The Multiscale Finite Element Method (MsFEM) and Its Applications

A Thesis by

Yalchin R. Efendiev

In Partial Fulfillment of the Requirements

for the Degree of

Doctor of Philosophy

California Institute of Technology

Pasadena, California

1999

(Submitted April 26, 1999)

Acknowledgements

I would like to thank my advisor, Prof. Thomas Hou, who suggested this work and guided me in this thesis to its completion. I am grateful to the other members of the thesis defense committee, Prof. O. Bruno, Prof. D. Meiron and Prof. M. Ortiz. I am also very grateful to Dr. X. H. Wu and Dr. Y. Zhang for their collaborations. I would like to thank Prof. Lou Durlofsky (Stanford) and Dr. Seong Lee (Chevron Petroleum Technology Company) for their suggestions and contributions to the last part of my thesis.

Many thanks are due to Peter Park for reading this thesis. I would also like to thank all the graduate students of the Applied Mathematics Department at Caltech, in particular Mayya Tokman, Gang Hu, Peter Park and Dr. Anatoly Baumstein. Thanks are also due to Sheila Shull and Georgia Frueh for making my stay at Caltech so pleasant.

Abstract

Multiscale problems occur in many scientific and engineering disciplines, in petroleum engineering, material science, etc. These problems are characterized by the great deal of spatial and time scales which make it difficult to analyze theoretically or solve numerically. On the other hand, the large scale features of the solutions are often of main interest. Thus, it is desirable to have a numerical method that can capture the effect of small scales on large scales without resolving the small scale details.

In the first part of this work we analyze the multiscale finite element method (MsFEM) introduced in [28] for elliptic problems with oscillatory coefficients. The idea behind MsFEM is to capture the small scale information through the base functions constructed in elements that are larger than the small scale of the problem. This is achieved by solving for the finite element base functions from the leading order of homogeneous elliptic equation. We analyze MsFEM for different situations both analytically and numerically. We also investigate the origin of the resonance errors associated with the method and discuss the ways to improve them.

In the second part we discuss flow based upscaling of absolute permeability which is an important step in the practical simulations of flow through heterogeneous formations. The central idea is to compute the upscaled, grid-block permeability from fine scale solutions of the flow equation. It is well known that the grid block permeability may be strongly influenced by the boundary conditions imposed on the flow equations and the size of grid blocks. We analyze the effects of the boundary conditions and grid block sizes on the computed grid block absolute permeabilities. Moreover, we employ the ideas developed in the analysis of MsFEM to improve the computed values of absolute permeability.

The last part of the work is the application of MsFEM as well as upscaling of absolute permeability on upscaling of two-phase flow. In this part we consider coarse models using MsFEM. We demonstrate the efficiency of these models for practical problems. Moreover, we show that these models improve the existing approaches.

Contents

List of Figures

List of Tables

Chapter 1 Overview

Many problems of fundamental and practical importance have multiple scale solutions. A complete analysis of these problems is difficult. For example the difficulty in analyzing groundwater transport is mainly caused by the heterogeneity of subsurface formations spanning over many scales. The heterogeneity is often represented by the multiscale fluctuations in the permeability of the media. A direct numerical solution of the multiple scale problems is difficult even with modern supercomputers. The major difficulty of direct solutions is the scale of computation. For groundwater simulations, it is common to have millions of grid blocks involved, with each block having a dimension of tens of meters, whereas the permeability measured from cores is in order of several centimeters. This gives more than 10^5 degrees of freedom per spatial dimension in computation. Therefore, a tremendous amount of computer memory and CPU time are required, and they can easily exceed the limits of today's computing resources. The situation can be relieved to some degree by parallel computing; however, the size of a discrete problem is not reduced. The load is shared by more processors with some memory. Whenever one can afford to resolve all the small scale features of a physical problem, direct solutions provide quantitative information of the physical processes at all scales. On the other hand, it is often sufficient to predict the macroscopic properties of the multiple-scale systems, such as effective permeability. Therefore, it is desirable to develop methods which can capture the effect of small scales on the large scales without resolving all the small scale features.

In this work we discuss the methods which can capture the small scale information on the large scales without resolving all details. The thesis consists of three parts: 1) The multiscale finite element method (MsFEM) and its analysis, 2) Upscaling of absolute permeability and 3) The applications of MsFEM. Each part is provided with introduction where we discuss the previous works and describe the results of the part. Here we briefly describe the problems considered in the work.

In the first part of this work we analyze the multiscale finite element method (MsFEM) introduced in [28] for elliptic problems with oscillatory coefficients. The idea of MsFEM is to capture the small scale information through the base functions constructed in the

elements whose sizes are larger than the sizes of the small scales of the problem. This is achieved by solving the finite element base functions from the leading order of homogeneous elliptic equation. The construction of the base functions is fully decoupled from element to element; thus, the method is perfectly parallel and is naturally adapted to massively parallel computers. For the same reason, the method has the ability to handle extremely large degrees of freedom due to highly heterogeneous media, which are intractable by the conventional finite element (difference) methods. In this part of the work we analyze MsFEM for different possible situations both analytically and numerically. MsFEM have been analyzed with emphasis on the resonance sampling effect. We investigate the origin of the resonance errors that occurs in MsFEM and the ways to improve them.

In the second part of the work, we discuss flow based upscaling of absolute permeability which becomes an important step in practical simulations of flow through heterogeneous formations. The central idea is to compute upscaled, grid-block permeability from fine scale solutions of the flow equation. It is well known that the grid block permeability may be strongly influenced by the boundary conditions imposed on the flow equations and the size of grid blocks. In this part of the work we employ the techniques developed in the analysis of MsFEM to improve the computed values of absolute permeability. We estimate the difference between the computed value and exact value of absolute permeability for periodic structures. The numerical examples for random permeability fields demonstrate that the new upscaling methods improve the accuracy of the existing upscaling methods in general. In this part of the work we also consider the effects of the boundary conditions and grid block sizes on the computed grid block absolute permeabilities.

The last part of the work is the application of MsFEM on upscaling of two-phase flow. In this part we consider coarse models using MsFEM. The main idea of these methods is to incorporate the higher order statistics of the problem into the coarse model. The calculations of higher moments require some approximation of the detailed behavior of the problem. This approximation can be effectively constructed using the multiscale base functions of MsFEM. We demonstrate the efficiency of these models for practical problems. Moreover, we show that these models improve the predictions of the existing approaches.

Part I

Multiscale Finite Element Methods and their analysis

Chapter 2 Introduction

In this part of the work we analyze MsFEM for different cases. Recently, a multiscale finite element method (MsFEM) has been developed [28, 26] for capturing the large scale solutions of multiscale problems on a coarse mesh (with mesh size larger than a certain cut-off scale of the problem). The main idea of the method is to build the local small scale information of the leading order differential operator into the finite element base functions. It is through these multiscale bases and the finite element formulation that the effect of small scales on the large scales are correctly captured. A key feature of MsFEM is that the construction of the base functions is a *local* operation within the elements. Thus, the construction in one element is decoupled from that in another element. In other words, a large scale computation is broken into many small and independent pieces. This results in many computational advantages [26], such as a saving in computer memory and good parallel efficiency. We remark that special base functions in finite element methods have been used by several authors in capturing multiscale solutions of PDE's. In particular, the works presented in [42, 5, 11, 29] are most relevant to our previous conforming multiscale finite element method [28].

The basic convergence property of the method has been established for a two-scale elliptic problem with periodic coefficients [28]. It is shown that the numerical solution converges to the homogenized solution in the limit of $\epsilon \to 0$ (ϵ is the small scale in the solution). The homogenization theory is used in the proof; however, it is neither required by the MsFEM formulation nor used in the computations. Our numerical experiments demonstrate that MsFEM, together with an over-sampling method, is well applicable to general elliptic problems with many or continuous scales [27]. It is shown that numerical solutions computed using MsFEM on a coarse grid give accuracy comparable to that of well-resolved solutions computed using conventional methods on a fine grid. The application of MsFEM to practical problems such as two-phase flows in porous media and other types of equations is currently under study. It is worth mentioning that MsFEM also gives convergent solutions when $h \ll \epsilon$, just like the conventional methods.

Previous analysis also reveals the resonance error between the grid scale and the scales

of the continuous problem [28]. This is a common difficulty in numerical upscaling methods. For the two-scale problem, the error due to the resonance manifests as a ratio between the wavelength of the small scale oscillation and the grid size; the error becomes large when the two scales are close. The resonance represents a *fundamental* difficulty due to the mismatch between the local construction of the multiscale base functions and the global nature of the elliptic problems. This mismatch between the local solution and the global solution produces a boundary layer in the first order corrector of the local solution which causes the resonance.

In this part of the work, we present the analytical and numerical analysis of the resonance error and the possible improvements. Following the formulations and overview of MsFEM in chapter 4, we analyze MsFEM for multiple scale problems. The convergence rate of the method for different choices of mesh sizes has been derived. Particularly, assuming ϵ_k be a sequence of decreasing separated scales we find that if the mesh h is between the scales ϵ_k and ϵ_{k+1}, i.e., $\epsilon_k \gg h \gg \epsilon_{k+1}$, then the convergence rate of MsFEM in H^1 norm is $C((h/\epsilon_k) + \sqrt{\epsilon_{k+1}/h})$. We see that the method has resonance effect towards both neighboring scales. The error towards a larger scale (larger than h) (h/ϵ_k) is the error of resolving larger scales and it is consistent with traditional FEM. The error towards a smaller scale (smaller than h) $\sqrt{\epsilon_{k+1}/h}$ is the error of capturing the smaller scales. The capturing error is specific for MsFEM and depends on the construction of base functions. Furthermore, we show how the error of resolving larger scales can be improved by introducing higher elements. We conclude the chapter with numerical justifications of our analytical results and some comments.

In chapter 5 we analyze MsFEM for problems with discontinuous coefficients. It is known that in general the convergence of finite element and finite difference methods deteriorate in the case of discontinuous parameters. The previous results (and later) on the analysis of MsFEM are derived under the assumption of smooth parameters. In chapter 5 we derive the convergence rate for MsFEM in the case of discontinuous parameters. The analysis requires a special treatment for discontinuous terms which is done in chapter 5. We show that MsFEM retains its convergence rate in the case of a two scale problem. We justify our analytical results with numerical experiments on checker-board structures.

The previous analysis of MsFEM has been performed for periodic structures. In these analyses we use homogenization theory which is developed and well studied for periodic

setting. The multiple scale expansion for general random problems has been worked out [32, 43]. It has been shown that the multiple scale expansion for general random cases and periodic cases has some similarities. But in general random cases, certain important functions of multiple scale expansion lose their periodicity and boundedness and gain some growth rate as $x \to \infty$. This growth rate influences the resonance error of MsFEM. In chapter 6, we derive the resonance error for MsFEM when the parameters of the problem have weak dependence.

As above analyses of MsFEM show that the resonance is due to the mismatch of the artificial boundary conditions of the base function and the oscillatory nature of the solution. Motivated by the analysis, the authors propose an *over-sampling* technique to overcome the difficulty due to the scale resonance in [27]. The idea is quite simple and easy to implement. Since the boundary layer in the first order corrector is thin, $O(\epsilon)$, we can sample in a domain with size larger than $h + \epsilon$ and use only the interior information to construct the bases (h is the mesh size). By doing this, the boundary layer in the larger domain has no influence on the base functions. Now the corresponding first order correctors are free of boundary layers. As a result, we obtain an improved rate of convergence.

In chapter 7, we analyze analytically and numerically the over-sampling version of MsFEM for the case of one-scale problem. Unfortunately, the over-sampling technique results in a *nonconforming* MsFEM method. The previous analyses need to be modified to take into account the nonconforming error. In chapter 7, we perform a careful estimate of the nonconforming errors in both H^1 norm and L^2 norm. The analysis shows that the nonconforming error is indeed small, consistent with the numerical results [26, 27]. Our analysis also reveals another type of resonance, which is the mismatch between the mesh size and the "perfect" sample size. In the case of a periodic structure, the "perfect" sample size is the length of an integer multiple of the period. We call the new resonance the "cell resonance." In the error expansion, this resonance effect appears as a *higher* order correction. Although the over-sampling helps eliminate the leading order resonance error, we find that over-sampling alone does not remove the cell resonance error, which dominates the nonconforming error. However, from our computational experience, the cell resonance errors seem to be generically small and are rarely observed in computations. This may be due to some subtle error cancellation in the convolution with the discrete Green's function (i.e., the inverse of the stiffness matrix). The discrete Green's function can be highly oscillatory

depending on the ratio h/ϵ, where h is the mesh size and ϵ is the small scale of the problem. In fact, in our tests with the worst resonance case, $h/\epsilon = 1.5$, the solution appears to converge. Only when $h = 1/1024$ does the effect of cell resonance become strong enough to stop the convergence. For more general problems, such as problems with random coefficients, we did not find resonance errors through numerical tests (see [27]). These results strongly indicate that the chance of having significant cell resonance in practical computations is small.

Finally we would like to note that for the practical problems MsFEM is a robust method. In practice the small scales are fixed and the constants in front of the resonance errors play an important role. As we observed numerically these constants are usually small and an overall error of the method is negligible for practical purposes.

Chapter 3 Formulation and overview

In this section we introduce the model problem and the multiscale finite element method. First we state some notations and conventions. In the following, the Einstein summation convention is used: summation is taken over repeated indices. Throughout the paper, we use the $L_2(\Omega)$ based Sobolev spaces $H^k(\Omega)$ equipped with norms and seminorms:

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

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

$H_0^1(\Omega)$ consists of those functions in $H^1(\Omega)$ that vanish on $\partial\Omega$. $H^{-1}(\Omega)$ is dual space of $H_0^1(\Omega)$, i.e., the set of all continuous linear functionals on $H_0^1(\Omega)$. We define $H^{1/2}(\partial\Omega)$ as the trace on $\partial\Omega$ of all functions in $H^1(\Omega)$ with the norm $\|v\|_{1/2,\partial\Omega} = \inf \|u\|_{1,\Omega}$ where the infimum is taken over all $u \in H^1(\Omega)$ with the trace v. In the paper the space $C^k(\Omega)$, continuous functions along with their k^{th} derivatives is equipped with the norm

$$\|u\|_{C^k(\Omega)} = \sum_{\alpha=0}^{k} \max_\Omega |D^\alpha u|.$$

Throughout, C (with or without subscripts) denotes a generic constant, which is independent of ϵ and h (mesh size), unless otherwise stated and $C + C = C$, $C \cdot C = C$.

3.1 Model problem and the multiscale finite element method

In this section we are going to investigate MsFEM on the following model problem

$$L_\epsilon u = f \quad \text{in } \Omega, \quad u = 0 \quad \text{on } \partial\Omega, \tag{3.1}$$

where

$$L_\epsilon = \frac{\partial}{\partial x_i} a_\epsilon^{ij} \frac{\partial}{\partial x_j}$$

is the linear elliptic operator, ϵ *is a small parameter, and* a_ϵ^{ij} *is symmetric and satisfies*

$$\alpha|\xi|^2 \le \xi_i a_\epsilon^{ij} \xi_j \le \beta|\xi|^2, \tag{3.2}$$

for all $\xi \in R^2$ and with $0 < \alpha < \beta$ where α and β are independent of ϵ. We assume that $f \in L_2(\Omega)$ and coefficients are smooth functions unless otherwise stated. As for the boundary $\partial\Omega$ we assume it is piecewise smooth. Under these conditions (3.1) is a well posed problem.

Variational problem of (3.1) is to seek $u \in H_0^1(\Omega)$ such that

$$a(u,v) = f(v), \qquad \forall v \in H_0^1(\Omega), \tag{3.3}$$

where

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

It is easy to see that the linear form $a(\cdot,\cdot)$ is elliptic and continuous, i.e.,

$$\alpha|v|_{1,\Omega}^2 \le a(v,v), \quad \forall v \in H_0^1 \tag{3.4}$$

and

$$|a(u,v)| \le \beta|u|_{1,\Omega}|v|_{1,\Omega}, \qquad \forall u,v \in H_0^1. \tag{3.5}$$

A finite element method is obtained by restricting the weak formulation (3.3) to a finite dimensional subspace of $H_0^1(\Omega)$. For $0 < h \le 1$, let $\mathbf{K}^h$ be a partition of Ω of elements K *(rectangles, triangles, etc)* with diameter $\le h$. For simplicity, we can assume that the partition consists of rectangular elements which are defined by an axi-parallel rectangular mesh. In each element $K \in \mathbf{K}^h$, we define a set of nodal basis $\{\phi_K^i, i = 1, \ldots, d\}$ with $d(= 4)$ being the number of nodes in the element. We will neglect the subscript K when

working in one element. In our multiscale method, ϕ^i satisfies

$$L_\epsilon \phi^i = 0 \qquad \text{in } K \in \mathbf{K}^h. \tag{3.6}$$

Let $x_j \in \overline{K}$ $(j = 1, \ldots, d)$ be the nodal points of K. As usual, we require $\phi^i(x_j) = \delta_{ij}$. One needs to specify the boundary condition of ϕ^i for well-poseness of (3.6). The boundary conditions play an important role in the convergence of MsFEM. In later sections we are going to specify and deal with different boundary conditions. For now, we assume that the base functions are linear along the boundaries of the elements, i.e. along the boundaries MsFEM base functions and traditional finite element base functions coincide. MsFEM with these base functions is conforming, i.e.,

$$V^h = \text{span}\{\phi_K^i : i = 1, \ldots, d; K \in \mathbf{K}^h\} \subset H_0^1(\Omega),$$

and the approximate solution of (3.3) in V^h, i.e., $u^h \in V^h$ is

$$a(u^h, v) = f(v), \qquad \forall v \in V^h. \tag{3.7}$$

As we see the only difference of MsFEM from traditional finite element method is the construction of the base functions.

Remark 3.1.1. The above formulation of the multiscale method is not restricted to the rectangular elements. It can be applied to triangular elements, which are more flexible in modeling more complicated geometries. In fact, in most of the analysis below, the shape of element is irrelevant. In some cases we find that the triangular elements have some advantages in discrete error cancellations.

Remark 3.1.2. As mentioned in the introduction, the purpose of the multiscale method is to capture the large scale solution. This general idea can be made more precise in the context of the above model problem when $a^{ij}(x/\epsilon)$ is a periodic function. In this case, there are two distinct scales in the solution, characterized by 1 and $1/\epsilon$. The large scale solution is the homogenized solution u_0, which is the limit of u as $\epsilon \to 0$. In fact, u equals u_0 up to $O(\epsilon)$ perturbations. It can be shown that u_0 is the solution of a homogenized, elliptic problem with constant coefficient. Thus u_0 is smooth. The oscillations at the ϵ-scale are contained in the perturbations. Because the base functions ϕ^i are defined by the same

operator L_ϵ, they are expected to have local structure similar to that of u.

The multiscale base functions are smooth if $h \ll \epsilon$ and can be well approximated by the standard continuous linear (bilinear) base functions. Thus, we expect the multiscale method to behave similarly as the linear finite element methods.

On the other hand, when $h \gg \epsilon$, ϕ^i contains a smooth part and an oscillatory part, which *cannot* be approximated by the linear (bilinear) functions. In this case, the multiscale method is very different from conventional finite element methods. In fact, we will show that the multiscale method gives solutions converging to u_0 in the limit of $\epsilon \to 0$ while the standard finite element method with piecewise polynomial base functions does not. An intuitive explanation is as follows. The effective coefficient a^*, which determines the homogenized operator, includes both the average of a and the averaged result of the interaction of small scale oscillations (see (3.10)). The polynomial base functions can only capture the first part because they do not characterize any oscillations. In contrast, the multiscale base functions contain the small scale information *in the same fashion* as u does. Therefore, they are able to accurately capture a^* through the variational formulation.

3.2 Homogenization for one periodic case and related estimates

In this section, we review the homogenization theory of (3.1) in the case when $a^{ij}(x/\epsilon)$ is a one periodic function. We provide some estimates which are important for understanding of MsFEM. These results reveal the structure of the solution and the multiscale functions.

It is known [8, 32] that the solution of (4.1) can be expanded as

$$u_\epsilon = u_0 + \epsilon\chi^i(\frac{x}{\epsilon})\nabla_i u_0 + \epsilon\theta^u. \tag{3.8}$$

In the expansion, u_0 is the solution of the homogenized equation

$$a^{ij}_* \nabla_i \nabla_j u_0 = f \qquad \text{in } \Omega, \tag{3.9}$$

satisfying $u_0 = 0$ on $\partial\Omega$. The constant homogenized coefficients a^{ij}_* are given by

$$a^{ik}_* = \frac{1}{|Y|}\int_Y a^{ij}(\delta_{jk} + \nabla^y_j \chi^k)dy, \tag{3.10}$$

where ∇^y is the gradient with respect to the "fast" variable $y = x/\epsilon$, and χ^k is the periodic

solution in the unit cell Y of

$$\nabla_i^y a^{ij} \nabla_j^y \chi^k = -\nabla_i^y a^{ik}. \tag{3.11}$$

It can be shown [8, 32] that (3.11) has a unique solution in H^1_{per} which is defined up to a constant. For θ^u we have

$$\begin{aligned} L_\epsilon \theta^u &= \frac{1}{\epsilon} L_\epsilon (u_\epsilon - u_0 - \epsilon \chi^i \nabla_i u_0) \qquad \text{in } \Omega, \\ \theta^u &= -\chi^i \nabla_i u_0 \qquad \text{on } \partial\Omega. \end{aligned}$$

The heuristic derivation of the average equation can be obtained by expanding u_ϵ

$$u_\epsilon(x,y) = u_0(x,y) + \epsilon u_1(x,y) + \dots$$

Introducing $\nabla = \nabla_x + 1/\epsilon \nabla_y$, and substituting the expansion into the above system of equations, and collecting terms with the same power of ϵ, we get

$$O(\epsilon^{-2}): \quad \nabla_{y_i} a^{ij}(y) \nabla_{y_j} u_0 = 0 \tag{3.12}$$

$$O(\epsilon^{-1}): \quad \nabla_{x_i} a^{ij}(y) \nabla_{y_j} u_0 + \nabla_{y_i} a^{ij}(y) \nabla_{x_j} u_0 + \nabla_{y_i} a^{ij}(y) \nabla_{y_j} u_1 = 0 \tag{3.13}$$

$$O(\epsilon^0): \quad \nabla_{x_i} a^{ij}(y) \nabla_{x_j} u_0 + \nabla_{y_i} a^{ij}(y) \nabla_{x_j} u_1 + \nabla_{x_i} a^{ij}(y) \nabla_{y_j} u_1 + \nabla_{y_i} a^{ij}(y) \nabla_{y_j} u_2 = 0. \tag{3.14}$$

From (3.12) we obtain that $u_0(x,y) = u_0(x)$. Assuming $u_1(x,y) = \chi^p(y) \nabla_p u_0(x)$ in the second equation (3.13), we get (3.11) for $\chi^p(y)$. Eventually, taking an average over the period Y in (3.14) we obtain (3.9).

The convergence of u_0 or $u_0 + \epsilon \chi^p \nabla_p u_0$ to the exact solution u_ϵ under different regularity conditions and in different norms have been thoroughly investigated.

3.3 Expansion of base functions

We have the similar expansion for the base functions when the mesh size is much bigger than the period ϵ:

$$\phi_\epsilon = \phi_0 + \epsilon\chi^i(\frac{x}{\epsilon})\nabla_i\phi_0 + \epsilon\theta^\epsilon, \tag{3.15}$$

where ϕ_0, χ^i, and θ^ϵ are defined in the same way as above. The equation for θ^ϵ can be further simplified by taking into account the fact that ϕ_0 is linear along the boundaries of the triangular elements K. It follows that ϕ_0 is linear in K. Thus by (3.6) and the equations for θ^ϵ and χ^k, we have

$$\begin{aligned} L_\epsilon\theta^\epsilon &= \frac{1}{\epsilon}L_\epsilon(\phi_\epsilon - \phi_0 - \epsilon\chi^k\nabla_k\phi_0) \\ &= -\frac{1}{\epsilon}\nabla_i a_\epsilon^{ij}\nabla_j(\phi_0 + \epsilon\chi^k\nabla_k\phi_0) \\ &= -(\nabla_i a_\epsilon^{ij}\nabla_j\chi^k + \frac{1}{\epsilon}\nabla_i a_\epsilon^{ik})\nabla_k\phi_0 - \nabla_i\chi^k a_\epsilon^{ij}\nabla_j\nabla_k\phi_0 \\ &= 0. \end{aligned}$$

and $\theta^\epsilon = -\chi^p\nabla_p\phi_0$ on the boundary of K. θ^ϵ can be written as $\theta^\epsilon = \eta^p\nabla_p\phi_0$, where η^p is defined as the solution of

$$\begin{aligned} L_\epsilon\eta^p &= 0 \quad \text{in } K \\ \eta^p &= -\chi^p \quad \text{on } \partial K. \end{aligned} \tag{3.16}$$

3.4 Some estimates in finite element methods

Cea's lemma

Assume that we are approximating the solution u of the variational equations

$$a(u,v) = f(v), \quad \forall v \in V \tag{3.17}$$

where the space V, the bilinear form $a(u,v)$ and the linear form f satisfy the assumptions of the Lax-Milgram lemma. Consider a family V_h of subspaces of space V. With each finite

element space V_h is associated the discrete solution u_h which satisfies

$$a(u_h, v_h) = f(v_h), \quad \forall v_h \in V_h.$$

Then Cea's lemma states:

Lemma 3.4.1 *There exists a constant C independent of the subspace V_h such that*

$$\|u - u_u\| \leq C \inf_{v_h \in V_h} \|u - v_h\|.$$

Consequently, a sufficient condition for convergence is that there exists a family V_h of subspaces of space V such that, for each $u \in V$,

$$\lim_{h \to 0} \inf_{v_h \in V_h} \|u - v_h\| = 0.$$

We note that the norm in the formulation of Cea's lemma is the norm associated with bilinear form $a(u, v)$ (Lax-Milgram lemma).

Strang's lemma

The continuity of the finite element approximation of the solution of a boundary value problem and possibly of its derivatives have been used in conforming finite element bases. However, conforming elements, particularly for higher order problems, are usually more complex [12]. Hence, it is natural to enquire if one can relax these interelement continuity requirements and still obtain a convergent finite element solution. These kinds of finite element methods are called nonconforming finite element methods. To show Strang's lemma we introduce a nonconforming discrete space $\hat{H}^h$ with the norm $\|\cdot\|_h$

$$\|w_\epsilon^h\|_{h,\Omega} = \Big(\sum_{K \in \mathbf{K}^h} \int_K |\nabla w_\epsilon^h|^2 dx \Big)^{1/2}. \tag{3.18}$$

The discrete solution $u^h \in \hat{H}^h$ is defined by solving

$$a(u^h, v^h) = f(v^h), \quad \forall v^h \in \hat{H}^h$$

where

$$a(u^h, v^h)| = |\sum_{K\in\mathbf{K}^h}\int_K a_\epsilon^{ij}\nabla_i u^h \nabla_j v^h dx|.$$

We would like to note that it is required $\|\cdot\|_h$ be a norm in $\hat{H}^h$. Then Strang's lemma asserts [12] that

$$\|u_\epsilon - u_\epsilon^h\|_{h,\Omega} \leq C(\inf_{v_\epsilon^h\in\hat{H}^h}\|u_\epsilon - v_\epsilon^h\|_{h,\Omega} + \sup_{w_\epsilon^h\in\hat{H}^h}\frac{|f(w_\epsilon^h) - a(u_\epsilon, w_\epsilon^h)|}{\|w_\epsilon^h\|_h}), \tag{3.19}$$

where $\hat{H}^h$ is the finite dimensional space generated by the nonconforming basis functions (in general $\hat{H}^h \not\subset H^1$).

3.5 Overview of MsFEM convergence

MsFEM has been introduced and analyzed in the case of one periodic smooth coefficients a_ϵ^{ij} in [28]. The analysis has been carried out for the different choices of the mesh size h: the case when the mesh size h is larger than the characteristic length scale ϵ, $h >> \epsilon$ and the case $h << \epsilon$. In the case $h \ll \epsilon$, MsFEM approximates the solution as traditional finite element methods with the second order accuracy $O((h/\epsilon)^2)$. The case $h \gg \epsilon$ is of the main interest since this case shows how effectively MsFEM is capturing the small scale information of the global problem. The L_2 convergence rate obtained in [28] is

$$C_1 h + C_2\frac{\epsilon}{h}. \tag{3.20}$$

This error estimate shows that small scale information in the base functions leads to the correct large scale information and numerical solution approaches exact solution as $\epsilon \to 0$. This convergence rate has been confirmed with numerical results. The term ϵ/h in the convergence rate indicates the resonance between the physical scale and the mesh size. We would like to note that the resonance phenomenon is common in upscaling methods. The analysis of MsFEM presented in [28] is omitted since more general cases will be investigated in this work. We would like to point out that the cause of the resonance is in (3.20). It was found [28] that θ^ϵ in the expansion of the base functions (see (3.15)) causes the resonance. We note that this term corrects the oscillations on the boundary given with

$\phi_0 + \epsilon\chi^i(x/\epsilon)\nabla_i\phi_0$ to the linear boundary conditions. Thus, θ^ϵ has large gradients in the neighborhood of the boundary. Numerically it was found that these large gradients are in ϵ neighborhood of the boundary. As the values of the mesh size and the small scale of the problem get closer the boundary layers of θ^ϵ occupy the larger portions of the elements causing the deterioration of the convergence rate of MsFEM.

Based on the analysis of this method the authors also proposed [53] an improved version of MsFEM, MsFEM-os (MsFEM with over-sampling) which practically removes the resonance. The main idea of this method is to use interior information of the base functions sampled in larger domains (larger than our elements). The authors have tested this method on different examples and demonstrated its numerical robustness. In a later section we are going to analyze this method.

For the practical purposes, the constants C_1 and C_2 in (3.20) play an important role. Numerous numerical examples for random cases and the cases with continuous spectrum of scales demonstrate that MsFEM is a robust method and the error of MsFEM does not exceed 2 percent in these problems. In Fig. 3.1 and Fig. 3.2 the discrete L_2 error estimates are plotted for random coefficients. In particular, these coefficients have the continuous spectrum of scales between 1/32 and 1/256. The domain is $[0,1]^2$, the ratio of maximum and minimum of the coefficients is 400 and $f = -1$. In these figures the horizontal lines represent the resolved solution. In Fig. 3.1 and Fig. 3.2 LFEM stands for FEM where the mesh size changes from 1/32 to 1/512, MFEM-L denotes MsFEM with linear boundary conditions on the mesh between 1/32 and 1/512, and the lines with "x" correspond to MsFEM with over-sampling. These examples demonstrate that MsFEM is a robust method.

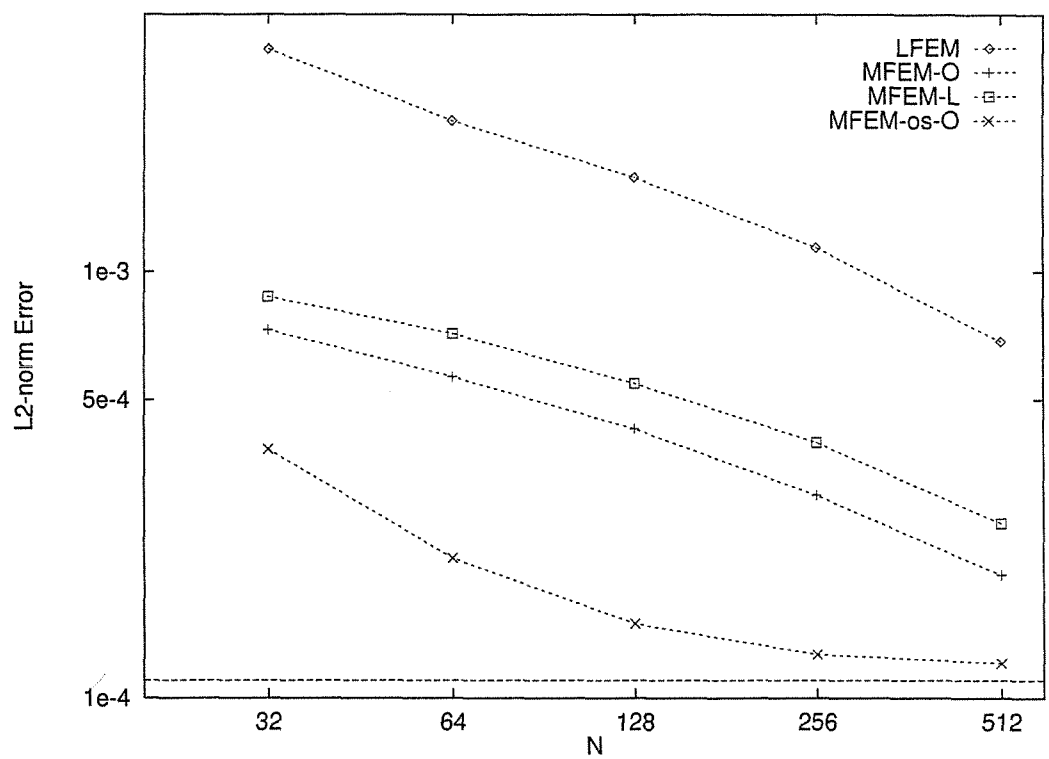

Figure 3.1: The discrete L_2 norm error of the solutions using various schemes for a fractally distributed permeability field. The horizontal dash line indicates the error of traditional FEM solution with $N = 2048$. Along the N axis we plot the resolution (number of elements in each direction)

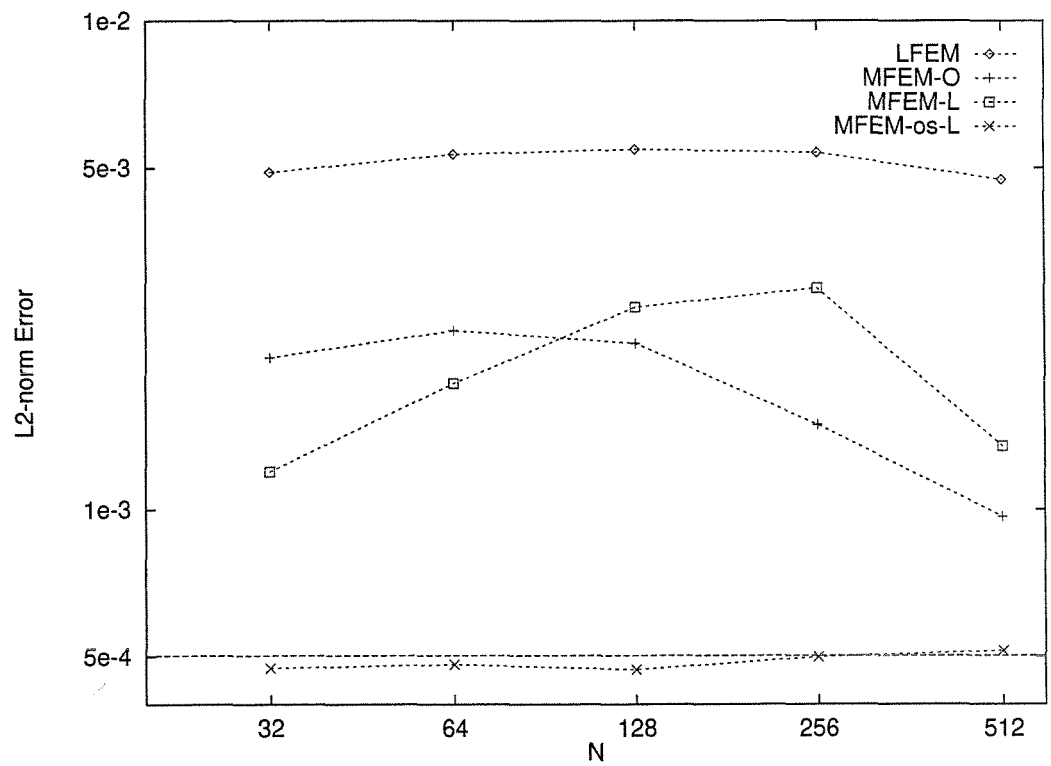

Figure 3.2: The discrete L_2 norm error of the solutions using various schemes for a log-normally distributed permeability field. The horizontal dash line indicates the error of traditional FEM solution with $N = 2048$. Along the N axis we plot the resolution (number of elements in each direction).

Chapter 4 MsFEM for problems with many separated scales

4.1 Summary

In this section we analyze the convergence of MsFEM for problems with many separated scales. In particular, we are interested in the convergence rate of the method for different mesh sizes. We establish that if the mesh size is between the scales ϵ_k and ϵ_{k+1}, i.e., $\epsilon_k \gg h \gg \epsilon_{k+1}$, then the H^1 norm of the error is

$$C(\sqrt{\frac{\epsilon_{k+1}}{h}} + \frac{h}{\epsilon_k})$$

The first term in the estimate is due to the way of capturing of the scales smaller than ϵ_{k+1} through the multiscale base functions. The second term is the error of resolving scales larger than ϵ_k with the homogenized part of the multiscale base functions.

We introduce an improved version of MsFEM. This method improves the resonance error corresponding to h/ϵ_k, i.e., the error of resolving the larger scales (larger than the mesh size). This is done by changing the boundary conditions of MsFEM to higher order polynomials. The corresponding error is $h + (h/\epsilon_k)^m + \sqrt{\epsilon_{k+1}/h}$ with integer $m > 1$. We note that this version of MsFEM is still conforming. We discuss in later sections how to improve the resonance error $\sqrt{\epsilon_{k+1}/h}$, the error corresponding to capturing the small scale information.

We also present the L_2 estimates for MsFEM. The structure of the estimates are similar to that of the H^1 norm estimates. Because of some small terms in the H^1 estimate, which cannot be expressed through the L_2 norm of the source term of the problem, the L_2 norm estimate we derived in the Aubin-Nitsche fashion contains some overestimated terms. These terms, however, contain no resonance effect. Therefore, our L_2 estimate of the resonance error is tight. Moreover, the overestimated terms can indeed be eliminated by using the method introduced in [28], which compares the solution and its numerical counterpart at discrete nodal points.

Our error estimates are confirmed by the numerical experiments. The computations are extremely large and are done on parallel computers, e.g., the Intel Paragon computer. Even so, we can only test two scale problems. We confirm the L_2 estimate when the mesh size is between the physical scales ϵ_1 and ϵ_2, i.e., $\epsilon_1 \gg h \gg \epsilon_2$. The computations encounter difficulties because memory limitation prevents us from choosing well separated small scales in the tests. Consequently, it is difficult to separately verify each resonance error in the L_2 estimates. But different numerical examples demonstrate that in some cases either one of them can be dominating, whereas in other cases both may be important. Results for $h \ll \epsilon_2 \ll \epsilon_1$ are also presented.

The rest of the chapter is organized as follows. The brief formulations of the 2-D problem introduced in the next section. In Section 3 we estimate the first order correctors for partially homogenized solutions. In sections 4 and 5 we derive H^1 and L_2 error estimates for MsFEM. The numerical results are presented in Section 6. The higher order MsFEM and other possible generalizations are discussed in the concluding remarks.

4.2 Formulation

In this section we investigate MsFEM for

$$\nabla_i a_\epsilon^{ij} \nabla_j u = f, \quad u = 0 \text{ on } \partial\Omega, \tag{4.1}$$

where

$$a_\epsilon^{ij} = a^{ij}(\frac{x}{\epsilon_1}, \frac{x}{\epsilon_2}, \ldots, \frac{x}{\epsilon_n}) \tag{4.2}$$

where $\epsilon_1 \gg \epsilon_2 \gg \cdots \gg \epsilon_n$ is a set of n ordered length scales, which all depend on a single parameter ϵ. For example, ϵ_i, $i = 1, \ldots, n$ are some powers of ϵ, ϵ^{p_i} with $p_1 < p_2 < \cdots < p_n$. Moreover, for simplicity we assume $a^{ij}(y_1, y_2, \ldots, y_n)$ to be sufficiently smooth periodic functions in y_i $(i = 1, \ldots, n)$ in a unit cube Y. Here, the smoothness assumption is convenient but not crucial for our analysis. It is sufficient to assume that $a^{ij} \in C^1$. The convergence of MsFEM for problems with discontinuous coefficients will be analyzed in following sections.

MsFEM is implemented as in the previous section. In each element $K \in \mathbf{K}^h$, we define

a set of nodal basis $\{\phi_K^i\}$, $i = 1, \ldots, d$, with $d(= 4)$ being the number of nodes of the element. We will neglect the subscript K when working in one element. The base functions ϕ^i satisfies

$$L_\epsilon \phi^i = 0 \quad \text{in} \quad K = \mathbf{K}^h. \tag{4.3}$$

We assume that the base functions are linear on each side of the boundary and $\phi^i(x_j) = \delta_{ij}$ where $x_j \in K$ $(j = 1, \ldots, d)$ are the nodal points of K . So we have:

$$V^h = \text{span}\{\phi_K^i; \quad i = 1, \ldots, d, \quad K \subset \mathbf{K}^h\} \subset H_0^1(\Omega).$$

In the following we study the approximate solution of (3.3) in V^h, i.e., $u^h \in V^h$ such that

$$a(u^h, v) = f(v), \quad \forall v \in V^h. \tag{4.4}$$

4.3 Estimates for first order correctors

In this section we review the homogenization theory of Eq. (4.1) [8, 2] and estimate the first order correctors, namely the difference between the solution and its H^1 approximation. We found the convergence results for the problems with many scales but we could not find the estimates for the convergence rate. We will derive these estimates in this section. The main difficulty in this estimate is to express the H^1 norm of the first order corrector through right-hand side of (4.1), i.e., $\|f\|_{0,\Omega}$. This is essential for the use of Aubin-Nitsche trick in the L_2 analysis of MsFEM. In Lemma 3.1 we prove such a result for smooth domains. However, for convex polygons we could not obtain the similar estimate. As shown later, this leads to a slightly overestimated L_2 norm error.

First, let us consider the case with two scales:

$$\nabla_i a^{ij}(\frac{x}{\epsilon_1}, \frac{x}{\epsilon_2}) \nabla_j u_\epsilon = f \quad \text{in } \Omega, \qquad u^\epsilon = 0 \quad \text{on } \partial\Omega \tag{4.5}$$

where $f \in L_2(\Omega)$ and $\epsilon_1 \gg \epsilon_2$. For convenience we take $y = \frac{x}{\epsilon_1}$ and $z = \frac{x}{\epsilon_2}$. Fixing $\frac{x}{\epsilon_1} = \lambda$ as a parameter, we consider $a^{ij}(\frac{x}{\epsilon_1}, \frac{x}{\epsilon_2})$ to be a family of functions $a^{ij}(\lambda, z)$ where λ is a parameter. By the assumption made in the previous section, $a(\lambda, z)$ is z periodic for any λ and $\alpha|\xi|^2 \leq \xi_i a^{ij}(\lambda, z)\xi_j \leq \beta|\xi|^2$ for all $\xi \in \mathbf{R}^2$ with $0 < \alpha < \beta < \infty$. Then the family of

operators,

$$A_\lambda^\epsilon = \frac{\partial}{\partial x_i} a^{ij}(\lambda, \frac{x}{\epsilon_2}) \frac{\partial}{\partial x_j}, \tag{4.6}$$

can be homogenized by the standard homogenization rule, λ being a parameter. Next we homogenize A_λ^ϵ with respect to ϵ_1. Thus

$$\frac{\partial}{\partial x_i} a^{ij}(\frac{x}{\epsilon_1}, \frac{x}{\epsilon_2}) \frac{\partial}{\partial x_j}, \tag{4.7}$$

is homogenized in two steps by the reiterated homogenization. More specifically, we define $\chi_\lambda^k(z)$ on $Z = (0,1) \times (0,1)$ as the periodic solution of

$$\frac{\partial}{\partial z_i} a^{ij}(\lambda, z) \frac{\partial(\chi_\lambda^k(z) + z_k)}{\partial z_j} = 0, \tag{4.8}$$

such that

$$\int_Z \chi_\lambda^k(z) = 0.$$

Then the homogenized operator for A_λ^ϵ is given by

$$A_\lambda = \frac{\partial}{\partial x_i} a_\lambda^{ij} \frac{\partial}{\partial x_j} \tag{4.9}$$

where

$$a_\lambda^{ij} = \frac{1}{|Z|} \int_Z a_{kl}(\lambda, z) \frac{\partial(\chi_\lambda^i + z_i)}{\partial z_k} \frac{\partial(\chi_\lambda^j + z_j)}{\partial z_l} dz. \tag{4.10}$$

Now taking into account that $\lambda = y = \frac{x}{\epsilon_1}$, A_λ can be homogenized as

$$A_0 = \frac{\partial}{\partial x_i} a_0^{ij} \frac{\partial}{\partial x_j}$$

where

$$a_0^{ij} = \frac{1}{|Y|} \int_Y a^{kl}(y) \frac{\partial(\chi^i + y_i)}{\partial y_k} \frac{\partial(\chi^j + y_j)}{\partial y_l} dy \tag{4.11}$$

and χ^j is the periodic solution of $A_\lambda(\chi^j - y_j) = 0$, such that $\int_Y \chi^j = 0$.

Remark 4.3.1. The reiterated homogenization procedure can be used for the n scale case. Denoting H as the partial homogenization operator we have

$$a_0 = H \circ H \circ \cdots \circ H \circ a(\frac{x}{\epsilon_1}, \frac{x}{\epsilon_2}, \ldots, \frac{x}{\epsilon_n}).$$

Following the homogenization steps represented above we can approximate the solution of (4.5) in H^1 norm as

$$u_\epsilon = u_0^\lambda + \epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda + \theta_\epsilon \tag{4.12}$$

where u_0^λ is the solution of

$$A_\lambda u_0^\lambda = f \quad \text{in } \Omega, \qquad u_0^\lambda = 0 \quad \text{on } \partial\Omega. \tag{4.13}$$

Lemma 4.3.1 *Let θ_ϵ be the solution of (4.12), then under the assumption that $\partial\Omega$ is sufficiently smooth we have*

$$\|\theta_\epsilon\|_{H^1(\Omega)} \leq (C_1 \epsilon_2^{1/(2+\eta)} + C_2 \frac{\epsilon_2}{\epsilon_1})\|f\|_{0,\Omega} \tag{4.14}$$

where $\eta > 0$ is an arbitrary positive number.

Proof. Denoting $u_\epsilon^1 = u_0^\lambda + \epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda$ we can write $a_\epsilon^{ij}\nabla_j u_\epsilon^1$ as:

$$\begin{aligned} a_\epsilon^{ij}\nabla_j u_\epsilon^1 &= (a_\epsilon^{ij} + a_\epsilon^{ik}\nabla_k^z \chi_\lambda^j)\nabla_j u_0^\lambda + \epsilon_2 a_\epsilon^{ij}\chi_\lambda^m \nabla_j \nabla_m u_0^\lambda \\ &= a_\lambda^{ij}\nabla_j u_0^\lambda + (a_\epsilon^{ij} + a_\epsilon^{ik}\nabla_k^z \chi_\lambda^j - a_\lambda^{ij})\nabla_j u_0^\lambda + \epsilon_2 a_\epsilon^{ij}\chi_\lambda^m \nabla_j \nabla_m u_0^\lambda \\ &= a_\lambda^{ij}\nabla_j u_0^\lambda + g_i^j \nabla_j u_0^\lambda + \epsilon_2 a_\epsilon^{ij}\chi_\lambda^m \nabla_j \nabla_m u_0^\lambda \end{aligned} \tag{4.15}$$

where $g_i^j = a_\epsilon^{ij} + a_\epsilon^{ik}\nabla_k^z \chi_\lambda^j - a_\lambda^{ij}$ with ∇_k^z denoting $\partial/\partial z_k$. Using (4.8) and the fact that $\nabla_i a_\lambda^{ij}\nabla_j$ is a partially homogenized operator for (4.7) we have

$$\int_Z g_i^j dz = 0 \quad \text{and} \quad \frac{\partial}{\partial z_i} g_i^j = 0. \tag{4.16}$$

Thus, g_i^k is a periodic solenoidal vector with average zero. It can be expressed as [32]

$$g_i^k = \frac{\partial}{\partial z_j}\alpha_{ij}^k(y,z)$$

where $\alpha_{ij}^k = -\alpha_{ji}^k$ and $\alpha_{ij}^k \in (L_2(Y), H^1(Z))$. Using this representation we can write (4.15) as

$$\begin{aligned} a_\epsilon^{ij}\nabla_j u_\epsilon^1 = a_\lambda^{ij}\nabla_j u_0^\lambda + \epsilon_2\frac{\partial}{\partial x_j}(\alpha_{ij}^k(y,z)\frac{\partial}{\partial x_k}u_0^\lambda) - \frac{\epsilon_2}{\epsilon_1}\frac{\partial}{\partial y_j}(\alpha_{ij}^k\frac{\partial}{\partial x_k}u_0^\lambda) \\ - \epsilon_2\alpha_{ij}^k(y,z)\frac{\partial^2}{\partial x_j\partial x_k}u_0^\lambda + \epsilon_2 a_\epsilon^{ij}\chi_\lambda^m\nabla_j\nabla_m u_0^\lambda. \end{aligned} \tag{4.17}$$

Denoting the last two terms on the r.h.s. as r_ϵ^i, by (4.5), (4.12), and (4.13) we have

$$\begin{aligned} &\nabla_i a_\epsilon^{ij}\nabla_j\theta_\epsilon = -div\ \mathbf{r}_\epsilon \quad in\ \Omega, \\ &\theta_\epsilon = \epsilon_2\chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda \quad on\ \ \partial\Omega. \end{aligned} \tag{4.18}$$

Equation (4.18) along with the regularity properties for the solution of the elliptic PDE gives

$$\|\theta_\epsilon\|_{H^1(\Omega)} \le C\|div\mathbf{r}_\epsilon\|_{-1,(\Omega)} + C\|\epsilon_2\chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{H^{1/2}(\partial\Omega)}. \tag{4.19}$$

Using the fact that $\|div\mathbf{p}\|_{H_{-1}} \le C\|p\|_{L_2}$ for any $p \in L_2(\Omega))$, we have

$$\begin{aligned} \|div\mathbf{r}_\epsilon\|_{-1,\Omega} \le C\|\mathbf{r}_\epsilon\|_{0,\Omega} \le C\|\epsilon_2\alpha_{ij}^k(y,z)\frac{\partial^2}{\partial x_j\partial x_k}u_0^\lambda\|_{0,\Omega} \\ + C\|\epsilon_2 a_\epsilon^{ij}\nabla_j\nabla_m u_0^\lambda\|_{0,\Omega} \le C\epsilon_2|u_0^\lambda|_{2,\Omega}. \end{aligned} \tag{4.20}$$

Following [36] it can be shown that

$$|u_0^\lambda|_{2,\Omega} \le \frac{C}{\epsilon_1}\|f\|_{0,\Omega}. \tag{4.21}$$

Combining (4.20) and (4.21) we have

$$\|div\mathbf{r}_\epsilon\|_{-1,\Omega} \le C\frac{\epsilon_2}{\epsilon_1}\|f\|_{0,\Omega}. \tag{4.22}$$

For the estimate of the second term on the r.h.s. of (4.19) we use

$$\|\epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{H^{1/2}(\partial\Omega)} = \inf \|\phi\|_{1,\Omega}, \tag{4.23}$$

where the inf is over all ϕ satisfying $\phi = \epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda$ on the boundary $\partial\Omega$ [32]. For the construction of the continuation of $\epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda$ onto Ω, we introduce a family of functions τ^ϵ satisfying the following conditions.

1. $\tau^\epsilon \in C_0^\infty(\Omega)$, $0 \le \tau^\epsilon \le 1$, $\tau^\epsilon = 1$ outside the ϵ_2 neighborhood of $\partial\Omega$.

2. $\epsilon_2|\nabla\tau^\epsilon| \le C$ in Ω, where the constant C does not depend on ϵ_i (i=1,2). Such functions can be constructed for any domain with Lipschitz boundary. Then we have

$$\begin{aligned}
\|\epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{H^{1/2}(\partial\Omega)} &\le \|(1-\tau^\epsilon)\epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{H^1(\Omega)} \\
&\le C\|(1-\tau^\epsilon)\epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{0,\Omega} + C\|(\nabla\tau^\epsilon)\epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{0,\Omega} \\
&\quad + C\|(1-\tau^\epsilon)\epsilon_2 \nabla\chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{0,\Omega} + C\|(1-\tau^\epsilon)\epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla\nabla_m u_0^\lambda\|_{0,\Omega} \\
&\le C\epsilon_2\|u_0^\lambda\|_{1,\Omega} + C\|(\nabla\tau^\epsilon)\epsilon_2 \chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{0,\Omega} \\
&\quad + C\|(1-\tau^\epsilon)\epsilon_2 \nabla\chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{0,\Omega} + C\epsilon_2\|u_0^\lambda\|_{2,\Omega} \\
&\le C\epsilon_2\|u_0^\lambda\|_{1,\Omega} + \|\epsilon_2\nabla\tau^\epsilon\|_{L^{2+\eta}(\Omega)}|u_0^\lambda|_{W^{1,2+4/\eta}(\Omega)} \\
&\quad + C\|1-\tau^\epsilon\|_{L^{2+\eta}(\Omega)}|u_0^\lambda|_{W^{1,2+4/\eta}(\Omega)} + C\epsilon_2|u_0|_{2,\Omega}
\end{aligned} \tag{4.24}$$

where $\eta > 0$. Furthermore using the inequalities

$$\begin{aligned}
\|u_0^\lambda\|_{2,\Omega} \le \frac{C}{\epsilon_1}\|f\|_{0,\Omega}, \qquad \|\epsilon_2\nabla\tau^\epsilon\|_{L^{2+\eta}(\Omega)} \le C\epsilon_2^{1/(2+\eta)} \\
\|1-\tau^\epsilon\|_{L^{2+\eta}(\Omega)} \le C\epsilon_2^{1/(2+\eta)}, \qquad \|\nabla u_0^\lambda\|_{L^{2+4/\eta}(\Omega)} \le C\|f\|_{0,\Omega}
\end{aligned} \tag{4.25}$$

we conclude (4.14). Note that the last inequality follows from Theorem 4(i) of [3]. □

Remark 4.3.2. For convex polygon domains we assume that u_ϵ (or any partially homogenized part of it) is in $C^2(\Omega)$ for the fixed ϵ and

$$\|u_\epsilon\|_{C^1(\Omega)} \le C, \quad \|u_\epsilon\|_{C^2(\Omega)} \le \frac{C}{\epsilon_n} \tag{4.26}$$

where ϵ_n is the smallest scale in (4.2) (or in the partially homogenized problem). This

assumption is true for fixed ϵ_i $(i = 1, \ldots, n)$ under the compatibility conditions stated in [4]. Under the assumptions in (4.26), for convex polygonal domains we have

$$\|\theta_\epsilon\|_{H^1(\Omega)} \leq C\frac{\epsilon_2}{\epsilon_1} + C_1\sqrt{\epsilon_2}. \tag{4.27}$$

This can be derived by letting $\eta = 0$ in (4.24).

Remark 4.3.3. For further convenience the quantities which depend on $\epsilon_i/\epsilon_{i-1}$ with $2 \leq i \leq n$ and ϵ_i, $i \geq 1$, we denote $O(\epsilon)$. It indicates that $O(\epsilon)$ is a small quantity independent of the mesh size.

The above procedure can be applied for homogenization of many scale problems. Instead of (4.12), using the reiterated homogenization we find

$$u_\epsilon = u_0(x, \frac{x}{\epsilon_1}, \ldots, \frac{x}{\epsilon_m}) + u_\epsilon^1 + \theta. \tag{4.28}$$

Here u_0 is the partially homogenized part of u_ϵ over the scales $\epsilon_{m+1}, \ldots, \epsilon_n$; u_ϵ^1 is given by

$$u_\epsilon^1 = \sum_{i=1}^{n-m+1} \Big(\sum_{m_k, m<m_1\leq\cdots\leq m_i\leq n} \epsilon_{m_i}\chi^{k_i}(\frac{x}{\epsilon_{m_i}})(\epsilon_{m_{i-1}}\nabla_{k_i}\chi^{k_{i-1}}(\frac{x}{\epsilon_{m_{i-1}}}))\ldots(\epsilon_{m_1}\nabla_{k_2}\chi^{k_1}(\frac{x}{\epsilon_{m_1}}))\nabla_{k_1}u_0\Big); \tag{4.29}$$

θ is the remaining part. Furthermore, similar to a two-scale case it can be shown shown that for smooth domains Ω

$$\|\theta_\epsilon\|_{H^1(\Omega)} \leq O(\epsilon)\|f\|_{0,\Omega}$$

with

$$O(\epsilon) = C_1\epsilon_{m+1}^{1/(2+\eta)} + C_2 \max_{i\geq m+1} \frac{\epsilon_i}{\epsilon_{i-1}} \quad (\eta > 0).$$

Moreover, we use the regularity estimate

$$\|u_\epsilon\|_{2,\Omega} \leq C\|f\|_{0,\Omega}/\epsilon_n, \tag{4.30}$$

which can be obtained, for example, by following the derivation in [36]. For convex polygonal

domains, the estimate is similar to (4.27):

$$\|\theta_\epsilon\|_{H^1(\Omega)} \leq C_1 \epsilon_{m+1}^{1/2} + C_2 \max_{i \geq m+1} \frac{\epsilon_i}{\epsilon_{i-1}}. \tag{4.31}$$

4.4 H^1 estimates

For simplicity, we first present the estimates for problems with two scales in detail. The estimate for the multiscale case can be obtained following the same approach. The only difference is that the expansion of the oscillatory solution in the multiscale case contain more terms which complicates the proof.

4.4.1 H^1 estimates for two scale case

In this section we analyze the MsFEM for three different cases : (1) $\epsilon_1 \gg h \gg \epsilon_2$, (2) $h \gg \epsilon_1 \gg \epsilon_2$, and (3) $\epsilon_1 \gg \epsilon_2 \gg h$. As in the standard FEM we have Cea's lemma [28]:

Lemma 4.4.1 *Let u and u^h be the solutions of (4.1) and (4.4) respectively. then*

$$\|u - u^h\|_{1,\Omega} \leq C \frac{\beta}{\alpha} \|u - v\|_{1,\Omega}, \quad \forall v \in V^h \tag{4.32}$$

Below we need the following lemma:

Lemma 4.4.2 *For a rectangular domain with the sides of order h we have*

$$\|f\|_{L_2(\partial K)} \leq C h^{1/2} |f|_{H^1(K)} + C h^{-1/2} \|f\|_{L_2(K)}. \tag{4.33}$$

This lemma can be derived from the standard trace inequality [1] using the scaling argument. We omit the proof of the lemma.

Case 1: $\epsilon_1 \gg h \gg \epsilon_2$.

Theorem 4.4.3 *Let u_ϵ and u_ϵ^h be the solution and MsFEM solution of (4.1) respectively. Then*

$$\|u_\epsilon - u_\epsilon^h\|_{1,\Omega} \leq C_1 \frac{h}{\epsilon_1} + C_2 (\frac{\epsilon_2}{h})^{1/2}. \tag{4.34}$$

Proof. Define $v_\epsilon^h \in V^h$ such that in each $K \in \mathbf{K}^h$

$$v_\epsilon^h(x) = (I_k u)(x) = \sum_{j=1}^{d} \alpha_j \phi^j(x) \tag{4.35}$$

with $\alpha_j = u_0(x_j)$ where u_0 is the partially homogenized part of u_ϵ over the scale ϵ_2 and x_j are the nodal points of K. Then in any element $K \in \mathbf{K}^h$, we have

$$\nabla_i a_\epsilon^{ij} \nabla_j (u_\epsilon - v_\epsilon^h) = f \quad in \quad K, \tag{4.36}$$

where v_ϵ^h on ∂K is a piecewise linear function whose values at the nodal points are $u_0(x_j)$.

We divide the solution of (4.36) into two parts: $(u_\epsilon - v_\epsilon^h) = (u_\epsilon - v_\epsilon^h)_1 + (u_\epsilon - v_\epsilon^h)_2$, where $(u_\epsilon - v_\epsilon^h)_1$ and $(u_\epsilon - v_\epsilon^h)_2$ satisfy

$$\begin{aligned} \nabla_i a_{ij}^\epsilon \nabla_j (u_\epsilon - v_\epsilon^h)_1 &= f \quad in \quad K, \\ (u_\epsilon - v_\epsilon^h)_1 &= 0 \quad on \quad \partial K; \end{aligned} \tag{4.37}$$

$$\begin{aligned} \nabla_i a_\epsilon^{ij} \nabla_j (u_\epsilon - v_\epsilon^h)_2 &= 0 \ \text{ in } \ K, \\ (u_\epsilon - v_\epsilon^h)_1 &= (u_\epsilon - v_\epsilon^h) \ \text{ on } \ \partial K; \end{aligned} \tag{4.38}$$

respectively. $(u_\epsilon - v_\epsilon^h)_1$ can be estimated from (4.37) and the Poincare inequality:

$$\|(u_\epsilon - v_\epsilon^h)_1\|_{H^1(K)} \le Ch \|f\|_{L_2(K)}. \tag{4.39}$$

As for $(u_\epsilon - v_\epsilon^h)_2$, using the expansion over the scale ϵ_2, $u_\epsilon = u_0 + \epsilon_2 \chi_\lambda^m \nabla_m u_0 + \theta_\epsilon$, and (4.38) we have

$$\begin{aligned} \|(u_\epsilon - v_\epsilon^h)_2\|_{H^1(K)} &\le \|u_\epsilon - v_\epsilon^h\|_{H^{1/2}(\partial K)} \\ &\le C\|u_0 + \epsilon_2 \chi_\lambda^m (\frac{x}{\epsilon_2}) \nabla_m u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} + C\|\theta_\epsilon\|_{H^{1/2}(\partial K)} \\ &\le C\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} + C\|\epsilon_2 \chi_\lambda^m (\frac{x}{\epsilon_2}) \nabla_m u_0\|_{H^{1/2}(\partial K)} + C\|\theta_\epsilon\|_{H^1(K)}. \end{aligned} \tag{4.40}$$

Combination of (4.39) and (4.40) gives

$$\begin{aligned}\|u_\epsilon - v_\epsilon^h\|_{H^1(K)} \le C_1 h\|f\|_{0,K} + C\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} \\ + C\|\epsilon_2\chi^m(\frac{x}{\epsilon_2})\nabla_m u_0\|_{H^{1/2}(\partial K)} + C\|\theta_\epsilon\|_{H^1(K)}.\end{aligned} \tag{4.41}$$

The third term on the r.h.s. of this inequality can be estimated using the interpolation inequality as

$$\begin{aligned}\|\epsilon_2\chi^m(\frac{x}{\epsilon_2})\nabla_m u_0\|_{H^{1/2}(\partial K)} &\le \|\epsilon_2\chi^m(\frac{x}{\epsilon_2})\nabla_m u_0\|^{1/2}_{L_2(\partial K)}\|\epsilon_2\chi^m(\frac{x}{\epsilon_2})\nabla_m u_0\|^{1/2}_{H^1(\partial K)} \\ &\le C\epsilon_2^{1/2}\|\nabla u_0\|^{1/2}_{L_2(\partial K)}(\|\nabla u_0\|^{1/2}_{L_2(\partial K)} + \epsilon_2^{1/2}|\nabla u_0|^{1/2}_{H^1(\partial K)}) \\ &= C\epsilon_2^{1/2}\|\nabla u_0\|_{L_2(\partial K)} + C\epsilon_2\|u_0\|^{1/2}_{L_2(\partial K)}|\nabla u_0|^{1/2}_{H^1(\partial K)} \\ &\le C\epsilon_2^{1/2}\|\nabla u_0\|_{L_2(K)}h^{-1/2} + C(\epsilon_2 h)^{1/2}|\nabla u_0|_{H^1(K)} + C\frac{\epsilon_2}{\epsilon_1^{1/2}}h^{1/2}.\end{aligned} \tag{4.42}$$

In the last step we have used Lemma 4.4.2 for the first term and the assumption that $\|u_0\|_{C^2(\Omega)} \le C/\epsilon_1$ (see (4.26)) for the second term.

For the second term on the r.h.s. of (4.41) we can write

$$\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} = \inf_{\phi|_{\partial K} = u_0 - v_\epsilon^h} \|\phi\|_{H^1(K)}. \tag{4.43}$$

Since v_ϵ^h is the linear function on ∂K we can extend it as a bilinear function $\tilde{v}_\epsilon^h$ onto K. Noticing the fact that because of (4.35) u_0 and $\tilde{v}_\epsilon^h$ coincide at the nodal points. Then we have

$$\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} \le \|u_0 - \tilde{v}_\epsilon^h\|_{H^1(K)} \le Ch|u_0|_{2,K}. \tag{4.44}$$

Combining (4.42) and (4.44) we have

$$\begin{aligned}\|u_\epsilon - v_\epsilon^h\|_{1,K} \le Ch\|f\|_{0,K} + Ch|u_0|_{2,K} + C\epsilon_2^{1/2}h^{-1/2}\|\nabla u_0\|_{L_2(K)} \\ + C\epsilon_2^{1/2}h^{1/2}|\nabla u_0|_{H^1(K)} + C\frac{\epsilon_2}{\epsilon_1^{1/2}}h^{1/2} + C\|\theta_\epsilon\|_{H^1(K)}.\end{aligned} \tag{4.45}$$

Summing (4.45) over all $K \subset \mathbf{K}$ and using Cea's lemma we get

$$\begin{aligned}\|u_\epsilon - u_\epsilon^h\|_{1,\Omega} &\leq Ch\|f\|_{0,\Omega} + Ch|u_0|_{2,\Omega} + C\epsilon_2^{1/2}h^{-1/2}\|\nabla u_0\|_{L_2(\Omega)} \\ &\quad + C\epsilon_2^{1/2}h^{1/2}|\nabla u_0|_{H^1(\Omega)} + C\frac{\epsilon_2}{\epsilon_1^{1/2}h^{1/2}} + C_4\|\theta_\epsilon\|_{H^1(\Omega)} \\ &\leq (Ch + C\frac{h}{\epsilon_1} + C_2\sqrt{\frac{\epsilon_2}{h}} + C_3\frac{\sqrt{\epsilon_2 h}}{\epsilon_1})\|f\|_{0,\Omega} + C\frac{\epsilon_2}{\sqrt{\epsilon_1 h}} + C\frac{\epsilon_2}{\epsilon_1} + C\sqrt{\epsilon_2}.\end{aligned} \tag{4.46}$$

In the last step we have used (4.30) and the estimate for $\|\theta_\epsilon\|_{H^1(\Omega)}$ (see (4.27)). Note that in (4.46) $\sqrt{\epsilon_2 h}/\epsilon_1$, $\epsilon_2/\sqrt{\epsilon_1 h}$, and $\sqrt{\epsilon_2}$ are much smaller than $\sqrt{\epsilon_2/h}$, and $\epsilon_2/\epsilon_1 \ll h/\epsilon_1$. Thus, dropping the lower order terms and using Cea's lemma we get (4.34). □

Case 2: $h \gg \epsilon_1 \gg \epsilon_2$

Theorem 4.4.4 *Let u_ϵ and u_ϵ^h be the solution and MsFEM solution of (4.1) respectively. Then*

$$\|u_\epsilon - u_\epsilon^h\|_{1,\Omega} \leq C_1 h + C_2(\frac{\epsilon_1}{h})^{1/2} + C_3\frac{\epsilon_2}{\epsilon_1}. \tag{4.47}$$

Proof. We define v_ϵ^h as in (4.35) with u_0 denoting the fully homogenized solution over the scales ϵ_1 and ϵ_2. Using the partition $u_\epsilon - v_\epsilon^h = (u_\epsilon - v_\epsilon^h)_1 + (u_\epsilon - v_\epsilon^h)_2$ as in case 1, and taking into account the inequality (4.39) for $(u_\epsilon - v_\epsilon^h)_1$ and the expansion of u_ϵ over the scales ϵ_1 and ϵ_2 (see (4.29))

$$u_\epsilon = u_0 + \epsilon_1\chi^m(\frac{x}{\epsilon_1})\nabla_m u_0 + \epsilon_2\chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0 + \epsilon_2\chi_\lambda^m(\frac{x}{\epsilon_2})(\epsilon_1\nabla_m\chi^l(\frac{x}{\epsilon_1})\nabla_l u_0) + \theta',$$

we have

$$\begin{aligned}\|u_\epsilon - v_\epsilon^h\|_{H^1(K)} &\leq \|(u_\epsilon - v_\epsilon^h)_1\|_{H^1(K)} + \|(u_\epsilon - v_\epsilon^h)_2\|_{H^1(K)} \\ &\leq Ch\|f\|_{0,K} + C\|u_\epsilon - v_\epsilon^h\|_{H^{1/2}(\partial K)} \\ &\leq C\|f\|_{0,\Omega} + C(\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} + \|\epsilon_1\chi^m(\frac{x}{\epsilon_1})\nabla_m u_0\|_{H^{1/2}(\partial K)} \\ &\quad + \|\epsilon_2\chi_\lambda^m(\frac{x}{\epsilon_2})\nabla_m u_0\|_{H^{1/2}(\partial K)} + \|\epsilon_2\chi_\lambda^m(\frac{x}{\epsilon_2})(\epsilon_1\nabla_m\chi^l(\frac{x}{\epsilon_1})\nabla_l u_0)\|_{H^{1/2}(\partial K)} \\ &\quad + \|\theta'\|_{H^1(K)}).\end{aligned} \tag{4.48}$$

Treating the third, fourth, and fifth terms on the r.h.s of (4.48) similarly as in (4.42) (i.e., applying the interpolation inequality and Lemma 4.4.2), we get

$$\begin{aligned}\|u_\epsilon - v_\epsilon^h\|_{H^1(K)} \le\ & Ch\|f\|_{0,K} + C\|u_\epsilon - v_\epsilon^h\|_{H^{1/2}(\partial K)} \\ & + C\epsilon_1 h^{-1/2}\|\nabla u_0\|_{0,K} + C\epsilon_1^{1/2}h^{1/2}|\nabla u_0|_{1,K} + C\epsilon_1 h^{1/2} \\ & + C\sqrt{\frac{\epsilon_2}{h}}\|\nabla u_0\|_{0,K} + C\epsilon_2^{1/2}h^{1/2}|\nabla u_0|_{1,K} + C\epsilon_2 h^{1/2} + C\|\theta'\|_{1,K}.\end{aligned} \tag{4.49}$$

For the second term on the r.h.s. of (4.49) since v_ϵ^h is a linear function on ∂K we continue it as a bilinear function $\tilde{v}_\epsilon^h$ onto K. Noticing that the values of u_0 and $\tilde{v}_\epsilon^h$ coincide at the nodal points we have

$$\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} \le \|u_0 - \tilde{v}_\epsilon^h\|_{H^1(K)} \le Ch\|u_0\|_{2,K}. \tag{4.50}$$

Finally, summing (4.49) over $K \in \mathbf{K}^h$ and using (4.26), (4.50), and the fact $|u_0|_{2,\Omega} \le C\|f\|_{0,\Omega}$ we get

$$\begin{aligned}\|u_\epsilon - v_\epsilon^h\|_{1,\Omega} \le\ & C(h + \sqrt{\frac{\epsilon_1}{h}} + (\epsilon_1 h)^{1/2} + \sqrt{\frac{\epsilon_2}{h}} + (\epsilon_2 h)^{1/2})\|f\|_{0,\Omega} \\ & + C\epsilon_1 h^{-1/2} + C\epsilon_2 h^{-1/2} + C\frac{\epsilon_2}{\epsilon_1} + C\sqrt{\epsilon_1}.\end{aligned} \tag{4.51}$$

Here we have used the estimate for $\|\theta'\|_{1,\Omega}$ (4.31). In (4.51) we may neglect $\sqrt{\epsilon_1 h}$, $\epsilon_1/\sqrt{h}$, and $\sqrt{\epsilon_2}$ as they are lower order comparing with $\sqrt{\epsilon_1/h}$. Thus, (4.47) follows from (4.51) and Cea's lemma . □

Case 3: $h \ll \epsilon_2 \ll \epsilon_1$

Theorem 4.4.5 *Let u_ϵ and u_ϵ^h be the solution and MsFEM solution of (4.1) respectively. Then*

$$\|u_\epsilon - u_\epsilon^h\|_{1,\Omega} \le C\frac{h}{\epsilon_2}\|f\|_{0,\Omega}. \tag{4.52}$$

Proof. Define v_ϵ^h as in (4.35) and take α_j to be the nodal value of u_ϵ at x_j. Again using the partition of $u_\epsilon - v_\epsilon^h$, we obtain

$$\|u_\epsilon - v_\epsilon^h\|_{H^1(K)} \leq Ch\|f\|_{0,K} + \|u_\epsilon - v_\epsilon^h\|_{H^{1/2}(\partial K)}. \tag{4.53}$$

Continuing v_ϵ^h from ∂K onto K as a bilinear interpolant of u_ϵ in K, $\tilde{v}_\epsilon^h$, yields

$$\|u_\epsilon - v_\epsilon^h\|_{H^{1/2}(\partial K)} \leq \|u_\epsilon - \tilde{v}_\epsilon^h\|_{H^1(K)} \leq Ch\|u_\epsilon\|_{2,K}. \tag{4.54}$$

Summing (4.53) over all K, using (4.54), (4.30), and Cea's lemma we get (4.52). □

4.4.2 H^1 error estimates for many scale case

Without loss of generality we assume the order of h is between the scales ϵ_{m+1} and ϵ_m for $m \in [1, n]$, i.e.,

$$\epsilon_1 \gg \cdots \gg \epsilon_m \gg h \gg \epsilon_{m+1} \cdots \gg \epsilon_n.$$

We have

Theorem 4.4.6 *Let u_ϵ and u_ϵ^h be the solution and MsFEM solution of (4.1) respectively. Then*

$$\|u_\epsilon - u_\epsilon^h\|_{1,\Omega} \leq C_1\frac{h}{\epsilon_m} + C_2(\frac{\epsilon_{m+1}}{h})^{1/2} + C_3 \max_{i \geq m+1} \frac{\epsilon_i}{\epsilon_{i-1}}. \tag{4.55}$$

Proof. The proof follows the same steps as in Section 4.1. First, we define v_ϵ^h as in (4.35) and $\alpha_j = u_0(x_j)$, u_0 being the partially homogenized part of u_ϵ over the scales $\epsilon_{m+1}, \ldots, \epsilon_n$. In this case, the expansion of u_ϵ is given by (4.28):

$$u_\epsilon = u_0(x, x/\epsilon_1, \cdots, x/\epsilon_m) + u_\epsilon^1 + \theta,$$

where u_ϵ^1 is given by (4.29). It is easy to show that

$$\begin{aligned}\|u_\epsilon - v_\epsilon^h\|_{H^1(K)} &\le Ch\|f\|_{0,K} + C\|u_\epsilon - v_\epsilon^h\|_{H^{1/2}(\partial K)} \\ &\le Ch\|f\|_{0,K} + C\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} + C\|u_\epsilon^1\|_{H^{1/2}(\partial K)} + C\|\theta\|_{1,K} \qquad (4.56) \\ &\le Ch\|f\|_{0,K} + Ch|u_0|_{2,K} + C\|u_\epsilon^1\|_{H^{1/2}(\partial K)} + C\|\theta\|_{1,K}.\end{aligned}$$

In the last step we extended v_ϵ^h as $\tilde{v}_\epsilon^h$, the bilinear interpolant of u_0 onto K. Then we have

$$\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} \le Ch|u_0|_{2,K}. \qquad (4.57)$$

Now we only need to estimate the third term on the r.h.s. of (4.56). For this we need the following estimate:

$$|u_\epsilon^1|_{1,\partial K} \le C\|\nabla u_0\|_{0,\partial K} + C\epsilon_{m+1}|\nabla u_0|_{1,\partial K}. \qquad (4.58)$$

In fact, by (4.29) u_ϵ^1 can be written in the form of $C_\epsilon(x)\cdot\nabla u_0$, where $C_\epsilon(x)$ contains the linear combinations of products of $\chi_k(\frac{x}{\epsilon_k})$ $(k = m+1, \dots, n)$ and their gradients. Furthermore, it can be verified that

$$\max_x |\nabla C_\epsilon(x)| \le C + \max_{i\ge m+1}(\epsilon_i/\epsilon_{i-1}) \le C \quad \text{and} \quad \max_x |C_\epsilon(x)| \le C\epsilon_{m+1},$$

and hence ∇u_ϵ^1 is bounded. Therefore,

$$\begin{aligned}\|\nabla u_\epsilon^1\|_{0,\partial K} &\le \|\nabla C_\epsilon(x)\cdot\nabla u_0\|_{0,\partial K} + \|C_\epsilon(x)\cdot\nabla^2 u_0\|_{0,\partial K} \\ &\le \max_x |\nabla C_\epsilon(x)|\|\nabla u_0\|_{0,\partial K} + \max_x |C_\epsilon(x)|\|\nabla^2 u_0\|_{0,\partial K} \\ &\le C\|\nabla u_0\|_{0,\partial K} + C\epsilon_{m+1}\|\nabla^2 u_0\|_{0,\partial K} + C\max_{i\ge m+1}(\epsilon_i/\epsilon_{i-1})\|\nabla u_0\|_{0,\partial K} \\ &\le C\|\nabla u_0\|_{0,\partial K} + C\epsilon_{m+1}|\nabla u_0|_{1,\partial K}.\end{aligned}$$

From (4.58), the interpolation inequality, the trace inequality (4.33), and (4.26) it follows that

$$\begin{aligned}
\|u_\epsilon^1\|_{H^{1/2}(\partial K)} &\le \|u_\epsilon^1\|_{L_2(\partial K)}^{1/2}\|u_\epsilon^1\|_{H^1(\partial K)}^{1/2} \\
&\le C\epsilon_{m+1}^{1/2}\|\nabla u_0\|_{L_2(\partial K)}^{1/2}(\|\nabla u_0\|_{L_2(\partial K)}^{1/2} + \epsilon_{m+1}^{1/2}|\nabla u_0|_{H^1(\partial K)}) \\
&\le C\epsilon_{m+1}^{1/2}\|\nabla u_0\|_{L_2(\partial K)} + C\epsilon_{m+1}{\epsilon_m}^{-1/2}h^{1/2} \\
&\le C\epsilon_{m+1}^{1/2}h^{-1/2}\|\nabla u_0\|_{0,K} + C\epsilon_{m+1}^{1/2}h^{1/2}|\nabla u_0|_{1,K} + C\epsilon_{m+1}h^{1/2}\epsilon_m^{-1/2}.
\end{aligned} \tag{4.59}$$

Substituting (4.59) into (4.56) and then summing (4.56) over all $K \in \mathbf{K}^h$, we have

$$\begin{aligned}
\|u_\epsilon - v_\epsilon^h\|_{1,\Omega} &\le C(h + \sqrt{\frac{\epsilon_{m+1}}{h}} + \frac{h}{\epsilon_m} + \frac{\epsilon_{m+1}^{1/2}h^{1/2}}{\epsilon_m})\|f\|_{0,\Omega} + C\frac{\epsilon_{m+1}}{\sqrt{\epsilon_m h}} + C\|\theta_\epsilon\|_{1,\Omega} \\
&\le C(h + \sqrt{\frac{\epsilon_{m+1}}{h}} + \frac{h}{\epsilon_m} + \frac{\epsilon_{m+1}^{1/2}h^{1/2}}{\epsilon_m})\|f\|_{0,\Omega} + C\frac{\epsilon_{m+1}}{\sqrt{\epsilon_m h}} + C\max_{i\ge m+1}\frac{\epsilon_i}{\epsilon_{i-1}}.
\end{aligned} \tag{4.60}$$

The estimate (4.55) follows from (4.60), (4.31) (estimate of $\|\theta_\epsilon\|_{1,\Omega}$), and Cea's lemma. The lower order terms with h ($\ll h/\epsilon_m$) and with $\sqrt{\epsilon_{m+1}h}/\epsilon_m$, $\epsilon_{m+1}/\sqrt{\epsilon_m h}$, and $\sqrt{\epsilon_{m+1}}$ (much smaller than $\sqrt{\epsilon_{m+1}/h}$) are neglected. □

Remark 4.4.1. The estimate (4.55) shows that the error of the MsFEM solution becomes larger as h approaches either ϵ_m or ϵ_{m+1}. This is the resonance phenomenon mentioned in the introduction. Furthermore, we see from (4.60) that both resonance errors come from the first term on the r.h.s of the last inequality when $h \sim \epsilon_m$ or $h \sim \epsilon_{m+1}$, in which cases the other two terms are of lower order.

4.5 L_2 estimates

In this section we derive the L_2 estimates using the H^1 estimates. Because some terms in the H^1 estimate (4.60) cannot be expressed through $\|f\|_{0,\Omega}$, the estimates we obtain from the duality argument is not optimum when h is not comparable with the physical scales. The estimates do capture the correct order of the resonance error, which are the leading order error when h is comparable to the physical scales. Employing the method introduced in [28], which compares the discrete values of the solution and the numerical solution, we can show that the order of the method is not much affected by those small terms not expressed through $\|f\|_{0,\Omega}$. The present approach, however, is more concise to present. We use the

following abstract lemma.

Lemma 4.5.1 *Let u and u^h be the solutions of (3.3) and (4.4) respectively with V^h consisting of conforming base functions. If*

$$\|u - u^h\|_{1,\Omega} \leq C_1\gamma\|f\|_{0,\Omega} + C_2\delta \tag{4.61}$$

where γ and δ are small positive quantities, $C_1 \geq 0$, $C_2 \geq 0$, then

$$\|u - u^h\|_{0,\Omega} \leq C_3\gamma^2\|f\|_{0,\Omega} + C_4\sqrt{\gamma\delta\|f\|_{0,\Omega}} + C_5\delta.$$

Proof. We use the Aubin-Nitshe trick as follows. Let w be the solution of (3.3) with $f = u - u^h$, i.e., $w \in H_0^1(\Omega)$ satisfies

$$a(w, v) = (u - u^h, v), \quad \forall v \in H_0^1(\Omega). \tag{4.62}$$

Let $w_h \in V^h$ be the interpolant of w. Then (4.61) implies

$$\|w - w^h\|_{1,\Omega} \leq C_1\gamma\|u - u^h\|_{0,\Omega} + C_2\delta. \tag{4.63}$$

Choosing $v = u - u^h$ in (4.62) we have

$$\begin{aligned}\|u - u^h\|_{0,\Omega}^2 = a(u - u^h, w) = a(u - u^h, w - w_h) &\leq \beta\|u - u^u\|_{1,\Omega}\|w - w_I\|_{1,\Omega} \\ &\leq \beta(C_1\gamma\|f\|_{0,\Omega} + C_2\delta)(C_1\gamma\|u - u^h\|_{0,\Omega} + C_2\delta) \\ &\leq \|u - u^h\|_{0,\Omega}(C_1\gamma^2\|f\|_{0,\Omega} + C_2\delta\gamma) + C\delta\gamma\|f\|_{0,\Omega} + C_1\delta^2.\end{aligned} \tag{4.64}$$

Therefore,

$$\begin{aligned}\|u - u^h\|_{0,\Omega} &\leq |C_1\gamma^2\|f\|_{0,\Omega} + C_1\delta\gamma| + \sqrt{(C_1\gamma^2\|f\|_{0,\Omega} + C_1\delta\gamma)^2 + 4(C\delta\gamma\|f\|_{0,\Omega} + C_1\delta^2)} \\ &\leq C\gamma^2\|f\|_{0,\Omega} + C_1\delta\gamma + C\gamma^2\|f\|_{0,\Omega} + C_1\delta\gamma + C_2\sqrt{\delta\gamma\|f\|_{0,\Omega}} + C_3\delta \\ &\leq C\gamma^2\|f\|_{0,\Omega} + C_1\sqrt{\gamma\delta\|f\|_{0,\Omega}} + C_3\delta.\end{aligned} \tag{4.65}$$

□

In our H^1 estimates, γ contains the resonance errors which may become $O(1)$ depending on h. But δ is a small quantity independent of the mesh size. In particular, $\gamma = C(h + \epsilon_{m+1}^{1/2}/h^{1/2} + h/\epsilon_m + \epsilon_{m+1}^{1/2}h^{1/2}/\epsilon_m)$ and $\delta = C\epsilon_{m+1}\epsilon_m^{-1/2}h^{-1/2} + O(\epsilon)$ (see (4.60)). Note that when h becomes comparable to any physical scale, γ becomes of order one due to either h/ϵ_m or $\sqrt{\epsilon_{m+1}/h}$ while δ remains small. Consequently, the term $\sqrt{\gamma\delta}$ is a small quantity which does not resonate at any scale of the problem, i.e., $\sqrt{\gamma\delta} = O(\epsilon)$, since it depends only on $\epsilon_i/\epsilon_{i-1}$, $2 \leq i \leq n$, and ϵ_i, $i \geq 1$. Thus γ^2 is the dominating resonance error for the case when h becomes comparable with the physical scales. We note that some terms of γ and δ (see(4.60) do not change when we decrease the mesh size and the physical scales at the same time, but remain negligible with respect to h/ϵ_m or $\sqrt{\epsilon_{m+1}/h}$ when h is comparable with any physical scale. These terms may be called lower order resonance terms.

Applying Lemma 4.5.1 to the cases analyzed in section 4.4 we have the following resonance errors:

Theorem 4.5.2 *Let u_ϵ and u_ϵ^h be the solution and MsFEM solution of (4.1) respectively. Then*

1. *Case $\epsilon_1 \gg h \gg \epsilon_2$.*

$$\|u_\epsilon - u_\epsilon^h\|_{0,\Omega} \leq C_1(\frac{h}{\epsilon_1})^2 + C_2\frac{\epsilon_2}{h} \tag{4.66}$$

2. *Case $h \gg \epsilon_1 \gg \epsilon_2$.*

$$\|u_\epsilon - u_\epsilon^h\|_{0,\Omega} \leq C_1h^2 + C_2\frac{\epsilon_1}{h} \tag{4.67}$$

3. *Case $\epsilon_1 \gg \epsilon_2 \gg h$.*

$$\|u_\epsilon - u_\epsilon^h\|_{0,\Omega} \leq C(\frac{\epsilon_2}{h})^2 \tag{4.68}$$

4. *Case $\epsilon_1 \gg \cdots \gg \epsilon_m \gg h \gg \epsilon_{m+1} \cdots \gg \epsilon_n$.*

$$\|u_\epsilon - u_\epsilon^h\|_{0,\Omega} \leq C_1(\frac{h}{\epsilon_m})^2 + C_2\frac{\epsilon_{m+1}}{h} \tag{4.69}$$

4.6 Numerical experiments

In this section we study the convergence and the accuracy of the multiscale method through numerical experiments. The model problem is solved using the multiscale method with the base functions defined by (4.3) and the linear boundary conditions. Since it is very difficult to construct a test problem with both exact solution and sufficient generality, we use resolved numerical solutions in place of exact solutions. The numerical results are compared with the theoretical analysis.

The implementation of the multiscale method has been given in [26]. Here we outline the implementation and define some notations to be used below. All computations are performed on a unit square domain $\Omega = (0,1) \times (0,1)$. Let N be the number of elements in x and y directions. The mesh size is thus $h = 1/N$. To compute the base functions, each element is discretized into $M \times M$ subcell elements with size of $h_s = h/M$. Rectangular elements are used in all numerical tests.

To solve the subcell problem, we use the standard linear finite element method. After solving the base functions, the local stiffness matrix and the right hand side are computed using numerical quadrature rules. We compute the gradients of a base function at the center of a subcell element and use two-dimensional centered trapezoidal rule for the volume integration. This procedure ensures that the entries of the stiffness matrix are computed with second order accuracy. In our computations, we only solve three base functions, ϕ^i $(i = 1, 2, 3)$. The fourth one is obtained from $\phi^4 = 1 - \phi^1 - \phi^2 - \phi^3$.

In all examples below, the resolved solutions are obtained using linear FEM. Given the wave length of small scales ϵ_1 and ϵ_2, we solve the model problem twice on two meshes with one mesh size being twice of other. Then the Richardson extrapolation is used to approximate the exact solutions from numerical solutions on two meshes. Throughout our numerical experiments, both of the mesh sizes used to compute the well resolved solution are less than $\epsilon/8$, so that the error of the extrapolated solutions is less that 10^{-6}.

The main difficulty in our tests is that the choices of well separated physical small scales for test problems are severely limited by the available computer memory. Our parallel implementation of the multiscale method on an Intel Paragon computer with 512 processors enables us to test the two scale problem. Even so, we are limited to testing the cases with $\epsilon_1 \gg h \gg \epsilon_2$ and $\epsilon_1 \gg \epsilon_2 \gg h$. The case with $h \gg \epsilon_1 \gg \epsilon_2$ gives very large subcell

problems for constructing the base functions, which cannot be fitted into a single processor. Therefore, it cannot be handled by our current implementation of MsFEM.

The case with h in between the two physical scales is generic and important for practical purposes. As indicated by Theorem 6.2, the L_2 norm error is given by

$$C_1(\frac{h}{\epsilon_1})^2 + C_2\frac{\epsilon_2}{h}, \tag{4.70}$$

which consists of two resonance errors. One would expect that when h is close to ϵ_1 the first term dominates and when h is close to ϵ_2 the second term dominates. This asymptotic observation, however, is not always reflected by the numerical results shown below. There are two reasons. First, the constants C_1 and C_2 may differ by a large factor which is problem dependent. Second, it is difficult to choose well separated ϵ_1 and ϵ_2 in the numerical computations. Therefore, the two error may interact with each other. Instead of verifying each of them, in the following, we show that the numerical error does follow the estimate as a whole. Furthermore, we use the least square fitting to obtain the constants. These constants indicate the relative magnitude of the two terms.

Again, due to limitation of computing resources, in all tests below, ϵ_1 and ϵ_2 are fixed during the tests and we only allow h to vary.

Example 1. In this example, we solve (4.1) with $f = -1$, $u|_{\partial\Omega} = 0$, and

$$a^{ij} = \frac{\delta^{ij}}{3 + \sin(\frac{2\pi x}{\epsilon_1}) + \cos(\frac{2\pi y}{\epsilon_2})}, \tag{4.71}$$

where δ^{ij} are the Kroneker's symbols.

Fixing $\epsilon_1 = 0.5$ and $\epsilon_2 = 0.005$ and varying h between ϵ_1 and ϵ_2, we calculate the discrete l_2 norms of the errors between the MsFEM solution and the refined solution (Table 4.1). In this table, the presence of both h^2/ϵ_1^2 and ϵ_2/h of (4.70) are noticeable. In Fig. 4.1 we fit the l_2 errors using the least square method. We find $C_1 = 5e-4$ and $C_2 = 1.5e-3$, indicating that the second term in (4.70) is relatively large.

The result for the case $h \ll \epsilon_2 \ll \epsilon_1$ is shown in Table 4.2. For this test, $\epsilon_1 = 0.2$ and $\epsilon_2 = 0.08$ are chosen. We see that MsFEM is second order with respect to h as analyzed.

Example 2. In this example, we solve (4.1) with $f = 0$ and linear boundary conditions,

Table 4.1: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ for $\epsilon_1 \gg h \gg \epsilon_2$.

h	l_2	rate
1/8	9.42e-4	
1/16	1.61e-4	2.5
1/32	1.87e-4	
1/64	5.08e-4	-1.44
1/128	9.77e-4	-0.94
1/256	2.01e-3	-1.04

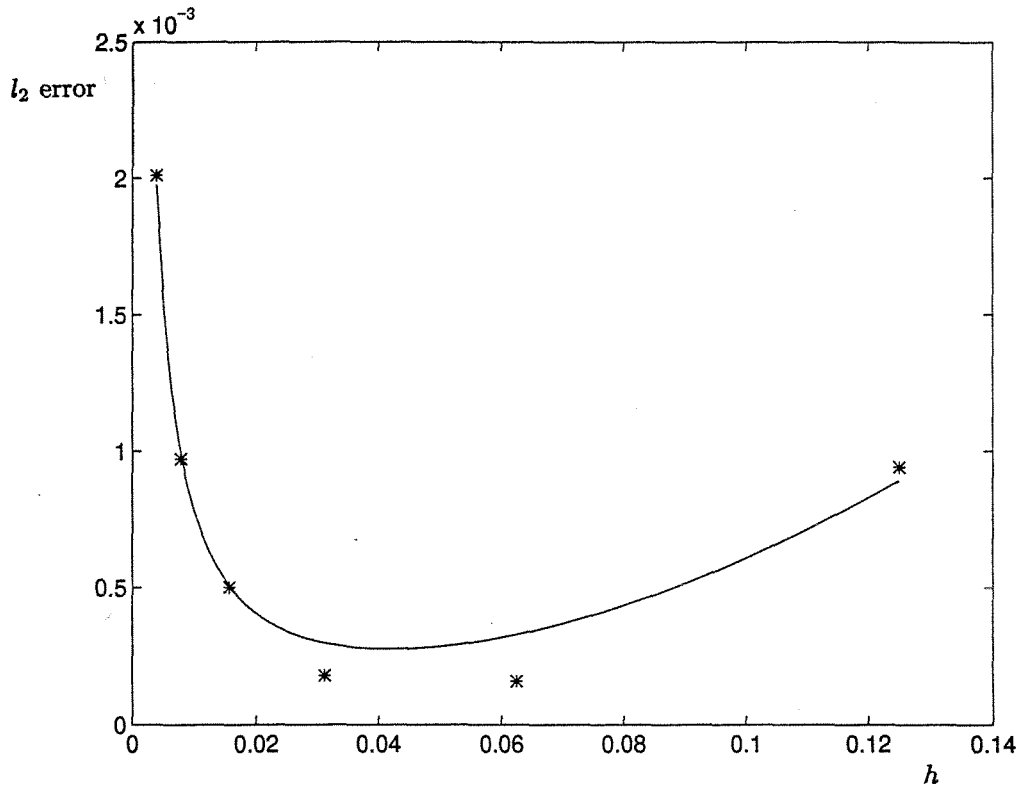

Figure 4.1: The least square fit in l_2 norm of error

Table 4.2: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ and $\|u_\epsilon - u_\epsilon^h\|_{l_\infty}$ for $h \ll \epsilon_2 \ll \epsilon_1$.

h	l_2	rate	l_∞	rate
1/32	1.10e-3		2.04e-3	
1/64	3.22e-4	1.77	5.89e-4	1.79
1/128	8.40e-5	1.98	1.52e-4	1.95
1/256	2.09e-5	2.00	3.79e-5	2.00
1/512	4.90e-6	2.09	8.93e-6	2.08
1/1024	1.22e-6	2.00	2.22e-6	2.00

Table 4.3: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ for $\epsilon_1 \gg h \gg \epsilon_2$.

h	l_2	rate
1/8	2.15e-3	
1/16	5.28e-4	2.02
1/32	3.28e-4	
1/64	1.75e-4	
1/128	1.60e-4	

Table 4.4: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ for $\epsilon_1 \gg h \gg \epsilon_2$.

h	l_2	rate
1/8	8.25e-5	
1/16	1.92e-5	2.10
1/32	1.77e-5	
1/64	1.17e-5	
1/128	3.04e-5	-1.38

$u|_{\partial\Omega} = x$, and

$$a^{ij} = \left(\frac{2+\sin(2\pi x/\epsilon_1)}{2+\cos(2\pi y/\epsilon_1)} + \frac{2+\cos 2\pi(y/\epsilon_1)}{2+\sin(2\pi x/\epsilon_1)}\right)\left(\frac{2+\sin(2\pi x/\epsilon_2)}{2+\cos(2\pi y/\epsilon_2)} + \frac{2+\cos(2\pi y/\epsilon_2)}{2+\sin(2\pi x/\epsilon_2)}\right)\delta^{ij} \quad (4.72)$$

where δ^{ij} are Kronoker's symbols.

Fixing $\epsilon_1 = 0.125$ and $\epsilon_2 = 0.0078125$, we vary h between ϵ_1 and ϵ_2. As we see from the calculated l_2 error of MsFEM (see Table 4.3), the method does not reveal the -1 order convergence for $h = O(\epsilon_2)$. Calculating C_1 and C_2 from the l_2 errors by using the least squares method, we find that $C_1 = 2.1 \times 10^{-3}$, which is much larger than $C_2 = 0.02 \times 10^{-2}$. Consequently, the h^2/ϵ_1^2 term in (4.70) has a larger weight than the ϵ_2/h term. Therefore, the $O(\epsilon_2/h)$ error is not noticeable in our test. The fitted curve as well as the data points for the numerical error are plotted in Fig. 4.2. It shows that the error varies consistently with the estimate (4.70).

Example 3. We consider another example with a^{ij} defined by (4.72), $f = -1$, and $u|_{\partial\Omega} = 0$. The calculated L_2 errors for the fixed $\epsilon_1 = 0.125$ and $\epsilon_2 = 0.0078125$ also demonstrate the presence of both h^2/ϵ_1^2 and ϵ_2/h terms in the l_2 error of the method (see Table 4.4) as h becomes close to ϵ_1 and ϵ_2, respectively. This example shows that C_1 and C_2 are indeed problem dependent. The l_2 error for $h \ll \epsilon_2 \ll \epsilon_1$ is shown in Table 4.5, where second order convergence of MsFEM is demonstrated.

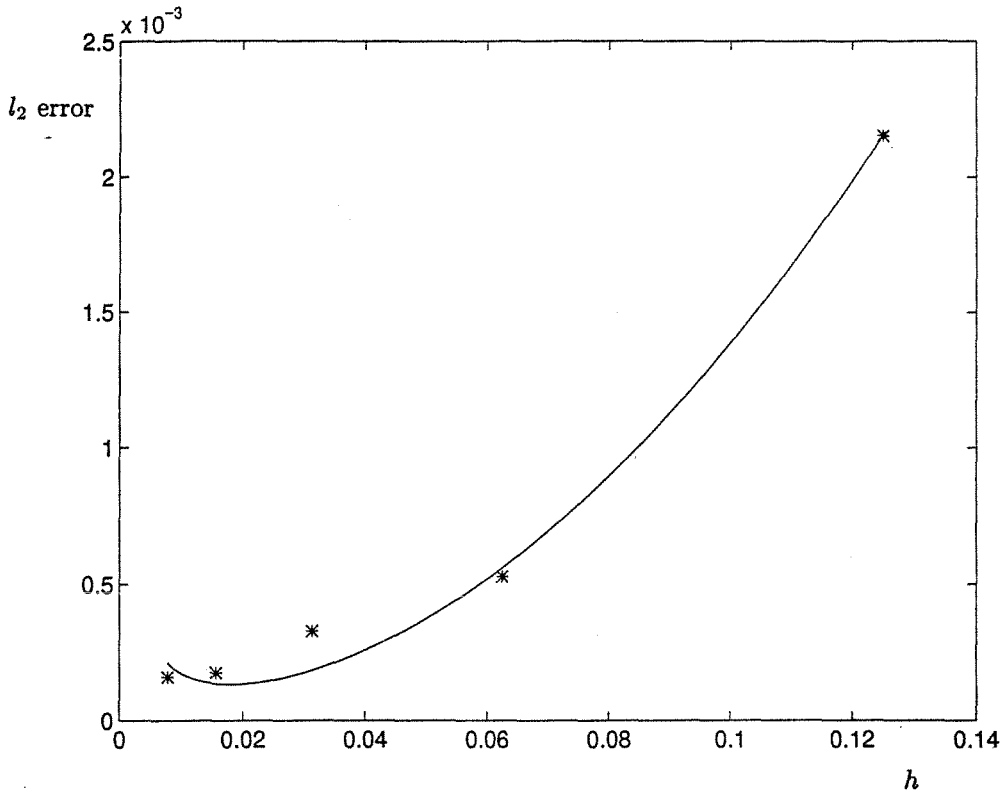

Figure 4.2: The least square fit in l_2 norm of error

Table 4.5: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ and $\|u_\epsilon - u_\epsilon^h\|_{l_\infty}$ for $h \ll \epsilon_2 \ll \epsilon_1$.

h	l_2	rate	l_∞	rate
1/64	1.14e-5		2.84e-5	
1/128	2.81e-6	2.02	5.79e-6	2.29
1/256	7.13e-7	1.97	1.56e-6	1.89
1/512	1.69e-7	2.07	3.91e-7	2.00
1/1024	3.33e-8	2.35	1.01e-7	1.95

4.7 Concluding remarks and generalizations

The purpose of the multiscale method is to provide a systematic approach to capture the small scale effect on large scales when we cannot afford to resolve all the small scale features in the physical solution. Our study shows that MsFEM is a robust method for practical multiple scale problems. In particular, the method works for multiple scale problems and when we choose the mesh size to be between the two physical scales. We note that there are two types of resonance error, one from the way of resolving the large scales, the other from capturing the small scales. The second type of error is caused by the artificial boundary layers in our base functions. This important issue and its numerical resolution, e.g., the over-sampling method, has been analyzed for problems with one small scale (see nonconforming MsFEM).

The first type of resonance error is common among traditional finite difference and finite element methods. The traditional approaches, however, cannot capture the small scales. To reduce this error, a natural idea is to generalize MsFEM to higher order in the sense that the large scales are more accurately resolved. The idea is to construct base functions such that their homogenized parts consist of higher order polynomials than linear (bilinear) functions. This can be achieved by changing the linear boundary condition of the base functions to higher order polynomials.

Denoting S^h as a finite dimensional subspace of $H_0^1(\Omega)$ such that for any

$$b(u,v) = \int_\Omega b^{ij}(x/\epsilon_1, \ldots, x/\epsilon_k)\nabla_i u \nabla_j v dx$$

with $b^{ij}(y_1, \ldots, y_k)$ are sufficiently smooth periodic functions in y_i $(i = 1, \ldots, k)$ in a unit cube, $\alpha|\xi|^2 \leq b^{ij}\xi_i\xi_j \leq \beta|\xi|^2$, and $\epsilon_1 \gg \epsilon_2 \gg \cdots \gg \epsilon_k \gg h$ we have

$$\|u - u^h\|_{1,\Omega} \leq C(\frac{h}{\epsilon_k})^n, \tag{4.73}$$

where $u \in H_0^1(\Omega)$ is the solution of $b(u,v) = (f,v)$ for any $v \in H_0^1(\Omega)$ and $u^h \in S^h$ is the solution of $b(u^h, v^h) = (f, v^h)$ for any $v^h \in S^h$. For example S^h can be the space spanned by high order polynomials of degree n [13]. In each element $K \in \mathbf{K}^h$ we denote $P_K^i \in S^h$ $(i = 1, \ldots, d)$ a set of nodal basis of S^h. Then the higher order multiscale base functions

are given by ϕ_K^i $(i = 1, \ldots, d)$, which satisfy (4.3) and the boundary conditions

$$\phi_K^i|_{\partial K} = P_K^i|_{\partial K}. \tag{4.74}$$

With these high order multiscale base functions we have

$$\|u_\epsilon - u_\epsilon^h\|_{1,\Omega} \leq Ch + C(\frac{h}{\epsilon_m})^n + C_3\sqrt{\frac{\epsilon_{m+1}}{h}} + C \max_{i \geq m+1} \frac{\epsilon_i}{\epsilon_{i-1}} \tag{4.75}$$

instead of (4.55). To show this we continue v_ϵ^h on to K as $\tilde{v}_\epsilon^h \in S^h$ in (4.57). The nodal values of the interpolant α_i in (4.35) can be defined as those of u^h satisfying (4.73). Finally taking into account the estimates for u_ϵ^1 and θ of section 4.2 we conclude (4.75). We note that S^h needs not to be a subspace of H^1. Such S^h may give nonconforming multiscale base functions.

For smooth domains Ω we enjoy the smoothness of u_ϵ (and any partially homogenized part of it) for fixed ϵ. We also have (4.26) and the estimate for $\|\theta_\epsilon\|_{1,\Omega}$ through $\|f\|_{0,\Omega}$, i.e., $\|\theta_\epsilon\|_{1,\Omega} \leq O(\epsilon)\|f\|_{0,\Omega}$. But the difficulty in deriving H^1 norm estimate for MsFEM is to define multiscale functions near the boundary. There are various ways to treat curve boundaries in the finite element methods. In fact, following the triangulation of Ω in [47] and using the nodal base functions constructed in that paper to provide the boundary conditions for the multiscale base functions, we can show that (4.75) holds on the smooth domain.

Finally we would like to note that the assumption (4.26) which requires some compatibility conditions for the problem is not necessary for deriving the H^1 estimates. Without this assumption we have

$$\|u_\epsilon - u_\epsilon^h\|_{1,\Omega} \leq C(h + \frac{h}{\epsilon_m} + (\frac{\epsilon_{m+1}}{h})^{1/2-1/p} + (\frac{\epsilon_{m+1}h}{\epsilon_m})^{1/2-1/p})\|f\|_{0,\Omega} + \|\theta_\epsilon\|_{1,\Omega} \tag{4.76}$$

where $p > 2$ is an arbitrary constant. Because of the insufficient smoothness of the homogenized parts of the solution near the corner points of the domain Ω, the rate of convergence of $\|\theta_\epsilon\|_{1,\Omega}$ to zero as $\epsilon \to 0$ deteriorates depending on the parameters of the problem.

In order to show the estimate, (4.76) we only need to reestimate $\|u_\epsilon^1\|_{H^{1/2}(\partial K)}$ in (4.56). We derive it only for the two scale case $\epsilon_1 \gg h \gg \epsilon_2$. The derivation for the general n scale case is similar.

Introducing the family of function τ_ϵ in K as in section 3 with the properties 1 and 2 we have

$$\begin{aligned}\|u_\epsilon^1\|_{H^{1/2}(\partial K)} &= \|\epsilon_2\chi^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{H^{1/2}(\partial K)} \le \|\epsilon_2(1-\tau_\epsilon)\chi^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{H^1(K)} \\ &\le \|\epsilon_2(1-\tau_\epsilon)\chi^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{L_2(K)} + \|\epsilon_2\nabla\tau_\epsilon\chi^m(\frac{x}{\epsilon_2})\nabla_m u_0^\lambda\|_{L_2(K)} \qquad (4.77) \\ &+\|\epsilon_2(1-\tau_\epsilon)\nabla(\chi^m(\frac{x}{\epsilon_2}))\nabla_m u_0^\lambda\|_{L_2(K)} + \|\epsilon_2(1-\tau_\epsilon)\chi^m(\frac{x}{\epsilon_2})\nabla\nabla_m u_0^\lambda\|_{L_2(K)}.\end{aligned}$$

Denoting the support of $1-\tau_\epsilon$ by S, we have $meas(S) \le C\epsilon_2 h$. Furthermore taking into account that $|\epsilon_2\nabla\chi| \le C$ and $|\epsilon_2\nabla\tau| \le C$, we can estimate the r.h.s of (4.77) as

$$\|u_\epsilon^1\|_{H^{1/2}(\partial K)} \le C\epsilon_2|u_0|_{H^1(K)} + C|u_0|_{H^1(S)} + C|u_0|_{H^1(S)} + C\epsilon_2|u_0|_{H^2(K)}. \qquad (4.78)$$

For the estimate of $\|\nabla u_0\|_{L_2(S)}$, we use the following facts [54]:

$$\|v\|_{L_2(S)} \le (meas(S))^{1/2-1/p}\|v\|_{L_p(S)}, \quad p > 2,$$

and

$$\|v\|_{L_p(K)} \le Ch^{-1+2/p}\|v\|_{L_2(K)} + Ch^{2/p}\sqrt{p}|v|_{H^1(K)}, \quad p \ge 1.$$

Then

$$\begin{aligned}|u_0|_{H^1(S)} \le \|\nabla u_0\|_{L_2(S)} \le C(\epsilon_2 h)^{1/2-1/p}\|\nabla u_0\|_{L_p(S)} \le \\ C(\epsilon_2 h)^{1/2-1/p}(h^{-1+2/p}\|\nabla u_0\|_{L_2(K)} + h^{2/p}\sqrt{p}|\nabla u_0|_{H^1(K)}) \le \qquad (4.79) \\ C(\epsilon_2/h)^{1/2-1/p}\|\nabla u_0\|_{L_2(K)} + C(\epsilon_2 h)^{1/2-1/p}h^{2/p}\sqrt{p}|\nabla u_0|_{H^1(K)}.\end{aligned}$$

Then using the estimates (4.78) and (4.79) for $\|u_\epsilon^1\|_{H^{1/2}(\partial K)}$ in the summation of (4.41) over all K we get (4.76). As we see the estimate (4.76) is slightly weaker than (4.60) but it does not contain the $\epsilon_{m+1}/\sqrt{\epsilon_m h}$ term of (4.60).

Chapter 5 MsFEM for discontinuous case

5.1 Summary

In this section we analyze MsFEM for elliptic equations with discontinuous coefficients. It is known that the convergence of traditional finite element methods deteriorates for the case of discontinuous parameters. The more singular the behavior of the solution at discontinuity points slower is the convergence rate of finite element method. There are different versions of finite element methods to handle the discontinuous coefficients case. Some of them include the singular functions (assuming that we know them) to the finite dimensional space of base functions [49]. These methods destroy the band structure of the stiffness matrix, and consequently require special fast linear matrix equation solvers. Some other finite element methods find the similarity solution around the discontinuity point by partitioning the domain in the neighborhood of this point. This method is known as Infinite Element Method which can effectively capture the singular behavior of the solution [55].

In this section we analyze the convergence of MsFEM in the case of discontinuous coefficients. The analysis has been performed for one periodic case, $a_\epsilon^{ij}(x) = a^{ij}(x/\epsilon)$, where $a^{ij}(y)$, $y = x/\epsilon$ is a discontinuous function. We note that the analysis for smooth coefficients case cannot be carried out in the cases of discontinuous coefficients.

We show that despite discontinuity of the parameters MsFEM retains its convergence rate

$$\|u - u^h\|_{H^1(\Omega)} \leq Ch + C_1(\frac{\epsilon}{h})^{1/2}.$$

The L_2 convergence rate has been also derived:

$$\|u - u^h\|_{L_2(\Omega)} \leq Ch^2 + C_1\frac{\epsilon}{h}.$$

This estimates demonstrate that MsFEM can also be used to capture the singular multiscale nature of the solution. We would like to note that as the number of singular points increase the complexity of Infinite Element methods increase substantially. The reason for it is that

the method needs to find similarity solution in the neighborhood of each singular point.

We confirm our estimates with numerical experiments. We present the numerical results for the cases when the boundaries of the elements are aligned with the discontinuity surfaces and when they are not. We choose checker-board models for the numerical examples.

5.2 Formulations

In this section we assume that the coefficients of the elliptic operator (3.1) a_ϵ^{ij} has the form:

$$a_\epsilon^{ij} = a^{ij}(\frac{x}{\epsilon}) \tag{5.1}$$

where $a^{ij}(y)$, $y = x/\epsilon$, is a periodic function in y in the unit cube Y . For the H^1 analysis we assume that $a^{ij}(y) \in L^\infty(Y)$. The assumption on $a^{ij}(y)$ in L_2 analysis is also sufficiently general (for example, it is sufficient for checker-board structures) and can be found in subsection 4 of this section.

The difficulty in analyzing this case is that the gradient of χ^i (3.11) is no longer continuous. this does not allow us to use the techniques we employed in the previous sections. For this reason, we treat the terms with χ^i in a separate way. As a consequence we cannot use the Aubin-Nitsche trick. We note that it was possible to use Aubin-Nitsche trick in the one periodic case.

The implementation of MsFEM is the same as described in the previous sections.

5.3 H^1 estimates

In this section we consider H^1 error estimate for MsFEM for

$$a_\epsilon^{ij}(x) = a^{ij}(\frac{x}{\epsilon})$$

where $a^{ij}(y)$ is a periodic function in Y. For H^1 analysis we assume that $a^{ij}(y) \in L^\infty(Y)$. For L_2 analysis the assumption on $a^{ij}(y)$ is sufficiently general (for example it is sufficient for checkerboard structures) and can be found in section 4.

The analysis differs from the analysis of the smooth coefficients case due to the fact that the generalized $\nabla\chi$ is not continuous anymore. As usual, we enjoy Cea's lemma.

Lemma 5.3.1 *Let u and u^h be the solutions of (4.1) and (4.4), respectively. then*

$$\|u - u^h\|_{1,\Omega} \leq C\frac{\beta}{\alpha}\|u - v\|_{1,\Omega}, \quad \forall v \in V^h. \tag{5.2}$$

Theorem 5.3.2 *Let u_ϵ and u_ϵ^h be the solution and MsFEM solution of (4.1) respectively. Then*

$$\|u_\epsilon - u_\epsilon^h\|_{1,\Omega} \leq C_1 h + C_2(\frac{\epsilon}{h})^{1/2}. \tag{5.3}$$

Proof. Define $v_\epsilon^h \in V^h$ such that in each $K \in \mathbf{K}^h$

$$v_\epsilon^h(x) = (I_k u)(x) = \sum_{j=1}^{d} \alpha_j \phi^j(x) \tag{5.4}$$

with $\alpha_j = u_0(x_j)$, where u_0 is the homogenized part of u_ϵ and x_j are the nodal points of K. Then in any element $K \in \mathbf{K}^h$, we have

$$\nabla_i a_\epsilon^{ij} \nabla_j (u_\epsilon - v_\epsilon^h) = f \quad in \quad K, \tag{5.5}$$

where v_ϵ^h on ∂K is a piecewise linear function whose values at the nodal points are $u_0(x_j)$.

Let's divide the solution of (5.5) into two parts: $(u_\epsilon - v_\epsilon^h) = (u_\epsilon - v_\epsilon^h)_1 + (u_\epsilon - v_\epsilon^h)_2$, where $(u_\epsilon - v_\epsilon^h)_1$ and $(u_\epsilon - v_\epsilon^h)_2$ satisfy

$$\begin{aligned} \nabla_i a_{ij}^\epsilon \nabla_j (u_\epsilon - v_\epsilon^h)_1 &= f \quad in \quad K, \\ (u_\epsilon - v_\epsilon^h)_1 &= 0 \quad on \quad \partial K; \end{aligned} \tag{5.6}$$

$$\begin{aligned} \nabla_i a_\epsilon^{ij} \nabla_j (u_\epsilon - v_\epsilon^h)_2 &= 0 \ \text{ in } \ K, \\ (u_\epsilon - v_\epsilon^h)_2 &= (u_\epsilon - v_\epsilon^h) \ \text{ on } \ \partial K; \end{aligned} \tag{5.7}$$

respectively. $(u_\epsilon - v_\epsilon^h)_1$ can be estimated from (5.6) and the Poincare inequality:

$$\|(u_\epsilon - v_\epsilon^h)_1\|_{H^1(K)} \leq Ch\|f\|_{L_2(K)}. \tag{5.8}$$

As for $(u_\epsilon - v_\epsilon^h)_2$, using the expansion of u_ϵ, $u_\epsilon = u_0 + \epsilon_2 \chi_\lambda^m \nabla_m u_0 + \theta^u$, and (5.7) we have

$$\begin{aligned}\|(u_\epsilon - v_\epsilon^h)_2\|_{H^1(K)} &\le \|u_\epsilon - v_\epsilon^h\|_{H^{1/2}(\partial K)} \\ &\le C\|u_0 + \epsilon\chi^m(\frac{x}{\epsilon})\nabla_m u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} + C\|\theta^u\|_{H^{1/2}(\partial K)} \\ &\le C\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} + C\|\epsilon\chi^m(\frac{x}{\epsilon})\nabla_m u_0\|_{H^{1/2}(\partial K)} + C\|\theta^u\|_{H^1(K)}.\end{aligned} \tag{5.9}$$

Combination of (5.8) and (5.9) gives

$$\begin{aligned}\|u_\epsilon - v_\epsilon^h\|_{H^1(K)} \le C_1 h\|f\|_{0,K} &+ C\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} \\ &+ C\|\epsilon\chi^m(\frac{x}{\epsilon})\nabla_m u_0\|_{H^{1/2}(\partial K)} + C\|\theta^u\|_{H^1(K)}.\end{aligned} \tag{5.10}$$

For the second term on the r.h.s. of (5.10) we can write

$$\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} = \inf_{\phi|_{\partial K} = u_0 - v_\epsilon^h} \|\phi\|_{H^1(K)}. \tag{5.11}$$

Since v_ϵ^h is the linear function on ∂K we can extend it as a linear function $\tilde{v}_\epsilon^h$ onto K. Noticing the fact that because of (5.4) u_0 and $\tilde{v}_\epsilon^h$ coincide at the nodal points we have

$$\|u_0 - v_\epsilon^h\|_{H^{1/2}(\partial K)} \le \|u_0 - \tilde{v}_\epsilon^h\|_{H^1(K)} \le Ch|u_0|_{2,K}. \tag{5.12}$$

For the third term on the r.h.s of (5.10) we have

$$\|\epsilon\chi^m\nabla_m u_0\|_{H^{1/2}(\partial K)} \le C\|\epsilon\chi^m\nabla_m u_0(1-\tau_\epsilon)\|_{1,K} \tag{5.13}$$

where τ_ϵ is a family of functions satisfying the following conditions:

1. $\tau_\epsilon \in C_0^\infty(K)$, $0 \le \tau_\epsilon \le 1$, $\tau_\epsilon = 1$ outside the ϵ neighborhood of K.
2. $\epsilon|\nabla\tau_\epsilon| \le C$ in K, where the constant C does not depend on ϵ. Such functions can be

constructed for any domain with Lipschitz boundary. Then

$$\begin{aligned}\|\epsilon\chi^m\nabla_m u_0\|_{H^{1/2}(\partial K)} &\le C\|\epsilon\chi^m\nabla_m u_0(1-\tau_\epsilon)\|_{0,K} + C\|\epsilon\nabla\chi^m\nabla_m u_0(1-\tau_\epsilon)\|_{0,K}\\ &+ C\|\epsilon\chi^m\nabla_m u_0\nabla\tau_\epsilon\|_{0,K} + C\|\epsilon\chi^m\nabla_m^2 u_0(1-\tau_\epsilon)\|_{0,K} \le C\epsilon|u_0|_{1,K}\\ &+ C\|\epsilon\nabla\chi^m\nabla_m u_0(1-\tau_\epsilon)\|_{0,K} + C\|\epsilon\chi^m\nabla_m u_0\nabla\tau_\epsilon\|_{0,K}\\ &+ C\|\epsilon\chi^m\nabla_m^2 u_0(1-\tau_\epsilon)\|_{0,K}.\end{aligned} \tag{5.14}$$

To estimate the r.h.s. of (5.14) we denote by S the support of $1-\tau_\epsilon$. Note that $meas(S) \le C\epsilon h$ and $\nabla\tau_\epsilon = 0$ in $K\setminus S$ (outside S). Then for the last two terms on the r.h.s. of this inequality using the fact that $|\chi| \le C$, and $|\epsilon\nabla\tau_\epsilon| \le C$ we have

$$\|\epsilon\chi^m\nabla_m u_0\nabla\tau_\epsilon\|_{0,K} \le C\|\nabla u_0\|_{0,S} \le C\sqrt{\epsilon h}, \tag{5.15}$$

$$\|\epsilon\chi^m\nabla_m^2 u_0(1-\tau_\epsilon)\|_{0,K} \le C\epsilon\|u_0\|_{2,K}. \tag{5.16}$$

For the second term on the r.h.s of (5.14) we have

$$\|\epsilon\nabla\chi^m\nabla_m u_0(1-\tau_\epsilon)\|_{0,K} \le C\|\epsilon\nabla\chi\|_{0,S}.$$

Here we used the fact that $|\nabla u_0| \le C$. To estimate $\|\epsilon\nabla\chi\|_{0,S}$ we cover S with the ϵ-periods Π_ϵ^i of χ such that

$$S \subset \bigcup_{i=1}^{N} \Pi_\epsilon^i$$

where $N \le Ch/\epsilon$. Then

$$\|\epsilon\nabla\chi\|_{0,S}^2 \le C\sum_{i=1}^{N}\|\epsilon\nabla\chi\|_{0,\Pi_\epsilon^i}^2. \tag{5.17}$$

To estimate $\|\epsilon\nabla\chi\|_{0,\Pi_\epsilon^i}$ we make the change of variables $y_i = x_i/\epsilon$, $(i=1,2)$, where y_i are

the new variables. Then

$$\|\epsilon\nabla\chi\|_{0,\Pi_\epsilon^i} = (\int_{\Pi_\epsilon^i} \epsilon^2(\nabla_x\chi)^2 dx)^{1/2} = (\int_{\Pi_1^i} \epsilon^2(\nabla_y\chi)^2 dy)^{1/2} = \epsilon\|\nabla\chi\|_{0,\Pi_1^i} \leq C\epsilon \tag{5.18}$$

where Π_1^i is the period of size 1. In the last step we have used the fact that $\|\nabla_y\chi\|_{0,\Pi_1^i} \leq C$. Furthermore, from (5.17) using (5.18) we have

$$\|\epsilon\nabla\chi\|_{0,S}^2 \leq CN\epsilon^2 \leq C\epsilon h \tag{5.19}$$

or

$$\|\epsilon\nabla\chi^m\nabla_m u_0(1-\tau_\epsilon)\|_{0,K} \leq C\sqrt{N}\epsilon \leq C\sqrt{\epsilon h} \tag{5.20}$$

(5.14) along with (5.15), (5.16), and (5.20) gives

$$\|\epsilon\chi^m\nabla_m u_0\|_{H^{1/2}(\partial K)} \leq C(\epsilon+h)\|u_0\|_{2,K} + C_1\sqrt{\epsilon h}. \tag{5.21}$$

Furthermore, taking into account (5.12) and (5.20) in (5.10), we have

$$\|u_\epsilon - v_\epsilon^h\|_{1,K} \leq C(h+\epsilon)\|u_0\|_{2,K} + C_1\sqrt{\epsilon h} + C_2\|\theta^u\|_{1,K}. \tag{5.22}$$

Summing (5.22) over all $K \in \mathbf{K}^h$ and noticing that $\|\theta^u\|_{1,\Omega}$ depends only on ϵ, and $\|u_0\|_{2,\Omega} \leq C\|f\|_{0,\Omega}$ along with Cea's lemma we get the desired result.

□

5.4 L_2 estimates

In this section, we derive the L_2 norm estimates for MsFEM. As we can see from the previous section, the use of Aubin-Ntsche trick becomes complicated because of the estimate (5.21).

Following [28], we use a discrete error analysis to overcome the difficulty. First, we show that the L_2-norm error can be determined from the discrete l^2 norm error. For the latter, we compare the discrete solution of (4.1) and that of the homogenized equation at the nodal points.

Specifically, let us denote by u_0^h the numerical solution of the homogenized equation

which is calculated using FEM. Then we have [28]

$$\begin{aligned} \|u_\epsilon - u_\epsilon^h\|_{L_2(\Omega)} &\leq \|u_\epsilon - u_0\|_{L_2(\Omega)} + \|u_0 - u_0^h\|_{L_2(\Omega)} + \|u_0^h - u_\epsilon^h\|_{l_2(\Omega)} \\ &\leq C\epsilon + C_1 h^2 + C_2 \|u_0^h - u_\epsilon^h\|_{l_2(\Omega)}. \end{aligned} \tag{5.23}$$

Here we used estimates for the convergence of homogenized solution to the exact and the traditional finite element estimate for regular elliptic equations. The estimate $\|u_\epsilon - u_0\|_{L_2(\Omega)} \leq C\epsilon$ can be easily derived, while $\|u_0 - u_0^h\|_{L_2(\Omega)} \leq Ch^2$ is a traditional finite element estimate.

5.4.1 Asymptotic expansion of the discrete solution

Denote by U_ϵ^h and U_0^h the nodal values of u_ϵ^h and u_0^h, respectively. The linear system of equations for U_ϵ^h is

$$A_\epsilon^h U_\epsilon^h = f_\epsilon^h, \tag{5.24}$$

where A_ϵ^h and f_ϵ^h are obtained from $a(u^h, v)$ and $f(v)$ by using $v = \phi_\epsilon^i$ for $i = 1, \ldots, N$. Similarly, for U_0^h one has

$$A_0^h U_0^h = f_0^h, \tag{5.25}$$

where A_0^h and f_0^h are obtained by applying $v = \phi_0^i$ $(i = 1, \ldots, N)$ to $a^*(u_0^h, v) = f(v)$ with

$$a^*(u_0^h, v) = \int_\Omega a_*^{ij} v_{,i} u_{0,j}^h dx.$$

Note that the "comma" notation is used here and below for partial derivatives. We write

$$A_\epsilon^h = A_0^h + \epsilon A_1^h, \qquad f_\epsilon^h = f_0^h + \epsilon f_1^h; \tag{5.26}$$

then formal calculation shows that U_ϵ^h can be expanded as

$$U_\epsilon^h = U_0^h + \sum_{i \geq 1} \epsilon^i U_i^h, \tag{5.27}$$

where U_i^h $(i \geq 1)$ are given by

$$A_0^h U_1^h = f_1^h - A_1^h U_0^h; \tag{5.28}$$

$$A_0^h U_i^h = -A_1^h U_{i-1}^h \qquad (i > 1). \tag{5.29}$$

We note that the expansion enables us to avoid dealing with the inverse of A_ϵ^h whose entries are oscillatory. In contrast, A_0^h and its inverse are well understood.

5.4.2 Derivations and estimates for A_1^h and f_1^h

The basis of (5.26) is the expansion of the multiscale base functions. We rewrite it here for convenience:

$$\phi_\epsilon^j = \phi_0^j + \epsilon \chi^p \nabla_p \phi_0^j + \theta_\epsilon^j. \tag{5.30}$$

Note that ϕ_0^j are linear functions and $|\nabla \phi_0^j| \leq C/h$. Moreover, $\theta_\epsilon^j = \eta^p \nabla_p \phi_0^j$, where η^p is the solution of (3.16).

Now (5.26) can be derived. We note that the global stiffness matrix A_ϵ^h is assembled from the local stiffness matrices defined on $K \in \mathbf{K}^h$

$$A_{kl}^e = \int_K a_\epsilon^{ij} \phi_{\epsilon,i}^k, \phi_{\epsilon,j}^l dx,$$

over several neighboring elements, and f_ϵ^h is assembled from

$$f_k^e = \int_K f \phi_\epsilon^k dx.$$

Similarly, A_0^h and f_0^h are assembled from

$$A_{0kl}^e = \int_K a_*^{ij} \phi_{0,i}^k \phi_{0,j}^l dx, \qquad f_{0k}^e = \int_K f \phi_0^k dx,$$

respectively. Below, we consider the expansions of A_{kl}^e and f_k^e which lead to (5.26).

First, we review some relevant results proved in [28]. Letting σ^{ij} be

$$\sigma^{ij} = a^{ij}(\delta_{jk} + \nabla_j^y \chi^k).$$

we have $\langle \sigma^{ij} \rangle = a_*^{ij}$ and $\sigma^{ij}_{,y_i} = 0$. Using integration by parts, we further obtain $\langle \sigma^{pj} \chi^i_{,y_p} \rangle = 0$. Define $\tilde{\sigma}^{ij} = \sigma^{ij} - a_*^{ij} - \sigma^{pj}\chi^i_{,y_p}$, then $< \tilde{\sigma}^{ij} >= 0$.

Now following [28], from the definition of A^e_{kl} and f^e_k, we can derive (5.26) with A^h_1 given by

$$A^e_{1_{kl}} = \int_K a^{ij}_\epsilon \nabla_i(\phi^k_0 + \epsilon\chi^p \nabla_p \phi^k_0 + \epsilon\theta^k)\nabla_j(\phi^l_0 + \epsilon\chi^p\nabla_p\phi^l_0 + \epsilon\theta^l)dx - \int_K a^{ij}_* \nabla_i\phi^k_0 \nabla_j \phi^l_0 dx. \tag{5.31}$$

After some algebra we have

$$\begin{aligned} A^e_{1_{kl}} &= -\int_K \sigma^{ij}(\phi^k_{0,j}\theta^l_{,i} + \phi^l_{0,j}\theta^k_{,i})dx + \int_K \epsilon a^{ij}_\epsilon \theta^k_{,i}\theta^l_{,j}dx + \frac{1}{\epsilon}\int_K \tilde{\sigma}^{ij}\phi^l_{0,j}\phi^k_{0,i}dx \\ &\equiv \Lambda^e_{1_{kl}} + \Lambda^e_{2_{kl}} + \Lambda^e_{3_{kl}}, \end{aligned} \tag{5.32}$$

and for f^h_1

$$f^e_{1_i} = -\int_K f(\chi^p \phi^i_{0,p} + \theta^i)dx \equiv F^e_{1_i} + F^e_{2_i}. \tag{5.33}$$

Note that Λ^e_i $(i = 1, 2, 3)$ are the ith integrals in (5.32) and F^e_1 and F^e_2 are the first and second integrals in (5.33).

To estimate entries in A^h_1 and f^h_1, it is sufficient to estimate $A^e_{1_{kl}}$ and $f^e_{1_i}$. Let Λ^h_i be the matrix whose entries are assembled from Λ^e_i $(i = 1, 2)$, such that

$$A^h_1 = \Lambda^h_1 + \Lambda^h_2 + \Lambda^h_3.$$

Similarly we define vectors F^h_1 and F^h_2. We will use the estimates of $\|\Lambda^h_i\|$ and $|F^h_i|$, where $\|\cdot\|$ is the standard max norm of matrices in $\mathbf{R}^{N\times N}$ and $|\cdot|$ denotes the max norm of vectors in $\mathbf{R}^N$. Note that $N \sim 1/h^2$ is the number of nodal points.

We estimate the elements of A^h_1. For $\Lambda^e_{1_{kl}}$ using the integration by parts we have

$$\begin{aligned} |\Lambda^e_{1_{kl}}| &= |\int_K \sigma^{ij}(\phi^k_{0,j}\theta^l_{,i} + \phi^l_{0,j}\theta^k_{,i})dx| = |\int_{\partial K} \sigma^{ij}(\phi^k_{0,j}\theta^l + \phi^l_{0,j}\theta^k)ds| \\ &\le \frac{C}{h^2}\int_{\partial K} |\sigma^{ij}|ds + \frac{C}{h^2}\int_{\partial K} |\sigma^{ij}||\epsilon\nabla\chi|ds \end{aligned} \tag{5.34}$$

Here we have used the fact that $|\theta^l| \le C$ on ∂K and it is continuous. The first term on

the r.h.s of (5.34) is not necessarily defined if $a^{ij}(y) \in L^\infty(Y)$. For that reason, we assume that the intersection of ∂K and the interfaces of the discontinuity of the coefficients are countable number of points. In general this occurs when ∂K does not have a common part with the discontinuity interfaces. In this case, $a^{ij}(y)$ becomes piecewise smooth function on ∂K. Then (5.34) can be estimated as

$$|\Lambda^e_{1_{kl}}| \leq \frac{C}{h^2} \int_{\partial K} \epsilon |\nabla \chi| ds. \tag{5.35}$$

Moreover, we assume that the integral on the r.h.s. of (5.35) exists. This is for example the case for the checker-board models where ∂K does not coincide with the discontinuity interfaces.

To estimate $\int_{\partial K} \epsilon |\nabla \chi| ds$, we proceed as we did (5.21). Covering ∂K with the ϵ-periods of χ, Π^i_ϵ, we have

$$\partial K \subset \sum_{i=1}^{M} \partial \Pi^i_\epsilon,$$

where $\partial \Pi^i_\epsilon = \partial K \bigcap \Pi^i_\epsilon$. We note that $M \leq Ch/\epsilon$. Then

$$\int_{\partial K} \epsilon |\nabla \chi| ds \leq \sum_{i=1}^{M} \int_{\partial \Pi^i_\epsilon} \epsilon |\nabla \chi| ds. \tag{5.36}$$

Making change of variables $y = x/\epsilon$, where y is the new variables, for each $\int_{\partial \Pi^i_\epsilon} \epsilon |\nabla \chi| ds$, we have

$$\int_{\partial \Pi^i_\epsilon} \epsilon |\nabla \chi| ds_x = \epsilon \int_{\partial \Pi^i_1} |\nabla_y \chi| ds_y$$

where Π^i_1 is the period of size 1. Since the singularity of χ is of type r^α $(\alpha > 0)$, we get

$$\int_{\partial \Pi^i_1} |\nabla_y \chi| ds_y \leq C.$$

Then (5.36) becomes

$$\int_{\partial K} \epsilon |\nabla \chi| ds \leq CM\epsilon \leq Ch.$$

Consequently,

$$|\Lambda^e_{1_{kl}}| \leq \frac{C}{h}. \tag{5.37}$$

To estimate $|\Lambda^e_{2_{kl}}|$, we need to estimate $\|\nabla\theta^l\|_{0,K}$. From the equation for θ we have

$$\|\nabla\theta^l\|_{0,K} \leq \frac{C}{h}\|\chi\|_{1/2,\partial K} \leq \frac{C}{h}\|\chi(1-\tau_\epsilon)\|_{1,K}.$$

where τ_ϵ is a family of functions introduced in the previous section. The estimate for $\|\chi(1-\tau_\epsilon)\|_{1,K}$ can be obtained similar to (5.13). Then

$$\|\nabla\theta^l\|_{0,K} \leq \frac{C}{\sqrt{\epsilon h}}. \tag{5.38}$$

Here we used the assumption about the smoothness of u_0, $|u_0|_{C^2(\Omega)} \leq C$. Having the estimate (5.38) we can estimate $|\Lambda^e_{2_{kl}}|$ as following:

$$|\Lambda^e_{2_{kl}}| \leq C\epsilon\|\nabla\theta\|^2_{0,K} \leq \frac{C}{h}. \tag{5.39}$$

As for $\Lambda^e_{3_{kl}}$

$$|\Lambda^e_{3_{kl}}| = |\frac{1}{\epsilon}\int_K \tilde{\sigma}^{ij}\nabla_i\phi_0^k\nabla_j\phi_0^l dx| = |\frac{1}{\epsilon h^2}\int_K \tilde{\sigma}^{ij}\xi_{ij}^{kl}dx| \tag{5.40}$$

where ξ_{ij}^{kl} are constants, $\xi_{ij}^{kl} = h^2\nabla_i\phi_0^k\nabla_j\phi_0^l$, which are independent of h and ϵ, and $|\xi_{ij}^{kl}| \leq C$. Since $\tilde{\sigma}^{ij}\xi_{ij}^{kl}$ is a periodic function with the average zero, we can discard the periods of $\tilde{\sigma}^{ij}\xi_{ij}^{kl}$ from K without changing the value of the integral (5.40). Denoting the ϵ-periods of $\tilde{\sigma}^{ij}\xi_{ij}^{kl}$ by Π^i_ϵ, such that $\bigcup_{i=1}^L \Pi^i_\epsilon \subset K \subset \bigcup_{i=1}^{L+1}\Pi^i_\epsilon$ we have

$$\begin{aligned}|\Lambda^e_{3_{kl}}| &\leq |\frac{1}{\epsilon h^2}\sum_{i=1}^{L}\int_{\Pi^i_\epsilon}\tilde{\sigma}^{ij}\xi_{ij}^{kl}dx| + |\frac{1}{\epsilon h^2}\int_{K\setminus \mathrm{T}_{i=1}^{L}\Pi^i_\epsilon}\tilde{\sigma}^{ij}\xi_{ij}^{kl}dx| \\ &= |\frac{1}{\epsilon h^2}\int_{K\setminus \mathrm{T}_{i=1}^{L}\Pi^i_\epsilon}\tilde{\sigma}^{ij}\xi_{ij}^{kl}dx| \leq \frac{C}{\epsilon h^2}\int_S |\epsilon^2(\nabla\chi)^2|dx\end{aligned} \tag{5.41}$$

where $S = K\setminus\bigcap_{i=1}^L \Pi^i_\epsilon$. Here we have used the fact that $\tilde{\sigma}^{ij} = a^{ij} + a^{ik}\nabla_k\chi^j - a^{pj}\nabla_p\chi^i -$

$a^{pk}\nabla_p\chi^i\nabla_k\chi^j - a_*^{ij}$. The r.h.s of (5.41) has been estimated in (5.19) as

$$\int_S |\epsilon^2(\nabla\chi)^2|dx \leq C\epsilon h.$$

Consequently,

$$|\Lambda^e_{3_{kl}}| \leq \frac{C}{h}. \tag{5.42}$$

It is straightforward to check that

$$|F^e_{1_i}| \leq Ch, \quad |F^e_{2_i}| \leq Ch. \tag{5.43}$$

5.4.3 The error estimate

We consider U_1^h in some detail; the rest of U_i^h can be studied similarly. In [28], it was found that the conservative structure of the linear system (5.24) is the key to obtain discrete error cancellation using "summation by parts." In the following, we present a simpler derivation.

First, it is easy to see

$$\sum_{k=1}^{3}\phi_\epsilon^k = 1, \qquad \sum_{k=1}^{3}\phi_0^k = 1,$$

for all $K \in \mathbf{K}^h$. It follows that [28]

$$\sum_{j=1}^{N} A_1^{ij} = 0,$$

because $\nabla_i a_\epsilon^{ij}\nabla_j\phi_\epsilon^k = 0$, $\nabla_i a_0^{ij}\nabla_j\phi_0^k = 0$ for all k and $\sum_{k=1}^3\phi_\epsilon^k = 1$, $\sum_{k=1}^3\phi_0^k = 1$ on ∂K. Now, according to (5.28), we have

$$U_1^h = G_0^h f_1^h - G_0^h A_1^h U_0^h, \tag{5.44}$$

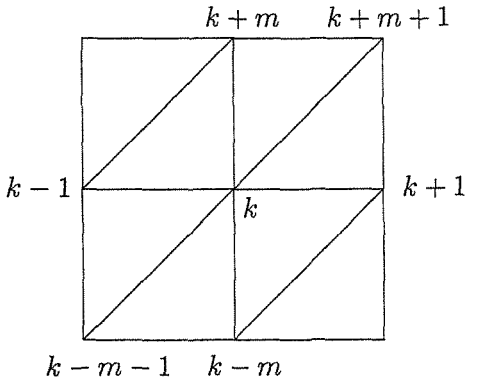

Figure 5.1: An example of neighboring nodes in a triangulation of a rectangular mesh with $m \times n$ nodal points. $O_k = \{k-m-1, k-m, k-1, k+1, k+m, k+m+1\}$.

where $G_0^h = (A_0^h)^{-1}$. For $G_0^h A_1^h U_0^h$ we note that

$$\sum_{j=1}^{N} A_1^{ij} U_0^j = \sum_{j=1}^{N} A_1^{ij} U_0^j - \sum_{j=1}^{N} A_1^{ij} U_0^i = \sum_{j=1,\ j\neq i}^{N} A_1^{ij}(U_0^j - U_0^i).$$

Furthermore by the symmetry of A_1^{ij} we can combine $A_1^{kj}(U_0^j - U_0^k)$ and $A_1^{jk}(U_0^k - U_0^j)$ in the sum $G_0^{ik} A_1^{kj} U_0^j$ and get

$$\sum_{k=1}^{N} G_0^{ik} \sum_{j=1,\ j\neq k}^{N} A_1^{kj}(U_0^j - U_0^k) = \sum_{k=1}^{N} \sum_{j=1}^{j<k} A_1^{kj}(U_0^j - U_0^k)(G_0^{ik} - G_0^{ij}). \tag{5.45}$$

We note that A_1^h is sparse, which is important in obtaining the error estimates below. The following definitions are helpful to describe the sparsity. We say two nodal points are neighbors if they are the vertices of the same element. For a nodal point with index k, we denote the set of the indices of its neighbors by O_k (see Fig. 5.1). We have $A_1^{ij} = 0$ if $j \notin O_i$. Thus, the r.h.s. of (5.45) can be written as

$$\sum_{k=1}^{N} \sum_{j\in O_k, j<k} A_1^{kj}(U_0^j - U_0^k)(G_0^{ik} - G_0^{ij}).$$

Since the summation of j depends only on the geometry of the mesh and independent of N (or h), we can safely ignore it when estimating the above expression. This would greatly

simplify the presentation. Introduce a difference operator D such that for any $V \in \mathbf{R}^N$, $(DV)_k \equiv V^k - V^j$ for some $j \in O_k$. Note that j is not specified in the definition. For a matrix, D applies to its row vectors unless otherwise stated. With this definition, from (5.28) and (5.45) we get

$$U_1^h = G_0^h f_1^h - DG_0^h A_1^h DU_0^h.$$

Here, the important features to notice are the differences on G_0^h and U_0^h; the details of these differences, however, are not important and are hidden in D.

Since a_*^{ij} are constants, the discrete Green's function G_0^h and the solution U_0^h are well understood (cf. [28, 23, 48]). We have

$$\begin{aligned} |DU_0^h| \le Ch, \qquad |D^2 U_0^h| \le Ch^2; \\ \|G_0^h\| \le C/h^2, \quad \|DG_0^h\| \le C/h, \quad \|D^2 G_0^h\| \le C \ln|h|. \end{aligned} \tag{5.46}$$

Then

$$|U_1^h| \le \|G_0^h\||f_1^h| + \|DG_0^h\||A_1^h||DU_0^h| \le \frac{C}{h}. \tag{5.47}$$

From this we can deduce that the first order correction in (5.27) is of order ϵ/h. Similarly we can estimate $|U_i^h|$, $(i \ge 2)$. For the estimate of U_2^h we estimate $|DU_1^h|$.

$$|DU_1^h| = |DG_0^h f_1^h - D^2 G_0^h A_1^h DU_0^h| \le \|DG_0^h\||f_1^h| + \|D^2 G_0^h\||A_1^h||DU_0^h| \le C|\log(h)|.$$

Then

$$|U_2^h| \le |DG_0^h A_1^h DU_1^h| \le \|DG_0^h\||A_1^h||DU_1^h| \le \frac{C|\log(h)|}{h^2}.$$

Consequently, the second order correction in (5.27) is of order $\epsilon^2|\log(h)|/h^2$. Let's show by induction that

$$|U_i^h| \le \frac{C}{h^i}|\log(h)|^{i-1}, \quad |DU_i^h| \le \frac{C}{h^{i-1}}|\log(h)|^{i-1}. \tag{5.48}$$

Assuming (5.48) holds for $i = k-1$, we can show that it holds for $i = k$.

$$|U_k^h| \leq \|DG_0^h\| \|A_1^h\| |DU_{k-1}^h| \leq \frac{C}{h^2} |DU_{k-1}^h| \leq \frac{C}{h^k} |\log(h)|^{k-1}$$

and

$$|DU_k^h| \leq \|D^2 G_0^h\| \|A_1^h\| |DU_{k-1}^h| \leq \frac{C}{h} |\log(h)| |DU_{k-1}^h| \leq \frac{C}{h^{k-1}} |\log(h)|^{k-1}.$$

Using these estimates for U_k^h in (5.27), we have

$$|U_\epsilon^h - U_0^h| \leq C\frac{\epsilon}{h}\left(\sum_{i\geq 0}\left(\frac{\epsilon}{h}|\log(h)|\right)^i\right) = C\frac{\epsilon}{h}\frac{1}{1-\epsilon|\log(h)|/h} \leq CC_1\frac{\epsilon}{h} \tag{5.49}$$

if $\epsilon|\log(h)|/h \leq 1 - 1/C_1$.

5.5 Numerical experiments

In this section we study the convergence and the accuracy of the multiscale method through numerical experiments. The model problem is solved by using the multiscale method with the base functions defined by (4.3) and the linear boundary conditions.

The calculation of the multiscale base functions involves solving PDE with discontinuous coefficients. The traditional FEM does not provide a robust approximation to the base functions. For that reason, there have been developed different numerical methods which robustly approximate the solutions of elliptic PDE with discontinuous coefficients. But because of the change of the mesh size, the elements may include several discontinuities of the coefficients. This causes difficulties in numerical codes and decreases their efficiency.

For the calculation of the base functions, we use finite element discretization aligned with the discontinuity interfaces of the coefficients with Black-Box multigrid. It is known that this method is robust for not large inhomogenities, i.e., for not large β/α (see (3.2)). We test problems with not large inhomogenities also. The numerical results confirm our theoretical rate.

For the numerical tests, we choose the checker-board model where the coefficients $a^{ij}(y)$,

Table 5.1: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ for $\epsilon = 1/32$, $\delta = 0.1$, aligned

h	l_2	rate
1/8	0.98e-2	
1/16	0.18e-1	-0.87
1/32	0.34e-1	-0.92

Table 5.2: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ for $\epsilon = 1/64$, $\delta = 0.1$, aligned

h	l_2	rate
1/16	0.11e-1	
1/32	0.19e-1	-0.79
1/64	0.34e-1	-0.84

$y \in [-1/2, 1/2]^2$ defined as

$$a^{ij}(y) = \begin{cases} 1, & \text{if } y_1 y_2 > 0 \\ \delta & \text{if } y_1 y_2 < 0. \end{cases} \tag{5.50}$$

As examples, we consider the homogeneous equation with the r.h.s zero, $f = 0$, and inhomogeneous equation with the r.h.s -1, $f = -1$. Also we test separately the cases when multiscale elements are aligned with the discontinuity interfaces of the coefficients and when they are not. The numerical tests show that the error estimate (5.49) holds for the case when multiscale elements are aligned with the discontinuity interfaces of the coefficients.

Example 1. In this example, we consider the case when a^{ij} is defined by (5.50), with $\delta = 0.1$, and $f = -1$, $u^{boundary} = 0$.

In the tabels 5.1 and 5.2 the numerical tests are done when the multiscale elements are aligned with the discontinuity interfaces of the coefficients. In particular, $a^{ij}(x/\epsilon)$ is defined as

$$a^{ij}(y) = \begin{cases} 1, & \text{if } \sin(\frac{2\pi x_1}{\epsilon}) \sin(\frac{2\pi x_2}{\epsilon}) > 0 \\ \delta & \text{if } \sin(\frac{2\pi x_1}{\epsilon}) \sin(\frac{2\pi x_2}{\epsilon}) < 0. \end{cases}$$

As we see from these tables, the error estimate is ϵ/h as predicted. Indeed the error increases with -1 order as we decrease h, and it resonates as the ratio ϵ/h is fixed.

In tables 5.3 and 5.4, we repeat the tests shown in the tables 5.1 and 5.2, for $\delta = 0.49$.

In the tables 5.5 and 5.6 we tested the case when the multiscale elements are not aligned

Table 5.3: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ for $\epsilon = 1/32$, $\delta = 0.49$, aligned

h	l_2	rate
1/8	0.31e-3	
1/16	0.63e-3	-1.02
1/32	0.16e-2	-1.3

Table 5.4: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ for $\epsilon = 1/64$, $\delta = 0.49$, aligned

h	l_2	rate
1/16	0.24e-3	
1/32	0.78e-3	-1.7
1/64	0.16e-2	-1.03

with the discontinuity interfaces of the coefficients. In this case $a^{ij}(x/\epsilon)$ is defined as

$$a^{ij}(y) = \begin{cases} 1, & \text{if } \sin(\frac{2\pi x_1}{\epsilon} + \frac{\pi}{2}) \sin(\frac{2\pi x_2}{\epsilon} + \frac{\pi}{2}) > 0 \\ \delta & \text{if } \sin(\frac{2\pi x_1}{\epsilon} + \frac{\pi}{2}) \sin(\frac{2\pi x_2}{\epsilon} + \frac{\pi}{2}) < 0. \end{cases}$$

These tables confirm MsFEM convergence rate derived in the previous section.

Table 5.5: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ for $\epsilon = 1/32$, $\delta = 0.1$, not aligned

h	l_2	rate
1/8	0.78e-2	
1/16	0.12e-1	-0.72
1/32	0.21e-1	-0.80

Table 5.6: $\|u_\epsilon - u_\epsilon^h\|_{l_2}$ for $\epsilon = 1/64$, $\delta = 0.1$, not aligned

h	l_2	rate
1/16	0.65e-2	
1/32	0.12e-1	-0.88
1/64	0.21e-1	-0.80

Chapter 6 MsFEM for problems with weakly dependent random coefficients

6.1 Discussion

The main difference in the analysis of MsFEM for elliptic problems with random and periodic coefficients is the behavior of the solution of cell problem (3.11). Unlike the periodic case in general (almost periodic or stationary ergodic cases) the "cell problem" [8, 32]

$$\nabla_i a^{ij} \nabla_j \chi^k = -\nabla_i a^{ik}, \ y \in R^d$$

does not have almost periodic or stationary solutions even when the r.h.s. has mean zero. It is nevertheless possible to construct a certain solution of this problem which possesses special properties [32]. It turns out that this special solution is not bounded [32]. The boundedness of the solution of cell problem plays an important role in the convergence estimate. In general random cases, it is known that a solution of the "cell problem" has a sublinear growth. This growth rate affects the resonance error and the resulting resonance error is $(\epsilon/h)^{1-a}$, $0 \leq a \leq 1$, where a depends on the growth rate of the solution of cell problem. If the solution of the cell problem is bounded as in the periodic case then it can be shown that the convergence rate is ϵ/h. Let's note that for almost periodic coefficients a_ϵ^{ij} satisfying Kozlov's diophantine condition ([34]), the solution of the cell problem is bounded. For example, the finite trigonometric polynomials satisfy Kozlov's diophantine condition. For general random cases the behavior of the solution of cell problem has not been extensively studied. In this work we use some estimates derived for the almost solutions of the cell problem [56] to study the resonance error of MsFEM. We note that the analysis of MsFEM has been limited to periodic structures so far.

6.2 Summary

Untill now the mathematical analysis of MsFEM has been limited by the case of periodic structures. We analyze MsFEM for problems with random coefficients whose correlation

decays at least by power law.

In other words, denoting $\Phi(A)$ to be the σ-algebra generated by the parameters of the problem in the physical domain A, we assume that

$$|E\xi\eta - E\xi E\eta| \leq \kappa(q)(E\xi^2)^{1/2}(E\eta^2)^{1/2}$$

where ξ is $\Phi(A)$ measurable, η is $\Phi(B)$ measurable and q is the distance between the domains A and B, i.e, $q = \inf\{\|x-y\|;\ x \in A,\ y \in B\}$. For the analysis of this paper we assume that

$$\kappa(q) \leq \frac{C}{q^A}. \tag{6.1}$$

The exponential decay of $\kappa(q)$ is often used in geostatistical models. Basic variogram models, exponential, Gaussian, Nugget [31] are the examples. But there are still basic variogram models (e.g., spherical model) which have power decay of $\kappa(q)$. This indicates the importance of including the power decay case in our analysis.

We derive the following estimate for MsFEM error:

$$E\|u-u^h\|^2_{1,D} \leq C\epsilon^\alpha + C_1(\frac{\epsilon}{h})^{\frac{Ad-2A}{5A+4d+2Ad}} \tag{6.2}$$

where d is the dimensionality of the physical space, A is the power in (6.1) and α is a positive constant. We see that this estimate in the case of $d = 2$ becomes of order 1. In $\mathbf{R}^3$, MsFEM convergence rate becomes

$$E\|u-u^h\|^2_{1,D} \leq C\epsilon^\alpha + C_1(\frac{\epsilon}{h})^{\frac{A}{11A+12}}.$$

Moreover, this estimate indicates that as A gets smaller the convergence rate of MsFEM deteriorates. In the case very large A the convergence rate of MsFEM approaches

$$E\|u-u^h\|^2_{1,D} \leq C\epsilon^\alpha + C_1(\frac{\epsilon}{h})^{1/11}.$$

The presence of the ratio $\frac{\epsilon}{h}$ in the error estimates of MsFEM has been observed eariler and it is called the resonance. This error is typical in upscaling problems.

The rest of the section is organized as follows. The formulation of the model problem

and MsFEM are presented in Section 2. In Section 3 the multiscale expansion of the solution is discussed. The main inequalities used in the article are in Section 4. Section 5 contains the derivation of H^1 estimates of MsFEM.

6.3 Formulations

Let (Ω, F, P) be a probability space and $a^{ij}(y,\omega)$, $y \in \mathbf{R}^3$, $\omega \in \Omega$ be a homogeneous ergodic random field.

We recall the concept of a homogeneous field. We assume that a dynamical system with d-dimensional time is given on Ω which satisfies: 1) the group property: $T(0) = I$ (I is the identity mapping), and $T(x+y) = T(x)T(y)$, $\forall x, y \in \mathbf{R}^d$; 2) the mapping $T(x) : \Omega \to \Omega$ preserve the measure μ on Ω, i.e., for every $x \in \mathbf{R}^d$, and every μ-measurable set $F \in \Omega$, we have $T(x)F$ is measurable, $\mu(T(x)F) = \mu(F)$; 3) for any measurable function $f(\omega)$ on Ω, the function $f(T(x)\omega)$ defined on $\Omega \times \mathbf{R}^d$ is also measurable. All functions on Ω are assumed to be F measurable; equality between random functions means equality almost everywhere (a.e.) with respect to P, and, as a rule, the notation a.e. is omitted.

A function $f(\omega)$ is called invariant if $f(T(x)\omega) = f(\omega)$ for any fixed $x \in \mathbf{R}^d$. The dynamical system is called ergodic if the set of invariant functions is exhausted by functions which are identically constant. A homogeneous random field is a function $f : \mathbf{R}^d \times \Omega \to \mathbf{R}^d$ for which

$$f(x+y,\omega) = f(x, T(y)\omega).$$

According to Birkhoff's theorem, the spatial average defined as

$$< f(x,\omega) >= \lim_{m\to\infty} \frac{1}{(2m)^d} \int_{|x|\le m} f(x,\omega)dx$$

exists for a homogeneous field, and $< f >$ is an invariant function. We observe that in the ergodic case there is the equality $< f >= Ef$; where E denotes the operator of the averaging with respect to the measure P (mathematical expectation).

Consider the following elliptic problem

$$\begin{aligned} \nabla_i a^{ij}(x/\epsilon,\omega)\nabla_j u_\epsilon = f \ \text{ in } \ D \\ u_\epsilon = 0 \ \text{ on } \ D \end{aligned} \tag{6.3}$$

where $a^{ij}(x/\epsilon,\omega)$ is a homogeneous ergodic random field, $f \in L_2(D)$ is a non-random function and ϵ is a small parameter. Furthermore assume that $a^{ij} = a^{ji}$ $(i,j = 1,2,3)$ have smooth realizations, uniformly bounded in the entire space along with their derivatives and

$$\alpha|\xi|^2 \le a^{ij}(y,\omega)\xi_i\xi_j \le \beta|\xi|^2 \tag{6.4}$$

for any $\xi \in \mathbf{R}^3$, with probablity 1. Here α and β are positive non-random numbers.

Variational problem of (6.4) is to seek $u \in H_0^1(D)$ s.t.

$$a(u,v) = f(v), \quad \forall v \in H_0^1(D), \tag{6.5}$$

where

$$a(u,v) = \int_D a^{ij}\frac{\partial u}{\partial x_i}(\frac{x}{\epsilon},\omega)\frac{\partial v}{\partial x_j}dx, \quad f(v) = \int_D fvdx. \tag{6.6}$$

It is easy to see that the bilinear form $a(\cdot,\cdot)$ is elliptic and continuous for almost each $\omega \in \Omega$.

As in previous sections, for $0 < h \le 1$, let $\mathbf{K}^h$ be a partition of D of regular triangles K with diameter less h. In each element $K \in \mathbf{K}^h$, we define a set of nodal basis $\{\phi_K^i\}$, $i = 1,\ldots,d$, with $d(=3)$ being the number of nodes of the element, satisfying

$$\nabla_k a^{kl}(\frac{x}{\epsilon},\omega)\nabla_l\phi_K^i = 0 \quad \text{in} \ \ K = \mathbf{K}^h. \tag{6.7}$$

We will neglect the subscript K when working in one element. Let $x_j \in K$ $(j = 1,\ldots,d)$ be the nodal points of K. As usual we require $\phi^i(x_j) = \delta_{ij}$ and that the base functions are linear on each side of on the boundary. So we have:

$$V^h = \text{span}\{\phi_K^i; \ \ i = 1,\ldots,d, \ \ K \subset \mathbf{K}^h\} \subset H_0^1(\Omega).$$

In the following we study the approximate solution of (6.5) in V^h, i.e., $u^h \in V^h$ such that

$$a(u^h, v) = f(v), \quad \forall v \in V^h. \tag{6.8}$$

6.4 Homogenization results

In this section we review the homogenization theory for equation(6.4) and give the estimate for the first order correctors. The behavior of the random field $u_\epsilon(x)$ as $\epsilon \to 0$ has been investigated under various assumptions on $a^{ij}(y,\omega)$. When $a^{ij}(y,\omega)$ is strictly stationary and ergodic it was found that there exists a^{ij}_* $(i,j = 1,2,3)$ such that if $u_0(x)$ is the solution of the deterministic Dirichlet problem

$$\begin{aligned} a^{ij}_* \nabla_i \nabla_j u_0 &= f \ \text{ in } \ D \\ u_0 &= 0 \ \text{ on } \ D \end{aligned} \tag{6.9}$$

then

$$\int_D E|u_\epsilon - u_0|^2 dx \to 0 \ \text{ as} \to 0 \tag{6.10}$$

where E denotes the expectation. The matrix a^{ij}_* is called the effective coefficient and can be calculated for some special cases.

Remark 6.4.1. For our analysis we assume that $u_0 \in C^2(D)$. This holds for smooth domains, as well as for convex polygons under some compatibility conditions [4].

When $a^{ij}(y)$ is a periodic function the solution of (6.3) can be approximated [32] as

$$u_\epsilon(x) = u_0 + \epsilon \chi^i(y) \nabla_i u_0(x) + O(\sqrt{\epsilon}) \tag{6.11}$$

where $y = x/\epsilon$ and $\chi^i(y)$ is the periodic solution of

$$\frac{\partial}{\partial y_i} a^{ij}(y) \frac{\partial}{\partial y_j} \chi^k = -\frac{\partial}{\partial y_j} a^{kj} \tag{6.12}$$

in a unit cell. The periodic solution of (6.12) exists and is unique up to a constant.

However, extension of the result on periodic structures to random homogeneous structures encounters a difficulty related to the insolvability of (6.12) in terms of homogeneous

functions. Indeed, the problem (6.12) is not Fredholm and has no solution in general. It is nevertheless possible to construct and analyze the solutions of approximate equations which is basic for our analysis. In this chapter following [56] we introduce almost solutions (in the sense of [58]) of (6.12) (for large $T > 0$)

$$-\nabla_i a^{ij}(y,\omega)\nabla_j \phi_T^k + \frac{\phi_T^k}{T} = D_i g_i \tag{6.13}$$

where $g_i = a^{ki}$. This equation has also played a central role in the homogenization of second order random elliptic equations in [35, 43]. It is known that the solution of this equation exists and it is a homogeneous random field (see [35]).

For the calculations of the convergence rate of MsFEM we require an assumption about the weak dependence of values of the field of coefficients of (6.13), $k = (a^{ij}, g_i)$ at remote points. Let $\Phi(A)$ be the σ-algebra generated by random vectors $k(x)$, $x \in A$. Then we assume that the family of σ-algebras ($\Phi(A)$, $A \in \mathbf{R}^3$) satisfy the condition of uniformly mixing

$$|E\xi\eta - E\xi E\eta| \leq \kappa(q)(E\xi^2)^{1/2}(E\eta^2)^{1/2} \tag{6.14}$$

where the random quantity ξ is $\Phi(A)$ measurable and η is $\Phi(B)$ measurable, and $q = \inf\{||x-y||,\ x \in A,\ y \in B\}$. For our analysis we assume that

$$\kappa(q) \leq \frac{C}{q^A}. \tag{6.15}$$

We note that (6.14) and (6.15) imply strong mixing for the coefficients a^{ij}.

Under the assumption of power decay of $\kappa(q)$, it can be shown [56] that u_ϵ can be expanded as

$$u_\epsilon = u_0 + \epsilon \phi_T^k \nabla_k u_0 + \theta \tag{6.16}$$

where ϕ_T is the solution of (6.13), u_0 is the solution of Dirichlet problem with constant coefficients, and θ is bounded

$$E\|\theta\|_{1,D}^2 \leq C\epsilon^\alpha \tag{6.17}$$

for some positive α. This estimate is obtained using nontrivial estimates for the correlations of the field ϕ_T and choosing the large T, $T = \epsilon^{-\beta}$ ($\beta > 0$). In [56] the author refine the claim of [35] about the existence of the limit ψ_{jk} of the field of derivatives of $\nabla_j \phi_T^k$ as $T \to \infty$. In particular, he shows

$$\langle |\psi - \nabla \phi_T|^2 \rangle \leq CT^{-\eta} \tag{6.18}$$

for some positive η. The effective coefficients a_*^{ij} in (6.9) are defined as

$$a_*^{ij} = \langle a^{ik} \psi_{jk} \rangle.$$

6.5 Estimates for the covariance of ϕ_T

Here we formulate the main theorem and the inequalities used in the derivation of the convergence rate of MsFEM. For further convinience, in the estimates for ϕ_T^k we omit the index k.

Theorem 6.5.1 *If (6.15) is satisfied then*

$$E[\phi_T; m]^2 \leq CT(\kappa(q) + (\frac{Tq^d}{m^d})^{1/2} \log(T) + \exp(-C(\log(T))^2) \tag{6.19}$$

where $[f; m] = 1/(2m)^d \int_{|x| \leq m} f(x)dx$, *and* m *and* q *are arbitrary large numbers* , $m \geq q+7$, $q \geq 5$, $T \geq 1$, *and* d *is the dimensionality of the physical space.*

The proof of this theorem is based on the analysis of the paths of the diffusion processes corresponding to (6.13) and it can be found in [56]. In the proof the author uses the following representation for ϕ_T

$$\phi_T = M_x \int_0^\infty \exp(-t/T) \nabla_j g_j(\eta(t)) dt$$

where M_x is the distribution of the diffusion process η with generating operator $1/2 \nabla_i (a^{ij}(y, \omega) \nabla_j)$. The parameter q can be choosen such that r.h.s of (6.19) achieves its minimum. With this choice of q, $q = (m^d/T)^{1/d+2A}$ (6.19) becomes

$$E[\phi_T; m]^2 \leq CT((\frac{T}{m^d})^{A/d+2A} \log(T) + \exp(-C(\log(T))^2). \tag{6.20}$$

In [56] the author uses the parameter q to define neighborhoods of the trajectories of a diffusion process corresponding to $1/2\nabla_i(a^{ij}(y,\omega)\nabla_j)$.

It has been also shown [35, 43] that the solution of (6.13) satisfies the inequality

$$\langle|\nabla\phi_T|^2\rangle + \frac{\langle\phi_T^2\rangle}{T} \leq C. \tag{6.21}$$

From where

$$\langle|\nabla\phi_T|^2\rangle \leq C \tag{6.22}$$

and

$$\langle\phi_T^2\rangle \leq CT. \tag{6.23}$$

6.6 H^1 estimates for MsFEM

In this section we obtain the convergence rate of MsFEM. Our main goal is to derive the resonance error in (6.2). For obtaining H^1 estimate, we need the following Cea's lemma:

Lemma 6.6.1 *Let u and u^h be the solutions of (6.3) and (6.8) respectively. Then*

$$E\|u-u^h\|_{1,D}^2 \leq C\ E\|u-v^h\|_{1,D}^2, \tag{6.24}$$

for any $v \in V^h$.

Theorem 6.6.2 *Let u and u^h be the solutions of (6.3) and (6.8) respectively. If $\kappa(q)$ satisfies (6.15), and $d=3$ then*

$$E\|u-u^h\|_{1,D}^2 \leq C(\frac{\epsilon}{h})^{\frac{A}{12+11A}} + C_1\epsilon^\alpha \tag{6.25}$$

where $\alpha \geq 0$, and A is defined in (6.15).

Proof. Define $v_\epsilon^h \in V^h$ such that in each $K \in \mathbf{K}^h$

$$v_\epsilon^h = (I_k u)(x) = \sum_{j=1}^{3} \alpha_j \phi_\epsilon^j(x)$$

with $\alpha_j = u_0(x_j)$ where u_0 is the homogenized solution, and x_j are the nodal points of K. Then

$$\begin{aligned} \nabla_i a_\epsilon^{ij}(x,\omega)\nabla_j v_\epsilon^h &= 0 \ \text{ in } K \\ v_\epsilon^h &= l^h \ \text{ on } \partial K \end{aligned} \tag{6.26}$$

where l^h is a linear function with the nodal values $u_0(x_j)$. The solution of (6.26) can be expanded similar to u_ϵ as

$$v_\epsilon^h = v_0^h + \epsilon\phi_t^k \nabla_k v_0^h + \theta^v. \tag{6.27}$$

Here v_0^h is a linear function because homogenized coefficients are constants. We note that the parameter t in (6.27) is not the same as the parameter T in (6.13). As we will see later t is scaled as $(\epsilon/t)^{-\eta}$, while T is scaled as $\epsilon^{-\eta}$ ($\eta > 0$), i.e., they approach to infinity at different rates. These rates represent the ratio of the small scale and the macroscopic scale of the problem. Having the expansion for v_ϵ^h, we can get the following H^1 estimate of the interpolant of MsFEM in K:

$$\begin{aligned} \|u_\epsilon - v_\epsilon^h\|_{1,K} &\le \|u_0 + \epsilon\phi_T^k\nabla_k u_0 + \theta^u - v_0 - \epsilon\phi_t^k\nabla_k v_0 - \theta^v\|_{1,K} \\ &\le \|u_0 - v_0\|_{1,K} + \|\epsilon\phi_T^k\nabla_k u_0 - \epsilon\phi_t^k\nabla_k v_0\|_{1,K} + \|\theta^v\|_{1,K} + \|\theta^u\|_{1,K} \\ &\le Ch|u_0|_{2,K} + \|\epsilon(\phi_T^k - \phi_t^k)\nabla_k v_0\|_{1,K} + \|\epsilon\phi_T^k\nabla(u_0 - v_0)\|_{1,K} + \|\theta^v\|_{1,K} + \|\theta^u\|_{1,K} \\ &\le Ch|u_0|_{2,K} + \|\epsilon(\nabla\phi_T^k - \nabla\phi_t^k)\|_{0,K} + \|\epsilon\phi_T^k\nabla^2(u_0 - v_0)\|_{0,K} \\ &+ \|\epsilon\nabla\phi_T^k\nabla_k(u_0 - v_0)\|_{0,K} + \|u_0 - v_0\|_{1,K} + \|\theta^v\|_{1,K} + \|\theta^u\|_{1,K} \le Ch|u_0|_{2,K} + \\ &\|\epsilon(\nabla\phi_T^k - \nabla\phi_t^k)\|_{0,K} + \|\epsilon\phi_T^k\|_{0,K} + \|\epsilon\nabla\phi_T^k\nabla_k(u_0 - v_0)\|_{0,K} + \|\theta^v\|_{1,K} + \|\theta^u\|_{1,K} \end{aligned} \tag{6.28}$$

Here we have used the facts that $\|u_0 - v_0\|_{1,K} \le Ch|u_0|_{2,K}$, $|u_0|_{C^2(D)} \le C$, and $|\nabla v_0| \le C$.

Then summing (6.28) over all $K \in \mathbf{K}^h$, we have

$$E\|u_\epsilon - v_\epsilon^h\|_{1,D}^2 \leq Ch^2|u_0|_{2,D}^2 + E\|\theta^u\|_{1,D} + E\|\epsilon(\nabla\phi_T^k - \nabla\phi_t^k)\|_{0,D}^2 + E\|\epsilon\phi_T^k\|_{0,D}^2 + E(\sum_{K\in D}\|\theta^v\|_{1,K}^2). \tag{6.29}$$

Here we have used

$$E(\sum_{K\in D}\|\epsilon\nabla\phi_T^k\nabla_k(u_0 - v_0)\|_{0,K}) \leq C\sum_{K\in D}\|\nabla(u_0 - v_0)\|_{0,K} \leq Ch|u_0|_{2,D}.$$

To estimate the r.h.s. of (6.29), we need to choose appropriate T and t. The choice of T is made in [56] and it was shown that for

$$T = \epsilon^{\frac{-2}{1+\gamma}} \tag{6.30}$$

we have (6.17) for some small γ. Consequently we need to estimate only the last three terms in (6.29), which we do in the following three lemma.

Remark 6.6.1. As we will see from the estimates the main resonance error is caused by the first order corrector, $\|\theta^v\|_{1,K}$ as in periodic case. We would like to note that the term $E\|\epsilon(\nabla\phi_T^k - \nabla\phi_t^k)\|_{0,D}^2$ is not present in periodic case.

Lemma 6.6.3

$$E(\sum_{K\in D}\|\theta^v\|_{1,K}^2) \leq C(\frac{\epsilon}{h})^{\frac{Ad-2A}{5A+4d+2Ad}} \tag{6.31}$$

Proof. θ^v satisfies

$$\begin{aligned}\nabla_i a_\epsilon^{ij}\nabla_j\theta^v &= \nabla_i(a_\epsilon^{ij}\nabla v_\epsilon^h - a_\epsilon^{ij}\nabla_j v_0^h - \epsilon a_\epsilon^{ij}\nabla_j\phi_t^k\nabla_k v_0^h) \\ &= -(\nabla_i a_\epsilon^{ik} + \epsilon\nabla_i a_\epsilon^{ij}\nabla_j\phi_t^k)\nabla_k v_0 = -\frac{\phi_t^k}{\epsilon t}\nabla_k v_0.\end{aligned} \tag{6.32}$$

In the last step, we have used the equation for ϕ_t. On the boundary of K, we have

$$\theta^v = -\epsilon\phi_t^k\nabla_k v_0.$$

To estimate θ^v we divide it into two parts for the convenience

$$\begin{aligned} \nabla_i a_\epsilon^{ij} \nabla_j \theta_1^v &= -\frac{\phi_t^k}{\epsilon t} \nabla_k v_0^h \ \text{ in } \ K \\ \theta_1^v &= 0 \ \text{ on } \ \partial K \end{aligned} \tag{6.33}$$

and

$$\begin{aligned} \nabla_i a_\epsilon^{ij} \nabla_j \theta_2^v &= 0 \ \text{ in } \ K \\ \theta_2^v &= -\epsilon \phi_t^k \nabla_k v_0^h \ \text{ on } \ \partial K. \end{aligned} \tag{6.34}$$

Then $\theta^v = \theta_1^v + \theta_2^v$. For θ_1^v we have

$$\int_K a_\epsilon^{ij} \nabla_j \theta_1^v \nabla_i g dx = -\int_K \frac{\phi_t^k}{\epsilon t} \nabla_k v_0 g dx$$

for any $g \in C_0^\infty(K)$.

Let $K = \bigcup_{i=1}^N K_i$ where K_i are non-intersecting open cubes with the sides $2m\epsilon$. Then denoting $< f; i >= 1/(2\epsilon m)^d \int_{K_i} f(x)dx$ we have

$$\begin{aligned} |\int_{K_i} \frac{\phi_t^k}{\epsilon t} \nabla_k v_0 g dx| \le |\int_{K_i} \frac{\phi_t}{\epsilon t} g dx| &\le |\int_{K_i} \frac{\phi_t}{\epsilon t} (g - < g; i >) dx| + |K_i| | < \frac{\phi_t}{\epsilon t} > | | < g; i > | \\ &\le C(\epsilon m) |g|_{1,K_i} \| \frac{\phi_t}{\epsilon t} \|_{0,K_i} + | < \frac{\phi_t}{\epsilon t}; i > | \|g\|_{0,K_i} \|1\|_{0,K_i}. \end{aligned} \tag{6.35}$$

Using the Cauchy inequality, Poincare inequality and the estimates (6.19) and (6.23) we get from (6.35)

$$\begin{aligned} E|\int_K \frac{\phi_T}{\epsilon T} g dx|^2 &\le E| \sum_{i=1}^N \{(\epsilon m) |g|_{1,K_i} \| \frac{\phi_t}{\epsilon t} \|_{0,K_i} + | < \frac{\phi_t}{\epsilon t}; i > | \|g\|_{0,K_i} \|1\|_{0,K_i} \}|^2 \\ &\le E(\epsilon m)^2 |g|_{1,K}^2 \| \frac{\phi_t}{\epsilon t} \|_{0,K}^2 + E(\sum_{i=1}^N | < \frac{\phi_t}{\epsilon t}; i > | \|g\|_{0,K_i} \|1\|_{0,K_i})^2 \\ &\le \frac{m^2}{t} h^d |g|_{1,K}^2 + (\frac{h}{(\epsilon m)})^d \epsilon^{-2} t^{-1} (\exp(-c(\log(t))^2) + \frac{1}{q^A} + (\frac{tq^d}{m^d})^{1/2} \log(t)) \|g\|_{0,K} (\epsilon m)^d \\ &\le \frac{m^2}{t} h^d |g|_{1,K}^2 + \epsilon^{-2} t^{-1} (\exp(-c(\log(t))^2) + \frac{1}{q^A} + (\frac{tq^d}{m^d})^{1/2} \log(t)) h^{d+2} |g|_{1,K}^2 \end{aligned} \tag{6.36}$$

Consequently from (6.33) using regularity estimates for PDE, and (6.36) we have

$$E\sum_{K\in D}\|\theta_1^v\|_{1,K}^2 \le \frac{m^2}{t} + h^2\epsilon^{-2}t^{-1}(\exp(-c(\log(t))^2) + \frac{1}{q^A} + \frac{tq^d}{m^d}^{1/2}\log(t)). \tag{6.37}$$

Equating the last two terms we can find the optimum value for q; $q \sim (m^d/t)^{1/(d+2A)}$. Then (6.37) becomes

$$E\sum_{K\in D}\|\theta_1^v\|_{1,K}^2 \le (\frac{\epsilon}{h})^{-2}t^{-1}((\frac{t}{m^d})^{\frac{A}{d+2A}}\log(t) + \exp(-c(\log(t))^2) + \frac{m^2}{t}. \tag{6.38}$$

Moreover, equating the first and the last terms we can find the optimum value for m

$$m \sim (\frac{\epsilon}{h})^{-\frac{2(d+2A)}{2d+4A+Ad}}t^{-\frac{2A+2d+Ad}{2d+4A+Ad}}$$

and the corresponding estimate

$$E\sum_{K\in D}\|\theta_1^v\|_{1,K}^2 \le (\frac{\epsilon}{h})^{-\frac{2(d+2A)}{2d+4A+Ad}}t^{\frac{2A}{2d+4A+Ad}-1} + (\frac{\epsilon}{h})^{-2}t^{-1}\exp(-C(\log(t))^2). \tag{6.39}$$

Now we show that

$$E\sum_{K\in D}\|\theta_2^v\|_{1,K}^2 \le C(\frac{\epsilon}{h})^{2/3}t^{1/3}. \tag{6.40}$$

From (6.34) we have

$$\|\theta_2^v\|_{1,K}^2 \le C\|\epsilon\phi_t\|_{H^{1/2}(\partial K)} \le C\|\epsilon\phi_t(1-\tau)\|_{1,K}$$

where $\tau \in C_0^\infty(K)$ is the cutoff function for domain K: $\tau(x) = 1$ if $\inf(|x-y|,\ y \notin K) \ge \delta$ and $\tau = 0$ on ∂K. Consequently,

$$|\nabla\tau| \le \frac{C}{\delta}.$$

Furthermore using the estimates (6.22) and (6.23) for ϕ_t, we have

$$E\|\theta_v^2\|_{1,K}^2 \leq E\int_K \epsilon^2\phi_t^2(\nabla\tau)^2dx + E\int_K \epsilon^2(\nabla\phi_t)^2\tau^2dx \leq Ch^d(\frac{\epsilon}{\delta})^2t + Ch^{d-1}\delta = Ch^{d-1}(\delta + h(\frac{\epsilon}{\delta})^2t). \tag{6.41}$$

Choosing an optimum value for δ we have $\delta \sim \epsilon^{2/3}t^{1/3}h^{1/3}$ and

$$E\|\theta_2^v\|_{1,K}^2 \leq Ch^{d-2/3}\epsilon^{2/3}t^{1/3}. \tag{6.42}$$

Summing (6.42) over all $K \in \mathbf{K}^h$, we get (6.40).

Combining the estimates for θ_1^v and θ_2^v we have

$$E\sum_{K\in\mathbf{K}^h}\|\theta^v\|_{1,K}^2 \leq (\frac{\epsilon}{h})^{-\frac{4(d+2A)}{2d+4A+Ad}}t^{-\frac{2A+2d+Ad}{2d+4A+Ad}} + (\frac{\epsilon}{h})^{2/3}t^{1/3}. \tag{6.43}$$

For the optimum t in (6.43)

$$t \sim (\frac{\epsilon}{h})^{-\frac{8d+16A+Ad}{5A+4d+2Ad}},$$

the estimate (6.43) becomes

$$E\sum_{K\in\mathbf{K}^h}\|\theta^v\|_{1,K}^2 \leq (\frac{\epsilon}{h})^{\frac{Ad-2A}{5A+4d+2Ad}}.$$

As we can see if $d = 2$ the upper bound for $E\sum_{K\in\mathbf{K}^h}\|\theta^v\|_{1,K}^2$ is of order 1. So the method does not converge in general if $d = 2$. But in the case $d = 3$ this estimate becomes

$$E\sum_{K\in\mathbf{K}^h}\|\theta^v\|_{1,K}^2 \leq (\frac{\epsilon}{h})^{\frac{A}{11A+12}}.$$

□

Lemma 6.6.4

$$E\|\epsilon\phi_T\|_{0,D}^2 \leq C\epsilon^\alpha. \tag{6.44}$$

Proof.

Using the estimate (6.23) we have

$$E\|\epsilon\phi_T\|^2_{0,D} \leq C\epsilon^2 T.$$

As we noted eariler for the convergence of u_ϵ, T has been chosen to be $T = \epsilon^{-2/(1+\gamma)}$. Then

$$E\|\epsilon\phi_T\|^2_{0,D} \leq C\epsilon^{\frac{2\gamma}{1+\gamma}}.$$

□

Lemma 6.6.5

$$E\|\epsilon\nabla\phi^k_T - \epsilon\nabla\phi^k_t\|^2_{0,D} \leq C(\frac{\epsilon}{h})^{\frac{Ad-2A}{5A+4d+2Ad}}. \tag{6.45}$$

Proof. For further convinience we omit the index k in the proof. To prove the lemma, we introduce ψ_∞, the limit of $\nabla\phi_T$ as $T \to \infty$. We will show that this limit exists. Then writing l.h.s. of (6.45) as

$$E\|\epsilon\nabla\phi_T - \epsilon\nabla\phi_t\|^2_{0,D} \leq E\|\epsilon\nabla\phi_T - \psi_\infty\|^2_{0,D} + E\|\epsilon\nabla\phi_t - \psi_\infty\|^2_{0,D}, \tag{6.46}$$

we just need to derive the convergence rate of $E\|\epsilon\nabla\phi_T - \psi_\infty\|^2_{0,D}$ as $T \to \infty$. Following to [56] we show that

$$E\|\epsilon\nabla\phi_T - \psi_\infty\|^2_{0,D} \leq CT^{-s}$$

for some positive s. Then using similar estimate for ϕ_t we get the desired result.

Comparing the equation (6.13) for different values of T, $T = T_1$ and $T = T_2$ $(T_2 \geq T_1)$, we have

$$E < |\nabla\phi_{T_1} - \nabla\phi_{T_2}|^2 > + E\frac{< (\phi_{T_1} - \phi_{T_2})^2 >}{T_2} \leq C|(\frac{1}{T_1} - \frac{1}{T_2})E < \phi_{T_1}(\phi_{T_1} - \phi_{T_2}) > |.$$

Furthermore, because of homogenity of the field ϕ_T

$$\begin{aligned} E < \phi_{T_1}(\phi_{T_1} - \phi_{T_2}) > &= E[\phi_{T_1}(\phi_{T_1} - \phi_{T_2})] = E[\phi_{T_1}][\phi_{T_1} - \phi_{T_2}] \\ &+ E[\phi_{T_1}(\phi_{T_1} - \phi_{T_2} - [\phi_{T_1} - \phi_{T_2}])] \end{aligned} \tag{6.47}$$

where $[f] = 1/(2m)^d \int_{|x|\leq m} f(x)dx$. The first term on the r.h.s of (6.47) can be estimated as

$$\begin{aligned} |E[\phi_{T_1}][\phi_{T_1} - \phi_{T_2}]| &\leq C(E[\phi_{T_1}]^2)^{1/2}(E[\phi_{T_1}]^2 + E[\phi_{T_2}^2)^{1/2} \\ &\leq CT_1^{1/2}T_2^{1/2}(\exp(-C(\log(T_1))^2) + \frac{1}{q^A} + (\frac{T_1 q^d}{m^d})^{1/2}\log(T_1)). \end{aligned} \tag{6.48}$$

Choosing an optimal value for q, $q \sim (m^d/T_1)^{1/(d+2A)}$ we have

$$|E[\phi_{T_1}][\phi_{T_1} - \phi_{T_2}]| \leq CT_1^{1/2}T_2^{1/2}(\exp(-C(\log(T_1))^2) + (\frac{T_1}{m^d})^{1/(2+d/A)}\log(T_1)). \tag{6.49}$$

The second term on the r.h.s of (6.47) can be estimated by a version of Poincare - Friedrichs inequality for functions with the average zero as

$$\begin{aligned} E[\phi_{T_1}(\phi_{T_1} - \phi_{T_2} - [\phi_{T_1} - \phi_{T_2}])] &\leq (E[\phi_{T_1}]^2)^{1/2}(E[\phi_{T_1} - \phi_{T_2} - [\phi_{T_1} - \phi_{T_2}])^{1/2} \\ &\leq CT_1^{1/2}m < |\nabla\phi_{T_1} - \nabla\phi_{T_2}|^2 >^{1/2} \end{aligned} \tag{6.50}$$

(6.47), (6.49) and (6.50) lead to the ineqality

$$< |\nabla\phi_{T_1} - \nabla\phi_{T_2}|^2 > \leq C(\frac{m^2}{T_1} + \frac{T_2}{T_1}^{1/2}(\exp(-C(\log(T_1))^2) + (\frac{T_1}{m^d})^{\frac{1}{2+d/A}}\log(T_1). \tag{6.51}$$

Taking $T_2 = T_1^{1+\gamma}$ and choosing the optimum value for m,

$$m \sim T_1^{\frac{A(2\gamma+2)+\gamma d}{8A+4d+2Ad}}$$

we have

$$< |\nabla\phi_{T_1} - \nabla\phi_{T_2}|^2 > \leq CT_1^{-\frac{2A+2d+Ad-2A\gamma-\gamma d}{4A+2d+Ad}} = CT^{-\eta} \tag{6.52}$$

where $\eta = \frac{2A+2d+Ad-2A\gamma-\gamma d}{4A+2d+Ad}$ and it is a positive number for small γ.

Now choosing $T(k) = \exp((1+\gamma)^k)$ ($T(k+1) = T(k)^{1+\gamma}$ as it is assumed) and choosing $T_1 = T(n)$ we have

$$< |\nabla\phi_{T_1} - \psi_\infty > \leq C\sum_{k\geq n} < |\nabla\phi_{T(k)} - \nabla\phi_{T(k+1)}|^2 > \leq C\sum_{k\geq n} T(k)^{-\eta} \leq CT(n)^{-\eta}. \tag{6.53}$$

Consequently, using (6.53) we can estimate the first term on the r.h.s. of (6.46) as

$$E\|\epsilon\nabla\phi_T - \psi_\infty\|^2_{0,D} \leq CT^{-\frac{2A+2d+Ad-2A\gamma-\gamma d}{4A+2d+Ad}}\log(T).$$

Similarly, the estimate for the second term on the r.h.s of (6.46) becomes

$$E\|\epsilon\nabla\phi_t - \psi_\infty\|^2_{0,D} \leq Ct^{-\frac{2A+2d+Ad-2A\gamma-\gamma d}{4A+2d+Ad}}\log(t).$$

Taking into accound the values for T and t , $T = \epsilon^{-2/(1+\gamma)}$, and

$$t = (\frac{\epsilon}{h})^{-\frac{8d+16A+Ad}{5A+4d+2Ad}}$$

we have

$$E\|\epsilon\nabla\phi_T - \psi_\infty\|^2_{0,D} \leq C\epsilon^{\tilde{\alpha}}$$

and

$$E\|\epsilon\nabla\phi_t - \psi_\infty\|^2_{0,D} \leq C(\frac{\epsilon}{h})^{\frac{(8d+16A+Ad)(2A+2d+Ad-2A\gamma-\gamma d)}{(5A+4d+2Ad)(4A+2d+Ad)}}.$$

It can be easily checked that

$$E\|\epsilon\nabla\phi_t - \psi_\infty\|^2_{0,D} \leq C(\frac{\epsilon}{h})^{\frac{Ad-2A}{5A+4d+2Ad}}.$$

□

Eventually, from (6.29) using (6.31), (6.44), and (6.45) we get the desired result (6.25) for $d = 3$.

□

Remark 6.6.2. We can observe numerically that the resonance error is no longer ϵ/h if the coefficients are random. Indeed, in Fig. 3.2 the line denoted with square represent the L_2 error of MsFEM for different values of the mesh size h in the case of log-normal field. The realizations of these random fields are used as coefficients a^{ij}_ϵ and the small scales range is between 1/64 and 1/512. We can see from this figure that the resonance error for

fixed small scales behaves approximately as $h^{-0.1}$. This indicates that the resonance error is $(\epsilon/h)^{0.1}$.

Chapter 7 Nonconforming MsFEM and its analysis

7.1 Summary

In this section we analyze the nonconforming method introduced in [53]. As we have seen from the analysis of conforming MsFEM presented in previous sections these methods suffer from the resonance error. This resonance is due to the mismatch between the local construction of multiscale base functions and the global nature of the elliptic problems. Motivated by the error analysis, an *over-sampling* technique was proposed to overcome the difficulty due to scale resonance in [26]. The idea is quite simple and easy to implement. Since the boundary layer in the first order corrector is $O(\epsilon)$ thin, we can sample in a domain with size larger than $h + \epsilon$ and use only the interior sampled information to construct the bases (h is the mesh size and ϵ is physical scale). By doing this, the boundary layer in the larger domain has no influence on the base functions. Now the corresponding first order correctors are free of boundary layers. As a result, we obtain an improved rate of convergence.

Unfortunately, the over-sampling technique results in a *nonconforming* MsFEM method. The previous analysis needs to be modified to take into account the nonconforming error. In [21] , we perform a careful estimate of the nonconforming errors in both H^1 norm and the L^2 norm. The analysis shows that the nonconforming error is small. Our analysis also reveals another source of resonance, which is the mismatch between the mesh size and the "perfect" sample size. In the case of a periodic structure, the "perfect" sample size is the length of an integer multiple of the period. We call the new resonance the "cell resonance." In the error expansion, this resonance effect appears as a *higher* order correction. Although the over-sampling helps eliminate the leading order resonance error, we find that over-sampling alone does not remove the cell resonance error, which dominates the nonconforming error. From our computational experience, the cell resonance errors seem to be generically small and are rarely observed in computations. Nonetheless, we propose an averaged over-sampling technique to eliminate this cell resonance error. This reduces the nonconforming error by eliminating the resonance error.

7.2 Description of nonconforming MsFEM

The analysis of nonconforming MsFEM is performed on one periodic problem (3.1), i.e., when $a_\epsilon^{ij} = a^{ij}(x/\epsilon)$ is a smooth periodic function with a period ϵ satisfying the ellipticity conditions.

The base functions for nonconforming MsFEM are constructed in the following way. We first construct base functions ψ_ϵ^j in a sampling domain $S \supset K$(see Fig. 8.4) by solving

$$\nabla_i a_{ik}^\epsilon \nabla_k \psi_\epsilon^j = 0 \quad \text{in S},$$

where ψ_ϵ^j is piecewise linear along ∂S and $\psi_\epsilon^j = \delta_{ij}$ at the nodal points of S. For simplicity, we assume S to be triangular and hence having the same number of nodal points as K. Moreover we choose S sufficiently large so that $\text{diam}(S) = h_1 > h$ and ∂S is away from ∂K at a distance of order ϵ. Next, the base functions ϕ_ϵ on K is constructed from the linear superposition of ψ_ϵ:

$$\phi_{\epsilon,i} = \sum_{j=1}^{d} c_{ij} \psi_\epsilon^j, \tag{7.1}$$

where i is the index of the nodal point and constants c_{ij} are determined by the condition $\phi_{0,i}(x_j) = \delta_{ij}$, x_j being the nodal points of K. Note that $\phi_{0,i} = \sum_{j=1}^{d} c_{ij} \psi_0^j$ and ψ_0^i, $(i = 1, 2, 3)$ are linear functions.

By this procedure, the boundary layer structure near ∂S is avoided. To make this more precise, we expand ψ_ϵ as

$$\psi_\epsilon = \psi_0 + \epsilon \chi^p \nabla_p \psi_0 + \epsilon \theta',$$

where ψ_0 is the homogenized part of ψ_ϵ, which is linear. For the corrector θ', we have $\theta' = \eta^p \nabla_p \psi_0$ with η^p being the solution of

$$\nabla_i a_\epsilon^{ij} \nabla_j \eta^p = 0 \quad \text{in } S \qquad \text{and} \qquad \eta^p = \chi^p \quad \text{on } \partial S. \tag{7.2}$$

The analysis of the behavior of η^p is a complicated problem, which is yet to be rigorously carried out. In [9] the problem (7.2) in the half space $S = \mathbf{R}_+^2$ whose boundary aligned

with the period was analyzed . Beside the fact that $\nabla\eta^p$ exponentially decays away from ∂S, the authors found out that there exists a constant χ_*^p such that η^p, the solution of (7.2) with boundary condition $\eta^p = \chi^p - \chi_*^p$ on ∂S, decays exponentially too. Moskow and Vogelius [40] investigated some special features of the behavior of η^p for polygon domains with specially oriented sides. It was shown that η^p has exponential boundary layers away from the corners. We note that these boundary layers are the cause of large H^1 norm of η^p and the resonance in MsFEM [28]. Using the over-sampling technique we eliminate these boundary layers from the base functions. The behavior of η^p in the interior of S has not been rigorously carried out yet. But it was numerically demonstrated [26] that $\nabla\eta^p$ has bounded oscillations in the interior of S. Therefore we assume:

Assumption A: $\nabla\eta^p \in L_\infty(K)$ when $K \subset S$ is away from ∂S at least at a distance ϵ.

Indeed, as we will see later from numerical results the computational error is not caused by the terms containing η^p. We note that this assumption is based on numerical result. Using the over-sampling technique we get rid of a large part of the error produced by θ.

An important consequence of the above construction is that the base functions ϕ are no longer continuous across the internal boundaries of the elements, nor are they zero on the external boundaries. Setting the base functions to be zero outside the external boundaries of the elements introduces first order discontinuities of ϕ along all sides of the elements. Thus, these base functions are nonconforming and V_h spanned by ϕ is no longer in H^1. This complicates the analysis of MsFEM. In the next section, we will provide the estimates of the nonconforming error. We remark that despite the complications, the nonconforming finite element methods have been widely used in practice when conforming elements that satisfy certain physical and numerical properties are too tedious or impossible to construct [13, 49].

In our computations, we usually choose a large sample domain, S, to contain many elements to improve the efficiency of computation [26]. Two implementations of the over-sampling method have been tested. In [26], we choose S to be non-overlapped, which is easy to implement. This removes the boundary layers in the interior elements, but the elements adjacent to ∂S are still under the influence of the boundary layers. The accuracy of solutions can be further improved by a fully implemented over-sampling method [27], in which the sample domains have some overlaps (with width of $O(\epsilon)$). For convenience and

clarity in the derivations below, let us introduce some notations here. Let

$$\mathbf{K}_h(S) = \{K : K \in \mathbf{K}, K \subset S\}$$

be the set of all elements contained in the sample domain S and

$$\tilde{S} = \bigcup_{K \in \mathbf{K}_h(S)} K \subseteq S \tag{7.3}$$

be union of these sets.

7.3 H^1 estimates

Because of non-conformity of the base functions, we no longer enjoy Cea's lemma. But we have the following estimate between the exact solution and the numerical solution in the energy norm [12] (see section 3.4):

$$\|u_\epsilon - u_\epsilon^h\|_{h,\Omega} \leq C(\inf_{v_\epsilon^h \in \hat{H}^h} \|u_\epsilon - v_\epsilon^h\|_{h,\Omega} + \sup_{w_\epsilon^h \in \hat{H}^h} \frac{|f(w_\epsilon^h) - a(u_\epsilon, w_\epsilon^h)|}{\|w_\epsilon^h\|_h}), \tag{7.4}$$

where $\hat{H}^h$ is the finite dimensional space generated by the nonconforming basis functions (in general $\hat{H}^h \not\subset H^1$),

$$\|w_\epsilon^h\|_{h,\Omega} = (\sum_{K \in \mathbf{K}^h} \int_K |\nabla w_\epsilon^h|^2 dx)^{1/2}, \tag{7.5}$$

and

$$|f(w_\epsilon^h) - a(u_\epsilon, w_\epsilon^h)| = |\int_\Omega f w_\epsilon^h dx - \sum_{K \in \mathbf{K}^h} \int_K a_\epsilon^{ij} \nabla_i u_\epsilon \nabla_j w_\epsilon^h dx|. \tag{7.6}$$

7.3.1 Case of $h \gg \epsilon$

Theorem 7.3.1 *Let u_ϵ be the solution of (4.1) and u_ϵ^h be the numerical solution computed using MsFEM with over-sampling. Assuming that Assumption A is valid and the homogenized part of u_ϵ, i.e., u_0, is in $W^{1,\infty}(\Omega)$, we have*

$$\|u_\epsilon - u_\epsilon^h\|_{h,\Omega} \leq C\frac{\epsilon}{h} + C_1\sqrt{\epsilon} + C_2 h \qquad (h > \epsilon). \tag{7.7}$$

Remark 7.3.1. The assumption $u_0 \in W^{1,\infty}(\Omega)$ holds for two dimensional convex polygon domains [36].

In the following, we first consider the "conforming error" on the r.h.s. of (7.4). Then, we analyze the nonconforming error given by the second term. For the analysis of nonconforming error we will need the following lemma.

Lemma 7.3.2 *Let D be a unit box in $\mathbf{R}^2$ and $N(y) \in L^\infty(D)$ be a 1-periodic function and $\langle N \rangle = 0$. Then for $\forall f \in H^1(K) \bigcap L^\infty(K)$, $K \subset \mathbf{R}^2$, and $diam(K) = h$, we have*

$$|\int_K f(x) N(\frac{x}{\epsilon}) dx| \leq C\epsilon h.$$

Proof. Define

$$f_i = \int_{Y_i} f\, dx,$$

where Y_i is a periodic cell of $N(x/\epsilon)$, $Y_i \in K$. Then

$$\|f - f_i\|_{L_2(Y_i)} \leq \epsilon \|\nabla f\|_{L_2(Y_i)}.$$

Denote $K' = \cup_{Y_i \subset K} Y_i$. We have

$$\begin{aligned}
|\int_K f(x) N(\frac{x}{\epsilon}) dx| &\leq |\sum_{Y_i \subset K} \int_{Y_i} (f - f_i) N(\frac{x}{\epsilon}) dx| + |\sum_{Y_i \subset K} \int_{Y_i} f_i N(\frac{x}{\epsilon}) dx| \\
&\quad + |\int_{K \setminus K'} f(x) N(\frac{x}{\epsilon}) dx| \\
&\leq \sum_{Y_i \subset K} \|f - f_i\|_{L_2(Y_i)} \|N(\frac{x}{\epsilon})\|_{L_2(Y_i)} + |\int_{K \setminus K'} f(x) N(\frac{x}{\epsilon}) dx| \\
&\leq C\epsilon \|\nabla f\|_{L_2(K')} \|N\|_{L_2(K')} + |\int_{K \setminus K'} f(x) N(\frac{x}{\epsilon}) dx| \\
&\leq C\epsilon h.
\end{aligned}$$

In the last two steps we have used the Cauchy-Schwarz inequality. □

Proof of Theorem 3.1. For the conforming part, we notice that in each element $K \in \mathbf{K}^h$, $v_\epsilon^h \in \hat{H}^h$ can be expanded as

$$v_\epsilon^h = v_0^h + \epsilon\chi^p \nabla_p v_0^h + \theta^h, \tag{7.8}$$

where $v_0^h = v_k c_j^k \psi_0^j$ with $v_k = u_0(x_k)$ and $\theta^h = v_k c_j^k \theta_j'$, ψ_ϵ^j are the base functions, θ_j' are corresponding correctors in the sample domain $S \supset K$ and c_j^k are chosen in order to give $c_j^k \psi_\epsilon^j$ the desirable values at the nodal points. So $\phi_0^k = c_j^k \psi_0^j$ are linear functions and $\phi_0^k(x_j) = \delta_{jk}$. Then by (3.8)

$$\|u_\epsilon - v_\epsilon^h\|_{h,K} \le \|u_0 - v_0^h\|_{h,K} + \|\epsilon\chi^p \nabla_p (u_0 - v_0^h)\|_{h,K} + \epsilon\|\theta^h\|_{h,K} + \epsilon\|\theta^u\|_{h,K}, \tag{7.9}$$

where $\|u\|_{h,K} = (\int_K (\nabla u)^2 dx)^{1/2}$. Taking into account that

$$\|u_0 - v_0^h\|_{h,K} \le Ch|u_0|_{2,K}, \tag{7.10}$$

and $\chi^p \in W^{1,\infty}(Y)$ (the latter follows from $a^{ij}(y) \in W^{1,p}$, $p > 2$, cf. Theorem 15.1 in [36]), we have

$$\|u_\epsilon - v_\epsilon^h\|_{h,K} \le Ch|u_0|_{2,K} + \epsilon\|\theta^h\|_{h,K} + \epsilon\|\theta^u\|_{h,K}, \tag{7.11}$$

where θ^h satisfies the following equation:

$$\nabla_i a_\epsilon^{ij} \nabla_j \theta^h = 0 \quad \text{in } S \qquad \text{and} \qquad \theta^h = \chi^p \nabla_p v_0^h \quad \text{on } \partial S.$$

Since v_0^h is linear in K, we have $\theta^h = \eta^p \nabla_p v_0^h$, where η^p is defined by (7.2) and $\nabla\eta^p$ is assumed to be in $L_\infty(K)$ (Assumption A). Then using $\|\nabla v_0^h\|_{L_2(K)} \le C\|\nabla u_0\|_{L_2(K)}$ (see Appendix B), we get

$$\|\epsilon\theta^h\|_{h,K} \le C\epsilon\|\nabla v_0^h\|_{L_2(K)} \le C\epsilon\|\nabla u_0\|_{L_2(K)}.$$

Summing (7.11) over all $K \in \mathbf{K}^h$ we have

$$\|u_\epsilon - v_\epsilon^h\|_{h,\Omega} \le Ch|u_0|_{2,\Omega} + \epsilon\|\nabla u_0\|_{L_2(\Omega)} + C\sqrt{\epsilon} \le C(h + \sqrt{\epsilon}).$$

Here we have used the fact that $|\epsilon\theta^u|_{1,\Omega} \leq C\sqrt{\epsilon}$ [32].

The nonconforming error in (7.4) can be written in the following way:

$$\begin{aligned} N.C. &\equiv \sup_{w_\epsilon^h \in \hat{H}^h} \frac{|f(w_\epsilon^h) - a(u_\epsilon, w_\epsilon^h)|}{\|w_\epsilon^h\|_h} = \sup_{w_\epsilon^h \in \hat{H}^h,\ \|w_\epsilon^h\|_h=1} |f(w_\epsilon^h) - a(u_\epsilon, w_\epsilon^h)| \\ &\leq \sup_{w_\epsilon^h \in \hat{H}^h,\ \|w_\epsilon^h\|_h=1} \{|f(w_\epsilon^h - l^h) - a(u_\epsilon, w_\epsilon^h - l^h)| + |f(l^h) - a(u_\epsilon, l^h)|\}, \end{aligned} \tag{7.12}$$

where l^h is the homogenized part of w_ϵ^h. Note that l^h is a linear function. Because of the conformity of the linear base functions, the second term on the r.h.s of (7.12) is zero. Then (7.12) becomes

$$\begin{aligned} N.C. &\leq C \sup_{w_\epsilon^h \in \hat{H}^h,\ \|w_\epsilon^h\|_h=1} \{|f(w_\epsilon^h - l^h) - a(u_\epsilon, w_\epsilon^h - l^h)|\} \\ &\leq C \sup_{w_\epsilon^h \in \hat{H}^h,\ \|w_\epsilon^h\|_h=1} |f(w_\epsilon^h - l^h)| + C \sup_{w_\epsilon^h \in \hat{H}^h,\ \|w_\epsilon^h\|_h=1} |a(u_\epsilon, w_\epsilon^h - l^h)|. \end{aligned} \tag{7.13}$$

First let us estimate the second term of the r.h.s. of (7.13). The first term can be estimated analogously. Since w_ϵ^h is a linear combination of the base functions w_ϵ^h satisfies $\nabla_i a_\epsilon^{ij} \nabla_j w_\epsilon^h = 0$ in S. Using the expansion for $w_\epsilon^h = l^h + \epsilon\chi^p \nabla_p l^h + \epsilon\theta$ we have

$$|\int_K a_\epsilon^{ij} \frac{\partial u^\epsilon}{\partial x_j} \frac{\partial (w_\epsilon^h - l^h)}{\partial x_j} dx| \leq |\int_K a_\epsilon^{ij} \frac{\partial u^\epsilon}{\partial x_j} (\epsilon \nabla_i \chi^p) \nabla_p l^h dx| + |\int_K a_\epsilon^{ij} \frac{\partial u^\epsilon}{\partial x_j} \frac{\partial (\epsilon\theta)}{\partial x_j} dx|. \tag{7.14}$$

Observe that $u^\epsilon = u_0 + \epsilon\chi^p \nabla_p u_0 + \epsilon\theta^u$ and

$$a_\epsilon^{ij} \nabla_j (u_0 + \epsilon\chi^p \nabla_p u_0) = a_*^{ij} \nabla_j u_0 + g^{ij} \nabla_j u_0 + \epsilon\chi^k a_\epsilon^{ij} \nabla_j \nabla_k u_0,$$

where $g^{ij} = a_\epsilon^{ij} + a_\epsilon^{ip} \nabla_p \chi^j - a_*^{ij}$. Then (7.14) becomes:

$$\begin{aligned} |\int_K a_\epsilon^{ij} \frac{\partial u^\epsilon}{\partial x_j} \frac{\partial (w_\epsilon^h - l^h)}{\partial x_j} dx| &\leq |\int_K a_\epsilon^{ij} \frac{\partial u^\epsilon}{\partial x_j} (\epsilon \nabla_i \chi^p) \nabla_p l^h dx| + |\int_K a_\epsilon^{ij} \frac{\partial u^\epsilon}{\partial x_j} \frac{\partial (\epsilon\theta)}{\partial x_j} dx| \\ &\leq |\nabla_p l^h| \, |\int_K a_*^{ij} \frac{\partial u_0}{\partial x_j} \epsilon \nabla_i \chi^p dx| + |\nabla_p l^h| \, |\int_K g^{ij} \frac{\partial u_0}{\partial x_j} \epsilon \nabla_i \chi^p dx| \\ &\quad + |\nabla_p l^h| \, |\int_K a_\epsilon^{ij} \frac{\partial^2 u_0}{\partial x_j \partial x_k} \epsilon\chi^k (\epsilon \nabla_i \chi^p) dx| \\ &\quad + |\nabla_p l^h| \, |\int_K a_\epsilon^{ij} \frac{\partial (\epsilon\theta^u)}{\partial x_j} \epsilon \nabla_i \chi^p dx| + |\int_K a_\epsilon^{ij} \frac{\partial u^\epsilon}{\partial x_j} \frac{\partial (\epsilon\theta)}{\partial x_j} dx|. \end{aligned} \tag{7.15}$$

Using Lemma 7.3.2 along with the facts that $u_0 \in H_0^2(\Omega) \bigcap W^{1,\infty}(\Omega)$ and $\chi^p \in W^{1,\infty}(Y)$ (the latter follows from $a^{ij} \in W^{1,p}(Y)$, $p > 2$, cf. Theorem 15.1 in [36]), we conclude that the first two terms on the r.h.s of (7.15) are less than $C\epsilon h|\nabla l|$ because $\langle g^{ij}\nabla_i\chi^p\rangle = 0$ ($|\nabla l| = \max_{p=1,2}|\nabla_p l^h|$). It can be easily shown that the third term on the r.h.s. of (7.14) is bounded by $C\epsilon h|\nabla l^h||u_0|_{2,K}$. Then

$$|\int_K a_\epsilon^{ij}\frac{\partial u^\epsilon}{\partial x_j}\frac{\partial(w_\epsilon^h - l^h)}{\partial x_j}dx| \le C\epsilon h|\nabla l^h| + C\epsilon h|\nabla l^h||u_0|_{2,K} + C|\nabla l^h||\epsilon\theta^u|_{1,K}|\epsilon\chi|_{1,K} + C|u_\epsilon|_{1,K}|\epsilon\theta|_{1,K}.$$

Noting that $h|\nabla l^h| \le C\|\nabla l^h\|_{L_2(K)}$ and taking into account (see Appendix B)

$$\|\nabla l^h\|_{L_2(K)} \le C\|\nabla w_\epsilon^h\|_{L_2(K)},$$

we have

$$\begin{aligned}|\int_K a_\epsilon^{ij}\frac{\partial u_\epsilon}{\partial x_j}\frac{\partial(w_\epsilon^h - l^h)}{\partial x_j}dx| \le & C\epsilon\|\nabla w_\epsilon^h\|_{L_2(K)} + C\epsilon\|\nabla w_\epsilon^h\|_{L_2(K)}|u_0|_{2,K} \\ & + C\|\nabla w_\epsilon^h\|_{L_2(K)}|\epsilon\theta^u|_{1,K} + C|u_\epsilon|_{1,K}|\epsilon\theta|_{1,K}.\end{aligned} \tag{7.16}$$

By the same argument for $\epsilon\theta$, one has

$$|\epsilon\theta|_{1,K} \le C\epsilon\|\nabla l^h\|_{L_2(K)} \le C\epsilon\|\nabla w_\epsilon^h\|_{L_2(K)}.$$

Summing (7.16) over all $K \in \mathbf{K}^h$ and using Cauchy-Schwarz inequality, we obtain

$$|\int_\Omega a_\epsilon^{ij}\frac{\partial u^\epsilon}{\partial x_j}\frac{\partial(w_\epsilon^h - l^h)}{\partial x_j}dx| \le C\frac{\epsilon}{h} + C_3\epsilon|u_0|_{2,\Omega} + C_2|\epsilon\theta^u|_{1,\Omega} + C_1\epsilon \le C\frac{\epsilon}{h} + C_2\sqrt{\epsilon} + C_1\epsilon,$$

where we have used $\|\nabla w_\epsilon^h\|_h = 1$ and $|\epsilon\theta^u|_{1,\Omega} \le \sqrt{\epsilon}$. Analogously,

$$\begin{aligned}|\sum_{K\in\mathbf{K}^h}\int_K f(x)(w_\epsilon^h - l^h)dx| \le & |\sum_{K\in\mathbf{K}^h}\int_K f\epsilon\chi^p\nabla_p l^h dx| + |\sum_{K\in\mathbf{K}^h}\int_K f\epsilon\theta dx| \\ \le & \sum_{K\in\mathbf{K}^h}\epsilon\|\nabla l^h\|_{L_2(K)}\|f\|_{L_2(K)} + \sum_{K\in\mathbf{K}^h}\|\epsilon\theta\|_{L_2(K)}\|f\|_{L_2(K)}.\end{aligned} \tag{7.17}$$

The second term on the r.h.s of (7.17) can be treated as in the previous case and the first term is bounded by $C\epsilon$. Combining the estimates for conforming and nonconforming errors we have (7.7).□

Remark 7.3.2. The estimate (7.7) is better than our previous estimate for MsFEM without over-sampling [28]. The leading order error without over-sampling is $(\epsilon/h)^{\frac{1}{2}}$ due to the boundary layers in θ. Here we notice that the resonance still exists. Inspecting the proof, we see that the ϵ/h term in (7.7) comes from the integrals containing χ (see (7.14)). These are the cell terms mentioned in §1. We see below (Section 4) that a similar term gives rise to the resonance error in the L_2 estimates.

7.3.2 Case of $h \ll \epsilon$.

Theorem 7.3.3 *Let u_ϵ be the solution of (4.1) and u_ϵ^h be the numerical solution computed using MsFEM with over-sampling. Assuming one element per sample, i.e., each sample contains only one element and $a^{ij}(y) \in W^{1,\infty}(Y)$, we have*

$$\|u_\epsilon - u_\epsilon^h\|_{h,\Omega} \leq C\frac{h}{\epsilon} \quad (h \ll \epsilon). \tag{7.18}$$

Remark 7.3.3. The theorem holds in the case of arbitrary number of elements per sample under the assumption $u_\epsilon \in W^{1,\infty}$. This assumption is true for smooth domains Ω.

As in the previous section, we will estimate the conforming and nonconforming error. For the proof of this case we assume that $\phi_{\epsilon,i}(x_j) = \delta_{ij}$, x_j being the nodal points of K. The main difficulty is that because the boundary of the sample element has corners, the base functions are not smooth in any subdomain of S. We overcome this difficulty by introducing auxiliary functions $v = (v_1, v_2)$, the solutions of (7.21) and (7.22). These functions are the same for the base functions of each sample.

Using (7.4) and (7.13) we obtain

$$\|u_\epsilon - u_\epsilon^h\|_{h,\Omega} \leq C(\inf_{v_\epsilon^h \in \hat{H}^h} \|u_\epsilon - v_\epsilon^h\|_{h,\Omega} + \sup_{w_\epsilon^h \in \hat{H}^h,\ \|w_\epsilon^h\|_h = 1} |f(w_\epsilon^h - l^h) - a(u_\epsilon, w_\epsilon^h - l^h)|). \tag{7.19}$$

Before deriving the estimates, we introduce the following notations:

$$|f(w_\epsilon^h) - a(u_\epsilon, w_\epsilon^h)|_{\tilde{S}} = |\int_{\tilde{S}} f w_\epsilon^h dx - \sum_{K \in \mathbf{K}^h(S)} \int_K a_\epsilon^{ij} \nabla_i u_\epsilon \nabla_j w_\epsilon^h dx|,$$

$$\|w_\epsilon^h\|_{h,\tilde{S}} = \left(\sum_{K \in \mathbf{K}^h(S)} \int_K |\nabla w_\epsilon^h|^2 dx \right)^{\frac{1}{2}},$$

where S is the sample domain and $\tilde{S} = \bigcup_{K \in \mathbf{K}_h(S)} K \subseteq S$ with $diam(S) = h_1$ and $diam(\tilde{S}) = \tilde{h}_1$ and h_1 and $\tilde{h}_1$ are the same order. In our case $\tilde{S} = K$ and $diam(\tilde{S}) = h$. Consequently, h and h_1 are the same order.

Remark 7.3.4. In the proof of the theorem, we keep the notation of $\tilde{S}$ despite the fact that the union of the elements in the sample element is K, i.e., $\tilde{S} = K$. We note that the proof of the theorem for the case of arbitrary number of elements per sample can be carried out following our proof with the notation $\tilde{S}$.

Proof. To estimate the second term on the r.h.s. of (7.19), we first estimate it in $\tilde{S}$, then summation over all S brings us the desired estimates. Consider a sample domain S. For the second term in $K \in \mathbf{K}_h(S)$, we have

$$\begin{aligned} |f(w_\epsilon^h - l^h) - a(u_\epsilon, w_\epsilon^h - l^h)|_K &\leq \|f\|_{0,K} \|w_\epsilon^h - l^h\|_{0,K} + C\|\nabla u_\epsilon\|_{0,K} \|\nabla(w_\epsilon^h - l^h)\|_{0,K} \\ &\leq Ch^2 \|f\|_{0,K} |w_\epsilon^h|_{2,K} + C_1 h \|\nabla u_\epsilon\|_{0,K} |w_\epsilon^h|_{2,K} \\ &\leq (Ch^2 \|f\|_{0,K} + C_1 h \|\nabla u_\epsilon\|_{0,K}) |w_\epsilon^h|_{2,K}. \end{aligned} \tag{7.20}$$

Here we have used the estimates for the linear interpolants l^h. To estimate $|w^h|_{2,K}$, notice that w^h is determined on K as a restriction of the solution of

$$\begin{aligned} \nabla_i a^{ij} \nabla_j w_\epsilon^h &= 0 \ \text{ in } \ S, \\ w_\epsilon^h &= l^h \ \text{ on } \partial S, \end{aligned}$$

where l^h is linear function whose values at the nodal points of K are the same as w_ϵ^h. Expanding $w_\epsilon^h = l^h + w_1^h$, for w_1^h we have

$$\nabla_i a^{ij} \nabla_j w_1^h = -(\nabla_i a^{ij})(\nabla_j l^h) \text{ in } S,$$
$$w_1^h = 0 \text{ on } \partial S,$$

and $|w_\epsilon^h|_{2,K} = |w_1^h|_{2,K}$.

Because of the conic points on ∂S, we cannot estimate $|w_\epsilon^h|_{2,K}$ for any subset $K \subset S$. To overcome this difficulty , we introduce vector $v = (v_1, v_2)$ such that for $k = 1, 2$,

$$\nabla_i a^{ij} \nabla_j v_k = -\nabla_i a^{ik} \text{ in } S, \tag{7.21}$$
$$v_k = 0 \text{ on } \partial S. \tag{7.22}$$

We can treat w_1^h of the same samples as linear combinations of v_1 and v_2 which are defined to be the same for each sample. We note that

$$|v|_{2,S} = \left(\sum_{k=1}^{2} |v_k|_{2,S}^2 \right)^{\frac{1}{2}} \leq C \|\nabla \cdot a\|_{0,S} \leq C \frac{h_1}{\epsilon}. \tag{7.23}$$

We get $w_1^h = (\nabla_k l^h) v_k$. Using (7.21), (7.23), the Cauchy-Schwarz inequality, and the fact that

$$\|\nabla l^h\|_{0,K} \leq \|\nabla w_\epsilon^h\|_{0,K},$$

(Appendix B), we have

$$|w_\epsilon^h|_{2,K} = |w_1^h|_{2,K} \leq |\nabla_k l^h| \ |v_k|_{2,K} \leq \|\nabla w_\epsilon^h\|_{0,K} \, |v|_{2,K}/h. \tag{7.24}$$

Now, from (7.20), using (7.23), (7.24), the fact that $\tilde{S} = K$, the triangle inequality, and the Cauchy-Schwarz inequality, we get

$$\begin{aligned} |f(w^h - l^h) - a(u, w^h - l^h)|_{\tilde{S}} &\leq \sum_{K \in \mathbf{K}_h(S)} |f(w^h - l^h) - a(u, w^h - l^h)|_K \\ \leq C \sum_{K \in \mathbf{K}_h(S)} h \|f\|_{0,K} \|\nabla w_\epsilon^h\|_{0,K} |v|_{2,K} &+ \sum_{K \in \mathbf{K}_h(S)} C \|\nabla u_\epsilon\|_{0,K} \|\nabla w_\epsilon^h\|_{0,K} |v|_{2,K} \\ \leq Ch \|f\|_{0,\tilde{S}} \|\nabla w_\epsilon^h\|_{0,\tilde{S}} \|\nabla \cdot a\|_{0,S} &+ C \|\nabla u_\epsilon\|_{0,\tilde{S}} \|\nabla w_\epsilon^h\|_{0,\tilde{S}} \|\nabla \cdot a\|_{0,S} \\ \leq Chh_1 \|f\|_{0,\tilde{S}} \|w_\epsilon^h\|_{h,\tilde{S}}/\epsilon &+ Ch_1 \|\nabla u_\epsilon\|_{0,\tilde{S}} \|\nabla w_\epsilon^h\|_{0,\tilde{S}}/\epsilon \end{aligned} \tag{7.25}$$

where we have used the fact that $\|\nabla \cdot a\|_\infty \leq C/\epsilon$ and $\text{diam}(S) = h_1$ in the last step. Summing (7.25) over all $\tilde{S} \subset \Omega$ and using the fact that h_1 and h are the same order and Cauchy-Schwarz inequality, we have

$$\begin{aligned} |f(w_\epsilon^h - l^h) - a(u, w_\epsilon^h - l^h)|_\Omega &\leq Chh_1 \|f\|_{0,\Omega} \|w_\epsilon^h\|_h/\epsilon + Ch \|\nabla u_\epsilon\|_{0,\Omega} \|w_\epsilon^h\|_h/\epsilon \\ &\leq Ch/\epsilon \|f\|_{0,\Omega}. \end{aligned} \tag{7.26}$$

For the first term on the r.h.s of (7.19), we note that l^h coincides with u_ϵ at the nodal points and $\|\nabla l^h\|_{L_2(K)} \leq \|\nabla u_\epsilon\|_{L_2(K)} \leq Ch$. This, along with the regularity estimate for (4.1), i.e., $|u_\epsilon|_{2,\Omega} \leq C\|f\|_{0,\Omega}/\epsilon$, gives

$$\begin{aligned} \|u_\epsilon - v_\epsilon^h\|_{h,\Omega} &\leq \|u_\epsilon - l^h\|_{h,\Omega} + \|v_\epsilon^h - l^h\|_{h,\Omega} \leq C\frac{h}{\epsilon}\|f\|_{0,\Omega} + \|v_\epsilon^h - l^h\|_{h,\Omega} \\ &\leq Ch(\|f\|_{0,\Omega}/\epsilon + |v_\epsilon^h|_{2,\Omega}), \end{aligned}$$

where

$$|v_\epsilon^h|_{2,\Omega} = \left(\sum_{K \in \mathbf{K}^h} |v_\epsilon^h|_{2,K}^2 \right)^{\frac{1}{2}}.$$

Note that v_ϵ^h can be handled in the same way as we did for w_ϵ^h. Thus, following the above procedure, it is straightforward to show that

$$\|u_\epsilon - v_\epsilon^h\|_{h,\Omega} \leq Ch\|f\|_{0,\Omega}/\epsilon. \tag{7.27}$$

Combining (7.26) and (7.27), we obtain the desired error the estimate for error:

$$\|u_\epsilon - u_\epsilon^h\|_{h,\Omega} \leq C\frac{h}{\epsilon} \qquad (h \ll \epsilon).$$

□

Remark 7.3.5. For the case of arbitrary number of elements per sample the estimate (7.27) also contains the term $Ch\|\nabla_y \cdot a\|_{0,\Omega}/\epsilon$ which is less than Ch/ϵ.

7.4 L_2 estimates

In this section, we will derive the L_2 norm estimates for MsFEM-os (over-sampling). Although the cell resonance has been shown in the H^1 analysis, the L_2 analysis reveals more clearly the source of this cell resonance. This enables us to design effective methods to remove it (see §7.5).

Following [28], we use a discrete error analysis to overcome the difficulty. First, we show that the L_2-norm error can be determined from the discrete l^2 norm error. For the latter, we compare the discrete solution of (4.1) with that of the homogenized equation (3.9) at the nodal points.

More specifically, let us denote by u_0^h the numerical solution of (3.9) which is calculated using the MsFEM-os method. Since the homogenized coefficients a_*^{ij} are constants, MsFEM-os reduces to the conventional finite element method with linear base functions. Then we have [28]

$$\begin{aligned} \|u_\epsilon - u_\epsilon^h\|_{L_2(\Omega)} &\leq \|u_\epsilon - u_0\|_{L_2(\Omega)} + \|u_0 - u_0^h\|_{L_2(\Omega)} + \|u_0^h - u_\epsilon^h\|_{L_2(\Omega)} \\ &\leq C\epsilon + C_1 h^2 + C_2\|u_0^h - u_\epsilon^h\|_{l_2(\Omega)}. \end{aligned} \tag{7.28}$$

Let $N \sim 1/h^2$ be the number of nodal points. Denote $\|\cdot\|$ the standard maximum norm of matrices in $\mathbf{R}^{N\times N}$ and $|\cdot|$ the maximum norm of vectors in $\mathbf{R}^N$.

The linear system of equations for U_ϵ^h is

$$A_\epsilon^h U_\epsilon^h = f_\epsilon^h, \tag{7.29}$$

where A_ϵ^h and f_ϵ^h are obtained from $a(u^h, v)$ and $f(v)$ by using $v = \phi_\epsilon^i$ for $i = 1, \ldots, N$.

Similarly, for U_0^h one has

$$A_0^h U_0^h = f_0^h, \tag{7.30}$$

where A_0^h and f_0^h are obtained by applying $v = \phi_0^i$ $(i = 1, \ldots, N)$ to $a^*(u_0^h, v) = f(v)$ with

$$a^*(u_0^h, v) = \int_\Omega a_*^{ij} v_{,i} u_{0,j}^h dx.$$

The "comma" notation is used here and below for partial derivatives. Note that U_ϵ^h is not exactly the nodal values of u_ϵ^h since the base functions ϕ_ϵ^h have different values at the same node corresponding to triangular elements with the common vertex at this nodal point. On the other hand the values of the base functions in different triangles at a common node point differ from each other by order ϵ. So we sometimes still refer to U_ϵ^h as nodal values of the numerical solution.

The main result of the section is summarized as follows:

L_2 Error Estimate: *Under Assumption A, we have*

$$|U_\epsilon^h - U_0^h| \le C_r \epsilon^2 / h^2 + C_1 \epsilon |\ln h| \qquad (h \gg \epsilon). \tag{7.31}$$

It follows from the above estimate and (7.28) that for $h \gg \epsilon$,

$$\|u_\epsilon - u_\epsilon^h\|_{L_2(\Omega)} \le C\epsilon + C_1 h^2 + C_r \epsilon^2 / h^2 + C_1 \epsilon |\ln h|.$$

The above L_2 estimate can be proved rigorously. To illustrate the main ideas more clearly, we will present the error analysis through the following steps.

Step One: Asymptotic expansion of the discrete operator

Using the asymptotic expansion of the base functions,

$$\phi_\epsilon^j = \phi_0^j + \epsilon \chi^p \nabla_p \phi_0^j + \epsilon \theta_\epsilon^j, \tag{7.32}$$

we can expand the stiffness matrix of the problem around the stiffness matrix of the ho-

mogenized problem as follows:

$$A_\epsilon^h = A_0^h + \epsilon A_1^h, \qquad f_\epsilon^h = f_0^h + \epsilon f_1^h; \tag{7.33}$$

where A_1^h and f_1^h are assembled from

$$A^e_{1_{kl}} = -\int_K \sigma^{ij}(\phi^k_{0,j}\theta^l_{,i} + \phi^l_{0,j}\theta^k_{,i})dx + \int_K \epsilon a^{ij}_\epsilon \theta^k_{,i}\theta^l_{,j}dx + \frac{1}{\epsilon}\int_K \tilde{\sigma}^{ij}\phi^l_{0,j}\phi^k_{0,i}dx, \tag{7.34}$$

and

$$f^e_{1_i} = -\int_K f(x)\chi^p\phi^i_{0,p}dx - \int_K f(x)\theta^i dx, \tag{7.35}$$

respectively, where $K \in \mathbf{K}^h$; σ^{ij} is the integrand in (3.10), i.e.,

$$\sigma^{ij} = a^{ij}(\delta_{jk} + \nabla^y_j \chi^k),$$

and $\tilde{\sigma}^{ij} = \sigma^{ij} - a^{ij}_* - \sigma^{pj}\chi^i_{,y_p}$. Note that from (3.10) and (3.11) we have $< \sigma^{ij} >= a^{ij}_*$ and $\sigma^{ij}_{,y_i} = 0$, respectively. Using integration by parts we further obtain $< \sigma^{pj}\chi^i_{,y_p} >= 0$, and hence $< \tilde{\sigma}^{ij} >= 0$. Later, we will use this property to extract a conservative structure in A_1^h operator.

Step Two: Asymptotic expansion of the discrete solution

Next we will obtain an asymptotic expansion of U_ϵ^h in powers of ϵ. This asymptotic expansion can be shown to be convergent later in this section.

$$U_\epsilon^h = U_0^h + \sum_{i\geq 1} \epsilon^i U_i^h, \tag{7.36}$$

where U_i^h $(i \geq 1)$ are given by

$$A_0^h U_1^h = f_1^h - A_1^h U_0^h; \tag{7.37}$$

$$A_0^h U_i^h = -A_1^h U_{i-1}^h \qquad (i > 1). \tag{7.38}$$

We note that the expansion enables us to avoid dealing with the inverse of A_ϵ^h whose entries

are oscillatory. In contrast, A_0^h and its inverse are well understood. Furthermore, we note that because the homogenized base functions are linear, the expansions of A_ϵ^h and f_ϵ^h have only two terms. This simplifies the equations for U_i^h $(i > 1)$ (i.e., (7.38)) and their analysis.

Step Three: Recognizing the conservative structure in A_1^h

Let $G_0^h = (A_0^h)^{-1}$. Then we have

$$U_1^h = G_0^h f_1^h - G_0^h A_1^h U_0^h, \tag{7.39}$$

$$U_i^h = -G_0^h A_1^h U_{i-1}^h \qquad (i > 1). \tag{7.40}$$

In [28], it was found that the linear system (7.39) has a conservative structure which leads to cancellation of the resonance error by using "summation by parts." Here we give a simpler proof.

Lemma 7.4.1 *Let $G_0^h = (A_0^h)^{-1}$. Then for any $V \in \mathbf{R}^N$*

$$(G_0^h A_1^h V)_i = \sum_{k=1}^{N} \sum_{j=1}^{j<k} A_1^{kj}(V^j - V^k)(G_0^{ik} - G_0^{ij}). \tag{7.41}$$

Proof. First, we note that

$$\sum_{k=1}^{3} \phi_\epsilon^k = 1, \qquad \sum_{k=1}^{3} \phi_0^k = 1,$$

for all $K \in \mathbf{K}^h$. Then it follows from (7.34) ([28]) that

$$\sum_{j=1}^{N} A_1^{ij} = 0.$$

Thus,

$$\sum_{j=1}^{N} A_1^{ij} V^j = \sum_{j=1}^{N} A_1^{ij} V^j - \sum_{j=1}^{N} A_1^{ij} V^i = \sum_{j=1,\ j\neq i}^{N} A_1^{ij}(V^j - V^i).$$

By the symmetry of A_1^{ij} (cf. (7.34)), we can combine $A_1^{kj}(V^j - V^k)$ and $A_1^{jk}(V^k - V^j)$ in

the sum of $G_0^{ik}A_1^{kj}V^j$ and get

$$\sum_{k=1}^{N} G_0^{ik} \sum_{j=1,\ j\neq k}^{N} A_1^{kj}(V^j - V^k) = \sum_{k=1}^{N}\sum_{j=1}^{j<k} A_1^{kj}(V^j - V^k)(G_0^{ik} - G_0^{ij}).$$

Thus, (7.41) follows immediately. □

We note that A_1^h is sparse. The following definitions are helpful to describe the sparsity. We say two nodal points are neighbors if they are the vertices of the same element. For a nodal point with index k, we denote the set of the indices of its neighbors by O_k (see Fig. 5.1). We have $A_1^{ij} = 0$ if $j \notin O_i$. Thus, the r.h.s. of (7.41) can be written as

$$\sum_{k=1}^{N} \sum_{j\in O_k, j<k} A_1^{kj}(V^j - V^k)(G_0^{ik} - G_0^{ij}).$$

Since the summation of j depends only on the geometry of the mesh and is independent of N (or h), we can safely ignore it when estimating the above expression. This will greatly simplifies the presentation. Introduce a difference operator D such that for any $V \in \mathbf{R}^N$, $(DV)_k \equiv V^k - V^j$ for some $j \in O_k$. Note that j is not specified in the definition. For a matrix, D applies to its row vectors unless otherwise stated. With this definition, from (7.41) and Lemma 7.4.1 we get

$$U_1^h = G_0^h f_1^h - DG_0^h A_1^h DU_0^h. \tag{7.42}$$

Here, the important features to notice are the differences on G_0^h and U_0^h; the details of these differences, however, are not important and are hidden in D. For example, for an interior nodal point k (Fig. 5.1) $DG_0^h A_1^h DU_0^h$ becomes

$$\begin{aligned} DG_0^h A_1^h DU_0^h = {} & D_{x+y}G_0^{ik} A_1^{kk-m-1} D_{x+y}U_0^k + D_y G_0^{ik} A_1^{kk-m} D_y U_0^k \\ & + D_x G_0^{ik} A_1^{kk-1} D_x U_0^k \end{aligned} \tag{7.43}$$

where $D_{x+y}G_0^{ik} = G_0^{ik-m-1} - G_0^{ik}$, $D_y G_0^{ik} = G_0^{ik-m} - G_0^{ik}$, $D_x G_0^{ik} = G_0^{ik-1} - G_0^{ik}$, and similarly for U_0^k. We would like to note that in DG_0 the difference operator is applied to the second index of G_0.

Step Four: Additional conservative structure in A_1^h and error cancellations

To further exploit the conservative structure in A_1^h, we break A_1^h into two parts, Λ_1^e and Λ_2^e, where Λ_1^e are those integrals in (7.34) containing θ_ϵ and Λ_2^e are the integral without θ_ϵ, i.e.,

$$\begin{aligned}\Lambda^e_{1_{kl}} &= -\int_K \sigma^{ij}(\phi^k_{0,j}\theta^l_{,i} + \phi^l_{0,j}\theta^k_{,i})dx + \int_K \epsilon a^{ij}_\epsilon \theta^k_{,i}\theta^l_{,j}dx \\ \Lambda^e_{2_{kl}} &= \frac{1}{\epsilon}\int_K \tilde{\sigma}^{ij}\phi^l_{0,j}\phi^k_{0,i}dx.\end{aligned} \tag{7.44}$$

Similarly, let F_1^e and F_2^e be the first and second integrals in (7.35). Moreover, let Λ_i^h be the matrix whose entries are assembled from Λ_i^e $(i = 1, 2)$, such that

$$A_1^h = \Lambda_1^h + \Lambda_2^h. \tag{7.45}$$

F_1^h and F_2^h are defined similarly.

A key observation made in [28, 26] is that the MsFEM can be improved if f_1^h and A_1^h can be written in difference forms. Such difference structures enable further error cancellation and hence reduce the error. It was found [28] that Λ_2^h can be written in difference forms. However, the boundary layer structures in θ_ϵ^j prevent Λ_1^h from having a difference structure. Therefore, the idea of over-sampling is introduced to remove the boundary layers [26]. In Appendix A, we will show Λ_2^h and f_1^h have a difference structure, i.e.,

$$\Lambda_2^h = D\lambda^h \quad \text{and} \quad f_1^h = D\hat{f}^h, \tag{7.46}$$

where λ^h is a matrix and D applies to its column vectors. Consequently, for the first term on the r.h.s. of (7.42), we have

$$|G_0^h f_1^h| = |G_0^h D\hat{f}^h| = |DG_0^h\hat{f}^h| \le \|DG_0^h\|\,|\hat{f}^h| \le C\epsilon/h + C_1. \tag{7.47}$$

It should be noted that in obtaining the second equality, we have used the summation by parts. The details in this step are not shown. We ignored the sums of the boundary terms generated by the summation by parts (analogous to the boundary integrals produced by the integration by parts). These sums are at a lower dimension and do not exceed the interior sums. Below, we use the summation by parts in the above symbolic fashion. We remark that it can be justified rigorously (cf. [28, 26]).

Using (7.49), (7.51), and (7.52) we obtain

$$|DG_0^h \Lambda_2^h DU_0^h| = |D^2 G_0^h \lambda^h DU_0^h + DG_0^h \lambda^h D^2 U_0^h| \leq C|\ln h| + C_1. \tag{7.48}$$

Step Five: Estimate for U_1^h

By the over-sampling, the error due to θ_ϵ^j is reduced to lower order. Using $\|\nabla \theta_\epsilon^j\|_{L_\infty(K)} \leq C/h$ (which follows from Assumption A) and the sparsity of A_1^h, we can show that

$$\|\Lambda_1^h\| \leq C, \quad \|\Lambda_2^h\| \leq C/h; \tag{7.49}$$

$$|F_1^h| \leq C\epsilon, \quad |F_2^h| \leq Ch. \tag{7.50}$$

Moreover, it can be shown that

$$|DU_0^h| \leq Ch, \qquad |D^2 U_0^h| \leq Ch^2; \tag{7.51}$$

$$\|G_0^h\| \leq C/h^2, \quad \|DG_0^h\| \leq C/h, \quad \|D^2 G_0^h\| \leq C \ln |h|. \tag{7.52}$$

Combining (7.48), (7.45), (7.47) and our estimates for Λ_1^h , we obtain the estimate for U_1^h

$$|\epsilon U_1^h| \leq C\epsilon |\ln h| + C_1 \epsilon. \tag{7.53}$$

Step Six: Estimate for higher order terms, U_i^h $(i > 1)$, and the cell resonance

To complete our L_2 estimate for U_ϵ^h, we need to estimate U_i^h $(i > 1)$. It turns out that the higher order terms behave differently from the leading order term, U_1^h. To illustrate this point, let us consider U_2^h. From (7.38) we have

$$U_2^h = DG_0^h A_1^h DU_1^h = DG_0^h (\Lambda_1^h + \Lambda_2^h) DU_1^h.$$

From the definition of A_1^h, it is clear that the leading order error comes from Λ_2^h. Moreover, the difference structure of Λ_2^h and summation by parts do not help reduce the error. The reason is that U_1^h, unlike U_0^h, is oscillatory and its change over a distance of h is no longer small. In fact, DU_1^h and the higher order differences $D^p U_1^h$ $(p > 1)$ are all $O(1)$. Thus

$|U_2^h| \leq 1/h^2$. Similarly, it can be shown that U_i^h $(i > 1)$ are of order $1/h^i$ or $1/h^i \ln(|h|)$. Consequently, the formal series (7.36) converges for small ϵ/h. Using estimates for A_ϵ^{-1} it can be shown that the formal series (7.36) converges to the solution.

It is evident from the derivation that the resonance error, $C_r\epsilon^2/h^2$, in (7.31) is due to Λ_2^h, which is not improved by the over-sampling. Below, we will see that the error is due to the cell resonance. Here, we would like to point out that (7.31) seems to be an overestimate for the actual computations. In our numerical experiments, we found that $\epsilon|\ln h|$ is the dominant error in most cases, even when h and ϵ are of the same order (see [26, 27] and also below). From our computational experiments it is difficult to observe the ϵ^2/h^2 error, although it does seem to have an important effect on the solutions for certain choices of h/ϵ. This subtle behavior of cell resonance is discussed further in the next section.

7.5 Cell resonance and averaged over-sampling

In this section, we study the effect of Λ_2^h, which gives the error of order $(\epsilon/h)^2$ in (7.31). To further understand Λ_2^h, we consider

$$\Lambda_{2_{kl}}^e = \frac{1}{\epsilon}\int_K \tilde{\sigma}^{ij}\phi_{0,j}^l\phi_{0,i}^k dx.$$

Because $< \tilde{\sigma}^{ij} >= 0$, applying Lemma A.0.1 to our triangulation (Fig. 5.1), we deduce that $\Lambda_2^h = 0$ if $\alpha \equiv h/\epsilon$ is an integer. It follows immediately that $|U_\epsilon^h - U_0^h| = O(\epsilon)$ and there is no resonance. In general, $\|\Lambda_2^h\|$ depends on the value of α: the norm is smaller if α is closer to an integer. Indeed, α is a measure of how good the element size matches the perfect sample sizes, and the resonance due to Λ_2^h is nothing but the cell resonance.

In the following, we first show the subtle behavior of the resonance error through numerical tests. Then we provide a partial analysis of the error followed by a numerical approach designed to remove the cell resonance.

Table 7.1: $\|U_\epsilon^h - U_0^h\|_{l_2}$. Note: $\alpha = h/\epsilon$ is irrational in the middle column.

h	$\alpha = 1.5$		$\alpha \approx 1.5396$		$\alpha = 1.5625$	
	l_2	rate	l_2	rate	l_2	rate
1/16	1.85e-4		1.75e-4		1.58e-4	
1/32	1.01e-4	0.87	9.70e-5	0.85	8.53e-5	0.89
1/64	5.60e-5	0.85	4.57e-5	1.09	4.44e-5	0.94
1/128	3.51e-5	0.68	2.45e-5	0.90	2.26e-5	0.97
1/256	2.76e-5	0.34	1.30e-5	0.91	1.14e-5	0.99
1/512	2.57e-5	0.10	6.74e-6	0.95	5.75e-6	0.99
1/1024	2.54e-5	0.02	3.31e-6	1.02	2.91e-6	0.98

7.5.1 Some observations

We first present a comparison between U_ϵ^h and U_0^h, using the same periodic test problem as in [26, 27]:

$$a_\epsilon = \frac{2 + P\sin(2\pi x/\epsilon)}{2 + P\cos(2\pi y/\epsilon)} + \frac{2 + P\sin(2\pi y/\epsilon)}{2 + P\sin(2\pi x/\epsilon)} \tag{7.54}$$

($P = 1.8$) and $f = -1$.

Table 7.1 shows the discrete l_2 norm of $U_\epsilon^h - U_0^h$ for three fixed α's. We see degrading convergence as h decreases when $\alpha = 1.5$. Moreover, the error stagnates at the bottom of the column. This result shows that the constant C_r in front of the resonance error (see (7.31)) is quite small. In this case, it is less than 6×10^{-5} (or 0.3% of the maximum value of U_0^h). For the other two α's, however, the convergence appears to be first order and C_r in both cases must be extremely small if not zero. Note that the differences in the values of α are small. Thus, C_r is sensitive to α and depends on α in a subtle way.

The fact that C_r is generically small has been demonstrated before [27, 26] by various examples including some random cases. In this section our goal is to understand the nature of C_r.

There could be several factors contributing to the subtle dependence of C_r on α. Among them we notice that the expansion of U_ϵ^h is strictly valid for $\epsilon \ll h$, while the cancellations among terms of different orders can be important when h and ϵ are close. Therefore, we study below the difference between U_ϵ^h and U_0^h directly.

Following the derivations in the previous section, we have

$$\begin{aligned} U_\epsilon^h - U_0^h &= \epsilon(DG_\epsilon^h A_1^h DU_0^h + G_\epsilon^h f_1^h) \\ &= \epsilon[DG_\epsilon^h(\Lambda_1^h + \Lambda_2^h)DU_0^h + D^2G_\epsilon^h\lambda^h DU_0^h + DG_\epsilon^h\lambda^h D^2U_0^h + DG_\epsilon^h\hat{f}^h], \end{aligned}$$

where $G_\epsilon^h = (A_\epsilon^h)^{-1}$. Note that we have used (7.46). Having obtained (7.49) and (7.50), our remaining problem is to estimate DG_ϵ^h and $D^2G_\epsilon^h$. Using the asymptotic expansion of G_ϵ^h with respect to ϵ we can show that $\|DG_\epsilon^h\| = O(1/h)$ for $\epsilon \ll h$. Our numerical tests indicate that this estimate is indeed valid even for $\alpha = h/\epsilon \sim 1$, e.g., those values in the Table 7.1. If this is the case, then we have

$$U_\epsilon^h - U_0^h = \epsilon D^2G_\epsilon^h\lambda^h DU_0^h + O(\epsilon). \tag{7.55}$$

The resonance error is contained in the first term on the right hand side. As usual, we can estimate $\|D^2G_\epsilon^h\|$, which gives the upper bound of the error. Indeed, noting that $\|\lambda^h\|$ has the same order as $\|\Lambda_2^h\|$, i.e., $1/h$, and that $|DU_0^h| \sim h$, we have

$$|U_\epsilon^h - U_0^h| \le C\epsilon\|D^2G_\epsilon^h\|. \tag{7.56}$$

It should be stressed that in this upper bound, we have ignored the subtle cancellation in $D^2G_\epsilon^h\lambda^h DU_0^h$. Therefore (7.56) tends to be an overestimate.

According to (7.56), the $|\ln h|$ behavior of $\|D^2G_\epsilon^h\|$ implies first order convergence of the solution but *not* vice versa. In fact, for $\alpha = 1.5625$ our numerical tests indicate that $\|D^2G_\epsilon^h\|$ behaves almost the same as in the case of $\alpha = 1.5$ (not shown here), however the solutions converge with first order (see Table 7.1). We attribute this delicate situation to the complicated error cancellation in $D^2G_\epsilon^h\lambda^h DU_0^h$, which depends on the coefficients a_ϵ and α. In the next subsection we investigate the behavior of $\|D^2G_\epsilon^h\|$ with respect to α. This behavior determines the regions of α for which $\|D^2G_\epsilon^h\|$ is of order $|\ln h|$. Moreover, the dependence of $\|D^2G_\epsilon^h\|$ on α might be helpful in understanding of the subtle dependence of C_r on α. The further error cancellation in $D^2G_\epsilon^h\lambda^h DU_0^h$ has not been fully understood yet. This error cancellation is difficult to analyze using the above approach of asymptotic expansion. Instead of exploring such special error cancellation, we study a much more general method for eliminating the resonance error in the section 6.

Table 7.2: Influence of magnitude of $\alpha = h/\epsilon$ on $\|D^2\mathcal{G}^h\|_{l_1}$.

h	$\alpha = 1.5$		$\alpha = 2.5$		$\alpha = 4.5$	
	l_1	increment	l_1	increment	l_1	increment
1/8	2.192		2.181		2.184	
1/16	2.724	0.532	2.661	0.480	2.643	0.459
1/32	3.478	0.754	3.240	0.579	3.137	0.494
1/64	4.749	1.271	4.091	0.851	3.746	0.609
1/128	7.154	2.405	5.583	1.492	4.658	0.912
1/256	11.887	4.733	8.443	2.860	6.285	1.627

7.5.2 The discrete Green's function and cell resonance

Now consider the behavior of the second difference of G^h. It is difficult to analyze the behavior of G^h analytically. We study it numerically and compute it for very small ϵ. We will combine analytical arguments with the numerical experiments. We use the asymptotic expansion as a guidance. After some algebra, we arrive at the following estimate

$$\|D^2 G_\epsilon^h\| \leq C|\ln h| + \frac{C_\alpha}{\alpha h}\sum_{i\geq 0}(C_\alpha|\ln h|/\alpha)^i + O(\alpha^{-1}). \tag{7.57}$$

The series is convergent when $\alpha \gg 1$.

From (7.57), we see that for fixed α if C_α/α is small, $|\ln h|$ may domintate the estimate for relatively large h. But, as $h \to 0$, the $1/h$ term will eventually dominate. This non-uniform behavior of $D^2 G_\epsilon^h$ explains the non-uniform convergence of solutions in Table 7.1 for $\alpha = 1.5$, where we see that the convergence rate gradually decreases.

Since G_ϵ^h is a dense matrix with large dimensions, it is difficult to compute. Thus, we test only one row of G_ϵ^h, denoted by $\mathcal{G}^h$. Note that $\|D^2 G_\epsilon^h\|$ is the maximum of $\|D^2\mathcal{G}^h\|_{l_1}$ over all rows. In our test, we choose $\mathcal{G}^h$ to be the one whose peak is at $(0.5, 0.5)$, the center of Ω. We use this row to show the qualitative behavior of $D^2 G_\epsilon^h$. For fixed α, the transition from $|\ln h|$ to $1/h$ in the behavior of $\|D^2\mathcal{G}^h\|_{l_1}$ as h decreases has indeed been observed numerically. This observation is consistent with (7.57). Moreover, since the fractional parts of α are the same in this test, one can deduce that C_α does not increase with α. Therefore, the $|\ln h|$ regime extends over a longer range of h for larger α.

The fractional parts of α has a strong effect on C_α. As mentioned above, $C_\alpha = 0$ for integer α. One can imagine that C_α should be small when α is close to an integer.

Table 7.3: Influence of the fractional parts of $\alpha = h/\epsilon$ on $\|D^2\mathcal{G}^h\|_{l_1}$.

h	$\alpha = 1.5$		$\alpha = 1.25$		$\alpha = 1.125$	
	l_1	increment	l_1	increment	l_1	increment
1/8	2.192		2.160		1.977	
1/16	2.724	0.532	2.642	0.482	2.677	0.700
1/32	3.478	0.754	3.192	0.550	3.177	0.500
1/64	4.749	1.271	3.956	0.764	3.790	0.613
1/128	7.154	2.405	5.221	1.265	4.704	0.914
1/256	11.887	4.733	7.563	2.342	6.295	1.591

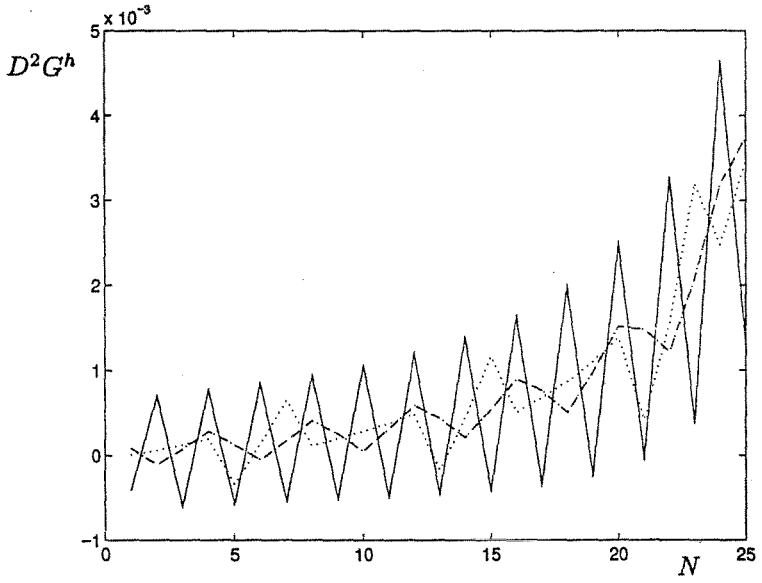

Figure 7.1: A cross-section of $D^2\mathcal{G}^h$ for $\alpha = 1.5$ (solid line), $\alpha = 1.25$ (dash line), and $\alpha = 1.125$ (dotted line). Note: only one part is shown here to avoid the singular point.

From (7.57), the $1/h$ term is less significant for smaller C_α. This has been confirmed with numerical results.

The influence of the fractional part of α on $\|D^2G_\epsilon^h\|$ can be viewed from another point of view. We observe that in general when $\alpha = 1.5$, the sign of Λ_2^{ij} (with fixed j) alternates for every other i. As a result, G_ϵ^h has small oscillations at about the same frequency. The oscillations in G_ϵ^h can be magnified after taking differences on G_ϵ^h. The frequency of oscillation is determined by the fractional part of α; the oscillation is less frequent when α is closer to an integer. As shown in Fig. 7.1, the period of oscillation, in terms of the number of nodal points, is 2 for $\alpha = 1.5$, 4 for $\alpha = 1.25$, and 8 for $\alpha = 1.125$. Moreover, $\|D^2G_\epsilon^h\|$ is smaller if G_ϵ^h oscillates less frequently. Thus, when $h \approx \epsilon$, both the magnitude of

C_α and the frequency of oscillation are important in determining the behavior of $\|D^2 G_\epsilon^h\|$, which in turn influences the convergence of solutions. We conclude that the cell resonance is weaker for α closer to an integer.

7.6 Averaged over-sampling method

To demonstrate the main idea of averaged over-sampling method we assume that the homogenized coefficients are constants and the mesh is uniform (Fig. 7.2). As we showed above the stiffness matrix of over-sampling method can be written as

$$A = A_0 + \epsilon A_1, \quad A_1 = A_1^{cell} + A_1^{\theta},$$

where A_1^{cell} and A_1^{θ} consist of the cell term and of terms containing θ of A_1 respectively. For further convenience let's introduce the following notation.

Definition. Denote $\diamond_{kl}$ the union of two triangular elements with the common side whose vertices are k and l.

For example, $\diamond_{kl}$ (Fig. 7.2) is the union of elements S_1 and S_2.

We assume that the nodal points are numerated lexicographically from 1 to N. Then the entries of A_0^{ij} $(i,j = 1,\ldots,N)$ are the same on the $2-D$ mesh along each diagonal with i being the index of interior nodal point. Indeed, it can be easily checked that $\xi^{kl} = a_*^{ij} \nabla_i \phi_0^k \nabla_j \phi_0^l$ is the same for k and l for which $\diamond_{kl}$ can be translated to each other (see example below).

For the same reason, the elements of the cell matrix A_1^{cell} have the same integrands for those k and l for which $\diamond_{kl}$ can be translated to each other, particularly the elements on the same diagonals corresponding to interior nodal points (but notice that they are integrated over different domains). For example, $\diamond_{kl}$ and $\diamond_{mn}$ (Fig. 7.2) can be translated to each other and

$$A_{cell}^{kl} = \frac{1}{h^2} \int_{S_1+S_2} \sigma^{12} dx, \qquad A_{cell}^{mn} = \frac{1}{h^2} \int_{S_3+S_4} \sigma^{12} dx.$$

Consequently, taking the arithmetical average over the diagonal entries of A^{ij} (i being the index of interior nodal point) and replacing them with the average, we obtain a new stiffness matrix with the same A_0. But by doing so we reduce ϵA_1^{cell} and get rid of resonance.

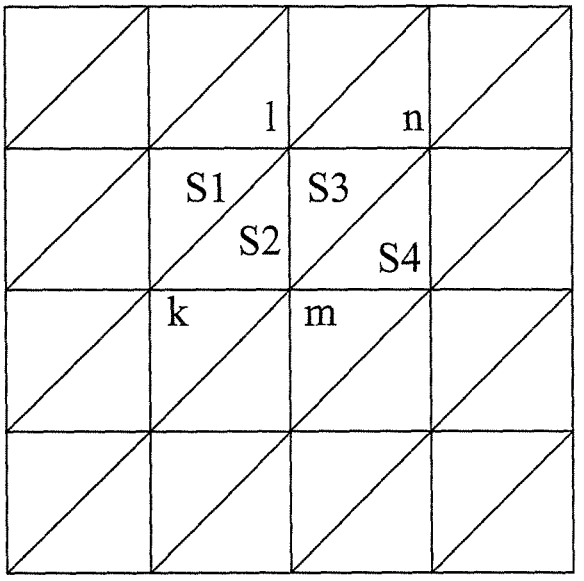

Figure 7.2: The neighboring elements in a triangulation

Indeed, the arithmetical average of the above mentioned elements of the cell matrix with the same integrand $h^{-2}\int_K \zeta(x/\epsilon)dx$ will be averaged to $\int_\Omega \zeta(x/\epsilon)dx$ where $diam(\Omega)$ is of order 1. The value of this integral is of order ϵ. It can be shown that the integrations' domains will sum up to a single connected domain with size of order 1. We remark that using these averages, we can find a_*^{ij} which is important in some applications. The impact of the averaging on A_1^θ is of order $O(h+\epsilon)$. Thus we can conclude that MsFEM with averaged over-sampling is first order for problems with periodic coefficients.

Remark 7.6.1. The error made in the homogenized coefficients of the operator contributes at the same rate as the H^1 norm of the difference between the exact homogenized solution and numerical solution using the homogenized coefficients obtained numerically.

Remark 7.6.2. It can be shown that the averaged over-sampling method converges is we can neglect the effect of θ for problems with stationary ergodic random coefficients [32, 8] independent of the ratio between h and ϵ. Indeed, averaging gives us the numerical solution which converges to the homogenized solution for a.e. realization in L_2 norm independent of the ratio between h and ϵ. Consequently, we obtain

$$E\|u_\epsilon^h - u_\epsilon\|^2_{L_2(\Omega)} \to 0$$

independent of the ratio between h and ϵ.

Part II

Upscaling of absolute permeability

Chapter 8 Upscaling of absolute permeability

8.1 Introduction

The direct numerical simulation of flows through porous formations is difficult due to the fine scale heterogeneity in the media. A common approach is to "scale up" a heterogeneous medium. For a single phase flow problem, the medium is solely described by the permeability field which can be very oscillatory. The goal of upscaling is to find an effective representation of the permeability on a coarse mesh so that the large scale flow can be correctly computed on this mesh. The computational cost is thus greatly reduced. The main result of upscaling is often the block permeability, a constant tensor computed in each grid block. Both analytical and numerical methods have been used in upscaling; see [51] and [44] for recent extensive reviews. While various formulations are proposed based on different physical and numerical considerations, we only consider flow based upscaling in this paper. The central theme is to compute the block permeability from certain averages of fine scale flow solutions. These fine scale flows can be obtained from either the global solutions of the flow equation in the whole reservoir [52] or the local solutions in each grid block. The latter approach, though, demands much less computing power. For convenience, we refer to the two approaches as *global* and *local* Laplacian, respectively.

Recently, the local Laplacian method has been extensively studied and successfully used in practical computations (see, e.g., the above review articles). However, some important issues associated with the method have not been addressed satisfactorily. For example, ideally the block permeability should only depend on the structure and the partition of the medium [30]. But in practice it has been observed that the block permeability can be strongly affected by the boundary conditions imposed for the local flow equation (cf. [15]). Furthermore, it is well-known that when the size of grid blocks is close to the scale of heterogeneity, the pressure and flow solutions computed on such a grid can have a large error. The understanding of these questions is indeed crucial for a quantification of the upscaling error.

Here we provide a rigorous analysis of the upscaling error. We assume that the under-

lying medium is periodic at small scales; however, the upscaling formulations we analyze are fully general. The periodic assumption enables us to explore the detailed structures of the local fine scale solutions using the homogenization theory [8, 32]; therefore accurate *a priori* estimates can be obtained. We show that the upscaling error appears as a *resonance* between the small physical scales of the medium and the artificial mesh scale (size). Indeed the error is given by the ratio between the two scales: it increases as the size of grid blocks gets close to the small physical scales. We also show that the effect of different boundary conditions lies in a narrow region near the boundaries of grid blocks and it contributes to the resonance error. Therefore, by using an *over-sampling* technique [26], we can remove the boundary layer effect and obtain the upscaled grid-block permeability. In this case, the scale-up accuracy is also improved. We provide numerical experiments to demonstrate these analytical findings. The upscaling of general heterogeneous media is much more difficult to analyze. However, the insights obtained through analyzing the model problem are useful in understanding the upscaling error in problems with more general random media. This is demonstrated through numerical tests.

The analysis here is motivated by the recent development of the multiscale finite element method for solving porous media flows [28, 26, 27]. In this method, the permeability is not explicitly upscaled. The focus is on the final solutions: the pressure and velocity on the coarse grids. Similar resonance error was found [28]. The problem has been further studied in more detail in the first part of the work. Although the mechanisms of resonance are the same, the effect of resonance is stronger for the upscaling methods analyzed here (see Section 8.4).

The rest of this chapter is organized as follows. In Section 8.2, we present a coherent account of upscaling formulations. Different formulations are shown to be equivalent under proper conditions. Section 8.4 starts with a brief review of the homogenization theory for media with periodic small scales. The analysis of resonance error is given subsequently. In Section 8.5 we discuss the use of over-sampling technique in upscaling. The following section is devoted to the discussion on the averaged over-sampling technique. Numerical results are given in the last two sections.

8.2 Formulations. Effective and grid block permeability

The central goal of upscaling of absolute permeability is to approximate a highly heterogeneous media with fine scale oscillations by an effective or equivalent medium with only large scale variations. Consider single phase Darcy's flows in a medium Ω:

$$u^\epsilon = -K^\epsilon(x)\nabla p^\epsilon, \qquad \nabla \cdot u^\epsilon = f, \tag{8.1}$$

where u^ϵ is the velocity, K^ϵ is the permeability tensor which is symmetric and positive definite (assuming unit viscosity), p^ϵ is the pressure, and f is the source. Note that ϵ is a small parameter indicating the length of the small scales. To solve (8.1) numerically, one need to cover Ω with a mesh consisting of finite number of grid blocks. An accurate solution is obtained if $h \ll \epsilon$ (h being the size of grid blocks). This requirement is often too restrictive for practical simulations.

The ultimate goal of upscaling is to compute solutions on a mesh with mesh size $h \geq \epsilon$. The approach considered in this paper is to replace $K^\epsilon(x)$ with the grid-block permeabilities, $\tilde{K}$, a constant tensor defined in each grid block. By definition, $\tilde{K}$ is a *discrete* quantity relying on the discretization of the medium. In particular, $\tilde{K}$ depends on the location and geometry of the grid block in which it is computed. The essential requirements for $\tilde{K}$ is that it leads to pressure and velocity solutions with desired accuracy. Moreover, one hopes that $\tilde{K}$ depends *only* on the heterogeneous permeability field and the discretization of the medium, so that it can be used in different flow scenarios once it is computed.

Different definitions of $\tilde{K}$ have been proposed (cf. [51]). Following [45], we define $\tilde{K}$ in a given grid block V such that

$$\tilde{K}\langle \nabla p^\epsilon \rangle_V = -\langle u^\epsilon \rangle_V, \tag{8.2}$$

where p^ϵ and u^ϵ are solutions of (8.1) in V (with appropriate boundary conditions, see Section 8.3.1), and

$$\langle \cdot \rangle_V = \frac{1}{V}\int_V (\cdot)\, dx$$

is the volume average over V. In 3-D, three pressure solutions are sufficient in order to

determine $\tilde{K}$ from (8.2), provided that the volume averages of the pressure gradients are linearly independent. It is evident that different sets of pressure solutions in general lead to different $\tilde{K}$s. This makes it difficult to analyze the use of $\tilde{K}$ in more general flow simulations. It seems that an analysis of upscaling error is feasible only if $\tilde{K}$ can be related to an intrinsic property of the medium. Such an intrinsic property is the effective permeability of the medium, K^*.

Physically, K^* exists when there exists an elementary representative volume (REV) such that the averages of the pressure gradient and velocity over this volume have a unique well-defined relation. An REV should be large enough to contain small scale information for a meaningful average and yet sufficiently small to reflect the large scale heterogeneity of the medium [6]. These ideas are expressed more rigorously in the homogenization theory. Mathematically, K^* is defined through the following criteria [8]: for any measurable $V \subset \Omega$,

$$\lim_{\epsilon \to 0} \langle \nabla p^\epsilon \rangle_V = \langle \nabla p \rangle_V, \quad \lim_{\epsilon \to 0} \langle u^\epsilon \rangle_V = \langle u \rangle_V, \tag{8.3}$$

where p and u are the solutions of the effective (or homogenized) equation

$$u = -K^*(x) \nabla p, \qquad \nabla \cdot u = f \tag{8.4}$$

in Ω. In (8.3), we have replaced the weak convergence of ∇p^ϵ and u^ϵ by their equivalent definitions (cf. [22]). It can further be shown that the sufficient condition for the existence of K^*, p, and u is that for any V the limits on the left-hand sides of (8.3) exist (see [32] for more rigorous details). This condition is very general; it does not assume any structure of the small scales of K^*.

The nice properties of K^* are [32]: (1) it is unique; (2) it is independent of the source term f and of the boundary condition on $\partial\Omega$; and (3) it can be determined locally, i.e., to determine K^* at a point $x \in \Omega$, one needs only to consider (8.1) in the neighborhoods of x. This last property of K^* is the foundation of all local Laplacian methods.

It is easy to see that $\tilde{K}$ is an approximation of K^*. More specifically, consider a point

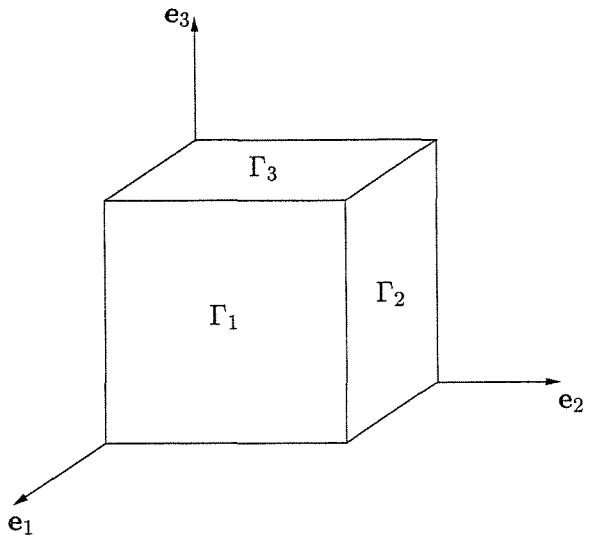

Figure 8.1: A 3D grid block.

$x \in \Omega$ and grid block $V \in \Omega$ containing x. From (8.2), (8.3) and (8.4), we have

$$\langle u^\epsilon \rangle_V \approx \lim_{\epsilon \to 0} \langle u^\epsilon \rangle_V = -\langle K^* \nabla p \rangle_V \approx -K^*(x) \nabla p,$$
$$\langle \nabla p^\epsilon \rangle_V \approx \lim_{\epsilon \to 0} \langle \nabla p^\epsilon \rangle_V = \langle \nabla p \rangle_V.$$

Thus, $\tilde{K} \approx K^*(x)$. In later sections we analyze the accuracy of this approximation in details. Then we analyze the difference between the upscaled solution and the homogenized solution of (8.4) and hence, indirectly, the difference between the upscaled solution and the fine scale solution of (8.1).

8.3 Overview

8.3.1 Local Laplacian formulations

In these methods, $\tilde{K}$ is determined from the local flow solutions in the grid blocks. The main differences among various formulations are the boundary conditions imposed on the flow equation and the averaging processes for computing $\tilde{K}$. Here we provide a unified view of the local Laplacian approach and clarify some misconceptions.

Consider a cubic grid block V of size h in a d-dimensional space (see Fig. 8.1). To

determine $\tilde{K}$ from (8.2), we need d sets of fine scale flow solutions in V, u_i^ϵ and p_i^ϵ ($i = 1, \ldots, d$), such that $\langle \nabla p_i^\epsilon \rangle_V$ are linearly independent. These fine scale solutions are solved from

$$u_i^\epsilon = -K^\epsilon \nabla p_i^\epsilon, \quad \nabla \cdot u_i^\epsilon = 0 \tag{8.5}$$

in the grid block V. Note that the source term is set to zero because of the second property of K^* mentioned above. Equation (8.5) is well posed with suitable boundary conditions on ∂V. Define $w_i^\epsilon = p_i^\epsilon - x_i$. Then

$$u_i^\epsilon = -K^\epsilon (\nabla w_i^\epsilon + e_i), \tag{8.6}$$

e_i being the unit vector in the ith direction. Thus, the linear pressure drop condition [24], the periodic boundary condition [46, 33, 15], and the widely used pressure-drop no-flow condition can then be formulated respectively as

$$w_i^\epsilon = 0 \qquad \text{on } \partial V; \tag{8.7}$$

$$w_i^\epsilon \text{ being periodic on } V; \tag{8.8}$$

$$w_i^\epsilon = 0 \text{ on } \Gamma_i, \quad n \cdot u_i^\epsilon = 0 \text{ on } \Gamma_j \ (j \neq i), \tag{8.9}$$

where Γ^i are the faces of ∂V normal to e_i. We note that (8.7) can be conveniently imposed on ∂V if V is not a rectangular box. (8.8) and (8.9) are not as flexible, but one can embed V in a larger rectangular box V' and solve (8.5) in V'. This strategy indeed has some advantages; see Section 8.5.

Conditions (8.7) and (8.8) guarantee the linear independence of the averages of pressure gradients. In fact, if either (8.7) or (8.8) holds, the Green's Theorem gives

$$\langle \nabla p_i^\epsilon \rangle_V = e_i + \frac{1}{V} \int_{\partial V} w_i^\epsilon n \, ds = e_i. \tag{8.10}$$

Thus, (8.2) can be simplified as

$$\tilde{K} e_i = -\langle u_i^\epsilon \rangle_V. \tag{8.11}$$

Another nice property of (8.7) and (8.8) is that they lead to symmetric and positive definite $\tilde{K}$. Indeed, (8.10) and (8.11) yield

$$e_i \cdot \tilde{K} e_j = -e_i \cdot \langle u_j^\epsilon \rangle_V = -\langle (\nabla p_i^\epsilon - \nabla w_i^\epsilon) \cdot u_j^\epsilon \rangle_V.$$

By (8.5), (8.7) (or (8.8)), and integration by parts,

$$\int_V \nabla w_i^\epsilon \cdot u_j^\epsilon \, dx = \int_{\partial V} w_i^\epsilon n \cdot u_j^\epsilon \, ds = 0. \tag{8.12}$$

Therefore,

$$e_i \cdot \tilde{K} e_j = \langle \nabla p_i^\epsilon \cdot K^\epsilon \nabla p_j^\epsilon \rangle_V, \tag{8.13}$$

and hence $\tilde{K}$ is symmetric and positive definite. Eq. (8.13) gives us another way to compute $\tilde{K}$. In fact, it gives symmetry up to the round-off error in numerical computations.

In contrast to (8.7) and (8.8), the condition (8.9) does not enjoy the above properties. Because of (8.9), in general (8.10) is invalid. Indeed, we have

$$\int_{\partial V} w_i^\epsilon n ds = \sum_{j \neq i} C_j e_j \neq 0,$$

C_j being some constants. Thus, $\langle \nabla p_i^\epsilon \rangle_V$ is not aligned with e_i. Consequently, (8.11) *and* (8.13) *are invalid under the boundary condition* (8.9). In this case, the original equation (8.2) should be used. Without (8.10), it is more difficult to show the linear independence of $\langle \nabla p_i^\epsilon \rangle_V$. In Section 4 using numerical examples we show that (8.9) works as well as (8.7) and (8.8).

There are many other choices of boundary conditions. Here we only consider the ones listed above because they are simple and easy to use in practice. Some other simple boundary conditions may generate singularities in the local fine scale solutions which are difficult to analyze and compute; thus they should be avoided. An example is to let $p^\epsilon = 1$ on one face of ∂V and $p^\epsilon = 0$ on the other faces.

We remark that several popular upscaling formulations can be included in the above formulation based on (8.2) and (8.5). For example, with the periodic boundary condition (8.8), the formulations in [46, 15] are included. The block permeability Kitanidis [33]

derived using the method of moments under the periodic boundary condition is identical to the symmetric part of $\tilde{K}$, which equals $\tilde{K}$ itself. Moreover, the formulation based on the conservation of dissipation [30] is equivalent to the above formulations under the conditions (8.7) and (8.8) (see Appendix C).

8.3.2 Volume vs. surface averages

In many practical upscaling computations, instead of using the volume averaged flux on the right hand side of (8.2), one often averages the velocity over the outflow surface. This approach seems appealing physically, but it may give incorrect results.

Durlofsky [15] noticed that under the periodic boundary condition (8.8), the volume average of velocity can by replaced by the averaged outflow. This is also true for the Dirichlet condition (8.7). Indeed, using (8.5), (8.13), and integration by parts we have

$$\tilde{K}_{ij} = \frac{1}{V} \int_{\partial V} p_i^\epsilon n \cdot u_j^\epsilon \, ds. \tag{8.14}$$

By using (8.7) or (8.8), it can be further reduced to

$$\tilde{K}_{ij} = \frac{1}{\Gamma_i} \int_{\Gamma_i} e_i \cdot u_j^{\epsilon+} \, ds; \tag{8.15}$$

the plus sign indicates that the flux is taken at the outflow boundary. We see that the volume averaged flow is equivalent to the averaged outflow under the condition (8.7) or (8.8) for rectangular grid blocks. Furthermore, since u_j^ϵ is divergence free, under (8.8) one can show that the averaged flux is identical over any cross-section of V parallel to Γ_i. This property, however, is not shared by (8.7).

The boundary condition (8.7) (or (8.8)) is crucial in the derivation of (8.15) from (8.2). The derivation breaks down under the pressure-drop no-flow condition (8.9). As pointed out in the literature, only when $\tilde{K}$ is diagonal can one obtain the correct result using (8.9) together with (8.15). This can also be seen from the above derivation.

A common misunderstanding is that one cannot compute the off-diagonal entries of $\tilde{K}$ if (8.9) is used. This conclusion is often deduced mistakenly by confusing the volume with the surface averaged flows. Instead, under (8.9) the right-hand side of (8.15) vanishes for $i \neq j$, but the right-hand side of (8.2) may be nonzero. A simple example is demonstrated

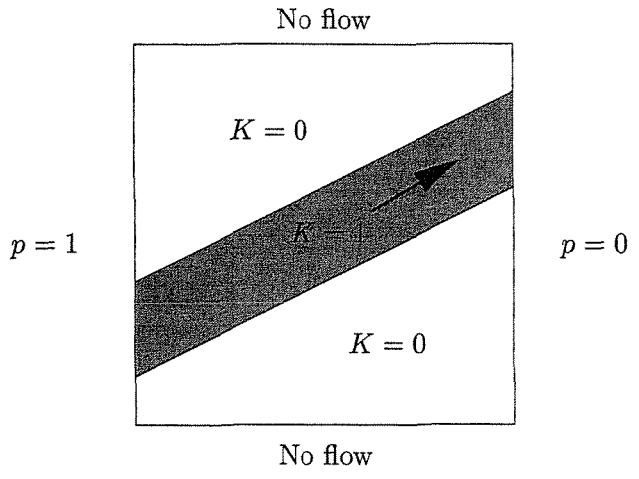

Figure 8.2: Flow through a channel. There is no flow through the top and bottom surfaces, but the volume averaged flow has a vertical component.

in Fig. 8.3.2 (see also Section 8.7.1).

8.4 Accuracy of upscaling

We now study the accuracy of the upscaling method given in the last section. We assume the medium is periodic at the small scale, i.e., K^ϵ has the form $K(x, x/\epsilon)$, where $K(x, y)$ is periodic in the y variable in a unit cube Y. For this type of media, the homogenization theory [8] makes it possible to understand the upscaling precisely. For simplicity, we assume that K is smooth in both x and y below.

In the following, unless otherwise stated, Eistein's summation convention is used, i.e., summation is taken over repeated indices; V is a cubic block of size h; C denotes a generic constant independent of h and ϵ.

8.4.1 Some estimates

With K^ϵ given above, the homogenized effective permeability is given by [8, 32]

$$K^*_{ij}(x) = \langle (e_i + \nabla_y \chi^i(x, y)) \cdot K(x, y)(e_j + \nabla_y \chi^j(x, y)) \rangle_Y, \tag{8.16}$$

where $\langle\cdot\rangle_Y = \int_Y(\cdot)dy/Y$, ∇_y is the gradient with respect to y, and χ^j is the periodic solution of

$$\nabla_y \cdot K(x,y)[\nabla_y \chi^j(x,y) + e_j] = 0 \tag{8.17}$$

in Y and satisfies $\int_Y \chi^j dy = 0$. Note that x is just a parameter in the above equations. Since $K(x,y)$ is smooth, it is easy to see that $K^*(x)$ is also smooth.

Assuming Dirichlet boundary condition, $p^\epsilon = g(x)$ (g is independent of ϵ) on $\partial\Omega$, the solution of (8.1) has the following expansion [8]:

$$p^\epsilon(x) = p(x) + \epsilon p'(x,y) + \epsilon\theta^\epsilon + u_2. \tag{8.18}$$

Here $p(x)$ is the solution of the homogenized equation (8.4) with K^* given by (8.16) and boundary condition $p = g$ on $\partial\Omega$; p' is given by

$$p'(x,y) = \frac{\partial p}{\partial x_i}(x)\chi^i(x,y),$$

and θ^ϵ satisfies

$$\nabla \cdot K^\epsilon \nabla\theta^\epsilon = 0 \quad \text{in } \Omega, \qquad \theta^\epsilon = -p' \quad \text{on } \partial\Omega. \tag{8.19}$$

We note that the main role of θ^ϵ is to correct the discrepancy in the boundary conditions between p^ϵ and its first order expansion $p + \epsilon p'$. The discrepancy is caused by the fact that p' is periodic in y and hence generally nonzero along $\partial\Omega$.

For convex polygonal domains, it can be shown that [40]

$$\|\nabla p^\epsilon - \nabla(p + \epsilon p' + \epsilon\theta^\epsilon)\|_{L_2(\Omega)} \le C\epsilon\|p\|_{H^2,\Omega}, \tag{8.20}$$

where C is a constant independent of ϵ and the size of Ω. (8.20) is equivalent to

$$\|\nabla u_2\|_{L_2(\Omega)} \le C\epsilon\|p\|_{H^2(\Omega)} \le C\epsilon.$$

In the last step we have used the fact that $\|p\|_{H^2(\Omega)}$ is bounded (this is true under the regularity assumptions on K^ϵ and g).

Remark 8.4.1. We note that in the case of $g(x) = x_i$ $(i = 1, \ldots, d)$ $u_2 = 0$.

The following estimates can be readily derived by elliptic regularity

$$\begin{aligned} \|\nabla p_i^\epsilon\|_{L_2(\Omega)} \leq Ch^{d/2}, \quad \|\nabla p_i\|_{L_2(\Omega)} \leq Ch^{d/2}, \\ \|\nabla \theta_i^\epsilon\|_{L_2(\Omega)} \leq Ch^{\frac{d-1}{2}}\epsilon^{\frac{-1}{2}}, \quad \|\theta_i^\epsilon\|_{L_2(\Omega)} \leq Ch^{\frac{d}{2}}. \end{aligned} \tag{8.21}$$

The following lemma can be proved from Lemma 7.3.2 with minor modifications

Lemma 8.4.1 *Given domain V with $diam(V) = h$, let $g(x, y)$ be a Y-periodic function in y, where Y is the unit cube. Assume $g(x, y) \in C^1(V) \cap C^0(Y)$. Then for any $f(x) \in H^1(V) \cap C^0(\bar{V})$, we have*

$$\left| \int_V f(x) g(x, x/\epsilon)\, dx - \int_V f(x) \langle g \rangle_Y(x)\, dx \right| \leq C\epsilon h^{d-1},$$

where

$$\langle g \rangle_Y(x) = \frac{1}{Y} \int_Y g(x, y)\, dy$$

is the average of g.

8.4.2 Estimates for upscaled permeability and solutions

Now consider the formulation for $\tilde{K}$ given by (8.2) and (8.5) with boundary condition (8.7). Note that in this case (8.2) is equivalent to (8.13). Below, we use (8.13) instead of (8.2) in the analysis. With minor modifications, the following derivation also holds if the periodic boundary condition (8.8) is used. All estimates derived below are for $h > \epsilon$, which is often the case in scale-up. Our main result is the following estimate for the grid-block permeability:

$$\left| \tilde{K}_{ij} - \langle K_{ij}^*(x) \rangle_V \right| \leq C_1 \frac{\epsilon}{h} + C_2 h + C_3 \epsilon. \tag{8.22}$$

Since $K^*(x)$ is smooth, $|K_{ij}^*(x) - \langle K_{ij}^* \rangle_V| = O(h)$, and (8.22) still holds if we replace $\langle K_{ij}^* \rangle_V$ by $K_{ij}^*(x)$ with $x \in V$.

To simplify the presentation, it is convenient to introduce a vector $\chi = (\chi^i)$ $(i =$

$1, \ldots, d$). Define tensor $E(x,y) = I + \nabla_y \chi(x,y)$. Eqs. (8.16) and (8.17) can be written as

$$K^*(x) = \langle E^T(x,y) K(x,y) E(x,y) \rangle_V, \tag{8.23}$$

$$\nabla_y \cdot (K(x,y)E(x,y)) = 0, \tag{8.24}$$

respectively (T indicates transpose). Moreover, by (8.18)

$$\nabla p_i^\epsilon = E \nabla p_i + \epsilon \nabla \theta_i^\epsilon + \nabla u_2, \tag{8.25}$$

where p_i is the homogenized solution given by

$$\nabla \cdot K^*(x) \nabla p_i = 0 \quad \text{in } V, \qquad p_i = x_i \quad \text{on } \partial V; \tag{8.26}$$

and θ_i^ϵ satisfies

$$\nabla \cdot K^\epsilon \nabla \theta_i^\epsilon = 0 \quad \text{in } V, \qquad \theta_i^\epsilon = -\chi \cdot \nabla p_i \quad \text{on } \partial V. \tag{8.27}$$

From (8.13), (8.25) and (8.20) we have

$$\begin{aligned} \tilde{K}_{ij} = \langle (E\nabla p_i + \nabla \theta_i^\epsilon) \cdot K^\epsilon \nabla p_j^\epsilon \rangle_V + \langle \nabla u_2^i K^\epsilon \nabla p_j^\epsilon \rangle_V = \\ \langle \nabla u_2^i K^\epsilon (E\nabla p_j + \nabla \theta_j^\epsilon) \rangle_V + \langle \nabla u_2^i K^\epsilon \nabla u_2^j \rangle_V = \\ \langle (E\nabla p_i + \nabla \theta_i^\epsilon) \cdot K^\epsilon (E\nabla p_j + \nabla \theta_j^\epsilon) \rangle_V + O(\epsilon). \end{aligned} \tag{8.28}$$

Expanding the integrand of (8.28) and using the symmetry of K^ϵ, we get

$$\begin{aligned} \tilde{K}_{ij} = \langle (E\nabla p_i) \cdot K^\epsilon (E\nabla p_j) \rangle_V + 2\epsilon \langle (\nabla \theta_i^\epsilon) \cdot K^\epsilon (E\nabla p_j) \rangle_V \\ + \epsilon^2 \langle (\nabla \theta_i^\epsilon) \cdot K^\epsilon \nabla \theta_j^\epsilon \rangle_V + O(\epsilon). \end{aligned} \tag{8.29}$$

We estimate the two terms with θ_i^ϵ first. Due to the fact that $\|\nabla \theta_i^\epsilon\|_V \leq C h^{\frac{d-1}{2}} \epsilon^{-\frac{1}{2}}$ (8.21) and that the volume of V is $O(h^d)$, the last term is of order ϵ/h. For the second term, integration by parts gives

$$\int_V (\nabla \theta_i^\epsilon) \cdot K^\epsilon (E\nabla p_j) = \int_{\partial V} \theta_i^\epsilon n \cdot (K^\epsilon E \nabla p_j)\, dS - \int_V \theta_i^\epsilon \nabla \cdot (K^\epsilon E \nabla p_j)\, dx.$$

Since ∇p_j and E are bounded on ∂V, the surface integral is bounded by $C\epsilon h^{d-1}$. For the volume integral on the right-hand side, noting that $\nabla = \nabla_x + (1/\epsilon)\nabla_y$ and using (8.24), we have

$$\nabla \cdot (K^\epsilon E \nabla p_i) = \nabla_x \cdot (K(x,y)E(x,y))\nabla p_i + (K^\epsilon E) : (\nabla \nabla p_i)$$

with ':' indicating double contraction between tensors. By the estimates of $\|\theta_i^\epsilon\|_V$, $\|\nabla p_i\|_V$, and $\|p_i\|_{2,V}$, and the fact that K and E are smooth functions in x, the volume integral is $O(\epsilon h^d)$. It follows that

$$\left|\tilde{K}_{ij} - \langle (E\nabla p_i) \cdot K^\epsilon (E \nabla p_j)\rangle_V\right| \leq C_1 \frac{\epsilon}{h} + C_2 \epsilon. \tag{8.30}$$

Now, consider the first term in (8.29). Because

$$\langle (E\nabla p_i) \cdot K^\epsilon (E \nabla p_j)\rangle_V = \langle \nabla p_i \cdot (E^T K^\epsilon E) \nabla p_j \rangle_V,$$

from (8.23) and Lemma 8.4.1 we get

$$\langle (E\nabla p_i) \cdot K^\epsilon (E \nabla p_j)\rangle_V = \langle (\nabla p_i) \cdot K^* (\nabla p_j)\rangle_V + O(\frac{\epsilon}{h}). \tag{8.31}$$

Using (8.26) and integration by parts, we further have shown that

$$\langle (\nabla p_i) \cdot K^* (\nabla p_j)\rangle_V = \langle K_{ij}^* \rangle_V - \langle w_i \nabla \cdot (K_{ij}^* e_j)\rangle_V,$$

where $w_i = p_i - x_i$. Thus, it follows from the Cauchy-Schwartz inequality and (8.21) that

$$\begin{aligned} \left|\langle (\nabla p_i) \cdot K^* (\nabla p_j)\rangle_V - \langle K_{ij}^* \rangle_V\right| &\leq C\frac{1}{V}\|w_i\|_V \, \|\nabla \cdot (K^* e_j)\|_V \\ &\leq Ch. \end{aligned} \tag{8.32}$$

Therefore, (8.22) follows immediately from (8.30), (8.31), and (8.32). It should be noted that in the above derivation, we only require V to be a convex block with reasonable aspect ratios. Thus (3.10) applies to general unstructured meshes.

Using (8.22) we can derive the estimates for upscaled pressure and velocity solutions. Suppose $\tilde{K}$ is computed on all grid blocks. Then we have a piecewise constant upscaled

permeability field, which we still denote by $\tilde{K}$. Let $\tilde{p}$, $\tilde{u}$ be the solutions of the upscaled flow equation

$$\tilde{u} = \tilde{K}\nabla\tilde{p}, \qquad \nabla \cdot \tilde{u} = f \tag{8.33}$$

in Ω. For simplicity, we assume p and $\tilde{p}$ equal zero on $\partial\Omega$. From (8.33) and (8.4) we have

$$\nabla \cdot \tilde{K}(\nabla p - \nabla\tilde{p}) = \nabla \cdot (\tilde{K} - K^*)\nabla p.$$

Since $\tilde{K}$ is positive definite, multiplying both sides by $\nabla p - \nabla\tilde{p}$ and integrating by parts yield

$$\|\nabla p - \nabla\tilde{p}\|_\Omega \leq C\|(\tilde{K} - K^*)\nabla p\|_\Omega \leq C_1\frac{\epsilon}{h} + C_2 h + C_3\epsilon. \tag{8.34}$$

For the last inequality, we have used (8.22) and the fact that $\|\nabla p\|_\Omega$ is bounded. As a result, we have

$$\begin{aligned} \|u - \tilde{u}\|_\Omega &\leq \|\tilde{K}(\nabla p - \nabla\tilde{p})\|_\Omega + \|(K^* - \tilde{K})\nabla p\|_\Omega \\ &\leq C_1\frac{\epsilon}{h} + C_2 h + C_3\epsilon. \end{aligned} \tag{8.35}$$

Moreover, from (8.34) and the Poincaré inequality we get

$$\|p - \tilde{p}\|_\Omega \leq C_1\frac{\epsilon}{h} + C_2 h + C_3\epsilon. \tag{8.36}$$

Note that $\|p - p^\epsilon\|_\Omega \leq C\epsilon$ (cf. [40]). Thus by the triangle inequality

$$\|p^\epsilon - \tilde{p}\|_\Omega \leq C_1\frac{\epsilon}{h} + C_2 h + C_3\epsilon. \tag{8.37}$$

Unlike the pressure solutions, the homogenized velocity, u, does not approximate u^ϵ in the L_2 norm. Their L_2 norm difference is in fact $O(1)$. Similarly, $\tilde{u}$ does not approximate u^ϵ in the L_2 norm. In fact, we roughly have $\langle u^\epsilon - u\rangle_V = O(\epsilon/h)$ for $V \subset \Omega$, which is consistent with (8.3). This estimate holds also for $\langle u^\epsilon - \tilde{u}\rangle_V$. Thus, the velocity solution of the upscaled equation approximates the volume average of the fine scale velocity in the grid-blocks as $\epsilon \to 0$. In comparison, the velocity computed from MsFEM approximates

u^ϵ [28]. We note that the small scale fluctuations of u^ϵ are often important in simulating transport phenomena in multiphase flows (see e.g., [27]). Moreover, it has been observed the magnitude of the error using MsFEM is smaller than the magnitude of the error using upscaling methods described above.

8.4.3 Remarks

Several observations can be made from (8.22). First, (8.22) implies that we can obtain K^* at a given point by computing the limit of $\tilde{K}$ as $\epsilon \to 0$ in a series of shrinking blocks containing the point. This is consistent with the general argument in Section 2.1. Moreover, if K^* is constant, the derivation of (8.32) indicates that the $O(h)$ error in (8.22) vanishes. An example is the case of pure periodic media described by $K^\epsilon(x) = K(x/\epsilon)$. In general, however, K^* contains large scale heterogeneity; (8.22) shows that the size of grid-blocks should be small in comparison. This explains the observation that local upscaling is usually preferred against global upscaling (cf. [16]).

As noted above, (8.22) is valid for $h > \epsilon$. If $h < \epsilon$, a different derivation using Taylor expansion of the pressure field would show that $|\tilde{K}_{ij} - \langle K^\epsilon_{ij} \rangle_V|$ is $O(h/\epsilon)$. Thus, for fixed ϵ, $\tilde{K} \to K^\epsilon$ as $h \to 0$. Thus, the "upscaled" solution approximates the fine scale solution for $h \ll \epsilon$.

The $h \sim \epsilon$ case deserves further examination. It is seen that the error increases as h approaches ϵ. This scale "resonance" phenomenon is fundamental in upscaling. The derivation of (8.22) reveals two sources of the resonance effect. One is due to the first order corrector θ^ϵ_i , whose main role is to enforce the boundary conditions of p^ϵ_i on ∂V; the other comes from the volume average of functions with ϵ-periodic small scales (see (8.31)).

The scale resonance is a direct consequence of the objective of upscaling, namely dividing a globally coupled fine scale problem into many *decoupled* local problems in the coarse grid blocks. The decoupling is achieved by *artificial* local boundary conditions imposed on ∂V, such as (8.7). As pointed out in [28], the optimum boundary condition would be the ones consistent with the fine scale oscillations of the differential operator, which are solely determined by $\chi(x, x/\epsilon)$ for the model problem. However, boundary conditions (8.7) and (8.9) are inconsistent with χ since they enforce non-oscillatory Dirichlet conditions on ∂V or part of it. The periodic condition (8.8), on the other hand, allows oscillations on ∂V; however, the oscillations match those of χ only when h coincides with multiples of ϵ. Any

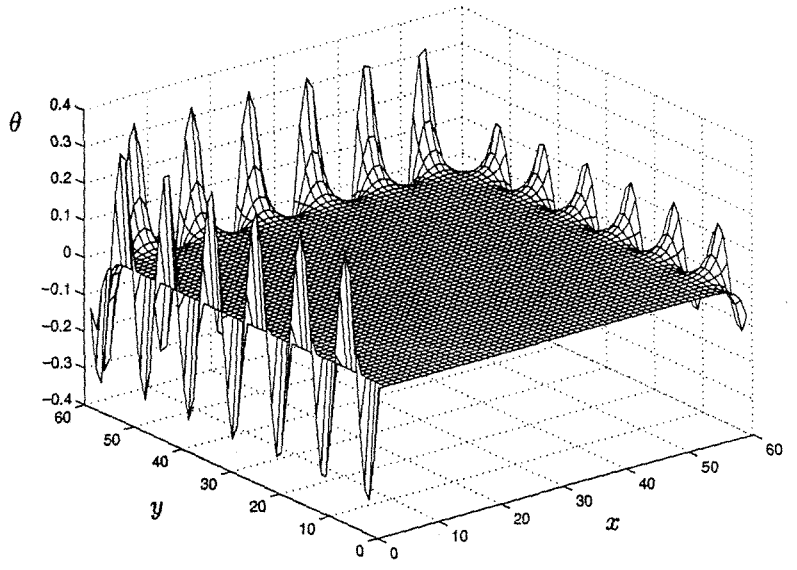

Figure 8.3: The cross-section of θ

mismatch at ∂V is to be corrected by θ_i^ϵ. It can be shown that the correction occurs in a thin layer with a thickness about $O(\epsilon)$ near ∂V [7]. This boundary layer yields large gradient of θ_i^ϵ near ∂V, which is the main cause of the $O(\epsilon/h)$ error in (8.30) (in Fig. 8.3 we plot the cross-section of θ for a periodic oscillatory boundary condition). In the interior of V, however, $\nabla\theta_i^\epsilon$ is much smaller.

The accuracy of upscaling is also strongly influenced by the size of the sample based on which the average is taken. The sample size is given by the size of the grid blocks. Intuitively, for media with ϵ-periodic small scale, the perfect sample sizes are multiples of the period ϵ. Otherwise, error occurs due to the mismatch between the block size and the perfect sample size. Mathematically, the mismatch gives rise to $O(\epsilon h^{d-1})$ bound in Lemma 8.4.1. This type of resonance is referred to as the "cell resonance". Eq. (8.32) shows that the cell resonance error is $O(\epsilon/h)$ for the present upscaling formulation.

8.5 Over-sampling method

It is possible to employ the over-sampling idea in upscaling of absolute permeability. The over-sampling technique has been developed in previous section to remove the resonance due to the boundary layer of θ_i^ϵ. It is shown to be very effective when used together with the multiscale finite element method. This technique can be borrowed here to improve the

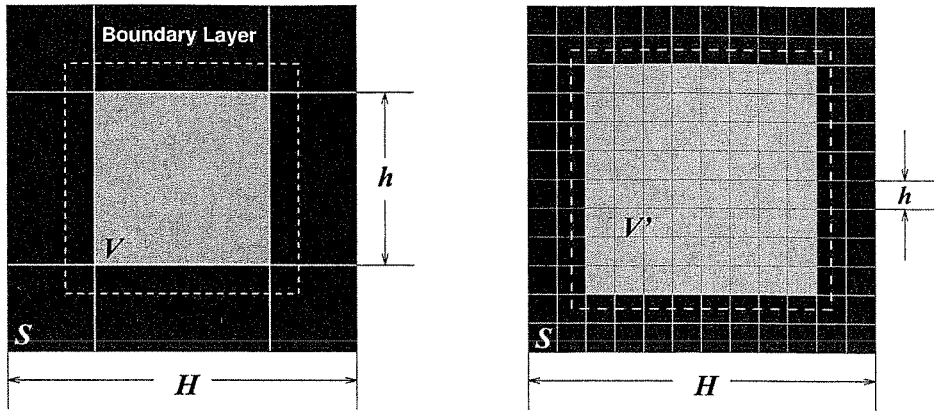

Figure 8.4: Over-sampling for one (left) and multiple (right) grid blocks. The boundary layer is outside the white dash line. V' denotes the union of the set of grid blocks in gray region.

upscaling of permeability. The idea is to use a larger sampling block $S \supset V$, such that the distance between ∂S and V is at least $O(\epsilon)$. Then we solve (8.5) in S and compute $\tilde{K}$ using (8.2) in V. Since V is away from ∂S and is not "polluted" by the boundary layer of θ_i^ϵ, the resulting $\tilde{K}$ is free from the resonance error due to θ_i^ϵ. The method is shown schematically in Fig. 8.4 (left). We further notice that the boundary conditions given in Section 2.2 may give different θ_i^ϵ. However, the main difference among these θ_i^ϵs lie in the boundary layer region. Thus, by removing the boundary layers, different boundary conditions give rise to almost identical $\tilde{K}$. In other words, $\tilde{K}$ depends mainly on K^ϵ and the partition of the domain.

We note that S can be as large as Ω. In fact, choosing large S that contains many grid blocks has the advantage of greatly reducing redundant calculations in the overlapping region near ∂S [26] (see the right figure in Fig. 8.4). In practice, however, it is difficult to use Ω as the sampling domain since the computation would be too expensive. Therefore, one may choose S as large as possible according to the computing resource. In a parallel implementation, S may be chosen so that the computation on S fits into the memory of each processor. In this way, the inter-processor communication is minimum.

Interestingly, in an attempt to reduce the effect of local boundary conditions, White and Horne [52] used the entire domain to sample fine scale solutions with different boundary conditions. The over-determined system of equation for $\tilde{K}$ was solved by using the

least square method. Inspired by their work, Gómez-Hernández [24] proposed a method of Laplacian with skin, where the width of the skin is arbitrarily set to half of the grid block size. These methods can now be unified and understood in the over-sampling framework.

It should be noted that the analysis in the previous section needs to be modified for the over-sampling method. Due to over-sampling, the equivalence between (8.2) and (8.13) no longer holds. Thus, one needs to return to (8.2) for the computation of $\tilde{K}$ and for the analysis of upscaling error. The basic ingredients of the analysis are the same as those given above. The main difference lies in θ_i^ϵ, whose gradient is bounded independent of ϵ , in contrast to the estimate of $\|\nabla\theta_i^\epsilon\|$ in (8.21). On the other hand, the over-sampling is not helpful for reducing the cell resonance, which is governed by the grid block size. Therefore, the estimate (8.22) remains true even with the over-sampling. Nevertheless, we demonstrate below that numerically the error due to the cell resonance is small compared to that due to θ_i^ϵ. Thus, the over-sampling is still very effective in reducing the resonance error.

In comparison, the cell resonance error in the multiscale finite element method with over-sampling is $O(\epsilon^2/h^2)$ because of some additional error cancellation.

To derive the error estimate for upscaling method with over-sampling we make an assumption in order to neglect the effect of θ_i^ϵ. As we have seen from the previous chapter the effect of theta is small when we use the over-sampling technique. This assumption has been made in the chapter of Nonconforming MsFEM, $\nabla\eta^p \in L_\infty(K)$ when $K \subset S$ is away from ∂S at least at a distance ϵ, where η^p is defined in (3.16). Then in (8.29)

$$\tilde{K}_{ij} = \langle (E\nabla p_i)\cdot K^\epsilon(E\nabla p_j)\rangle_V + 2\epsilon\langle(\nabla\theta_i^\epsilon)\cdot K^\epsilon(E\nabla p_j)\rangle_V + \epsilon^2\langle(\nabla\theta_i^\epsilon)\cdot K^\epsilon\nabla\theta_j^\epsilon\rangle_V + O(\epsilon). \tag{8.38}$$

In (8.38) we can neglect the terms containing θ^ϵ. Consequently, taking into account the Lemma 8.4.1 from (8.38) we have

$$|\tilde{K}_{ij} - \langle K_{ij}^*\rangle_V| = |\langle (E\nabla p_i)\cdot K^\epsilon(E\nabla p_j) - K_{ij}^*(x)\rangle_V| + C\epsilon \le C\frac{\epsilon}{h} + C_1 h. \tag{8.39}$$

As we see the cause of the resonance error is the mismatch between volume V and the period of the problem. When volume V is an integer union of periods of the problem, then

the resonance error is absent in above formulation. We call this resonance error "the cell resonance."

Remark 8.5.1. In general random cases if we neglect the effects from θ^ϵ terms, the cell resonance is the dominating error. We can show that in this case this error is $o(\epsilon/h)$ (i.e., it goes to zero as $\epsilon/h \to 0$. This shows the existence of the resonance between two intrinsic scales ϵ and h.

8.6 Averaged over-sampling in upscaling of absolute permeability

Averaged over-sampling idea introduced in section 7.6 can be applied in upscaling of absolute permeability in order to reduce the effect of the cell resonance. It can get rid of the cell resonance when K^*_{ij} are constants throughout the entire domain Ω.

To demonstrate the main idea of the method we denote by $\tilde{K}^V_{ij}$, $(i,j = 1,\ldots,d)$, the local upscaled permeability in V calculated using over-sampling method. Averaged over-sampling method consists of the arithmetic averaging $\tilde{K}^V_{ij}$ over all $V \subset \Omega$. Assuming K^*_{ij} are constants in Ω from (8.38) we have

$$Ch^d \sum_{V\subset\Omega} \tilde{K}^V_{ij} - K^*_{ij} = Ch^d \sum_{V\subset\Omega} \langle (E\nabla p_i)\cdot(K^\epsilon - K^*_{ij})(E\nabla p_j)\rangle_V + C\epsilon + C_1 h \qquad (8.40)$$

where Ch^d in front of $\sum$ indicates the number of V's in Ω. Furthermore we assume that ∇p_i are the same for all $V \subset \Omega$. This is true, for example, if the elements V's are the same. In the typical situation when V's are square elements and x_i boundary conditions have been used for p^ϵ_i, all of ∇p_i are equal constants. Consequently, denoting $\tilde{K}^{av}_{ij}$ absolute permeability obtained using averaged over-sampling, we have

$$\tilde{K}^{av}_{ij} - K^*_{ij} = C\int_\Omega (E\nabla p_i)\cdot(K^\epsilon - K^*_{ij})(E\nabla p_j) + C\epsilon + C_1 h \le C\epsilon + C_1 h. \qquad (8.41)$$

We note that we also assume that the size of Ω is of order 1.

In the case when K^*_{ij} are not constants throughout the domain we can apply the averaged over-sampling method for absolute permeability in each local domain V. If the size of V is of order h, the error estimate of averaged over-sampling method is $C\epsilon/h + C_1 h$. $C_1 h$ error

comes from approximation of $K^*_{ij}(x)$ by constants in domain V. As we see this error rate can be obtained just using over-sampling in domains V's without further partitioning V for averaged over-sampling. So the question is whether we need to partition local domains V further for averaged over-sampling or just use the over-sampling method without further partitioning. The answer of this question depends on available computer resources and on ϵ and h. If the available computer resources is limited then the further partitioning of the local domains is advantageous if ϵ is too small and Ω (the whole domain) is of size 1.

8.7 Numerical results

8.7.1 Numerical results for 2-D

In this section, we provide some numerical results demonstrating the estimate (8.22) and the effect of over-sampling. For this purpose, we use a periodic K^ϵ without large scale heterogeneity. Thus, the $O(h)$ error in (8.22) vanishes, and we can focus on the resonance error. Tests with more general random permeability field are given in §8.7.1.

Periodic case

In the following, we compute the upscaled permeability $\tilde{K}$ on one grid block $V = (0,1)^2$ from

$$K^\epsilon(x,y) = 1/[2 + P\sin(2\pi(2x-y)/\epsilon)].$$

We fix $P = 1.8$. The exact effective permeability can be calculated analytically

$$K^*_{11} = 0.62942, \quad K^*_{12} = K^*_{21} = 0.25883, \quad K^*_{22} = 1.01766.$$

We solve (8.5) with boundary conditions (8.7), (8.8), and (8.9) on uniform square grids using a Galerkin finite element method with bilinear base functions. $\tilde{K}$ is computed from (8.2). For convenience, we denote $\tilde{K}^i$ $(i = 1,2,3)$ to be the numerical results obtained by using (8.7), (8.8), and (8.9), respectively.

First, we compute $\tilde{K}$ with $\epsilon = 1$. In this case, the periodic condition (8.8) gives $\tilde{K} \equiv K^*$. The only error in this case is the discretization error. However, according to (8.22), using (8.7) and (8.9) gives rise to the resonance error. The error of $\tilde{K}^1$ and $\tilde{K}^3$ compared to K^*

Table 8.1: Convergence of the numerical solution of $\tilde{K}^1$ to K^* ($\epsilon/h = 1$).

N	$\lvert\tilde{K}^1_{11} - K^*_{11}\rvert$	$\lvert\tilde{K}^1_{12} - K^*_{12}\rvert$	$\lvert\tilde{K}^1_{21} - K^*_{21}\rvert$	$\lvert\tilde{K}^1_{22} - K^*_{22}\rvert$
16	1.021e-1	1.229e-1	1.664e-1	2.095e-2
32	9.307e-2	1.130e-1	1.229e-1	1.564e-2
64	9.103e-2	1.107e-1	1.132e-1	1.439e-2
128	9.055e-2	1.101e-1	1.107e-1	1.408e-2

Table 8.2: Convergence of the numerical solution of $\tilde{K}^3$ to K^* ($\epsilon/h = 1$).

N	$\lvert\tilde{K}^3_{11} - K^*_{11}\rvert$	$\lvert\tilde{K}^3_{12} - K^*_{12}\rvert$	$\lvert\tilde{K}^3_{21} - K^*_{21}\rvert$	$\lvert\tilde{K}^3_{22} - K^*_{22}\rvert$
16	1.612e-2	2.001e-1	2.528e-1	5.278e-2
32	2.001e-2	1.961e-1	2.338e-1	5.615e-2
64	2.080e-2	1.951e-1	2.283e-1	5.690e-2
128	2.097e-2	1.949e-1	2.269e-1	5.708e-2

are presented in the next two tables (table 8.1 and 8.2) , where N is the number of elements in the x and y directions. From the tables we see that the error in both cases does not converge to zero as grid refines. Evidently, the resonance error is dominating in these cases (note that $\epsilon/h = 1$); the error in the off-diagonal terms is especially large.

As shown by (8.22), one way of reducing the resonance error is to reduce to ratio ϵ/h. This is demonstrated in Table 8.3 for $\tilde{K}^1$. The result for $\tilde{K}^3$ is similar. The errors in $\tilde{K}^1_{11}$ and $\tilde{K}^1_{12}$ decrease almost in the order of ϵ/h, but the errors in $\tilde{K}^1_{21}$ and $\tilde{K}^1_{22}$ decrease more slowly. Note that the discretization error is fixed in the test because $N\epsilon$ is kept constant; hence, the error reduction is mainly due to the decrease of the resonance error (cf. (8.22)). Thus, faster error reduction can be seen for a component of $\tilde{K}$ with more dominant resonance error. Numerically, we find that the resonance error is indeed much larger than the discretization error for $\tilde{K}^1_{11}$ and $\tilde{K}^1_{12}$ but not so for $\tilde{K}^1_{21}$ and $\tilde{K}^1_{22}$.

The next table shows the improvement of resonance error due to over-sampling. The over-sampling method depicted in Fig. 8.4 is implemented. For convenience, we denote the distance between ∂S and V or V' by d_s. We take $\epsilon = 0.8$ in the tests. In this case, the

Table 8.3: Variation of $\lvert\tilde{K}^1 - K^*\rvert$ versus ϵ/h ($h = 1$).

N	ϵ	$\lvert\tilde{K}^1_{11} - K^*_{11}\rvert$	$\lvert\tilde{K}^1_{12} - K^*_{12}\rvert$	$\lvert\tilde{K}^1_{21} - K^*_{21}\rvert$	$\lvert\tilde{K}^1_{22} - K^*_{22}\rvert$
16	1.0	1.021e-1	1.229e-1	1.664e-1	2.095e-2
32	0.5	5.249e-2	6.358e-2	1.021e-1	1.297e-2
64	0.25	2.691e-2	3.299e-2	6.886e-2	8.843e-3
128	0.125	1.397e-2	1.752e-2	5.205e-2	6.754e-3
256	0.0625	7.474e-3	9.751e-3	4.360e-2	5.704e-3

Table 8.4: Resonance error reduction by over-sampling (N is the total number of elements in the x and y directions in S).

	$\lvert\tilde{K}_{12} - K^*_{12}\rvert$			$\lvert\tilde{K}_{22} - K^*_{22}\rvert$		
N	BC (8.7)	BC (8.8)	BC (8.9)	BC (8.7)	BC (8.8)	BC (8.9)
64	4.751e-2	4.748e-2	4.744e-2	9.953e-3	9.950e-3	1.018e-2
128	1.445e-2	1.443e-2	1.424e-2	1.952e-3	1.949e-3	2.203e-3
256	3.745e-3	3.725e-3	3.523e-3	4.771e-4	4.745e-4	7.351e-4

over-sampling removes much but not all of the resonance error, because there is the cell resonance. We choose $S = (0,4)^2$ and $V = (1,3)^2$ at the center of S. Thus, $d_s = 1 > \epsilon$. The errors in $\tilde{K}_{21}$ and $\tilde{K}_{22}$ are reported as they are larger than the error of the other two components. The results of using different upscaling boundary conditions appear to be quite similar, indicating that the influence of boundary conditions are small due to over-sampling. We note that in this particular test, the cell resonance is small as indicated by the decrease of error. Moreover, we note that (8.9) does give the correct off-diagonal terms of $\tilde{K}$ if it is used correctly; otherwise, even the diagonal entries of $\tilde{K}$ may be wrong.

Random cases

Here, we present two tests of upscaling randomly generated permeability fields. The random field generator is based on the superposition of random modes in Fourier domain and the fast Fourier transform to give K^ϵ in the physical domain. The details of the generator has been described in [26]. The purpose of these tests is to show that the difference in the upscaled permeability due to different upscaling boundary conditions can be effectively reduced by using the over-sampling. Thus, $\tilde{K}$ depends solely on the geometry of the grid block and the underlying fine scale permeability.

In these numerical tests, we compute the relative differences of $\tilde{K}$s. More specifically, we compute $D^i_{jk} = \|\tilde{K}^i_{jk} - \tilde{K}^2_{jk}\|/\|\tilde{K}^2_{jk}\|$ $(i = 1,3)$, where the norm is the discrete max or l_2 norm over Ω. The relative error of the solution of $\tilde{p}$ is also checked against p^ϵ. For computing the pressure, we assume uniform injection $f = 1$ and $p = 0$ on $\partial\Omega$. The reference solution of p^ϵ is calculated from two fine-grid solutions using the Richardson extrapolation. The pressure error is denoted by E^i_p $(i = 1,2,3)$, where i means the same as in $\tilde{K}^i$. For all tests below, uniform square mesh is used. Moreover, $\Omega = [0,1]^2$, the permeability fields are generated on the 1024×1024 mesh, and the solutions of $\tilde{K}$ and $\tilde{p}$ are computed on the

Table 8.5: Test of over-sampling using log-normal permeability Difference and error are shown in percentage.

	No Over-sampling		Over-sampling ($d_s = h$)	
	Max norm	l_2 norm	Max norm	l_2 norm
D_{11}^1	22.8125	17.2754	1.2233	0.5775
D_{11}^3	6.0135	2.8384	2.0317	0.5072
D_{22}^1	50.5283	18.3060	1.6727	0.6139
D_{22}^3	9.7429	2.9306	1.4441	0.4937
E_p^1	13.0553	12.7723	3.6848	1.6446
E_p^2	4.7253	2.2043	3.6459	1.6203
E_p^3	4.2316	2.0360	3.6182	1.6198

32×32 coarse mesh. A 16×16 subcell mesh is used for upscaling of permeability.

In Case 1, a realization of a log-normal permeability field is generated. The highest wave number of the random modes used in generating the field is 32. The permeability field is in fact smooth and can be resolved by the 512×512 finite element mesh. It is scaled so that its contrast (i.e., $K^\epsilon_{max}/K^\epsilon_{min}$) is 400. In this case, the size of the grid blocks, $h = 1/32$, is about the correlation length of the permeability field (which often happens in practical simulations). Note that the correlation length is analogous to ϵ in the periodic case and we choose $d_s = h$ for the over-sampling. As shown by Table 8.5, our choice of d_s appears to be sufficiently wide for eliminating most of the effect of boundary conditions. Using smaller d_s, e.g., $h/2$, gives more error in $\tilde{p}^1$ and larger difference between $\tilde{K}^1$ and $\tilde{K}^2$. In Table 8.5, only the diagonal entries of $\tilde{K}$ are compared, because the off-diagonal entries are two orders of magnitude smaller than the diagonal ones and have much less effect on the pressure solution. Nevertheless, our numerical tests show that over-sampling has similar effect on the off-diagonal entries.

In some sense, the above problem is similar to the periodic problem because there is a distinctive small scale in permeability which is characterized by the correlation length. In practice, permeability distributions often exhibit multiple scales or no intrinsic small scale. In the next test, we use a permeability field whose logarithm has a fractal dimension of 2.8. The image of the field is shown in Fig. 8.5. In this case, how to choose d_s is not obvious; we determine it through the numerical experiment. The results for D^i and E_p^i are shown in Table 8.6. Two sets of over-sampling results are presented. We see that by using $d_s = 2h$, the difference between $\tilde{K}^1$ and $\tilde{K}^2$ and the error in the pressure solutions are reduced to

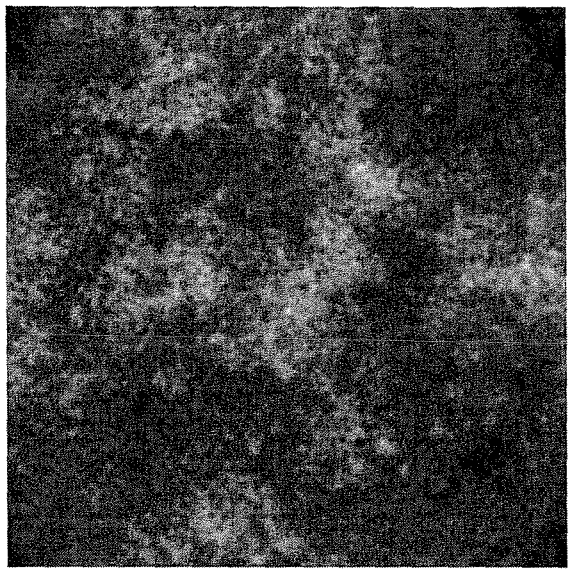

Figure 8.5: $\ln K^\epsilon$ with fractal dimension 2.8 (contrast of K^ϵ is 10^4).

Table 8.6: Test of over-sampling using permeability field shown in Fig. 8.5. Difference and error are shown in percentage.

	No Over-sampling		Over-sampling ($d_s = h$)		Over-sampling ($d_s = 2h$)	
	Max norm	l_2 norm	Max norm	l_2 norm	Max norm	l_2 norm
D_{11}^1	10.8141	7.3283	1.5048	1.2033	1.1942	0.7092
D_{11}^3	1.3060	1.0102	0.6799	0.4553	0.4875	0.2932
D_{22}^1	8.2273	7.4026	2.3588	1.6885	1.2986	0.7532
D_{22}^3	1.5108	1.0178	0.4657	0.4300	0.5565	0.4068
E_p^1	7.1424	6.5983	2.8717	1.7653	2.9171	1.4517
E_p^2	3.8179	1.5503	2.9465	1.3993	3.0436	1.3698
E_p^3	3.5359	1.5143	2.9519	1.3932	3.1065	1.3863

the same level as in Table 8.5.

As mentioned in §8.5, over-sampling may lead to significant overhead in computations. In practice, such overhead can be greatly reduced by scaling up many grid blocks together (Fig. 8.4). For the present computations, the sample domains are square regions containing 4×4 and 8×8 grid blocks for over-sampling with $d_s = h$ and $2h$, respectively. Note that due to the use of larger sample domain, the improved accuracy in the last column of Table 8.6 comes at no additional cost compared to the computation for the second column.

Interestingly, both Tables 8.5 and 8.6 show that the upscaling with the periodic and the pressure-drop no-flow boundary conditions (i.e., (8.9)) give very similar results. The

pressure error using these boundary conditions are also quite small even without over-sampling. However, as indicated by the tests for the periodic problem, this is not always the case and should not be generalized. On the other hand, the use of the Dirichlet boundary condition (8.7) should always be accompanied by over-sampling. The advantage of (8.7) is that it is easy to apply on grid blocks with general geometry, i.e., those arisen from unstructured grids.

8.7.2 Averaged over-sampling. Numerical results.

In this section we present the numerical results. The main difficulty here is the large amount of computation which is required to solve the state equation. This difficulty has been overcome by avoiding to allocate memory for the stiffness matrix. Instead, we define a routine which multiplies the stiffness matrix to the vector without explicitly defining stiffness matrix. For the calculation of the effective coefficients the cell problem with the fine resolution has been solved numerically.

In the computations we vary the three parameters for the upscaling method with averaged over-sampling. They are, the mesh size ($diam(K)$), h; the sample size ($diam(S)$), H; and the small parameter of the problem, ϵ. For the upscaling with only over-sampling, h stays unchanged. Moreover we take $h = 1$ since the error estimate is derived for a single block. But we vary H and ϵ in order to see the effect of the boundary layer term. As it is clear when we vary all three parameters at the same time, the boundary layer term (the term involving θ) does not change its value since the solution of the equation in the sample domain just gets translated. Consequently, the value of the boundary layer terms stays unchanged. This indicates that this term resonates when we vary all three parameters. On the other hand the error made by this term can be reduced by fixing the H, and decreasing h and ϵ. The reason for this is that as we go further away from the boundaries the effect of the layers gets smaller.

As we see from the numerical results of the upscaling method with averaged over-sampling, the value of the boundary layer term is small and the method converges even when we decrease all three parameters h, ϵ, and H at the same time. This indicates that the value of boundary layers value is small and is comparable to $C\epsilon$. But as we see in some cases, the effect of the small resonance term shows up and it reduces the order of the convergence of the upscaling method with averaged over-sampling.

Table 8.7: $|a_{*,h} - a_*|$ for the upscaling with averaged over-sampling.

h	$a_{*,h} - a_*$	rate
1	0.0132	
0.5	0.003	2.1

In our computations we use the values 1 and 0.5 for h. The values of other two parameters are $H = 1.6$, $H = 0.8$ and for ϵ, $\epsilon = 0.3$ and $\epsilon = 0.15$. In these cases $d_s = 0.6h$. The error corresponding to $h = 1$ is the error of the over-sampling method. We present the values of this error for different coarse blocks in the case of $h = 0.5$. As we will see, these values change and their average value is the better approximant for the effective permeability of the block.

Example 1.

In this example we consider L_ϵ with the following coefficients

$$a_\epsilon^{ij} = \left(\frac{2 + \cos(\frac{2\pi x}{\epsilon})}{2 + \cos(\frac{2\pi y}{\epsilon})} + \frac{2 + \cos(\frac{2\pi y}{\epsilon})}{2 + \cos(\frac{2\pi z}{\epsilon})} + \frac{2 + \cos(\frac{2\pi z}{\epsilon})}{2 + \cos(\frac{2\pi x}{\epsilon})} \right) \delta^{ij}$$

where δ^{ij} are the Kronecker symbols. The effective coefficient for this problem is $a_* = 3.4599$.

The numerical tests are performed for $H = 1.6$, $h = 1$, $\epsilon = 0.3$ and for $H = 0.8$, $h = 0.5$ $\epsilon = 0.15$. This is a case when we decrease all three parameters at the same time. As we discussed above the over-sampling approach does not converge and the error resonates in this case. However we see from the numerical results presented in the table 8.7, the averaged over-sampling method converges. This indicates that the effect of the boundary layer term is negligible.

In table 8.8 we demonstrate the values of a_{*,h,K_q} for the possible $K_q \subset \Omega$ in the case of $h = 0.5$, $\epsilon = 0.15$, $H = 0.8$. Here q changes from 1 to 8. The K_q are numerated lexicographically in the table. The value of $a_{*,h}$ in the case of $h = 1$, $\epsilon = 0.3$, $H = 1.6$ is the same as the value of a_{*,h,K_q} corresponding to $(1, 1, 1)$. These values represent the effective block permeability using the over-sampling approach. As we see they change from block to block and their average value approximates the effective permeability more robustly. We also would like to note that in this case the error made by averaged over-sampling method is less than 3 percent.

Table 8.8: a_{*,h,K_q} for the upscaling with averaged over-sampling.

number	a_{*,h,K_q}
(1,1,1)	3.4467
(1,1,2)	3.4620
(1,2,1)	3.4614
(1,2,2)	3.4696
(2,1,1)	3.4580
(2,1,2)	3.4699
(2,2,1)	3.4692
(2,2,2)	3.4741

Table 8.9: $|a_{*,h} - a_*|$ for the upscaling with averaged over-sampling.

h	$a_{*,h} - a_*$	rate
1	0.011	
0.5	0.007	0.7

Example 2.

In this example we consider L_ϵ with the following coefficients

$$a_\epsilon^{ij} = \left(\frac{1}{(2+\sin(\frac{2\pi x}{\epsilon}))(2+\sin(\frac{2\pi y}{\epsilon}))(2+\sin(\frac{2\pi z}{\epsilon}))} \right) \delta^{ij}$$

where δ^{ij} are the Kronecker symbols. The effective coefficient for this problem is $a_* = 0.1667$.

As in the example 1 the numerical test are performed for $H = 1.6$, $h = 1$, $\epsilon = 0.3$ and for $H = 0.8$, $h = 0.5$ $\epsilon = 0.15$. We see from the numerical results presented in table 8.9 that the averaged over-sampling method converges. We would like to note that the error of averaged over-sampling method is less than 4 percent in this case.

In table 8.10 we demonstrate the values of a_{*,h,K_q} for the possible $K_q \subset \Omega$ in the case of $h = 0.5$, $\epsilon = 0.15$, $H = 0.8$. Here q changes from 1 to 8. The K_q are numerated lexicographically in the table. The value of $a_{*,h}$ in the case of $h = 1$, $\epsilon = 0.3$, $H = 1.6$ is the same as the value of a_{*,h,K_q} corresponding to $(1,1,1)$.

Example 3

In this example we consider L_ϵ with the following coefficients

$$a_\epsilon^{ij} = \left(4 + \sin(\frac{2\pi x}{\epsilon}) + \sin(\frac{2\pi y}{\epsilon}) + \sin(\frac{2\pi z}{\epsilon}) \right) \delta^{ij}$$

Table 8.10: a_{*,h,K_q} for the upscaling with averaged over-sampling.

number	a_{*,h,K_q}
(1,1,1)	0.1553
(1,1,2)	0.1569
(1,2,1)	0.1569
(1,2,2)	0.1584
(2,1,1)	0.1581
(2,1,2)	0.1597
(2,2,1)	0.1597
(2,2,2)	0.1613

Table 8.11: $|a_{*,h} - a_*|$ for the upscaling with averaged over-sampling.

h	$a_{*,h} - a_*$	rate
1	0.15	
0.5	0.09	0.7

where δ^{ij} are the Kronecker symbols. The effective coefficient for this problem is $a_* = 4.0227$. The numerical tests as in the previous examples performed for $H = 1.6$, $h = 1$, $\epsilon = 0.3$ and for $H = 0.8$, $h = 0.5$ $\epsilon = 0.15$. The numerical results presented in the table 8.11 which indicates that the averaged over-sampling method converges. We would like to note that the error of averaged over-sampling method is about 2 percent.

In table 8.12 we demonstrate the values of a_{*,h,K_q} for the possible $K_q \subset \Omega$ in the case of $h = 0.5$, $\epsilon = 0.15$, $H = 0.8$ as we did in the previous examples. Here q changes from 1 to 8. The K_q are numerated lexicographically in the table.

Example 4

Table 8.12: a_{*,h,K_q} for the upscaling with averaged over-sampling.

number	a_{*,h,K_q}
(1,1,1)	4.0227
(1,1,2)	3.9971
(1,2,1)	3.9971
(1,2,2)	3.9715
(2,1,1)	3.9905
(2,1,2)	3.9648
(2,2,1)	3.9648
(2,2,2)	3.9391

Table 8.13: $|a_{*,h} - a_*|$ for the upscaling with averaged over-sampling.

h	$a_{*,h} - a_*$	rate
1	0.06	
0.5	0.03	1.0

Table 8.14: a_{*,h,K_q} for the upscaling with averaged over-sampling.

number	a_{*,h,K_q}
(1,1,1)	3.9410
(1,1,2)	3.9344
(1,2,1)	3.9440
(1,2,2)	3.9375
(2,1,1)	3.9438
(2,1,2)	3.9379
(2,2,1)	3.9469
(2,2,2)	3.9410

In this example we consider L_ϵ with the following coefficients

$$a_\epsilon^{ij} = \left(4 + \sin(\frac{2\pi(x-y)}{\epsilon}) + \sin(\frac{2\pi(y-z)}{\epsilon})\right)\delta^{ij}$$

where δ^{ij} are the Kronecker symbols. The effective coefficient for this problem is $a_* = 3.9358$. As in the previous examples, numerical results have been obtained for $H = 1.6$, $h = 1$, $\epsilon = 0.3$ and for $H = 0.8$, $h = 0.5$ $\epsilon = 0.15$. As we see from the numerical results presented in table 8.13, the method converges. The absolute error in this case is about 1 percent.

In table 8.14 we demonstrate the values of a_{*,h,K_q} for the possible $K_q \subset \Omega$ in the case of $h = 0.5$, $\epsilon = 0.15$, $H = 0.8$. Here q changes from 1 to 8. The K_q are numerated lexicographically in the table. The value of $a_{*,h}$ in the case of $h = 1$, $\epsilon = 0.3$, $H = 1.6$ is the same as the value of a_{*,h,K_q} corresponding to $(1,1,1)$.

Part III

Applications of MsFEM to upscaling of displacements in heterogeneous porous media

Chapter 9 Applications of MsFEM to upscaling of displacements in heterogeneous porous media

9.1 Introduction

Enhanced oil recovery methods generally involve the injection of fluids that alter the flow properties of the natural rock-fluid system in reservoir. To study the motion of this combined mixture given the distribution of pores is a goal of a reservoir engineer. In this chapter we consider the coarse models for two-phase immiscible flow. A typical example considered here are one injection well and one production well.

Through the use of sophisticated geological and geostatistical modeling tools, engineers and geologists can now generate highly detailed, three dimensional representations of reservoir properties. Such models can be particularly important for reservoir management, as fine scale details in formation properties, such as thin, high permeability layers or thin shale barriers, can dominate reservoir behavior. The direct use of these highly resolved models for reservoir simulation is not generally feasible because their fine level of detail (several millions) places prohibitive demands on computational resources. Therefore, the ability to coarsen these highly resolved geologic models to levels of detail appropriate for reservoir simulation (tens of thousands grid blocks), while maintaining the integrity of the model for purposes of flow simulation (i.e., avoiding the loss of important details), is clearly needed.

In this chapter we discuss the applications of MsFEM to the scale-up of displacement processes in heterogeneous cross-sectional models (2-D). Moreover, the disadvantages of these scale-up models and their improvements will be considered. We note that the coarse models described in this chapter are not the exact homogenized limits of detailed equations. The homogenization of transport phenomena is a complicated problem. Our goal in this chapter is to approximate the average characteristics of transport flow (e.g., production rate) on a coarser grid. For example, given the fine description of the reservoir in 2000×2000 grid we would like to describe the reservoir displacements in a coarser grid, for example 100×100 grid.

The coarse models discussed in this chapter designed to generate a coarsened model that

is capable of providing simulation predictions in close agreement with results using the original, detailed reservoir description. More specifically, it requires an agreement in the global pressure-flow rate behavior of the reservoir, the breakthrough (the time when oil reaches the production well) characteristics of the displacing fluid and, the post-breakthrough fractional flows (the production rate of oil later times) of all reservoir fluids.

The coarse model introduced in [19] achieves an efficient scale-up result by identifying high fluid velocities (via single phase flow calculation). These high flow areas lead to an early breakthrough of displacing fluids. The non-uniform coarsening idea developed in [19] coarsens non-uniformly the fine grid geological descriptions in such a way that the high flow regions are finely gridded and the low flow regions are more coarsely gridded. The resulting coarsened reservoir description is able to model both average reservoir behavior and some important effects due to extremes in reservoir properties (such as early breakthrough of injected fluids), without prior knowledge of the global flow field. But there is a rather definite limit to the scale up that can be achieved through non-uniform coarsening alone. Indeed, given many high flow areas, this coarse model needs to resolve all these areas. This decreases the efficiency of the scale-up. Let's note that MsFEM can be used for the calculation of average velocity field in this coarse model.

The coarse models we discuss in this chapter for two-phase flow speed up the scale up process and can be efficiently combined with MsFEM to describe the flow properties of the reservoir. The main idea of this method is to incorporate higher order moments (correlations) into coarse models. These ideas have been used before in material sciences and turbulent flow problems. The calculation of higher moments require fine detailed information of the velocity field or some robust approximation of it. Using the base functions of MsEM which do contain a robust approximation of details, we can calculate these higher moments. Numerical experiments show that our coarse models improve the results of existing non-uniform coarsening approach. In this chapter we also simplify rigorous upscaled models derived in [50, 20] for periodic (or layered) non-ergodic flow and apply the results for flow in a typical reservoir cross-section.

This chapter is organized as follows. In the next section we present the governing equations for two-phase flow. In section 3 we discuss initial and boundary conditions for governing equations and a special case of two-phase flow, unit mobility case, used in this chapter. The flow features and their simulation have been discussed in section 4. In section 5

we present the existing non-uniform coarsening approach along with some numerical examples. We discuss the coarse model in section 6. In section 7 we derive rigorous homogenized equations for layered and non-ergodic periodic flow using the results of [50, 20]. Sections 8, 9, 10, 11 are devoted to numerical implementations and results for the coarse model in the unit mobility case and their comparisons with the existing non-uniform coarsening approach. In sections 12 and 13 we extend the results of the unit mobility case to more general two-phase flow when the velocity weakly depend on time and present numerical results. We conclude the chapter with section 14 where my current research in upscaling of two-phase flow is discussed.

9.2 Governing equations

We consider a heterogeneous system which represent two-phase immiscible flow. Our interest is in the effect of permeability heterogeneity on two-phase flow. Therefore, we neglect the effect of gravity, compressibility, and capillary pressure, and consider porosity to be constant.

This system can be described by writing Darcy's law for each phase (all quantities are dimensionless):

$$v_j = \frac{k_{rj}(S)}{\mu_j} k \nabla p, \tag{9.1}$$

where v_j are the Darcy's velocity for the phase j ($j = o, w$ oil, water), p is pressure, S is water saturation, k is the permeability tensor, k_{rj} is the relative permeabilities of each phase and μ_j is the viscosity of the phase j. The Darcy's law for each phase coupled with mass conservation, can be manipulated to give the pressure and saturation equations:

$$\begin{aligned} \nabla(\lambda(S) k \nabla p) &= 0, \\ \frac{\partial S}{\partial t} + v_j \nabla_j f(S) &= 0, \end{aligned} \tag{9.2}$$

which can be solved subject to the boundary and initial conditions (see next sections). The

parameters in the equation (9.2) are given as:

$$\begin{aligned} \lambda &= \frac{k_{rw}(S)}{\mu_w} + \frac{k_{ro}(S)}{\mu_o}, \\ f(S) &= \frac{k_{rw}(S)/\mu_w}{k_{rw}(S)/\mu_w + k_{ro}(S)/\mu_o}, \\ v &= v_w + v_o = -\lambda(S) k \nabla p. \end{aligned} \tag{9.3}$$

$k_{rj}(S)$ are referred as the relative permeabilities of the fine scale. A single set of relative permeability curves are assumed to describe the entire domain. Permeability is usually highly variable with the different value in each fine grid block.

9.3 Unit mobility case

The unit mobility case is a special case of two-phase flow. The relative permeabilities in this case are defined as

$$k_{rw} = S, \quad k_{ro} = 1 - S, \quad \mu_o = \mu_w.$$

This case is commonly used for the upscaling of flow in natural reservoirs. Using these relations for the relative permeabilities, we can calculate the flux $f(S)$, $f(S) = S$. The final system becomes

$$\begin{aligned} v_i = k^{ij} \nabla_j p, \quad \nabla_i v_i &= 0 \\ \frac{\partial S}{\partial t} + v^j \nabla_j S &= 0. \end{aligned} \tag{9.4}$$

This system is well posed with some boundary and initial conditions. The velocity field in this model does not evolve in time unlike the general two-phase model

For our analysis we consider the following boundary and initial conditions for (9.4) and (9.2). Assuming the cross-section of the reservoir is $\Omega = [0,1] \times [0,1]$, we impose the following boundary conditions for pressure equation:

$$\begin{aligned} p &= 1 \quad \text{if } x = 0, \\ p &= 0 \quad \text{if } x = 1, \\ k^{ij} \nabla_j p &= 0 \quad \text{if } y = 0 \text{ or } y = 1. \end{aligned} \tag{9.5}$$

These boundary conditions represent the flow from $x = 0$ to $x = 1$ with no flow boundary conditions on $y = 0$ and $y = 1$. For saturation equation, we only impose

$$S(x, y, t = 0) = \{1 \text{ if } x \leq 0;\ 0 \text{if } x > 0\}.$$

As we see $S(x, y, t = 0) = H(x)$ where $H(x)$ is a Heaviside function. We note that the velocity in the x direction (v_1) at $x = 0$ and $x = 1$ is positive and we do not need boundary conditions for S on $x = 0$ and $x = 1$.

9.4 The flow features in reservoir and their modeling

The reservoir cross-sections used in this section are identical, except for the statistically generated absolute permeability fields. The flow domain is a square, of equal length and thickness. The fine grid in all cases is a uniform 100×100 rectangular grid (unless otherwise stated).

Geostatistical models often suggest [25, 41] that the logarithm of permeability field is weakly or second order stationary in space so that the mean log permeability is constant and its covariance only depends on the relative distance of two points rather than their actual locations. One commonly used covariance model is the exponential one

$$C_Y(r) = \sigma^2 exp(-(\frac{r_1^2}{l_x^2} + \frac{r_2^2}{l_y^2})$$

where r is the separation vector between two points, σ^2 is the variance of log permeability, and l_x and l_y are the correlations lengths.

We generate log-normal absolute permeability fields using a simple " moving ellipse" averaging technique. We refer to the major and minor radii of the ellipse as the correlation lengths, though we note that other papers sometimes use the ellipse diameters. We choose to define the correlation lengths in terms of the radii because they provide the closest correspondence to the use of the term for other correlation functions, such as exponential. We assume that the axes of ellipse lie along the x and y directions. The permeability fields are then characterized by three parameters: l_x, l_y, and σ, where l_x and l_y designate the correlation lengths in x and y directions, and σ^2 is the variance of logarithm of permeability field. We will express l_x and l_y in dimensionless units, so that l_x (l_y) is the ratio of the

correlation length to the length (thickness) of the cross section.

The moving ellipse algorithm is as follows. One first overlays the grid of the reservoir cross-section with a larger grid. To each cell i of this larger grid one assigns an independent Gaussian random variable x_i. For each cell i of the cross-section, a new random variable y_i is generated from x_i as follows: An ellipse with specified major and minor axes is centered on that cell, the cells of the larger grid that fall within the ellipse are marked, and y_i is set equal to the average of the x_i within marked cells. The y_i then constitute a correlated Gaussian field. A log-normal permeability field is obtained by exponentiating y_i; the mean and variance are set by the coefficients of the exponential.

The realizations we generated are all strongly heterogeneous and highly layered. The heterogeneity is determined by σ, which is equal to 1 or 2. The high degree of layering was ensured by taking $l_x \ll l_y$. We note that the layered permeability fields are a good approximation for natural reservoirs. Under the gravitational forces natural reservoirs tend to develop thin layers in x direction.

In Fig. 9.1 we depicted a realization of permeability field with $l_x = 0.3$, $l_y = 0.01$ and $\sigma = 2$. As we see high permeability areas are long thin layers. The corresponding streamline field is contoured in Fig. 9.2.

Throughout the chapter we are interested in the fractional curves. Fractional curve at time t, $F(t)$, is defined as a production rate at $x = 1$ by the following expression:

$$F(t) = \frac{\int_0^1 v_x(x=1,y,t) f(S(x=1,y,t))dy}{\int_0^1 v_x(x=1,y,t)dy}.$$

The integral in numerator represents flux at $x = 1$ and the integral in denominator normalizes this flux. We also use dimensionless time PVI throughout the chapter. For problems of this chapter

$$PVI = \frac{tv^{average}L_y}{L_xL_y} = tv^{average},$$

where L_x and L_y are the lengths of the cross-section of the reservoir in x and y directions ($L_x = L_y = 1$), t is time and $v^{average}$ is the total average velocity. As we see PVI is proportional to the flux rate of injecting fluid.

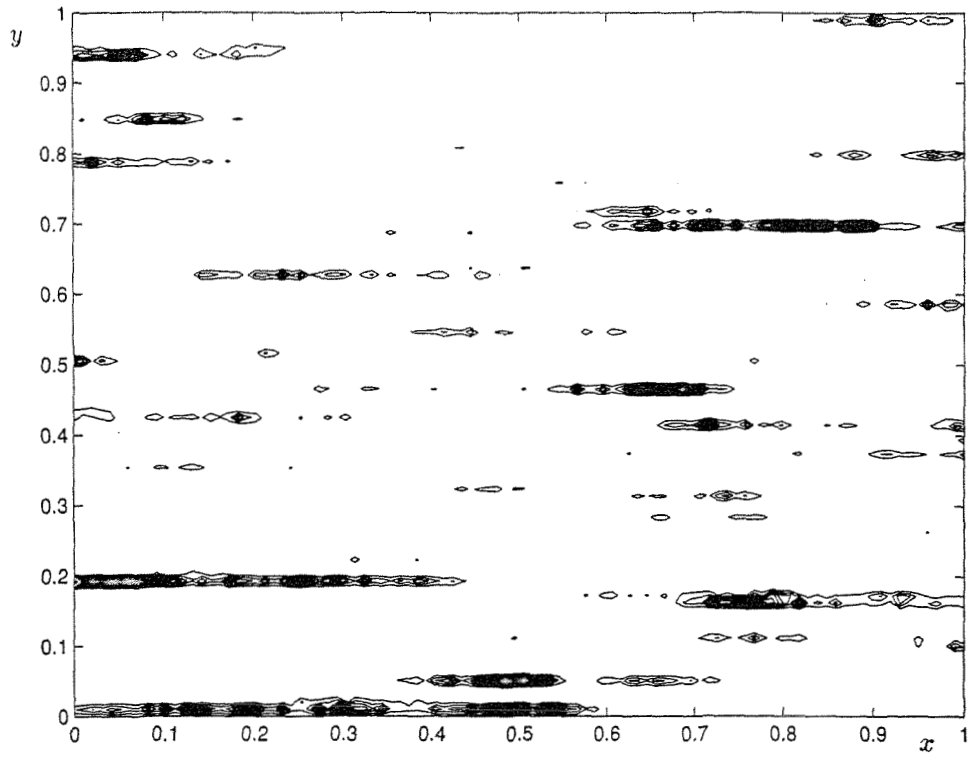

Figure 9.1: Log normally distributed permeability field with $l_x = 0.3$, $l_y = 0.01$, and $\sigma = 2$. The ratio of max to min is $2e + 4$

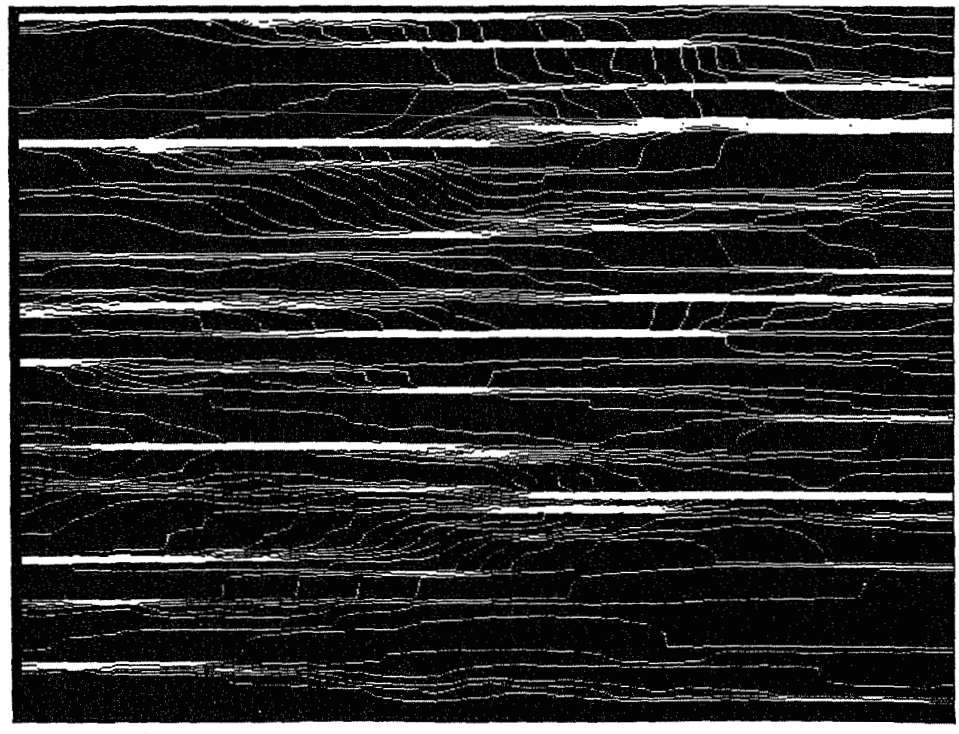

Figure 9.2: Streamlines for log normally distributed permeability field with $l_x = 0.3$, $l_y = 0.01$, and $\sigma = 2$. The ratio of max to min is $2e + 4$

9.5 Non-uniform coarsening method

Many different methods can be devised for coarsening a grid non-uniformly. In this chapter we use the term "non-uniform coarsening" to refer to the specific algorithm formulated in [19]. The characteristic feature of this algorithm is that simulations on the coarsened grid reproduce the fine grid results without any upscaling of the fine grid relative permeabilities. Only the absolute permeabilities are upscaled. The coarsenings identified by the non-uniform coarsening algorithm tend to be finely gridded in regions of high flux, and more coarsely gridded in the regions of low flux. For the unit mobility case, denoting k^* upscaled absolute permeability (see previous chapters), we have the following formulation for our model

$$\begin{aligned} &\nabla \cdot k^* \nabla p^* = 0, v^* = k^* \nabla p^* \\ &\frac{\partial S^*}{\partial t} + v^* \cdot \nabla S^* = 0 \end{aligned} \tag{9.6}$$

with the same initial conditions as the fine saturation field.

A straightforward implementation of this scale-up model fails to give a robust approximation. Indeed, if v^* is a constant, then we miss the important features of the flow. The idea of non-uniform coarsening incorporated into this model does indeed make a difference. The typical velocity field as we discussed in the previous section has thin layers of high flow regions. Resolving these areas non-uniformly the coarse method increases the accuracy of the method. In other words the coarse blocks in high velocity regions are smaller than the coarse blocks away from them. The accuracy of this method depends on the degree of resolving high flow areas. Consequently, the method can be almost as expensive as solving the fine scale problem for problems with many high flow layers. For example in Fig. 9.3 we plot fractional flow curves for permeability field (see Fig. 9.1) with $l_x = 0.3$, $l_y = 0.01$, and $\sigma = 2$. As we see coarse models on 10×10 and 15×15 coarse grids do not give a satisfactory agreement while the coarse model on 23×24 coarse grid has a good accuracy. This indicates that there is a definite limit to the scale up through non-uniform coarsening. In order to increase the efficiency of the method, we need to improve it. In the next section, we discuss a method which speeds up the upscaling process.

The non-uniform coarsening method can be applied to general two-phase flow. In this case, as we noted in the beginning of the section, the fine relative permeability curves retain

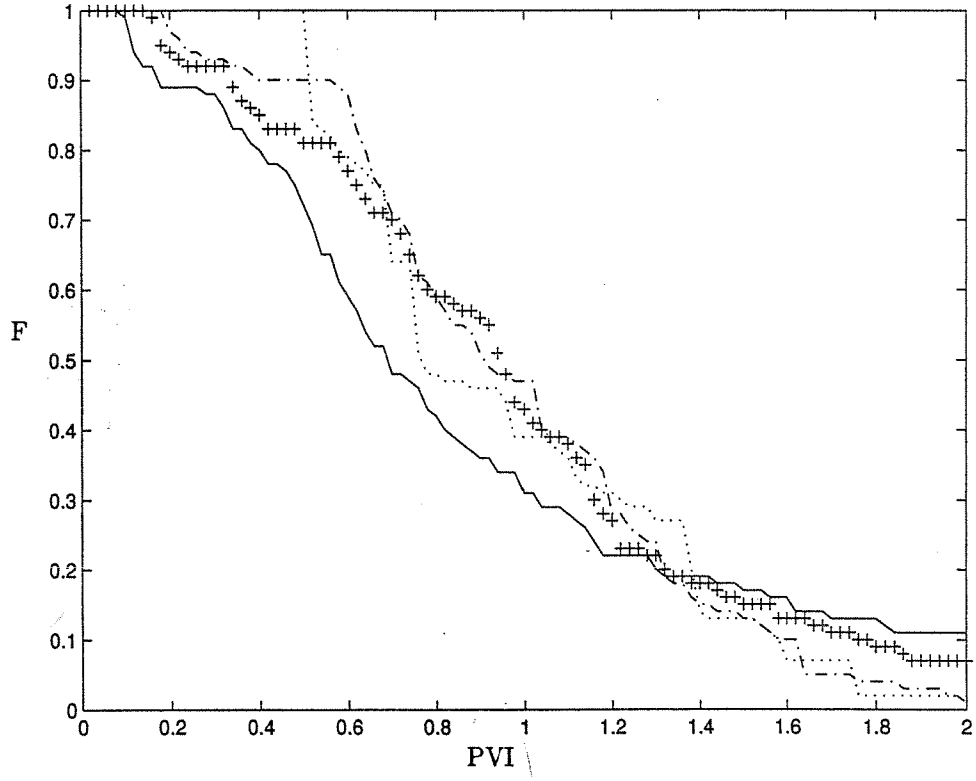

Figure 9.3: The comparison of three different scale-up results using non-uniform coarsening method. Solid line is the fractional flow curve for the fine model 100×100. Dotted line is for 10×10, Dashed-dotted line is for 15×15, '+" line is for 23×24 coarse model

their form:

$$\begin{aligned} \nabla(\lambda(S^*)k^*\nabla p^*) &= 0, \\ \frac{\partial S^*}{\partial t} + v_j^*\nabla_j f(S^*) &= 0, \end{aligned} \tag{9.7}$$

where

$$\begin{aligned} \lambda(S^*) &= \frac{k_{rw}(S^*)}{\mu_w} + \frac{k_{ro}(S^*)}{\mu_o}, \\ f(S^*) &= \frac{k_{rw}(S^*)/\mu_w}{k_{rw}(S^*)/\mu_w + k_{ro}(S^*)/\mu_o}, \\ v^* &= \lambda(S^*)k^*\nabla p^*. \end{aligned} \tag{9.8}$$

9.6 Derivation of the coarse model equations for unit mobility case

The coarse model we are going to derive in this section is known in literature [14, 57]. We would like to repeat this derivation in this section. In the following sections, we will describe how to use this model in a bounded reservoir and how this model can be effectively combined with MsFEM. In the next section, we will also present a rigorous homogenization model and compare it with this coarse model.

For the derivation of the coarse model, we use an expansion of the velocity and the saturation field around their averages. We note that this derivation is not rigorous and assumes that the variations of the perturbed fields are small. The averages of perturbed fields could be their volume or ensemble averages. These approaches are commonly used in turbulence and other fields.

$$S(x,y) = S_0(x,y) + S', \quad v_i(x,y) = v_i^0 + v_i'(x,y).$$

Substituting them into the saturation equation of (9.4) we have

$$\frac{\partial S_0}{\partial t} + \frac{\partial S'}{\partial t} + v_0^j\nabla_j S_0 + v_0^j\nabla_j S' + v_j'\nabla_j S^0 + v_j'\nabla_j S' = 0. \tag{9.9}$$

Next we average this equation and take into account that $\overline{v'} = 0$, and $\overline{S'} = 0$:

$$\frac{\partial S_0}{\partial t} + v_0^j \nabla_j S_0 + \overline{v_j' \nabla_j S'} = 0. \tag{9.10}$$

The fluctuating equations is obtained by subtracting the average equations from the (9.9):

$$\frac{\partial S'}{\partial t} + v_0^j \nabla_j S' + v_j' \nabla_j S^0 + v_j' \nabla_j S' = \overline{v_j' \nabla_j S'}. \tag{9.11}$$

Our goal is to derive an equation for $\overline{S' v_j'}$ from (9.11) in order to have a closed system with equation (9.10). If we multiply (9.11) by v_j' and average (neglecting third order terms), we find that the term $v_0^j \nabla_j S'$ causes a difficulty in closing the above system. The contribution of this term is not negligible which can be checked with a simple computation. To deal with this term further, we notice that it can be included in $\partial S'/\partial t$ along the trajectories

$$dx_j/dt = v_j^0.$$

Indeed, if we project the first two terms on to this trajectory, we have

$$\frac{dS'}{dt} = \frac{\partial S'}{\partial t} + v_0^j \nabla_j S'.$$

As we see from here $v_0^j \nabla_j S'$ causes the non-locality in the problem. The fact that the averaged equations have the nonlocal and memory dependent diffusion has been known from the mathematical theory of homogenization (see next section). Consequently, we have

$$\frac{dS'(t, x(t))}{dt} + v_j' \nabla_j S^0 + v_j' \nabla_j S' = \overline{v_j' \nabla_j S'} \tag{9.12}$$

where $x(t)$ is a trajectory of $dx/dt = v_0(x)$. Then for each (x, t), s.t. $x(t) = x$ integrating the equation (9.12) over $(0, t)$, we have

$$S'(t, x) = -\int_0^t \{v_j'(x(\tau) \nabla_j S^0(\tau, x(\tau)) + v_j'(x(\tau)) \nabla_j S'(\tau, x(\tau))\} d\tau + \int_0^t \overline{v_j'(x(\tau)) \nabla_j S'}(\tau, x(\tau)) d\tau. \tag{9.13}$$

Multiplying this equation by $v_k'(x)$ and averaging we get

$$\overline{S'(t,x)v_k'(x)} = -\int_0^t \{\overline{v_k'(x)v_j'(x(\tau))}\nabla_j S^0(\tau,x(\tau)) + \overline{v_k'(x)v_j'(x(\tau))\nabla_j S'(\tau,x(\tau))}\}d\tau + \int_0^t \overline{v_k'(x)\overline{v_j'(x(\tau))\nabla_j S'}(\tau,x(\tau))}d\tau. \tag{9.14}$$

The last term is equal to zero because of $\overline{v'} = 0$. We neglect the term containing $\overline{v'v'S'}$ as a higher order term. Then we have

$$\overline{S'(t,x)v_k'} = -\int_0^t \overline{v_k'(x)v_j'(x(\tau)}\nabla_j S^0(\tau,x(\tau))d\tau. \tag{9.15}$$

Consequently the homogenized equation derived in this heuristic way is

$$\frac{\partial S^0}{\partial t} + v_j^0\nabla_j S^0 = \nabla_k \int_0^t \overline{v_k'(x)v_j'(x(\tau)}\nabla_j S^0(\tau,x(\tau))d\tau. \tag{9.16}$$

As we see this averaged equation does have a diffusive behavior with non-local diffusion coefficients. In the next sections we discuss the boundary conditions for this equation. We note that this equation can be further approximated with

$$\frac{\partial S^0}{\partial t} + v_j^0\nabla_j S^0 = \nabla_k \int_0^t \overline{v_k'(x)v_j'(x(\tau)}d\tau\nabla_j S^0(t,x). \tag{9.17}$$

To show this we need to show that

$$\int_0^t \overline{v_k'(x)v_j'(x(\tau)}\nabla_j S^0(\tau,x(\tau))d\tau = \int_0^t \overline{v_k'(x)v_j'(x(\tau))}\nabla_j d\tau S^0(,x). \tag{9.18}$$

Indeed,

$$\int_0^t \overline{v_k'(x)v_j'(x(\tau)}\nabla_j S^0(\tau,x(\tau))d\tau = \int_0^t \frac{d}{d\tau}\left(\int_0^\tau \overline{v_k'(x)v_j'(x(s)}ds\right)\nabla_j S^0(\tau,x(\tau))d\tau =$$
$$\int_0^t \overline{v_k'(x)v_j'(x(\tau)}\nabla_j d\tau S^0(t,x) - \int_0^t \left(\int_0^\tau \overline{v_k'(x)v_j'(x(s)}ds\right)\nabla_j \frac{d}{d\tau}S^0(\tau,x(\tau))d\tau =$$
$$\int_0^t \overline{v_k'(x)v_j'(x(\tau)}\nabla_j d\tau S^0(t,x) + O(v'v'v'v'). \tag{9.19}$$

In the last step, we have used the fact that the change of $S_0(t,x(t))$ along the trajectory is a higher order term according to (9.16). For the further convenience we denote

$$R^{ij}(x,x(t)) = \overline{v_i'(x)v_j'(x(t))}$$

and

$$D^{ij}(x,t) = \int_0^t R^{ij}(x,x(\tau))d\tau.$$

Then the final equation becomes

$$\frac{\partial S_0}{\partial t} + v_0^j\nabla_j S_0 + \nabla_i D^{ij}(x,t)\nabla_j S_0 = 0. \tag{9.20}$$

9.7 Rigorous derivation of upscaled equations for layered system

In this section we will derive a rigorous upscaled equation for layered media and compare the results with the results of the previous section. One reason for that is the heuristic derivation of the previous section assumes small variations for perturbed field while we are interested in the cases when the variances are not very small.

We consider

$$\frac{\partial S^\epsilon(x,y,t)}{\partial t} + v(\frac{y}{\epsilon})\frac{\partial S^\epsilon(x,y,t)}{\partial x} = 0$$
$$S^\epsilon(x,y,0) = H(x) \tag{9.21}$$

where $H(x)$ is Heaviside function and $v(z)$, $z = y/\epsilon$ is a periodic function in $[0,1]$. We

assume that $v > 0$. The solution of this equation is

$$S^\epsilon(x,y,t) = H(x - tv(\frac{y}{\epsilon})).$$

As $\epsilon \to 0$ we have

$$S_0(x,t) = \int_0^1 H(x - tv(z))dz.$$

The homogenized equation for (9.21) was studied in [50]. We will briefly describe it here. Applying Fourier transform in x and Laplace transform in t (9.21) gives

$$LF_x S^\epsilon(\xi,y,p) = \frac{F_x H(\xi)}{p + 2i\pi v(y/\epsilon)}$$

which is valid for $\Re p > 0$. Denoting $d\nu_y$ Young measures associated with $v(y/\epsilon)$ we find the weak limit S_0 of the sequence S^ϵ

$$LF_x S_0(\xi,y,p) = F_x H(\xi) \int \frac{d\nu_y(\lambda)}{p + 2i\pi\xi\lambda}. \tag{9.22}$$

It can be shown that there exists a nonnegative measure $d\mu(\lambda)$ suct that

$$\left(\int \frac{d\nu(\lambda)}{z+\lambda}\right)^{-1} = z + v_0 - \int \frac{d\mu(\lambda)}{z+\lambda} \tag{9.23}$$

holds for z outside the real segment $[-max(v), -min(v)]$. Then taking inverse Fourier and Laplace transform of (9.22), it can be shown that the homogenized equation satisfy

$$\frac{\partial S_0}{\partial t} + v_0 \frac{\partial S_0}{\partial x} = \int_0^t \int \frac{\partial^2}{\partial x^2} S_0(x - \lambda(t-s), s) d\mu(\lambda) ds. \tag{9.24}$$

To understand this homogenized equation we assume that there are finite number of layers with given volume fractions, i.e. $v(z) = \{v_1, \ldots, v_n\}$, and $f_i = meas\{z, v(z) = v_i\}$. Let's note that the layered geostatistical models are built based on finite number of layers. In this case l.h.s of (9.23) becomes

$$\int \frac{d\nu(\lambda)}{z+\lambda} = \sum_{i=1}^n \frac{f_i}{z+v_i}.$$

Consequently,

$$\int \frac{d\mu(\lambda)}{z+\lambda} = z + v_0 - \left(\sum_{i=1}^{n} \frac{f_i}{z+v_i} \right)^{-1}$$

where $v_0 = \sum_{i=1}^{n} v_i f_i$. In the case of $n = 2$ we can explicitly calculate the diffusion. Indeed, in this case we have

$$\int \frac{d\mu(\lambda)}{z+\lambda} = \frac{v_1 f_1^2 + v_2 f_2^2 - v_0^2}{z + v_1 f_2 + v_2 f_1},$$

where the layers with velocities v_1 and v_2 have volume fractions f_1 and f_2. This indicates that

$$d\mu(\lambda) = (v_1 f_1^2 + v_2 f_2^2 - v_0^2)\delta(\lambda - v1 f_2 - v_2 f_1).$$

It can be easily checked that $v_1 f_1^2 + v_2 f_2^2 - v_0^2$ is equal to the variance of v. Then the final homogenized equation is

$$\frac{\partial S_0}{\partial t} + v_0 \frac{\partial S_0}{\partial x} = var(v) \int_0^t \frac{\partial^2}{\partial x^2} S_0(x - (v_1 f_2 + v_2 f_1)(t-s), s) ds. \tag{9.25}$$

The diffusion in the averaged equation derived in the previous section (9.16) can be calculated explicitly since the correlation of the velocity field is constant and equal to $var(v)$. Then (9.16) becomes

$$\frac{\partial S_0}{\partial t} + v_0 \frac{\partial S_0}{\partial x} = var(v) \int_0^t \frac{\partial^2}{\partial x^2} S_0(x - v_0(t-s), s) ds. \tag{9.26}$$

As we see the only difference in equations (9.25) and (9.26) is in the "average" velocity field which counts for memory effect. We note that in the case $f_1 = f_2$ the equations are the same.

Next we will try to describe the behavior of the homogenized equation in the general n layered case. In this case denoting $P_m(z) = \sum_{i=1}^{m} a_m z^m$ to be mth order polynomial we

have

$$-\int \frac{d\mu(\lambda)}{z+\lambda} = \left(\sum_{i=1}^{n} \frac{f_i}{z+v_i}\right)^{-1} - z - \sum_{i=1}^{n} f_i v_i = \frac{\prod_{i=1}^{n}(z+v_i)}{\sum_{k=1}^{n} f_k \prod_{i=1,\ i\neq k}^{n}(z+v_i)} - z - \sum_{i=1}^{n} f_i v_i = \frac{B_n z^n + B_{n-1} z^{n-1} + B_{n-2} z^{n-2} + P_{n-3}(z)}{\sum_{k=1}^{n} f_k \prod_{i=1,\ i\neq k}^{n}(z+v_i)}. \tag{9.27}$$

The coefficients B_n, B_{n-1}, and B_{n-2} can be readily calculated

$$B_n = 1 - \sum_{i=1}^{n} f_i = 0$$

$$B_{n-1} = \sum_{i=1}^{n} v_i - \sum_{k=1}^{n} f_k \sum_{i=1,\ i\neq k}^{n} v_i - \sum_{i=1}^{n} f_i v_i = 0$$

$$B_{n-2} = \sum_{i,j=1,\ i\neq j}^{n} v_i v_j - \sum_{k=1}^{n} f_k \sum_{i,j=1,\ i\neq j,\ i\neq k,\ j\neq k}^{n} v_i v_j - (\sum_{i=1}^{n} f_i v_j) \sum_{k=1}^{n} f_k \sum_{i=1,\ i\neq k}^{n} v_i = \sum_{i,j=1,\ i\neq j}^{n} v_i v_j - \sum_{k=1}^{n} f_k \sum_{i,j=1,\ i\neq j,\ i\neq k,\ j\neq k}^{n} v_i v_j - (\sum_{i=1}^{n} f_i v_j)(\sum_{i=1}^{n} v_i - \sum_{i=1}^{n} f_i v_i) = \sum_{i,j=1,\ i\neq j}^{n} v_i v_j - \sum_{k=1}^{n} f_k \sum_{i,j=1,\ i\neq j,\ i\neq k,\ j\neq k}^{n} v_i v_j - \sum_{k=1}^{n} \sum_{i=1,\ i\neq k}^{n} f_k v_i v_k + (\sum_{i=1}^{n} f_i v_j)^2 - \sum_{i=1}^{n} f_i v_i^2 = (\sum_{i=1}^{n} f_i v_j)^2 - \sum_{i=1}^{n} f_i v_i^2 = var(v). \tag{9.28}$$

So we have

$$-\int \frac{d\mu(\lambda)}{z+\lambda} = \frac{var(v) * z^{n-2}}{\sum_{k=1}^{n} f_k \prod_{i=1,\ i\neq k}^{n}(z+v_i)}. \tag{9.29}$$

Furthermore, it can be shown that $R(z) = \sum_{k=1}^{n} f_k \prod_{i=1,\ i\neq k}^{n}(z+v_i)$ has exactly $n-1$ roots $-u_i$, such that $v_i < u_i < v_{i+1}$ assuming that $0 < v_1 < \cdots < v_n$. To show that we just

need to look at the change of the sign of $R(z)$. We can easily see that

$$R(-v_i) = f_i \prod_{k=1,\ k\neq i}^{n} (-v_i + v_k).$$

Consequently,

$$R(-v_i)R(-v_{i+1}) = (-v_i + v_{i+1})(-v_{i+1} + v_i) \prod_{k=1,\ k\neq i\ k\neq i+1}^{n} (-v_i + v_k) \prod_{k=1,\ k\neq i\ k\neq i+1}^{n} (-v_{i+1} + v_k) < 0.$$

From above computations we conclude that

$$-\int \frac{d\mu(\lambda)}{z+\lambda} = \sum_{i=1}^{n-1} \frac{\alpha_i}{z+u_i} \tag{9.30}$$

where $v_i < u_i < v_{i+1}$, and $\sum \alpha_i = var(v)$. The last equality follows from the fact that $B_{n-2} = var(v)$. Consequently,

$$d\mu(\lambda) = \sum_{i=1}^{n-1} \alpha_i \delta(\lambda - u_i)$$

and the homogenized equation is

$$\frac{\partial S_0}{\partial t} + v_0 \frac{\partial S_0}{\partial x} = var(v) \sum_{i=1}^{n-1} \int_0^t \beta_i \frac{\partial^2}{\partial x^2} S_0(x - u_i(t-s), s) ds, \tag{9.31}$$

where $\sum_{i=1}^{n-1} \beta_i = 1$.

Remark 9.7.1. As we described in the previous section the cross-section of a reservoir is modeled with permeability field which is a log-normal random field with correlation lengths in x and y directions l_x and l_y such that $l_x \gg l_y$. As a consequence, the velocity field has long and thin layers concentrated on these high permeability regions. Generally (it can be checked numerically) the corresponding velocity fields have a high fraction of low velocity areas and a low fraction of high velocity areas. In this case sum of β_i in (9.31) corresponding to high velocities is close to 1. For simplicity if we assume there are two ranges of the velocity v and V, where $V \gg v$, we have the following approximation of our

coarse model

$$\frac{\partial S_0}{\partial t} + v_0 \frac{\partial S_0}{\partial x} = var(v) \int_0^t \frac{\partial^2}{\partial x^2} S_0(x - u(t-s), s) ds. \tag{9.32}$$

Let's note that if the volume fraction of v is larger, u is close to V. In the next sections we will show numerically that this model can be used successfully for almost layered cases.

Remark 9.7.2. The coarse model (9.31) can be derived for general non-ergodic periodic flow with no stagnation point using the continuous model [20] similar to the one for layered flow. Let's note that the velocity field corresponding to the flow described in section 3 does not have stagnation point. Denoting the periodic velocity field by $(v_1(x/\epsilon), v_2(x/\epsilon))$ we define

$$\overline{v_i} = \frac{1}{|Y|} \int_Y v_i(y) dy$$

where Y is the unit cube (period). The slope is often referred to as the asymptotic slope of streamline, or the rotation number. Furthermore, following [20] we define $\tilde{v}_i$ such that

$$\tilde{v}_i(y) = \lim_{t \to \infty} \frac{1}{t} \int_0^t v_i(X(s,y)) ds$$

where $dX/dt = v(X)$, $X(0,y) = y$. It has been shown that if θ is the angle such that $-\tilde{v}_1(y)\sin(\theta) + \tilde{v}_2(y)\cos(\theta) = 0$ (if $(\overline{v_1}, \overline{v_2})$ is not equal to zero then θ is the directional angle of $(\overline{v_1}, \overline{v_2})$), then the homogenized equation is

$$\frac{\partial S_0}{\partial t} + \overline{v_i} \frac{\partial S_0}{\partial x_i} =$$
$$\int_0^t \int (\cos(\theta) \frac{\partial}{\partial x_1} + \sin(\theta) \frac{\partial}{\partial x_2})^2 S_0(x_1 - \lambda(t-s)\cos(\theta), x_1 - \lambda(t-s)\sin(\theta), s) d\mu(\lambda) ds. \tag{9.33}$$

Here the measure $d\mu(\lambda)$ is defined similar to the layered case, i.e.,

$$\int \frac{d\mu(\lambda)}{z+\lambda} = z + \int_Y b(y) dy - \left(\int \frac{dy}{z + b(y)} \right)^{-1}$$

where $b(y) = \tilde{v}_1(y)\cos(\theta) + \tilde{v}_2(y)\sin(\theta)$. In the case of non-ergodic flow $(\tilde{v}_1(y), \tilde{v}_2(y))$ is constant along the streamlines. Assuming there are n distinct streamlines with velocities

$\tilde{u}_k$ and the corresponding volume fractions f_k, we can show (similar to the layered case) that the homogenized equation is

$$\frac{\partial S_0}{\partial t}+\overline{v_i}\frac{\partial S_0}{\partial x_i} =$$
$$var(\tilde{v})\sum_{i=1}^{n-1}\int_0^t \beta_i(\cos(\theta)\frac{\partial}{\partial x_1}+\sin(\theta)\frac{\partial}{\partial x_2})^2 S_0(x_1-u_i(t-s)\cos(\theta), x_2-u_i(t-s)\sin(\theta), s)ds. \tag{9.34}$$

where $\sum_{i=1}^{n-1}\beta_i = 1$, and u_i are between $\tilde{u}_i$ and $\tilde{u}_{i+1}$ as in the layered case.

9.8 The formulations of the diffusion in coarse models

In this section we present a formulation for the diffusion in the coarse models. Our goal with formulating the coarse models is to show that these coarse models improve the results of existing non-uniform coarsening models and can be efficiently combined with MsFEM. The coarse model we are going to implement and analyze in this chapter is (9.20)

$$\frac{\partial S_0}{\partial t}+v_0^j\nabla_j S_0+\nabla_i D^{ij}(x,t)\nabla_j S_0 = 0 \tag{9.35}$$

where $D^{ij}(x,t) = \int_0^t \overline{v_i'(x)v_j'(x(\tau))}d\tau$. The derivation of this model presented in previous sections is carried out under the assumption of small variations for perturbed fields. This assumption does not hold for the whole reservoir since reservoirs are very heterogeneous. But increasing the number of coarse blocks, we can still accept this assumption. Numerical experiments show that for the variation $\sigma = 2$ of permeability field, the coarse model (9.35) gives satisfactory accuracy. As we will show in later sections, this coarse model improves the existing non-uniform coarsening approach. In our numerical calculations, we are going to compare a non-uniform coarsening model and our coarse model (9.35) on the same coarse grid. We will see that our coarse model makes a substantial improvement over the existing method. Note that this coarse model can be effectively combined with MsFEM since the calculations of two-point correlations $D^{ij}(x,t)$ require the approximation of fine velocity field.

Remark 9.8.1. In this remark we would like to show some approximations for the diffusion coefficients $D^{ij}(x,t)$. We note that the diffusion coefficients is approximately equal to the correlation between the length of the fine trajectory in $(0,t)$ which starts at point x and $v'(x)$. Indeed,

$$D^{ij}(x,t) = \int_0^t \overline{v_i'(x)v_j'(x(\tau))}d\tau = \overline{v_i'(x)\int_0^t v_j(x(\tau))d\tau} \approx \overline{v_i'(x)L_j} \tag{9.36}$$

where L_j is displacement in the j direction of the coarse trajectory which is started at x at time 0 and traveled with the fine velocity $-v$ until time t. Next we propose to use the following simplifications of the final result of (9.36)

$$D^{ij}(x,t) \approx \sqrt{\overline{v_i'(x)v_i'(x)}}L_j^0 \tag{9.37}$$

where $\sqrt{\overline{v'(x)v'(x)}}$ is the standard deviation of v' at the coarse block with the center x, L_j^0 is displacement in the j $(j=1,2)$ direction of the coarse trajectory which is started at x at time 0 and traveled with the coarse velocity $-v_0$ until time t. Note that the numerical experiments show that this approximation works as well as the previously suggested one for flows with large correlation length in x direction and small correlation length in y. The advantage of this approximation is that we calculate the variance of the velocity in each coarse block instead of two-point correlation. This reduces the computational cost of the problem.

Remark 9.8.2. The numerical examples we present are for flow scenarios with $l_x \gg l_y$, where l_x and l_y are correlation lengths in x and y directions. This flow filed is almost layered and we can try to use (9.31). To use this model in its general form is cumbersome. We can simplify this model, noticing that the velocity fields corresponding to log-normal permeability fields with large ratio l_x/l_y tend to have low velocity areas with high volume fraction and high velocity areas with low volume fraction. In this case we can assume that the velocity field has a two characteristics scale. As we noticed in the previous section, in this case the coarse model (9.31) can be approximated with

$$\frac{\partial S_0}{\partial t} + v_0\frac{\partial S_0}{\partial x} = var(v)\int_0^t \frac{\partial^2}{\partial x^2}S_0(x-u(t-s),s)ds \tag{9.38}$$

where u is close to v_{max}. A few numerical results for this model show that this coarse model has a good accuracy for small l_y and as we increase the value of l_y the accuracy of the method decreases. The comparison with non-uniform coarsening approach shows that (9.38) can give comparable accuracy for fractional curves in a very much coarser grid. We would like to note that this coarse model can be effectively combined with MsFEM to calculate v_0 and $var(v)$, where these quantities require an approximation of the detailed velocity field in general.

9.9 The use of MsFEM in coarse models

Our coarse models as we see from the previous section require average characteristics of the flow. These average characteristics (variation, two-point correlations) can be calculated with the knowledge of the detailed behavior of the flow. For some simple cases, we can recover these average characteristics without knowing the detailed behavior of the velocity field, but in general we need some approximation of fine velocity field. We use MsFEM to construct the approximation for the detailed flow field, which is further used for calculations of average characteristics of the flow. Indeed, as we see from the first part of the work, multiscale bases functions contain the small scale information. To construct the approximation of a fine pressure field in a given coarse block, we just need

$$p_\epsilon^h = \sum p_0^k \phi_\epsilon^k$$

where ϕ_ϵ^k are the base functions in this coarse block, p_0^k are the nodal values calculated from coarse matrix equation, and p_ϵ^h is the approximation of the pressure field in the coarse block. The H^1 convergence for MsFEM presented in previous sections indicates the robustness of our construction. We would like to note that the calculation of the variance of velocity field in a coarse block uses multiscale finite element base functions of this coarse block once. Consequently this base function can be discarded from memory after we calculate the average information in this coarse block.

9.10 Boundary conditions for coarse models

Adding the diffusion term into the coarse model requires additional boundary conditions since we increase the order of our partial differential equation. In this section, we describe what boundary conditions we use for our coarse models. We start describing boundary conditions for the coarse model

$$\frac{\partial S_0}{\partial t} + v_0^j \nabla_j S_0 + \nabla_i D^{ij}(x,t) \nabla_j S_0 = 0 \tag{9.39}$$

where $D^{ij}(x,t) = \int_0^t v_i'(x) v_j'(x(\tau)) d\tau$. Because the y component of the velocity is equal to zero on $y = 0$ and $y = 1$, we conclude that $v_2' = 0$. Consequently, $D^{22} = 0$ on $y = 0$ and $y = 1$ for all t and we do not need to impose any boundary conditions on $y = 0$ and $y = 1$ since there is no convective or diffusive flux. To impose boundary conditions on $x = 0$, we note that $D^{11} = 0$ on $x = 0$ for all t because $x(\tau) = x - v_0 t > 0$, or $t < x/v_0$. This means that the integration path shrinks to zero as we approach to $x = 0$, and therefore $D^{11}(0, y, t) = 0$. Consequently, there is no diffusive flux on $x = 0$ and we do not need to impose any extra boundary conditions. On $x = 1$ we have positive outgoing flux. To formulate boundary conditions on $x = 1$, we note that in dimensionless time PVI the amount of fluid in the reservoir at a given time is fixed and equal to the amount of fluid of the detailed flow. Since we do not have any diffusive flux entering into the reservoir on $x = 0$, we require that there is only convective flux on $x = 1$, or diffusive flux is equal to zero on $x = 1$.

Remark 9.10.1. The same boundary conditions are imposed for

$$\frac{\partial S_0}{\partial t} + v_0 \frac{\partial S_0}{\partial x} = var(v) \int_0^t \frac{\partial^2 S}{\partial x^2}(x - u(t-s), s) ds. \tag{9.40}$$

9.11 Numerical results for unit mobility case

In this section we analyze numerically the coarse models we proposed in the previous section. The hyperbolic equations in these models used second order ENO schemes in order to avoid numerical diffusions. We consider the coarse model (9.35)

$$\frac{\partial S_0}{\partial t} + v_0^j \nabla_j S_0 + \nabla_i D^{ij}(x,t) \nabla_j S_0 = 0 \tag{9.41}$$

where $D^{ij}(x,t) = \int_0^t \overline{v_i'(x)v_j'(x(\tau))}d\tau$. The numerical examples we present here compare the fractional flow curves of the coarse model (9.41) and the non-uniform coarsening method on the same coarse grid. Coarse grids for the examples are chosen to be non-uniform in order to have the best performance for the non-uniform coarsening approach. In the numerical examples, we use the layered permeability fields as discussed before. As our numerical experiments Fig. 9.4 - Fig. 9.10 demonstrate, on the non-uniform coarse grid our coarse model with diffusion improves the results of the non-uniform coarsening method. Moreover, our method is capable of accurately predicting the fractional flow curve on as coarse as 10×10 grid.

In Fig. 9.4 we consider the layered case with very large correlation length in x direction. The post-breakthrough fractional flow curves for non-uniform coarsening method and the coarse model (9.41) are compared on 10×10 coarse grid and 100×100 fine grid. As this figure shows that our coarse model gives a good accuracy and improves the existing approach. In Fig. 9.5 we decrease l_x to $l_x = 0.3$. This flow deviates from layered flow since the correlation in x direction is less than the length of the reservoir in the x direction. We see from this figure that our method improves the result of the non-uniform coarsening method.

In numerical examples presented in Fig. 9.6 - Fig. 9.8, we increase the value of l_y. As the value of l_y increases, the flow field becomes less layered and the effect of cross flow becomes noticable. As a result the non-uniform coarsening approach, gives more accurate predictions in comparison to the layered case. The reason is that is the approximation of the flow in each coarse block with single average velocity becomes more robust since the flow loses its individual features in each coarse block. Despite that, we see from Fig. 9.6 - Fig. 9.8 that our coarse model improves the predictions of the non-uniform coarsening approach.

Numerical examples Fig. 9.9 - Fig. 9.10 are performed for flow with smaller correlation length in the x direction with $l_x = 0.15$. As we see from these numerical results, our coarse model (9.41) retains good accuracy and improvement in all cases.

Remark 9.11.1. We would like to present a few numerical results for the coarse model

$$\frac{\partial S_0}{\partial t} + v_0 \frac{\partial S_0}{\partial x} = var(v) \int_0^t \frac{\partial^2 S}{\partial x^2}(x - u(t-s), s)ds. \tag{9.42}$$

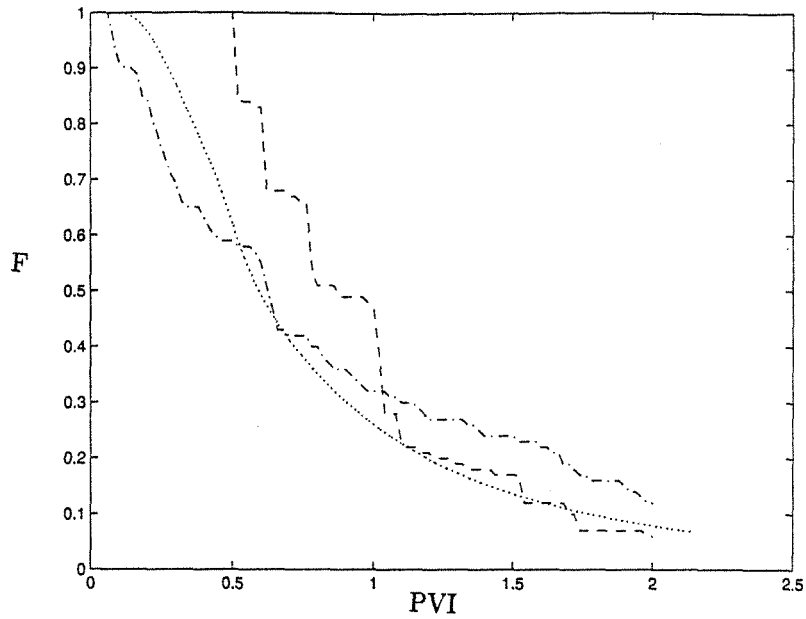

Figure 9.4: Case $l_x = 0.6$, $l_y = 0.01$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 10×10 coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

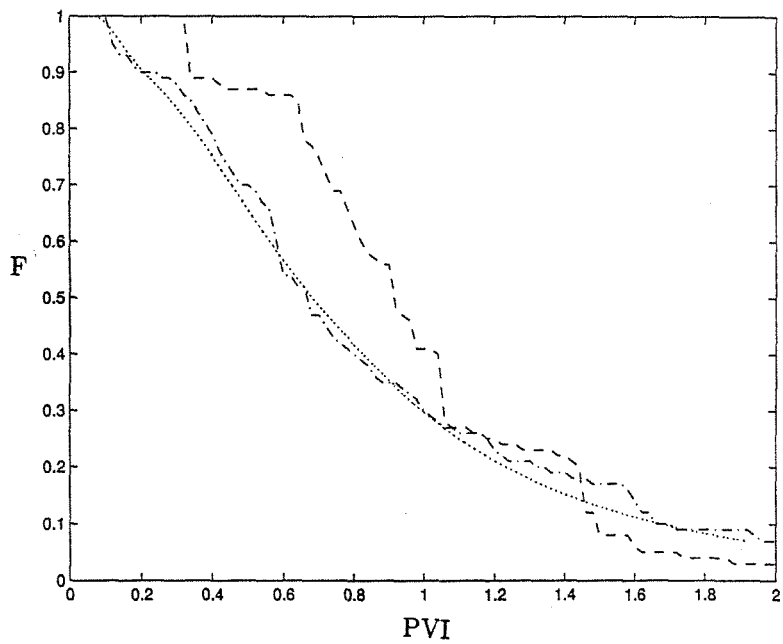

Figure 9.5: Case $l_x = 0.3$, $l_y = 0.005$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 12×12 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

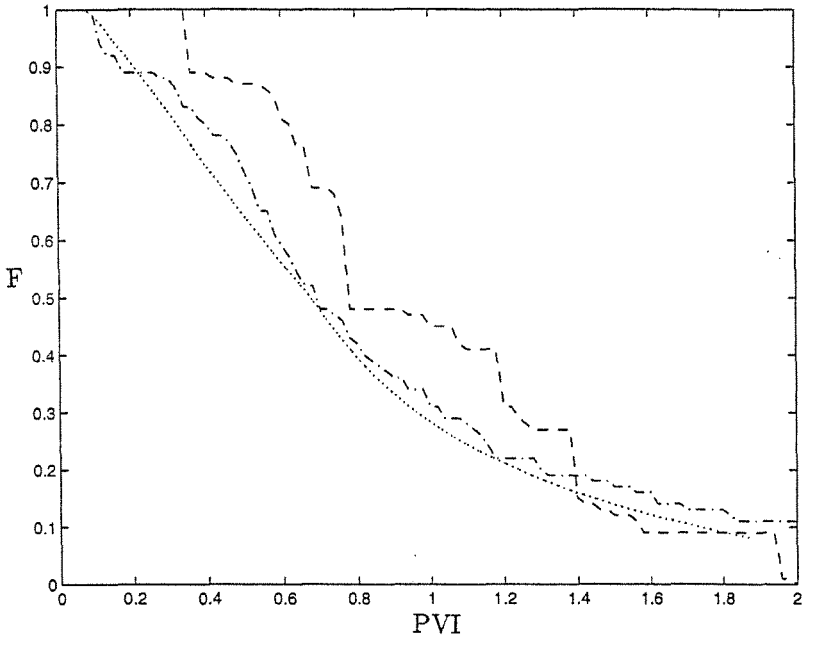

Figure 9.6: Case $l_x = 0.3$, $l_y = 0.01$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 10×10 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

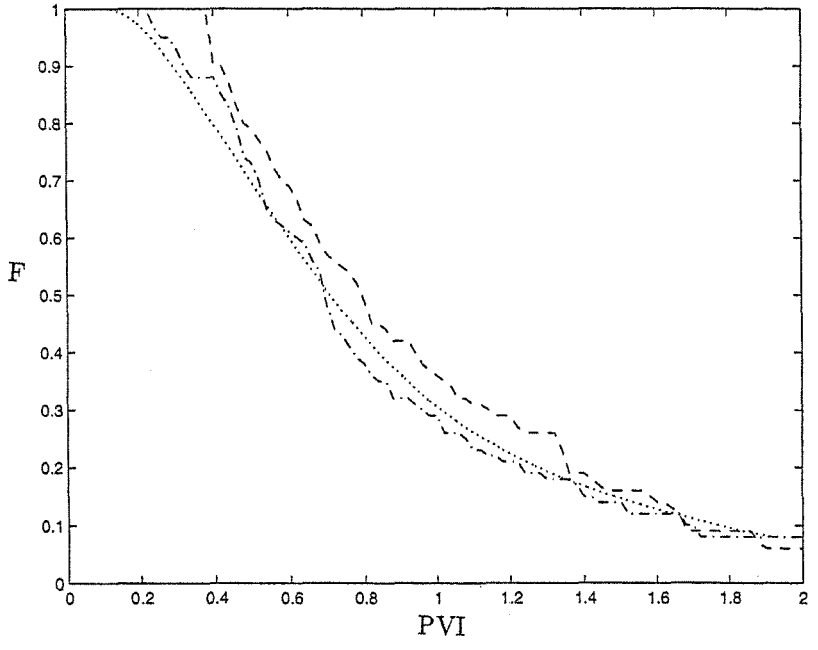

Figure 9.7: Case $l_x = 0.3$, $l_y = 0.05$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 11×11 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

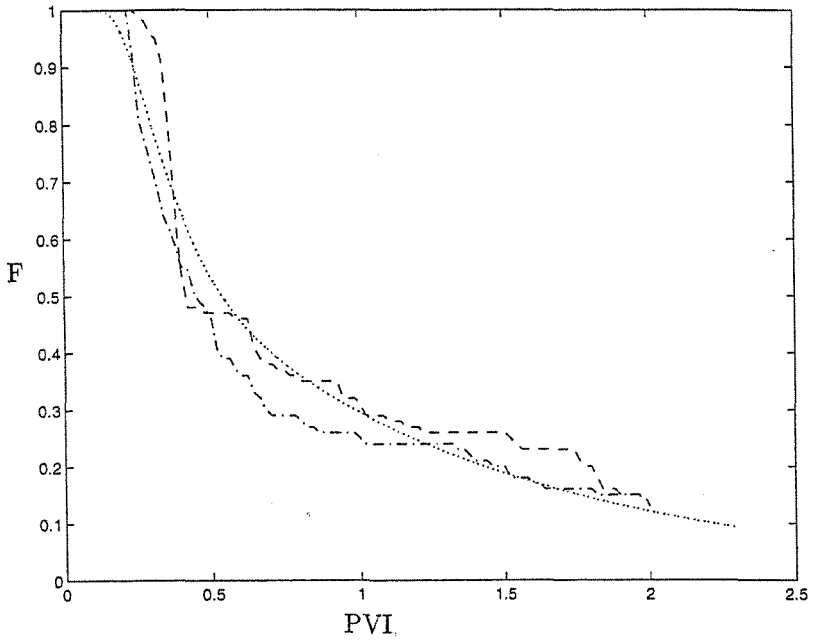

Figure 9.8: Case $l_x = 0.3$, $l_y = 0.2$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 9×9 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

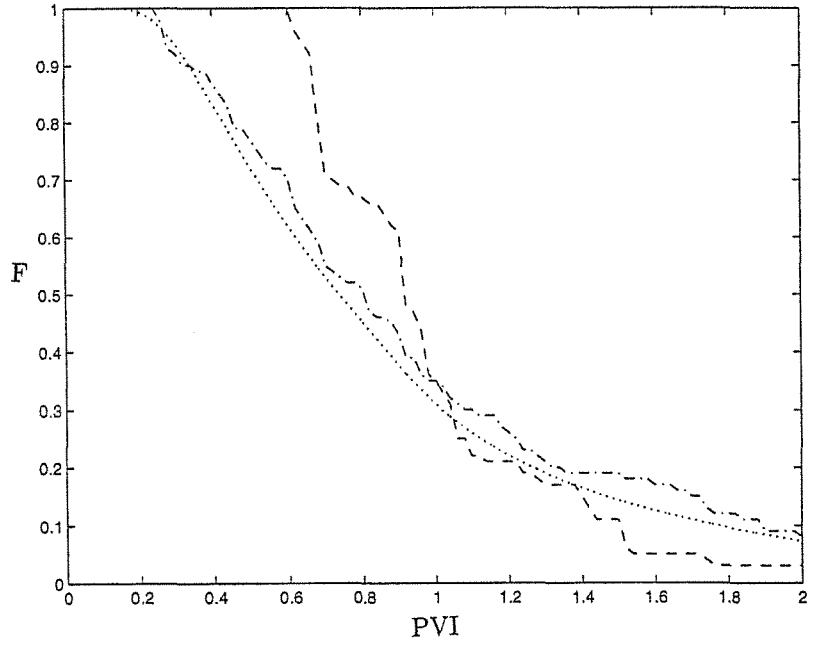

Figure 9.9: Case $l_x = 0.15$, $l_y = 0.005$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 10×10 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

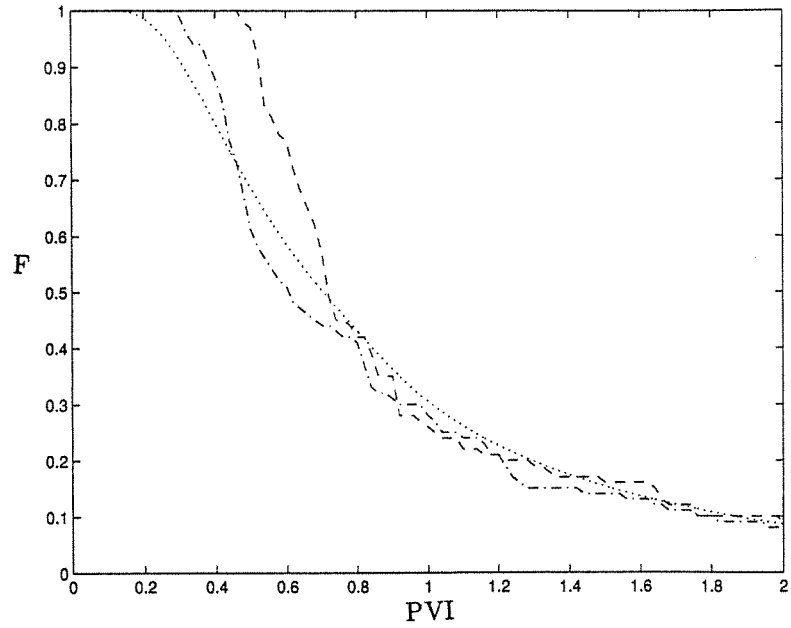

Figure 9.10: Case $l_x = 0.15$, $l_y = 0.025$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 10×10 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

As we discussed in the previous section, this model can be used to approximate the detailed flow. We are going to compare the results of this coarse model with the non-uniform coarsening model mentioned before. In the numerical example presented in Fig. 9.11, the velocity field is nearly layered. Such a high value for l_x is not commonly used in practical examples. As we see from fractional flow curves, the coarse model (9.42) has very good accuracy for this case. We also depict in this figure the fractional curve for 10×10 coarse model using the non-uniform coarsening approach. As we see from this figure our coarse model on 1×1 coarse grid gives better accuracy. In Fig. 9.12 we present a numerical result for the layered flow with moderate value of l_x, $l_x = 0.15$. We see that for this case also, the coarse model (9.42) gives a better approximation to the fractional flow curve of the detailed flow than the non-uniform coarsening on 10×10 coarse grid. If we increase the correlation length in the y direction then the performance of the coarse model (9.42) deteriorates (see Fig. 9.13).

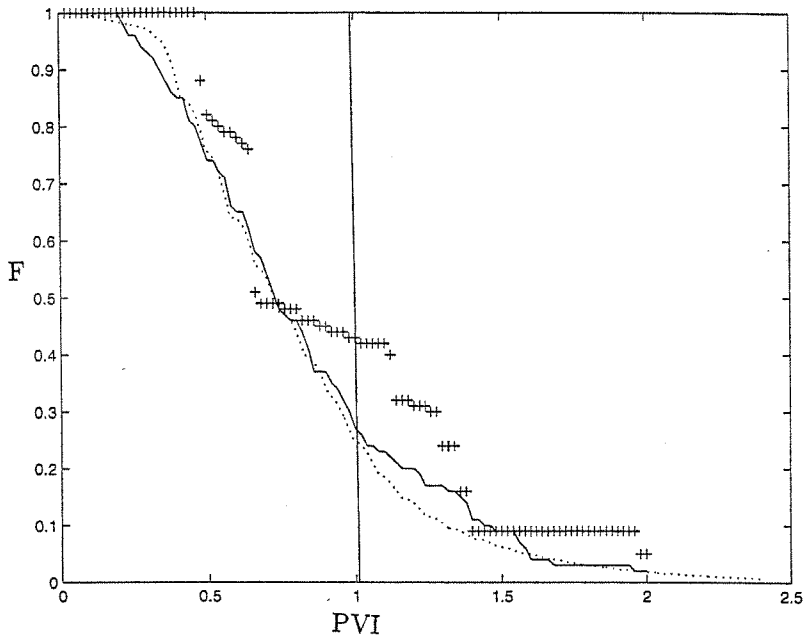

Figure 9.11: Case $l_x = 0.7$, $l_y = 0.005$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, '+' is for coarse model using non-uniform coarsening method on 10×10 coarse grid, dotted line is for the coarse model with diffusion on 1×1 coarse grid, solid line represents the coarse fractional flow curve on 1×1 coarse grid

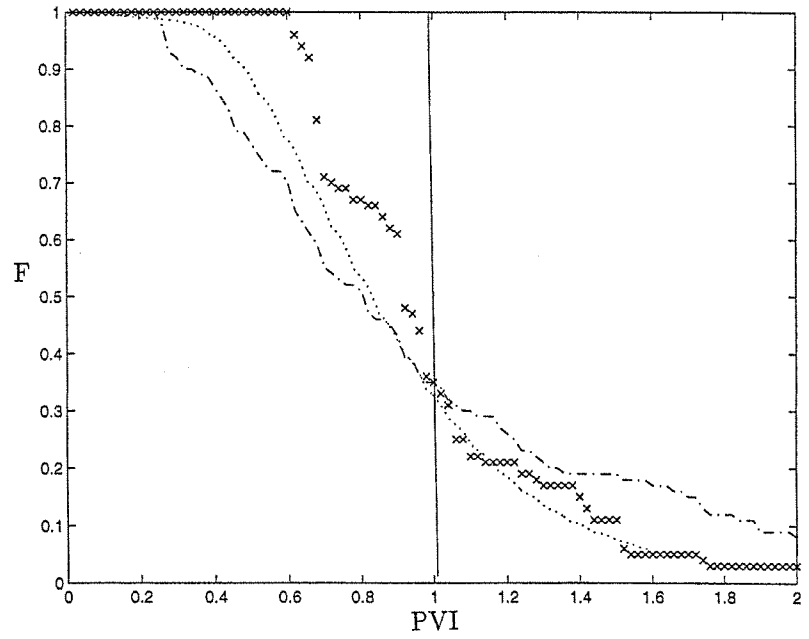

Figure 9.12: Case $l_x = 0.15$, $l_y = 0.005$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, '+' is for coarse model using non-uniform coarsening method on 10×10 coarse grid, dotted line is for the coarse model with diffusion on 1×1 coarse grid, solid line represents the coarse fractional flow curve on 1×1 coarse grid

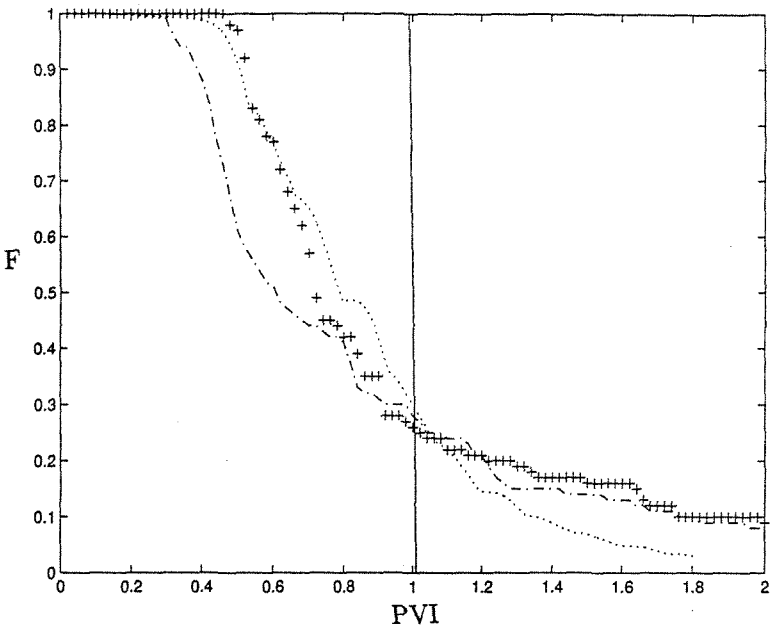

Figure 9.13: Case $l_x = 0.15$, $l_y = 0.025$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, '+' is for coarse model using non-uniform coarsening method on 10×10 coarse grid, dotted line is for the coarse model with diffusion on 1×1 coarse grid, solid line represents the coarse fractional flow curve on 1×1 coarse grid

9.12 Upscaling of two-phase flow

In this section we discuss a coarse model for two-phase flow. This model has been derived in [38, 37]. I have also derived this model independently during my summer internship at Chevron. The derivation of the model assumes that the velocity field weakly depends on time. This kind of assumption has been made in a number of publications. Indeed, the saturation dependence in the velocity has been examined before [38, 37] and it has been shown that it does not have a large influence on heterogeneous fingering for modest mobility ratio $\lambda(S)$ (9.2). Under this assumption, our fine scale model is

$$\frac{\partial S}{\partial t} + v_j \nabla_j f(S) = 0, \tag{9.43}$$

where $v_i = k^{ij}\nabla_j p$, and $\nabla_i k^{ij} \nabla_j p = 0$. The derivation of the coarse model is based on perturbation of saturation and the velocity fields around their averages. Using the expansion of these quantities as we did in the previous section

$$S = S_0 + S', \quad v_j = v_j^0 + v_j'$$

in the saturation equation, we have

$$\frac{\partial S_0}{\partial t} + \frac{\partial S'}{\partial t} + (v_j^0 + v_j')\nabla_j \left(f(S_0) + A(S_0)S' + B(S_0){S'}^2 \right) = 0. \tag{9.44}$$

Here

$$A(S_0) = f'(S_0), \quad B(S_0) = f''(S_0), \quad C(S_0) = f'''(S_0).$$

To get (9.44) we have used

$$f(S) = f(S_0 + S') = f(S_0) + A(S_0)S' + B(S_0){S'}^2.$$

(9.44) is equivalent to

$$\frac{\partial S_0}{\partial t} + \frac{\partial S'}{\partial t} + (v_j^0 + v_j') \left(A(S_0)\nabla_j S_0 + A(S_0)\nabla_j S' + S'B(S_0)\nabla_j S_0 + B(S_0)\nabla_j {S'}^2 + C(S_0){S'}^2 \nabla_j S_0 \right) = 0. \tag{9.45}$$

Averaging (9.45) and neglecting the terms of third and higher orders, we get

$$\frac{\partial S_0}{\partial t} + v_j^0 A(S_0)\nabla_j S_0 + A(S_0)\overline{v_j' \nabla_j S'} + B(S_0)\nabla_j S_0 \overline{v_j' S'} + v_j^0 B(S_0)\nabla_j \overline{{S'}^2} + v_j^0 \overline{{S'}^2} C(S_0)\nabla_j S_0 = 0. \tag{9.46}$$

This equation for averaged saturation is equivalent to

$$\frac{\partial S_0}{\partial t} + v_j^0 A(S_0)\nabla_j S_0 + \nabla_j A(S_0)\overline{v_j' S'} + \nabla_j v_j^0 B(S_0)\overline{{S'}^2} = 0. \tag{9.47}$$

In [17, 18] the author numerically investigated the effect of velocity-saturation covariance ($\overline{v_j' S'}$) and the variance of the saturation on the upscaled relative permeabilities. These upscaled relative permeabilities show the strongest dependence on the velocity-saturation covariance. For this reason we only concentrate on the effect of velocity-saturation covariance onto the average saturation. In [38, 37] the authors use a different argument for

neglecting the effects of saturation variance on the averaged properties of the flow. Then the approximate equation for averaged saturation becomes

$$\frac{\partial S_0}{\partial t} + v_j^0 A(S_0)\nabla_j S_0 + \nabla_j A(S_0)\overline{v_j' S'} = 0. \tag{9.48}$$

To approximate the velocity-saturation covariance, the fluctuating equation (i.e., equation for S') has been used. To derive this equation we subtract (9.46) from (9.45) and neglect the higher order terms:

$$\frac{\partial S'}{\partial t} + v_j^0 S' B(S_0)\nabla_j S_0 + v_j^0 A(S_0)\nabla_j S' + v_j' A(S_0)\nabla_j S_0 = A(S_0)\overline{v_j'\nabla_j S'}B(S_0)\nabla_j S_0\overline{v_j' S'} + v_j^0 B(S_0)\nabla_j \overline{S'^2} + v_j^0 \overline{S'^2} C(S_0)\nabla_j S_0. \tag{9.49}$$

Denoting the r.h.s of (9.49) as $\Phi(x,t)$, since $\Phi(x,t)$ is non-oscillatory function, we can write (9.49) in a more compact form:

$$\frac{\partial S'}{\partial t} + v_j^0 S' B(S_0)\nabla_j S_0 + v_j^0 A(S_0)\nabla_j S' + v_j' A(S_0)\nabla_j S_0 = \Phi(x,t). \tag{9.50}$$

In [38, 37] the authors using certain assumptions propose the following approximation for velocity-saturation covariance

$$\overline{S'(t,x)v_k'(x)} = -A(S_0(x,t))\int_0^t R^{jk}(x,x(\tau))d\tau\nabla_j S_0(x,t) \tag{9.51}$$

where $R^{jk}(x,x(\tau)) = \overline{v_j'(x)v_k'(x(\tau))}$, two point correlation of the velocity field as in the unit mobility case. With this approximation the equation for averaged saturation becomes

$$\frac{\partial S_0}{\partial t} + v_j^0 A(S_0)\nabla_j S_0 = \nabla_k A^2(S_0(x,t))\int_0^t R^{jk}(x,x(\tau))d\tau\nabla_j S_0(x,t). \tag{9.52}$$

This result is consistent with the unit mobility case where $A(S_0) = 1$. We use this model in our numerical examples. To calculate second order statistics, we use multiscale base functions. Even though to derive the coarse model (9.52) a number of assumptions have been used, the numerical experiments demonstrate that with this coarse model we can get substantial improvements over the existing non-uniform coarsening approach. Also as we

will see from numerical results of the next section the coarse model (9.52) is capable of approximating accurately the averaged flow properties on the relatively coarse grid.

9.13 Numerical results for two-phase flow

In this section we analyze numerically the coarse model (9.52) for two-phase flow assuming that the velocity field does not (weakly) depends on time. Second order ENO schemes have been used to approximate the nonlinear hyperbolic equations in order to avoid numerical diffusions. As for relative permeabilities, we use the commonly used flux for two-phase immiscible flow

$$k_{rw}(S) = (1-S)^2, \quad k_{ro}(S) = S^2$$

where S is the oil saturation. We assume the viscosity ratio to be 5, i.e., $\mu_o/\mu_w = 5$.

In Fig. 9.14 - Fig. 9.19, we compare the fractional flow curves for our coarse model with the non-uniform coarsening method on the same coarse grid for different realizations of permeability fields as we did in the unit mobility case.

As in the unit mobility case, we start with flow field with large correlation length in the x direction, $l_x = 0.6$. In Fig. 9.14 we present the numerical result for this case. As we see, our coarse model improves the predictions of non-uniform coarsening approach and gives an accurate approximation on 10×10 coarse grid.

In the next figures, we compare our method with non-uniform coarsening approach for various values of l_x and l_y. These values of l_x and l_y have been used in the numerical examples for the unit mobility case. For this reason we do not comment on the corresponding velocity fields. As we see from these figures, when the ratio of l_x/l_y increases the non-uniform coarsening approach becomes more robust. The reason is that in this case the velocity field loses its individual features. We see from the previous numerical results that our coarse model is capable of predicting the flow rate on 10×10 coarse grid. Moreover, the coarse model with diffusion improves the predictions of the non-uniform coarsening method.

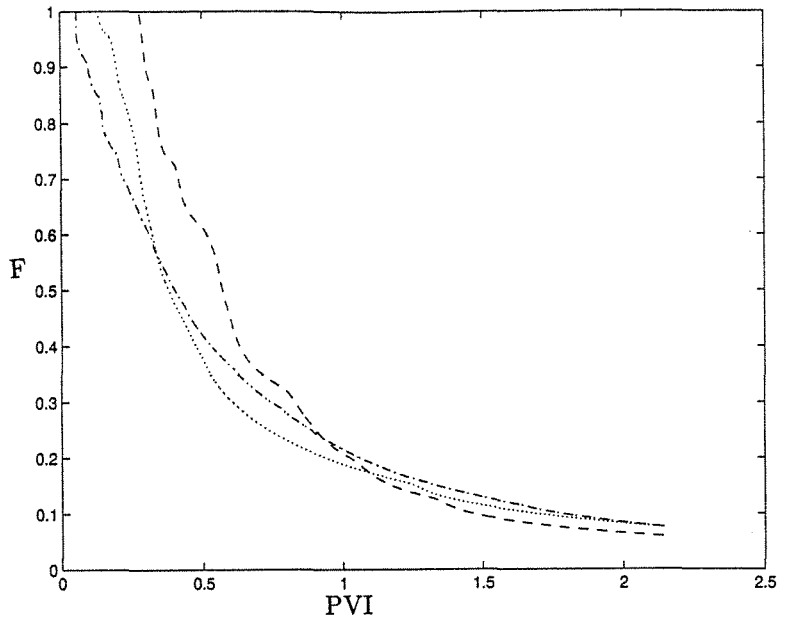

Figure 9.14: Case $l_x = 0.6$, $l_y = 0.01$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, Dashed-dashed line is for coarse model using non-uniform coarsening method on 10×10 coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

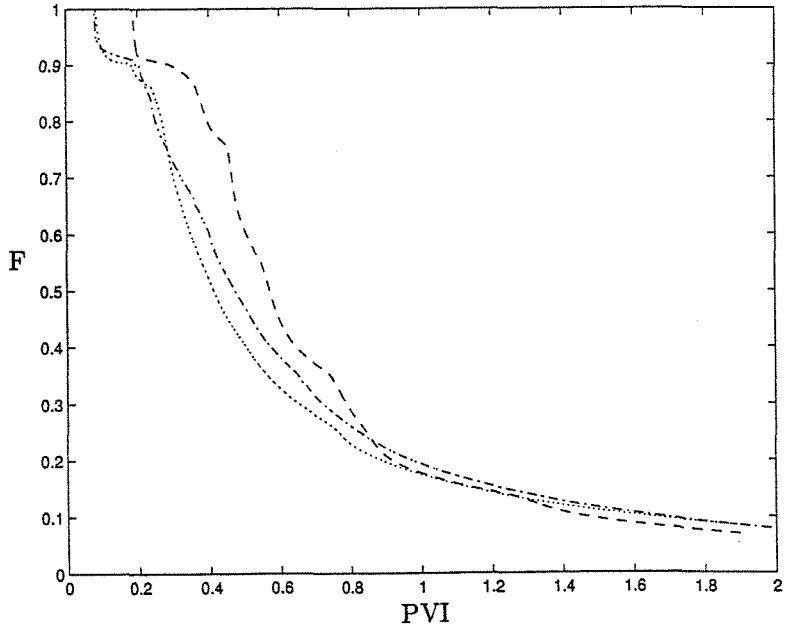

Figure 9.15: Case $l_x = 0.3$, $l_y = 0.005$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 12×12 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

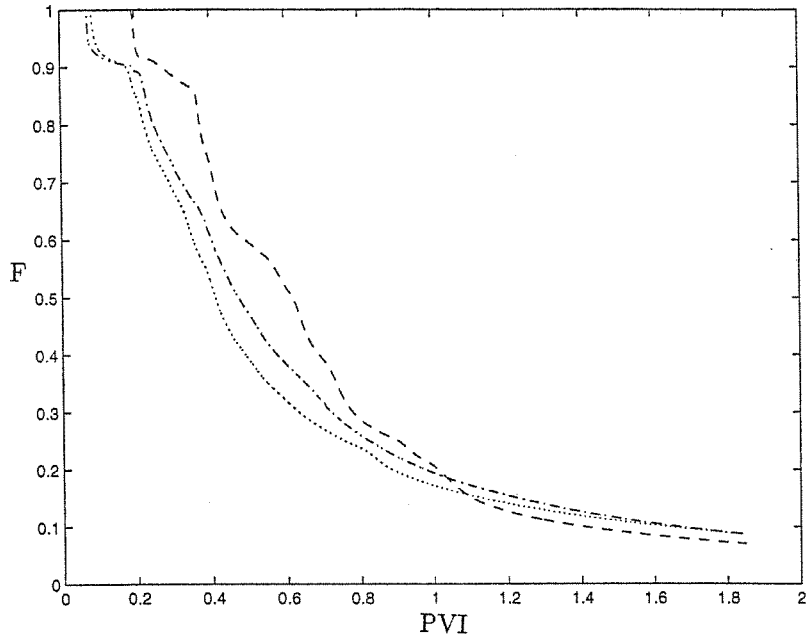

Figure 9.16: Case $l_x = 0.3$, $l_y = 0.01$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 10×10 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

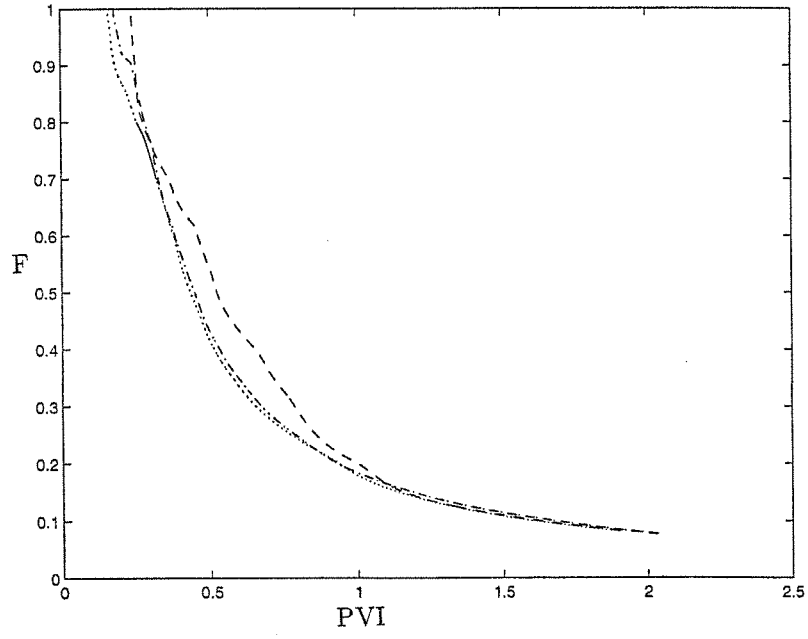

Figure 9.17: Case $l_x = 0.3$, $l_y = 0.05$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 11×11 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

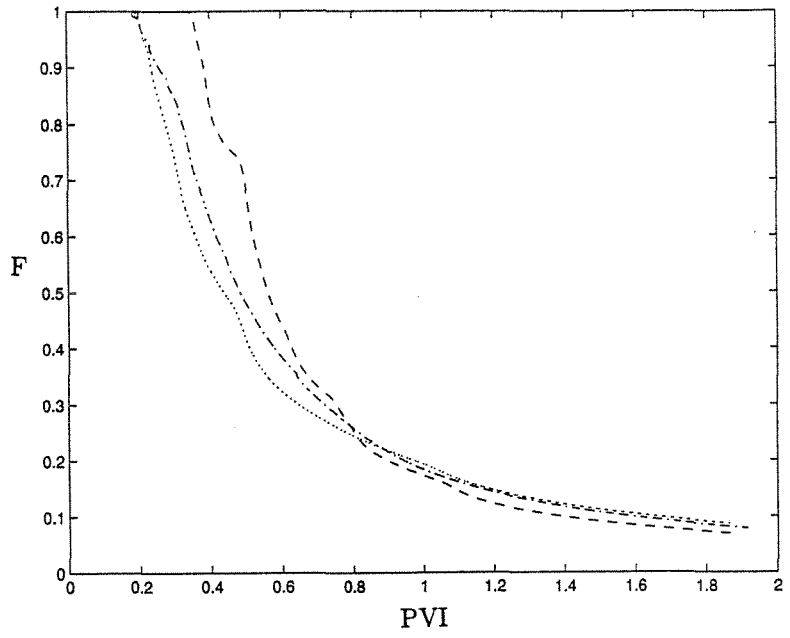

Figure 9.18: Case $l_x = 0.15$, $l_y = 0.005$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 10×10 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

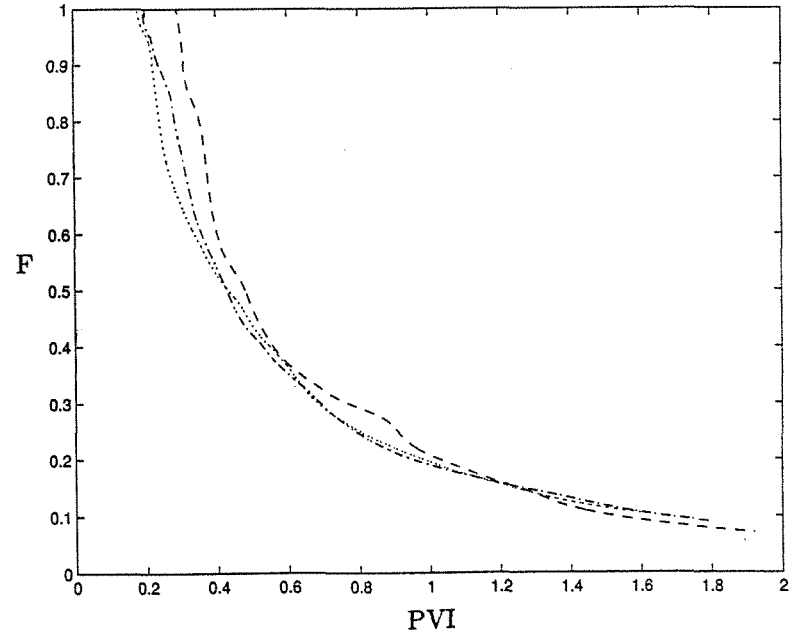

Figure 9.19: Case $l_x = 0.15$, $l_y = 0.025$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, dashed-dashed line is for coarse model using non-uniform coarsening method on 10×10 non-uniform coarse grid, dotted line is for the coarse model with diffusion on the same coarse grid

9.14 Concluding remarks

The implementation of the coarse model formulated in previous section to the general two-phase flow is currently under investigation.

Developing new effective coarse models for two-phase flow problem is one of my ongoing researches. Our goal is to find coarse models which are process independent and capable of predicting the flow properties on a relatively coarse grid. The main idea of the coarse models I am currently working on is to assign three velocity fields (for practical purposes) to each coarse block rather than one as in the non-uniform coarsening approach. In each coarse block, assigning three velocity values our goal is to represent the high, average and low flow areas. Let's note that for each velocity we also assign a weight, calculated from the MsFEM approximation of the velocity field. In the non-uniform coarsening approach, we just represent the average flow while missing the important features of the flow. Introducing the representatives for high and low velocities, we are capable to predict the average flow properties on a relatively coarse grid. In Fig. 9.20 - Fig. 9.22, we plot the fractional flow curves for the coarse model with three velocity representatives on 2×2 coarse grid. As we see from these figures, the predictions on a 2×2 coarse grid is better than a 5×5 coarse approximation of the fractional flow curves using non-uniform coarsening.

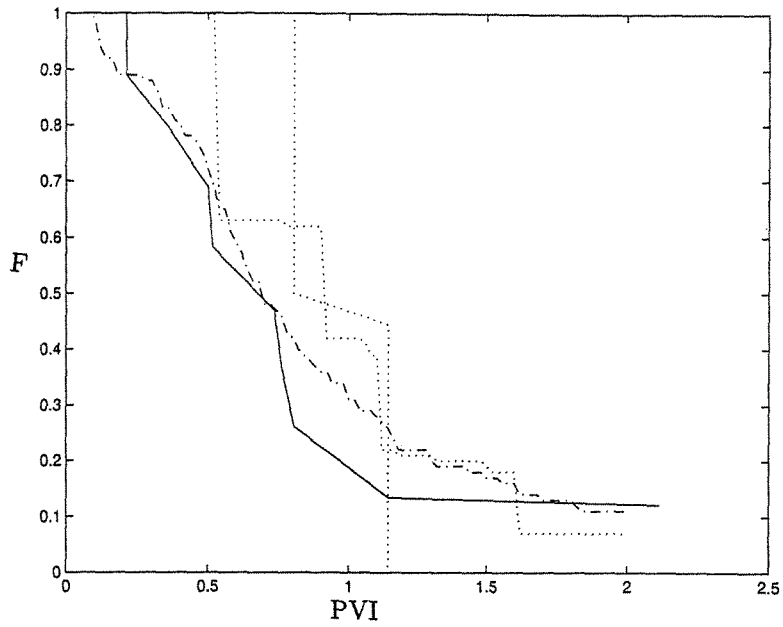

Figure 9.20: Case $l_x = 0.3$, $l_y = 0.01$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, solid line is for coarse model using three velocity values in each coarse block on 2×2 coarse grid, dotted lines is for the non-uniform coarsening approach on 2×2 and 5×5 coarse grid (2×2 is the worst one)

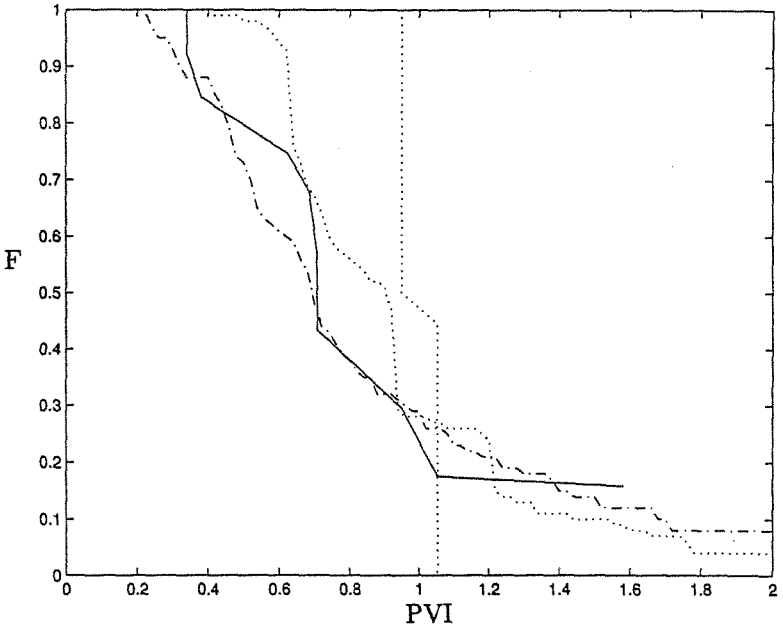

Figure 9.21: Case $l_x = 0.3$, $l_y = 0.05$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, solid line is for coarse model using three velocity values in each coarse block on 2×2 coarse grid, dotted lines is for the non-uniform coarsening approach on 2×2 and 5×5 coarse grid (2×2 is the worst one)

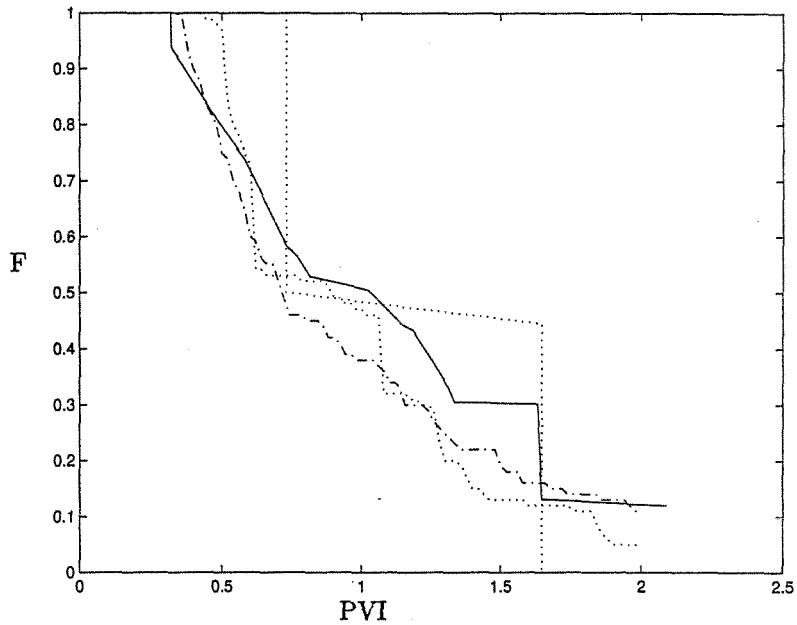

Figure 9.22: Case $l_x = 0.3$, $l_y = 0.2$, $\sigma = 2$. Dashed-dotted line is for fine fractional flow curve on 100×100 grid, solid line is for coarse model using three velocity values in each coarse block on 2×2 coarse grid, dotted lines is for the non-uniform coarsening approach on 2×2 and 5×5 coarse grid (2×2 is the worst one)

Appendix A Difference form of the cell term

To show that Λ_2^h can be written in the difference form we need just to prove that $\phi_{0,1}^k\phi_{0,1}^l$, $\phi_{0,2}^k\phi_{0,2}^l$ and $\phi_{0,1}^k\phi_{0,2}^l + \phi_{0,2}^k\phi_{0,1}^l$ are the same constants in each of two triangular elements with the common side kl. In this way the integrand is the same in the union of these two elements. Since the average of the integrand of Λ_2^h is zero, it is the divergence of a periodic field. Consequently, Λ_2^h can be written as a difference of integrals over the boundaries.

To show that these constants are the same for our configuration we need the following lemma.

Lemma A.0.1 *If $S_1 \bigcup S_2$ is a parallelogram then $\phi_{0,j}^k\phi_{0,p}^l + \phi_{0,j}^l\phi_{0,p}^k$ is the same in S_1 and S_2.*

Proof. In this case we have

$$\phi_{0,j}^l = -\phi_{0,j}^k \quad (j = 1, 2).$$

Note that the line segment kA is parallel to the line segment lB, which is also parallel to the line segment Ox, see Fig. A.1. Moreover, since $|kA| = |lB|$ and $\phi_0^k|_{S_1}(k) = 1$, $\phi_0^k|_{S_1}(A) = 0$, $\phi_0^l|_{S_2}(l) = 1$, $\phi_0^l|_{S_2}(B) = 0$, we have

$$\phi_{0,1}^k|_{S_1} = -\phi_{0,1}^l|_{S_2}.$$

The same can be shown for $j = 2$. □

Remark A.0.1. It can be shown that this difference structure leads to the summation by parts in (7.48).

Now let us show that $\int_K f\theta^k dx$ can be written in the difference form ($\int_K f\chi^p\phi_{0,p}^k dx$ is similar). Consider the configuration illustrated in Fig. A.2. By reordering the terms, we obtain

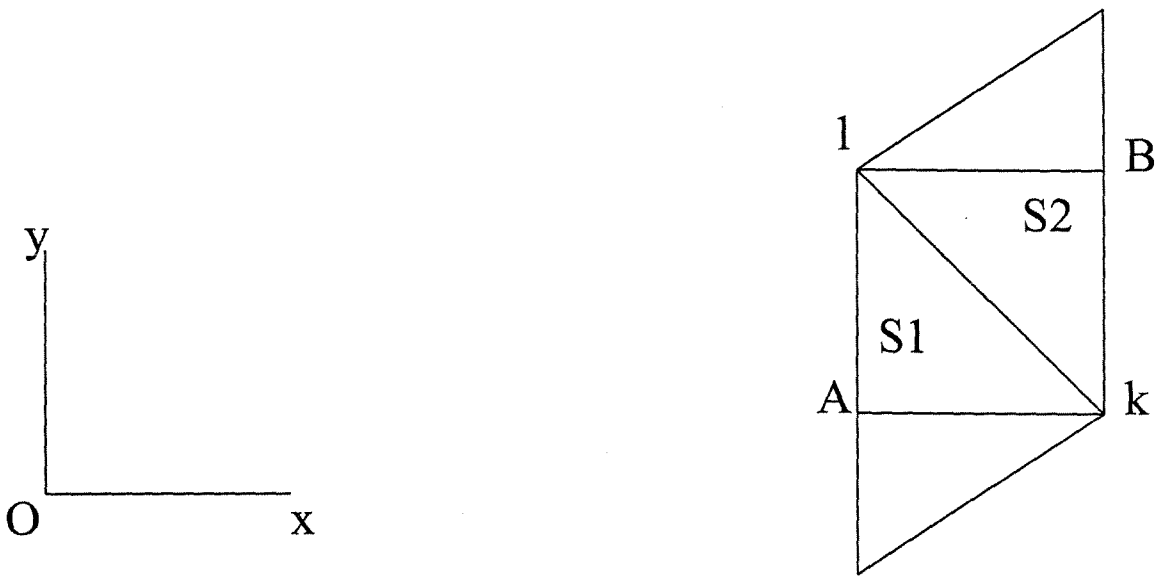

Figure A.1: Segments in a triangulation

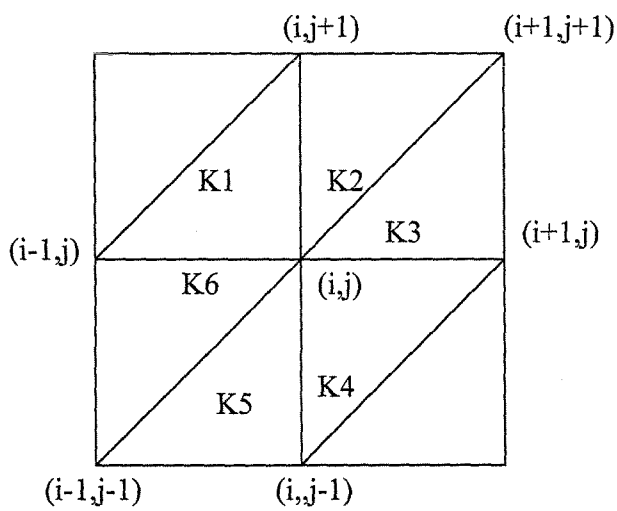

Figure A.2: Element nodes in a triangulation

$$\int f\theta^{ij}dx = \sum_{i=1}^{6}\int_{K_i} f\theta^{ij}dx = (\int_{K_5} f\theta^{ij}dx - \int_{K_1} f\theta^{ij+1}dx)$$
$$+(\int_{K_3} f\theta^{ij}dx - \int_{K_1} f\theta^{i-1j}dx) + (\int_{K_4} f\theta^{ij}dx - \int_{K_6} f\theta^{i-1j}dx) \qquad \text{(A.1)}$$
$$+(\int_{K_2} f\theta^{ij}dx - \int_{K_6} f\theta^{i-1j-1}dx).$$

Here we have used $\sum_{k=1}^{3}\theta^k = 0$ on $K \in \mathbf{K}^h$. This difference structure leads to the summation by parts in $G^h f_1^h$.

Appendix B The estimate for linear functions

In this section we show that on any triangular element K

$$\|\nabla l\|_{L_2(K)} \leq C\|\nabla u\|_{L_2(K)}, \tag{B.1}$$

where l is a linear function and u satisfies the following equation on a triangular domain $S \supset K$

$$\begin{aligned} \nabla_i a^{ij} \nabla_j u = 0 \quad & in \quad S \\ u = l \quad & on \ \partial S. \end{aligned} \tag{B.2}$$

Let us consider the difference of u and l: $\phi = u - l$. Clearly ϕ satisfies the following equation:

$$\begin{aligned} \nabla_i a^{ij} \nabla_j \phi = -\nabla_i a^{ij} \nabla_j l \quad & in \quad S \\ \phi = 0 \quad & on \ \partial S. \end{aligned} \tag{B.3}$$

Introducing the auxiliary function v_i $(i = 1, 2)$ defined by

$$\begin{aligned} \nabla_i a^{ij} \nabla_j v_k = -\nabla_i a^{ik} \quad & in \quad S, \\ v_k = 0 \quad & on \ \partial S. \end{aligned} \tag{B.4}$$

We can express the solution of (B.3) as a linear combination of v_i. This gives

$$u = l + v_i \nabla_i l = l + v_i \alpha_i,$$

where $\alpha_i = \nabla_i l$ are constants. Then we obtain

$$\begin{aligned} \int_K (\nabla u)^2 dx = & \alpha_1^2 \int_K [(1 + \nabla_1 v_1)^2 + (\nabla_2 v_1)^2] dx + \\ & \alpha_2^2 \int_K [(1 + \nabla_2 v_2)^2 + (\nabla_1 v_2)^2] dx + \\ & 2\alpha_1 \alpha_2 \int_K [(1 + \nabla_1 v_1)\nabla_1 v_2 + (1 + \nabla_2 v_2)\nabla_2 v_1] dx = (A\alpha, \alpha), \end{aligned}$$

where A is the matrix with the following elements

$$A^{11} = \int_K [(1+\nabla_1 v_1)^2 + (\nabla_2 v_1)^2]dx,$$
$$A^{12} = \int_K [(1+\nabla_1 v_1)\nabla_1 v_2 + (1+\nabla_2 v_2)\nabla_2 v_1]dx,$$
$$A^{21} = \int_K [(1+\nabla_1 v_1)\nabla_1 v_2 + (1+\nabla_2 v_2)\nabla_2 v_1]dx,$$
$$A^{22} = \int_K [(1+\nabla_2 v_2)^2 + (\nabla_1 v_2)^2]dx.$$

It can be checked that

$$A^{11} = \int_K [(1+\nabla_1 v_1)^2 + (\nabla_2 v_1)^2]dx > 0,$$

$$A^{22} = \int_K [(1+\nabla_2 v_2)^2 + (\nabla_1 v_2)^2]dx > 0,$$

and

$$(A\alpha, \alpha) > 0$$

are the sufficient conditions for $(A\alpha, \alpha) \geq d\alpha^2$, for some $d > 0$. For example under these conditions d can be chosen to be $2d = A^{11} + A^{22} - \sqrt{(A^{11} - A^{22})^2 + 4(A^{12})^2}$. Note that if $d > 0$ then $d > Ch^2$ from which it follows that $d\alpha^2 \geq C\|\nabla l\|_{L_2(K)}$. Assuming the opposite, i.e., any of the inequalities does not hold we have that $v_1 + x_1$ or $v_2 + x_2$ or u is constant in K. Let's note that $v_i + x_i$ $(i = 1, 2)$ satisfy

$$\begin{aligned} \nabla_i a^{ij} \nabla_j (v_k + x_k) &= 0 \quad in \;\; S, \\ v_k + x_k &= x_k \quad on \; \partial S. \end{aligned} \tag{B.5}$$

Therefore $v_i + x_i$ $(i = 1, 2)$ cannot be constant in K [39]. Consequently, A^{11} and A^{22} are strictly positive. Also $\int_K (\nabla u)^2 dx > 0$ if $|\nabla l| > 0$ [39] which guarantees $(A\alpha, \alpha) > 0$. If $|\nabla l| = 0$ then (B.1) satisfies. This completes the proof of (B.1).

Appendix C Formulation based on dissipation energy

Indelman and Dagan [30] suggested the use of averaged dissipation energy for determining the grid block permeability, i.e.,

$$\langle \nabla p \cdot \tilde{K} \nabla p \rangle_V = \langle \nabla p^\epsilon \cdot K^\epsilon \nabla p^\epsilon \rangle_V, \tag{C.1}$$

where p^ϵ is the solution of (8.5) and p is the solution of

$$\nabla \cdot \tilde{K} \nabla p = 0 \qquad \text{in } V. \tag{C.2}$$

This formulation may be viewed as an approximation to the energy convergence in the homogenization theory (cf. [32]). Note that $\tilde{K}$ cannot be uniquely determined from (C.1) since adding any anti-symmetric tensor to $\tilde{K}$ does not change the equality. Thus, we enforce $\tilde{K}$ to be symmetric.

Equation (C.1) is useful for calculating $\tilde{K}$ only when ∇p is known in advance. This can be achieved by specifying special boundary conditions. Let $p = w + e \cdot x$ be the solution of (C.2), where e is a constant vector. Then under the condition $w = 0$ on ∂V or w being periodic in V, we have $\nabla p = e$ on V from (C.2) since $\tilde{K}$ is a constant tensor in V. Thus (C.1) reduces to

$$e \cdot \tilde{K} e = \langle \nabla p^\epsilon \cdot K^\epsilon \nabla p^\epsilon \rangle_V. \tag{C.3}$$

This explicit formula is in fact equivalent to (8.13). We briefly outline the proof here.

First, because (C.3) holds for arbitrary e, choosing $e = e_i$ $(i = 1, \ldots, d)$ and denoting the corresponding p^ϵ by p_i^ϵ, we obtain (8.13) for $i = j$. Now, choose $e = e_i + e_j$ $(i \neq j)$. By using the symmetry of $\tilde{K}$ (as enforced) and K^ϵ, as well as the previous result for $i = j$, it is easy to show that (8.13) holds for $i \neq j$. On the other hand, since any vector e can be written as a linear combination of e_i, we obtain (C.3) from (8.13) by simple algebra. For this, we use the facts that (8.5) is linear and homogeneous, (8.7) (or (8.8)) is invariant under linear superposition, and that the two sides of (8.13) are bilinear forms.

With the boundary condition (8.9), $\nabla p = e$ only when $\tilde{K}$ is diagonal; otherwise, ∇p and $\tilde{K}$ are coupled together and (C.1) could be difficult to use in actual computations. We mention that reference [10] showed the above equivalence under the periodic boundary condition with a different approach, but the conclusion for the linear pressure drop condition (corresponding to $w = 0$ on ∂V) was incorrect.

Bibliography

[1] S. Agmon. *Lectures on elliptic boundary value problems.* Van Nostrand, Princeton, N.J., 1965.

[2] G. Allaire and M. Briane. Multiscale convergence and reiterated homogenization. *Proceedings of the royal society of Edinburgh A Mathematics*, 126:297–342, 1996.

[3] M. Avellaneda and F-H. Lin. Compactness methods in the theory of homogenization. *Comm. Pure Appl. Math.*, XL:803–843, 1989.

[4] A. Azzam. Smoothness properties of bounded solution of dirichlet problem for elliptic equations in regions with corners on the boundaries. *Canadian Mathematical Bulletin*, 23, 1980.

[5] I. Babuška, G. Caloz, and E. Osborn. Special finite element methods for a class of second order elliptic problems with rough coefficients. *SIAM J. Numer. Anal.*, 31:945–981, 1994.

[6] J. Bear. *Dynamics of Fluids in Porous Media.* Elsevier, 1972.

[7] A. Bensoussan, J. L. Lion, and G. Papanicolaou. Boundary layer analysis in homogenization of diffusion equations with dirichlet conditions on the half space. *Proceedings of the International Symposium on Stochastic Differential equations*, pages 21–40, 1976.

[8] A. Bensoussan, J. L. Lion, and G. Papanicolaou. *Asymptotic Analysis for Periodic Structure*, volume 5 of *Studies in Mathematics and Its Applications.* North-Holland Publ., 1978.

[9] A. Bensoussan, J.L. Lions, and G. Papanicolaou. Boundary layer analysis in homogenization of diffusion equations with dirichlet conditions in the half space. In Proceedings of the International Symposium on Stochastic Differential Equations, Kyota 1976, K.Ito, ed., Willey 1978.

[10] O. Boe. Analysis of an upscaling method based on conservation of dissipation. *Transport in Porous Media*, pages 77–86, 1994.

[11] F. Brezzi, L. Franca, T. Hughes, and A. Russo. b is equal to integral of g. *Comput. Methods Appl. Mech. Engrg*, 145:329–339, 1997.

[12] Graham Carey and Tinsley Oden. *Finite Elements: A second course. Volume 2.* Prentice-Hall, 1983.

[13] P. G. Ciarlet. *The finite element method for elliptic problems.* Horth-Holland, Amsterdam, 1978.

[14] G. Dagan. Solute transport in heterogeneous porous formations. *J. Fluid Mech.*, 145:151, 1984.

[15] L. J. Durlofsky. Numerical-calculation of equivalent grid block permeability tensors for heterogeneous porous media. *Water Resour. Res.*, 27:699–708, 1991.

[16] L. J. Durlofsky. Representation of grid block permeability in coarse scale models of randomly heterogeneous porous-media. *Water Resour. Res.*, 28:1791–1800, 1992.

[17] L. J. Durlofsky. Use of higher moments for the description of upscaled, process independent relative permeabilities. *SPE paper 37987*, 1997.

[18] L. J. Durlofsky. Coarse scale models of two phase flow in heterogeneous reservoirs: Volume averaged equations and their relationship to the existing upscaling techniques. *Computational Geosciences*, 1999. to appear.

[19] L. J. Durlofsky, R. C. Jones, and W. J. Milliken. A nonuniform coarsening approach for the scale up of displacement processes in heterogeneous media. *Advances in Water Resources*, 20:335–347, 1997.

[20] Weinan E. Homogenization of linear and nonlinear transport equations. XLV:301–326, 1992.

[21] Y. R. Efendiev, T. Y. Hou, and X. H. Wu. The convergence of non-conforming multiscale finite element methods. 1998. submitted to SIAM in Num. Anal.

[22] L. C. Evans. Weak convergence methods for nonlinear partial differential equations. Volume 74 of regional conference series in mathematics, AMS, Providence, R.I., 1990.

[23] J. Frehse and R. Rannacher. Eine l^1-fehlerabschätzung für diskrete Grundlösungen in der Methode der finiten Elemente. In J. Frehse, editor, *Finite Elemente*, number 89 in Bonn. Math. Schrift., pages 92–114. 1975.

[24] J. J. Gomez-Hernandez. A stochastic approach to the simulation of block conductivity fileds conditioned upon data measured ata smaller scale. Ph.D. thesis, Stanford University, 1991.

[25] R. Hoeksema and P. Kitanidis. Analysis of the spatial structure of properties of selected aquifers. *Water Resour. Res.*, 21:563–572, 1985.

[26] T. Y. Hou and X. H. Wu. A multiscale finite element method for elliptic problems in composite materials and porous media. *Journal of Computational Physics*, 134:169–189, 1997.

[27] T. Y. Hou and X. H. Wu. A multiscale finite element method for PDEs with oscillatory coefficients. In *Numerical Treatment of Multi-Scale Problem*. Vieweg-Verlag, 1997. to appear.

[28] T. Y. Hou, X. H. Wu, and Z. Cai. Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients. 1996. submitted to Math. Comp.

[29] T. Hughes, G. Feijoo, L. Mazzei, and J. Quincy. The variational multiscale method - a paradigm for computational mechanics. *Comput. Methods Appl. Mech. Engrg*, 166:3–24, 1998.

[30] P. Indelman and G. Dagan. Upscaling of permeability of anisotropic heterogeneous formations 1. *Water Resour. Res.*, 26(6):917–923, 1993.

[31] E.H. Isaaks and R. Mohan Srivastava. *An introduction to applied geostatistics*. Oxford University Press, 1989.

[32] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik. *Homogenization of differential operators and integral functionals*. Springer-Verlag, 1994.

[33] P. K. Kitanidis. Effective hydraulic conductivity for gradually varying flow. *Water Resour. Res.*, 26:1197–1208, 1990.

[34] S. M. Kozlov. Averaging differential operators with almost periodic, rapidly oscillating coefficients. *Math. USSR Sbornik*, 35:481–498, 1978.

[35] S.M. Kozlov. Averaging of random operators. *Math. USSR Sbornik*, 37:167–180, 1980.

[36] O. A. Ladyzhenskaia and N. N. Uraltseva. *Linear and quasilinear elliptic eqautions.* Academic Press, New York, 1968.

[37] P. Langlo. Macrodispersion for two-phase, immisible flow in heterogeneous media. 1992.

[38] P. Langlo and M. Espedal. Macrodispersion for two-phase, immisible flow in porous media. *Advances in Water Resources*, 17:297–316, 1994.

[39] F-H. Lin. Nodal sets of solutions of elliptic and parabolic equations. *Comm. Pure Appl. Math.*, XLIV:287–308, 1991.

[40] S. Moskow and M. Vogelius. First order corrections to the homogenized eigenvalues of a periodic composite medium. *Proc. Roy. Soc. Edinburgh.* submitted.

[41] S. Neuman and J. Depner. Use of variable-scale pressure test data to estimate the log hydraulic conductivity covariance and dispersivity of fractured granites near oracle, arizona. *Journal of Hydrology*, 102:475–501, 1988.

[42] E. Oriordan and M. Stynes. A globally uniformly convergent finite element method for a singularly perturbed elliptic problem in 2 dimensions. *Math. Comp.*, 57:47–62, 1991.

[43] G. Papanicolaou and S. R. S. Varadhan. Boundary value problems with rapid oscillating random coefficients. *Seria Colloquia Mathematica Societatis Janos Bolyai*, 27:835–873, 1981.

[44] Ph. Renard and G. de Marsily. Calculating effective permeability: a review. *Advances in Water Resources*, 20:253–278, 1997.

[45] Y. Rubin and J.J. Gomez-Hernandez. A stochastic approach to the problem of upscaling of conductivity in disordered media: Theory and unconditional numerical simulations. *Water Resor. Res.*, 26(4):691–701, 1990.

[46] A. Saez, C. J. Otero, and I. Rusinek. The effective homogeneous behavior of heterogeneous media. *Transport in Porous Media*, pages 213–238, 1989.

[47] R. Scott. Interpolated boundary conditions in the finite element method. *SIAM J. Numerical Analysis*, 12:404–427, 1975.

[48] R. Scott. Optimal l^∞ estimates for the finite element method on irregular meshes. *Math. Comp.*, 30:681–697, 1976.

[49] G. Strang and G. J. Fix. *An analysis of the finite element.* Prentice-Hall, 1973.

[50] L. Tartar. Nonlocal effects induced by homogenization. *PDE and calculus of variations*, pages 925–938, 1989. Boston.

[51] X. H. Wen and J.J. Gomez-Hernandez. Upscaling hydraulic conductivities in heterogeneous media. *Journal of Hydrology*, 183:9–32, 1996.

[52] C. D. White and R. N. Horne. Computing absolute transmissivity in the presence of fine scale heterogeneity. *SPE*, 16011:209–221, 1987.

[53] X. H. Wu and T. Y. Hou. Numerical simulation of two phase flow in porous media using a multiscale finite element method. 1997. in preparation.

[54] L. Ying. *Infinite Element Methods.* Peking University Press, Beijing, 1995.

[55] L. A. Ying. Infinite element method for elliptic problems. *Science in China*, 14:1438–1447, 1991.

[56] V.V. Yurinskii. Averaging of symmetric diffusion in random media. *Sibirskii Mathematicheskii Zhurnal*, 27:167–180, 1984.

[57] Qiang Zhang. The asymptotic scaling behavior of mixing induced by a random velocity field. *Advances in Applied Math*, 16:23–58, 1995.

[58] V. V. Zhikov, S. M. Kozlov, O. A. Oleinik, and Kha T'en Ngoan. Averaging and g-convergence of differential operators. *Russian Math. Surveys*, 34:69–147, 1979.