Finite Element Method in Solving Elliptic Problems With Rapidly Oscillating Coefficitents and Its Effective Coefficients

Shengtong Zhang
Tsinghua University
Instructor:
Wenjia Jing Zhongyi Huang

Contents

1 Introduction

In this paper, we mainly discuss the second order elliptic equations with rapidly oscillating coefficients. Such equations ofter arise in materials science when two or more materials are composited together or in flows in porous media. In traditional finite element method, the cost of the memory is huge. Therefore, it is nearly not possible and not efficient to use the traditional way to calculate the solution to the partial differential equations with highly oscillatory coefficients. On the other hand, it is not the properties in small scale that is of the main interest, but is the features of he solution in large scale and the averaged effect of small scales that we care about most. Thus, it is reasonable to use a new way to construct the element that can catch the average properties in solving the original functions which give rise to the use of multiscale finite element method.

In [1], Thomas Y.Hou and its partners come up with the idea to use the multiscale finite element method in solving the elliptic problems with rapidly oscillating coefficients and prove its convergence rate. We mainly follow the way in [1] using multiscale finite element method.

The main idea is to construct finite element base functions from the leading order homogenous elliptic equation. Typically, the scale of the element is larger than the scale of the oscillatory coefficient. By choosing the boundary condition of the base functions properly, the information in smaller scale can be emerged into the base functions. In other words, we substitute the traditional bilinear base functions with the oscillatory base functions, and the small scale information within each element is brought into the large scale solution. The scale of the global stiffness matrix is rather small using the multiscale finite element method, and the large scale solution is computed correctly using the oscillatory base functions.

This paper is organized as follows. In the first part of this paper, we will mainly state the basic homogenization theory in second order partial differential equations and prove the solution to the equations with highly oscillatory coefficients will converge to a equation with constant coefficient in some sense. In the second part, we will state the idea to choose the boundary conditions to the base functions and present the convergence rate of both bilinear base functions and oscillatory base funcitons. We will see that, by choosing the boundary conditions properly, it is possible to eliminate the boundary layer in the first order corrector function, which would give rise to a nice conservative difference structure in the discretization. Thus the rate is acceptable dut to the cancellation of errors, which is proved in details in [1]. In the third part, we calculate a series of effective coefficents according to the theorem in homogenizations. In this procedure, we have to calculate the first order corrector functions first, and the basic finite element method is used in this part with triangular elements. By choosing a series of rapidly oscillating coefficients which is piecewise constant function with different space and shape, we will see the changes of the effective coefficients, which picture the properties of the composite material in large scale.

2 Basic Formulations

In this section, we mainly introduce the Sobolev space and weak derivatives, and define the weak solution to the model problem. For the convenience, the Einstein sumation convention is used : summation is taken over repeated indices.

In many cases, the traditional solution to the partial differential equations does not exist, since the coefficients may not satisfy enough smoothness conditions. Thus we need to extend the definition of the solution to the equation, which is the weak solution. Before giving the rigorous definition of the weak solution, the weak derivative is needed first.

2.1 Weak derivatives

Notation. Let $C_c^\infty(\Omega)$ be the set of infinitely differentiable functions with compact support in Ω. We call a function belonging to $C_c^\infty(\Omega)$ a *test function*.

Definition. Suppose $u, v \in L^1_{\text{loc}}(\Omega)$, and α is a multiindex. $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_n)$, $|\alpha| = \sum_{k=1}^{n} \alpha_k$. We say that v is the α^{th}-weak derivative of u, written

$$D^\alpha u = v,$$

provided

$$\int_\Omega u D^\alpha \phi dx = (-1)^\alpha \int_\Omega v\phi dx$$

for all test functions $\phi \in C_c^\infty(\Omega)$.

The uniqueness of the weak derivatives can be found in [2].

2.2 Sobolev spaces

Now, we introduce the sobolec space $H^k(\Omega)$ equipped with norms and semi-norms:

$$||u||_{k,\Omega} = (\int_\Omega \sum_{|\alpha|\le k} |D^\alpha u|^2)^{\frac{1}{2}},$$

$$|u|_{k,\Omega} = (\int_\Omega \sum_{|\alpha|=k} |D^\alpha u|^2)^{\frac{1}{2}}.$$

And the sobolev space $H^k(\Omega)$ is defined as:

$$H^k(\Omega) = \{u : \Omega \to \mathbb{R} |\, \exists D^\alpha u,\ D^\alpha u \in L^2\}$$

We denote by $H_0^k(\Omega)$ the closure of $C_c^\infty(\Omega)$ in $H^k(\Omega)$.

In the following, we always assume Ω is a unit square in $\mathbb{R}^2$ unless mentioned specially and the derivaties are all weak derivatives.

2.3 Weak solution

We call u is the weak solution to the equation

$$\begin{cases} \nabla \cdot (A(x)\nabla u) = f & \text{in } \Omega \\ u = 0 & \text{on } \partial\Omega \end{cases}$$

if it satisfies:

$$\int_\Omega A(x)\nabla u \nabla \phi = \int_\Omega f\phi \qquad \forall \phi \in H_0^1(\Omega).$$

The existence and uniqueness of weak solution derive from Lax-Milgram Theorem. And this equation is our principal concern.

2.4 Model problem and basic idea of multiscale method.

Our elliptic model problem is:

$$L_\epsilon u_\epsilon = f \text{ in } \Omega, \qquad u_\epsilon = 0 \qquad \text{on } \partial\Omega, \tag{2.1}$$

where

$$L_\epsilon = \nabla \cdot A(x/\epsilon)\nabla$$

is the linear elliptic operatpr, ϵ is a small parameter, and $a(x) = (a_{ij}(x))$ is symmetric and satisfies the elliptic conditions. More specifically, $\alpha|\xi|^2 \le \xi_i a_{ij}\xi_j \le \beta|\xi|^2$, for all $\xi \in \mathbb{R}^2$ and with $0 < \alpha < \beta$. Furthermore, $a_{ij}(\mathbf{y})$ are periodic functions in $\mathbf{y}$ in a unit square Y and $a_{ij}(\mathbf{y}) \in W^{1,p}(Y)$ $(P > 2)$. We use u instead of u_ϵ for simplicity, but the solution u depends on ϵ.

Due to the definition of weak solution, problem (2.1) equals to the variational problem

$$a(u, v) = f(v), \qquad \forall v \in H_0^1(\Omega), \tag{2.2}$$

where

$$a(u, v) = \int_\Omega A_{ij}\frac{\partial u}{\partial x_j}\frac{\partial v}{\partial x_i}dx \quad \text{and} \quad f(v) = \int_\Omega fv dx.$$

Therefore, the bilinear functional $a(\cdot,\cdot)$ is elliptic and continuous, i.e.,

$$\alpha|v|_{1,\Omega}^2 \le a(v, v), \qquad \forall v \in H_0^1, \tag{2.3}$$

and

$$|a(u,v)| \leq \beta |u|_{1,\Omega}^2 |v|_{1,\Omega}^2, \qquad \forall u,v \in H_0^1. \tag{2.4}$$

The idea of finite element method is to restrict the weak formulation (2.2) to a finite dimensional subspace of $H_0^1(\Omega)$. Let K^h be a partition of Ω by elements K with diameter less than h. In each element $K \in K^h$, we define a set of nodal basis $\{\phi_K^i,\ i = 1, \cdots, d\}$ with d being the number of nodes of the element. The idea of the multiscale method in [1] is to choose ϕ^i to satisfy

$$L_\epsilon \phi^i = 0 \qquad \text{in } K \in K^h. \tag{2.5}$$

$x_j \in \overline{K}(j = 1, \cdots, d)$ denotes the nodal points of K. We require $\phi^i(x_j) = \delta_{ij}$. Our goal is to specify the boundary condition of ϕ^i to make (2.5) well-posed. Moreover, we assume that the base functions are continuous across the boundaries of the elements, so that

$$V^h := span\{\phi_k^i : i = 1, \cdots, d;\ K \in K^h\} \subset H_0^1(\Omega)$$

Now, we only need to solve the approximation solution of (2.2) in V^h, i.e., find $u^h \in V^h$ such that

$$a(u^h, v) = f(v), \qquad \forall v \in V^h \tag{2.6}$$

Remark 2.1. In the following, we solve (2.2) using rectangle elements. Thus, the number of nodal points $d = 4$.

3 Homogenization theory in elliptic equations

3.1 two-scale expansion

Homogenization theory studies the effetive quality of solutions to highly oscillatory coefficients equations. In our model problem, there exists two scales: a macroscopic scale of order 1 and a microscopic scale of order ϵ. For fixed small scale $\epsilon > 0$, the solution u^ϵ will be rather complicated and have different behaviors on the two length scales.

Homogenization theory studies the limiting behavior $u^\epsilon \to u_0$ as $\epsilon \to 0$, which is also known as "H-convergence".

Let us first assume $u^\epsilon \to u_0$ as $\epsilon \to 0$ in a reasonable sense, and try to determine the equation which u satisfies. Here we use the trick of *two-scale expansion*. Suppose

$$u^\epsilon(x) = u_0(x, x/\epsilon) + \epsilon u_1(x, x/\epsilon) + \epsilon^2 u_2(x, x/\epsilon) + \cdots, \tag{3.1}$$

where $u_i : T \to \mathbb{R}$ $(i = 0, 1, \cdots),\ u_i = u_i(x, y)$. Here T denotes the unit square in $\mathbb{R}^2$ with periodic structure, which can also been treated as a torus. We also need u_i be 1-periodic in T.

Now, we plug this ansatz into the original equation (2.1). Since we double the variables, $y = x/\epsilon$, we have

$$\nabla = \nabla_x + \frac{1}{\epsilon}\nabla_y.$$

And (2.1) becomes

$$-(\nabla_x + \frac{1}{\epsilon}\nabla_y) \cdot (A(y)(\nabla_x + \frac{1}{\epsilon}\nabla_y)u^\epsilon) = f. \tag{3.2}$$

Equating the same powers of ϵ, we deduce

$$\nabla_y \cdot (A(y)\nabla_y u^0(x,y)) = 0 \qquad \text{for } O(\epsilon^{-2}). \tag{3.3}$$

By the uniqueness of the equation, we get

$$u_0(x,y) = u_0(x). \tag{3.4}$$

Do the same thing to ϵ^{-1} and we deduce

$$-\nabla_y \cdot (A(y)\nabla_y u_1) = \nabla_y \cdot (\nabla_x u_0(x)). \tag{3.5}$$

We can seperate variables to represent u_1 more simply. For $i = 1, 2, \cdots, n$, let $\chi_i = \chi_i(y)$ solves

$$-\nabla_y \cdot (A(y)\nabla\chi_k(y)) = \nabla_y \cdot (A(y)e_k) \qquad k = 1, \cdots, n. \tag{3.6}$$

According to the *Fredholem alternative theory*, the solution $\chi_k,\ k = 1, \cdots, n$ exist and unique up to an additive constant. What's more, the solution χ_k are also 1-periodic.

We call the equation (3.6) is *cell problem* to the model equation and χ_k are called *corrector function*. Under the condition

$$\int_T \chi_k(y) = 0, \qquad k = 1, \cdots, n,$$

cell problem has only one solution. Therefore, we can represent u_1 as

$$u_1(x,t) = \partial_k u_0(x)\chi_k(y)$$

.

Equating the zero-th term of ϵ, and take the integral in T, we deduce

$$\begin{cases} -\overline{A}_{ij} u_{x_i x_j} = f & \text{in } T \\ u = 0 & \text{on } \partial T \end{cases} \tag{3.7}$$

where

$$\overline{A}_{ij} := \int_T A_{ij}(y) - A_{ik}(y)\frac{\partial \chi_j(y)}{\partial y_k}dy, \qquad 1 \le i,j \le n \tag{3.8}$$

are the *homogenized coefficients* and $\chi_j,\ j = 1, \cdots, n$ solves the *cell problem* (3.6). We now also need to prove (3.7) is also a elliptic equation, which means we need to prove $\overline{A} = (\overline{A}_{ij})$ is positive definite. This can be seen from the weak formula of cell problem (3.6). Here we take the test function $\chi_i,\ 1 \le i \le n$. And we have

$$\int_\Omega (A(y)\nabla(\chi_j(y) + y_j) \cdot \nabla\chi_i(y) = 0$$

Therefore, the homogenized coefficients $\overline{A}_{ij}$ have the symmetric form:

$$\begin{array}{rcl} \overline{A}_{ij} & = & \int_T A_{ij}(y) - A_{ik}(y)\frac{\partial \chi_j(y)}{\partial y_k}dy \\ & = & \int_T (A(y)\nabla(\chi_j + y_j)) \cdot \nabla y_i \\ & = & \int_T (A(y)\nabla(\chi_j + y_j)) \cdot \nabla(\chi_i + y_i). \end{array} \tag{3.9}$$

And the ellipticity property can be easily seen from this symmetric form of the homogenized coefficents. Thus, the limit solution u_0 solves a second-order constant efficient elliptic problem (3.7).

Example. In a special case of dimension one, the homogenized coefficients $\overline{A}$ can be written analytically. In this case, the cell problem (3.6) becomes

$$-\frac{d}{dx}(A(y)(1 + \chi')) = 0 \qquad \text{in } T. \tag{3.10}$$

We assume

$$c = A(y)(1 + \chi').$$

Since χ is periodic in $T = [0, 1]$, we get

$$c = \frac{1}{\int_0^1 \frac{1}{A}dy}.$$

Thus

$$\overline{A} = \int_0^1 A(y)(1 + \chi')dy = c = \frac{1}{\int_0^1 \frac{1}{A}dy}. \tag{3.11}$$

3.2 Oscillating test function method

Suppose Ω is a bounded open set, u^ϵ satisfies (2.2). By the ellipticity of coefficients $A(y)$ and the Poincaré's inequality, we have a priori bound

$$||\nabla u^\epsilon||_{L^2(\Omega)} \le C||f||_{L^2(\Omega)}.$$

Using Poincaré's inequality again, we get

$$||u||_{1,\Omega} \le C'||f||_{L^2(\Omega)}.$$

By the Rellich's Theorem that $H^1(\Omega)$ can be compactly injected into $L^2(\Omega)$, up to a subsequence, we get the convergence:

$$\begin{aligned} \nabla u^{\epsilon_k} &\rightharpoonup \nabla u_0 && \text{weakly in } L^2(\Omega), \\ u^{\epsilon_k} &\to u_0 && \text{strongly in } L^2(\Omega), \\ \xi^{\epsilon_k} := a(x/\epsilon_k)\nabla u^{\epsilon_k} &\rightharpoonup \xi_0 && \text{weakly in } L^2(\Omega). \end{aligned} \tag{3.12}$$

Passing to the limit in (2.2), we get for all $v \in H_0^1(\Omega)$,

$$\int_\Omega \xi_0 \cdot \nabla v = \int_\Omega fv. \tag{3.13}$$

Now we take a specific test function in (2.2), namely *oscillating test function.* Let $v^\epsilon = v^\epsilon(x) := x_k + \epsilon\chi_k(x/\epsilon)$. For $\phi \in C_0^\infty(\Omega)$, we take $v = \phi v^\epsilon$. From (3.6), i.e. the cell problem, we have the following observation:

$$\nabla \cdot (A(x/\epsilon)\nabla v^\epsilon(x)) = \nabla \cdot (A(x/\epsilon)(e_k + \nabla\xi_k(x/\epsilon)) = 0. \tag{3.14}$$

Substitute the test function v in (2.2) with ϕ^ϵ, we get

$$\int_\Omega A(x/\epsilon)\nabla u^\epsilon \cdot \nabla\phi v^\epsilon + \int_\Omega A(x/\epsilon)\nabla u^\epsilon \cdot \nabla v^\epsilon \phi = \int_\Omega f\phi v^\epsilon. \tag{3.15}$$

Since the corrector function χ_k is bounded in Ω, we can easily pass to the limit in the right hand side of the previous equation

$$\int_\Omega f\phi v^\epsilon \to \int_\Omega f\phi x_k.$$

By the strong convergence of v_ϵ and the weak convergence of $A(x/\epsilon)\nabla u^\epsilon$, we get

$$\int_\Omega A(x/\epsilon)\nabla u^\epsilon \cdot \nabla\phi v^\epsilon \to \int_\Omega \xi_0 \cdot \nabla\phi x_k.$$

Now we only need to pass to the limit of the second term of (3.15) in the left hand side. According to (3.14), we choose the test function $u^\epsilon\phi$, and we have the weak formula

$$\int_\Omega A(x/\epsilon)\nabla v^\epsilon \cdot \nabla(u^\epsilon\phi) = 0.$$

Thus,

$$\int_\Omega A(x/\epsilon)\nabla u^\epsilon \cdot \nabla v^\epsilon \phi = -\int_\Omega A(x/\epsilon)u^\epsilon\nabla v^\epsilon \cdot \nabla\phi.$$

Moreover,

$$A(x/\epsilon)\nabla v^\epsilon = A(x/\epsilon)(e_k + \epsilon\nabla\chi_k(x/\epsilon)) \to \int_\Omega A(x)(e_k + \nabla\chi_k(y))dy = \overline{A}e_k.$$

Here $\Omega = T$, and the convergence above is in the $L^2(T)$ sense.

Now we can take (3.15) to the limit and get

$$\int_\Omega \xi_0 \cdot \nabla\phi x_k + \int_\Omega \overline{A}ue_k \cdot \nabla\phi = \int_\Omega f\phi x_k. \tag{3.16}$$

And we take $v = x_k\phi$ in (3.13), we get

$$\int_\Omega f\phi x_k = \int_\Omega \xi_0 \cdot (e_k\phi + x_k\nabla\phi).$$

Thus,

$$\int_\Omega \overline{A}\nabla u \cdot e_k\phi = \int_\Omega \overline{A}ue_k \cdot \nabla\phi = \int_\Omega \xi_0 \cdot e_k\phi. \tag{3.17}$$

Now we conclude that

$$\xi_0 = \overline{A}u. \tag{3.18}$$

And according to (3.9), $\overline{A}$ is elliptic. For now we have proved the following theorem.

Theorem. The family of solutions u^ϵ converges weakly in $H_0^1(\Omega)$ and strongly in $L^2(\Omega)$ to the solution u_0 of the problem

$$\begin{cases} \nabla \cdot \overline{A}\nabla u_0 = f, & x \in \Omega, \\ u_0 = 0, & x \in \partial\Omega. \end{cases} \tag{3.19}$$

There is also a proof of error estimate of the convergence in [3]. It can be proved that the remainder

$$r^\epsilon(x) := u^\epsilon(x) - u_0(x) - \epsilon\chi(x/\epsilon) \cdot \nabla u_0(x).$$

is of order one of the scale parameter ϵ. In other words, there exists a constant C, independent of u_0 and ϵ, such that

$$||u^\epsilon - u_0||_{0,T} \leq C\epsilon||u_0||_{2,T}. \tag{3.20}$$

4 Convergence of multiscale method

Subtracting (2.6) from (2.2), we get the prthogonality property

$$a(u - u^h, v) = 0, \qquad \forall v \in V^h. \tag{4.1}$$

4.1 Convergence for $h < \epsilon$

Let u_l be the standard bilinear interpolant of u in $K \in K^h$. From traditional finite element method, we have the estimates

$$\begin{aligned} ||u - u_l||_{0,K} &\leq C_1 h^2 |u|_{2,K}, \\ ||u - u_l||_{1,K} &\leq C_2 h |u|_{2,K}. \end{aligned} \tag{4.2}$$

According *Theorem 4.4* in [1], suppose u and u_h be the solution of (2.1) and (2.6) respectively. Then there exist constants $C_1, C_2 > 0$ independent of h, such that

$$\begin{aligned} ||u - u_h||_{0,\Omega} &\leq C_1 (h/\epsilon)^2 ||f||_{0,Omega}, \\ ||u - u_h||_{1,\Omega} &\leq C_2 (h/\epsilon) ||f||_{0,\Omega}. \end{aligned} \tag{4.3}$$

For $h < \epsilon$, the situation is similar to the normal finite element method. We can use a linear boundary condition for the base function ϕ^i. And the L^2 norm of the remainder $u - u^h$ is of order two of the relative scale parameter h/ϵ.

4.2 Convergence for $h > \epsilon$

In the section above, we choose the linear boundary condition for the base functionϕ^i. It behaves similar to the standard finite element method. But in the situation of $h > \epsilon$, the two methods behave very differently, especially in the case as $\epsilon \to 0$. This is because on each side of ∂K, the coefficients have oscillations. Thus the linear boundary condition cannot characterize the oscillation of the solution any longer. We need to choose a diffrent boundary condition, which is the oscillatory condition.

The basic idea is to solve the reduced elliptic problems on each side of ∂K with boundary condition 1 and 0 at the two end points, and use the resulting solution as the boundary condition for the base function ϕ^i. The reduced problems are obtained from (2.5) by setting one variable as constant, and solve the 1-dimension problem

$$L_\epsilon \phi^i = 0$$

with the boundary condition 1 and 0 at the two end points of ∂K. We call such boundary conditions for ϕ oscillatory boundary conditions. In case of a being a seperable in space, i.e. $a(\mathbf{x}) = a_1(x/\epsilon)a_2(y/\epsilon)$, the problem can be reduced to two 1-dimension problem, and the base function ϕ^i with oscillatory boundary conditions can be computed analytically by forming a tensor product. Let μ^i be the oscillatory boundary conditions for ϕ^i. Thus $\sum_{i=1}^{d} \mu^i = 1$ on ∂K, and hence

$$\sum_{i=1}^{d} \phi_K^i = 1 \qquad \forall K \in K^h. \tag{4.4}$$

According to *theorem 5.1* in [1], we have the following error estimate:

Theorem. Let u and u^h be the solutions of (2.1) and (2.6) respectively. Then there exist constants C_1 and C_2, independent of ϵ and h, such that

$$||u - u^h||_{1,\Omega} \leq C_1 h ||f||_{0,\Omega} + C_2(\epsilon/h)^{\frac{1}{2}}. \tag{4.5}$$

From the theorem above, we can see that the H^1 norm of the error is bounded by the grid parameter h and the relative scale parameter ϵ/h. In the case that ϵ/h is constant, the error is of order h.

And we also have the L^2 estimates using the Aubin-Nitsche trick.

$$||u - u^h||_{0,\Omega} \leq C_1 h^2 ||f||_{0,\Omega} + C_2(\epsilon/h)^{\frac{1}{2}}. \tag{4.6}$$

Note that the order is h^2 if ϵ/h is constant. But the convergence is still dominated by the $\sqrt{\epsilon/h}$ term.

Remark. The error estimates can be improved further in 1-D and 2-D problem by the error cancellation in [1].

5 Algorithm of multiscale method

We use the *example 7.1* in [1]. The coefficient of the problem is a diagonal matrix with

$$A(\mathbf{x}/\epsilon) = \frac{1}{2 + Psin(2\pi(x-y)/\epsilon)}. \tag{5.1}$$

Here we take $P = 1.8$, and the right hand side function $f(x, y)$ is given by

$$f(x,y) = -\frac{1}{2}[(6x^2 - 1)(y^4 - y^2) + (6y^2 - 1)(x^4 - x^2)]. \tag{5.2}$$

We impose $u = 0$ on $\partial\Omega$. In such case, the effective coefficients $\overline{A}$ is a full 2×2 matrix.

In the first step, we shall solve the base function ϕ^i. Here we use the standard finite element method, and the elements are all identical with boundaries parallel to the axises. ϕ^i has either linear boundary or oscillatory boundary, which can be calculated analytically. We denote μ_i be the boundary of ϕ^i. According to (2.5), the base function ϕ^i satisfies the equation below with Dirichelet boundary condition.

$$-\nabla \cdot (A(\mathbf{x}/\epsilon)\nabla\phi^i) = 0. \qquad \text{in } \Omega \tag{5.3}$$

$$\phi^i = \mu_i \qquad \text{on } \partial\Omega. \tag{5.4}$$

We subtract μ_i from ϕ^i and the function become homogeneous. Then we can apply the finite element method and get the base function
phi^i

The second step is using the finite element method solving the problem (2.2) with the base function ϕ_i. In this case, the boundary of u is 0. We only need to calculate the value on the nodal points inside Ω.

We use the standard finite element method with mesh size less than 1/10 of the mesh size in our multiscale method. Then we solve it again on he grid with the mesh size half of the first one. We apply the Richardson extrapolation to get a reference solution on the coarse grid. Now we can calculate the error with different norms. the numerical results can be found in [1].

Let N be the number of elements in the x and y directions, and M is the number of elements in each cell. Thus there are totally $M \times M$ subelements in each element. $h = \frac{1}{N}$ be the size of mesh.

First we take $M = 8$ and $h < \epsilon$. We can use the multiscale method to calculate the error. Here we use MFEM-L as a short of multiscale finite element method with linear boundary condition, MFEM-L as a short of multiscale finite element method with oscillatory boundary condition, and LFEM as the linear finite element method.

ϵ	N	MFEM-O		MFEM-L		LFEM	
		$\|E\|_{l^2}$	rate	$\|E\|_{l^2}$	rate	$\|E\|_{l^2}$	rate
0.08	64	5.60e-4		6.90e-5		2.55e-4	
	128	2.32e-4	1.3	1.58e-5	2.1	6.65e-5	1.9
	256	7.11e-5	1.7	3.58e-6	2.1	1.66e-5	2.0
	512	1.87e-5	1.9	7.09e-7	2.3	3.92e-6	2.1
0.04	128	5.82e-4		5.65e-5		2.39e-4	
	256	2.39e-4	1.3	1.23e-5	2.2	6.20e-5	1.9
	512	7.33e-5	1.7	2.71e-6	2.2	1.55e-5	2.0
0.02	256	5.92e-4		5.10e-5		2.32e-4	
	512	2.42e-4	1.3	1.08e-5	2.2	5.98e-5	2.0
	1024	7.42e-5	1.7	2.32e-6	2.2	1.50e-5	2.0

Then we fix $\epsilon = 0.005$ small enough to see the rates of the error.

Mesh		MFEM-O			MFEM-L		
N	M	$\|E\|_\infty$	$\|E\|_{l^2}$	rate	$\|E\|_\infty$	$\|E\|_{l^2}$	rate
32	64	2.01e-4	8.29e-5		2.73e-4	1.15e-4	
64	32	5.07e-4	2.07e-4	-1.3	5.81e-4	2.44e-4	-1.1
128	16	9.38e-4	3.87e-4	-0.90	1.12e-3	4.74e-4	-0.96
256	8	2.20e-3	9.13e-4	-1.2	2.02e-3	8.63e-4	-0.86

Moreover, if we fix the ratio of ϵ and h, we can see the error is at the same order. Here we take $M = 32$.

N	ϵ	MFEM-O		MFEM-L		LFEM	
		$\|E\|_{l^2}$	rate	$\|E\|_{l^2}$	rate	MN	$\|E\|_{l^2}$
8	0.04	1.30e-4		1.36e-4		256	6.20e-5
16	0.02	1.63e-4	-0.33	1.98e-4	-0.54	512	5.98e-5
32	0.01	1.94e-4	-0.25	2.31e-4	-0.22	1024	5.88e-5
64	0.005	2.07e-4	-0.09	2.44e-4	-0.08	2048	5.83e-5

6 The effective coefficients

6.1 Effective coefficients on unit square

We have introduced a multiscale method to solve the elliptical equation with highly osillatory coefficients. Due to the homogenization theory, the solution u^ϵ will converge to u, which solves a constant elliptical partial differential equation with coefficients $\overline{A}$. The coefficient $\overline{A}$ is defined by (3.8)

$$\overline{A}_{ij} := \int_T a_{ij}(y) - a_{ik}(y)\frac{\partial \chi_j(y)}{\partial y_k}dy, \qquad 1 \le i, j \le n \tag{6.1}$$

Here χ_j is the corrector function solving the cell problem (3.6).

$$-\nabla_y \cdot (A(y)\nabla\chi_k(y)) = \nabla_y \cdot (A(y)e_k) \qquad k = 1, \cdots, n. \tag{6.2}$$

Notice that the cell problem containing no coefficients with ϵ scale. Thus, we can solve the corrector function χ_i using standard finite element method. Then, we can get the effective coefficients $\overline{A}$. But $\overline{A}$ is not a simple integral average of the coefficients a. It contains the term

$$\int_T a_{ik}(y)\frac{\partial \chi_j(y)}{\partial y_k}dy.$$

Different coefficient a may result in different effective coefficients. In the real world, we always take a being diagonal matrix, which means

$$A(\mathbf{y}) = A(y)I.$$

But the effective coeffciets $\overline{A}$ does not neccesarily be the diagonal. In the case of (5.1), $\overline{A}$ is a full 2×2 matrix. Thus we cannot take it for granted that $\overline{A}$ is diagonal. We have to calculate each element of $\overline{A}$ to determine its diagonal property.

In the following case, we mainly concentrate on the case that a is piecewise constant, which is exactly the case that two isotropy materials composite together.

The coefficients of the both materials differ and our solution still solves the equation (2.1) with $\mathbb{Z}$-periodicity. More specifically, we choose

$$A(y) = \begin{cases} 1, & y \in T/\Omega, \\ 2, & y \in \Omega, \end{cases} \tag{6.3}$$

where $\Omega \subset T$, and T denotes the torus $[0,1] \times [0,1]$.

For simplicity, we set Ω center at the certer of T, i.e., $(0.5, 0.5)$. We change the shape and the diameter of Ω to check the influence on the effective coefficients $\overline{A}$.

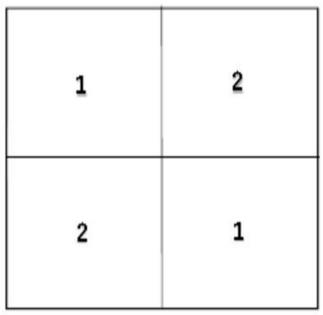
Due to the discontinuity of A on the $\partial\Omega$, the corrector function χ_k, $(k = 1, 2)$ has interior layer near $\partial\Omega$. In the region away from $\partial\Omega$, χ_i solves a Poisson equation. Under the restriction of periodicity and the zero integral average of χ_i, we have $\chi_i = 0$ in those regions. It is the interior layer of χ_i that contributes to the great value of gradient of χ_i. The definition of $\overline{A}$ contains the term of $\nabla\chi_i$. Thus the interior layer of χ_i has great influence on the effective coefficients.

Notice that in each example below, we use the finite element method on the triangle elements constructed accordingly with periodic boundary condition.

Example 1. Let's consider a simple version of coefficient. Suppose

$$A(x, y) = \begin{cases} 2 & 0 \leq x, y \leq 0.5 \text{ or } 0.5 < x, y < 1; \\ 1 & \text{other.} \end{cases} \tag{6.4}$$

The coefficients can be seen intuitively in the picture below.



Notice that the region T is a big square, and it is divided into four identical small squares. Thus, it is of quite convenience to use rectangle element in finite element method. The corrector functions χ^i, $(i = 1, 2)$ is shown in the two pictures below. The left one is χ^1, and the right one is χ^2. Here we choose the length of the mesh grids to be $\frac{1}{64}$.

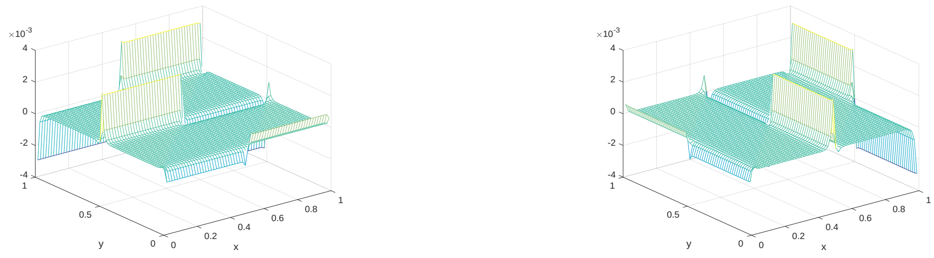


Figure 1: Corrector functions

As we can see from the picture, the corrector function has interior layer near the boundary of each small triangle, which is also the area that the coefficient is discontinuous. This is the reason why $\nabla\chi^i$ is not zero in some region. It results in that the effective coefficient $\overline{A}$ may not be the simple integral average of a. But in this simple case, the effective coefficient converges to a, which can be seen from the table below. Here $\overline{A}$ is a 2×2 matrix.

n	$\overline{A}$	
64	1.4980e+000	295.5840e-006
	295.5840e-006	1.4980e+000
128	1.4990e+000	73.8960e-006
	73.8960e-006	1.4990e+000
256	1.4995e+000	18.4740e-006
	18.4740e-006	1.4995e+000

Table 1: effective coefficient

Here n denotes the number of element in one direction. As we can see from the table, the effective coefficient $\overline{A}$ is still a diagonal matrix and the diagonal elements converge to the integral average of a, which is 1.5 in this case. And the nondiagonal elements converge to zero. This is partly of the reason that χ^i is symmetric to the center of the big square, which can be seen from figure 1 directively.

Example 2. Now we consider a more difficult case that the region Ω is no longer the square parallel to the axis. First we consider the case that Ω is a rhombus which has the same center with the unit square T. Let r denote the distance between the center of Ω and the vertex. The coefficient has been displayed in figure 2.

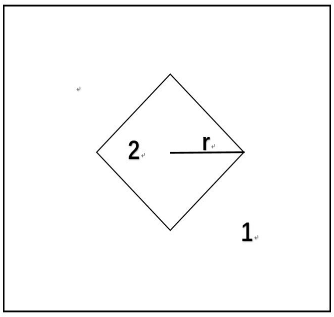


Figure 2: rhombus

Since the boundary of Ω is no longer parallel to the axises, the error would be nonnegligible if we still use the rectangle element since the integral on the element which cross the boundary of Ω would not converge. Thus we choose to use the triangle element which would not cross the boundary. In the first step, we do the triangulation on T. Then at each step, we take the midpoints of each side, and divide the triangle into four small triangles. Let N be the number of steps to refine the grid. We take $r = 0.25$ as an example to display the original triangulation.

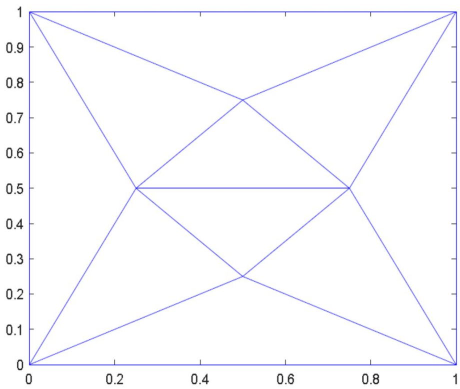


Figure 3: $N = 0$

The corrector function is displayed below. We did some rotations to best interpret the quality of χ^i.

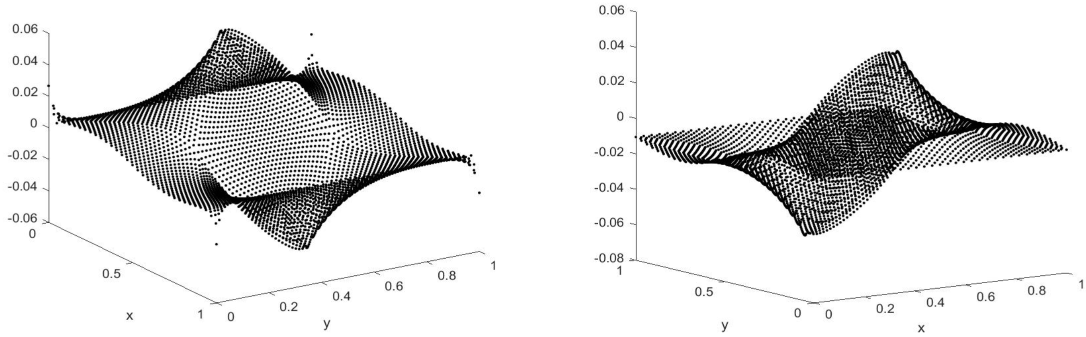


Figure 4: Left: χ^1; right: χ^2.

As we can see from the image above, χ^i has a great change of value near the boundary of Ω. We can see the shape of rhombus clearly through the image of χ^i. Under this circumstances, the gradient of χ^i is nonzero and has contribution to the effective coefficient $\overline{A}$. We take N large enough that every edge of the triangulation is less than 0.01. And we change the value of r keeping the rhombus is strictly inside the unit square T. And the table below is our numerical results. Denote $< A >$ be the integral average of the original coefficient A.

r	$\overline{A}$		$\overline{A}- < A >$	
0.1	1.012981847	-1.17558E-17	-0.007018153	-1.17558E-17
	1.1994E-18	1.013357598	1.1994E-18	-0.006642402
0.2	1.055526484	-6.27567E-17	-0.024473516	-6.27567E-17
	-6.10812E-17	1.055917441	-6.10812E-17	-0.024082559
0.3	1.129722511	-4.08012E-17	-0.050277489	-4.08012E-17
	-6.00648E-17	1.130139234	-6.00648E-17	-0.049860766
0.4	1.242117867	-1.44853E-16	-0.077882133	-1.44853E-16
	-1.64636E-16	1.242574581	-1.64636E-16	-0.077425419

Table 2: effective coefficient on rhombus

From the table above, we can see that the effective coeffcients is still diagonal, but the diagonal elements are no longer the simple integral average of a. We can see this from the diagonal value of $\overline{A}- < A >$, and it will decrease along the increase of r. It means that the larger the area of Ω is, the smaller the error between $\overline{A}$ and $<A>$ is. Moreover, we notice that the error between $\overline{A}(1,1)$ and $\overline{A}(2,2)$ is less than 0.01. Thus we can say that the effective coefficient also hold an isotropy type, i.e.

$$\overline{A}(\mathbf{y}) = \overline{A}(y)I. \tag{6.5}$$

Example 3. Let us consider a even more complicated case that Ω is a circle centered at the certer of the unit square T. Since the boundary of Ω is not a straight line, there must be some triangle element crossing the boundary. In this case, we take a different way to construct the elements. We employ the function "distmesh2D" in Matlab to construct uniform triangle elements. The image below is an example with the maximum length of each element less than 0.1.

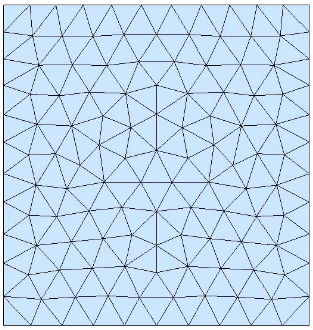


Figure 5: Uniform triangle element

Notice that each triangle element has almost the same length on each edge. So there would not be any triangle with badly acute angle, which makes the error smaller.

The next step in to delete the nodal points on the boundary of the square and substitute them with equally spaced nodes. Under the periodic setting, we know that the corrector function has the same value on the nodes at the opposite position on the boundary of the unit square. Then we number these points again and apply the basic finite element method to find the value of the corrector function χ^i. And finally calculate the effective coefficient $\overline{A}$. The corrector function is displayed below. Here we take $r = 0.25$ as an example.

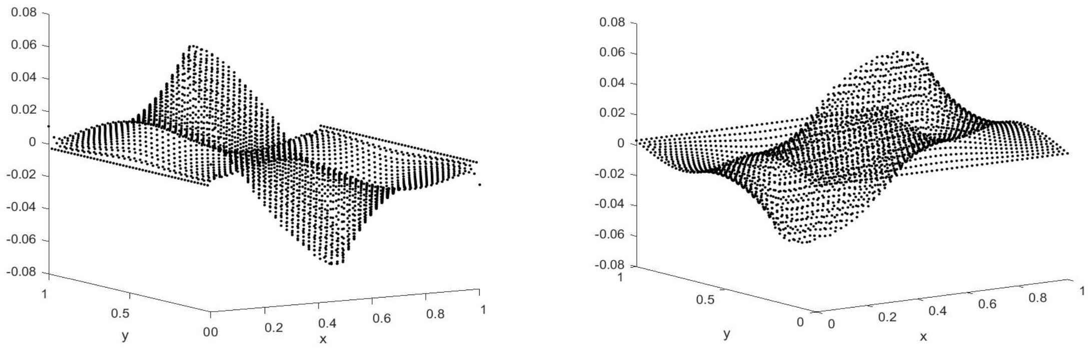


Figure 6: Left: χ^1; right: χ^2.

As shown on the image above, the corrector function has a clear interior layer on the boundary of Ω. We can see a clear shape of circle inside the unit square. On the region away from the boundary of Ω, the corrector function tends to be zero. Thus the result of the corrector function χ^i is credible. We take different radius r such that the region Ω is strictly contained in T. The numerical result of the effective coefficients are displayed at the following form.

r	$\overline{A}$		$\overline{A} - < A >$	
0.1	1.021164537	1.02363E-13	-0.01025139	1.02363E-13
	1.02364E-13	1.0213861	1.02364E-13	-0.010029827
0.2	1.087469302	8.97449E-13	-0.038194404	8.97449E-13
	8.97441E-13	1.087772944	8.97441E-13	-0.037890762
0.3	1.208486779	3.12725E-12	-0.07425656	3.12725E-12
	3.12721E-12	1.2087199	3.12721E-12	-0.074023439
0.4	1.404399579	4.13826E-12	-0.098255246	4.13826E-12
	4.13834E-12	1.404481935	4.13834E-12	-0.09817289

Table 3: effective coefficient on circle

We can see clearly from the table above that the effective coefficient $\overline{A}$ also holds an isotropy type. Due to the affection of $\nabla\chi^i$, the effective coefficient $\overline{A}$ differs from the integral average $< A >$. And the value of the error between them will decrease along with the increase of the radius r. But this is not a simple linear relationship. If we discretize the radius r thinner in $(0, 0.5)$, we can draw a image of the relationship between r and $e(1,1), e(2,2)$. Here e denotes the error between $\overline{A}$ and $< A >$, i.e.

$$e = \overline{A} - < A > . \tag{6.6}$$

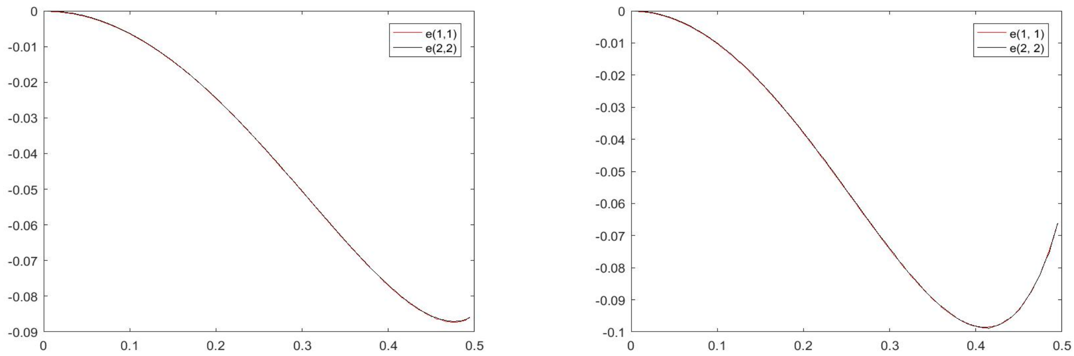
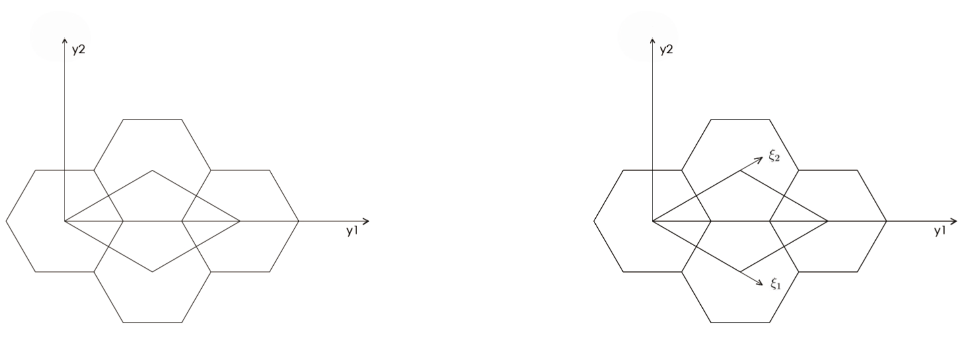


Figure 7: Left: rhombus; right: circle.

As shown on the figure 7, the diagonal elements of e are also identical, and when r is rather small or large, the error is not proportional to r. Thus they do not hold a linear relationship. But in the region near 0.25, the error is nearly proportional to r, and in most region, the error will decrease along the increase of r.

6.2 Effective coefficients on honeycomb

In crystal or other regular material, we often come across the honeycomb region, which have slightly difference with the normal square region. Now we extend our sight to the honeycomb problem. In this section, the region T denotes a regular hexagon. Ω is also a circle centered at the center of the hexagon. We choose the center of the hexagon to be the origion. Firstly we need to specify the cell problem in this setting and then transform it to the problem we have studied before. The region is shown in the picture below.



If we connect the center of four adjacent hexagon and construct a rhombus, we can see the corrector function is periodic on this rhombus. Thus we can define our cell problem on it. Let T_0 be the rhombus shown above. We need to deduce the equation of the corrector function. The main idea is transformation of coordinates.

Remind the model problem is (2.1). For $\mathbf{y} = (y_1, y_2)$,

$$-\nabla_{\mathbf{y}} \cdot (A(\mathbf{y}/\epsilon)\nabla_{\mathbf{y}} u^{\epsilon}(\mathbf{y})) = f(\mathbf{y}). \tag{6.7}$$

Here we take the two sides of T_0 being the new axises: ξ_1, ξ_2. Suppose

$$\xi = Jy.$$

Here J is the Jacobi matrix from ξ to y. Specifically

$$J = \begin{pmatrix} \frac{\sqrt{3}}{2} & \frac{\sqrt{3}}{2} \\ -\frac{1}{2} & \frac{1}{2} \end{pmatrix}. \tag{6.8}$$

Thus, the function (6.7) becomes to

$$-\nabla_{\xi} \cdot [(J^T A(J^{-1}\xi/\epsilon)J)\nabla_{\xi} u^{\epsilon}(J^{-1}\xi)] = f(J^{-1}\xi). \tag{6.9}$$

Denote

$$A'(\xi) = J^T A(J^{-1}\xi/\epsilon)J \tag{6.10}$$
$$v^{\epsilon}(\xi) = u^{\epsilon}(J^{-1}\xi) \tag{6.11}$$

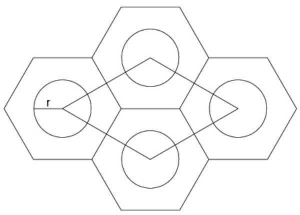
and we have the conclusion that v^{ϵ} solve the same model problem with u^{ϵ}.

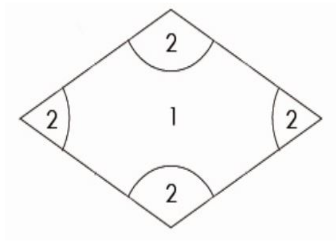
$$-\nabla_{\xi} \cdot [A'(\xi)\nabla_{\xi} v^{\epsilon}(\xi)] = f(J^{-1}\xi). \tag{6.12}$$

Follow the same deduction on homogenization theory, we deduce that the cell problem on the honeycomb is

$$-\nabla_{\xi} \cdot (A'(\xi)\nabla \chi_k(\xi)) = \nabla_{\xi} \cdot (A'(\xi)e_k) \qquad k = 1, 2 \tag{6.13}$$

Here we choose the coefficient the same with (6.3), and Ω is chosen to be the circle centered at the origion with radius $r0$. Suppose the length of the rhombus is 1, and $r \in (0, 0.5)$ to keep the circle inside the hexagon. It is displayed in the figure on the left hand side below. And inside the cell T_0, the coefficient is shown on the right hand side below.





Using the same method of Example 3 from the last subsection, we can calculate the corrector function of the new cell problem on rhombus T_0.

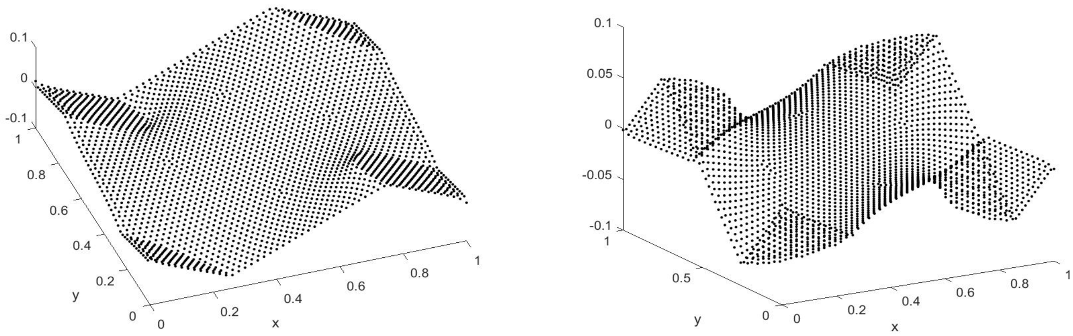


Figure 8: Left: χ^1; right: χ^2.

As the figure shows, the corrector function has four interior layer near the boundary of the four arcs. Thus, χ^i has contribution on the effective coefficients. And since the coefficient in rhombus is no longer 2×2 diagonal matrix. Here

$$A' = J^T AJ.$$

is a 2×2 full matrix. Thus the effective coefficient $\overline{A}'$ is also a 2×2 full matrix.

Since the length of the rhombus T_0 is chosen to be 1, the length of hexagon is $\frac{\sqrt{3}}{3}$. We choose $r0$ properly to let the area of Ω and area of T_0 has the same proportion. Here we choose

$$r0 = \frac{k}{10} * \sqrt{\frac{\sqrt{3}}{2}}, \qquad k = 1, 2, 3, 4.$$

And we use the standard finite element method to calculate the effective coefficient $\overline{A}'$. Then we divide it by the area of the hexagon T so that the coefficient has the same integral average on the diagonal. Suppose

$$\overline{A_0} = \frac{\overline{A}'}{\frac{\sqrt{3}}{2}}$$

. We compare it with the result in example 3. Here $\overline{A}$ is picked up from table 3. The numerical results are displayed below. Here we use the same elements with example 3.

r	$\overline{A}$		$\overline{A_0}$	
0.1	1.021164537	1.02363E-13	1.016289356	0.504554253
	1.02364E-13	1.0213861	0.504554253	1.016414197
0.2	1.087469302	8.97449E-13	1.068265257	0.518766723
	8.97441E-13	1.087772944	0.518766723	1.068351504
0.3	1.208486779	3.12725E-12	1.168149807	0.548958431
	3.12721E-12	1.2087199	0.548958431	1.168227178
0.4	1.404399579	4.13826E-12	1.346926871	0.623956074
	4.13834E-12	1.404481935	0.623956074	1.346837406

Table 4: Comparison between square and hexagon

From the table above we can clearly see that under the same proportion, the effective coefficient on honeycomb is smaller than it on unit square. And since the coefficient A' is no longer diagonal, the effective coefficient $\overline{A}'$ is not diagonal either. But it is still a symmetric positive definite matrix.

In a word, in the honeycomb problem, we do the transformation of coordinates and use the theory of homogenization. Then we deduce the cell problem in a rhombus with periodic boundary condition. Then we can solve the problem in the new region. Using the same element in finite element method, we also deduce that the effective coefficient in honeycomb problem is smaller than the square region.

References

[1] Thomas Y. Hou, Xiao-Hui Wu and Zhiqiang Cai, *Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients*, Math. Comp. 68 (1999), 913-943.

[2] Lawrence C. Evans, *Partial Differential Equations*, Graduate Studies in Mathematics, Volume 19, American Mathematical Society.

[3] Christophe Prange, *The Oscillating Test Function Method*, February 15, 2016.

[4] Lu Jin-Fu, Guan Zhi, *Numerical resolution of partial differential equations*, the third edition, Tsinghua publisher, November, 2016.