of the element patches $S_t(K)$ with $t = 1, 2$.

the cell I_δ is also triangulated by *EasyMesh* with $h = \delta/M$. In terms of N and M, the error bound (3.32) changes to

$$\begin{aligned} \big\|\nabla(U_0 - U_H)\big\|_{L^2(D)} &\le C\bigg(N^{-k} + N^{-m-1} + \delta + \frac{\varepsilon}{\delta} + \frac{\delta^{2k'}}{(M\varepsilon)^{2k'}}\bigg), \\ \big\|U_0 - U_H\big\|_{L^2(D)} &\le C\bigg(N^{-k-1} + N^{-m-1} + \delta + \frac{\varepsilon}{\delta} + \frac{\delta^{2k'}}{(M\varepsilon)^{2k'}}\bigg). \end{aligned} \tag{4.3}$$

4.1. **Details for least-squares reconstruction.** To recover $\mathcal{A}_H$ on an element K, the patch $S_t(K)$ is built by aggregating the elements around K and collecting their vertices as $\mathcal{I}_t(K)$. Three examples of $S_t(K)$ are shown in Fig. 2.

A direct consequence of Assumption A is

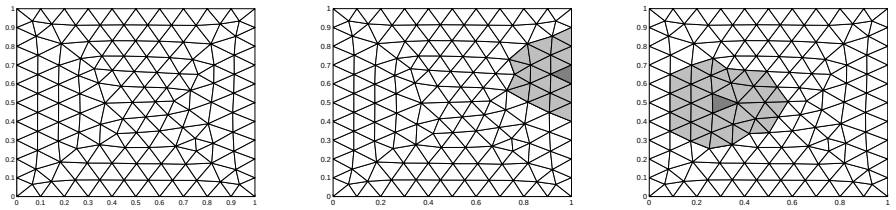
$$\#\mathcal{I}_t(K) \ge \binom{m+d}{d}. \tag{4.4}$$

For large m or the elements near the boundary, the cardinality $\#\mathcal{I}_t(K)$ may be smaller than $\binom{m+d}{d}$ when $t = 1$. This obviously contradicts with (4.4). A natural way out of this difficult is to include more layers into $S_t(K)$. For example, third order reconstruction requires at least ten nodes. For the shadowed element in Fig. 2(b), $\#\mathcal{I}_1(K) = 7$. In order to sample enough nodal values, we use a larger patch as in Fig. 2(c), in which $\#\mathcal{I}_2(K) = 13$. Actually, an even larger patch is required to guarantee the uniform boundedness of $\Lambda\big(m, \mathcal{I}_t(K)\big)$ because the condition (3.14) essentially requires that $\#\mathcal{I}_t(K)$ is of $\mathcal{O}(m^{2d})$ for mth order reconstruction. However, it is worthwhile to note that $\#\mathcal{I}_t(K)$ is irrelevant to the cost of solving the linear system arising from (2.2) or (2.5). Moreover, it is observed that the reconstruction is more stable with a bigger $\#\mathcal{I}_t(K)$.

In what follows, we compute $\Lambda\big(m, \mathcal{I}_t(K)\big)$ for three examples. First we consider $\mathcal{A}(\boldsymbol{x}) = a(\boldsymbol{x})I$ with

$$\begin{aligned} a(\boldsymbol{x}) = {} & (x_1 + x_2 - 0.5)^2(x_1 - x_2 + 1)^2\left(5 - 4(x - 0.2)^2 - 4(x_2 - 0.6)^2\right) \\ & + \frac{1}{2.1 + \sin\left(3\pi x_1/2 - 3\right) + \cos(2\pi x_2)}. \end{aligned}$$

We choose two types of elements in the triangulation as shown in Fig. 3(a), one is on the boundary, and the other is in the interior of the domain. The patches $S_2(K)$ for both elements are shown in Fig. 3(b) and Fig. 3(c), respectively. The corresponding values of $\Lambda\big(m, \mathcal{I}_t(K)\big)$ are reported in Table 1 and Table 2, respectively. It is clear that $\Lambda\big(m, \mathcal{I}_t(K)\big)$ is only slightly bigger than 1.0. The results for other patches are similar. Therefore, the quantity $\Lambda\big(m, \mathcal{I}_t(K)\big)$ can be controlled. We note that both element patches $S_2(K)$ in Fig. 3(b) and Fig. 3(c) are non-convex, it would be interesting to know whether the convexity assumption on the element patch in Lemma 3.5 and Lemma A.1 can be removed.



(a) mesh size = 0.1 (b) K is on the boundary (c) K is inside the domain

FIG. 3. The mesh and two examples of non-convex element patch.

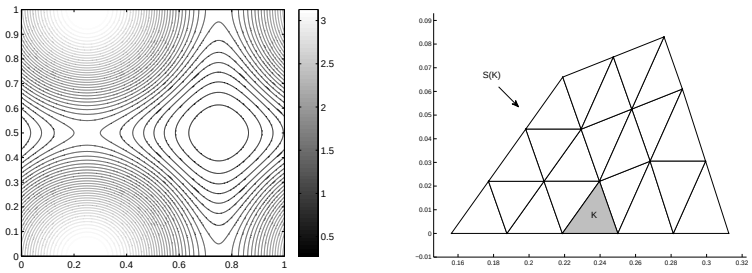
TABLE 1. Constant $\Lambda\big(m, \mathcal{I}_t(K)\big)$ when K is on the boundary.

m	$\big\|p\big\|_{L^\infty\left(S_2(K)\right)}$	$\big\|p\big\|_{\ell^\infty}$	$\Lambda\big(m, \mathcal{I}_t(K)\big)$
2	6.2361	6.1981	1.0061
3	6.2534	6.2210	1.0052
4	6.2575	6.2229	1.0059

TABLE 2. Constant $\Lambda\big(m, \mathcal{I}_t(K)\big)$ when K is inside the domain.

m	$\big\|p\big\|_{L^\infty\left(S_2(K)\right)}$	$\big\|p\big\|_{\ell^\infty}$	$\Lambda\big(m, \mathcal{I}_t(K)\big)$
2	4.8370	4.7774	1.0125
3	4.9150	4.8217	1.0193
4	6.9050	6.3542	1.0867

Next we compute $\Lambda\big(m, \mathcal{I}_t(K)\big)$ for the effective coefficient (4.2). The contour line of the effective coefficient is shown in Fig. 4(a). In view of the graph we may find that $\max_{\boldsymbol{x}\in D} \mathcal{A}(\boldsymbol{x}) = 3.2$ and the maximum is achieved at the points $(0.25, 0)$ and $(0.25, 1)$. Due to the symmetry of $\mathcal{A}(\boldsymbol{x})$, we only consider $\Lambda\big(m, \mathcal{I}_t(K)\big)$ over $S_2(K)$ that is near the point $(0.25, 0)$, where $S_2(K)$ is shown in Fig. 4(b). Table 3 shows



(a) Contour line of the effective matrix (b) $S_2(K)$ around the point $(0.25, 0)$

FIG. 4. Contour line and $S_2(K)$.

that $\Lambda\big(m, \mathcal{I}_t(K)\big)$ equals to 1 for $m = 2, 3, 4$. Due to the property of $\mathcal{A}(x)$, we may conclude that $\Lambda\big(m, \mathcal{I}_t(K)\big)$ is also uniformly bounded by a constant that is smaller than 2.

TABLE 3. Constant $\Lambda\big(m, \mathcal{I}_t(K)\big)$.

m	$\|p\|_{L^\infty\left(S_2(K)\right)}$	$\|p\|_{\ell^\infty}$	$\Lambda\big(m, \mathcal{I}_t(K)\big)$
2	3.1995	3.1995	1
3	3.1998	3.1998	1
4	3.20	3.20	1

Next the effective coefficient $\mathcal{A}(\boldsymbol{x}) = a(\boldsymbol{x})I$ is taken as the probability density function of normal distribution

$$a(\boldsymbol{x}) = 0.01 + \exp\left(-\frac{(x_1 - 0.85)^2 + (x_2 - 0.85)^2}{2\sigma^2}\right)$$

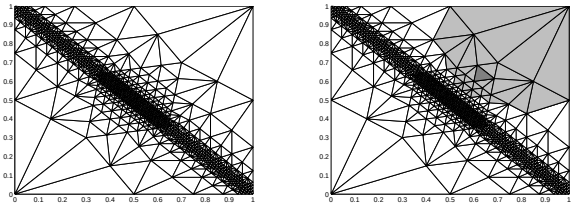
with $\sigma = 0.12$.

The mesh is generated by refining an initial grid near the diagonal, $S_2(K)$ is located at the position where the element size varies dramatically, while each element is shape-regular. We refer to Fig. 5(b) for an example of $S_2(K)$.

Now we compute $\Lambda\big(m, \mathcal{I}_t(K)\big)$ for the above coefficient $\mathcal{A}(\boldsymbol{x})$, and take $S_2(K)$ from Fig. 5(b). It follows from Table 4 that $\Lambda\big(m, \mathcal{I}_t(K)\big)$ grows with the reconstruction order m.

It follows from Table 5 that $\Lambda\big(m, \mathcal{I}_t(K)\big)$ is smaller than 2 when t is bigger than 3. This is consistent with Lemma 3.5.

4.2. **Numerical examples for the method.** In order to examine the convergence order of the reconstruction, we use the nodal values of the homogenized coefficients (4.2) as the input data of the least-squares reconstruction, the error



(a) adaptive mesh (b) K is inside the domain

FIG. 5. The mesh and $S_2(K)$.

TABLE 4. Constant $\Lambda\big(m, \mathcal{I}_t(K)\big)$ with m grows.

m	$\|p\|_{L^\infty\left(S_2(K)\right)}$	$\|p\|_{\ell^\infty}$	$\Lambda\big(m, \mathcal{I}_t(K)\big)$
2	0.2257	0.2234	1.0101
3	0.3903	0.2193	1.7794
4	0.8162	0.2198	3.7131

TABLE 5. Constant $\Lambda\big(m, \mathcal{I}_t(K)\big)$ with t grows.

	$t=2$	$t=3$	$t=4$	$t=5$	$t=6$
$m=2$	1.0101	1.0000	1.0000	1.0000	1.0000
$m=3$	1.7794	1.4165	1.1808	1.0768	1.0341
$m=4$	3.7131	1.6604	1.9327	1.8771	1.7860

bound (3.32) changes to

$$\begin{aligned} \big\|\nabla(U_0 - U_H)\big\|_{L^2(D)} &\le C\left(N^{-k} + N^{-m-1}\right), \\ \big\|U_0 - U_H\big\|_{L^2(D)} &\le C\left(N^{-k-1} + N^{-m-1}\right). \end{aligned} \tag{4.5}$$

In view of (4.5), we conclude that, in terms of H^1/L^2 error, $(k-1)$th/kth order reconstruction has to be used to match the kth order macrosopic solver, respectively. Table 6 shows clearly that the estimate (4.5) is optimal when $m = k = 1$. The results in Tables 7–8 illustrate the necessity to use second order reconstruction in order to obtain the optimal L^2 error estimate for the quadratic macroscopic solver. The results in Tables 6–8 are based on the least-squares reconstruction with constraints.

It is interesting to compare the reconstruction procedures with or without constraints. We report the results in Table 8 and Table 9, respectively. It agrees with the expectation that both methods achieve full order accuracy, while the constrained reconstruction slightly outperforms the one without constraints.

TABLE 6. P_1 element, 1st order reconstruction.

N	L^2 error	order	H^1 error	order
4	0.1622		0.3472	
8	0.0479	1.76	0.1742	0.99
16	0.0126	1.92	0.0875	0.99
32	0.0032	1.99	0.0436	1.01

TABLE 7. P_2 element, 1st order reconstruction.

N	L^2 error	order	H^1 error	order
4	0.0670		0.1208	
8	0.0193	1.80	0.0375	1.69
16	0.0050	1.95	0.0098	1.93
32	0.0013	1.99	0.0025	1.98

TABLE 8. P_2 element, 2nd order reconstruction.

N	L^2 error	order	H^1 error	order
4	0.0201		0.0798	
8	0.0026	2.97	0.0230	1.79
16	0.0004	3.29	0.0059	1.96
32	2.99e-05	3.14	0.0015	1.98

TABLE 9. P_2 element, 2nd order reconstruction without constraints.

N	L^2 error	order	H^1 error	order
4	0.0769		0.1758	
8	0.0093	3.05	0.0461	1.93
16	0.0007	3.68	0.0082	2.49
32	7.31e-05	3.31	0.0017	2.29

In what follows, we consider the case when the nodal values of the effective coefficients are obtained by solving the cell problems. Beside the macro-micro discretization error, there is another term, namely $\delta + \varepsilon/\delta$, stands for the so called resonance error. Extensive numerical experiments have been carried out to illustrate the influence of the resonance error in HMM-FEM; see [30, 41, 15, 23] and references therein. We shall not repeat it here, and we solve the periodic cell problems with $\delta = \varepsilon$. If the macroscopic solver is the quadratic element, the microscopic solver is linear element, and the second order reconstruction with constraints are used, then

the error estimate changes to

$$\left\|\nabla(U_0 - U_H)\right\|_{L^2(D)} \leq C(N^{-2} + \varepsilon + M^{-2}),$$
$$\left\|U_0 - U_H\right\|_{L^2(D)} \leq C(N^{-3} + \varepsilon + M^{-2}).$$

Equating the terms in the right-hand side of the above inequalities except ε since it is very small, i.e., $\varepsilon = 10^{-6}$, we get the following refinement strategy on the microcell, which has been proposed in [15]:

$$M = \begin{cases} N, & H^1 \text{ error}, \\ N^{3/2}, & L^2 \text{ error}, \end{cases}$$

and we take $M = N^{3/2}$ in the simulation.

A method based on the mid-point quadrature rule is proposed in [15] for quadratic macroscopic solver (P_2-edge for short). We report the results for our method and P_2-edge in Table 10 and Table 11, respectively. The CPU time of the new method asymptotically approaches to one third of the CPU time of P_2-edge. The saving is due to the fact that the number of the cell problems for P_2-edge is proportional to the total number of the edges, while the number of the cell problems for our method is proportional to the total number of the vertices, which is asymptotically one third of the number of the edges in a 2D simplex mesh as the mesh size tends to zero.

TABLE 10. Result of the new method.

N	M	CPU time(s)	L^2 error	H^1 error
4	8	0.31	0.0210	0.0805
8	32	11.85	0.0027	0.0230
16	64	165.88	0.0003	0.0059

TABLE 11. P_2-edge in [15].

N	M	CPU time(s)	L^2 error	H^1 error
4	8	0.46	0.0193	0.0809
8	32	27.75	0.0021	0.0236
16	64	445.65	0.0003	0.0060

In the last example the refinement strategy $M = N^{3/2}$ on the microcell makes the overall cost increasing rather rapidly as the macroscopic mesh is refined. The situation is even worse if we use cubic element as the macroscopic solver. In this case,

$$\left\|\nabla(U_0 - U_H)\right\|_{L^2(D)} \leq C(N^{-3} + \varepsilon + M^{-2k'}),$$
$$\left\|U_0 - U_H\right\|_{L^2(D)} \leq C(N^{-4} + \varepsilon + M^{-2k'}).$$

If $k' = 1$, then $M = N^2$, which leads to a sharp growth of the overall cost. We use high order microsopic solver to reduce the cost. For example, if $k' = 2$, then $M = N$. The advantage of high order microscopic solver is quite significant by the results in Table 12 and Table 13, at least for the problem with smooth microstructures. Higher order microscopic solver has been advocated in [3].

TABLE 12. Macro P_3 element, micro P_1 element, 3rd order reconstruction.

N	M	CPU time(s)	L^2 error	order	H^1 error	order
4	16	1.08	0.0200		0.0489	
8	64	49.72	0.0031	2.67	0.0063	2.94
16	256	> 3000	0.0002	3.82	0.0006	3.44

TABLE 13. Macro P_3 element, micro P_2 element, 3rd order reconstruction.

N	M	CPU time(s)	L^2 error	order	H^1 error	order
4	4	0.28	0.0190		0.0486	
8	8	2.99	0.0030	2.65	0.0063	2.95
16	16	46.03	0.0002	3.80	0.0006	3.43
32	32	803.89	1.4e-05	3.97	7.9e-05	2.89

5. Conclusion

In this paper, we have proposed a new high order HMM-FEM based on a local least-squares reconstruction of the effective coefficients. Theoretical and numerical results show that the method is more efficient than the high order HMM-FEM appeared in [17] and [15]. We also gave a unified analysis of the discretization error for the locally periodic problems when the cell problem is subject to the Dirichlet, the Neumann, or the periodic boundary condition.

Noticing that the proposed method is problem-independent, it can be readily extended to the nonlinear problem, the parabolic problem, the wave equation; see [17, 21, 22, 31, 18, 4]. We believe that the present idea is not limited to HMM, it may be used in other multiscale method; see [44], which will be a subject of our future work.

Appendix A. An explicit bound for $\Lambda\big(m, \mathcal{I}_t(K)\big)$

In this appendix, we give an explicit condition on m for the validity of the uniform upper bound of $\Lambda\big(m, \mathcal{I}_t(K)\big)$ in case of $d = 2$. We proceed essentially along the same line of Lemma 3.5 with certain modifications.

Lemma A.1. *If $S_t(K)$ is a convex polygon, and*

$$m < \frac{3\sqrt{3}(t^2+4\nu)}{8\pi(t+1)}\nu^{-3/2}, \tag{A.1}$$

then

$$\Lambda(m, \mathcal{I}_t(K)) \le 2. \tag{A.2}$$

Proof. For any polynomial p of degree m, denote by $\widetilde{\boldsymbol{x}} \in S_t(K)$ such that $|p(\widetilde{\boldsymbol{x}})| = \max_{\boldsymbol{x}\in S_t(K)}|p(\boldsymbol{x})|$. If $\widetilde{\boldsymbol{x}} \in \mathcal{I}_t(K)$, then we have $\Lambda\big(m, \mathcal{I}_t(K)\big) = 1$. Otherwise, we denote by $\widetilde{\boldsymbol{x}}_\ell = \arg\min_{\boldsymbol{y}\in\mathcal{I}_t(K)}|\boldsymbol{y} - \widetilde{\boldsymbol{x}}|$. If $\widetilde{\boldsymbol{x}}$ is on the boundary of $S_t(K)$, then $|\widetilde{\boldsymbol{x}}_\ell - \widetilde{\boldsymbol{x}}| \le H_K/2$. By Taylor expansion,

$$p(\widetilde{\boldsymbol{x}}_\ell) = p(\widetilde{\boldsymbol{x}}) + (\widetilde{\boldsymbol{x}}_\ell - \widetilde{\boldsymbol{x}}) \cdot \nabla p(\xi_{\boldsymbol{x}})$$

with $\xi_{\boldsymbol{x}}$ a point on the line with end points $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{x}}_\ell$. This gives

$$|p(\widetilde{\boldsymbol{x}})| \le |p(\widetilde{\boldsymbol{x}}_\ell)| + \frac{H}{2}\max_{\boldsymbol{x}\in S_t(K)}|\nabla p(\boldsymbol{x})|. \tag{A.3}$$

If $\widetilde{\boldsymbol{x}}$ is in the interior of $S_t(K)$, then $\nabla p(\widetilde{\boldsymbol{x}}) = 0$. By Taylor expansion,

$$p(\widetilde{\boldsymbol{x}}_\ell) = p(\widetilde{\boldsymbol{x}}) + \frac{1}{2}(\widetilde{\boldsymbol{x}}_\ell - \widetilde{\boldsymbol{x}})^2 \cdot \nabla^2 p(\xi_{\boldsymbol{x}})$$

with $\xi_{\boldsymbol{x}}$ a point on the line with end points $\widetilde{\boldsymbol{x}}$ and $\widetilde{\boldsymbol{x}}_\ell$. This implies

$$|p(\widetilde{\boldsymbol{x}})| \le |p(\widetilde{\boldsymbol{x}}_\ell)| + \frac{H^2}{2}\max_{\boldsymbol{x}\in S_t(K)}|\nabla^2 p(\boldsymbol{x})|. \tag{A.4}$$

Applying the Markov inequality to (A.3) and (A.4), respectively, we obtain

$$\max_{\boldsymbol{x}\in S_t(K)}|p(\boldsymbol{x})| \le \max_{\boldsymbol{x}\in\mathcal{I}_t(K)}|p(\boldsymbol{x})| + \frac{2m^2H}{w(K)}\max_{\boldsymbol{x}\in S_t(K)}|p(\boldsymbol{x})|,$$

and

$$\max_{\boldsymbol{x}\in S_t(K)}|p(\boldsymbol{x})| \le \max_{\boldsymbol{x}\in\mathcal{I}_t(K)}|p(\boldsymbol{x})| + 2\left(\frac{2m^2H}{w(K)}\right)^2\max_{\boldsymbol{x}\in S_t(K)}|p(\boldsymbol{x})|,$$

respectively.

Since K is a convex polygon, we have

$$p_t(K) \ge 2t\min_{K\in S_t(K)} H_K \ge 2t\min_{K\in S_t(K)}\rho_K.$$

Notice that

$$p_t \le (\#\mathcal{I}_t(K) - \#\mathcal{I}_{t-1}(K))\, H.$$

Combing the above two inequalities, we obtain the recursive relation

$$\#\mathcal{I}_t(K) - \#\mathcal{I}_{t-1}(K) \ge 2t/\nu,$$

which together with $\#\mathcal{I}_0(K) = 3$ gives

$$\#\mathcal{I}_t(K) \ge 3 + t(t+1)/\nu.$$

Next by *Euler's formula*,

$$\#S_t(K) = \#\mathcal{I}_t(K) + \#\mathcal{I}_{t-1}(K) + 2.$$

Combining the above three inequalities, we get

$$\begin{aligned}\#S_t(K) &= \#\mathcal{I}_t(K) - \#\mathcal{I}_{t-1}(K) + 2\#\mathcal{I}_{t-1}(K) + 2\\ &\geq 8 + 2t^2/\nu.\end{aligned}$$

For any vertex $\boldsymbol{x}$ of K, we have $\mathrm{dist}(\boldsymbol{x}, \partial S_t(K)) \leq (t+1)H$, which implies

$$p_t(K) \leq 2\pi(t+1)H.$$

By Finsler-Hadwiger inequality [19], we have

$$\mathrm{mes}K \geq \frac{3\sqrt{3}}{4}\rho_K^2.$$

Combining the above three inequalities and (3.17), we get the following lower bound of the width $w(K)$:

$$w(K) \geq \frac{3\sqrt{3}\left(4 + t^2\nu^{-1}\right)}{2\pi(t+1)\nu} \min_{K\in S_t(K)} \rho_K.$$

Using the condition (A.1), we have

$$\max_{\boldsymbol{x}\in S_t(K)} |p(\boldsymbol{x})| \leq \max_{\boldsymbol{x}\in \mathcal{I}_t(K)} |p(\boldsymbol{x})| + \frac{1}{2} \max_{\boldsymbol{x}\in S_t(K)} |p(\boldsymbol{x})|,$$

which implies (A.2). □

References

1. A. Abdulle, *On a priori error analysis of fully discrete heterogeneous multiscale FEM*, Multiscale Model. Simul. **4** (2005), 447–459.
2. ———, *The finite element heterogeneous multiscale method: a computational strategy for multiscale pdes*, Multiple Scales Problems in Biomathematics, Mechanics, Physics and Numerics, GAKUTO Internat. Ser. Math. Sci. Appl., 31, Gakkotosho, Tokyo, Japan, 2009, pp. 133–181.
3. A. Abdulle and B. Engquist, *Finite element heterogeneous multiscale methods with near optimal computational complexity*, Multiscale Model. Simul. **6** (2007), 1059–1084.
4. A. Abdulle and M.J. Grote, *Finite element heterogeneous multiscale method for the wave equation*, Multiscale Model. Simul. **9** (2011), 766–792.
5. R.A. Adams and J.J.F. Fournier, *Sobolev Spaces*, 2nd eds., Academic Press, New York, 2003.
6. G. Allaire and R. Brizzi, *A multiscale finite element method for numerical homogenization*, Multiscale Model. Simul. **4** (2005), 790–812.
7. I. Babuška, U. Banerjee, and J.E. Osborn, *Survey of meshless and generalized finite element methods: A unified approach*, **12** (2003), 1–125.
8. A Bensoussan, J.L. Lions, and G.C. Papanicolaou, *Asymptotic Analysis for Periodic Structures*, North-Holland, Amsterdam, 1978.
9. A. Berger, R. Scott, and G. Strang, *Approximate boundary conditions in the finite element method*, Symposia Mathematica **X** (1972), 295–313.
10. F. Brezzi, R. Marini, and E. Suli, *Residual-free bubbles for advection-diffusion problems: the general error analysis*, Numer. Math. **85** (1999), 31–47.
11. P.G. Ciarlet, *The Finite Element Method for the Elliptic Problems*, North-Holland, Amsterdam, 1978.

12. D. Coppersmith and T.J. Rivlin, *The growth of polynomials bounded at equally spaced points*, SIAM J. Math. Anal. **23** (1992), 970–983.
13. M. Crouzeix and V. Thomée, *The stability in L^p and $W^{1,p}$ of the L^2-projection onto finite element function spaces*, Math. Comput. **48** (1987), 321–332.
14. R. Du and P.B. Ming, *Convergence of the heterogeneous multiscale finite element method for elliptic problem with nonsmooth microstructures*, Multiscale Model. Simul. **8** (2010), 1770–1783.
15. ———, *Heterogeneous multiscale finite element method with novel numerical integration schemes*, Commun. Math. Sci. **8** (2010), 863–885.
16. W. E and B. Engquist, *The heterogeneous multi-scale methods*, Commun. Math. Sci. **1** (2003), 87–132.
17. W. E, P.B. Ming, and P.W. Zhang, *Analysis of the heterogeneous multiscale method for elliptic homogenization problems*, J. Amer. Math. Soc. **18** (2005), 121–156.
18. B. Engquist, H. Holst, and O. Runborg, *Multiscale methods for the wave equation*, Commun. Math. Sci. **9** (2011), 33–56.
19. P. Finsler and H. Hadwiger, *Einige relationen im dreieck*, Commentarii Mathematici Helvetici **10 (1)** (1937), 316–326.
20. D. Gilbarg and N.S. Trudinger, *Elliptic Partial Differential Equations of Second Order*, Reprint of the 1998 edition, Springer-Verlag Berlin Heidelberg, 2001.
21. A. Gloria, *An analytical framework for the numerical homogenization of monotone elliptic operators and quasiconvex energies*, Multiscale Model. Simul. **5** (2006), 996–1043.
22. ———, *An analytical framework for the numerical homogenization–part II: windowing and oversampling*, Multiscale Model. Simul. **7** (2008), 275–293.
23. ———, *Reduction of the resonance error–part 1: approximation of homogenized coefficients*, Math. Model. Method in Appl. Sci., **21** (2011), 1601–1630.
24. S Hazanov and C. Huet, *Order relationships for boundary conditions effect in heterogeneous bodies smaller than representative volume*, J. Mech. Phys. Solids **42** (1994), 1995–2011.
25. 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), 169–189.
26. V.A. Kondrat'ev, *Boundary value problems for elliptic equations in domains with conical or angular points*, Trans. Moscow Math. Soc. **16** (1967), 227–313.
27. T. Kubota, *Einige ungleischheitsbezichungen über eilinien und eiflähen*, Sci. Rep. of the Töhoku Univ. Ser. (1) **12** (1923), 45–65.
28. P.D. Lax, *Linear Algebra and Its Applications*, Enlarged second edition. Pure and Applied Mathematics (Hoboken). Wiley-Interscience [John Wiley & Sons], Hoboken, NJ, 2007.
29. B. Lipman, F. John, and M. Schechter, *Partial Differential Equations*, Intersciences Publishers, 1964.
30. P.B. Ming and X.Y. Yue, *Numerical methods for multiscale elliptic problems*, J. Comput. Phys. **214** (2006), 421–445.
31. P.B. Ming and P.W. Zhang, *Analysis of the heterogeneous multiscale method for parabolic homogenization problems*, Math. Comp. **76** (2007), 153–177.
32. M.J.D. Powell, *Approximation Theory and Methods*, Cambridge University Press, 1981.
33. P. Rabinowitz and N. Richter, *Perfectly symmetric two-dimensional integration formulas with minimal numbers of points*, Math. Comput. **23** (1969), 765–779.
34. E.A. Rakhmanov, *Bounds for polynomials with a unit discrete norm*, Ann. of Math. (2) **165** (2007), 55–88.
35. L. Reichel, *On polynomial approximation in the uniform norm by the discrete least squares method*, BIT **26** (1986), 349–368.

36. L. Tartar, *H-convergence*, Course Peccot, Collége de France. Partially written by F. Murat. Séminaire d'Analyse Fonctionnelle et Numérique de l'Université d'Alger, 1977–1978, March 1977.
37. V. Venkatakrishnan, *Convergence to steady state solutions of the Euler equations on unstructured grids with limiters*, J. Comput. Phys. **118** (1995), 120–130.
38. P. Šolín, K. Segeth, and I. Doležel, *Higher-Order Finite Element Methods*, Chapman & Hall/CRC, Boca Raton, F.L., 2004.
39. K. Wang, *Numerical simulation for 3d heterogeneous multiscale elliptic equation*, Master thesis, Institute of Computational Mathematics and Scientific/Engineering Computing, Chinese Academy of Sciences, 2010.
40. D.R. Wilhelmsen, *A markov inequality in several dimensions*, J. Approx. Theory **11** (1974), 216–220.
41. X.Y Yue and W. E, *The local microscale problem in the multiscale modeling of strongly heterogeneous media: effects of boundary conditions and cell size*, J. Comput. Phys. **222** (2007), 556–572.
42. Z.M. Zhang and A. Naga, *A new finite element gradient recovery method: superconvergence property*, SIAM J. Sci. Comput. **26** (2005), 1192–1213.
43. O.C. Zienkiewicz and J.Z. Zhu, *The superconvergence patch recovery and a posteriori error estimates. part 1: The recovery technique*, Internat. J. Numer. Methods Engrg. **33** (1992), 1331–1364.
44. T.I. Zohdi and P. Wriggers, *Introduction to Computational Micromechanics*, Springer-Verlag Berlin Heidelberg, 2005.

CAPT, LMAM and School of Mathematical Sciences, Peking University, Beijing 100871, People's Republic of China

E-mail address: rli@math.pku.edu.cn

LSEC, Institute of Computational Mathematics and Scientific/Engineering Computing, AMSS, Chinese Academy of Sciences, No. 55, Zhong-Guan-Cun East Road, Beijing, 100190, People's Republic of China

E-mail address: mpb@lsec.cc.ac.cn

School of Mathematical Sciences, Peking University, Beijing 100871, People's Republic of China

E-mail address: tfy283@126.com