COMPUTATIONAL UPSCALED MODELING OF HETEROGENEOUS POROUS MEDIA FLOW UTILIZING FINITE VOLUME METHOD

A Dissertation

by

VICTOR ERALINGGA GINTING

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2004

Major Subject: Mathematics

COMPUTATIONAL UPSCALED MODELING OF HETEROGENEOUS POROUS MEDIA FLOW UTILIZING FINITE VOLUME METHOD

A Dissertation

by

VICTOR ERALINGGA GINTING

Submitted to Texas A&M University
in partial fulfillment of the requirements
for the degree of

DOCTOR OF PHILOSOPHY

Approved as to style and content by:

Raytcho Lazarov (Co-Chair of Committee)	Yalchin Efendiev (Co-Chair of Committee)
Richard Ewing (Member)	Joseph Pasciak (Member)
Akhil Datta-Gupta (Member)	Al Boggess (Head of Department)

May 2004

Major Subject: Mathematics

ABSTRACT

Computational Upscaled Modeling of Heterogeneous Porous Media Flow

Utilizing Finite Volume Method. (May 2004)

Victor Eralingga Ginting, B.S., Institute of Technology Bandung, Indonesia;

M.S., Texas A&M University

Co–Chairs of Advisory Committee: Dr. Raytcho Lazarov
Dr. Yalchin Efendiev

In this dissertation we develop and analyze numerical method to solve general elliptic boundary value problems with many scales. The numerical method presented is intended to capture the small scales effect on the large scale solution without resolving the small scale details, which is done through the construction of a multiscale map. The multiscale method is more effective when the coarse element size is larger than the small scale length. To guarantee a numerical conservation, a finite volume element method is used to construct the global problem.

Analysis of the multiscale method is separately done for cases of linear and nonlinear coefficients. For linear coefficients, the multiscale finite volume element method is viewed as a perturbation of multiscale finite element method. The analysis uses substantially the existing finite element results and techniques. The multiscale method for nonlinear coefficients will be analyzed in the finite element sense. A class of correctors corresponding to the multiscale method will be discussed. In turn, the analysis will rely on approximation properties of this correctors. Several numerical experiments verifying the theoretical results will be given.

Finally we will present several applications of the multiscale method in the flow

in porous media. Problems that we will consider are multiphase immiscible flow, multicomponent miscible flow, and soil infiltration in saturated/unsaturated flow.

This dissertation is dedicated to my wonderful parents,

Elfrida, Sri, Olga, Isca,

and in loving memory of Marge Haislet

ACKNOWLEDGMENTS

I would like to express my sincerest thanks to my advisors, Dr. Raytcho Lazarov and Dr. Yalchin Efendiev, for their guidance and thoughtfulness during the course of my studies at Texas A&M University. Dr. Lazarov confidently invited me to join the numerical analysis group despite my lack of mathematical background. He was always patient and constantly encouraged me to strive for better progress. Dr. Efendiev introduced me to the realm of multiscale modeling. He gladly shared many interesting ideas and helped me gain invaluable insights. In addition, I would like to thank Dr. Richard Ewing, Dr. Joseph Pasciak, and Dr. Akhil Datta-Gupta for serving on the committee and for their input and suggestions.

I thank Dr. Panagiotis Chatzipantelidis for many useful discussions. I am grateful to Dr. Wen Chen, Dr. Hamdi Tchelepi, and Dr. Liyong Li for giving me the opportunity for an internship at ChevronTexaco Exploration and Production Technology Company, San Ramon, CA during the summer of 2002.

I wish to express my gratitude to the Department of Mathematics, especially to Dr. Jay Walton and Dr. Thomas Schlumprecht, who were departmental graduate advisors during my course of study. Also, thanks to Ms. Monique Stewart. I thank all my friends in the Department of Mathematics and in the numerical analysis group for their friendship and memories.

I would like to acknowledge the financial support that I received during my study. I am grateful to the Department of Mathematics and to the Institute for Scientific Computation, Texas A&M University for providing my graduate teaching and research assistantships throughout my study. This research was supported in part by the National Science Foundation under award numbers DMS-0327713 and DMS-0218229.

TABLE OF CONTENTS

CHAPTER Page

LIST OF TABLES

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

Most important flow and transport problems in porous media involve processes that occur over wide range of length and time scales. The numerical modeling of coarse scale features of such processes forces the researchers to understand the behavior and coupling of various physical, chemical and biological processes on different length and time scales. This coupling is often more complicated by the appearance of additional phases, species, and uncertainties.

Subsurface formations typically exhibit heterogeneities over a wide range of length scales. Laboratory studies are performed that can characterize rock samples at the micron scale. Flow experiments on these samples may give permeability estimates for core plugs of two inches long. Indirect measurements of reservoir properties give data that varies on a scale of approximately a foot (well logs), tens to hundreds of feet (well tests) and hundreds to thousands of feet (tracer tests, interference tests, production data). Geophysical (seismic) data and geological data can characterize the basins that reservoirs and aquifers are found in over scales of several miles. Flow models are constructed which aim to honor available data. It is very desirable if flow simulations can honor as many of the scales underlying the available data as possible.

For this reason, some type of coarsening, or upscaling, of the detailed geologic model must be performed before the model can be used for flow simulation. The upscaling is in general nontrivial because heterogeneities at all scales have a significant effect, and these must be captured in the coarsened subsurface description. For example, in multi-fluid systems, fluid-fluid interfaces at the pore scale support pressure

---

This dissertation follows the style and format of *SIAM Journal of Numerical Analysis*.

differences between the fluids systems which lead to larger scale capillary pressure concepts; in contaminant transport problems, biofilms that exist at the pore scale can significantly alter the overall transport of contaminants, leading to nonlinear reaction terms at the larger scales; and geological heterogeneities arising from natural depositional processes lead to variations in larger scale parameters at the field scale. For these, and many other examples, the underlying physical, chemical, and biological processes that ultimately determine the fate of subsurface fluids and contaminants occur over the wide range of scales. To simulate such processes efficiently one needs approaches that can capture the effects of small scales on the large ones.

Besides utilizing the parallel computing technology, there have been significant efforts to develop methods of obtaining effective parameters that are defined on coarser models. This is also in conjunction with engineering work that often requires only the knowledge of the processes on the large scale. Among the many literatures that deal with these issues are [14, 18, 19, 32, 61] and references cited therein. In general, the more simplified mathematical descriptions were motivated by homogenization theory (see, e.g. [45]).

A somewhat new direction in tackling this problem has been developed recently (cf. e.g. [1, 3, 11, 23, 42, 43, 44]). The numerical methods presented in these works have the ability to capture the small scales effect on the large scale solution without resolving the small scale details. In [42, 43] for example, this is implemented by devising the so called oscillatory basis functions which are incorporated into the finite element formulation on the coarse grid, hence the name multiscale finite element method. The basis functions serve as the building block of all small scales structures inherited from the original problem, so that they are set to satisfy the leading order homogeneous elliptic equation in each coarse element. It should be noted that the effectiveness of the multiscale finite element method is more significant when the

coarse element size is substantially larger than the small scale length.

In addition to capturing small scale effects on the large one, many engineering and physical applications such as those arising in the petroleum reservoir simulations, groundwater hydrology and environmental remediation, desire to develop numerical methods that have certain conservation features. This may be achieved by using mixed finite element, discontinuous Galerkin finite element, and finite volume methods. The finite volume method (box schemes) has the simplicity of the finite difference method [33], and at the same time enjoys the flexibility of the finite element method. For this reason this method is referred to as finite volume element method [29, 48].

The preceeding discussion gives motivation for the objectives of the dissertation. The study will concentrate on solving the following class of partial differential equation:

$$-\nabla \cdot \left(a_\epsilon \left(x, u_\epsilon, \nabla u_\epsilon\right)\right) + b_\epsilon \left(x, u_\epsilon, \nabla u_\epsilon\right) = f \quad \text{in } \Omega, \tag{1.1}$$

with some boundary conditions. Here ϵ represents the small scale in the domain $\Omega \subset \mathbb{R}^2$. The objectives of the study are threefold:

(1) Develop a multiscale finite volume element method (MsFVEM) for solving (1.1)

(2) Conduct an analysis to investigate the convergence of the multiscale methods

(3) Implement MsFVEM in various applications of porous media flow problems.

The method proposed in this dissertation will be a combination of two ingredients, one is related to quantifying the multiscale effects and the other is related to producing a conservative feature on the solution. Traditional approaches for scale up of linear elliptic boundary value problems generally involve the calculation of effective media properties. In these approaches the fine scale information is built into the

effective media parameters, and then the problem on the coarse scale is solved. We refer to [20, 31, 32, 10] for more discussions on upscaled modeling. Recently, a number of approaches have been introduced where the coupling of small scale information is performed through a numerical formulation of the global problem by incorporating the fine features of the problem into coarse elements. In this work we follow a similar approach using finite volume framework. The methodology is similar to multiscale finite element methods proposed in [42] for linear problems.

In Chapter II we will introduce a multiscale map that will devise the quantification of the multiscale effect on the numerical solution. Generally, this multiscale map represents the fluctuation of the solution which is obtained by solving a leading order homogeneous elliptic equation in each element. Obviously, one needs to impose certain boundary condition on this local problem. A piecewise linear function is used for this purpose. Having constructed this multiscale map, we may readily formulate the global problem by setting up the conservation expression on each of the control volume.

As mentioned before, we have imposed piecewise linear Dirichlet boundary conditions on each coarse element when constructing the multiscale map. Previous analysis of multiscale finite element for linear elliptic problem [28] suggested that this kind of treatment produces a resonance error which is due to the mismatch between the physical scale against the grid size. The authors of [28] proposed an oversampling technique that overcomes this drawback. Using this technique, the local problem associated with the multiscale map is solved on a domain substantially larger than the coarse element and in turn use only the information pertaining to it. We will apply similar ideas to solve (1.1).

Next we briefly describe the approach used on the convergence analysis for the proposed multiscale method. We will theoretically study a Dirichlet boundary value

problem associated with (1.1) with the lower order term neglected, and the elliptic coefficient is assumed to be periodic.

Chapter III gives convergence analysis of the linear MsFVEM. The procedures used in the analysis will be similar to the ones that have been employed in the standard finite volume element method [7, 8, 12]. The key issue is to view the finite volume element method as a perturbation of finite element method using a certain interpolation operator. This way, analysis of the method uses substantially the existing finite element results and techniques. Using this procedure, we will rely on the existing analysis of the linear multiscale finite element method. The linear MsFVEM will be written as a Petrov-Galerkin formulation and will be compared against the Petrov-Galerkin finite element formulation [62]. The Petrov-Galerkin setting is applied because of the specific construction of the multiscale map on which the numerical solution is sought. In addition, several results from theory of homogenization (see [45]) will be used.

Convergence analysis for the case of nonlinear coefficients will be presented in Chapter IV. To the best of our knowledge, there has not been a thorough analysis available on the multiscale method for nonlinear elliptic problem with periodic coefficient. Thus the nonlinear problem will be analyzed in the finite element sense. The elliptic coefficient is assumed to exhibit certain properties, namely polynomial growth, monotonicity with respect to the gradient of the solution, coercivity, and continuity. We will distinguish the analysis by whether the resulting operator is monotone or pseudomonotone. We will construct a class of correctors corresponding to the multiscale method, where in the process, the analysis will rely on several approximation properties of these correctors. We note that in general, we may not be able to produce a rate of convergence, but for monotone operators, a certain convergence rate can be deduced.

The last objective of this dissertation is on the applications of the multiscale methods for various problems of flow in porous media. The results are given in Chapter V. There are three main applications that will be investigated, i.e., multiphase immiscible flow, multicomponent miscible flow, and infiltration in saturated/unsaturated porous media.

For multiphase flow in petroleum reservoir simulation, the fine model is the usual pressure equation (elliptic equation) combined with a first order transport equation. The transport quantity is referred to as saturation. This set of equations models the displacement of non-wetting fluid under given pressure on the wetting fluid. An implicit pressure and explixit saturation (IMPES) is employed to solve this set of equations.

The MsFVEM is used to solve the pressure equation from which the velocity field can be recovered to be used in the transport equation. Moreover, two different coarse models are implemented for the saturation equation. One of them is a simple/primitive model where we use only the coarse scale velocity to update the saturation field on the coarse grid. In this case no upscaling of the saturation equation is performed. This kind of technique in conjunction with the upscaling of absolute permeability is commonly used in applications (e.g., [22, 21, 20]). The difference of our approach is that the coupling of the small scales is performed through using the MsFVEM for the global problem and the small scale information of the velocity field can be easily recovered.

In addition to the coarse model described above, we will also revisit a coarse model for the saturation proposed in [27], which was derived using a perturbation argument for the saturation equation. This will result in a diffusion term in addition to the coarse saturation equation that represents the effects of the small scales on the large ones. Note that the diffusion coefficient yields a correlation between the velocity

perturbation and the particle's displacement. Using the MsFVEM for the pressure equation, we are able recover the small scale features of the velocity field that allows us to compute the fine scale displacement. A similar procedure may be performed for the nonlinear flux in the saturation equation. All these macro-diffusion models will be presented in Section 5.1 of Chapter V.

The governing equations for the the multicomponent flow are similar to the ones in multiphase flow. Consequently, we may apply similar upscaling procedures. Again, the MsFVEM is used to solve the pressure equation which is then used to obtain the velocity field. This velocity field is used as an input to the transport equation to obtain the concentration dynamics. As in the multiphase flow, we may perform a macro-diffusion model in the transport equation to get its upscaled version. The only difference is that in the multicomponent flow, the velocity now depends on time (through its dependence on the concentration). Thus, we need to formulate a different approach to get the macro-diffusion coefficient. This is done in Section 5.2 of Chapter V.

Another important class of flow in porous media problems is the unsaturated and/or saturated water flow governed by Richards' equation [54, 6]. This application will be given in Section 5.3 of Chapter 5.2. We note that this equation comes up from the simplification of the two-phase water-air flow problem, where it is assumed that the temporal variation of the water saturation/water content is significantly larger than the temporal variation of the water pressure, and that the air phase is infinitely mobile so that the air pressure remains constant in the atmospheric level. The non-linearity of the equation comes from the dependence of the hydraulic conductivity (the elliptic coefficient) on the pressure.

Finally, Chapter VI is reserved for summary and conclusions and possible future research.

CHAPTER II

GENERAL FORMULATION

The goal of this chapter is to introduce general notations and terminology that will be used throughout the dissertation. The boundary value problem that will be the base model for the proposed numerical model is briefly described in Section 2.2. A brief summary on the solution existence and uniqueness and its aysmptotic behaviors will also be given. Finally in Section 2.3, we will develop the multiscale finite volume element method which solves the model problem.

2.1. Notations

Let K be a domain in $\mathbb{R}^2$. We denote by $L_p(K)$, the space of p integrable real functions over K, for $p = 2$, $(\cdot,\cdot)_K$ is the inner product in $L_2(K)$, $\|\cdot\|_{H^m(K)}$ and $|\cdot|_{H^m(K)}$, the norm and seminorm of the Sobolev space $H^m(K)$ for $m \in \mathbb{N}$. We also introduce the "broken" norm $\|\cdot\|_{m,h}$ such that $\|v\|_{m,h} = \{\sum_{K\in T_h} \|v\|^2_{H^m(K)}\}^{1/2}$, and its corresponding seminorm $|\cdot|_{m,h}$ such that $|v|_{m,h} = \{\sum_{K\in T_h} |v|^2_{H^m(K)}\}^{1/2}$. Also we denote by $\|\cdot\|_{W^{m,p}(K)}$ and $|\cdot|_{W^{m,p}(K)}$, respectively the norm and seminorm of the Sobolev space $W^{m,p}(K)$, $m \in \mathbb{N}$, $p \geq 1$. We note that we suppress the K in the notations whenever $K = \Omega$, and suppress the index m whenever $m = 0$, i.e., $H^0(K) = L_2(K)$. Throughout the paper, C and c (sometimes with indices) will denote generic constants independent of h and ϵ.

2.2. Model Problem

We consider the following elliptic boundary value problem:

$$\begin{aligned} -\nabla \cdot (a_\epsilon(x, u_\epsilon, \nabla u_\epsilon)) + b_\epsilon(x, u_\epsilon, \nabla u_\epsilon) &= f \quad \text{in } \Omega, \\ u_\epsilon &= g_D \quad \text{on } \Gamma_D, \\ a_\epsilon(x, u_\epsilon, \nabla u_\epsilon) \cdot n + b_\epsilon(x, u_\epsilon, \nabla u_\epsilon) &= g_N \quad \text{on } \Gamma_N, \end{aligned} \tag{2.1}$$

where ϵ represents the small scale in the domain $\Omega \subset \mathbb{R}^2$, a bounded polygonal domain, Γ_D and Γ_N are the Dirichlet and Neumann boundary, respectively, $\Gamma_D \cup \Gamma_N = \partial\Omega$, with the measure Γ_D always positive. This boundary value problem is a typical conservation law of the quantity represented by u_ϵ. There are many applications of (2.1), among which are the heat variation, diffusion/dispersion of certain material concentration, and pressure distribution, radiation transport, biological dynamics, and phase transition in biochemistry. The function $a_\epsilon\left(x, u_\epsilon, \nabla u_\epsilon\right)$ represents a vector of flux, while the lower order term $b_\epsilon\left(x, u_\epsilon, \nabla u_\epsilon\right)$ determines the amount of convection. In the case of $a_\epsilon(x, u_\epsilon, \nabla u_\epsilon) \equiv A_\epsilon(x, u_\epsilon)\nabla u_\epsilon$, and $b_\epsilon\left(x, u_\epsilon, \nabla u_\epsilon\right) \equiv b(x)u_\epsilon$, for some tensor A_ϵ and vector b_ϵ, then (2.1) is a typical combination of conservation law and the well known generalized Darcy's Law (in some applications it is referred to as Fick's Law), $v = -A_\epsilon(x, u_\epsilon)\,\nabla u_\epsilon + b(x)u_\epsilon$. In the next two subsections we briefly discuss the existence and uniqueness of the solution, along with its asymptotic behavior.

2.2.1. Existence and Uniqueness of the Solution

There have been a great number of efforts devoted on the existence and uniqueness of the solution to (2.1) (see for example [37]). Using monotone operator theories, the existence and uniqueness of the solution can be established with the price of imposing several restrictions on the nonlinear coefficients. Often, the nonlinear coefficients are

assumed to satisfy the Caratheodory condition, the growth condition, monotonicity condition and coercivity condition [17]. In turn, a nonlinear operator associated with the original boundary value problem may be constructed. Having the assumptions abovementioned, this operator is well defined, continuous, and monotone in Sobolev spaces, and hence its solvability is readily established. Problems may arise in the analysis of this PDEs when the coefficients are singular and/or degenerate. To tackle these difficulties, the authors of [17] have used the weighted Sobolev space method. The usual Sobolev spaces are devised with certain weight functions that are used in the definition of the space's norm; they help to develop certain procedures to show existence of the solution of the differential problem.

2.2.2. Asymptotic Behavior of the Solution

Next we briefly summarize the existing asymtotic analyses of (2.1), i.e., the behavior of the solution as the parameter ϵ vanishes. The homogenization theory for (2.1) relies on one basic assumption, that the nonlinear coefficients are periodic. Furthermore, it is also assumed that the coefficients exhibit certain Holder's continuity, in addition to the assumptions mentioned in the previous subsection. Under these assumptions, an existence of the solutions can be established. This is done by first showing an *a-priori* estimate of the solution which is independent of ϵ [35, 34]. Furthermore, using this *a-priori* estimate one can deduce that the nonlinear coefficients are uniformly bounded in an appropriate dual space. Consequently, a weak convergence of the solution and the nonlinear coefficients are established in the corresponding spaces.

As mentioned above, our model governing equations (2.1) are derived from the conservation law. Hence it is only natural that the numerical models aimed to approximate the solution enjoy certain local numerical conservation properties. Furthermore, we would like to be able to include the multiscale effects associated with ϵ

in the solution, a subject discussed in the next section.

2.3. The Multiscale Finite Volume Element Method (MsFVEM)

The finite volume method has certain local conservative properties (see [33] for an extensive survey of the method). Unlike the finite element method that relies on a global variational formulation, the finite volume method is derived from a local relation, namely the balance equation/conservation expression on a number of sub-domains which are called control volumes. In what follows, we describe the finite volume discretization of (2.1) that leads to its numerical solution.

Let T_h be the collection of quasiuniform triangulations of $\Omega \subset \mathbb{R}^2$, and X^h be the piecewise linear finite element space that lives in T_h, i.e.,

$$X^h = \left\{ \chi \in H_0^1(\Omega) : \chi \mid_K \text{ is linear, } \chi \mid_{\partial\Omega} = 0 \right\}. \tag{2.2}$$

Given the triangulation T_h, we describe the construction of the control volumes as follows. Consider a triangle $K \in T_h$, and let z_K be its barycenter. The triangle K is divided into three quadrilaterals of equal area by connecting z_K to the midpoints of its three edges. We denote these quadrilaterals by K_z, where $z \in Z_h(K)$, are the vertices of K. Also we denote $Z_h = \bigcup_K Z_h(K)$, and Z_h^0 are all vertices that do not lie in Γ_D. The control volume V_z is defined as the union of the quadrilaterals K_z sharing the vertex z (see Figure 2.1).

Next consider an element v^h that belongs to X^h. We denote by v_ϵ a function that satisfies the boundary value problem:

$$\begin{aligned} -\nabla \cdot \left(a_\epsilon \left(x, \eta^h, \nabla v_\epsilon \right) \right) = 0 \quad &\text{in } K \in T_h, \\ v_\epsilon = v^h \quad &\text{on } \partial K, \end{aligned} \tag{2.3}$$

with $\eta^h(x) = \sum_{K \in T_h} \Psi_K(x) \frac{1}{|K|} \int_K v^h \, dx$, Ψ_K being the characteristic function of the

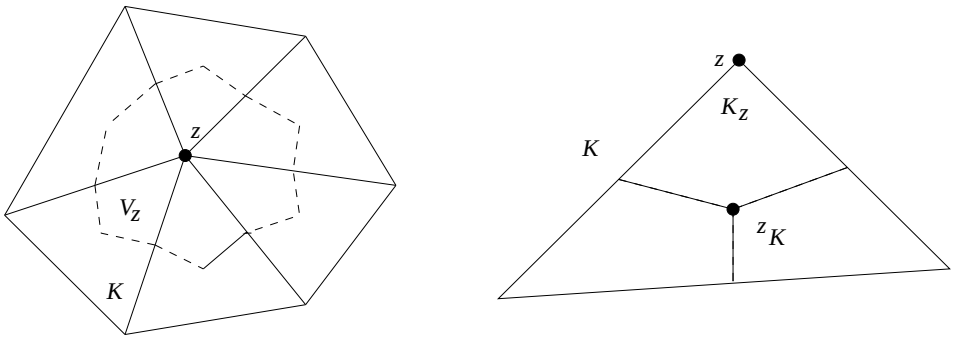

Fig. 2.1. *Left*: Portion of triangulation sharing a common vertex z and its control volume. *Right*: Partition of a triangle K into three quadrilaterals

element K. Then we denote by V_ϵ^h the space of all functions satisfying (2.3). Also we denote by $E : X^h \to V_\epsilon^h$ the corresponding multiscale map associated with (2.3).

The multiscale finite volume element method (MsFVEM) for (2.1) is to find $u^h \in X^h$ such that

$$-\int_{\partial V_z} a_\epsilon\left(x, \eta^h, \nabla u_\epsilon^h\right) \cdot n \, dS + \int_{V_z} b_\epsilon\left(x, \eta^h, \nabla u_\epsilon^h\right) \, dx = \int_{V_z} f \, dx \quad \forall z \in Z_h^0, \tag{2.4}$$

where $u_\epsilon^h = E(u^h)$. It is obvious that the number of control volumes that satisfies (2.4) is the same as the dimension of X^h. In the case of linear coefficients, namely, $a_\epsilon\left(x, u_\epsilon, \nabla u_\epsilon\right) \equiv A_\epsilon(x)\nabla u_\epsilon$, and $b_\epsilon\left(x, u_\epsilon, \nabla u_\epsilon\right) \equiv b_\epsilon(x) \cdot \nabla u_\epsilon$, the multiscale map E is a linear operator, and thus V_ϵ^h is a linear space. Then given a set of basis functions $\{\phi^i\}$ of X^h we may construct a set of multiscale basis functions $\{\phi_\epsilon^i\}$ of V_ϵ^h that satisfy

$$\begin{aligned} -\nabla \cdot \left(A_\epsilon(x) \, \nabla \phi_\epsilon^i\right) = 0 \quad &\text{in } K \in T_h, \\ \phi_\epsilon^i = \phi^i \quad &\text{on } \partial K. \end{aligned} \tag{2.5}$$

One drawback inherent in the proposed method is the error resulting from elements' boundary layers. This discrepancy is quantified by the ratio of the physical scale ϵ to the mesh size h. An analysis of the linear MsFEM has been done (see[43]), which shows that the convergence depends on this ratio.

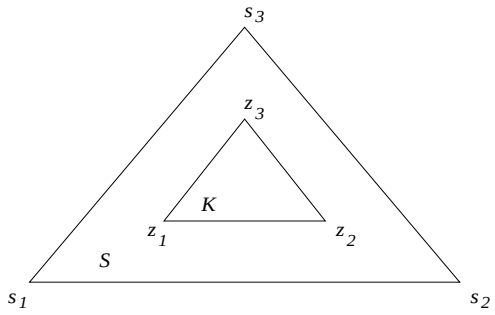

Fig. 2.2. Oversampling of (2.3) on a substantially larger domain than triangle K.

To overcome this drawback, an oversampling strategy is employed, in that the local problem (2.3) (correspondingly (2.5) for linear problem) is solved in domain of size larger than $h+\epsilon$ (see Figure 2.2). This procedure has been proposed and analyzed in [28] for linear MsFEM.

In the next chapter, we will explore in detail the convergence analysis of the MsFVEM for boundary value problems with linear coefficients.

CHAPTER III

CONVERGENCE ANALYSIS OF MSFVEM FOR A LINEAR ELLIPTIC PROBLEM

3.1. Preliminaries

In this chapter we present a convergence analysis of the MsFVEM for a special case of (2.1), namely a linear elliptic boundary value problem with $a_\epsilon(x, u_\epsilon, \nabla u_\epsilon) \equiv A_\epsilon(x)\nabla u_\epsilon$ and $b_\epsilon(x, u_\epsilon, \nabla u_\epsilon) \equiv b_\epsilon(x)\cdot\nabla u_\epsilon$. Using the notations and settings described in Chapter II, the MsFVEM formulation is defined as to find $u_\epsilon^h \in V_\epsilon^h$ such that

$$-\int_{\partial V_z} A_\epsilon(x)\,\nabla u_\epsilon^h \cdot n\,ds + \int_{\partial V_z} u_\epsilon^h\, b_\epsilon(x)\cdot n\,dx = \int_{V_z} f\,dx \quad V_z \subset \Omega. \tag{3.1}$$

It is obvious that for linear problems, the multiscale map E defined in Chapter 3.4 is a linear operator, and consequently V_ϵ^h is a linear space. Hence we may construct a set of basis functions belonging to V_ϵ^h satisfying the local problem (2.5). This description will be explored in detail in the next section.

For the analysis that follows we will impose several assumptions. First we denote by Y a unit square $(0,1)\times(0,1)$. Let $A(x) = \{A_{ij}(x)\}$ be a 2×2 matrix for $x \in \mathbb{R}^2$, satisfying the following properties:

L1 A_{ij} is Y-periodic in $\mathbb{R}^2$ for every $i, j = 1, 2$;

L2 there exist constants $\beta > \alpha > 0$ such that $\alpha|\xi|^2 \le \xi^t A(x)\xi \le \beta|\xi|^2$ a.e. $x \in \mathbb{R}^2$ and for any vector $\xi \in \mathbb{R}^2$;

L3 $A_{ij} \in W^{1,p}(\mathbb{R}^2)$ for some $p > 2$.

We then define that the elliptic coefficient in (3.1) by

$$A_\epsilon(x) = A(x/\epsilon), \tag{3.2}$$

where $\epsilon > 0$ is the positive number representing the small scale feature. By definition (3.2) we know that the functions A_{ij} are $\epsilon\, Y$-periodic in $\mathbb{R}^2$. In the analysis we will neglect the lower order term b_ϵ, and concentrate on solving the Dirichlet problem, namely to seek $u_\epsilon = u(x, x/\epsilon) \in H_0^1(\Omega)$ satisfying

$$\begin{aligned} -\nabla \cdot (A(x/\epsilon)\nabla u_\epsilon) &= f(x) \quad \text{in } \Omega \subset \mathbb{R}^2, \\ u_\epsilon &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{3.3}$$

for some $f \in L_2(\Omega)$. In this analysis we assume that $\Omega = (0,1) \times (0,1) \subset \mathbb{R}^2$. We note that by the Lax-Milgram lemma, assumption **L2** imply that there exists a unique weak solution of (3.3).

Next, we denote the following finite-dimensional space:

$$Y^h = \left\{ \xi \in L_2(\Omega) : \xi \mid_{V_z} \text{ is constant, } z \in Z_h^0, \, \xi \mid_{V_z} = 0 \text{ if } z \in \partial\Omega \right\}. \tag{3.4}$$

In the sections that follow we use the following interpolation operator $I_h : X^h \to Y^h$ such that $\forall\, \chi \in X^h$

$$I_h \chi = \sum_{z \in Z_0^h} \chi(z) \Psi_z, \tag{3.5}$$

where Ψ_z is the characteristic function of the control volume V_z. Below we list several properties of I_h (see [7, 12] for details):

$$(\chi, I_h \phi) = (\phi, I_h \chi), \;\; \forall\, \chi, \phi \in X^h, \tag{3.6}$$

$$c_1 \|\chi\|^2 \le |||\chi|||^2 \le c_2 \|\chi\|^2, \;\; \forall\, \chi \in X^h, \;\; c_2 > c_1 > 0, \;\; |||\chi|||^2 = (\chi, I_h \chi), \tag{3.7}$$

$$\int_K I_h \chi \, dx = \int_K \chi \, dx, \;\; \forall\, \chi \in X^h, \;\; \text{for any } K \in T_h, \tag{3.8}$$

$$\int_e I_h \chi \, ds = \int_e \chi \, ds, \;\; \forall\, \chi \in X^h, \;\; \text{for any side } e \text{ of } K \in T_h, \tag{3.9}$$

$$\|I_h \chi\|_{L_\infty(e)} \le \|\chi\|_{L_\infty(e)}, \;\; \forall\, \chi \in X^h, \;\; \text{for any side } e \text{ of } K \in T_h, \tag{3.10}$$

$$\|\chi - I_h\chi\|_{L_p(K)} \leq Ch_K|\chi|_{W^{1,p}(K)}, \ \ \forall\, \chi \in X^h, \ \ 1 \leq p < \infty. \tag{3.11}$$

3.2. An Overview of Homogenization Theory

Here we review some results from homogenization theory [45]. We will use the Einstein's summation wherever it applies, namely, summation is taken over repeated indices. First we define $N_k(y)$, $k = 1, 2$ to be the periodic solution in the unit square Y with $\langle N_k \rangle_Y = 0$ that satisfies the equation

$$\nabla_y \cdot (A(y)\nabla_y N_k(y)) = -\nabla_y^i A_{ik}(y). \tag{3.12}$$

Here ∇_y is the gradient with respect to the variable y and ∇_y^i is the i-th component of the ∇_y. By homogenization theory (cf. [45]), the solution of (3.3) can be expanded as

$$u_\epsilon(x, x/\epsilon) = u_0(x) + \epsilon\, N_k(x/\epsilon)\, \nabla_k u_0(x) + \epsilon\, \theta_\epsilon^u(x, x/\epsilon), \tag{3.13}$$

where ∇_k is the k-th component of ∇. The function u_0 is the solution of the following homogenized boundary value problem [45]:

$$\begin{aligned} -\nabla \cdot (A^*\nabla u_0) &= f \quad \text{in } \Omega, \\ u_0 &= 0 \quad \text{on } \partial\Omega, \end{aligned} \tag{3.14}$$

where the entries of A^*, denoted by A_{ij}^*, is expressed as

$$A_{ij}^* = \int_Y A_{ik} \left(\delta_{kj} + \nabla_y^k N_j\right) dy. \tag{3.15}$$

Regarding θ_ϵ^u, we have the following estimate [45]:

Lemma 3.1. *Let θ_ϵ^u be the corrector in* (3.13)*. Assume that the solution of* (3.14) *$u_0 \in C^2(\bar{\Omega})$ and property* **L3** *holds. Then there exists a constant $C > 0$ independent*

of ϵ such that

$$\epsilon\,|\theta^u|_1 \leq C\,\sqrt{\epsilon}. \tag{3.16}$$

3.3. Oversampling and Construction of the Solution Space V_ϵ^h

In this section we present the oversampling strategy that will be combined with the finite volume element method. As mentioned earlier, the space V_ϵ^h for linear problems is a linear space, and thus a set of basis functions satisfying (2.3) may be constructed. A particular construction of such a basis is explained in detail as follows.

We construct an intermediate set of functions $\{\psi_\epsilon^i, i = 1, 2, 3\}$ in an oversampled triangle domain $S \supset K$, $\mathrm{diam}(S) > 2h_K$ by solving

$$L_\epsilon\,\psi_\epsilon^i = -\nabla\cdot\left(A(x/\epsilon)\nabla\psi_\epsilon^i\right) = 0 \quad \text{in } S, \tag{3.17}$$

where ψ_ϵ^i is piecewise linear along ∂S, and $\psi_\epsilon^i(s_j) = \delta_{ij}$, with s_j, $j = 1, 2, 3$ being the vertices of S (see Figure 2.2). It follows from this construction that ψ_ϵ^i exhibit similar structure to u_ϵ. Furthermore, with respect to ϵ, ψ_ϵ^i has the following asymptotic expansion:

$$\psi_\epsilon^i(x, x/\epsilon) = \psi_0^i(x) + \epsilon\,N_k(x/\epsilon)\,\nabla_k\psi_0^i(x) + \epsilon\,\theta^i(x, x/\epsilon), \tag{3.18}$$

where ψ_0^i is the linear homogenized part of ψ_ϵ^i, $\theta^i = \eta_k\nabla_k\psi_0^i$, where η_k, $k = 1, 2$, satisfy the following problem:

$$\begin{aligned} \nabla\cdot(A(x/\epsilon)\nabla\eta_k) &= 0 \quad \text{in } S \\ \eta_k &= -N_k \quad \text{on } \partial S. \end{aligned} \tag{3.19}$$

The function η_k has the following property [28]:

Lemma 3.2. *Let $K \in T_h$ such that $K \subset S$ is away at least at a distance h_K. Then*

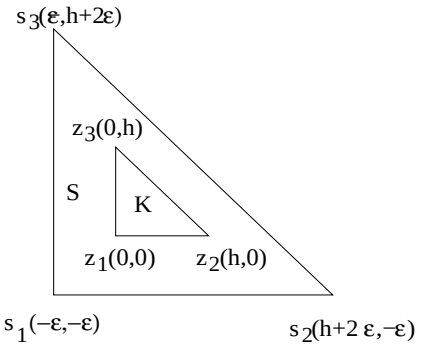

Fig. 3.3. Right triangle of size h and its oversampled counterpart.

there exists a constant $C > 0$ *independent of* ϵ *and* h_K *such that*

$$\|\nabla \eta_k\|_{L_\infty(K)} \leq \frac{C}{h_K}. \tag{3.20}$$

Next, the restrictions of the basis functions $\phi_\epsilon^i, i = 1, 2, 3$ on K are taken as linear combinations of $\psi_\epsilon^i, i = 1, 2, 3$, i.e.,

$$\phi_\epsilon^i = \sum_{j=1}^{3} c_{ij} \psi_\epsilon^j. \tag{3.21}$$

Substituting (3.18) to (3.21), we see that ϕ_ϵ^i can be expanded as follows:

$$\phi_\epsilon^i(x, x/\epsilon) = \phi_0^i(x) + \epsilon\, N_k(x/\epsilon)\, \nabla_k \phi_0^i(x) + \epsilon\, c_{ij}\, \theta^j(x, x/\epsilon), \tag{3.22}$$

where

$$\phi_0^i = \sum_{j=1}^{3} c_{ij}\, \psi_0^j. \tag{3.23}$$

The constants c_{ij} are obtained by setting $\phi_0^i(z_j) = \delta_{ij}$ which gives a system of linear equations. To see that the constants c_{ij} exist, without loss of generality, we consider a right triangle along with its oversampled counterpart shown in Figure 3.3. It is obvious that for this setting we have $\psi_0^2 = (x_1 + \epsilon)/(h + 3\epsilon)$, $\psi_0^3 = (x_2 + \epsilon)/(h + 3\epsilon)$, and $\psi_0^3 = 1 - \psi_0^1 - \psi_0^2$. Setting $\phi_0^i(z_j) = \delta_{ij}$ in (3.23) using these equations we obtain

the following linear system:

$$\frac{1}{h+3\epsilon}\begin{bmatrix} h+\epsilon & \epsilon & \epsilon \\ \epsilon & h+\epsilon & \epsilon \\ \epsilon & \epsilon & h+\epsilon \end{bmatrix}\begin{bmatrix} c_{i1} \\ c_{i2} \\ c_{i3} \end{bmatrix} = \begin{bmatrix} \delta_{i1} \\ \delta_{i2} \\ \delta_{i3} \end{bmatrix}, \quad i = 1, 2, 3. \tag{3.24}$$

It is straightforward to see that the 3×3 matrix in this linear system is invertible. Furthermore, since the oversampled domain S is not much larger than the triangle K, the matrix is well conditioned. We note that by this construction, the basis functions are continuous at the vertex points $z \in Z_h$, but in general they are not continuous across ∂K. We also set the basis functions to be zero on $\partial\Omega$. Consequently V_ϵ^h is no longer a subset of $H^1(\Omega)$. Now we have the tool to expand the functions that belong to the space of our approximate solution V_ϵ^h. So consider $v_\epsilon^h \in V_\epsilon^h$. Since the expansion of the basis functions was conducted on a triangle K, we will also have the asymptotic expansion for v_h^ϵ on K. First we write $v_0^h = v_0^h(z_i)\,\phi_0^i$, $z_i \in Z_h(K)$. Moreover, since $\theta^j = \eta_k \nabla_k \psi_0^j$ we may define θ^h using the following equivalent representations:

$$\theta^h = v_0^h(z_i)\, c_{ij}\, \theta^j = v_0^h(z_i)\, c_{ij}\, \eta_k\, \nabla_k \psi_0^j = v_0^h(z_i)\, \eta_k\, \nabla_k \phi_0^i = \eta_k\, \nabla_k v_0^h. \tag{3.25}$$

Then by setting $v_\epsilon^h = v_0^h(z_i)\,\phi_\epsilon^i$ for $z_i \in Z(K)$, and using (3.22) and (3.25), in each triangle $K \in T_h$, we have the following asymptotic expansion for $v_\epsilon^h \in V_\epsilon^h$:

$$v_\epsilon^h(x, x/\epsilon) = v_0^h(x) + \epsilon N_k(x/\epsilon)\nabla_k v_0^h(x) + \epsilon\theta^h(x, x/\epsilon). \tag{3.26}$$

We note that the function v_0^h in (3.26) is piecewise linear, since it is defined as a linear combination of the homogenized basis functions ϕ_0^i that are linear.

3.4. Reformulation of the Method

Now, we are in a position to formulate the two-scale finite volume element method for (3.3) that incorporates the small scale features: find $u_\epsilon^h \in V_\epsilon^h$ that satisfies the following equation expressing local conservation:

$$-\int_{\partial V_z} (A(x/\epsilon)\nabla u_\epsilon^h) \cdot n \, ds = \int_{V_z} f \, dx, \quad \forall \, z \in Z_h^0, \tag{3.27}$$

Obviously, this construction requires that the number of control volumes V_z to be equal to the dimension of V_ϵ^h. We note that this formulation may be equivalently written as the following variational problem: Find $u_\epsilon^h \in V_\epsilon^h$ such that

$$a_{FV}(u_\epsilon^h, \chi) = \sum_{z \in Z_h^0} \chi(z) \int_{V_z} f \, dx \quad \forall \, \chi \in X^h, \tag{3.28}$$

where the form $a_{FV}(\cdot,\cdot) : \tilde{H}_h^2 \times H_h^2 \to \mathbb{R}$ is defined by

$$a_{FV}\,(v, \chi) = -\sum_{z \in Z_h^0} \chi(z) \int_{\partial V_z} (A(x/\epsilon)\nabla v) \cdot \vec{n} \, ds, \tag{3.29}$$

with $\tilde{H}_h^2 = H^2(\Omega) + V_\epsilon^h$ and $H_h^2 = H^2(\Omega) + X^h$. In [7] it has been shown that using the interpolation operator I_h in (3.5), we have

$$\sum_{z \in Z_h^0} \chi(z) \int_{V_z} f \, dx = (f, I_h\chi). \tag{3.30}$$

Now we can give another equivalent representation of $a_{FV}(v, \chi)$. Consider a triangle K and a control volume V_z such that $K \cap V_z \neq \emptyset$. Then using Green's formula we get

$$\int_{K \cap V_z} L_\epsilon v \, dx = -\int_{\partial K \cap V_z} (A(x/\epsilon)\nabla v) \cdot \vec{n} \, ds - \int_{\partial V_z \cap K} (A(x/\epsilon)\nabla v) \cdot \vec{n} \, ds. \tag{3.31}$$

This equality and the interpolation operator I_h allows us to get

$$\begin{aligned} a_{FV}(v,\chi) &= -\sum_{K\in T_h}\sum_{z\in Z_h(K)}\int_{\partial V_z\cap K}(A(x/\epsilon)\nabla v)\cdot\vec{n}\, I_h\chi\, ds \\ &= \sum_{K\in T_h}\Big\{\big(L_\epsilon v, I_h\chi\big)_K + \big((A(x/\epsilon)\nabla v)\cdot\vec{n}, I_h\chi\big)_{\partial K}\Big\}. \end{aligned} \tag{3.32}$$

By combining all these identities we may write the following equivalent Petrov-Galerkin formulation of the *two-scale finite volume element* problem: Find $u_\epsilon^h \in V_\epsilon^h$ such that

$$a_{FV}(u_\epsilon^h,\chi) = (f, I_h\chi) \quad \forall\,\chi\in X^h, \tag{3.33}$$

with $a_{FV}(\cdot,\cdot)$ as in (3.32).

In the finite volume element method, there is well developed technique for the error analysis based on the existing results from its standard finite element counterpart (see [7, 12] for detail investigation). The main idea is to view the finite volume element as a perturbation of the finite element method with the help of the interpolation operator I_h. This way, one can tap into existing analysis in the Galerkin finite element method to derive the error estimates for the finite volume element method.

We will follow a similar procedure. However, due to the specific construction of the basis functions and the corresponding finite-dimensional space of the approximate solution V_ϵ^h, that accounts for the scale features, we will emphasize the Petrov-Galerkin formulation. First, we introduce the Petrov-Galerkin formulation of the *two-scale finite element* problem associated with (3.3) (cf. [62]): Find $\tilde{u}_\epsilon^h \in V_\epsilon^h$ such that

$$a_{FE}(\tilde{u}_\epsilon^h,\chi) = (f,\chi) \quad \forall\,\chi\in X^h, \tag{3.34}$$

where

$$a_{FE}(v_\epsilon^h,\chi) = \sum_{K\in T_h}\big(A(x/\epsilon)\nabla v_\epsilon^h, \nabla\chi\big)_K, \quad \forall\, v_\epsilon^h\in V_\epsilon^h, \chi\in X^h. \tag{3.35}$$

By Green's formula we may write $a_{FE}(\cdot,\cdot)$ as

$$\begin{aligned} a_{FE}(v_\epsilon^h,\chi) &= \sum_{K\in T_h}\Big\{\big(L_\epsilon v_\epsilon^h,\chi\big)_K + \big(A(x/\epsilon)\nabla v_\epsilon^h\cdot\vec{n},\chi\big)_{\partial K}\Big\} \\ &= \sum_{K\in T_h}\big(A(x/\epsilon)\nabla v_\epsilon^h\cdot\vec{n},\chi\big)_{\partial K} \quad \forall\, v_\epsilon^h\in V_\epsilon^h,\chi\in X^h. \end{aligned} \tag{3.36}$$

The first term in (3.36) vanishes since by the construction of the basis functions of V_ϵ^h, we have

$$L_\epsilon v_\epsilon^h = L_\epsilon\left(\sum_{i=1}^{3} v_0^h(z_i)\,\phi_\epsilon^i\right) = \sum_{i=1}^{3} v_0^h(z_i)\,L_\epsilon\,\phi_\epsilon^i = 0, \tag{3.37}$$

where as before $L_\epsilon\phi = -\nabla\cdot(A(x/\epsilon)\nabla\phi)$. Using (3.36) and (3.32), we may define the following bilinear form $D: V_\epsilon^h\times X^h\to\mathbb{R}$:

$$\begin{aligned} D(v_\epsilon^h,\chi) &= a_{FE}(v_\epsilon^h,\chi) - a_{FV}(v_\epsilon^h, I_h\chi) \\ &= \sum_{K\in T_h}\big((A(x/\epsilon)\nabla v_\epsilon^h)\cdot\vec{n},\chi - I_h\chi\big)_{\partial K}. \end{aligned} \tag{3.38}$$

This bilinear form characterizes the two-scale finite volume element method as a perturbation of the two-scale finite element method. Our aim now is to estimate (3.38), by using the existing results of the two-scale finite element method and then to obtain the convergence of the two-scale finite volume element method.

3.5. Convergence Analysis of the Method for Case $\epsilon \ll h$

As mentioned earlier, the analysis proceeds with quantification of the perturbation between the two-scale finite volume element method and its finite element counterpart. In this section we estimate (3.38), show the inf-sup condition of the bilinear form guaranteeing the existence and uniqueness of the solution, and prove an error estimate in the broken norm $\|\cdot\|_{1,h}$. First we establish the following lemma that will be used in the subsequent proof.

We define a 2×2 matrix $B(x/\epsilon) = \{B_{ij}(x/\epsilon)\}$ such that

$$B_{ij}(x/\epsilon) = A_{ij}(x/\epsilon) + \epsilon\, A_{ik}(x/\epsilon)\, \nabla_k N_j(x/\epsilon), \tag{3.39}$$

where N_j is as in Section 3.2.

Lemma 3.3. *Assume that there exist constants $c_2 > c_1 > 0$, such that*

$$c_1|\xi|^2 \leq \xi_i B_{ij} \xi_j \leq c_2|\xi|^2 \quad \forall\, \xi \in \mathbb{R}^2. \tag{3.40}$$

Then there exist constants $C_2 > C_1 > 0$ such that

$$C_1|\nabla v_\epsilon^h|_K \leq |\nabla v_0^h|_K \leq C_2|\nabla v_\epsilon^h|_K \tag{3.41}$$

for every $v_\epsilon^h \in V_\epsilon^h$ and for each $K \in T_h$.

Proof. In what follows all the estimates are taken over the triangle K. Using (3.26) and (3.25) and noting that v_0^h is linear in K, we have the following equality:

$$\begin{aligned} A_{ij}\nabla_j v_\epsilon^h &= A_{ij}\nabla_j v_0^h + A_{ij}\nabla_j \left(\epsilon N_k \nabla_k v_0^h\right) + \epsilon A_{ij}\nabla_j \theta^h \\ &= \left(A_{ij} + \epsilon A_{ik}\nabla_k N_j + \epsilon A_{ik}\nabla_k \eta_j\right) \nabla_j v_0^h \\ &= \left(B_{ij} + \epsilon A_{ik}\nabla_k \eta_j\right) \nabla_j v_0^h. \end{aligned} \tag{3.42}$$

Multiplying (3.42) by $\nabla_i v_0^h$ we have

$$\nabla_i v_0^h A_{ij} \nabla_j v_\epsilon^h = \nabla_i v_0^h \left(B_{ij} + \epsilon A_{ik}\nabla_k \eta_j\right) \nabla_j v_0^h. \tag{3.43}$$

Now by Lemma 3.2 we may apply the assumption (3.40) to the term $B_{ij} + \epsilon A_{ik}\nabla_k \eta_j$, so that

$$\beta\, |\nabla v_0^h|\, |\nabla v_\epsilon^h| \geq \nabla_i v_0^h A_{ij} \nabla_j v_\epsilon^h = \nabla_i v_0^h \left(B_{ij} + \epsilon A_{ik}\nabla_k \eta_j\right) \nabla_j v_0^h \geq c_1\, |\nabla v_0^h|^2, \tag{3.44}$$

from which we obtain the right hand side inequality of (3.41). Similarly, multiplying

(3.42) by $\nabla_i v_\epsilon^h$, and by the positive definiteness of A, we obtain the result for the left hand side of (3.41). □

3.5.1. Estimate on the form $D(v_\epsilon^h, \chi)$

In this subsection we estimate the form $D(v_\epsilon^h, \chi)$ defined in (3.38). First of all, we would like to rewrite the form $D(v_\epsilon^h, \chi)$ in (3.38) such that it will be easier to estimate. To this end, we note that by taking the partial derivative of the v_ϵ^h expansion in (3.26), and using the fact that v_0^h is piecewise linear (and hence its derivative is zero), we have:

$$\nabla_j v_\epsilon^h = \nabla_j v_0^h + \epsilon (\nabla_j N_k) \nabla_k v_0^h + \epsilon \nabla_j \theta^h, \quad j = 1, 2 \tag{3.45}$$

where as before, we have used the Einstein summation for $k = 1, 2$. Multiplying the matrix A to the vector ∇v_ϵ^h and applying (3.45) we obtain the following:

$$\begin{aligned}
(A \nabla v_\epsilon^h)_i &= \sum_{j=1}^{2} A_{ij} \nabla_j v_\epsilon^h \\
&= \sum_{j=1}^{2} A_{ij} \left(\nabla_j v_0^h + \epsilon \left((\nabla_j N_1) \nabla_1 v_0^h + (\nabla_j N_2) \nabla_2 v_0^h \right) + \epsilon \nabla_j \theta^h \right) \\
&= \sum_{j=1}^{2} \left(A_{ij} + \epsilon \sum_{k=1}^{2} (A_{ik} \nabla_k N_j) \nabla_j v_0^h + \epsilon A_{ij} \nabla_j \theta^h \right).
\end{aligned} \tag{3.46}$$

Notice that on the last line of (3.46), the first term is the the entry of the matrix B defined in (3.39). Combining all these derivations, we may substitute v_ϵ^h expansion in (3.26) to the form $D(v_\epsilon^h, \chi)$ in (3.38) to obtain

$$\begin{aligned}
D(v_\epsilon^h, \chi) = &\sum_{K \in T_h} \Big(\big(B(x/\epsilon) \nabla v_0^h \big) \cdot \vec{n}, \chi - I_h \chi \Big)_{\partial K} \\
&+ \sum_{K \in T_h} \Big(\big(A(x/\epsilon) \epsilon \nabla \theta^h \big) \cdot \vec{n}, \chi - I_h \chi \Big)_{\partial K},
\end{aligned} \tag{3.47}$$

for any $\chi \in X^h$. The following two lemmas are devoted to estimate the two terms in (3.47).

Lemma 3.4. *Assume that the entries of the matrix $A(y)$ are 1-periodic functions along each edge e of a triangle $K \in T_h$. Then for every $\chi \in X^h$ there exists a constant $C > 0$ independent of ϵ and h such that*

$$\int_e (B(x/\epsilon)\nabla v_0^h) \cdot \vec{n}\,(\chi - I_h\chi) ds \leq C\,\frac{\epsilon}{h}\,|v_\epsilon^h|_{H^1(K)}\,|\chi|_{H^1(K)} \tag{3.48}$$

for every edge e of the triangle K.

Proof. Since the matrix A is 1-periodic along the edge, so is the matrix B defined by (3.39). Choose a constant matrix $\widetilde{B}$ whose entries will be determined later. Since $\nabla v_0^h \cdot \vec{n}$ is constant on e, by (3.9) we have

$$\begin{aligned}\int_e \big(B(x/\epsilon)\nabla v_0^h\big) \cdot \vec{n}\;(\chi - I_h\chi)\,ds &= \int_e \big((B(x/\epsilon) - \widetilde{B})\nabla v_0^h\big) \cdot \vec{n}\;(\chi - I_h\chi)\,ds \\ &= n_i\,\nabla_j v_0^h \int_e (B_{ij}(x/\epsilon) - \widetilde{B_{ij}})\;(\chi - I_h\chi)\,ds.\end{aligned} \tag{3.49}$$

Note, that we have used Einstein's summation on the last line. Recall that $I_h\chi$ is discontinuous along the edge e. Let z_l and z_r be the two vertices connected by edge e, and z_m be the the midpoint of e. The integration in (3.49) may be broken up into integration along (z_l, z_m) plus integration along (z_m, z_r). Starting from z_m we may break up the segment (z_l, z_m) into a number of sub-segments Y_ϵ each of which has size ϵ and possibly one sub-segment Y_δ of size $\delta < \epsilon$ (see Figure 3.4). A similar partition may be implemented for segment (z_m, z_r). This partition implies that the integration in (3.49) may be broken up into the sum of integral over all the sub-segments. Now it is obvious that the matrix B is periodic with respect to the sub-segment Y_ϵ. In what follows we will estimate the integral (3.49) over the sub-segments Y_ϵ and Y_δ. We

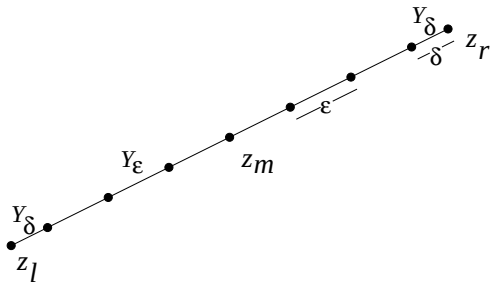

Fig. 3.4. Partition of an edge e into sub-segments Y_ϵ of size ϵ and possibly two segments Y_δ of size less than ϵ.

choose the matrix $\widetilde{B}$ to have the following entries:

$$\widetilde{B_{ij}} = \frac{1}{|Y_\epsilon|}\int_{Y_\epsilon} B_{ij}\, ds. \tag{3.50}$$

Obviously the estimate for integral over Y_δ is straightforward, since we have $|B_{ij}-\widetilde{B_{ij}}|$ bounded, and $|\chi - I_h\chi| \le C\,|\nabla\chi|\,\epsilon$ in Y'_ϵ. Hence,

$$\int_{Y_\delta}(B_{ij}-\widetilde{B_{ij}})\;(\chi - I_h\chi)\;ds \le C\frac{\epsilon^2}{h}\,|\chi|_{H^1(K)}, \tag{3.51}$$

where we have used the inverse inequality for χ. Moreover, by choosing $\widetilde{B}$ as in (3.50), we have the following indentity:

$$\int_{Y_\epsilon}(B_{ij}-\widetilde{B_{ij}})\;(\chi - I_h\chi)\;ds = \int_{Y_\epsilon}(B_{ij}-\widetilde{B_{ij}})\,\chi\,ds = \int_{Y_\epsilon}(B_{ij}-\widetilde{B_{ij}})\;(\chi-\widetilde{\chi})\;ds, \tag{3.52}$$

where

$$\widetilde{\chi} = \frac{1}{|Y_\epsilon|}\int_{Y_\epsilon}\chi\, ds. \tag{3.53}$$

By the Cauchy-Schwarz inequality, from (3.52) we have

$$\int_{Y_\epsilon}(B_{ij}-\widetilde{B_{ij}})\;(\chi - I_h\chi)\;ds \le \|B_{ij}-\widetilde{B_{ij}}\|_{L_2(Y_\epsilon)}\,\|\chi-\widetilde{\chi}\|_{L_2(Y_\epsilon)}, \tag{3.54}$$

so we need to estimate the two norms in this inequality. The Poincaré-Friedrich

inequality and a scaling argument gives us

$$\|B_{ij} - \widetilde{B_{ij}}\|_{L_2(Y_\epsilon)} \leq C\,\sqrt{\epsilon}\,\|\nabla^y B_{ij}\|_{L_2(0,1)} \leq C\,\sqrt{\epsilon}. \tag{3.55}$$

Furthermore, due to the fact that χ is linear on the edge e, we know that $|\chi - \widetilde{\chi}| \leq \epsilon\,|\nabla\chi|$, and thus using the fact that $\nabla\chi$ is constant on the edge e, applying inverse inequality to $|\nabla\chi|$ we have

$$\begin{aligned}
\|\chi - \widetilde{\chi}\|_{L_2(Y_\epsilon)} &\leq \left(\int_{Y_\epsilon} \epsilon^2\,|\nabla\chi|^2\,ds\right)^{1/2} \\
&\leq \epsilon\,|\nabla\chi|\left(\int_{Y_\epsilon} ds\right)^{1/2} \\
&\leq C\,\frac{\epsilon^{3/2}}{h}\,|\nabla\chi|_{H^1(K)}.
\end{aligned} \tag{3.56}$$

Combining (3.55), and (3.56) we have the following estimate:

$$\int_{Y_\epsilon} (B_{ij} - \widetilde{B_{ij}})\;(\chi - I_h\chi)\,ds \leq C\,\frac{\epsilon^2}{h}\,|\chi|_{H^1(K)}. \tag{3.57}$$

Putting our attention back to the integration over (z_l, z_m), now we may sum over all Y_ϵ and Y'_ϵ and note that all terms on the (3.51) and (3.57) are independent of ϵ except the ϵ itself. Thus,

$$\begin{aligned}
\int_{z_l}^{z_m} (B_{ij} - \widetilde{B_{ij}})\;(\chi - I_h\chi)\;ds &= \sum_{Y_\epsilon, Y'_\epsilon} \int_{Y_\epsilon} (B_{ij} - \widetilde{B_{ij}})\;(\chi - I_h\chi)\;ds \\
&\leq C\,\frac{\epsilon}{h}\,|\chi|_{H^1(K)} \sum_{Y_\epsilon, Y'_\epsilon} \epsilon \\
&\leq C\,\epsilon\,|\chi|_{H^1(K)}.
\end{aligned} \tag{3.58}$$

The same procedure described above may be implemented for (z_m, z_r) so that summing up the results from these two segments and an applying inverse inequality to

v_0^h give us

$$n_i \nabla_j v_0^h \int_e (B_{ij}(x/\epsilon) - \widetilde{B_{ij}}) \ (\chi - I_h\chi) \ ds \leq C \frac{\epsilon}{h} |v_0^h|_{H^1(K)} |\chi|_{H^1(K)}. \tag{3.59}$$

Then the right hand side of (3.41) in Lemma 3.3 finishes up the proof. □

Lemma 3.5. *Let e be an edge of triangle $K \in T_h$, and θ^h be as in* (3.25)*. Then for every $\chi \in X^h$ there a exists constant $C > 0$ independent of ϵ and h such that*

$$\int_e A(x/\epsilon) \, \epsilon \nabla \theta^h \cdot \vec{n} \ (\chi - I_h\chi) \ ds \leq \ C \frac{\epsilon}{h} |v_\epsilon^h|_{H^1(K)} |\chi|_{H^1(K)}. \tag{3.60}$$

Proof. Using (3.25), Lemma 3.2, and the fact that χ is linear on e, we have

$$\begin{aligned} \int_e A(x/\epsilon) \, \epsilon \nabla \theta^h \cdot \vec{n} \ (\chi - I_h\chi) \ ds &\leq \ C \frac{\epsilon}{h} |\nabla v_0^h| \int_e |\chi - I_h\chi| ds \\ &\leq \ C \frac{\epsilon}{h} |\nabla v_0^h| \, h \, \|\chi - I_h\chi\|_{L_\infty(e)} \\ &\leq \ C \frac{\epsilon}{h} |v_0^h|_{H^1(K)} |\nabla \chi| \, h \\ &\leq C \frac{\epsilon}{h} |v_0^h|_{H^1(K)} |\chi|_{H^1(K)}, \end{aligned} \tag{3.61}$$

where we have used inverse inequalities for ∇v_0^h and $\nabla \chi$. Using the right hand side inequality of (3.41) in Lemma 3.3 the proof is complete. □

Theorem 3.1. *For $v_\epsilon^h \in V_\epsilon^h$ we have*

$$|D(v_\epsilon^h, \chi)| \leq \ C_d \frac{\epsilon}{h} |v_\epsilon^h|_{1,h} |\chi|_{H^1} \quad \forall \, \chi \in X^h. \tag{3.62}$$

Proof. Considering (3.47), we may break up the integral over ∂K into a sum of integral over the edges e. Then the estimate is obtained by straightforward application of Lemmas 3.4 and 3.5. □

3.5.2. Inf-Sup Conditions and Error Estimates

We start by establishing the inf-sup condition of the finite element bilinear form. A similar proof has also been presented in [62]. Moreover, in [62] the authors derived an L_2-error estimate for the two-scale nonconforming Petrov-Galerkin finite element and demonstrated the smallness of the nonconforming error.

Lemma 3.6. *Assume that* (3.40) *holds. Then the finite element bilinear form* (3.35) *satisfies the inf-sup condition, i.e., for* $v_\epsilon^h \in V_\epsilon^h$ *we have*

$$\sup_{\chi \in X^h} \frac{a_{FE}(v_\epsilon^h, \chi)}{\|\chi\|_{H^1}} \geq C_{fe} \, \|v_\epsilon^h\|_{1,h} \tag{3.63}$$

for some constant $C_{fe} > 0$ *independent of* ϵ, h.

Proof. Let $M(x/\epsilon) = \{M_{ij}(x/\epsilon)\}$ be a 2×2 matrix such that its entries are defined as

$$M_{ij} = B_{ij} + \epsilon A_{ik} \nabla_k \eta_j, \tag{3.64}$$

where B_{ij} are as defined in (3.39), η_j is defined in (3.19), and as before we have used the Einstein summation appropriately. Using (3.46) and the fact that $\theta^h = \eta_k \, \nabla_k v_0^h$, cf. (3.25), we may rewrite $a_{FE}(v_\epsilon^h, \chi)$ in (3.36) as

$$a_{FE}(v_\epsilon^h, \chi) = \sum_{K \in T_h} \big(M(x/\epsilon) \nabla v_0^h, \nabla \chi \big)_K. \tag{3.65}$$

Now consider an arbitrary nonzero vector $\xi \in \mathbb{R}^2$. For sufficiently small ϵ we may use Lemma 3.2 and the assumption in (3.40) on B_{ij} to obtain the following estimate:

$$\begin{aligned} \xi_i \, M_{ij} \, \xi_j &= \xi_i \, (B_{ij} + \epsilon A_{ik} \nabla_k \eta_j) \, \xi_j \\ &\geq (c_1 - c \, \frac{\epsilon}{h}) \, |\xi|^2 \\ &\geq C \, |\xi|^2, \end{aligned} \tag{3.66}$$

where C in the last line is independent of ϵ and h. Thus, by taking $\chi = v_0^h$ an using

(3.66) we have

$$\sup_{\chi \in X^h} \frac{a_{FE}(v_\epsilon^h, \chi)}{\|\chi\|_{H^1}} \geq C \, \frac{|v_0^h|_{1,h}^2}{\|v_0^h\|_{1,h}}. \tag{3.67}$$

Left hand side inequality of (3.41) in Lemma 3.3 and the Poincaré-Friedrichs inequality complete the proof. □

The inf-sup condition (3.63) guarantees that there exists a unique solution of the two-scale finite element problem (3.34). Next lemma is devoted to establishing the inf-sup condition of the bilinear form of the two-scale finite volume element. The proof uses a standard procedure for the finite volume element method perturbation argument [50].

Lemma 3.7. *For sufficiently small ratio ϵ/h, the finite volume element bilinear form* (3.32) *satisfies inf-sup condition, i.e., for $v_\epsilon^h \in V_\epsilon^h$ there exists a constant $C_{fv} > 0$ such that*

$$\sup_{\chi \in X^h} \frac{a_{FV}(v_\epsilon^h, I_h\chi)}{\|\chi\|_{H^1}} \geq C_{fv} \, \|v_\epsilon^h\|_{1,h}. \tag{3.68}$$

Proof. Using (3.38) we may write

$$a_{FV}(v_\epsilon^h, I_h\chi) = a_{FE}(v_\epsilon^h, \chi) - D(v_\epsilon^h, \chi). \tag{3.69}$$

By Lemma 3.6 and Theorem 3.1 we have

$$\sup_{\chi \in X^h} \frac{a_{FV}(v_\epsilon^h, I_h\chi)}{\|\chi\|_{H^1}} \geq \left(C_{fe} - C_d \, \frac{\epsilon}{h} \right) \, \|v_\epsilon^h\|_{1,h}. \tag{3.70}$$

Thus for sufficiently small ϵ/h we have $C_{fv} = C_{fe} - C_d \, \epsilon/h$ positive. □

Hence, as in the finite element case, we guarantee the existence and uniqueness of the two-scale finite volume element solution by this inf-sup condition. We note that the following lemma is a consequence of Lemma 3.7.

Lemma 3.8. *Let $u_\epsilon^h \in V_\epsilon^h$ be the solution of* (3.33) *associated with* (3.3)*. Then*

$$\|u_\epsilon^h\|_{1,h} \leq C\|f\|_{L_2}. \tag{3.71}$$

Proof. By Lemma 3.7 and (3.33) we have

$$C_{fv}\|u_\epsilon^h\|_{1,h} \leq \sup_{\chi \in X^h} \frac{(f, I_h\chi)}{\|\chi\|_{H^1}}. \tag{3.72}$$

Now using the Cauchy-Schwarz inequality and (3.7) we have the result. □

Next we show that the difference between the two-scale finite volume element and two-scale finite element solutions is small.

Lemma 3.9. *Let $u_\epsilon^h \in V_\epsilon^h$ be the solution of* (3.33)*, and $\tilde{u}_\epsilon^h \in V_\epsilon^h$ be the solution of* (3.34)*, both associated with* (3.3)*. Then we have*

$$\|\tilde{u}_\epsilon^h - u_\epsilon^h\|_{1,h} \leq \left(C_1\, h + C_2\, \frac{\epsilon}{h}\right) \|f\|_{L_2}. \tag{3.73}$$

Proof. First we introduce a bilinear form

$$d(f, \chi) = (f, \chi - I_h\chi) \quad \forall\, f \in L_2, \chi \in X^h. \tag{3.74}$$

This bilinear form has the following approximation property [7, Lemma 5.1]:

$$|d(f, \chi)| \leq C\, h\, \|f\|_{L_2}\, \|\chi\|_{H^1}. \tag{3.75}$$

Using (3.74) and (3.38), we may write

$$a_{FE}(\tilde{u}_\epsilon^h - u_\epsilon^h, \chi) = d(f, \chi) - D(u_\epsilon^h, \chi) \quad \forall\, \chi \in X^h. \tag{3.76}$$

The terms on the right hand side of this equation are estimated in (3.75) and Theorem

3.1. Dividing both sides by $\|\chi\|_{H^1}$ and taking supremum over all χ we have

$$\sup_{\chi\in X^h} \frac{a_{FE}(\tilde{u}_\epsilon^h - u_\epsilon^h, \chi)}{\|\chi\|_{H^1}} \leq \left(C_1 h \|f\|_{L_2} + C_d \frac{\epsilon}{h} \|u_\epsilon^h\|_{1,h} \right). \tag{3.77}$$

But Lemma 3.8 guarantees the boundedness of u_ϵ^h, and thus by Lemma 3.7 we have the result. □

In the next two theorems we establish variations of Céa's Lemma, one for the two-scale finite element solution, and the other for the two-scale finite volume element solution.

Theorem 3.2. *Let u_ϵ and $\tilde{u}_\epsilon^h$ be the exact solution of boundary value problem* (3.3) *and the solution of two-scale finite element* (3.34)*, respectively. Then*

$$\|u_\epsilon - \tilde{u}_\epsilon^h\|_{1,h} \leq (1 + C_{fe}) \inf_{v_\epsilon^h \in V_\epsilon^h} \|u_\epsilon - v_\epsilon^h\|_{1,h}. \tag{3.78}$$

Proof. Let $v_\epsilon^h \in V_\epsilon^h$ and $\chi \in X^h$. We have $a_{FE}(\tilde{u}_\epsilon^h - v_\epsilon^h, \chi) = a_{FE}(u_\epsilon - v_\epsilon^h, \chi) + a_{FE}(\tilde{u}_\epsilon^h, \chi) - (f, \chi)$, where the last two terms cancel each other. Using this fact and in view of Lemma 3.6 we have

$$\begin{aligned} \|\tilde{u}_\epsilon^h - v_\epsilon^h\|_{1,h} &\leq C_{fe} \sup_{\chi\in X^h} \frac{a_{FE}(u_\epsilon - v_\epsilon^h, \chi)}{\|\chi\|_{H^1}} \\ &\leq C_{fe} \|u_\epsilon - v_\epsilon^h\|_{1,h}. \end{aligned} \tag{3.79}$$

The result follows from the triangle inequality $\|u_\epsilon - \tilde{u}_\epsilon^h\|_{1,h} \leq \|u_\epsilon - v_\epsilon^h\|_{1,h} + \|\tilde{u}_\epsilon^h - v_\epsilon^h\|_{1,h}$ and by taking the infimum over all elements of V_ϵ^h. □

Theorem 3.3. *Let u_ϵ and u_ϵ^h be the exact solution of boundary value problem* (3.3) *and the solution of the two-scale finite volume element* (3.33)*, respectively. Then,*

$$\|u_\epsilon - u_\epsilon^h\|_{1,h} \leq \left(C_1\, h + C_2\, \frac{\epsilon}{h} \right) \|f\|_{L_2} + C_3 \inf_{v_\epsilon^h \in V_\epsilon^h} \|u_\epsilon - v_\epsilon^h\|_{1,h}. \tag{3.80}$$

Proof. Let $\tilde{u}_\epsilon^h$ be be the solution of (3.34). Using triangle inequality we have

$$\|u_\epsilon - u_\epsilon^h\|_{1,h} \leq \|u_\epsilon - \tilde{u}_\epsilon^h\|_{1,h} + \|\tilde{u}_\epsilon^h - u_\epsilon^h\|_{1,h}.$$

The results follow directly from Theorem 3.2 and Lemma 3.9. □

As we can see from Theorems 3.2 and 3.3, we need to estimate the minimizing value of $\|u_\epsilon - v_\epsilon^h\|_{1,h}$ which is taken over all elements of the space V_ϵ^h. For this purpose we take an element v_ϵ^h of V_ϵ^h that its homogenized part v_0^h interpolates the homogenized part of the exact solution of (3.3).

Lemma 3.10. *Let u_ϵ be the exact solution of* (3.3)*, and u_0 be its homogenized part. Choose v_ϵ^h an element of V_ϵ^h such that for each triangle $K \in T_h$, $v_0^h(z) = u_0(z)$, $z \in Z_h(K)$, i.e., the homogenized part of v_ϵ^h coincides with the homogenized part of u_ϵ on the vertices of triangles $K \in T_h$. Then there exists a constant $C > 0$ independent of ϵ and h such that*

$$\|u_\epsilon - v_\epsilon^h\|_{1,h} \leq C \left(h\,|u_0|_{H^2} + \frac{\epsilon}{h}\,|u_0|_{H^1} + \sqrt{\epsilon} \right). \tag{3.81}$$

Proof. By definition of the "broken" energy norm, it suffices to establish the estimate over a triangle K. Using the expansions (3.13) and (3.26), we have

$$u_\epsilon - v_\epsilon^h = (u_0 - v_0^h) + \epsilon N_k \nabla_k (u_0 - v_0^h) + \epsilon\,\theta_u + \epsilon\,\theta^h. \tag{3.82}$$

It is well known that since v_0^h is linear on K, the following estimate holds:

$$|u_0 - v_0^h|_{H^1(K)} \leq C\,h\,|u_0|_{H^2(K)}. \tag{3.83}$$

Now since $A_{ij} \in W^{1,p}(Y)$, $p > 2$, we have that $\epsilon\,\nabla N_k$ is locally bounded. Hence

$$\begin{aligned} |\epsilon N_k \nabla_k (u_0 - v_0^h)|_{H^1(K)} &\leq \max\left\{ \epsilon\,\|\nabla N_1\|_{L_\infty(K)}, \epsilon\,\|\nabla N_2\|_{L_\infty(K)} \right\} |u_0 - v_0^h|_{H^1(K)} \\ &\leq C\,h\,|u_0|_{H^2(K)}. \end{aligned} \tag{3.84}$$

Next using (3.25) and applying Lemma 3.2, we have

$$\epsilon^2 \, |\theta^h|^2_{H^1(K)} \leq \epsilon^2 \int_K |\nabla \eta_k|^2 \, |\nabla_k v_0^h|^2 \, dx \leq C \, \frac{\epsilon^2}{h^2} \, |v_0^h|^2_{H^1(K)}. \tag{3.85}$$

Moreover, it is clear that using triangle inequality, (3.83) we have

$$|v_0^h|_{H^1(K)} \leq C \, h |u_0|_{H^2(K)} + |u_0|_{H^1(K)}.$$

Finally, summing up over all triangles $K \in T_h$ and using Lemma 3.1 to estimate θ^u, we obtain the desired estimate.

□

From Theorem 3.2, Theorem 3.3, and Lemma 3.10 we immediately obtain the following:

Corollary 3.1. *Let u_ϵ and $\tilde{u}_\epsilon^h$ be the solutions of* (3.3) *and* (3.34)*, respectively. Then there exist constants $C_i > 0$, $i = 1, 2, 3$, independent of ϵ and h such that*

$$\|u_\epsilon - \tilde{u}_\epsilon^h\|_{1,h} \leq C_1 \, h + C_2 \, \frac{\epsilon}{h} + C_3 \, \sqrt{\epsilon}. \tag{3.86}$$

Corollary 3.2. *Let u_ϵ and u_ϵ^h be the solutions of* (3.3) *and* (3.33)*, respectively. Then there exist constants $C_i > 0$, $i = 1, 2, 3$, independent of ϵ and h such that*

$$\|u_\epsilon - u_\epsilon^h\|_{1,h} \leq C_1 \, h + C_2 \, \frac{\epsilon}{h} + C_3 \, \sqrt{\epsilon}. \tag{3.87}$$

Therefore, both finite element and finite volume element for two-scale method have the same asymptotic convergence rates.

3.6. Numerical Examples

In this section we present numerical experiments to assess the performance of the two-scale finite volume element method. A convergence test of the method is reported

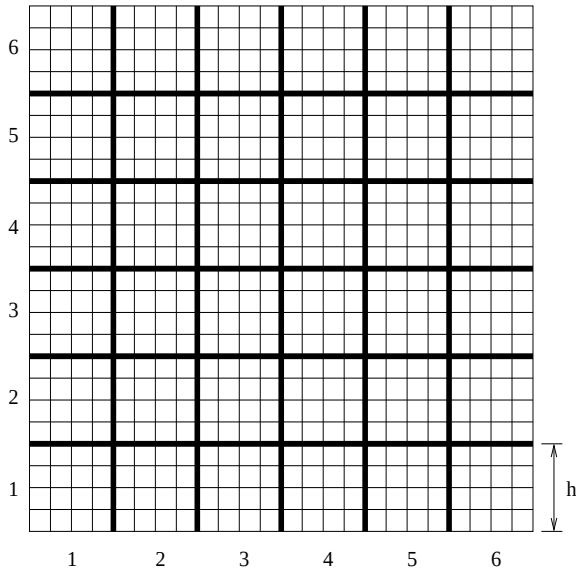

Fig. 3.5. Discretization of the domain into two-scale meshes.

which is followed by an application to flow in porous media. In all of these computations, we have used finely resolved numerical solutions obtained using finite volume method as reference solutions. This is because it is extremely hard to come up with a two-scale boundary value problem which has an exact solution. All the examples below use a unit square domain, $\Omega = (0,1) \times (0,1)$. The computation is implemented on a uniform rectangular mesh. The construction of the two-scale mesh is described as follows. Suppose the domain Ω is discretized into rectangular elements with step size $1/n_f$ in each direction. Then the rectangular coarse elements are constructed with the size $h = 1/N$ in each direction (represented by the bold lines in Figure 3.5). Now, each coarse element consists of the fine rectangular sub-elements of size h/n, where we have the relation $1/n_f = (1/N)(1/n)$. Having this kind of construction, we have N is the number of coarse elements and n the number of sub-elements in a coarse element, both for each direction in the domain.

3.6.1. Convergence Test

For the convergence test, the methods are tested by solving (3.3) with the periodic coefficient cf. [42]

$$A(x/\epsilon) = \frac{2 + 1.8\sin(2\pi x_1/\epsilon)}{2 + 1.8\cos(2\pi x_2/\epsilon)} + \frac{2 + \sin(2\pi x_2/\epsilon)}{2 + 1.8\cos(2\pi x_1/\epsilon)}$$

and

$$f = -1.$$

In the following, the error is denoted by $e = u_\epsilon - u_\epsilon^h$, TS-FV denotes the two-scale finite volume element method using conforming basis functions and TS-FV-O denotes the two-scale finite volume element method with oversampling. To investigate the interaction between the components of error written in Corrolary 3.2, we show three sets of scenario whose results are listed in three different tables. The first scenario deals with a constant ϵ while varying the number of coarse elements N. For this, the reference solution is resolved on a very fine mesh with a step size $1/2048 = 2^{-11}$ in each direction. The second scenario deals with a constant ratio ϵ/h and a constant number of sub-elements n. Thus, once an ϵ is given, the number of coarse element may be obtained from the specified ratio. Consequently, for each case in the second scenario, the number of total elements used to resolve the reference solution would be different. Finally, the third scenario uses a constant number of coarse elements N, while varying the ϵ (and consequently varying n also).

All results pertaining to the error of the solution in H^1 norm are listed in Tables 3.1, 3.2, and 3.3. In general we see a significant improvement using the oversampling strategy (TS-FV-O). Table 3.1 shows comparison of the H^1 norm of the error of the approximation taken against the number of elements N and n with a constant ϵ equal to 0.005. Obviously, TS-FV gives the worst results for fixed $n_f = N \times n$ with n

Table 3.1. Comparison of H^1 seminorm of solution error for $\epsilon = 0.005$.

N	n	TS-FV	TS-FV-O	
		$\|e\|_1$	$\|e\|_1$	Rate
32	64	2.142806×10^{-2}	8.188009×10^{-3}	-
64	32	2.926952×10^{-2}	4.114026×10^{-3}	0.99
128	16	4.473100×10^{-2}	2.288907×10^{-3}	0.85
256	8	5.951678×10^{-2}	1.911220×10^{-3}	0.26

decreasing since we have introduced more intercourse finite element boundaries, which in turn generate some errors. It may be seen from Table 3.1, that when preserving the ϵ, and letting the coase step size h decreases, the convergence in TS-FV-O (also in TS-FV) deteriorates when $\epsilon/h \approx 1$. It should be pointed out that for this regime, Corrolary 3.2 might not be true anymore. In Table 3.2 we present the corresponding error in the case of the ratio $\epsilon/h = 0.64$ and $n = 16$. From Table 3.2 we see that the first order convergence for TS-FV-O is relatively maintained irrespective of the value of h. This phenomenon gives a hint that the $O(\epsilon)$ constant (C_3) in Corrolary 3.2 is smaller than the $O(h)$ constant (C_1). Finally Table 3.3 gives the comparison for a constant $N = 32$ and varying ϵ. The table indicates that the TS-FV-O errors do not change significantly compared to TS-FV errors, which suggests that the oversampling strategy has reduced the resonance error inherent in the original two-scale method. Similar comparisons for L_2 norm errors are presented in Tables 3.4, 3.5, and 3.6. It is apparent that they exhibit similar behaviors as in H_1 norm. This finding is consistent with the investigation conducted in [62].

Table 3.2. Comparison of H^1 seminorm of solution error for $\epsilon/h = 0.64$ and $n = 16$.

N	ϵ	TS-FV	TS-FV-O	
		$\|e\|_1$	$\|e\|_1$	Rate
16	0.040	5.031755×10^{-2}	2.419640×10^{-2}	-
32	0.020	4.510508×10^{-2}	8.427971×10^{-3}	1.52
64	0.010	4.475054×10^{-2}	4.388929×10^{-3}	0.94
128	0.005	4.473100×10^{-2}	2.288907×10^{-3}	0.94

Table 3.3. Comparison of H^1 seminorm of solution error for $N = 32$.

ϵ	n	TS-FV	TS-FV-O
		$\|e\|_1$	$\|e\|_1$
0.020	16	4.510508×10^{-2}	8.427971×10^{-3}
0.010	32	2.975713×10^{-2}	8.195283×10^{-3}
0.005	64	2.142806×10^{-2}	8.188009×10^{-3}

Table 3.4. Comparison of L_2 norm of solution error for $\epsilon = 0.005$.

N	n	TS-FV	TS-FV-O	
		$\|\|e\|\|$	$\|\|e\|\|$	Rate
32	64	8.735775×10^{-5}	1.938853×10^{-5}	-
64	32	1.720292×10^{-4}	4.812917×10^{-6}	2.01
128	16	2.941193×10^{-4}	2.336342×10^{-6}	1.04
256	8	3.683877×10^{-4}	6.251241×10^{-7}	1.90

Table 3.5. Comparison of L_2 norm of solution error for $\epsilon/h = 0.64$, and $n = 16$.

N	ϵ	TS-FV	TS-FV-O	
		$\|e\|$	$\|e\|$	Rate
16	0.040	3.898167×10^{-4}	8.649052×10^{-5}	-
32	0.020	3.172062×10^{-4}	2.146057×10^{-5}	2.01
64	0.010	2.986045×10^{-4}	5.077901×10^{-6}	2.08
128	0.005	2.941193×10^{-4}	2.336342×10^{-6}	1.12

Table 3.6. Comparison of L_2 norm of solution error for $N = 32$.

ϵ	n	TS-FV	TS-FV-O
		$\|e\|$	$\|e\|$
0.020	16	3.172062×10^{-4}	2.146057×10^{-5}
0.010	32	2.086535×10^{-4}	1.886330×10^{-5}
0.005	64	8.735775×10^{-5}	1.938853×10^{-5}

3.6.2. Application to Flow in Porous Media

In this subsection we present an application of the two-scale finite volume element method to a flow in porous medium. The problem considered is typical representation of a cross section of a subsurface. In this case, (3.3) governs a pressure distribution over the domain. As before we set our domain $\Omega = (0,1)\times(0,1)$, with a given pressure on the left and right boundaries, i.e., $u(x_1 = 0, x_2) = 1$, and $u(x_1 = 1, x_2) = 0$ while the top and bottom boundaries are closed to flow, i.e. $u_{x_2}(x_1, x_2 = 0) = u_{x_2}(x_1, x_2 = 1) = 0$. As an exact solution we have used a fine solution with step size $1/1024 = 2^{-10}$.

Moreover, the matrix A is set to be a diagonal matrix with $A_{ii}(x) = k(x)$. Instead of using a periodic functions, we use $k(x)$ as a set of randomly generated numbers realized in 1025×1025 grid points, given its correlation structures (l_{x_1}, and l_{x_2}), covariance model and overall variance quantified via σ^2 which is the variance of $\log k$. We consider a GSLIB model developed in [16].

In the examples below, we concentrate on the anisotropic case, which in practical applications is the most difficult to upscale. We have used $l_{x_1} = 0.4$ and $l_{x_2} = 0.01$ with an exponential covariance model, and $\sigma = 1.0$. Table 3.7 presents the pressure error $e = \|u - u^h\|$ and the error of the velocity in the horizontal direction, $e_{x_1} = \|k(u_{x_1} - u^h_{x_1})\|$, and their corresponding relative errors, $e_r = e/\|u\|$, and $e_{x_1,r} = e_{x_1}/\|ku_{x_1}\|$. We note that many engineering problems require an accurate prediction of the velocity. On the second example shown in Table 3.8, we use the same correlation lengths and structures, and $\sigma = 1.5$. In all examples, we see that as h decreases, the errors decrease as well. Comparison of the visualized horizontal velocities are shown in Figure 3.6.

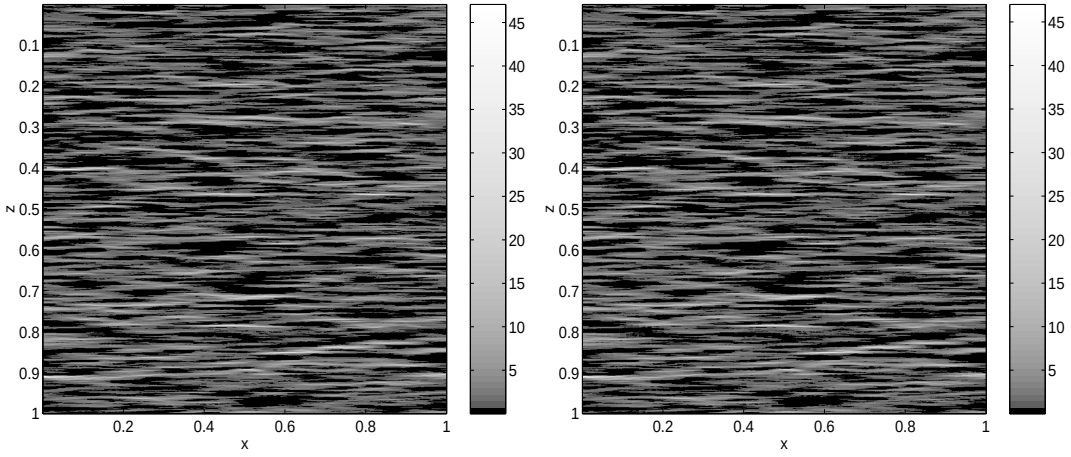

Fig. 3.6. Comparison of horizontal velocity for anisotropic absolute permeability with $\sigma = 1.5$: (left) finely resolved model with 1024×1024 elements, (right) two-scale FVE with 64×64 coarse elements

Table 3.7. Results for anisotropic case, $l_{x_1} = 0.40$, $l_{x_2} = 0.01$, $\sigma = 1.0$.

N	n	e	e_r (%)	e_{x_1}	$e_{x_1,r}$ (%)
32	32	2.260724×10^{-4}	0.04	3.532075×10^{-2}	1.97
64	16	1.198503×10^{-4}	0.02	2.741758×10^{-2}	1.53
128	8	8.155836×10^{-5}	0.01	2.305173×10^{-2}	1.28
256	4	5.907592×10^{-5}	0.01	1.928818×10^{-2}	1.07

Table 3.8. Results for anisotropic case, $l_{x_1} = 0.40$, $l_{x_2} = 0.01$, $\sigma = 1.5$.

N	n	e	e_r (%)	e_{x_1}	$e_{x_1,r}$ (%)
32	32	8.140234×10^{-4}	0.14	1.197157×10^{-1}	3.80
64	16	4.406654×10^{-4}	0.08	8.694687×10^{-2}	2.76
128	8	3.198741×10^{-4}	0.06	6.950237×10^{-2}	2.20
256	4	2.022701×10^{-4}	0.04	5.643796×10^{-2}	1.79

CHAPTER IV

ANALYSIS OF NUMERICAL HOMOGENIZATION FOR A NONLINEAR ELLIPTIC PROBLEM

4.1. The Framework

In this chapter we investigate the convergence of the numerical homogenization for the nonlinear elliptic equation (2.1). We will present a convergence analysis for numerical homogenization designed for the finite element variational formulation. For simplicity we will confine ourselves to the following boundary value problem: find $u_\epsilon \in W_0^{1,p}(\Omega)$, with $p \geq 2$, satisfying

$$-\nabla \cdot (a_\epsilon(x, u_\epsilon, \nabla u_\epsilon)) = f \quad \text{in } \Omega \subset \mathbb{R}^2, \tag{4.1}$$

where ∇ denotes the gradient and $a_\epsilon : \mathbb{R}^2 \times \mathbb{R} \times \mathbb{R}^2 \to \mathbb{R}^2$.

As before, let T_h be the collection of finite elements partitioning Ω and X^h be a standard finite element space that lives on T_h as described in Chapter II. The numerical homogenization scheme in the finite element setting associated with problem (4.1) is formulated as to seek $u^h \in X^h$ such that

$$\langle A_\epsilon^h\, u^h, w^h \rangle = \int_\Omega f\, w^h\, dx, \quad \forall\ w^h \in X^h, \tag{4.2}$$

where

$$\langle A_\epsilon^h\, v^h, w^h \rangle = \sum_{K \in T_h} \int_K (a_\epsilon(x, \eta^h, \nabla v_\epsilon), \nabla w^h)\, dx,$$

$$\eta^h(x) = \sum_{K \in T_h} \eta_K\, \Psi_K(x) \quad \text{with} \quad \eta_K = \frac{1}{|K|} \int_K v^h\, dx.$$

Here Ψ_K is the characteristic function of K and v_ϵ satisfies the following problem:

$$-\nabla \cdot (a_\epsilon(x, \eta^h, \nabla v_\epsilon)) = 0 \text{ in } K \in T_h \quad \text{and} \quad v_\epsilon = v^h \text{ on } \partial K. \tag{4.3}$$

As in Chapter II, we denote by V_ϵ^h the space of all functions v_ϵ satisfying (4.3), and by $E : X^h \to V_\epsilon^h$ the multiscale map associated with (4.3).

Remark 4.1. *For general elliptic problems, which include the lower-order term as in* (2.1)*, the corresponding numerical homogenization in the finite element setting is to seek an* $u^h \in X^h$ *such that*

$$\langle A_\epsilon^h\, u^h, w^h\rangle = \int_\Omega f\, w^h\, dx, \quad \forall\ w^h \in X^h,$$

where

$$\langle A_\epsilon^h\, v^h, w^h\rangle = \sum_{K\in T_h} \int_K (a_\epsilon(x, \eta^h, \nabla v_\epsilon), \nabla w^h)\, dx + \int_K b_\epsilon(x, \eta^h, \nabla v_\epsilon)\, w^h\, dx.$$

The analysis presented below can be extended to treat this numerical homogenization as well.

4.2. Main Results from the Convergence Analysis

4.2.1. Setting for the Analysis

One main assumption for the analysis is that the mesh size is greater than and depending on ϵ, i.e., $h = h(\epsilon) \gg \epsilon$, with $h(\epsilon) \to 0$ as $\epsilon \to 0$. Regarding the elliptic coefficient a_ϵ, we set several assumptions, namely, that $a_\epsilon(x, \cdot, \cdot) = a(x/\epsilon, \cdot, \cdot)$, where $a(y, \cdot, \cdot)$ is a periodic function in a unit square Y and satisfies the following properties:

N1 polynomial growth:

$$|a(\cdot, \eta, \xi)| \le c_0\,(1 + |\eta|^{p-1} + |\xi|^{p-1}), \quad \forall\ \eta \in \mathbb{R},\ \xi \in \mathbb{R}^2; \tag{4.4}$$

N2 monotonicity with respect to ξ:

$$(a(\cdot, \cdot, \xi_1) - a(\cdot, \cdot, \xi_2), \xi_1 - \xi_2) \ge c_1\, |\xi_1 - \xi_2|^p, \quad \forall\ \xi_1,\ \xi_2 \in \mathbb{R}^2; \tag{4.5}$$

N3 coercivity:

$$(a(\cdot,\cdot,\xi),\xi) \geq c_2|\xi|^p, \quad \forall\ \xi \in \mathbb{R}^2; \tag{4.6}$$

N4 continuity:

Denote

$$H(\eta_1,\xi_1,\eta_2,\xi_2,r) = (1+|\eta_1|^r+|\eta_2|^r+|\xi_1|^r+|\xi_2|^r), \tag{4.7}$$

for arbitrary $\eta_1,\ \eta_2 \in \mathbb{R}$, $\xi_1,\ \xi_2 \in \mathbb{R}^2$, and $r > 0$. Then

$$\begin{aligned} |a(\cdot,\eta_1,\xi_1) - a(\cdot,\eta_2,\xi_2)| \leq c_3\, H(\eta_1,\xi_1,\eta_2,\xi_2,p-1)\,\nu(|\eta_1-\eta_2|) + \\ c_4\, H(\eta_1,\xi_1,\eta_2,\xi_2,p-1-s)\,|\xi_1-\xi_2|^s, \end{aligned} \tag{4.8}$$

where $s > 0$, $s \in (0, \min(p-1,1))$ and ν is the modulus of continuity, which is bounded, concave, and continuous in $\mathbb{R}_+$, and $\nu(0) = 0$, $\nu(t) = 1$ for $t \geq 1$ and $\nu(t) > 0$ for $t > 0$.

Homogenization theory, e.g. [52], states that u_ϵ converges weakly to $u \in W_0^{1,p}(\Omega)$ as $\epsilon \to 0$, which satisfies

$$-\nabla \cdot (a^*(u,\nabla u)) = f, \tag{4.9}$$

i.e., $u \in W_0^{1,p}(\Omega)$ satisfies

$$\langle A^*\, u, w\rangle = \int_\Omega f\, w\, dx \quad \forall\ w \in W_0^{1,q}(\Omega), \tag{4.10}$$

where

$$\langle A^*\, v, w\rangle = \sum_{K\in T_h} \int_K (a^*(v,\nabla v),\nabla w)\, dx. \tag{4.11}$$

Here a^* is the homogenized coefficient defined as

$$a^*(\eta,\xi) = \frac{1}{|Y|}\int_Y a(y,\eta,\xi+\nabla_y N_{\eta,\xi}(y))\, dy, \tag{4.12}$$

and $N_{\eta,\xi} \in W^{1,p}_{per}(Y)$ is the solution of

$$-\nabla \cdot (a(y, \eta, \xi + \nabla_y N_{\eta,\xi}(y))) = 0. \tag{4.13}$$

Note that the solution of this equation is uniquely defined up to a constant.

Remark 4.2. *It is known [52] that the homogenized coefficient $a^*(\eta, \xi)$ satisfies the following properties:*

NH1 *polynomial growth:*

$$|a^*(\eta, \xi)| \leq c_0 \, (1 + |\eta|^{p-1} + |\xi|^{p-1}), \quad \forall \ \eta \in \mathbb{R}, \ \xi \in \mathbb{R}^2; \tag{4.14}$$

NH2 *monotonicity with respect to ξ:*

$$(a^*(\cdot, \xi_1) - a^*(\cdot, \xi_2), \xi_1 - \xi_2) \geq c_1 \, |\xi_1 - \xi_2|^p, \quad \forall \ \xi_1, \ \xi_2 \in \mathbb{R}^2; \tag{4.15}$$

NH3 *coercivity:*

$$(a^*(\cdot, \xi), \xi) \geq c_2 |\xi|^p, \quad \forall \ \xi \in \mathbb{R}^2; \tag{4.16}$$

NH4 *continuity:*

$$\begin{aligned} |a^*(\eta_1, \xi_1) - a^*(\eta_2, \xi_2)| \leq c_3 \, H(\eta_1, \xi_1, \eta_2, \xi_2, p-1) \, \nu(|\eta_1 - \eta_2|) + \\ c_4 \, H(\eta_1, \xi_1, \eta_2, \xi_2, p-1-\tilde{s}) \, |\xi_1 - \xi_2|^{\tilde{s}}, \end{aligned} \tag{4.17}$$

for arbitrary $\eta_1, \ \eta_2 \in \mathbb{R}$, $\xi_1, \ \xi_2 \in \mathbb{R}^2$, where $\tilde{s} \in (0, \min(p-1, 1))$ and H is as in the property **N4**.

4.2.2. Alternative Formulation

Taking advantage of the periodicity of the coefficients, it is possible to solve the local problem (4.3) in a single period domain Y_ϵ instead of in the element K. We can carefully set an appropriate periodic boundary condition on ∂Y_ϵ. This obviously

gives a significant CPU saving in the numerical computation. To write the alternative formulation, we denote by $I_\epsilon^K = \{i \in \mathbb{Z}^n : Y_\epsilon^i \subset K\}$ for each $K \in T_h$. Then the numerical homogenization associated with (4.1) is the same as in (4.2), except now the operator A_ϵ^h is written as

$$\langle A_\epsilon^h v^h, w^h \rangle = \sum_{K \in T_h} \sum_{i \in I_\epsilon^K} \int_{Y_\epsilon^i} (a_\epsilon(x, \eta^h, \nabla v_\epsilon), \nabla w^h)\, dx.$$

We note that the proofs presented in the next subsections are directly applicable to this case. In general, it is also a common practice to solve the local problem in a representative elementary volume for each element K. However, since our objective is to perform numerical homogenization for elliptic equations with general heterogeneous coefficients, we refrain from using this alternative formulation.

4.2.3. Main Results

The first theorem states the convergence of the numerical homogenization solution to the exact homogenized solution.

Theorem 4.1. *Let u^h be the solution of numerical homogenization* (4.2) *and u be the solution of the homogenized problem* (4.9). *Then* $\lim_{\epsilon \to 0} \|u^h - u\|_{W^{1,p}(\Omega)} = 0$.

Furthermore, the following theorem gives the convergence of the fluctuation $v_\epsilon \in V_\epsilon^h$ to the exact solution u_ϵ.

Theorem 4.2. *Let u_ϵ be the solution of boundary value problem* (4.1) *and $u^h \in X^h$ and $v_\epsilon \in V_\epsilon^h$ with $v_\epsilon = E(u^h)$ be the numerical homogenization solution (homogenized and fluctuating components, respectively)* (4.2), *Then* $\lim_{\epsilon \to 0} \|\nabla v_\epsilon - \nabla u_\epsilon\|_{L_p(\Omega)} = 0$.

Now we consider a problem associated with (4.1) when the coefficient a depends only on the gradient of solution, i.e., $a(\cdot, \eta, \xi) \equiv a(\cdot, \xi)$, with $\xi \in \mathbb{R}^2$. Obviously the

monotonicity property **N2** implies that the corresponding operator A_ϵ^h is monotone. Then it is possible to derive an order of convergence for the numerical homogenization solution.

Theorem 4.3. *Let u and u^h be the solutions of the homogenized problem* (4.9) *and numerical homogenization* (4.2)*, respectively, with the coefficient a independent of u_ϵ, i.e., $a(\cdot,\eta,\xi) \equiv a(\cdot,\xi)$. Then there exist constants $C_j > 0$, $j = 1, \cdots, 3$ independent of ϵ and h such that*

$$\|\nabla u^h - \nabla u\|_{L_p(\Omega)} \leq C_1 \left(\frac{\epsilon}{h}\right)^{\frac{s}{(p-1)(p-s)}} + C_2 \left(\frac{\epsilon}{h}\right)^{\frac{1}{p}} + C_3\, h^{\frac{1}{p-1}}.$$

Remark 4.3. *If $p \leq p_0$, $p_0 \approx 2.6$, then the second term on the convergence rate dominates the first one and the convergence rate is given by*

$$C_2 \left(\frac{\epsilon}{h}\right)^{\frac{1}{p}} + C_3\, h^{\frac{1}{p-1}}$$

while if $p \geq p_0$ the convergence rate is given by

$$C_1 \left(\frac{\epsilon}{h}\right)^{\frac{s}{(p-1)(p-s)}} + C_3\, h^{\frac{1}{p-1}}.$$

In particular, for large p the resonance error is defined by $(\epsilon/h)^{1/p^2}$.

4.3. Proofs of the Theorems

4.3.1. Several Auxiliary Results

The following coercivity holds for A_ϵ^h:

Lemma 4.1. *There exists a constant $c > 0$ such that*

$$\langle A_\epsilon^h\, v^h, v^h \rangle \geq c\, \|\nabla v^h\|_{L_p(\Omega)}^p \;\forall\, v^h \in X^h.$$

Proof. Let $\tilde{v_\epsilon} = v_\epsilon - v^h$. It follows that $\tilde{v_\epsilon}$ satisfies the following problem:

$$-\nabla \cdot (a(x/\epsilon, \eta^h, \nabla \tilde{v_\epsilon} + \nabla v^h)) = 0 \text{ in } K \quad \text{and} \quad \tilde{v_\epsilon} = 0 \text{ on } \partial K. \tag{4.18}$$

Using (4.18), applying Green's Theorem, and coercivity property **N3**, we have the following estimate:

$$\begin{aligned}
\langle A^h_\epsilon\, v^h, v^h \rangle &= \sum_K \int_K (a(x/\epsilon, \eta^h, \nabla \tilde{v_\epsilon} + \nabla v^h), \nabla \tilde{v_\epsilon} + \nabla v^h)\, dx \\
&\geq c \sum_K \int_K |\nabla \tilde{v_\epsilon} + \nabla v^h|^p\, dx \\
&= c \sum_K \int_K |\nabla v_\epsilon|^p\, dx.
\end{aligned}$$

For $p = 2$, we may use Green's Theorem, the fact that ∇v^h is constant in K and $\tilde{v_\epsilon} = 0$ on ∂K to obtain the following:

$$\begin{aligned}
\langle A^h_\epsilon\, v^h, v^h \rangle &\geq c \sum_K \int_K |\nabla v^h|^2\, dx + 2 \int_K (\nabla v^h, \nabla \tilde{v_\epsilon})\, dx + \left(\int_K |\nabla \tilde{v_\epsilon}|^2\, dx \right) \\
&\geq c \sum_K \left(\int_K |\nabla v^h|^2\, dx - 2 \int_K \nabla(\nabla v^h)\, \tilde{v_\epsilon}\, dx + 2 \int_{\partial K} (\nabla v^h, \vec{n})\, \tilde{v_\epsilon}\, dx \right) \\
&= c \|\nabla v^h\|^2_{L_2(\Omega)}.
\end{aligned}$$

For $p > 2$ we note that since v^h is piecewise linear on ∂K we may write $v_\epsilon|_{\partial K} = v^h = \beta + (\nabla v^h, x)$, for some constant β. We set $\tilde{v_\epsilon} = v_\epsilon - \beta$. Then by change of variable and homogeneity argument, and applying Trace Theorem we have

$$\begin{aligned}
\langle A^h_\epsilon\, v^h, v^h \rangle &\geq c \sum_K \int_K |\nabla v_\epsilon|^p\, dx \\
&\geq c \sum_K \frac{h^n}{h^p} \int_{K_r} |\nabla_y \tilde{v_\epsilon}|^p\, dl_y \\
&\geq c \sum_K \frac{h^n}{h^p} \int_{\partial K_r} |(\nabla v^h, y\, h)|^p\, dl_y \\
&= c \sum_K h^n\, |\nabla v^h|^p\, C(e_{\nabla v^h}),
\end{aligned}$$

where K_r is a reference triangle, $e_{\nabla v^h}$ is the unit vector in the direction of ∇v^h, and

$$C(e_{\nabla v^h}) = \int_{\partial K_r} |(e_{\nabla v^h}, y)|^p \, dl_y.$$

To complete the proof, we need only to establish that $C(e_\xi)$ is bounded from below independent of ξ and h. By contradiction suppose the claim is not true. Then there exists a sequence $\{e_{\xi_n}\}$ which has a subsequence (denoted by the same notation) such that $e_{\xi_n} \to e_*$ and $C(e_{\xi_n}) \to 0$ as $n \to \infty$. Since $C(e_\xi)$ is continuous it follows that $C(e_*) = 0$. This further implies that $(e_*, y) = 0$ on ∂K_r, and hence $e_* = 0$. This is a contradiction. □

The following two estimates will be used in the next subsection.

Lemma 4.2. *Let $v_\epsilon - v_0 \in W_0^{1,p}(K)$ satisfies the following problem:*

$$-\nabla \cdot (a(x/\epsilon, \eta, \nabla v_\epsilon)) = 0 \text{ in } K$$

where η is constant in K. Then

$$\|\nabla v_\epsilon\|_{L_p(K)} \leq c\,(|K|^{\frac{1}{p}} + \|\eta\|_{L_p(K)} + \|\nabla v_0\|_{L_p(K)}).$$

Proof. Let $\tilde{v_\epsilon} = v_\epsilon - v_0$. It follows that $\tilde{v_\epsilon}$ satisfies the following problem:

$$-\nabla \cdot (a(x/\epsilon, \eta, \nabla(\tilde{v_\epsilon} + v_0))) = 0 \text{ in } K \quad \text{and} \quad \tilde{v_\epsilon} = 0 \text{ on } \partial K. \tag{4.19}$$

Multiplying (4.19) with v_ϵ, applying Green's Theorem, and using the fact that $\tilde{v_\epsilon} = 0$ on ∂K, we immediately obtain the following equality:

$$\int_K (a(x/\epsilon, \eta, \nabla v_\epsilon), \nabla v_\epsilon)\, dx = \int_K (a(x/\epsilon, \eta, \nabla v_\epsilon), \nabla v_0)\, dx. \tag{4.20}$$

Next we use coercivity **N3** and polynomial growth **N1** properties to bound (4.20) from below and above, respectively. Thus by applying Holder's and Young's inequalities

we have

$$\begin{aligned} c_2\|\nabla v_\epsilon\|^p_{L_p(K)} &\le c_1 \int_K (1+|\eta|^{p-1}+|\nabla v_\epsilon|^{p-1})\,|\nabla v_0|\,dx \\ &\le c_1 \left(\int_K (1+|\eta|^p+|\nabla v_\epsilon|^p)\,dx\right)^{\frac{1}{q}} \|\nabla v_0\|_{L_p(K)} \\ &\le \frac{c_1\,\delta}{q}\int_K (1+|\eta|^p+|\nabla v_\epsilon|^p)\,dx + \frac{c_1}{p\,\delta}\|\nabla v_0\|^p_{L_p(K)}. \end{aligned}$$

The claim in this lemma is obtained from this inequality by choosing $\delta > 0$ appropriately. □

Lemma 4.3. *Let $v_\epsilon - v_0 \in W^{1,p}_0(K)$ and $w_\epsilon - w_0 \in W^{1,p}_0(K)$ satisfy the following problems, respectively:*

$$-\nabla\cdot(a(x/\epsilon,\eta,\nabla v_\epsilon)) = 0 \text{ in } K,$$

$$-\nabla\cdot(a(x/\epsilon,\eta,\nabla w_\epsilon)) = 0 \text{ in } K,$$

where η is constant in K. Then the following estimate holds:

$$\|\nabla(v_\epsilon - w_\epsilon)\|^p_{L_p(K)} \le C\,H_0\,\|\nabla(v_0-w_0)\|^{\frac{p}{p-s}}_{L_p(K)},$$

where

$$H_0 = \left(|K| + \|\eta\|^p_{L_p(K)} + \|\nabla v_0\|^p_{L_p(K)} + \|\nabla w_0\|^p_{L_p(K)}\right)^{\frac{p-s-1}{p-s}}.$$

Proof. Let $\tilde{v_\epsilon} = v_\epsilon - v_0$ and $\tilde{w_\epsilon} = w_\epsilon - w_0$. It follows that $\tilde{v_\epsilon}$ and $\tilde{w_\epsilon}$ satisfy the following problems respectively:

$$-\nabla\cdot(a(x/\epsilon,\eta,\nabla(\tilde{v_\epsilon}+v_0))) = 0 \text{ in } K \quad \text{and} \quad \tilde{v_\epsilon} = 0 \text{ on } \partial K,$$

$$-\nabla\cdot(a(x/\epsilon,\eta,\nabla(\tilde{w_\epsilon}+w_0))) = 0 \text{ in } K \quad \text{and} \quad \tilde{w_\epsilon} = 0 \text{ on } \partial K.$$

Using monotonicity property **N2** and applying Green's Theorem along with the fact

that $\tilde{v_\epsilon} = \tilde{w}_\epsilon = 0$ on ∂K, we immediately obtain the following inequality:

$$\begin{aligned}
& c_1 \left\| \nabla(v_\epsilon - w_\epsilon) \right\|^p_{L_p(K)} \\
& = c_1 \left\| \nabla(\tilde{v_\epsilon} + v_0) - \nabla(\tilde{w}_\epsilon - w_0) \right\|^p_{L_p(K)} \\
& \leq \int_K \left(a(x/\epsilon, \eta, \nabla v_\epsilon) - a(x/\epsilon, \eta, \nabla w_\epsilon), \nabla(v_0 - w_0) \right) dx \\
& \leq \; c_4 \int_K H(\eta, \nabla v_\epsilon, \eta, \nabla w_\epsilon, p-1-s) \left| \nabla(v_\epsilon - w_\epsilon) \right|^s \left| \nabla(v_0 - w_0) \right| dx,
\end{aligned}$$

where on the last line we have used continuity property **N4**. Applying Holder's and Young's inequalities appropriately we have

$$\begin{aligned}
& \left\| \nabla(v_\epsilon - w_\epsilon) \right\|^p_{L_p(K)} \\
& \leq \; c \left(\int_K H(\eta, \nabla v_\epsilon, \eta, \nabla w_\epsilon, p) \, dx \right)^{\frac{p-s-1}{p}} \left\| \nabla(v_0 - w_0) \right\|_{L_p(K)} \left\| \nabla(v_\epsilon - w_\epsilon) \right\|^s_{L_p(K)} \\
& \leq c \frac{\delta \, s}{p} \left\| \nabla(v_\epsilon - w_\epsilon) \right\|^p_{L_p(K)} + c \frac{p-s}{\delta \, p} \left(\int_K H(\eta, \nabla v_\epsilon, \eta, \nabla w_\epsilon, p) \, dx \right)^{\frac{p-s-1}{p-s}} \\
& \quad \left\| \nabla(v_0 - w_0) \right\|^{\frac{p}{p-s}}_{L_p(K)}.
\end{aligned}$$

Applying Lemma 4.2 and choosing $\delta > 0$ appropriately, we obtain the desired estimate. □

Regarding η^h, we note that Jensen's inequality implies

$$\left\| \eta^h \right\|_{L_p(\Omega)} \leq c \left\| v^h \right\|_{L_p(\Omega)}. \tag{4.21}$$

In addition, the following estimates hold for η^h:

$$\left\| v^h - \eta^h \right\|_{L_p(\Omega)} \leq c \, h \left\| \nabla v^h \right\|_{L_p(\Omega)}. \tag{4.22}$$

4.3.2. A Closer Look at the Corrector for Monotone Operators

In this subsection we investigate a class of corrector for the monotone operators that appear in (4.3). The results that follow differ from the previous ones obtained by Dal

Maso et al. [15] in two respects. First, the correctors in their paper are for the fixed domain whose size is independent of ϵ. Second, the technique introduced in [15] can not provide rate of the convergence. The approach we introduce allows us to derive the corrector result in a manner similar to that of the linear case and to obtain the convergence rate.

Let $\eta \in \mathbb{R}$ and $\xi \in \mathbb{R}^2$ be given. We denote a function P defined by

$$P_{\eta,\xi}(y) = \xi + \nabla_y N_{\eta,\xi}(y), \tag{4.23}$$

where $N_{\eta,\xi}(y)$ satisfies (4.13). Obviously,

$$\int_Y (a(y, \eta, P_{\eta,\xi}(y)), P_{\eta,\xi}(y))\, dy = \int_Y (a(y, \eta, P_{\eta,\xi}(y)), \xi)\, dy. \tag{4.24}$$

Before constructing the corrector for the monotone operator and discussing its approximation property, we show the boundedness of $P_{\eta,\xi}$.

Lemma 4.4. *For every $\eta \in \mathbb{R}$ and $\xi \in \mathbb{R}^2$ we have*

$$\|P_{\eta,\xi}\|^p_{L_p(Y_\epsilon)} \leq c\,(1 + |\eta|^p + |\xi|^p)\,|Y_\epsilon|.$$

Proof. By change of variables, it is sufficient to show that

$$\|P_{\eta,\xi}\|^p_{L_p(Y)} \leq c\,(1 + |\eta|^p + |\xi|^p),$$

where Y is the unit square. Applying monotonicity **N2** and polynomial growth **N1**

properties we have

$$\begin{aligned}\|P_{\eta,\xi}\|^p_{L_p(Y)} &= \int_Y |P_{\eta,\xi} - 0|^p \, dy \\ &\leq c \int_Y (a(y,\eta,P_{\eta,\xi}) - a(y,\eta,0), P_{\eta,\xi}) \, dy \\ &= c \int_Y (a(y,\eta,P_{\eta,\xi}),\xi) \, dy - c \int_Y (a(y,\eta,0), P_{\eta,\xi}) \, , dy \\ &\leq c \int_Y (1 + |\eta|^{p-1} + |P_{\eta,\xi}|^{p-1}) \, |\xi| \, dy + c \int_Y (1 + |\eta|^{p-1}) \, |P_{\eta,\xi}| \, dy.\end{aligned}$$

Next we use Holder's inequality with $r_1 = p/(p-1)$ and $r_2 = p$ on both terms and afterward apply Young's inequality, so that for some $\beta > 0$ we have

$$\|P_{\eta,\xi}\|^p_{L_p(Y)} \leq c_1(\beta) \, (1 + |\eta|^p + |\xi|^p) + c_2(\beta) \, \|P_{\eta,\xi}\|^p_{L_p(Y)}.$$

Here $c_2(\beta) \to 0$ as $\beta \to 0$. Choosing β appropriately, we obtain the desired estimate. □

An easy consequence of this lemma is the following estimate on $N_{\eta,\xi}$ which has been defined in (4.13).

Corollary 4.1. *For every $\eta \in \mathbb{R}$ and $\xi \in \mathbb{R}^2$ we have*

$$\|\nabla_y N_{\eta,\xi}\|^p_{L_p(Y_\epsilon)} \leq c \, (1 + |\eta|^p + |\xi|^p) \, |Y_\epsilon|.$$

Proof. By the triangle inequality,

$$\|\nabla_y N_{\eta,\xi}\|_{L_p(Y_\epsilon)} \leq \|\xi + \nabla_y N_{\eta,\xi}\|_{L_p(Y_\epsilon)} + \|\xi\|_{L_p(Y_\epsilon)},$$

from which the estimate follows immediately by applying Lemma 4.4 to the first term. □

At this stage we denote by $\mathcal{P}$ a corrector associated with v_ϵ in the local problem

(4.3):

$$\mathcal{P}(x, x/\epsilon) = P_{\eta^h, \nabla v^h}(x/\epsilon) = \nabla v^h(x) + \nabla_y N_{\eta^h, \nabla v^h}(x/\epsilon), \tag{4.25}$$

where as in the local problem (4.3) v^h belongs to X^h. Moreover, given the values $\eta = \eta^h$ and $\xi = \nabla v^h$, the function $N_{\eta,\xi}(y)$ is the solution of the periodic problem (4.13). We note that using this setting, in general, each element K will have different $N_{\eta,\xi}$ depending on the value of $\eta = \eta^h$ and $\xi = \nabla v^h$ in the corresponding element K. The next lemma states an approximation property of the corrector $\mathcal{P}$.

Lemma 4.5. *Let v_ϵ satisfies* (4.3) *and assume that ∇v^h is uniformly bounded in $L_p(\Omega)$. Then*

$$\|\nabla v_\epsilon - \mathcal{P}\|_{L_p(\Omega)} \leq C \left(\frac{\epsilon}{h}\right)^{\frac{1}{p(p-s)}}.$$

Proof. Recall that by definition $\mathcal{P} = \nabla v^h + \nabla_y N_{\eta^h, \nabla v^h}(x/\epsilon) = \nabla v^h + \epsilon \nabla N_{\eta^h, \nabla v^h}(x/\epsilon)$, where $N_{\eta^h, \nabla v^h}$ is a zero-mean periodic function satisfying the following:

$$-\nabla \cdot (a(y, \eta^h, \nabla v^h + \nabla_y N_{\eta^h, \nabla v^h})) = 0. \tag{4.26}$$

We may expand v_ϵ as

$$v_\epsilon = v_\epsilon(x, x/\epsilon) = v^h(x) + \epsilon N_{\eta^h, \nabla v^h}(x/\epsilon) + \theta(x, x/\epsilon).$$

Next we denote by $w_\epsilon = w_\epsilon(x, x/\epsilon) = v^h(x) + \epsilon N_{\eta^h, \nabla v^h}(x/\epsilon)$. Obviously w_ϵ satisfies (4.26). Taking all these into account, the claim in the lemma is the same as to proving

$$\|\nabla \theta\|_{L_p(\Omega)} = \|\nabla(v_\epsilon - w_\epsilon)\|_{L_p(\Omega)} \leq C \left(\frac{\epsilon}{h}\right)^{\frac{1}{p(p-s)}}.$$

Here we may write w_ϵ as a solution of the following boundary value problem:

$$-\nabla \cdot (a(x/\epsilon, \eta^h, \nabla w_\epsilon)) = 0 \text{ in } K \quad \text{and} \quad w_\epsilon = v^h + \epsilon \tilde{N}_{\eta^h, \nabla v^h} \text{ on } \partial K,$$

with $\tilde{N}_{\eta^h, \nabla v^h} = N_{\eta^h, \nabla v^h} \varphi$, where φ is a sufficiently smooth function whose value is 1

on a strip of width ϵ adjacent to ∂K, and 0 elsewhere. We denote this strip by S_ϵ. This idea has been used in [45]. By Lemma 4.3 we have the following estimate:

$$\begin{aligned}\|\nabla\theta\|^p_{L_p(K)} &= \|\nabla(v_\epsilon - w_\epsilon)\|^p_{L_p(K)} \\ &\leq C\, H_0\, \|\nabla(v^h - v^h - \epsilon\, \tilde{N}_{\eta^h,\nabla v^h})\|^{\frac{p}{p-s}}_{L_p(K)} \\ &\leq C\, H_0\, \|\epsilon\, \nabla \tilde{N}_{\eta^h,\nabla v^h}\|^{\frac{p}{p-s}}_{L_p(K)},\end{aligned} \tag{4.27}$$

where

$$H_0 = \Big(|K| + \|\eta^h\|^p_{L_p(K)} + \|\nabla v^h\|^p_{L_p(K)} + \|\nabla(v^h + \epsilon\, \tilde{N}_{\eta^h,\nabla v^h})\|^p_{L_p(K)}\Big)^{\frac{p-s-1}{p-s}}. \tag{4.28}$$

We need to show that H_0 is bounded and $\|\epsilon\, \nabla \tilde{N}_{\eta^h,\nabla v^h}\|^p_{L_p(\Omega)}$ uniformly vanishes as $\epsilon \to 0$. For this purpose, we use the following notations: let $J^K_\epsilon = \{i \in \mathbb{Z}^n : Y^i_\epsilon \bigcap K \neq 0, K \backslash Y^i_\epsilon \neq 0\}$ and $F^K_\epsilon = \cup_{i\in J^K_\epsilon} Y^i_\epsilon$. In other words, F^K_ϵ is the union of all periods Y^i_ϵ that covers the strip S_ϵ. Using these notations and since φ is zero everywhere in K, except in the strip S_ϵ, we may write the following:

$$\begin{aligned}\|\epsilon\, \nabla \tilde{N}_{\eta^h,\nabla v^h}\|^p_{L_p(K)} &= \epsilon^p \int_K |\nabla(N_{\eta^h,\nabla v^h}\, \varphi)|^p\, dx \\ &= \epsilon^p \int_{S_\epsilon} |\nabla(N_{\eta^h,\nabla v^h}\, \varphi)|^p\, dx \\ &\leq \epsilon^p \int_{F^K_\epsilon} |\nabla(N_{\eta^h,\nabla v^h}\, \varphi)|^p\, dx \\ &= \epsilon^p \sum_{i\in J^K_\epsilon} \int_{Y^i_\epsilon} |\nabla(N_{\eta^h,\nabla v^h}\, \varphi)|^p\, dx \\ &\leq \epsilon^p \sum_{i\in J^K_\epsilon} \int_{Y^i_\epsilon} \left(|\nabla\, N_{\eta^h,\nabla v^h}|^p\, |\varphi|^p + |N_{\eta^h,\nabla v^h}|^p\, |\nabla\varphi|^p\right)\, dx,\end{aligned} \tag{4.29}$$

where we have used the product rule on the partial derivative in the last line of (4.29). Our aim now is to show that the sum of integrals in the last line of (4.29) is uniformly bounded. We note that (see [55] and also Corollary 4.1)

$$\|\nabla_y\, N_{\eta^h,\nabla v^h}\|^p_{L_p(Y^i_\epsilon)} \leq C(1 + |\eta^h|^p + |\nabla v^h|^p)\, |Y^i_\epsilon|,$$

from which, using Poincaré-Friedrich inequality we have

$$\|N_{\eta^h,\nabla v^h}\|^p_{L_p(Y^i_\epsilon)} \leq C(1+|\eta^h|^p+|\nabla v^h|^p)\,|Y^i_\epsilon|.$$

We note also that η^h and ∇v^h are constant in K. Since φ is sufficiently smooth whose value is one on the strip S_ϵ and zero elsewhere, we know that $|\nabla\varphi| \leq C/\epsilon$ (cf. [45]). Applying all these facts to (4.29) we have

$$\begin{aligned}\|\epsilon\,\nabla\tilde{N}_{\eta^h,\nabla v^h}\|^p_{L_p(K)} &\leq C\,\epsilon^p\,(1+|\eta^h|^p+|\nabla v^h|^p)\sum_{i\in J^K_\epsilon}(1+\epsilon^{-p})\,|Y^i_\epsilon| \\ &= C\,(\epsilon^p+1)\,(1+|\eta^h|^p+|\nabla v^h|^p)\sum_{i\in J^K_\epsilon}|Y^i_\epsilon| \\ &\leq C\,(1+|\eta^h|^p+|\nabla v^h|^p)\sum_{i\in J^K_\epsilon}|Y^i_\epsilon|.\end{aligned}$$

Moreover, since all Y^i_ϵ, $i\in J^K_\epsilon$, cover the strip S_ϵ, we know that $\sum_{i\in J^K_\epsilon}|Y^i_\epsilon| \leq C\,h\,\epsilon$. Hence using local inverse inequality for η^h and ∇v^h, we have

$$\begin{aligned}\|\epsilon\,\nabla\tilde{N}_{\eta^h,\nabla v^h}\|^p_{L_p(K)} &\leq C\,\frac{h^2}{h^2}\,\left(1+|\eta^h|^p+|\nabla v^h|^p\right)\,h\,\epsilon \\ &\leq C\,\frac{\epsilon}{h}\,\left(|K|+\|\eta^h\|^p_{L_p(K)}+\|\nabla v^h\|^p_{L_p(K)}\right)\end{aligned} \tag{4.30}$$

Furthermore, using this estimate and noting that $\epsilon/h<1$ we obtain from (4.28) that

$$H_0 \leq C\,\left(|K|+\|\eta^h\|^p_{L_p(K)}+\|v^h\|^p_{L_p(K)}+\|\nabla v^h\|^p_{L_p(K)}\right)^{\frac{p-s-1}{p-s}}. \tag{4.31}$$

Summarizing the results, from (4.27) which is combined with (4.31) and (4.30) we get

$$\begin{aligned}\|\nabla\theta\|^p_{L_p(K)} &\leq C\,H_0\,\|\epsilon\,\nabla\tilde{N}_{\eta^h,\nabla v^h}\|^{\frac{p}{p-s}}_{L_p(K)} \\ &\leq C\,\left(\frac{\epsilon}{h}\right)^{\frac{1}{p-s}}\,\left(|K|+\|\eta^h\|^p_{L_p(K)}+\|v^h\|^p_{L_p(K)}+\|\nabla v^h\|^p_{L_p(K)}\right)\end{aligned}$$

Finally summing over all $K \in T_h$ and applying (4.21) to $\sum_{K\in T_h} \|\eta^h\|^p_{L_p(K)}$, we obtain

$$\begin{aligned}
\|\nabla\theta\|^p_{L_p(\Omega)} &= \sum_{K\in T_h} \|\nabla\theta\|^p_{L_p(K)} \\
&\leq C\left(\frac{\epsilon}{h}\right)^{\frac{1}{p-s}} \sum_{K\in T_h} \left(|K| + \|v^h\|^p_{L_p(K)} + \|\nabla v^h\|^p_{L_p(K)}\right) \\
&= C\left(\frac{\epsilon}{h}\right)^{\frac{1}{p-s}} \left(|\Omega| + \|v^h\|^p_{L_p(\Omega)} + \|\nabla v^h\|^p_{L_p(\Omega)}\right).
\end{aligned}$$

Obviously, this last inequality uniformly vanishes as ϵ approaching zero, and thus we have completed the proof of the Lemma 4.5. □

The approximation property that we have just proved is used to show the following lemma:

Lemma 4.6. *Suppose $v^h, w^h \in X^h$ with ∇v^h and ∇w^h uniformly bounded in $L_p(\Omega)$ and $L_{p+\alpha}(\Omega)$, respectively, for some $\alpha > 0$. Then $\lim_{\epsilon\to 0} \langle A^h_\epsilon\, v^h - A^*\, v^h, w^h\rangle = 0$.*

Proof. Given $v^h \in X^h$, we set the corrector $\mathcal{P}$ as in (4.25). By adding and subtracting terms we have the following equality:

$$\langle A^h_\epsilon\, v^h - A^*\, v^h, w^h\rangle = \sum_K (I_K + II_K + III_K),$$

where

$$\begin{aligned}
I_K &= \int_K (a(x/\epsilon, \eta^h, \nabla v_\epsilon) - a(x/\epsilon, \eta^h, \mathcal{P}), \nabla w^h)\, dx, \\
II_K &= \int_K (a(x/\epsilon, \eta^h, \mathcal{P}) - a^*(\eta^h, \nabla v^h), \nabla w^h)\, dx, \\
III_K &= \int_K (a^*(\eta^h, \nabla v^h) - a^*(v^h, \nabla v^h), \nabla w^h)\, dx.
\end{aligned}$$

Step 1: estimate of I_K

Using continuity property **N4** and Holder's inequality, I_K is estimated in the following

way:

$$I_K \le c \int_K |\nabla v_\epsilon - \mathcal{P}|^s \, H(\eta^h, \nabla v_\epsilon, \eta^h, \mathcal{P}, p-1-s) \, |\nabla w^h| \, dx$$
$$\le c \, \|\nabla v_\epsilon - \mathcal{P}\|^s_{L_p(K)} \left(\int_K H(\eta^h, \nabla v_\epsilon, \eta^h, \mathcal{P}, p) \, dx \right)^{\frac{p-1-s}{p}} \|\nabla w^h\|_{L_p(K)}.$$

It follows that

$$\sum_K I_K \le c \, \|\nabla v_\epsilon - \mathcal{P}\|^s_{L_p(\Omega)} \left(\int_\Omega H(\eta^h, \nabla v_\epsilon, \eta^h, \mathcal{P}, p) \, dx \right)^{\frac{p-1-s}{p}} \|\nabla w^h\|_{L_p(\Omega)}.$$

By Lemma 4.5 the last inequality vanishes as ϵ approaching zero.

Step 2: estimate of II_K

Let $I^K_\epsilon = \{i \in \mathbb{Z}^n : Y^i_\epsilon \subset K\}$ and $J^K_\epsilon = \{i \in \mathbb{Z}^n : Y^i_\epsilon \bigcap K \neq 0, K \backslash Y^i_\epsilon \neq 0\}$. Let $E^K_\epsilon = \cup_{i \in I^K_\epsilon} Y^i_\epsilon$ and $F^K_\epsilon = \cup_{i \in J^K_\epsilon} Y^i_\epsilon$. Then we may break up the integration II_K into the sum of integral over E^K_ϵ and $K \backslash E^K_\epsilon$. By (4.12) and the fact that ∇w^h is constant in K, we have the following estimate:

$$II_K = \sum_{i \in I^K_\epsilon} \int_{Y^i_\epsilon} (a(x/\epsilon, \eta^h, \mathcal{P}) - a^*(\eta^h, \nabla v^h), \nabla w^h) \, dx$$
$$+ \int_{K \backslash E^K_\epsilon} (a(x/\epsilon, \eta^h, \mathcal{P}) - a^*(\eta^h, \nabla v^h), \nabla w^h) \, dx$$
$$\le \int_{F^K_\epsilon} |(a(x/\epsilon, \eta^h, \mathcal{P}) - a^*(\eta^h, \nabla v^h), \nabla w^h)| \, dx.$$

It follows by applying Holder's inequality appropriately and using Lemma 4.4 that

$$\begin{aligned}
\sum_K II_K &\le \sum_K \sum_{i\in J_\epsilon^K} \int_{Y_\epsilon^i} \left(|(a(x/\epsilon,\eta^h,\mathcal{P}),\nabla w^h)| + |(a^*(\eta^h,\nabla v^h),\nabla w^h)|\right)\, dx \\
&\le c \sum_K \sum_{i\in J_\epsilon^K} \int_{Y_\epsilon^i} H(\eta^h,\mathcal{P},\eta^h,\nabla v^h,p-1)\,|\nabla w^h|\,dx \\
&\le c \left(\sum_K \sum_{i\in J_\epsilon^K} \int_{Y_\epsilon^i} H(\eta^h,\mathcal{P},\eta^h,\nabla v^h,p)\,dx\right)^{\frac{1}{q}} \|\nabla w^h\|_{L_p(\Omega)} \\
&\le c \left(\sum_K \sum_{i\in J_\epsilon^K} \left(|\eta^h|^p + |\nabla v^h|^p\right)\,|Y_\epsilon^i|\right)^{\frac{1}{q}} \|\nabla w^h\|_{L_p(\Omega)} \\
&\le c \left(\sum_K |K|\,\left(|\eta^h|^p + |\nabla v^h|^p\right)\,\frac{|F_\epsilon^K|}{|K|}\right)^{\frac{1}{q}} \|\nabla w^h\|_{L_p(\Omega)} \\
&\le c \max_K \left(\frac{|F_\epsilon^K|}{|K|}\right)^{\frac{1}{q}} \left(\|v^h\|^p_{L_p(\Omega)} + \|\nabla v^h\|^p_{L_p(\Omega)}\right)^{\frac{1}{q}} \|\nabla w^h\|_{L_p(\Omega)} \\
&\le c \left(\frac{\epsilon}{h}\right)^{\frac{1}{q}} \left(\|v^h\|^p_{L_p(\Omega)} + \|\nabla v^h\|^p_{L_p(\Omega)}\right)^{\frac{1}{q}} \|\nabla w^h\|_{L_p(\Omega)}
\end{aligned}$$

Obviously, this expression vanishes as ϵ approaching zero.

Step 3: estimate of III_K

Using (4.17) and Holder's inequality we estimate III_K in the following way:

$$\begin{aligned}
III_K &\le c \int_K H(\eta^h,\nabla v^h,v^h,\nabla v^h,p-1)\,\nu(|\eta^h - v^h|)\,|\nabla w^h|\,dx \\
&\le c \left(\int_K H(\eta^h,\nabla v^h,v^h,\nabla v^h,p)\,\nu(|\eta^h - v^h|)^q\,dx\right)^{\frac{1}{q}} \|\nabla w^h\|_{L_p(K)}.
\end{aligned}$$

It follows that

$$\sum_K III_K \le c \left(\int_\Omega H(\eta^h,\nabla v^h,v^h,\nabla v^h,p)\,\nu(|\eta^h - v^h|)^q\,dx\right)^{\frac{1}{q}} \|\nabla w^h\|_{L_p(\Omega)}.$$

If $\nabla w^h \in L_{p+\alpha}(\Omega)$, then we may use Lemma 4.3 to conclude that $\sum_K III_K$ vanishes as $\epsilon \to 0$. The fact that $\nabla w^h \in L_{p+\alpha}(\Omega)$ for some $\alpha > 0$ has been shown in [24]. $\square$

4.3.3. Proof of Theorem 4.1

Since A_ϵ^h is coercive, it follows that u^h is bounded, which implies that it has a subsequence (which we also denote by u^h) such that $u^h \rightharpoonup u$ in $W^{1,p}(\Omega)$ as $\epsilon \to 0$, for some $u \in W^{1,p}(\Omega)$. We note that u^h depends on ϵ which makes the convergence makes sense. Since the homogenized operator A^* is of type S_+ [55], then by its definition, the strong convergence would be true if we can show that $\limsup_{\epsilon\to 0} \langle A^* u^h, u^h - u\rangle \to 0$. Moreover, by adding and subtracting term, we have the following equality:

$$\begin{aligned}\langle A^* u^h, u^h - u\rangle &= \langle A^* u^h - A_\epsilon^h u^h, u^h - u\rangle + \langle A_\epsilon^h u^h, u^h - u\rangle \\ &= \langle A^* u^h - A_\epsilon^h u^h, u^h\rangle - \langle A^* u^h - A_\epsilon^h u^h, u\rangle + (f, u^h - u).\end{aligned}$$

Lemma 4.6 implies that the first and second term vanish as $\epsilon \to 0$ provided ∇u_h is uniformly bounded in $L_{p+\alpha}$ for $\alpha > 0$, while the last term vanishes as $\epsilon \to 0$ by the weak convergence of u^h. One can assume additional not restrictive regularity assumptions [49] for input data and obtain Meyers type estimates, $\|\nabla u\|_{L_{p+\alpha}(\Omega)} \le C$, for the homogenized solutions. In this case it is reasonable also to assume that the discrete solutions are uniformly bounded in $L_{p+\alpha}(\Omega)$. Similar Meyers type estimates for the approximate solutions in the case of $p = 2$ have been obtained in [24]. Finally since A^* is also of type M, all these conditions imply that $A^* u = f$, hence we have the claim of the theorem.

4.3.4. Proof of Theorem 4.2

We define an operator approximating the identity map in $L_p(\Omega)$ by

$$M_\epsilon \varphi(x) = \sum_{i \in I_\epsilon} \Psi_{Y_\epsilon^i}(x) \, \frac{1}{|Y_\epsilon^i|} \int_{Y_\epsilon^i} \varphi(y)\, dy, \tag{4.32}$$

where Y_ϵ^i, a square of size ϵ^2 for $i \in \mathbb{Z}^n$, and $I_\epsilon = \{i \in \mathbb{Z}^n : Y_\epsilon^i \subset \Omega\}$. Next we denote $P = P_{M_\epsilon u, M_\epsilon \nabla u}(x, x/\epsilon) = M_\epsilon \nabla u(x) + \nabla_y N_{M_\epsilon u, M_\epsilon \nabla u}(x/\epsilon)$., where u is the solution of

homogenized problem (4.9). The function P is a corrector associated with the original boundary value problem (4.1). Now by triangle inequality we have

$$\|\nabla v_\epsilon - \nabla u_\epsilon\|_{L_p(\Omega)} \leq \|\nabla v_\epsilon - \mathcal{P}\|_{L_p(\Omega)} + \|\mathcal{P} - P\|_{L_p(\Omega)} + \|P - \nabla u_\epsilon\|_{L_p(\Omega)},$$

where $\mathcal{P} = \nabla u^h + \nabla_y N_{\eta^h, \nabla u^h}$ as defined in (4.25). Lemma 4.5 gives the convergence of the first term. For the second and third terms, we need to establish approximation properties of the corrector P. This will be described in detail in section 4.4.

4.3.5. Proof of Theorem 4.3

For the following proof, we note that as the assumption in Theorem 4.3 says, the nonlinearity of the coefficient depends only on the gradient of the solution, i.e., $a(x/\epsilon, \eta, \xi) \equiv a(x/\epsilon, \xi)$, for $\xi \in \mathbb{R}^2$. The same is true for the homogenized coefficient, i.e., $a^*(\eta, \xi) \equiv a^*(\xi)$.

Let $P_h u \in X^h$ denotes the finite element solution of the homogenized problem (4.9). By triangle inequality we have

$$\|\nabla u^h - \nabla u\|_{L_p(\Omega)} \leq \|\nabla u^h - \nabla P_h u\|_{L_p(\Omega)} + \|\nabla P_h u - \nabla u\|_{L_p(\Omega)}. \tag{4.33}$$

Regarding $P_u u$, we have an existing result that states [13]

$$\|\nabla P_h u - \nabla u\|_{L_p(\Omega)} \leq C\, h^{\frac{1}{p-1}}.$$

The rest of the proof is concentrated on the first part of (4.33). Using the homogenized operator (4.11), and applying the monotonicity property **NH2** of the homogenized coefficient, we have

$$\begin{aligned}\|\nabla u^h - \nabla P_h u\|^p_{L_p(\Omega)} &\leq C\, \langle A^* u^h - A^* P_h u, u^h - P_h u\rangle \\ &= C\, \langle A^* u^h - A^h_\epsilon u^h, u^h - P_h u\rangle + C\, \langle A^h_\epsilon u^h - A^* P_h u, u^h - P_h u\rangle,\end{aligned}$$

Using steps 1 and 2 in the proof of Lemma 4.6 the first term can be estimated as:

$$\begin{aligned}\langle A^* u^h - A^h_\epsilon u^h, u^h - P_h u\rangle \leq\ & C_1 \left(\frac{\epsilon}{h}\right)^{\frac{s}{p-s}} \|\nabla u^h - \nabla P_h u\|_{L_p(\Omega)} \\ & + C_2 \left(\frac{\epsilon}{h}\right)^{\frac{p-1}{p}} \|\nabla u^h - \nabla P_h u\|_{L_p(\Omega)} \\ & \leq C_1 \left(\frac{\epsilon}{h}\right)^{\frac{sp}{(p-1)(p-s)}} + C_2 \frac{\epsilon}{h} + \delta \|\nabla u^h - \nabla P_h u\|^p_{L_p(\Omega)},\end{aligned}$$

where we have used Young's inequality with some $\delta > 0$. Furthermore it is straightforward to see that using (4.2) and (4.10) that

$$\langle A^h_\epsilon\, u^h, w^h\rangle = \int_\Omega f\, w^h\, dx = \langle A^*\, u, w^h\rangle \quad \forall\ w^h \in X^h.$$

Then applying to continuity property **NH4** of the homogenized coefficient we have

$$\begin{aligned}\langle A^h_\epsilon u^h - A^* P_h u, u^h - P_h u\rangle &= \langle A^* u - A^* P_h u, u^h - P_h u\rangle \\ &\leq C_3 \|\nabla u^h - \nabla P_h u\|_{L_p(\Omega)} \|\nabla u - \nabla P_h u\|^{\frac{p}{s}}_{L_p(\Omega)} \\ &\leq C_3 \|\nabla u - \nabla P_h u\|^{\frac{p^2}{s(p-1)}}_{L_p(\Omega)} + \delta \|\nabla u^h - \nabla P_h u\|^p_{L_p(\Omega)}.\end{aligned}$$

Choosing δ appropriately, we have the desired result.

4.4. Estimate on Corrector $P_{M_\epsilon u, M_\epsilon \nabla u}$

In this section we present a convergence property of the corrector P described in the proof of Theorem 4.2. There we also have defined an operator M_ϵ in $L_p(\Omega)$. This operator enjoys the following properties (e.g. [15]):

$$\lim_{\epsilon \to 0} \|M_\epsilon \varphi - \varphi\|_{L_p(\Omega)} = 0, \tag{4.34}$$

$$\|M_\epsilon \varphi\|_{L_p(\Omega)} \leq C \|\varphi\|_{L_p(\Omega)}. \tag{4.35}$$

Now we state the main result of this section.

Theorem 4.4. *Let u be the solution of homogenized problem* (4.9)*, and M_ϵ be the*

operator defined by (4.32) *and denote by*

$$P = P_{M_\epsilon u, M_\epsilon \nabla u}(x, x/\epsilon) = M_\epsilon \nabla u(x) + \nabla_y N_{M_\epsilon u, M_\epsilon \nabla u}(x/\epsilon).$$

Then $\lim_{\epsilon \to 0} \|P - \nabla u_\epsilon\|_{L_p(\Omega)} = 0$.

We recall that in the previous section we have shown (cf. Lemma 4.4) that given $\eta \in \mathbb{R}$ and $\xi \in \mathbb{R}^2$, $\|P_{\eta,\xi}\|^p_{L_p(Y_\epsilon)} \leq c\,(1 + |\eta|^p + |\xi|^p)\,|Y_\epsilon|$. In fact, we may obtain the similar estimate for the $L_{p+\tau}$ for some $\tau > 0$. The result is stated in the following corollary (see [15] for detail proof).

Corollary 4.2. *There exists* $\tau > 0$ *independent of* ϵ *such that*

$$\|P_{\eta,\xi}\|^{p+\tau}_{L_{p+\tau}(Y_\epsilon)} \leq c\,(1 + |\eta|^p + |\xi|^p)\,|Y_\epsilon|.$$

Lemma 4.7. *For every* $\eta_1, \eta_2 \in \mathbb{R}$ *and* $\xi_1, \xi_2 \in \mathbb{R}^2$ *we have*

$$\|P_{\eta_1,\xi_1} - P_{\eta_2,\xi_2}\|^p_{L_p(Y)} \leq\ c\,(H\,\nu(|\eta_1 - \eta_2|) + H^{\frac{p-1-s}{p-s}}\,|\xi_1 - \xi_2|^{\frac{p}{p-s}} + |\xi_1 - \xi_2|^p),$$

where

$$H = H(\eta_1, \xi_1, \eta_2, \xi_2, p) = 1 + |\eta_1|^p + |\eta_2|^p + |\xi_1|^p + |\xi_2|^p. \tag{4.36}$$

Proof. For simplicity of notation we denote $P_i = P(y, \eta, \xi_i)$, $i = 1, 2$. Using monotonicity property **N2** and by adding and subtracting terms we have

$$\begin{aligned} c_1 \|P_1 - P_2\|^p_{L_p(Y)} \leq & \int_Y (a(y, \eta_1, P_1) - a(y, \eta_2, P_2), P_1 - P_2)\,dy \\ & + \int_Y (a(y, \eta_2, P_2) - a(y, \eta_1, P_2), P_1 - P_2)\,dy \\ & = I_1 + I_2. \end{aligned}$$

Using (4.24) and continuity property **N4**, I_1 is estimated as follows:

$$\begin{aligned} I_1 &\leq c \int_Y \nu(|\eta_1 - \eta_2|)\, H(\eta_1, P_1, \eta_2, P_2, p-1)\, |\xi_1 - \xi_2|\, dy \\ &\quad + c \int_Y H(\eta_1, P_1, \eta_2, P_2, p-1-s)\, |P_1 - P_2|^s\, |\xi_1 - \xi_2|\, dy \\ &= I_{11} + I_{12}. \end{aligned}$$

Similarly, using continuity property **N4**,

$$I_2 \leq c \int_Y \nu(|\eta_1 - \eta_2|)\, H(\eta_1, P_1, \eta_2, P_2, p-1)\, |P_1 - P_2|\, dy.$$

Now we may use Holder's and Young's inequalities with $r_1 = p/(p-1) = q$ and $r_2 = p$ and Lemma 4.4 to get the following

$$I_{11} \leq c\, \nu(|\eta_1 - \eta_2|)^q\, H(\eta_1, \xi_1, \eta_2, \xi_2, p) + c\, |\xi_1 - \xi_2|^p.$$

Similarly, using Holder's and Young's inequalities with $r_1 = p/(p-1-s)$, $r_2 = p/s$ and a $\beta > 0$ such that we have

$$\begin{aligned} I_{12} &\leq c\, H(\eta_1, \xi_1, \eta_2, \xi_2, p)^{\frac{p-1-s}{p}}\, \|P_1 - P_2\|^s_{L_p(Y)}\, |\xi_1 - \xi_2| \\ &\leq c\, \beta^{\frac{p}{p-s}}\, H(\eta_1, \xi_1, \eta_2, \xi_2, p)^{\frac{p-1-s}{p-s}}\, |\xi_1 - \xi_2|^{\frac{p}{p-s}} + c\, \beta^{-\frac{p}{s}}\, \|P_1 - P_2\|^p_{L_p(Y)}. \end{aligned}$$

Using the same procedure as above me may estimate I_2 similarly, so that we have

$$I_2 \leq c\, \beta^q\, \nu(|\eta_1 - \eta_2|)^q\, H(\eta_1, \xi_1, \eta_2, \xi_2) + c\, \beta^{-p}\, \|P_1 - P_2\|^p_{L_p(Y)}.$$

Now choosing β appropriately, we have the desired estimate. □

For the next lemma, we need the following partition. Let $\Omega_j \subset \Omega$ be a partition of Ω such that $|\partial\Omega_j| = 0$, $\Omega_j \bigcap \Omega_k = 0$ for $j \neq k$. Furthermore, let χ and ψ be functions of the form

$$\chi(x) = \sum_{j=1}^{m} a_j\, \Psi_{\Omega_j}(x) \quad \text{and} \quad \psi(x) = \sum_{j=1}^{m} b_j\, \Psi_{\Omega_j}(x), \tag{4.37}$$

for some $a_j \in \mathbb{R}$ and $b_j \in \mathbb{R}^2$.

Lemma 4.8. *Let $\phi \in L_p(\Omega)$, $\varphi \in (L_p(\Omega))^2$, and let χ and ψ be as in (4.37). Then*

$$\begin{aligned}
&\limsup_{\epsilon\to 0} \|P(x/\epsilon, M_\epsilon\phi, M_\epsilon\varphi) - P(x/\epsilon, \chi, \psi)\|_{L_p(\Omega)} \\
&\le c\left(\int_Q \nu(|\phi-\chi|)^q\, H(\phi,\varphi,\chi,\psi,p)\,dx\right)^{\frac{1}{p}} \\
&+ c\left(|\Omega|^{\frac{1}{p}} + \|\phi\|_{L_p(\Omega)} + \|\varphi\|_{L_p(\Omega)} + \|\chi\|_{L_p(\Omega)} + \|\psi\|_{L_p(\Omega)}\right)^{\frac{p-1-s}{p-s}} \|\varphi-\psi\|_{L_p(\Omega)}^{\frac{1}{p-s}} \\
&+ c\,\|\varphi-\psi\|_{L_p(\Omega)}.
\end{aligned}$$

Proof. We use the following notations. Let $\Omega_0 = \Omega\setminus\bigcup_{j=1}^m \Omega_j$ with $a_0 = 0$, $b_0 = 0$, $\Omega_\epsilon = \bigcup_i \overline{Y_\epsilon^i}$ with $Y_\epsilon^i \subset \Omega$, $I_\epsilon^j = \{i \in I_\epsilon : Y_\epsilon^i \subset \Omega_j\}$, $J_\epsilon^j = \{i \in I_\epsilon : Y_\epsilon^i \bigcap \Omega_j \neq 0, \Omega_j\setminus Y_\epsilon^i \neq 0\}$, $E_\epsilon^j = \bigcup_{i\in I_\epsilon^j} \overline{Y_\epsilon^i}$, $F_\epsilon^j = \bigcup_{i\in J_\epsilon^j} \overline{Y_\epsilon^i}$, For sufficiently small ϵ we have $\Omega_j \subseteq \Omega_\epsilon$ for $j \neq 0$. Now by definition of M_ϵ, χ, and ψ we have

$$\|P(\cdot, M_\epsilon\phi, M_\epsilon\varphi) - P(\cdot,\chi,\psi)\|_{L_p(\Omega)}^p = \|P(\cdot, M_\epsilon\phi, M_\epsilon\varphi) - P(\cdot,\chi,\psi)\|_{L_p(\Omega_\epsilon)}^p \le \sum_{j=0}^m (e_j + f_j),$$

where

$$e_j = \|P(\cdot, M_\epsilon\phi, M_\epsilon\varphi) - P(\cdot, a_j, b_j)\|_{L_p(E_\epsilon^j)}^p$$

and

$$f_j = \|P(\cdot, M_\epsilon\phi, M_\epsilon\varphi) - P(\cdot, a_j, b_j)\|_{L_p(F_\epsilon^j)}^p.$$

Now we set $\eta_i = |Y_\epsilon^i|^{-1} \int_{Y_\epsilon^i} \phi(x)\,dx$ and $\xi_i = |Y_\epsilon^i|^{-1} \int_{Y_\epsilon^i} \varphi(x)\,dx$. By change of variable and applying Lemma 4.7 we get

$$\begin{aligned}
\sum_{j=0}^m e_j &= \sum_{j=0}^m \sum_{i\in I_\epsilon^j} \|P(\cdot,\eta_i,\xi_i) - P(\cdot,a_j,b_j)\|_{L_p(Y_\epsilon^i)}^p \\
&\le c\,(I_1 + I_2 + I_3),
\end{aligned}$$

where

$$I_1 = \sum_{j=0}^{m} \sum_{i \in I_\epsilon^j} H(\eta_i, \xi_i, a_j, b_j, p)\, \nu(|\eta_i - a_j|)^{\frac{p}{p-1}}\, |Y_\epsilon^i|$$
$$I_2 = \sum_{j=0}^{m} \sum_{i \in I_\epsilon^j} H(\eta_i, \xi_i, a_j, b_j, p)^{\frac{p-1-s}{p-s}}\, |\xi_i - b_j|^{\frac{p}{p-s}} |Y_\epsilon^i|$$
$$I_3 = \sum_{j=0}^{m} \sum_{i \in I_\epsilon^j} |\xi_i - b_j|^p |Y_\epsilon^i|$$

Now we may use Holder's and Jensen's inequalities appropriately to obtain the following:

$$I_1 \le c \sum_{j=0}^{m} \left(\int_{E_\epsilon^j} \nu(|M_\epsilon \phi - a_j|)^{\frac{p}{p-1}}\, H(\phi, \varphi, a_j, b_j, p)\, dx \right)$$
$$I_2 \le c \left(\sum_{j=0}^{m} \left(\|\phi\|^p_{L_p(E_\epsilon^j)} + \|\varphi\|^p_{L_p(E_\epsilon^j)} + (1 + |a_j|^p + |b_j|^p)\, |E_\epsilon^j| \right) \right)^{\frac{p-1-s}{p-s}} \|\varphi - \psi\|^{\frac{p}{p-s}}_{L_p(\Omega)}$$
$$I_3 \le c \sum_{j=0}^{m} \|\varphi - b_j\|^p_{L_p(E_\epsilon^j)}$$

Regarding I_1, we know that since $(M_\epsilon \phi - \chi) \to (\phi - \chi)$ in $L_p(\Omega)$ as $\epsilon \to 0$, and ϕ, φ, χ ψ are compact in $L_p(\Omega)$ it follows that $\int_\Omega \nu(|M_\epsilon \phi - \chi|)^{p/(p-1)}\, H\, dx$ converges to $\int_\Omega \nu(|\phi - \chi|)^{p/(p-1)}\, H\, dx$ where $H = H(\phi, \varphi, \chi, \psi, p)$. We note that the same estimate holds for $\sum_{j=0}^m f_j$, with J_ϵ^j replacing I_ϵ^j, and F_ϵ^j replacing E_ϵ^j. Furthermore, since $|\partial \Omega_j| = 0$ for $j \neq 0$, $|F_\epsilon^j|$ vanishes as $\epsilon \to 0$. This implies that all terms coming from $\sum_{j=0}^m f_j$ vanish as $\epsilon \to 0$, and hence we obtain the desired result. □

Proof of Theorem 4.4. To simplify notation we use $P_M = P(x/\epsilon, M_\epsilon u, M_\epsilon \nabla u)$. Using (4.5) and by adding and subtracting terms we write the following:

$$\begin{aligned} \int_\Omega |P_M - \nabla u_\epsilon|^p\, dx &\le c_1 \int_\Omega (a(x/\epsilon, u_\epsilon, P_M) - a(x/\epsilon, u_\epsilon, \nabla u_\epsilon), P_M - \nabla u_\epsilon)\, dx \\ &= c_1\, (I_1 - I_2 - I_3 + I_4 + I_5), \end{aligned} \tag{4.38}$$

where

$$I_1 = \int_\Omega (a(x/\epsilon, M_\epsilon u, P_M), P_M)\, dx, \quad I_2 = \int_\Omega (a(x/\epsilon, M_\epsilon u, P_M), \nabla u_\epsilon)\, dx,$$
$$I_3 = \int_\Omega (a(x/\epsilon, u_\epsilon, \nabla u_\epsilon), P_M)\, dx, \quad I_4 = \int_\Omega (a(x/\epsilon, u_\epsilon, \nabla u_\epsilon), \nabla u_\epsilon)\, dx,$$
$$I_5 = \int_\Omega (a(x/\epsilon, u_\epsilon, P_M) - a(x/\epsilon, M_\epsilon u, P_M), P_M - \nabla u_\epsilon)\, dx.$$

Next we will show that

$$I_k \to \int_\Omega (a^*(u, \nabla u), \nabla u)\, dx, \quad j = 1, \cdots, 4$$
$$I_5 \to c_\beta \int_\Omega |P_M - \nabla u_\epsilon|^p\, dx,$$

all as $\epsilon \to 0$. For this purpose we will use the following notation: $\eta_i = |Y_\epsilon^i|^{-1} \int_{Y_\epsilon^i} u\, dx$ and $\xi_i = |Y_\epsilon^i|^{-1} \int_{Y_\epsilon^i} \nabla u\, dx$. Also, we define $J_\epsilon = \{i \in \mathbb{Z}^n : Y_\epsilon^i \bigcap \Omega \neq 0, Y_\epsilon^i \backslash Q \neq 0\}$.

Step 1: $I_1 \to \int_\Omega (a^*(u, \nabla u), \nabla u)\, dx$ as $\epsilon \to 0$

By change of variable, we write

$$\begin{aligned}
I_1 &= \sum_{i \in I_\epsilon} \int_{Y_\epsilon^i} (a(x/\epsilon, M_\epsilon u, P_M), P_M)\, dx + \int_{\Omega \backslash \Omega_\epsilon} (a(x/\epsilon, M_\epsilon u, P_M), P_M)\, dx \\
&= \epsilon^n \sum_{i \in I_\epsilon} \int_Y (a(y, \eta_i, P_{\eta_i, \xi_i}), P_{\eta_i, \xi_i})\, dy + \int_{\Omega \backslash \Omega_\epsilon} (a(y, 0, P_{0,0}), P_{0,0})\, dx \\
&= \epsilon^n \sum_{i \in I_\epsilon} \int_Y (a(y, \eta_i, P_{\eta_i, \xi_i}), \xi_i)\, dy + \int_{\Omega \backslash \Omega_\epsilon} (a(y, 0, P_{0,0}), P_{0,0})\, dx \\
&= \sum_{i \in I_\epsilon} \int_\Omega 1_{Y_\epsilon^i}(x)\, (a^*(\eta_i, \xi_i), \xi_i)\, dx + \int_{\Omega \backslash \Omega_\epsilon} (a(y, 0, P_{0,0}), P_{0,0})\, dx \\
&= I_{11} + I_{12}.
\end{aligned}$$

We claim that

$$I_{11} = \int_\Omega (a^*(M_\epsilon u, M_\epsilon \nabla u), M_\epsilon \nabla u)\, dx \to \int_\Omega (a^*(u, \nabla u), \nabla u)\, dx \quad \text{as } \epsilon \to 0.$$

To this end, we take the difference between this two form:

$$I_{11} - \int_\Omega (a^*(u, \nabla u), \nabla u)\, dx = \int_\Omega (a^*(M_\epsilon u, M_\epsilon \nabla u) - a^*(u, \nabla u), M_\epsilon \nabla u)\, dx + \int_\Omega (a^*(u, \nabla u), M_\epsilon \nabla u - \nabla u)\, dx.$$

It is straightforward to see that the second term vanishes as $\epsilon \to 0$. We only need to apply Holder's inequality to it, and using (4.34), and that $a^*(u, \nabla u) \in L_q(\Omega)$ by (4.14). For the first term, we know that $\|M_\epsilon \nabla u\|_{L_p(\Omega)} \leq c \|\nabla u\|_{L_p(\Omega)}$ and thus by Holder's inequality, we only need to show that $\lim_{\epsilon \to 0}(a^*(M_\epsilon u, M_\epsilon \nabla u) - a^*(u, \nabla u)) = 0$ in $L_q(\Omega)$. Using (4.8) and Holder's inequality

$$\begin{aligned}
&\int_\Omega |a^*(M_\epsilon u, M_\epsilon \nabla u) - a^*(u, \nabla u)|^q\, dx \\
&\leq c \int_\Omega \nu(|M_\epsilon u - u|)^q\, H(M_\epsilon u, M_\epsilon \nabla u, u, \nabla u, p-1)^q\, dx \\
&+ c \int_\Omega |M_\epsilon \nabla u - \nabla u|^{sq}\, H(M_\epsilon u, M_\epsilon \nabla u, u, \nabla u, p-1-s)^q\, dx \\
&\leq c \int_\Omega \nu(|M_\epsilon u - u|)^q\, H(M_\epsilon u, M_\epsilon \nabla u, u, \nabla u, p)\, dx \\
&+ c \|M_\epsilon \nabla u - \nabla u\|^{sq}_{L_p(\Omega)} \\
&\quad \times \left(\|M_\epsilon u\|^{(p-1-s)q}_{L_p(\Omega)} + \|M_\epsilon \nabla u\|^{(p-1-s)q}_{L_p(\Omega)} + \|u\|^{(p-1-s)q}_{L_p(\Omega)} + \|\nabla u\|^{(p-1-s)q}_{L_p(\Omega)} \right)
\end{aligned}$$

The second term goes to zero as $\epsilon \to 0$ by (4.34). Furthermore, since

$$\lim_{\epsilon \to 0} \|M_\epsilon \nabla u - \nabla u\|_{L_p(\Omega)} = 0$$

and u, $M_\epsilon u$ are compact in $L_p(\Omega)$, ∇u, $M_\epsilon \nabla u$ are compact in $L_p(\Omega)$, by Lemma 4.3 the first term on the last inequality vanishes as $\epsilon \to 0$. Having this result, this step is completed if we can show that $I_{12} \to 0$ as $\epsilon \to 0$. Applying (4.4) and Holder's

inequality, we may estimate I_{12} in the following way:

$$\begin{aligned} I_{12} &\leq c \int_{\Omega\backslash\Omega_\epsilon} (1 + |P_{0,0}|^{p-1}) \, |P_{0,0}| \, dx \\ &\leq c \left(|\Omega\backslash\Omega_\epsilon|^{\frac{1}{q}} \, \|P_{0,0}\|_{L_p(\Omega\backslash\Omega_\epsilon)} + \|P_{0,0}\|^p_{L_p(\Omega\backslash\Omega_\epsilon)} \right) \end{aligned}$$

Thus it is enough to prove the vanishing of $\|P_{0,0}\|_{L_p(\Omega\backslash\Omega_\epsilon)}$ as $\epsilon \to 0$. Applying Holder's inequality with $r_1 = (p+\tau)/p$ and $r_2 = (p+\tau)/\tau$, where τ is as in Corollary 4.2, we have

$$\|P_{0,0}\|_{L_p(\Omega\backslash\Omega_\epsilon)} \leq |\Omega\backslash\Omega_\epsilon|^{\frac{\tau}{p+\tau}} \, \|P_{0,0}\|_{L_{p+\tau}(\Omega\backslash\Omega_\epsilon)}.$$

Now by breaking up the integration into sum of integral over Y^i_ϵ, $i \in J_\epsilon$, and by change of variable we have that

$$\|P_{0,0}\|_{L_{p+\tau}(\Omega\backslash\Omega_\epsilon)} \leq \left(\sum_{i \in J_\epsilon} |Y^i_\epsilon| \right)^{\frac{1}{p+\tau}} \|P_{0,0}\|_{L_{p+\tau}(Y)}.$$

By Corollary 4.2, $\|P_{0,0}\|_{L_{p+\tau}(Y)}$ is bounded independent of ϵ. Furthermore, $|\Omega\backslash\Omega_\epsilon| \to 0$, and $\sum_{i \in J_\epsilon} |Y^i_\epsilon| \to 0$ as $\epsilon \to 0$, and hence we have our result for Step 1.

Step 2: $I_2 \to \int_\Omega (a^*(u, \nabla u), \nabla u) \, dx$ as $\epsilon \to 0$

Let $\delta > 0$. Since $u \in W^{1,p}(\Omega)$, there exists simple functions $\chi(x) = \sum_{j=1}^m a_j 1_{\Omega_j}(x)$ and $\psi(x) = \sum_{j=1}^m b_j 1_{\Omega_j}(x)$ as in Lemma 4.8 such that

$$\|u - \chi\|_{L_p(\Omega)} \leq \delta \quad \text{and} \quad \|\nabla u - \psi\|_{L_p(\Omega)} \leq \delta.$$

Let us designate $P_S = P(x/\epsilon, \chi, \psi)$ and write I_2 as follows.

$$\begin{aligned} I_2 &= \int_\Omega (a(x/\epsilon, \chi, P_S), \nabla u_\epsilon) \, dx + \int_\Omega (a(x/\epsilon, M_\epsilon u, P_M) - a(x/\epsilon, \chi, P_S), \nabla u_\epsilon) \, dx \\ &= I_{21} + I_{22}. \end{aligned}$$

We claim that

$$I_{21} \to \int_\Omega (a^*(\chi,\psi),\nabla u)\,dx \quad \text{as } \epsilon \to 0.$$

We may write $I_{21} = \sum_{j=0}^m \int_{\Omega_j} (a(x/\epsilon, a_j, P_{a_j,b_j}),\nabla u_\epsilon)\,dx$. To this end, we note that using (4.4), Corrolary 4.2 $s = (p+\tau)/(p-1) > q$, then $\|a(\cdot,a_j,P_{a_j,b_j})\|_{L_s(\Omega)}$ is uniformly bounded with respect to ϵ. Moreover, $\|\nabla u_\epsilon\|_{L_p(\Omega)}$ is also bounded. Hence, we may set $t = ps/(p+s) > 1$ such that

$$\|(a(\cdot,a_j,P_{a_j,b_j}),\nabla u_\epsilon)\|_{L_t(\Omega)} \le \|a(\cdot,a_j,P_{a_j,b_j})\|_{L_s(\Omega)}\,\|\nabla u_\epsilon\|_{L_p(\Omega)},$$

which means that $(a(\cdot,a_j,P_{a_j,b_j}),\nabla u_\epsilon)$ is uniformly bounded with respect to ϵ. This implies that $(a(\cdot,a_j,P_{a_j,b_j}),\nabla u_\epsilon)$ converges weakly in $L_t(\Omega)$ as $\epsilon \to 0$. Furthermore $a(\cdot,a_j,P_{a_j,b_j})$ converges weakly to $a^*(a_j,b_j)$ in $L_q(\Omega)$, and $\nabla\cdot(a(x/\epsilon,a_j,P_{a_j,b_j})) = 0$. Then by compensated compactness theorem (e.g. [45]) we conclude that

$$(a(\cdot,a_j,P_{a_j,b_j}),\nabla u_\epsilon) \rightharpoonup (a^*(a_j,b_j),\nabla u) \quad \text{in} \quad L_t(\Omega).$$

Thus

$$I_{21} \to \sum_{j=0}^m \int_{\Omega_j} (a^*(a_j,b_j),\nabla u)\,dx = \int_\Omega (a^*(\chi,\psi),\nabla u)\,dx \quad \text{as } \epsilon \to 0.$$

Next, using (4.8) and Holder's inequality, I_{22} is estimated in the following way:

$$\begin{aligned} I_{22} \le\ & c_3 \int_\Omega \nu(|M_\epsilon u - \chi|)\,H(M_\epsilon u, P_M, \chi, P_S, p-1)\,|\nabla u_\epsilon|\,dx \\ & + c_4 \int_\Omega |P_M - P_S|^s\,H(M_\epsilon u, P_M, \chi, P_S, p-1-s)\,|\nabla u_\epsilon|\,dx \\ \le\ & c\left(\int_\Omega \nu(|M_\epsilon u-\chi|)^q\,H(M_\epsilon u, P_M,\chi,P_S,p)\,dx\right)^{\frac{1}{q}} \|\nabla u_\epsilon\|_{L_p(\Omega)} \\ & + c\left(\int_\Omega H(M_\epsilon u, P_M,\chi,P_S,p)\,dx\right)^{\frac{p-1-s}{p}} \|\nabla u_\epsilon\|_{L_p(\Omega)}\|P_M - P_S\|^s_{L_p(\Omega)}. \end{aligned} \tag{4.39}$$

Now we know that $M_\epsilon u$ and χ is compact in $L_p(\Omega)$, P_M, P_S, are uniformly bounded

in $L_{p+\tau}(\Omega)$, and ∇u_ϵ is bounded in $L_p(\Omega)$, by Corrolary 4.2. Then Lemma 4.3 implies that there exists a sequence (c_δ) converging to 0 as $\delta \to 0$. Using Lemma 4.8 we know there exists a constant $c > 0$ independent of δ such that

$$\limsup_{\epsilon \to 0} I_{22} \leq c(c_\delta + \delta^{\frac{s}{p-s}} + \delta). \tag{4.40}$$

Furthermore, similar to (4.39), we use (4.17) and applying Holder's inequality appropriately to obtain

$$\begin{aligned}
&\int_\Omega |(a^*(\chi,\psi) - a^*(u,\nabla u),\nabla u)|\, dx \\
&\leq c \left(\int_\Omega \nu(|\chi - u|)^q H(\chi,\psi,u,\nabla u,p)\, dx \right)^{\frac{1}{q}} \|\nabla u\|_{L_p(\Omega)} \\
&+ c \left(\int_\Omega H(\chi,\psi,u,\nabla u,p)\, dx \right)^{\frac{p-1-s}{p}} \|\nabla u\|_{L_p(\Omega)} \|\psi - \nabla u\|^s_{L_p(\Omega)}.
\end{aligned}$$

Using similar argument as in (4.39) we know that

$$\limsup_{\epsilon -> 0} \int_\Omega |(a^*(\chi,\psi) - a^*(u,\nabla u),\nabla u)|\, dx \leq c\,(c_\delta + \delta^s). \tag{4.41}$$

As δ approaches 0, (4.40) and (4.41) vanish, confirming the desired convergence.

Step 3: $I_3 \to \int_\Omega (a^*(u,\nabla u),\nabla u)\, dx$ as $\epsilon \to 0$

The proof for this step is similar to the proof in Step 2. So let us assume we have the simple functions χ and ψ as in Step 2 and use the notations accordingly. Then we may write I_3 in the following way:

$$\begin{aligned}
I_3 &= \sum_{j=0}^{m} \int_{\Omega_j} (a(x/\epsilon, u_\epsilon, \nabla u_\epsilon), P_{a_j,b_j})\, dx + \int_\Omega (a(x/\epsilon, u_\epsilon, \nabla u_\epsilon), P_M - P_S)\, dx \\
&= I_{31} + I_{32}.
\end{aligned}$$

By homogenization theory [52], $a(x/\epsilon, u_\epsilon, \nabla u_\epsilon)$ converges weakly to $a^*(u,\nabla u)$ in $L_q(\Omega)$. Also, P_{a_j,b_j} converges weakly to b_j in $L_p(\Omega)$, and by Corollary 4.2 P_{a_j,b_j}

is bounded in $L_{p+\tau}(\Omega)$. Consequently, we may find $t = (pq + q\tau)/(pq + \tau) > 1$ such that

$$\|(a(\cdot, u_\epsilon, \nabla u_\epsilon), P_{a_j,b_j})\|_{L_t(\Omega)} \leq \|a(\cdot, u_\epsilon, \nabla u_\epsilon)\|_{L_q(\Omega)} \|P_{a_j,b_j}\|_{L_{p+\tau}(\Omega)},$$

Taking into account (4.1), by compensated compactness theorem

$$I_{31} \to \sum_{j=0}^{m} \int_{\Omega_j} (a^*(u, \nabla u), b_j)\, dx = \int_{\Omega} (a^*(u, \nabla u), \psi)\, dx.$$

Furthermore, by Holder's inequality

$$I_{32} \leq \|a(\cdot, u_\epsilon, \nabla u_\epsilon)\|_{L_q(\Omega)} \|P_M - P_S\|_{L_p(\Omega)},$$

which by Lemma 4.8 and following the same argument as in Step 2, gives

$$\limsup_{\epsilon \to 0} I_{32} \leq c(c_\delta + \delta^{\frac{1}{p-s}} + \delta).$$

Finally, using Holder's inequality,

$$\int_{\Omega} |(a^*(u, \nabla u), \psi - \nabla u)|\, dx \leq \|a^*(u, \nabla u)\|_{L_q(\Omega)} \|\psi - \nabla u\|_{L_p(\Omega)} \leq c\,\delta.$$

Since δ is arbitrarily we have obtained the desired convergence.

Step 4: $I_4 \to \int_\Omega (a^*(u, \nabla u), \nabla u)\, dx$ as $\epsilon \to 0$

Using (4.1) and (4.9) along with Green's formula, it is straightforward to see that

$$\int_{\Omega} (a(x/\epsilon, u_\epsilon, \nabla u_\epsilon), \nabla u_\epsilon)\, dx = \int_{\Omega} (-\nabla \cdot (a(x/\epsilon, u_\epsilon, \nabla u_\epsilon)), u_\epsilon)\, dx = \int_{\Omega} f\, u_\epsilon\, dx,$$
$$\int_{\Omega} (a^*(u, \nabla u), \nabla u)\, dx = \int_{\Omega} (-\nabla \cdot (a^*(u, \nabla u)), u)\, dx = \int_{\Omega} f\, u\, dx.$$

But homogenization result tells us that u_ϵ converges weakly to u in $W^{1,p}(\Omega)$, which gives our claim.

Step 5: $I_5 \to c_\beta \int_\Omega |P_M - \nabla u_\epsilon|^p\, dx$,

Using (4.8) and Holder's and Young's inequalities with some constant $\beta > 0$ we estimate I_5 as follows.

$$\begin{aligned} I_5 &\leq c \int_\Omega \nu(|u_\epsilon - M_\epsilon u|)\, H(1, u_\epsilon, P_M, M_\epsilon u, P_M, p-1)\, |P_M - \nabla u_\epsilon|\, dx \\ &\leq c\, \beta^{-q}\, q^{-1} \int_\Omega \nu(|u_\epsilon - M_\epsilon u|)^q\, H(1, u_\epsilon, P_M, M_\epsilon u, P_M, p)\, dx \\ &\quad + c\, \beta^p\, p^{-1}\, \|P_M - \nabla u_\epsilon\|^p_{L_p(\Omega)}. \end{aligned}$$

By similar argument as in previous steps, we know that the first term vanishes as $\epsilon \to 0$. Now we may choose $\beta > 0$ such that this last term is absorbed to the left hand side of (4.38). Combining all results from the five steps we have proved the theorem. □

4.5. Numerical Implementations

In this section we present several ingredients pertaining to the implementation of the numerical homogenization. Obviously, we need to perform an iterative technique to tackle the nonlinearity. This is achieved by using an Inexact-Newton algorithm.

Moreover, the approximation property of the corrector $\mathcal{P}(x, x/\epsilon)$ (cf. Lemma 4.5) reveals the existence of a resonance error proportional to ϵ/h, which is resulted from the mismatch due to the imposed linear boundary conditions for the local problem in the multiscale map E. This drawback can be overcome by oversampling the multiscale map E on the element larger than $h+\epsilon$, and use only the information from the original element.

4.5.1. An Inexact-Newton Algorithm

For the numerical examples below we use $a_\epsilon(x, u_\epsilon, \nabla u_\epsilon) = a_\epsilon(x, u_\epsilon)\nabla u_\epsilon$. Let $\{\phi_i\}_{i=1}^d$ be the standard piecewise linear basis functions of X^h. Then the solution of numerical

homogenization (4.2) may be written as

$$u^h = \sum_{i=1}^{d} \alpha_i \, \phi_i$$

for some $\alpha = (\alpha_1, \alpha_2, \cdots, \alpha_d)^T$, where α_i depends on ϵ. Hence (4.2) can be viewed as to find α such that

$$F(\alpha) = 0,$$

where $F : \mathbb{R}^d \to \mathbb{R}^d$ is a nonlinear operator such that

$$F_i(\alpha) = \sum_{K \in T_h} \int_K (a_\epsilon(x, \eta^h) \nabla v_\epsilon), \nabla \phi_i) \, dx - \int_\Omega f \, \phi_i \, dx. \tag{4.42}$$

We note that in (4.42) α is implicitly buried in η^h, and v_ϵ. An inexact-Newton algorithm is a variation of Newton's iteration for nonlinear system of equations in that the system Jacobian is only solved approximately. To be specific, given an initial iterate α^0, for $k = 0, 1, 2, \cdots$ until convergence do the following:

- Solve $F'(\alpha^k)\delta^k = -F(\alpha^k)$ until by some iterative technique until $\|F(\alpha^k) + F'(\alpha^k)\delta^k\| \leq \beta_k \|F(\alpha^k)\|$.
- Update $\alpha^{k+1} = \alpha^k + \delta^k$.

In this algorithm $F'(\alpha^k)$ is the Jacobian matrix evaluated at iteration k. We note that when $\beta_k = 0$ then we have recovered the classical Newton iteration. Here we have used

$$\beta_k = 0.001 \left(\frac{\|F(\alpha^k)\|}{\|F(\alpha^{k-1})\|} \right)^2,$$

with $\beta_0 = 0.001$. Choosing β_k this way we avoid oversolving the Jacobian system when α^k is still considerably far from the exact solution.

Next we present the entries of the Jacobian matrix. For this purpose, we use the following notations. Let $T_h^i = \{K \in T_h : z_i \text{ is a vertex of K}\}$, $I^i = \{j :$

z_j is a vertex of $K \in T_h^i\}$, and $T_h^{ij} = \{K \in T_h^i : K$ shares $\overline{z_i z_j}\}$. We note that we may write $F_i(\alpha)$ as follows:

$$F_i(\alpha) = \sum_{K \in T_h^i} \left(\int_K (a_\epsilon(x, \eta^h) \nabla v_\epsilon, Dx\phi_i)\, dx - \int_K f\, \phi_i\, dx \right),$$

with

$$-\nabla \cdot (k(x, \eta^h)\, \nabla v_\epsilon) = 0 \text{ in } K \quad \text{and} \quad v_\epsilon = \sum_{z_m \in Z_K} \alpha_m\, \phi_m \text{ on } \partial K, \tag{4.43}$$

where Z_K is all the vertices of element K. It is apparent that $F_i(\alpha)$ is not fully dependent on all $\alpha_1, \alpha_2, \cdots, \alpha_d$. Consequently, $\frac{\partial F_i(\alpha)}{\partial \alpha_j} = 0$ for $j \notin I^i$. To this end, we denote $\psi_j = \frac{\partial v_\epsilon}{\partial \alpha_j}$. By applying chain rule of differentiation to (4.43) we have the following local problem for ψ_j:

$$-\nabla \cdot (a_\epsilon(x, \eta^h)\, \nabla \psi_j) = \frac{1}{3} \nabla \cdot (\frac{\partial a_\epsilon(x, \eta^h)}{\partial u}\, \nabla v_\epsilon) \text{ in } K \quad \text{and} \quad \psi_j = \phi_j \text{ on } \partial K. \tag{4.44}$$

Thus provided that v_ϵ has been computed, then we may compute ψ_j using (4.44). Using the above descriptions we have the expressions for the entries of the Jacobian matrix:

$$\frac{\partial F_i}{\partial \alpha_i} = \sum_{K \in T_h^i} \left(\frac{1}{3} \int_K (\frac{\partial a_\epsilon(x, \eta^h)}{\partial u}\, \nabla v_\epsilon, \nabla \phi_i)\, dx + \int_K (a_\epsilon(x, \eta^h)\, \nabla \psi_i, \nabla \phi_i)\, dx, \right)$$

$$\frac{\partial F_i}{\partial \alpha_j} = \sum_{K \in T_h^{ij}} \left(\frac{1}{3} \int_K (\frac{\partial a_\epsilon(x, \eta^h)}{\partial u}\, \nabla v_\epsilon, \nabla \phi_i)\, dx + \int_K (a_\epsilon(x, \eta^h)\, \nabla \psi_j, \nabla \phi_i)\, dx, \right)$$

for $j \neq i$, $j \in I^i$.

From this derivation it is obvious that the Jacobian matrix is not symmetric but sparse. Computation of this Jacobian matrix is similar to computing the stiffness matrix resulting from standard finite element, in that each entry is formed by accumulation of element by element contribution. Once we have the matrix stored in

memory, then its action to a vector is straightforward. Since it is a sparse matrix, devoting some amount of memory for entries storage is not terribly expensive.

4.5.2. An Oversampling Technique

First, we describe an oversampling technique, for general nonlinear elliptic problem. The idea is similar to linear ellptic problem, in that the multiscale map is solved on a domain larger than the element K (see Figure 2.2 in Chapter II). In general, given $v^h \in X^h$, where v^h is defined in K, we want to find v_ϵ that satisfies

$$\nabla \cdot a(x/\epsilon, \eta^h, \nabla v_\epsilon) = 0 \quad \text{in } S \tag{4.45}$$

such that $v_\epsilon(z_i) = v^h(z_i)$.

For special cases in which the gradient in the coefficient is linear, i.e., $a(x/\epsilon, \eta, \xi) = a(x/\epsilon, \eta)\,\xi$, given $v^h \in X^h$, we define

$$v_\epsilon = \sum_{i=1}^{3} c_i\, \phi_\epsilon^i,$$

where ϕ_ϵ^i satisfies

$$\nabla \cdot (a(x/\epsilon, \eta^h)\, \nabla \phi_\epsilon^i) = 0 \quad \text{in } S$$

$$\phi_\epsilon^i = \phi^i \quad \text{on } \partial S.$$

The constants c_i, $i = 1, 2, 3$ are determined by imposing the conditions

$$v_\epsilon^h(z_j) = v^h(z_j) \qquad j = 1, 2, 3.$$

We note that the piecewise constants in η^h are taken as the average over the element K. It is obvious that for this special case, the oversampling technique resembles its counterpart in linear elliptic problems.

Table 4.9. Numerical homogenization errors without oversampling

N	L_2 error	H^1 error	L_∞ error
32	4.2583×10^{-4}	8.2632×10^{-3}	1.0065×10^{-3}
64	6.6652×10^{-4}	1.2554×10^{-2}	1.1875×10^{-3}
128	7.6030×10^{-4}	1.6000×10^{-2}	1.3525×10^{-3}

4.5.3. Example

We want to solve the following problem:

$$-\nabla \cdot (a(x/\epsilon, u_\epsilon)\nabla u_\epsilon) = -1 \quad \text{in } \Omega \subset \mathbb{R}^2,$$

$$u_\epsilon = 0 \quad \text{on } \partial\Omega,$$

where $\Omega = [0,1] \times [0,1]$, $a(x/\epsilon, u_\epsilon) = k(x/\epsilon)/\left(1 + u_\epsilon\right)^{l(x/\epsilon)}$, with

$$k(x/\epsilon) = \frac{2 + 1.8\sin(2\pi x_1/\epsilon)}{2 + 1.8\cos(2\pi x_2/\epsilon)} + \frac{2 + \sin(2\pi x_2/\epsilon)}{2 + 1.8\cos(2\pi x_1/\epsilon)}$$

and $l(x/\epsilon)$ is generated from $k(x/\epsilon)$ such that the average of $l(x/\epsilon)$ over Ω is 2. Here we use $\epsilon = 0.01$. Since the exact solution for this problem is not available, we use a finely resolved numerical solution using standard finite element method as a reference solution. The discretization of the domain Ω follows the one in section 3.6 of Chapter III. The reference solution is solved on 512×512 mesh. Tables 4.9 and 4.10 present the errors of the solution with and without oversampling, respectively. In each table,the second, third, and fourth columns list the relative error in L_2, H^1, and L_∞ norm, respectively. As we can see from these two tables, the oversampling significantly improves the accuracy of the multiscale method.

Table 4.10. Numerical homogenization errors with oversampling

N	L_2 error	H^1 error	L_∞ error
32	2.6110×10^{-5}	2.4123×10^{-3}	1.1367×10^{-4}
64	3.5252×10^{-5}	1.3218×10^{-3}	6.9110×10^{-5}
128	1.6402×10^{-5}	6.2158×10^{-4}	3.2610×10^{-5}

CHAPTER V

APPLICATIONS TO POROUS MEDIA FLOW

In this chapter we will present applications of the multiscale method to several problems in porous media flow. First, we describe briefly various geometrical terminologies related to the method. We note that this description follows the setting in the numerical examples of Chapter III. Let T_h denote the collection of coarse elements/rectangles K, whose side lengths are h_1 and h_2 in the x_1- and x_2-directions, respectively, and the maximum of those two is h. We describe the construction of the control volumes as follows. Consider a coarse rectangular element K, and let ξ_K be its center. The element K is divided into four rectangles of equal area by connecting ξ_K to the midpoints of the element's edges. We denote these rectangulars by K_z, where $z \in Z_h(K)$ are the vertices of K. Also, we denote by $Z_h = \bigcup_K Z_h(K)$ the collection of all vertices and by $Z_h^0 \subset Z_h$ the vertices which do not lie on the Dirichlet boundary of Ω. The control volume V_z is defined as the union of the quadrilaterals K_z sharing the vertex z (see Figure 5.7).

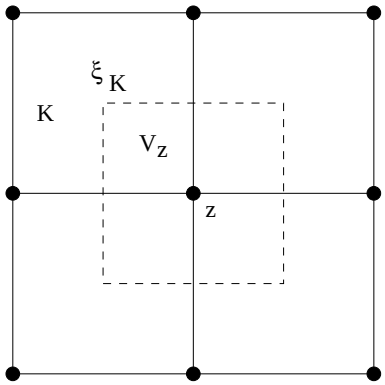

Fig. 5.7. Rectangular control volume

5.1. Two-Phase Flow in Oil Reservoir Simulation

We consider two-phase immiscible flow in a reservoir Ω under the assumption that the displacement is dominated by viscous effects; i.e., we neglect the effects of gravity, compressibility, and capillary pressure. Porosity will be considered to be constant. The two phases will be referred to as water and oil, designated by subscripts w and o, respectively.

5.1.1. Fine and Coarse Scale Models

The flow of two immiscible fluids in a porous medium Ω is governed by the mass balance equation for each fluid and the generalized Darcy's Law [10]:

$$\frac{\partial(\phi\,\rho_\alpha\,S_\alpha)}{\partial t} + \nabla\cdot(\rho_\alpha v_\alpha) = q_\alpha, \tag{5.1}$$

$$v_\alpha = -\frac{k k_{r\alpha}}{\mu_\alpha}(\nabla p_\alpha - \rho_\alpha\, g), \tag{5.2}$$

where $\alpha = w, o$, respectively denote the water phase and non-aqueous phase (for example oil). The variables ϕ and k are the porosity and the absolute permeability of the porous medium, ρ_α, μ_α, S_α, p_α, v_α, and $k_{r\alpha}$, are respectively the density, viscosity, saturation, pressure, velocity, and relative permeability of α phase. The variable g denotes the gravity acceleration. It is a common assumption that the two fluid phases filled all the void volume of porous medium, that is, $0 \le S_w, S_o \le 1$, and

$$S_w + S_o = 1. \tag{5.3}$$

We note that field and experimental observations show that the phase relative permeability is dependent on the phase saturation, i.e., $k_{r\alpha} = k_{r\alpha}(S_\alpha)$. Furthermore, the density ρ_α, and the viscosity μ_α can depend on the pressure p_α. Moreover, the pressures of the two phases are related to each other by the capillary pressure function,

which we denote by p_c:

$$p_c = p_w - p_o. \tag{5.4}$$

As mentioned at the beginning of this section, the fine model of the two-phase immiscible flow that we will upscale is derived from the governing equations described above under several assumptions, namely, effects of source/sink, gravity and capillary pressure are neglected ($q_\alpha = g = p_c = 0$), the fluids are incompressible (ρ_α is constant), and the porosity ϕ is constant.

Hence we may write $p = p_w = p_o$ which serves as a global pressure. Now we define the following total velocity v as the sum of each phase velocity,

$$v = v_w + v_o. \tag{5.5}$$

We denote by $\lambda = \lambda(S_w)$ the total mobility function which can be expressed as

$$\lambda(S_w) = \frac{k_{rw}(S_w)}{\mu_w} + \frac{k_{ro}(1 - S_w)}{\mu_o}. \tag{5.6}$$

Substitution of the Darcy's Law for each phase to (5.5), and using (5.6) gives:

$$v = -\lambda(S_w)k\nabla p. \tag{5.7}$$

Writing (5.1) for each phase $\alpha = w, o$ along with the related assumptions, and summing up the resulting equations give

$$\nabla \cdot v = 0. \tag{5.8}$$

Here we have used the fact that $S_w + S_0 = 1$ (thus the time derivative is zero), (5.5). Combining (5.8) with (5.7) gives the elliptic pressure equation

$$-\nabla \cdot \lambda(S_w)k\nabla p = 0. \tag{5.9}$$

Finally, the mass balance equation for the water phase is now written as

$$\frac{\partial S_w}{\partial t} + \nabla \cdot v_w = 0. \tag{5.10}$$

We will write the water phase velocity in terms of the total velocity v. For this purpose, we introduce the water phase relative mobility function denoted by $f(S_w)$:

$$f(S_w) = \frac{k_{rw}(S_w)/\mu_w}{\lambda(S_w)}. \tag{5.11}$$

Then we may write the water phase velocity as

$$v_w = f(S_w)v, \tag{5.12}$$

which gives

$$\frac{\partial S_w}{\partial t} + v \cdot \nabla f(S_w) = 0. \tag{5.13}$$

To summarize, denoting $S = S_w$, the two phase flow fine model is governed by the following pressure-saturation equations:

$$-\nabla \cdot \lambda(S) k \nabla p = 0, \tag{5.14}$$

$$\frac{\partial S}{\partial t} + v \cdot \nabla f(S) = 0. \tag{5.15}$$

We note that in this work, a single set of relative permeability curves is used and k is taken to be a diagonal tensor, $\mathrm{diag}(k_1, k_2)$.

Next, we wish to develop a coarse scale description for two-phase flow in heterogeneous porous media. Previous approaches for upscaling such systems are discussed by many authors; e.g., [14, 4, 20, 25]. In most upscaling procedures, the coarse scale pressure equation is of the same form as the fine scale equation (5.14), but with an equivalent grid block permeability tensor k^* replacing k. For a given coarse scale grid block, the tensor k^* is generally computed through the solution of the pressure

equation over the local fine scale region corresponding to the particular coarse block [18]. Coarse grid k^* computed in this manner have been shown to provide accurate solutions to the coarse grid pressure equation. We note that some upscaling procedures additionally introduce a different coarse grid functionality for λ, though this does not appear to be essential in our formulation.

In this work, the proposed coarse model is upscaling the pressure equation (5.14) to obtain the velocity field on the coarse grid and use it in (5.15) to solve the saturation on the coarse grid. A finite volume element method is implemented to upscale the pressure equation (5.14). Finite volume is chosen, because, by its construction, it enjoys the numerical local conservation which is important in groundwater and reservoir simulations.

As mentioned in Chapter III, the key idea of the method is the construction of basis functions on the coarse grids such that these functions capture the small scale information on each of these coarse grids. Here, these nodal basis functions are denoted by $\{\psi_z\}_{z\in Z_h^0}$. Having described the basis functions, we denote by V_ϵ^h the space of our approximate pressure solution which is spanned by the basis functions $\{\psi_z\}_{z\in Z_h^0}$. Now, we may formulate the finite-dimensional problem corresponding to finite volume element formulation of (5.14). A statement of mass conservation on a control volume V_z is formed from (5.14), where now the approximate solution is written as a linear combination of the basis functions. Assembly of this conservation statement for all control volumes would give the corresponding linear system of equations that can be solved accordingly. It is obvious that the number of control volumes V_z has to be equal to the dimension of the space V_ϵ^h. The resulting linear system has incorporated the fine scale information through the involvement of the nodal basis functions on the approximate solution. To be specific, the problem now is to seek $p^h \in V_\epsilon^h$ with

$p^h = \sum_{z \in Z_h^0} p_z \psi_z$ such that

$$\int_{\partial V_z} \lambda(S)\, k\, \nabla p^h \cdot \vec{n}\, dl = 0 \tag{5.16}$$

for every control volume $V_z \subset \Omega$. Here $\vec{n}$ defines the normal vector on the boundary of the control volume, ∂V_z and S indicates the fine scale saturation field. We note that concerning the basis functions, a vertex-centered finite volume difference is used to solve the local boundary value problem in each element K along with a harmonic average to approximate the permeability k at the edges of fine control volumes.

As mentioned earlier, the pressure solution may then be used to compute the total velocity field at the coarse scale level, denoted by $\overline{v} = (\overline{v}_1, \overline{v}_2)$ via (5.7). In general, the following equations are used to compute the velocities in the horizontal and vertical directions, respectively:

$$\overline{v}_1 = -\frac{1}{h_2} \sum_{z \in Z_h^0} p_z \left(\int_E \lambda(S) k_1 \frac{\partial \psi_z}{\partial x_1} dx_2 \right), \tag{5.17}$$

$$\overline{v}_2 = -\frac{1}{h_1} \sum_{z \in Z_h^0} p_z \left(\int_E \lambda(S) k_2 \frac{\partial \psi_z}{\partial x_2} dx_1 \right), \tag{5.18}$$

where E is the edge of V_z. Furthermore, for the control volumes V_z adjacent to Dirichlet boundary (which are half control volumes), we can derive the velocity approximation using the conservation statement derived from (5.14) on V_z. One of the terms involved is the integration along part of Dirichlet boundary, while the rest of the three terms are known from the adjacent internal control volumes calculations. The integration of forcing function may be approximated by midpoint rule. This way, we have the following equations (l, b, r, and t stand for left, bottom, right, and top,

respectively):

$$\begin{aligned} \overline{v}_1^l &= \overline{v}_1^r + 0.5\ h_1/h_2\left(\overline{v}_2^t - \overline{v}_2^b\right) \quad \text{for left Dirichlet boundary,} \\ \overline{v}_2^b &= \overline{v}_2^t + 0.5\ h_2/h_1\left(\overline{v}_1^r - \overline{v}_1^l\right) \quad \text{for bottom Dirichlet boundary.} \end{aligned} \tag{5.19}$$

The right and the top Dirichlet boundary conditions are defined similarly. It has been well known that these approximations give a second order accuracy to the velocity computation.

In this section we will consider two different coarse models for the saturation equation. One of them is a simple/primitive model where we use only the coarse scale velocity to update the saturation field on the coarse grid, i.e.,

$$\frac{\partial \overline{S}}{\partial t} + \overline{v} \cdot \nabla f(\overline{S}) = 0. \tag{5.20}$$

In this case no upscaling of the saturation equation is performed. This kind of technique in conjunction with the upscaling of absolute permeability is commonly used in applications (e.g., [22, 21, 20]). The difference of our approach is that the coupling of the small scales is performed through the finite volume element formulation of the global problem and the small scale information of the velocity field can be easily recovered using the multiscale basis functions. Within this upscaling framework, we use $\overline{S}$ instead of S in (5.16). If the saturation profile is smooth, this approximation is of first order. In the coarse blocks where the discontinuities of S are present, we need to modify the stiffness matrix corresponding to these blocks. The latter requires the values of the fine scale saturation. In our computation we will not do this. We simply use $\lambda(\overline{S})$ in (5.16). It has been demonstrated in previous findings [27] that such approach gives a reasonable accuracy.

5.1.2. Macro-Diffussion Model

In addition to the above described coarse model, we will also revisit a coarse model on the saturation proposed by [27], which uses $\lambda(S) = 1$ and $f(S) = S$. This model was derived using perturbation argument for (5.15), in which the saturation, S, and the velocity, v, on the fine scale are assumed to be the sum of their volume-averaged and fluctuating components,

$$v = \overline{v} + v', \quad S = \overline{S} + S'. \tag{5.21}$$

Here, the overbar quantities designate the average of fine scale quantities over the coarse control volume. Since our model uses rectangular control volumes, we may assume that (cf. [63])

$$\overline{\nabla f} = \nabla \overline{f}. \tag{5.22}$$

Substituting (5.21) into the saturation equation for single phase and averaging over coarse blocks, we obtain

$$\frac{\partial \overline{S}}{\partial t} + \overline{v} \cdot \nabla \overline{S} + \overline{v' \cdot \nabla S'} = 0. \tag{5.23}$$

The term $\overline{v' \cdot \nabla S'}$ represents subgrid effects due to the heterogeneities of convection. With the assumption that v' is divergence free, and using (5.22), this subgrid effects may be written as

$$\overline{v' \cdot \nabla S'} = \overline{\nabla \cdot (v'S')} = \nabla \cdot \overline{(v'S')}.$$

Our aim is to derive a representation for the cross term $\overline{v_i' S'}$, $i = 1, 2$. This term can be modeled using the equation for S' that is derived by subtracting (5.23) from the fine scale equation (5.15)

$$\frac{\partial S'}{\partial t} + \overline{v} \cdot \nabla S' + v' \cdot \nabla \overline{S} + v' \cdot \nabla S' = \overline{v' \cdot \nabla S'}. \tag{5.24}$$

The differential equation (5.24) can be solved along the characteristics $dx(\tau)/d\tau = \overline{v}$, $0 \le \tau \le t$. To be specific, using these characteristics, we rewrite (5.24) in terms of the total time derivative of S' for (x,t) with $x(t) = x$ as follows:

$$\frac{dS'(x,t)}{dt} + v' \cdot \nabla \overline{S} + v' \cdot \nabla S' = \overline{(v' \cdot \nabla S')}. \tag{5.25}$$

Integrating (5.25) over $(0,t)$ we obtain

$$\begin{aligned} S'(x,t) = &- \int_0^t v'(x(\tau)) \cdot \nabla \overline{S}(x(\tau),\tau)\, d\tau - \int_0^t v'(x(\tau)) \cdot \nabla S'(x(\tau),\tau)\, d\tau \\ &+ \int_0^t \overline{v'(x(\tau)) \cdot \nabla S'(x(\tau),\tau)}\, d\tau. \end{aligned} \tag{5.26}$$

Now, we only need to multiply (5.26) by $v_i'(x)$ and take the average over the control volume. We note that upon this multiplication, the second term in (5.26) will be neglected since it consists of higher order terms of the fluctuating components. Also upon taking average over the control volume (after the multiplication with v_i'), the corresponding third term vanishes since $\overline{v_i'} = 0$. To summarize, we now have the following representation of the cross term $\overline{v_i' S'}$:

$$\overline{v_i'(x) S'(x,t)} = -\overline{v_i'(x) \int_0^t v'(x(\tau)) \cdot \nabla \overline{S}(x(\tau),\tau)\, d\tau}, \quad i = 1,2.$$

Moreover, we assume that $\overline{S}$ does not significantly change along the characteristics. Thus,

$$\overline{v_i'(x) S'(x,t)} = -\sum_{j=1}^{2} \left(\int_0^t \overline{v_i'(x)\, v_j'(x(\tau))}\, d\tau \right) \frac{\partial \overline{S}(x,t)}{\partial x_j}, \quad i = 1,2.$$

Hence, using this last equation, we obtain the following coarse scale saturation equation which has taken into account the subgrid effects:

$$\frac{\partial \overline{S}}{\partial t} + \overline{v} \cdot \nabla \overline{S} - \nabla \cdot \big(D(x,t) \nabla \overline{S}(x,t) \big) = 0, \tag{5.27}$$

where $D(x,t)$ is the macro-diffusive tensor, whose entries are written as

$$D_{ij}(x,t) = \int_0^t \overline{v_i'(x)v_j'(x(\tau))}\, d\tau. \tag{5.28}$$

Next, we want to approximate the macro-diffusive tensor in a reasonable fashion. For this purpose, we denote by $L_j(x,t)$, $j = 1,2$, the displacement of the particle in x_j-direction that starts at point x and travels with velocity $-v_j$ (see Figure 5.8). Using the fact that $\overline{v_i'} = 0$ we have from (5.28) that

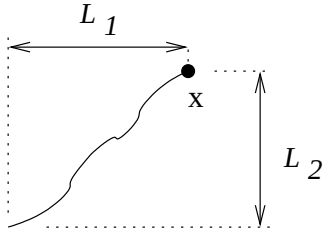

Fig. 5.8. The trajectory of particle x

$$\begin{aligned} D_{ij}(x,t) &= \int_0^t \overline{v_i'(x)(v_j'(x(\tau)) + \overline{v_j})}\, d\tau \\ &= \overline{v_i'(x) \int_0^t v(x(\tau))\, d\tau} \\ &= \overline{v_i'(x)\, L_j(x,t)}. \end{aligned} \tag{5.29}$$

The diffusion term in the coarse model for the saturation field (5.27) represents the effects of the small scales on the large ones. Note that the diffusion coefficient is a correlation between the velocity perturbation and the displacement. This is different from [27], where the diffusion is taken to be proportional to the length of the coarse scale trajectory. Using our upscaling methodology for the pressure equation, we can recover the small scale features of the velocity field that allows us to compute the fine scale displacement.

For the nonlinear flux, $f(S)$, we can use a similar argument by using Taylor

expansion around $\overline{S}$:

$$f(S) = f(\overline{S} + S') = f(\overline{S}) + f_S(\overline{S})S' + \dots.$$

In this expansion we will take into account only linear terms and assume that the flux is nearly linear. This case is similar to the linear case, and the analysis can be carried out in an analogous manner. For this case, we use the characteristics $dx(\tau)/d\tau = f_S(\overline{S})\,\overline{v}$ to obtain the corresponding equations for $S'(x,t)$ similar to (5.25) and (5.26). Furthermore similar trajectory as described in Figure 5.8 uses L_j as a displacement of a particle that travels with velocity $-f_S(\overline{S})v$. The resulting coarse scale equation has the form

$$\frac{\partial \overline{S}}{\partial t} + \overline{v} \cdot \nabla \overline{S} = \nabla \cdot (f_S(\overline{S})^2 D(x,t) \nabla \overline{S}(x,t)), \tag{5.30}$$

where $D(x,t)$ is the macro-diffusive tensor corresponding to the linear flow. This formulation has been derived within the stochastic framework in [46]. We note that the higher-order terms in the expansion of $f(S)$ may result in other effects that have not been studied extensively to the best of our knowledge. In [26] the authors use a similar formulation, although their implementation is different from ours. A couple of numerical examples for nonlinear flux $f(S)$ with $\lambda(S) = 1$ will be presented below.

5.1.3. Numerical Results

We now present numerical results that demonstrate the accuracy and limitation of our model compared to the fine scale model. As in [27], the systems considered are representative of cross sections in the subsurface. We therefore set the system length in the horizontal direction x (L_x) to be greater than the formation thickness (L_z); in the results presented below, $L_x/L_z = 5$. The problem that will be analyzed is typical in oil reservoir simulation, where a porous medium is initially occupied by oil. One

way to displace the oil out of the porous medium is by injecting water horizontally from the left boundary, and an immiscible displacement is assumed to occur. A no flow boundary condition is imposed on the upper and lower boundaries Γ_n. Figure 5.9 shows a description of this problem.

The fine model uses 120×120 rectangular elements. The absolute permeability is set to be $\mathrm{diag}(k, k)$. Thus, the fine grid permeability fields are 121×121 realizations of prescribed overall variance (quantified via σ^2, the variance of $\log k$), correlation structure, and covariance model. We consider models generated using GSLIB algorithms [16], characterized by spherical and exponential variograms [58, 16]. The dimension of the coarse models range from 10×10 to 40×40 elements and are generated using a uniform coarsening of the fine grid description.

For the spherical and exponential variogram models, the dimensionless correlation lengths (nondimensionalized by L_x and L_z, respectively) are designated by l_x and l_z. We set the relative permeabilities of oil and water to be simple quadratic functions of their respective saturations; i.e., $k_{rw} = S^2$ and $k_{ro} = (1 - S)^2$, where S is the water saturation. In all cases we fix pressure and saturation ($S = 1$) at the inlet edge of the model ($x = 0$) and also fix pressure at the outlet ($x = L_x$). The top and bottom boundaries are closed to flow. In this study, we applied our models to a variety of permeability fields.

Results are presented in terms of the fraction of oil in the production edge Γ_p, which is denoted by F, where $F = q_o/q$, with q_o being the volumetric flow rate of oil produced at the outlet edge and q the volumetric flow rate of total fluid produced at the production edge. It can expressed by the following equation:

$$F(t) = \frac{\int_0^{L_z} v_x(L_x, z, t)\,(1 - S(L_x, z, t))\,dz}{\int_0^{L_z} v_x(L_x, z, t)\,dz}, \tag{5.31}$$

where $1 - S$ is the saturation of oil. The fractional curve F will be plotted against

pore volumes injected (PVI). PVI is analogous to dimensionless time and is defined as qt/V_p, where t is dimensional time and V_p is the total pore volume of the system. It can be expressed as

$$\text{PVI} = \frac{t}{L_x\, L_z} \int_0^{L_z} v_x(0, z, t)\, dz, \tag{5.32}$$

where it is understood that PVI is the time required to fill all the domain by water injected on Γ_i. Our first example in Figure 5.10 is for the case $l_x = 0.4$, $l_z = 0.04$, and

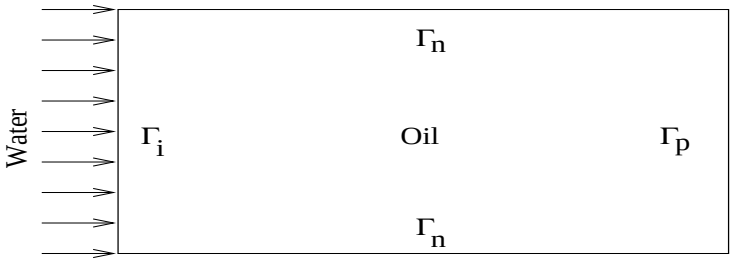

Fig. 5.9. Benchmark problem

$\sigma = 1.5$. An exponential variogram is used to generate the permeability realization. In the following two figures, the 120×120 fine model is represented by solid lines, while the coarse models are represented by the dashed lines and dotted lines, depending on the coarse model's dimension. On the left plot, the coarse model were run on 10×10 elements (dotted lines) and 30×30 elements (dashed lines). On the right plot, the coarse model were run on 20×20 elements (dotted lines) and 40×40 elements (dashed lines). In both of these plots, the coarse model overpredicts the breakthrough time and continues to overpredict the production of the displaced fluid until PVI ≈ 1. After that time the comparison shows that the coarse model agrees reasonably well with the fine model. Also, it can be observed that the finer coarse models are more accurate in general. For example, the 40×40 coarse scale model gives a reasonable approximation of the fine scale model.

For the second example, we consider an isotropic field. Figure 5.11 shows comparison of the fractional flow for case $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1.0$. Both plots in this figure show a good agreement between the fine model and coarse model, regardless of the coarse model dimensions. In conclusion, we would like to note that our coarse scale model tends to perform better for smaller correlation length. In particular, for the upscaling of high correlation length cases, we need larger coarse scale models. This difficulty can be relieved by introducing the nonuniform coarsening, which is a subject of further research.

Another important aspect that requires consideration is the ability of the coarse model to predict the saturation contour. In the following, we compare the saturation contours obtained from fine and coarse models with the same two permeability field scenarios as in the previous figures. The saturation contours are compared in the following fashion: the fine scale model result is averaged onto the coarse grid and then is overlapped with the result from the coarse model of 20×20 elements. In the subsequent figures, the following description is used: the upper plot shows $\overline{S} = 0.10$, the middle plot shows $\overline{S} = 0.30$, and the lower plot shows $\overline{S} = 0.50$.

Figure 5.12 gives comparison of saturation contours at PVI = 0.15, which is before breakthrough time. In general, the coarse model is able to predict the trends exhibited by the fine model, although for smaller values of saturation, it cannot quite follow the fingering indicated by the fine model as evident in upper and middle plots. For a higher value of saturation, however, the coarse model can follow the fingering indicated by the fine model as seen in lower plot. Similar behavior is shown in Figure 5.13 for isotropic field with $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1$. These comparisons also show that the coarse model predicts the contour of saturation better for lower correlation lengths compared to the case with higher correlation length along the main flow direction, $l_x = 0.4$, $l_z = 0.04$, and $\sigma = 1.5$.

At this stage, we present several numerical results of our coarse model with the macro-diffusion as described in subsection 5.1.1. Comparison is made between this transport coarse model with the primitive model, cf. (5.20). Contrary to the coarse model with macro-diffusion, by its nature, the primitive model does not account for the subgrid effects on the coarse grid. The macro-diffusion is computed using the approximation of the fine scale velocity field by sampling the basis functions.

The performance of this macro-diffusion model is exhibited in Figures 5.14 and 5.15. The following notation and terminology are used in those two figures. The solid line represents the fine model run on 120×120 elements, which as before, serves as a reference solution. The dashed line represents the primitive coarse model (D=0), while the dotted line represents the coarse model with macro-diffusion (with D). All coarse models are run on the 10×10 elements.

Figure 5.14 shows the macro-diffusion model performance in the case of a linear flux function, $f(S) = S$ and $\lambda(S) = 1$. The plot on the left corresponds to the isotropic permeability field with $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1.5$, and the plot on the right corresponds to permeability field with $l_x = 0.40$, $l_z = 0.04$, and $\sigma = 1.5$. For the isotropic case (left plot), it is evident from this figure that although the performance of the primitive coarse model seems to agree reasonably well with the fine model (specifically on the breakthrough time), the coarse model with macro-diffusion does improve the overall prediction. Conversely, when the correlation length is larger along the main flow direction (right plot), where now the diffusion caused by heterogeneity is stronger, the coarse model with macro-diffusion gives a better prediction compared to the primitive model.

The performance of the coarse model with macro-diffusion in the case of nonlinear flux function is shown in Figure 5.15. Here we have used $f(S) = 5S^2/(5S^2 + (1-S)^2)$ and $\lambda(S) = 1$. Again, the plot on the left corresponds to isotropic permeability

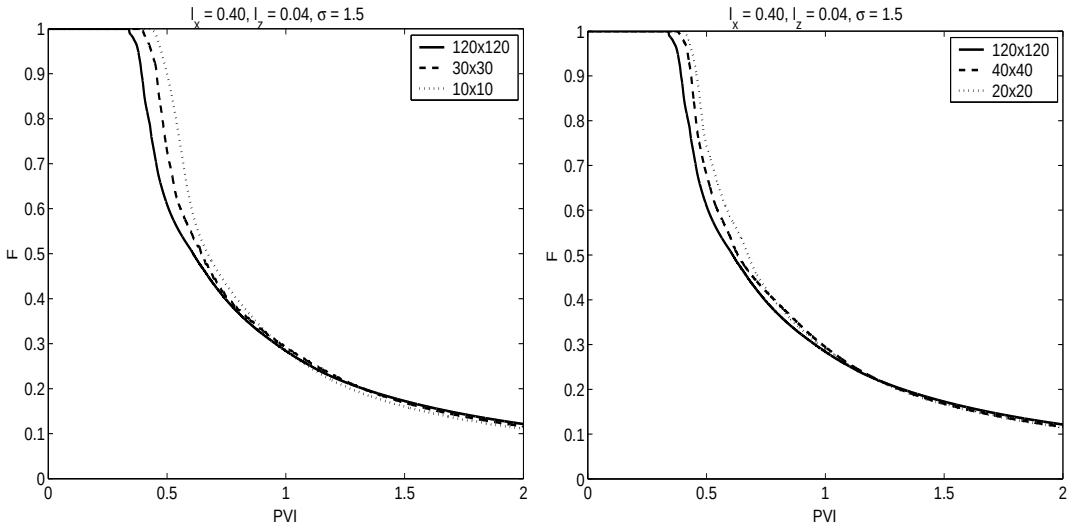

Fig. 5.10. Comparison of fractional flow of displaced fluid at the production edge for the case $l_x = 0.4$, $l_z = 0.04$, and $\sigma = 1.5$ with exponential variogram. Left plots are coarse model with 10×10 and 30×30 elements, right plots are coarse model with 20×20 and 40×40 elements.

field with $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1.5$, and the plot on the right corresponds to permeability field with $l_x = 0.40$, $l_z = 0.04$, and $\sigma = 1.5$. The significance of the macro-diffusion model in these two plots are obvious, in that the macro-diffusion model circumvents the primitive model in predicting the production on and shortly after the breakthrough. Also in this nonlinear flux function case, the model does not seem to be sensitive to the prescribed correlation structures.

To summarize, these computations reveal that the macro-diffusion resulting from the heterogeneity in the flow affects the coarse grid model, which may not be easily disregarded. Moreover, although solely based on the first order approximation, our proposed macro-diffusion model gives a reasonably well performance compared to the commonly used primitive model.

Finally, Figure 5.16 shows comparison of the average diffusion coefficient in the horizontal direction, where the average is taken over the domain. This comparison

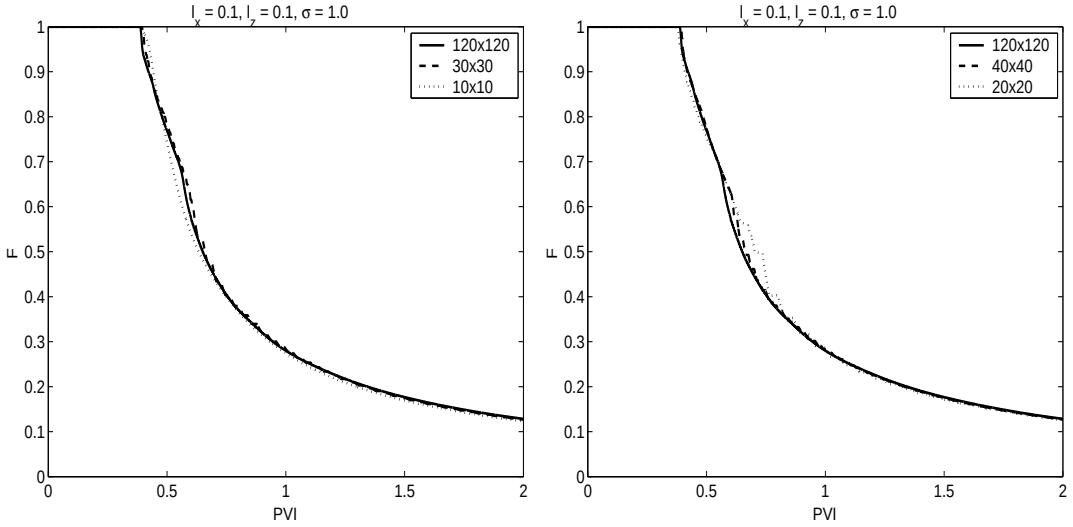

Fig. 5.11. Comparison of fractional flow of displaced fluid at the production edge for the case $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1.0$ with spherical variogram. Left plots are coarse model with 10×10 and 30×30 elements, right plots are coarse model with 20×20 and 40×40 elements.

shows that the more anisotropic the permeability then the larger the macro-diffusion coefficient is.

5.2. Two-Component Flow in Oil Reservoir Simulation

In addition to pumping water as the driving force as described in the previous section, a certain chemical substance is used that has an ability to perform some reactions with the trapped oil which in turn results in miscibility of the two-component of fluids. Consequently the reservoir fluids flow occurs in a single phase. In the following subsection we give an overview of the mathematical models for this technique.

5.2.1. Fine and Coarse Scale Models

We refer to [9] and [30] and for a detail derivation of the governing equations that follows. Let C denotes the concentration of injecting fluid component in the single

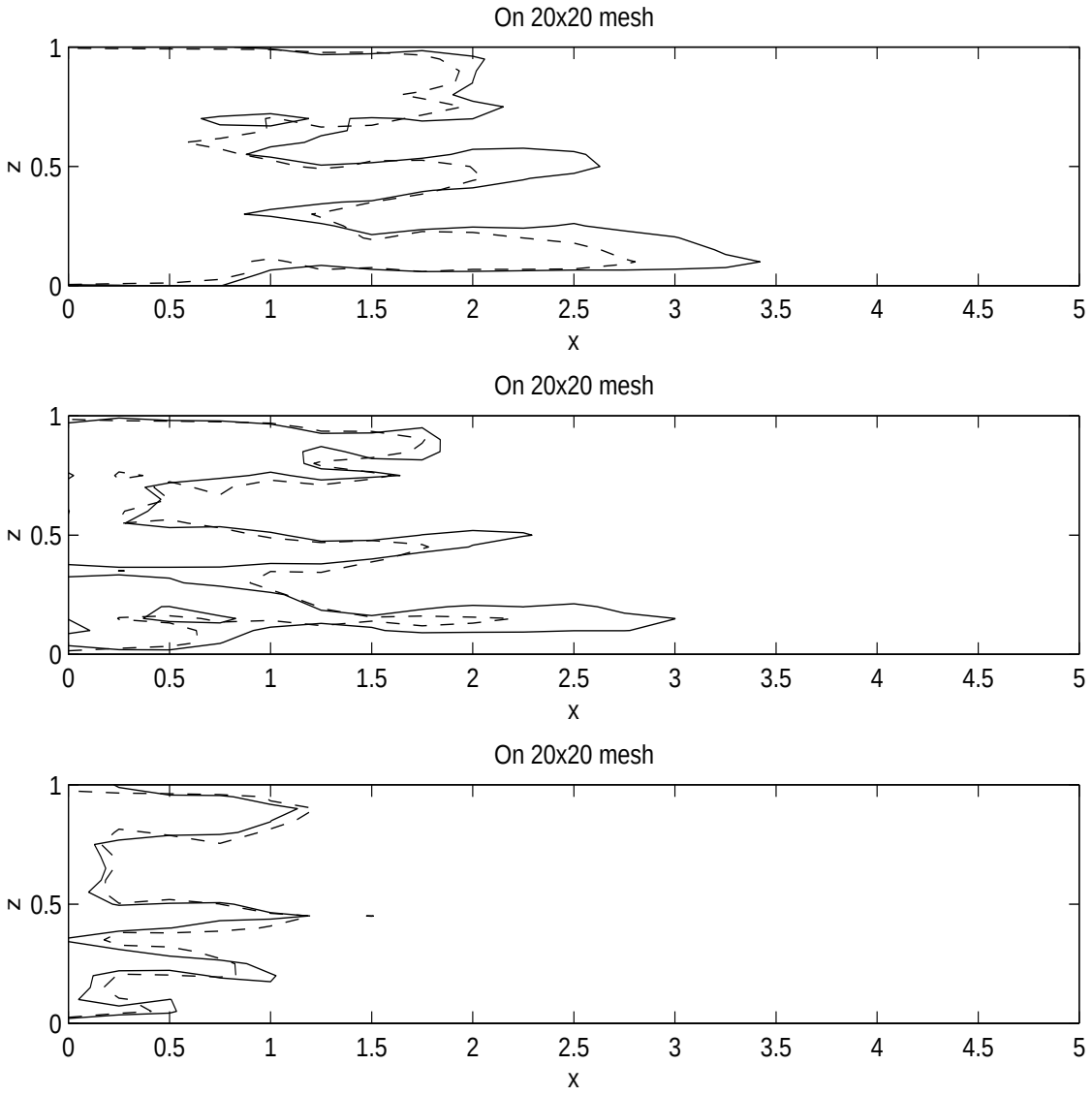

Fig. 5.12. Comparison of saturation contours at PVI = 0.15 for the case $l_x = 0.4$, $l_z = 0.04$, and $\sigma = 1.5$ with exponential variogram. The solid lines represent the fine grid saturation after averaging onto the coarse grid, while the dashed lines represent the coarse model with 20×20 elements. Upper plots are the contour of $\overline{S} = 0.10$, middle plots are the contour of $\overline{S} = 0.30$, and lower plots are the contour of $\overline{S} = 0.50$.

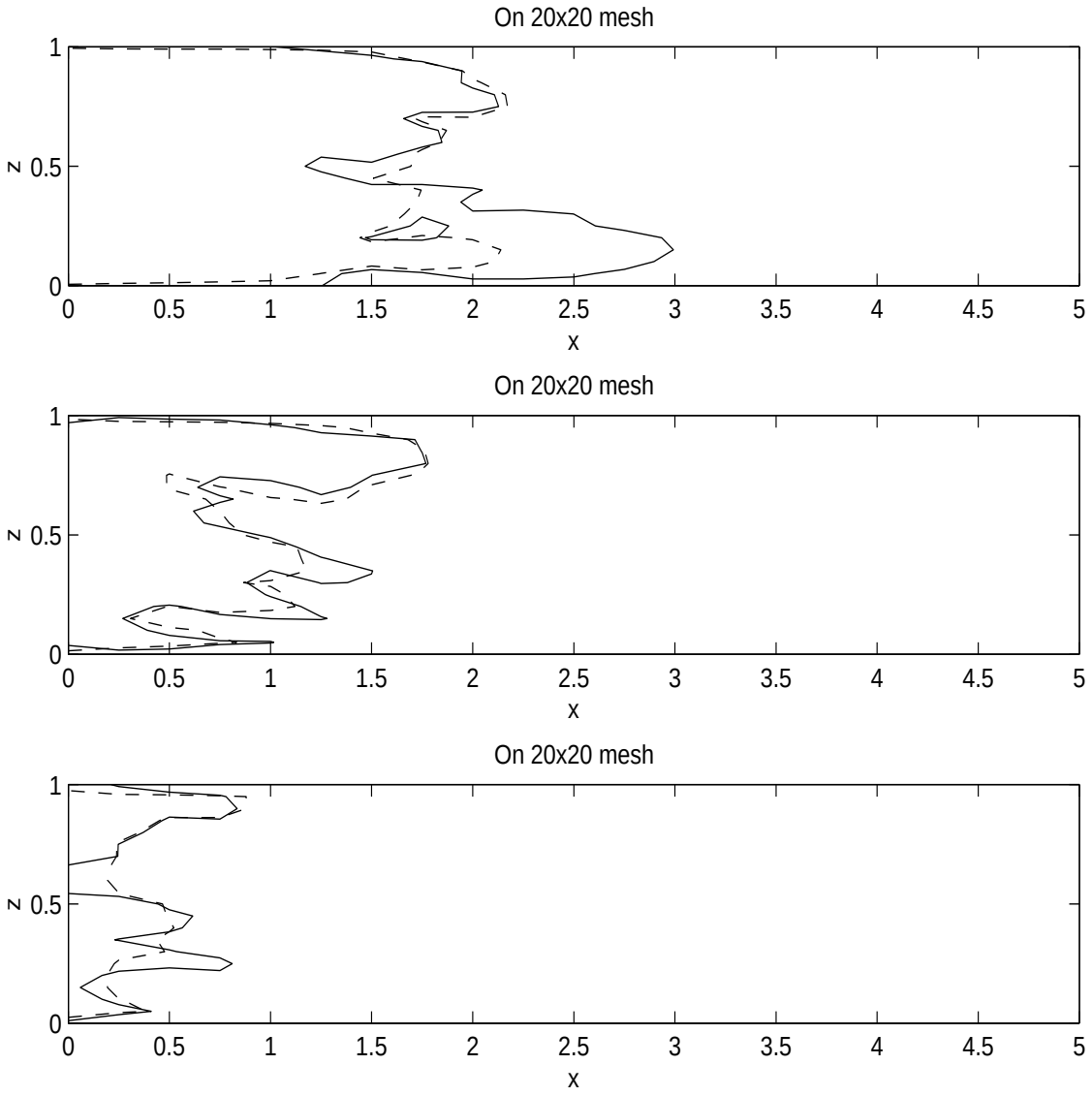

Fig. 5.13. Comparison of saturation contours at PVI = 0.15 for the case $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1.0$ with spherical variogram. The solid lines represent the fine grid saturation after averaging onto the coarse grid, while the dashed lines represent the coarse model with 20×20 elements. Upper plots are the contour of $\overline{S} = 0.10$, middle plots are the contour of $\overline{S} = 0.30$, and lower plots are the contour of $\overline{S} = 0.50$.

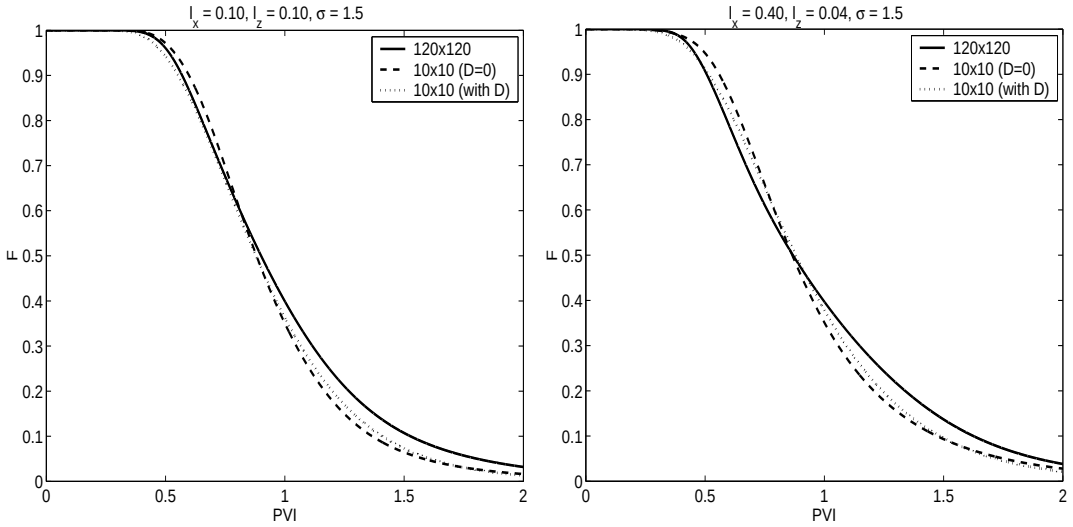

Fig. 5.14. Comparison of fractional flow of displaced fluid at the production edge. The flux function used is linear, $f(S) = S$. All coarse models are run on 10×10 elements. Plot on the left corresponds to $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1.5$ with spherical variogram. Plot on the right corresponds to $l_x = 0.40$, $l_z = 0.04$, and $\sigma = 1.5$ with spherical variogram.

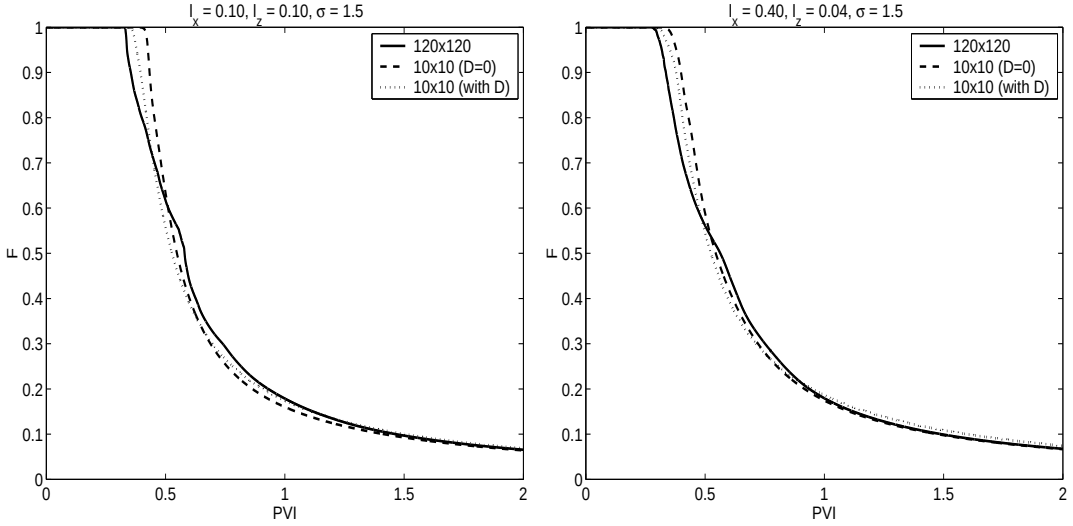

Fig. 5.15. Comparison of fractional flow of displaced fluid at the production edge. The flux function used is nonlinear, $f(S) = \frac{5S^2}{5S^2+(1-S)^2}$. All coarse models are run on 10×10 elements. Plot on the left corresponds to $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1.5$ with spherical variogram. Plot on the right corresponds to $l_x = 0.40$, $l_z = 0.04$, and $\sigma = 1.5$ with spherical variogram.

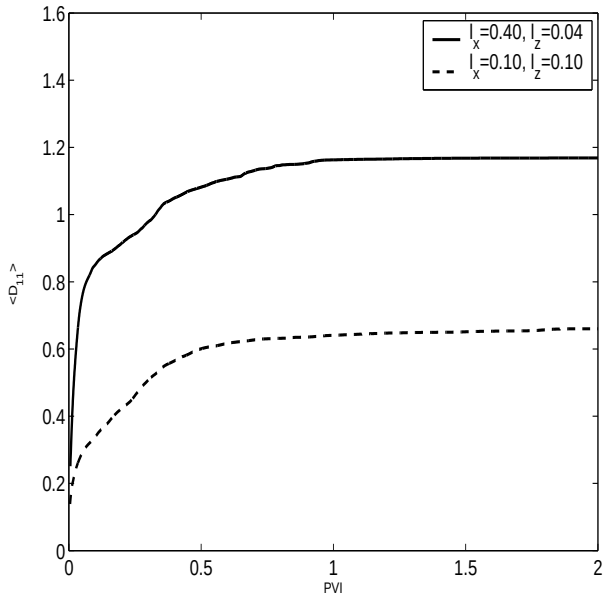

Fig. 5.16. The average of diffusion coefficient for different correlation length.

phase. The governing equations is derived from mass conservation of the fluid mixture, incompressibility condition, Darcy's Law, and mass conservation of the injecting fluid:

$$\begin{aligned} \nabla \cdot v &= q, \\ v &= -\frac{k}{\mu(C)} \nabla p, \end{aligned} \tag{5.33}$$

$$\frac{\partial C}{\partial t} + \nabla \cdot (vC) = \tilde{C}q, \tag{5.34}$$

where k is the absolute permeability tensor, $\tilde{C}q$ is some forcing function, and $\mu(C)$ is the viscosity of the fluid mixture that depends on the concentration. Typical dependency of this function is determined empirically using some mixing rule [30], such as

$$\mu(C) = \frac{\mu(0)}{\left(1 - C + M^{\frac{1}{4}} C\right)^4}, \tag{5.35}$$

where M is the mobility ratio between the resident and injected fluids, and $\mu(0)$ is the resident fluid viscosity. Another variation of the governing equation is by expanding

the divergence in (5.34) and substuting the appropriate term with (5.33), which result in the following transport equation for C:

$$\frac{\partial C}{\partial t} + v \cdot \nabla C = (\tilde{C} - C)q. \tag{5.36}$$

For our purpose we consider (5.33) and (5.36) as our fine model. Obviously, the pressure equation (5.33) and (5.36) are of the same form as (5.14) and (5.15) in section 5.1, hence the upscaled/coarse model for (5.33) employing the two-scale finite volume method is the same as in section 5.1. The primitive coarse model for (5.36) is the one using only the coarse scale velocity to update the concentration field on the coarse grid, i.e.,

$$\frac{\partial \overline{C}}{\partial t} + \overline{v} \cdot \nabla \overline{C} = (\tilde{C} - \overline{C})q, \tag{5.37}$$

so that no upscaling procedure is performed for the transport equation. As before the overbar variables denote the upscaled values on the coarse grid.

Next we describe an upscaling procedure for the transport equation via the macro-diffusion model. We will derive coarse scale equation for concentration C that resembles (5.27). Furthermore, since the velocity is time dependent (due to its concentration dependence), we propose a different approach to compute the macro-diffusion.

In similar way as in section 5.1, we use perturbation argument $C = \overline{C} + C'$ and $v = \overline{v} + v'$ to (5.36), and take an average of the resulting equation, which gives an upscaled version of the concentration equation:

$$\frac{\partial \overline{C}}{\partial t} + \overline{v} \cdot \nabla \overline{C} + \overline{v' \cdot \nabla C'} = (\tilde{C} - \overline{C})q, \tag{5.38}$$

where we have assumed that $\tilde{C}q$ is constant over the coarse control volume in which the average is taken. The term $\overline{v' \cdot \nabla C'}$ represents subgrid effects due to the hetero-

geneities of convection. This term can be modeled using the equation for C' that is derived by subtracting (5.38) from the fine scale equation (5.36)

$$\frac{\partial C'}{\partial t} + v \cdot \nabla C' + v' \cdot \nabla \overline{C} - \overline{v' \cdot \nabla C'} + qC' = 0. \tag{5.39}$$

Next we define the characteristics $dx(\tau)/d\tau = v$, thus with the notion of total derivative, we can write (5.39) as follows:

$$\frac{dC'}{dt} + qC' = -v' \cdot \nabla \overline{C} + \overline{v' \cdot \nabla C'}.$$

Multiplying both sides by e^{qt}, and integrating over $(0, t)$, we obtain the solution of this equation along the fine scale trajectory (x, t) such that $x(t) = x$:

$$C'(x,t) = e^{-qt} \int_0^t e^{q\tau} \left(-v'(x(\tau),\tau) \cdot \nabla \overline{C}(x(\tau),\tau) + \overline{v'(x(\tau),\tau) \cdot \nabla C'(x(\tau),\tau)} \right) d\tau.$$

The rest of the procedures follow those of section 5.1 such that the coarse scale concentration equation is written as

$$\frac{\partial \overline{C}(x,t)}{\partial t} + \overline{v} \cdot \nabla \overline{C}(x,t) - \nabla \cdot (D(x,t) \nabla \overline{C}(x,t)) = (\tilde{C} - \overline{C}(x,t))q, \tag{5.40}$$

where $D(x, t)$ is the macro-diffusive tensor, whose entries are written as

$$D_{ij}(x,t) = e^{-qt} \int_0^t e^{q\tau} \, \overline{v_i'(x,t) \, v_j'(x(\tau),\tau)} \, d\tau. \tag{5.41}$$

Furthermore, the dependency of D on the concentration $\overline{C}$ is obvious due to the fact that the velocity v depends on the concentration as governed by (5.33).

We now turn our attention to the procedure of computing D_{ij}. It is stated in the following proposition.

Proposition 5.1. *Let $L_j(x, t)$, $j = 1, 2$, be the trajectory length of the particle in*

x_j-direction that starts at point x (see Figure 5.8) computed as

$$L_j(x,t) = \int_0^t e^{q\tau}\, v_j'(x(\tau),\tau)\, d\tau.$$

Then

$$D_{ij}(x,t) \approx e^{-qt}\, \overline{v_i'(x,t)\, L_j(x,t)}.$$

Proof. The first term of the integrand in (5.41) is independent of τ, so we may take it out of the time integration:

$$D_{ij}(x,t) = e^{-qt}\, \overline{v_i'(x,t) \int_0^t e^{q\tau}\, v_j'(x(\tau),\tau)\, d\tau}. \tag{5.42}$$

We note that since the velocity depends on (x,t), so is the trajectory in (5.42), i.e., we have $x(\tau) = r(\tau|x,t)$ with $x(t) = r(t|x,t) = x$. Now let $\tau = t_p < t$. We assume that t_p is reasonably close to t. Then we may decompose the time integration in (5.42) as the sum of two integrations, namely,

$$\begin{aligned}\int_0^t e^{q\tau}\, v_j'(r(\tau|x,t),\tau)\, d\tau &= \int_0^{t_p} e^{q\tau}\, v_j'(r(\tau|x,t),\tau)\, d\tau + \int_{t_p}^t e^{q\tau}\, v_j'(r(\tau|x,t),\tau)\, d\tau \\ &= I_1 + I_2.\end{aligned}$$

Suppose we denote by y_p the particle location at time t_p. Then $r(\tau|x,t) = r(\tau|y_p,t_p)$, $0 \le \tau \le t_p$. Thus,

$$I_1 = \int_0^{t_p} e^{q\tau}\, v_j'(r(\tau|y_p,t_p),\tau)\, d\tau = L_j(y_p,t_p).$$

Furthermore, since we have assumed that t_p is reasonably close to t, the particle trajectory is still close to x, which gives

$$I_2 \approx e^{qt}\, (t - t_p)\, v_j'(x,t).$$

The proof of this proposition is completed by substituting these representations back

to (5.42), where now we have

$$L_j(x,t) = L_j(y_p,t_p) + e^{qt}\,(t-t_p)\,v_j'(x,t).$$

Thus the macro-diffusion coefficient may be computed as

$$D_{ij}(x,t) \approx e^{-qt}\,\overline{v_i'(x,t)\,L_j(y_p,t_p)} + (t-t_p)\,\overline{v_i'(x,t)\,v_j'(x,t)}.$$

This relation also gives a hint on how to numerically compute D_{ij}. We note that the fluctuation components v_i' are obtained by subtracting the average $\overline{v_i}$ from v_i, where v_i is constructed from the informations imbedded in the multiscale basis functions. Moreover, since $t_p < t$, $L_j(y_p,t_p)$ has been known. □

5.2.2. Numerical Results

In this section we present numerical results that give comparison between the fine and the coarse models presented in the previous subsections. The comparison will be made between the fine model, the primitive coarse model, and the coarse model with macro-diffusion that accounts for the subgrid effects on the coarse grid. Thus we can see possible improvement on the coarse model performance using this extension. As in section 5.1, the macro-diffusion coefficients are computed using the approximation of the fine scale velocity field by sampling the basis functions.

The case problem that we consider follows exactly the one in section 5.1 (cf. Figure 5.9), where the system is a cross section in the subsurface. As in section 5.1, he system length in the horizontal direction x (L_x) is greater than the formation thickness (L_z), with $L_x/L_z = 5$. Also the fine model uses 120×120 rectangular elements. The absolute permeability is set to be $\mathrm{diag}(k,k)$. In all the examples below we have used spherical variogram to generate the absolute permeability. We used the constitutive relation (5.35). As in section 5.1, we are interested in the fraction of oil

in the production edge F plotted against the dimensionless time PVI.

The first example is shown in Figure 5.17. The left plot uses isotropic field, i.e., $l_x = l_z = 0.10$, while the right plot uses anisotropic field of $l_x = 0.20$, $l_z = 0.02$. The solid line represents the fine model run on 120×120 elements, which as before, serves as a reference solution. The dashed line represents the primitive coarse model (D=0), while the dotted line represents the coarse model with macro-diffusion (with D). All coarse models are run on the 10×10 elements. For this example, we have used mobility ratio $M = 5$ and the variance of lognormal of permeability $\sigma = 1.5$. It is evident from this figure, that the coarse model with macro-diffusion made significant improvement compared to the primitive coarse model in both isotropic and anisotropic fields.

The second example is given in Figure 5.18. For this case we used the same parameters pertaining to the absolute permeability as in Figure 5.17. The only difference is we have used mobility ratio $M = 3$ in this example. Again, this example shows that the macro-diffusion model exhibit a better prediction than the primitive coarse model.

5.3. Infiltration in Saturated and Unsaturated Porous Media

We are interested in modeling the flow of water into a porous medium whose pore space is filled with air and some water. Several terminologies are in order. The fraction of the pore space volume to the porous medium total volume is called porosity, which is denoted by ϕ. The amount of water filling in the pore space of the medium is represented by the water saturation, S, i.e., it is defined as the fraction of the total pore space that is filled with water. In this connection, we say that the saturation varies between two values, namely, the residual water saturation, S_r, and the fully saturated value, S_s. These parameters are specific to different porous medium. Another

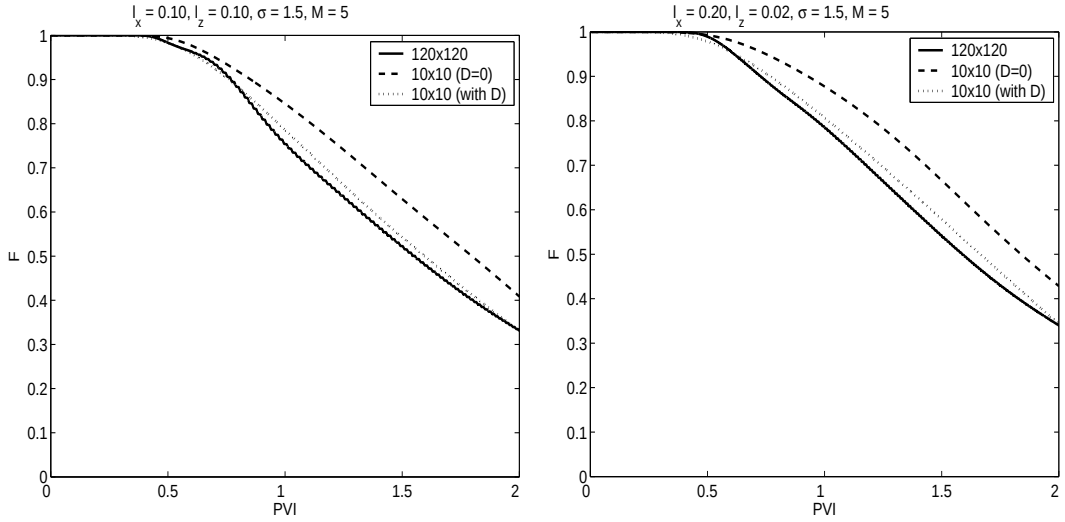

Fig. 5.17. Comparison of fractional flow of displaced fluid at the production edge for the two-component flow. All coarse models are run on 10×10 elements. Plot on the left corresponds to $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1.5$ with spherical variogram. Plot on the right corresponds to $l_x = 0.20$, $l_z = 0.02$, and $\sigma = 1.5$ with spherical variogram. In both plots viscosity ratio, $M = 5$.

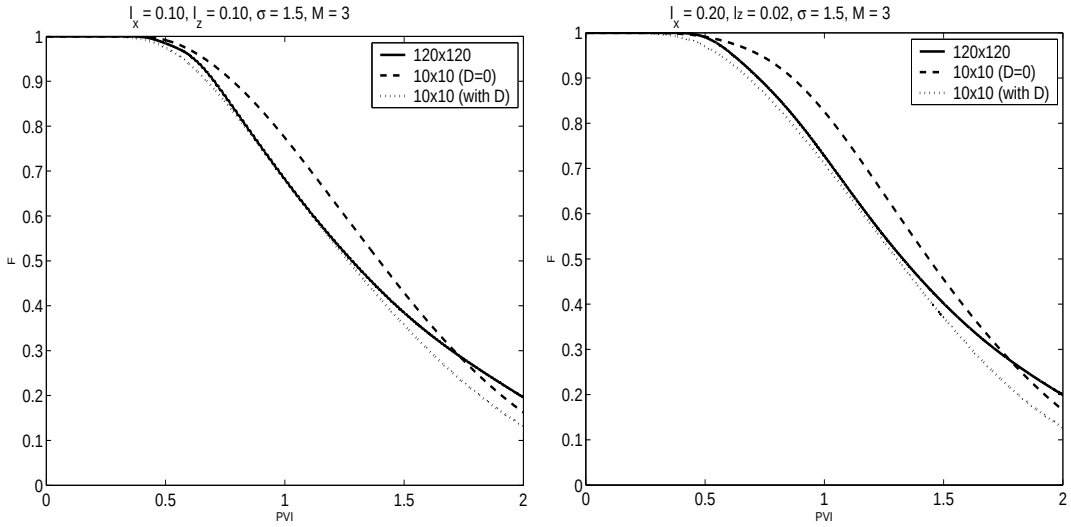

Fig. 5.18. Comparison of fractional flow of displaced fluid at the production edge for the two-component flow. All coarse models are run on 10×10 elements. Plot on the left corresponds to $l_x = 0.1$, $l_z = 0.1$, and $\sigma = 1.5$ with spherical variogram. Plot on the right corresponds to $l_x = 0.20$, $l_z = 0.02$, and $\sigma = 1.5$ with spherical variogram. In both plots viscosity ratio, $M = 3$.

measure of the amount of water in a porous medium that is closely related to water saturation is the so called volumetric water content, θ. The variable θ is defined as the fraction of porous medium total volume that is filled with water. In other words, the volumetric water content is related to the water saturation by $\theta = \phi S$.

The water flow into the porous medium is driven by the pressure gradient which is characterized by the empirical relation known as the Darcy's Law. Darcy's Law is basically a proportionality statement of the pressure gradient to the velocity vector. In the disciplines of hydrology and soil science, it is a common practice to use the term water pressure head which is defined as the amount of energy per unit weight of water. This normalization gives the pressure a dimension of length. Hence, the Darcy's Law is written as follows:

$$v = K(x, p)\,\nabla(p + x_3), \tag{5.43}$$

where v is the velocity vector, and K is called the unsaturated hydraulic conductivity tensor, which indicates the ability of the porous medium to transmit water under hydraulic gradients in unsaturated condition. Note that the variable x_3 represents the influence of gravity to the flow.

5.3.1. Richards' Equation

The following assumptions were proposed by Richards in [54] to give a simplified model for the fluid motion in unsaturated zone:

1. The porous medium and water are incompressible.

2. The temporal variation of the water saturation is significantly larger than the temporal variation of the water pressure.

3. Air phase is infinitely mobile so that the air pressure remains constant, in this

case it is atmospheric pressure which equals zero.

4. Neglect the source/sink terms.

The equation is written as follows:

$$\frac{\partial\theta(p)}{\partial t} - \nabla\cdot(K(x,p)\nabla(p+x_3)) = 0 \quad \text{in}\,\Omega, \tag{5.44}$$

where Ω is a bounded domain representing the porous medium.

Constitutive relations between θ and p, and between K and h are developed appropriately, which consequently gives nonlinearity behavior in (5.44). The relation between the water content and pressure head is referred to as moisture retention function. The equation written in (5.44) is called the *coupled-form* of Richards' Equation. In many other literatures this equation is also called the mixed form of Richards' Equation, due to the fact that there are two variables involved in it, namely, the water content θ and the pressure head p

Moreover, taking the advantage of the differentiability of the soil retention function, one may rewrite (5.44) as follows:

$$C(p)\frac{\partial p}{\partial t} - \nabla\cdot(K(x,p)\nabla(p+x_3)) = 0 \quad \text{in}\,\Omega, \tag{5.45}$$

where $C(p) = d\theta/dp$ is the specific moisture capacity. This version is referred to as the *head-form* (*h-form*) of Richards' Equation.

Another formulation of the Richards' Equation is based on the water content θ,

$$\frac{\partial\theta}{\partial t} - \nabla\cdot(D(x,\theta)\nabla\theta) - \frac{\partial K}{\partial x_3} = 0 \quad \text{in}\,\Omega, \tag{5.46}$$

where $D(\theta) = K(\theta)/(d\theta/dp)$ defines the diffusivity. This form is called the θ-*form* of Richards' Equation.

Richards' Equation is categorized as a nonlinear parabolic partial differential equation. There have been a great deal of efforts and investigations dedicated to Richards' Equation. It ranges from analyses of its mathematical properties, existence of solution, analytical and semi-analytical solutions with several restrictive conditions, to its numerical approximations along with the proposed algorithms. Richards' Equation enjoys the property of obeying the maximum principle [2], which is desirable for those who seek its numerical approximation.

Most of the earlier studies of existence and uniqueness of Richards' equation solution were implemented by assuming that the hydraulic conductivity is a power of the water content θ. Gilding and Peletier [39], for example, proposed some criteria for the weak solution of a one dimensional problem of (5.46), and showed its existence and uniqueness. The physical interpretation of this weak solution behaviors were investigated by Gilding in [38]. In particular, he showed the existence of the wetting front that serves as interface between adjacent wet and dry regions of a porous medium. The singularity of the wetting front were further studied and proved by Nakano in [51]. The regularity of the weak solution of multidimensional Richards' Equation was investigated by Aronson in [2], which he showed to be Holder continuous.

Several researchers have also tried to find analytical solutions of one dimensional Richards' Equation. Perhaps, the most classical results used and quoted in engineering fields are due to Gardner [36]. In his paper, he proposed an exponential and power relation of the hydraulic conductivity to the water saturation, such that a steady state solution may be obtained. Warrick and his associates [59, 60] studied analytical solutions of Richards' Equation for time-varying infiltration problems. Similar to Gardner, they also assumed exponential constitutive relations. Their solution takes the form of the time integration of the well known error function. Analytical solutions of problems in layered soils were examined in [56]. In this paper, Srivastava

et al. used the exponential constitutive relations to express the partial differential equation in terms of hydraulic conductivity K. Then, they employed the Laplace transform and inverse transform to obtain the solution.

The analytical solutions mentioned above are restrictive in nature and also limited to one dimensional problem. For more realistic cases, analytical solutions are in general not available. Consequently, numerical treatments are required to tackle the problems. The finite element, finite volume, and finite difference methods are most commonly used to generate the discretized equation. Results in [5, 6, 41, 47, 53] are several of the many works in numerical approximations of the equation. The most commonly used version of Richards' Equation is the *head-form* written in (5.45). Unfortunately, as found in [6, 53] this equation does not conserve the mass, and hence its numerical solution would suffer from this discrepancy.

The three versions of Richards' Equation written above have various advantages and disadvantages which in general depend upon the physical situations of the problems considered, and if used for numerical simulation also depend on the chosen numerical scheme. The θ-form for example, is by construction a conservative form, i.e., it follows the mass conservation law. However, this form only applies to the unsaturated zone, since for saturated condition the water content becomes constant and D approaches infinity. Furthermore, for multi-layered soils, θ cannot be guaranteed to be continuous across interfaces separating the layers. Thus, this form may be useful only for a homogeneous media.

On the other hand, due to the fact that the pressure head is continuous even for multi-layered soils, the head-form may be advantageous for heterogeneous soil condition. It is also applicable for both unsaturated and saturated media. Nevertheless, as described above the head-form does not maintain the global conservation of mass. Recently, Rathfelder et al. [53] proposed a method to solve the head-form equation

that still maintains the global mass balance. The key to their method is the different way of evaluating the specific moisture capacity C, in which they have used the so called the standard chord slope approximation.

The coupled-form of Richards' Equation is also mass conserved. It is applicable to both saturated and unsaturated porous media. The authors of [6] proposed the so-called modified Picard iteration to solve this equation, and made a comparison with results from the h-form. They showed that the coupled-form can maintain the mass conservation throughout the time marching of the simulation. These advantages have attracted many researchers and engineers to use this version for various practical problems.

5.3.2. Constitutive Relations

As has been mentioned in the Introduction, the sources of nonlinearity of Richards' Equation comes from the moisture retention and relative hydraulic conductivity functions, $\theta(p)$ and $K(x, p)$, respectively. Reliable approximation of these relations are in general tedious to develop and thus also challenging. Field measurements or laboratory experiments to gather the parameters are relatively expensive, and furthermore, even if one can come up with such relations from these works, they will be somehow limited to the particular cases under consideration.

Perhaps the most widely used empirical constitutive relations for the moisture content and hydraulic conductivity is due to the work of van Genuchten [57]. He proposed a method of determining the functional relation of relative hydraulic conductivity to pressure head by using the field observation knowledge of the moisture retention. In turn, the procedure would require curve-fitting the proposed moisture retention function with the experimental/observational data to establish certain parameters inherent to the resulting hydraulic conductivity model.

In attempts to formulate analytical solution of Richards' Equation, several researchers have employed exponential hydraulic parameters model to linearize the equation and applied some mathematical transformation to obtain the solution (see for example [56], and [59]). It is noted that although this approach may be very restrictive, it may be used to verify many numerical models.

There are several widely known formulations of the constitutive relations, among which are (see also Figures 5.19, 5.20, and 5.21):

1. **Haverkamp model** [41]:

$$\theta(p) = \frac{\alpha\,(\theta_s - \theta_r)}{\alpha + |p|^{\beta}} + \theta_r,$$

$$K(x,p) = K_s(x)\frac{A}{A + |p|^{\gamma}}$$

2. **van Genuchten model** [57]:

$$\theta(p) = \frac{\alpha\,(\theta_s - \theta_r)}{[1 + (\alpha|p|)^n]^m} + \theta_r,$$

$$K(x,p) = K_s(x)\frac{\left\{1 - (\alpha|p|)^{n-1}\,[1 + (\alpha|p|)^n]^{-m}\right\}^2}{[1 + (\alpha|p|)^n]^{m/2}}$$

3. **Exponential model** [59]:

$$\theta(p) = \theta_s\, e^{\beta p}$$

$$K(x,p) = K_s(x)\, e^{\alpha p}$$

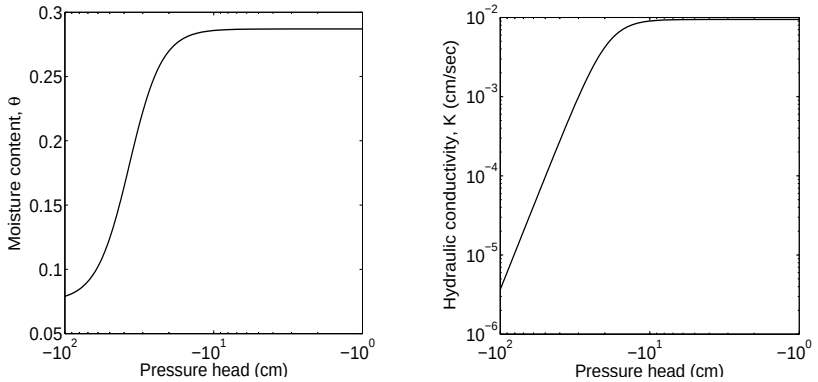

Fig. 5.19. Constitutive relations for Haverkamp model: (left) moisture content, (right) hydraulic conductivity

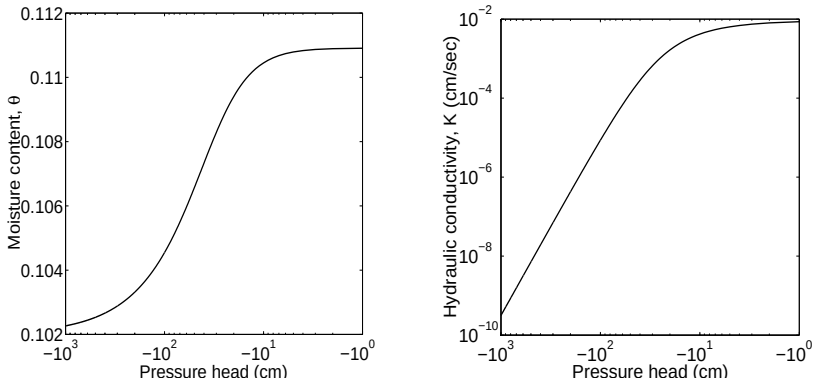

Fig. 5.20. Constitutive relations for van Genuchten model: (left) moisture content, (right) hydraulic conductivity

The variable K_s in the above models is also known as the saturated hydraulic conductivity. The figures indicate that the hydraulic conductivity has a broad range of values, which together with the functional forms presented above confirm the non-linear behavior of the process. It can also be seen that the water content and hydraulic conductivity approach zero as the pressure head goes to very large negative values.

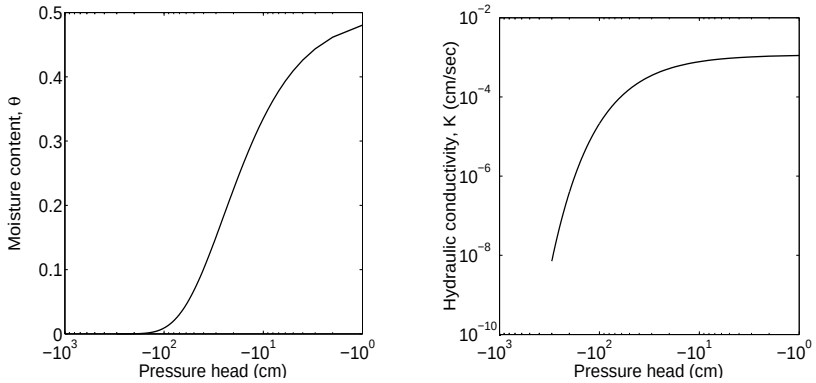

Fig. 5.21. Constitutive relations for exponential model: (left) moisture content, (right) hydraulic conductivity

In other words, the Richards' Equation has tendency to degenerate in a very dry condition, i.e., condition with the large negative pressure.

5.3.3. Fine and Coarse Scale Models

In this subsection we present a numerical homogenization for the coupled form of Richards' Equation (5.44). We will first describe the fine model used for comparison with the numerical homogenization. To simplify the presentation, we will neglect the gravity term in (5.44). By taking a backward Euler difference in time we have

$$\theta(p^n) - \theta(p^{n-1}) - \Delta t \, \nabla \cdot (K(x, p^n) \nabla p^n) = 0, \tag{5.47}$$

where the superscript n denotes the value of p computed at time t_n, and Δt is the time step. Obviously, for each time step n, we need to solve a nonlinear differential equation in p^n. For the fine model, we employ a procedure proposed by [6]. The idea is to linearize the equation in θ and K and solve the resulting equation iteratively. For simplicity of notation we denote by u the pressure that we want to solve in a time step n, i.e., $u = p^n$. Let us further denote by u^m the iterate of u at the iteration level

m. The first order Taylor expansion of θ may be written as

$$\theta(u^m) \approx \theta(u^{m-1}) + C(u^{m-1})\, r^m, \tag{5.48}$$

where $r^m = u^m - u^{m-1}$, and $C(u^{m-1})$ is the value of $d\theta/dp$ evaluated at u^{m-1}. By applying all these representations to (5.47) we have the following partial differential equation written in terms of r^m:

$$C(u^{m-1})\, r^m - \Delta t\, \nabla \cdot (K(x, u^{m-1}) \nabla r^m) = R^{m-1}, \quad m = 1, 2, 3, \cdot, \tag{5.49}$$

where

$$R^{m-1} = -(\theta(u^{m-1}) - \theta(p^{n-1})) + \Delta t\, \nabla \cdot (K(x, u^{m-1}) \nabla u^{m-1}). \tag{5.50}$$

The partial differential equation in (5.49) governs the residual of the solution at each iteration m. As the iteration converges in some fashion, we will have r^m vanishes and obtain the corresponding solution. Again, we note that this nonlinear iteration is done for each time step n. The preceeding description constitutes the fine model that we use to solve Richards' Equation (5.44).

We now turn our attention its numerical homogenization. As in the fine model, we are interested to numerically homogenize the Richards' equation after taking backward difference in time, i.e., the one written in (5.47). Thus for simplicity we designate as before u the solution p^n. Using the terminology in Chapter II, the MsFVEM for (5.44) is to find $u^h \in X^h$ such that

$$\int_{V_z} (\theta(\eta^h) - \theta^{n-1})\, dx - \Delta t \int_{\partial V_z} K(x, \eta^h)\, \nabla v_\epsilon \cdot n\, dl = 0 \quad \forall z \in Z_h^0, \tag{5.51}$$

where θ^{n-1} is the value of $\theta(\eta^h)$ evaluated at time step $n-1$, and $v_\epsilon \in V_\epsilon^h$ is a function

that satisfies the boundary value problem:

$$\begin{aligned} -\nabla \cdot (K(x, \eta^h) \nabla v_\epsilon) = 0 \quad \text{in } K \in T_h, \\ v_\epsilon = u^h \quad \text{on } \partial K, \end{aligned} \tag{5.52}$$

with $\eta^h(x) = \sum_{K \in T_h} \Psi_K(x) \frac{1}{|K|} \int_K u^h \, dx$. In general, the resulting finite dimensional equation obtained from (5.51) can be solved by direct application of the inexact Newton algorithm described in section 4.5 of Chapter IV.

A particular case in the numerical homogenization of (5.47) is also considered. When the nonlinearity and heterogeneity of $K(x, p)$ is separable, i.e.,

$$K(x, p) = k_s(x) \, k_r(p),$$

then we may use the linearization procedure implemented in the fine model to derive our coarse model. By this separability, and since in the formulation we always take the piecewise constant function η^h in replacement of u^h, the corresponding V_ϵ^h is a linear space, i.e., we may construct a set of basis functions $\{\psi_z\}_{z \in Z_h^0}$ such that they satisfy

$$\begin{aligned} -\nabla \cdot (k_s(x) \nabla \psi_z) = 0 \quad \text{in } K \in T_h, \\ \psi_z = \phi_z \quad \text{on } \partial K, \end{aligned} \tag{5.53}$$

where ϕ_z is a piecewise linear function. We note that if u^h has discontinuity or sharp front region, then the multiscale basis functions need to be updated in that region. Now, we may formulate the finite dimensional problem corresponding to (5.47). We want to seek $u^h \in V_\epsilon^h$ with $u^h = \sum_{z \in Z_h^0} p_z \psi_z$ such that

$$\int_{V_z} (\theta(\eta^h) - \theta^{n-1}) \, dx - \Delta t \int_{\partial V_z} k_s(x) \, k_r(\eta^h) \, \nabla u^h \cdot \vec{n} \, dl = 0, \tag{5.54}$$

for every control volume $V_z \subset \Omega$. To this equation we can directly apply the lineariza-

tion procedure described in the fine model (see (5.49)). Let us here denote

$$r^m = u^{h,m} - u^{h,m-1}, \quad m = 1, 2, 3, \cdot,$$

where $u^{h,m}$ is the iterate of u^h at the iteration level m. Thus we want to find $r^m = \sum_{z \in Z_h^0} r_z^m \psi_z$ such that for $m = 1, 2, 3, \cdot$ until convergence

$$\int_{V_z} C(\eta^{h,m-1})\, r^m \, dx - \Delta t \int_{\partial V_z} k_s(x)\, k_r(\eta^{h,m-1})\, \nabla r^m \cdot \vec{n}\, dl = R^{h,m-1}, \tag{5.55}$$

with

$$R^{h,m-1} = -\int_{V_z} (\theta(\eta^{h,m-1}) - \theta^{n-1})\, dx + \Delta t \int_{\partial V_z} k_s(x)\, k_r(\eta^{h,m-1})\, \nabla u^{h,m-1} \cdot \vec{n}\, dl. \tag{5.56}$$

As before the superscript m at each of the function means that the corresponding functions are evaluated at iteration level m.

5.3.4. Numerical Results

We present several numerical experiments that demonstrates the ability of the coarse models presented in the previous subsections. As in other applications in this chapter, the coarse models are compared with the fine model solved on a fine mesh. We have employed a finite volume difference to solve (5.49). This solution serves as a reference for the proposed coarse models. The problems that we consider are typical water infiltration into an initially dry soil. The porous medium that we consider is a rectangle of size $L_x \times L_z$ (see Figure 5.22. The fine model uses 256×256 rectangular elements, while the coarse model uses 32×32 rectangular elements. Similar to the cases in the previous sections, we generate a realization of the random variables with prescribed variance σ that represents the heterogeneity in the equation. We have used a spherical variogram for this purpose along with the correlation lengths that determine whether the realization is isotropic or anisotropic. All examples uses

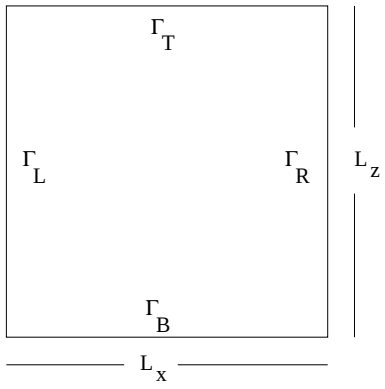

Fig. 5.22. Rectangular porous medium

$\sigma = 1.5$.

The first problem is a soil infiltration which was first analyzed by Haverkamp (cf. [6]). The porous medium dimension is $L_z = 40$ and $L_z = 40$. The boundary conditions are as follows: Γ_L and Γ_R are impermeable, while a Dirichlet conditions are imposed on Γ_B and Γ_T, namely $p_T = -21.7$ in Γ_T, and $p_B = -61.5$ in Γ_B. The initial pressure is $p_0 = -61.5$. The constitutive relations use **Haverkamp model** in section 5.3.2 [41]. The related parameters are as follows: $\alpha = 1.611 \times 10^6$, $\theta_s = 0.287$, $\theta_r = 0.075$, $\beta = 3.96$, $A = 1.175 \times 10^6$, and $\gamma = 4.74$. For this problem we assume that the nonlinearity and heterogeneity are separable, where the latter comes from $K_s(x)$ with $\overline{K_s} = 0.00944$. We assume that appropriate units for these parameters hold. There are two cases considered for this problem, namely, the isotropic heterogeneity with $l_x = l_z = 0.1$, and the anisotropic heterogeneity with $l_x = 0.01$ and $l_z = 0.20$. For the backward Euler scheme, we use $\Delta t = 10$. The comparison is shown in Figures 5.23 and 5.24, where the solutions are plotted at $t = 360$.

The second problem is a soil infiltration through a porous medium whose dimension is $L_x = 1$ and $L_z = 1$. The boundary conditions are as follows: Γ_L and Γ_R are impermeable. A Dirichlet conditions are imposed on Γ_B with $p_B = -10$.

The boundary Γ_T is divided into three parts. On the middle part a zero Dirichlet condition is imposed, and the rest are impermeable. The constitutive relations use **Exponential model** in section 5.3.2 with the following related parameters: $\beta = 0.01$, $\theta_s = 1$, $\overline{K_s} = 1$, and $\overline{\alpha} = 0.01$. The heterogeneity comes from $K_s(x)$ and $\alpha(x)$. It is obvious that for this problem the nonlinearity and heterogeneity are not separable. Again, isotropic and anisotropic heterogeneities are considered with $l_x = l_z = 0.1$, and $l_x = 0.20$, $l_z = 0.01$, respectively. For the backward Euler scheme, we use $\Delta t = 2$. The comparison is shown in Figures 5.25 and 5.26, where the solutions are plotted at $t = 10$.

We note that the problems that we have considered are vertical infiltration on the porous medium. Hence, it is also useful to compare the cross-sectional vertical velocity which will be plotted against the depth z. Here, the cross-sectional vertical velocity is obtained by taking an average over the horizontal direction (x-axis).

Figure 5.27 shows comparison of the cross-sectional vertical velocity for the Haverkamp model. The average is taken over all the horizontal span since the boundary condition on Γ_T (and also on Γ_B) is all Dirichlet condition. Both plots in this figure show a close agreement between the fine and coarse models.

For the Exponential model, as we have described above, there are three different segments for the boundary condition on Γ_T, i.e., a Neumann condition on the first and third part, and a Dirichlet condition on the second/middle part of Γ_T. Thus, we will compare the cross-sectional vertical velocity in each of these segments separately. Figures 5.28, 5.29, and 5.30 show the comparison for each of these segments, respectively. Contrary to the Haverkamp model, the vertical velocity seems to be more sensitive with respect to the anisotropy of the domain.

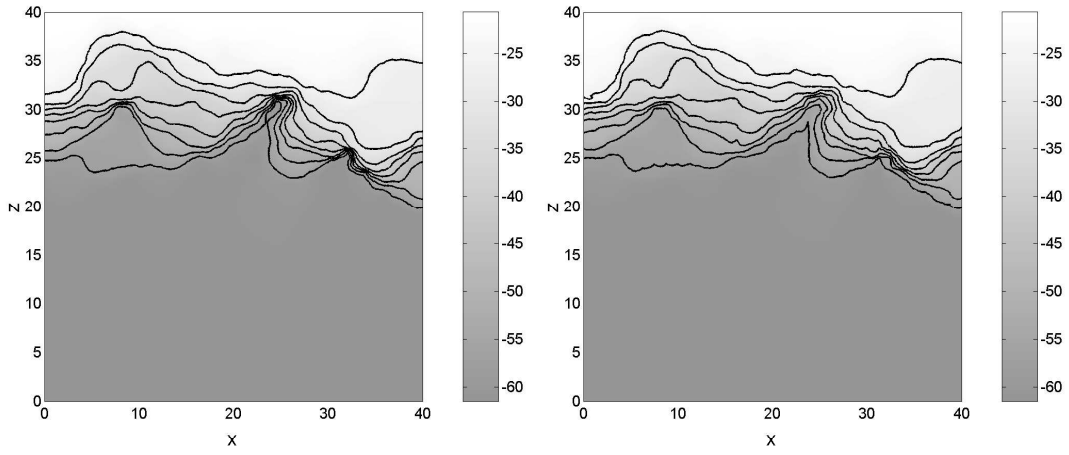

Fig. 5.23. Haverkamp model with isotropic heterogeneity. Comparison of water pressure between the fine model (left) and the coarse model (right).

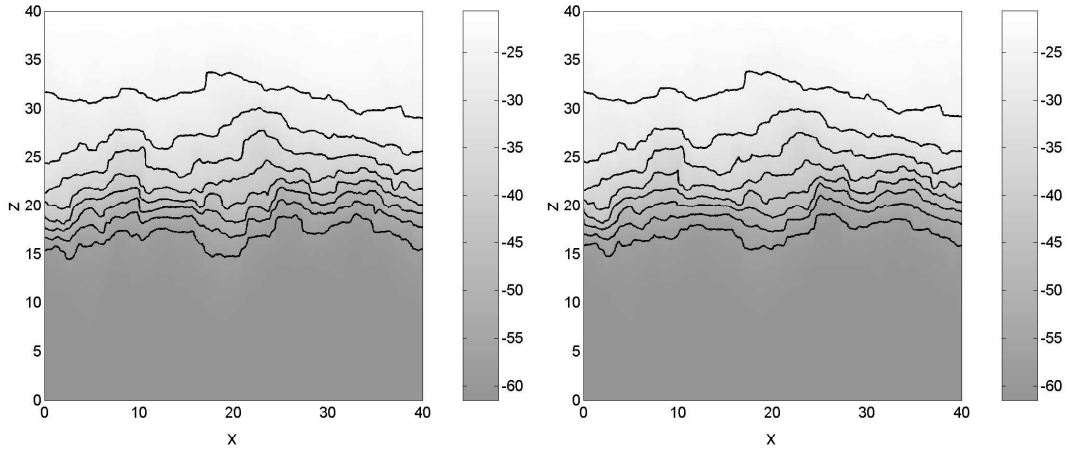

Fig. 5.24. Haverkamp model with anisotropic heterogeneity. Comparison of water pressure between the fine model (left) and the coarse model (right).

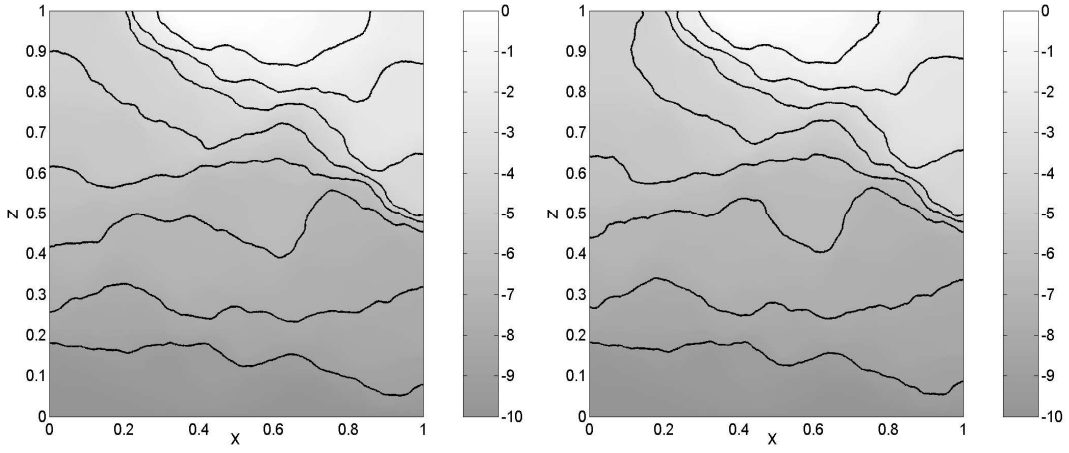

Fig. 5.25. Exponential model with isotropic heterogeneity. Comparison of water pressure between the fine model (left) and the coarse model (right).

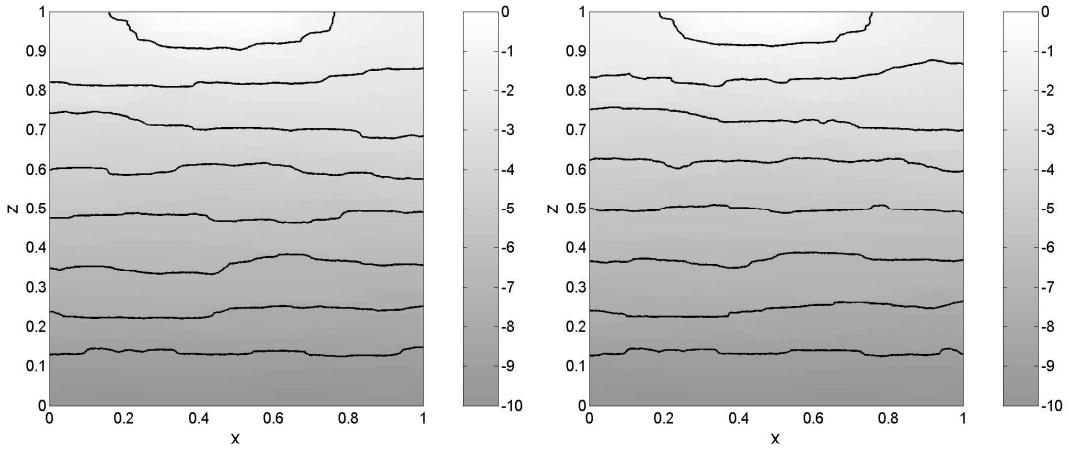

Fig. 5.26. Exponential model with anisotropic heterogeneity. Comparison of water pressure between the fine model (left) and the coarse model (right).

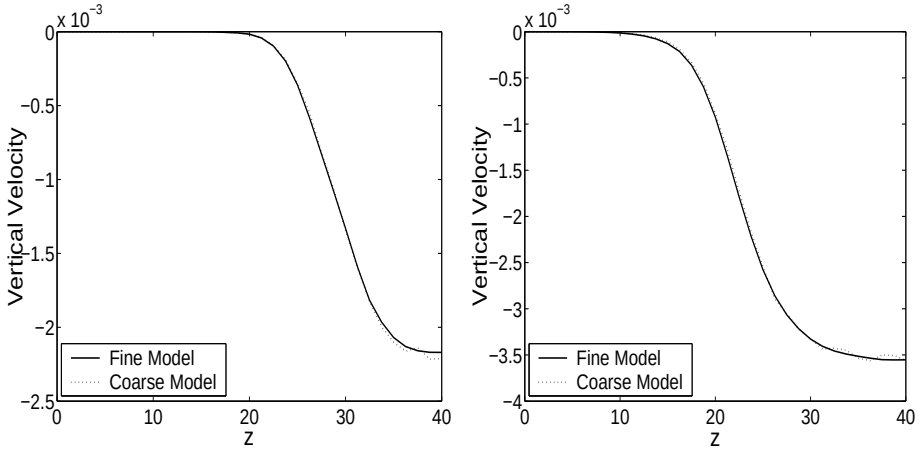

Fig. 5.27. Comparison of vertical velocity on the coarse grid for Haverkamp model: isotropic heterogeneity (left) and anisotropic heterogeneity (right).

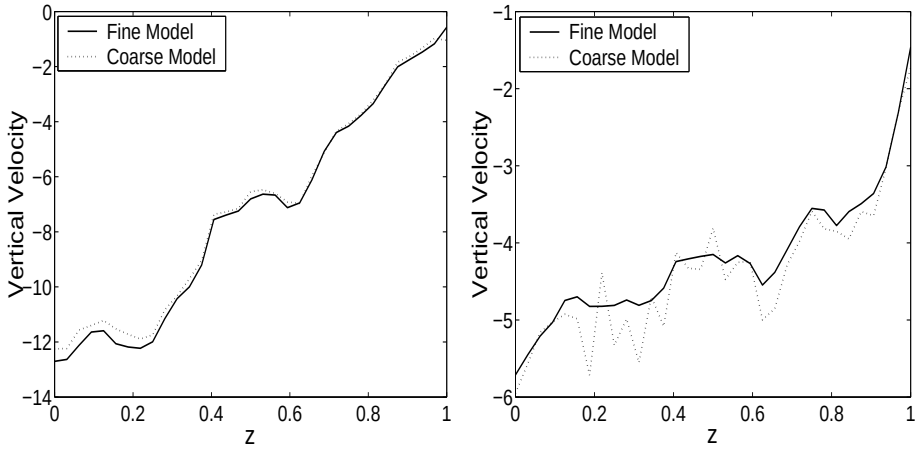

Fig. 5.28. Comparison of vertical velocity on the coarse grid for Exponential model: isotropic heterogeneity (left) and anisotropic heterogeneity (right). The average is taken over the first third of the domain.

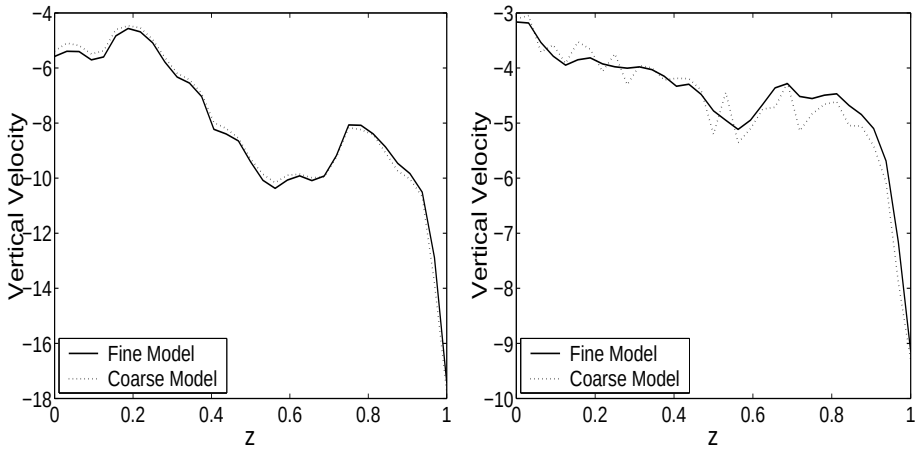

Fig. 5.29. Comparison of vertical velocity on the coarse grid for Exponential model: isotropic heterogeneity (left) and anisotropic heterogeneity (right). The average is taken over the second third of the domain.

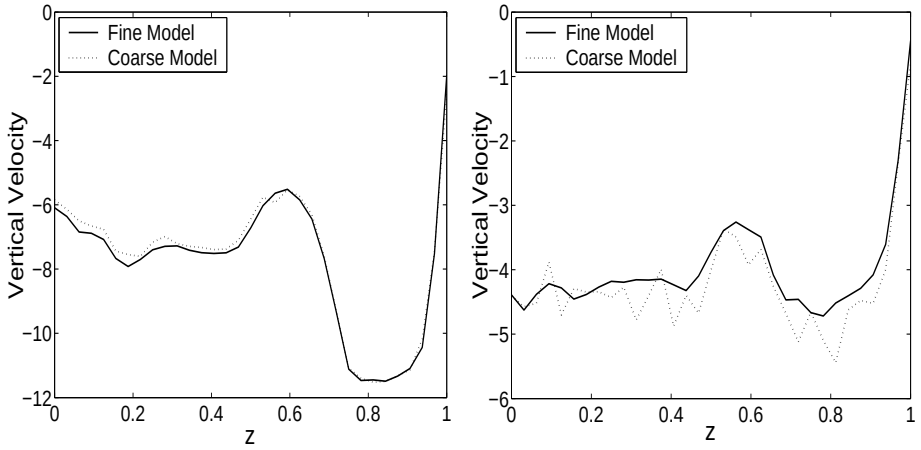

Fig. 5.30. Comparison of vertical velocity on the coarse grid for Exponential model: isotropic heterogeneity (left) and anisotropic heterogeneity (right). The average is taken over the last third of the domain.

CHAPTER VI

CONCLUSIONS

6.1. Summary

This dissertation concentrates on the development and analysis of multiscale methods for general elliptic boundary value problems. The formulation that has been presented in Chapter II is intended to cover nonlinearity in the coefficients. We have introduced a multiscale map E that serves as the quantification of the multiscale effect on the numerical solution. This multiscale map represents the fluctuation of the solution which is obtained by solving a leading order homogeneous elliptic equation in each element with a piecewise linear boundary conditions. To overcome resonance error inherent in imposing these boundary conditions, an oversampling technique has been used where the local problem associated with the multiscale map E is solved on a domain substantially larger than the coarse element and in turn use only the information pertaining to it. Then coarse scale problem is constructed through the conservation expression on each of the control volume. In this dissertation, this multiscale procedure is referred to as the multiscale finite volume element method (MsFVEM).

In Chapter III, we have investigated a convergence analysis of the linear MsFVEM, where we have the main assumption that the coefficient is periodic. A standard procedure that has been widely used in the analysis of finite volume method was used. The main idea is to view the finite volume element method as a perturbation of finite element method using certain interpolation operator. Analysis of the method uses substantially the existing finite element results and techniques. A Petrov-Galerkin formulation corresponding to the linear MsFVEM was used and

compared against the Petrov-Galerkin finite element formulation [62]. We conclude from the analysis presented in this chapter that the linear MsFVEM has the same convergence property as its finite element counterpart.

Chapter IV deals with the convergence analysis for the multiscale method for nonlinear elliptic problems. In addition to its periodicity, the elliptic coefficient is assumed to exhibit certain properties, i.e., polynomial growth, monotonicity with respect to the gradient of the solution, coercivity, and continuity. A clear distinction of whether the resulting operator is monotone or pseudomonotone has been made in the analysis. We have constructed a class of correctors corresponding to the multiscale method. The subsequent convergence of the method rely on approximation properties of this correctors. Particularly for monotone operators, we have been able to deduce a rate of convergence for the method.

Several applications of the multiscale methods to various problems of flow in porous media were presented in Chapter V. Three main applications that have been investigated are multiphase flow, multicomponent flow, and soil infiltration in saturated/unsaturated flow. In all of these applications, the MsFVEM is used to solve the pressure equation which can be elliptic or parabolic. Certain related variables, such as the velocity field can be recovered from the method. A macro-diffusion model was also presented that upscale transport equations. This macro-diffusion uses the small scale informations that may be gathered from the multiscale method.

6.2. Future Directions

Lastly, we would like to mention some future works. A more thorough analysis of the numerical homogenization for nonlinear elliptic problems needs to be pursued. This is especially crucial when the method is combined with the oversampling technique.

The analysis of numerical homogenization techniques within finite volume element framework is yet to be investigated.

For the applications, we need to make further assessment on the capability and limitation of the methods. For example, it is important to gain a clear understanding of the sensitivity of the methods with respect to the heterogeneity of the coefficients (quantified through the correlation structure) and its interaction with the nonlinearity. Furthermore, the two-phase immiscible flow model that we have used neglect features such as capillary pressure and gravity. Though these effects can be taken into account using our coarse scale methodologies without a great difficulty the detail numerical study of the obtained coarse scale models is yet to be carried out. Inclusion of these effects would certainly produce more realistic predictions.

A more effective procedure for the macro-diffusion in the upscaling of the transport equation is possible. Besides employing explicit/implicit scheme for the macro-diffusion models, splitting operator procedures may be used. The splitting may be done between the convective term and the diffusive term, where the operator involving the convective term is solved explicitly and the operator involving the diffusive term is solved implicitly.

Applications of the upscaling methods to inverse problems, such as subsurface characterization, are also of interest. Coarse scale models have advantages in subsurface characterization because (1) the mathematical inversion of flow equations is not computationally intensive and (2) additional dynamic data, such as production and pressure transient data, are responding to the spatial variation of larger-scale subsurface properties. Current practice is often limited to the use of the same form of the equations at the coarse level as those at the fine level. The adequate use of upscaled models at different coarse level will produce more accurate predictions.

REFERENCES

[1] T. Arbogast and S. L. Bryant, *Numerical subgrid upscaling for waterflood simulations*, Texas Institute for Computational and Applied Mathematics (TICAM) Report 01-23, http://www.ticam.utexas.edu/reports/2001/index.html.

[2] D. G. Aronson, *Regularity of flows in porous media: a survey*, in Nonlinear diffusion equations and their equilibrium states, I (Berkeley, CA, 1986), volume 12 of Math. Sci. Res. Inst. Publ., Springer, New York, 1988, pp. 35–49.

[3] 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(1994), pp. 945–981.

[4] J. W. Barker and S. Thibeau, *A critical review of the use of pseudo-relative permeabilities for upscaling*, SPE Res. Eng., 12(1997), pp. 138–143.

[5] E. T. Bouloutas, *Improved numerical methods for modeling flow and transport processes in partially saturated porous media*, Ph.D. thesis, Department of Civil Engineering, Massachusetts Institute of Technology, 1989.

[6] M. A. Celia, E. T. Bouloutas, and R. L. Zarba, *A general mass-conservative numerical solution for the unsaturated flow equation*, Water Resour. Res., 26(1990), pp. 1483–1496.

[7] P. Chatzipantelidis, *Finite volume methods for elliptic PDE's: a new approach*, M2AN Math. Model. Numer. Anal., 36(2002), pp. 307–324.

[8] P. Chatzipantelidis, R. Lazarov, and V. Thomée, *Error estimates for the finite volume element method for parabolic equations in convex polygonal domains*, ISC Technical Report Series, 3(2003) (to appear in J. Num. Meth. PDEs).

[9] G. Chavent and J. Jaffré, *Mathematical Models and Finite Elements for Reservoir Simulation*, Number 17 in Studies in Mathematics and its Applications, North-Holland, Amsterdam, 1986.

[10] Z. Chen, R. E. Ewing, and Z.-Ci Shi, eds., *Numerical treatment of multiphase flows in porous media*, volume 552 of Lecture Notes in Physics, Springer-Verlag, Berlin, 2000.

[11] Z. Chen and T. Y. Hou, *A mixed multiscale finite element method for elliptic problems with oscillating coefficients*, Math. Comp., 72(2003), pp. 541–576.

[12] S. Chou and Q. Li, *Error estimates in L^2, H^1 and L^∞ in covolume methods for elliptic and parabolic problems: a unified approach*, Math. Comp., 69(2000), pp. 103–120.

[13] S. S. Chow, *Finite element error estimates for nonlinear elliptic equations of monotone type*, Numer. Math., 54(1989), pp. 373–393.

[14] M.A. Christie, *Upscaling for reservoir simulation*, J. Pet. Tech., 1996, pp. 1004–1010.

[15] G. Dal Maso and A. Defranceschi, *Correctors for the homogenization of monotone operators*, Differential Integral Equations, 3(1990), pp. 1151–1166.

[16] C. V. Deutsch and A. G. Journel, *GSLIB: Geostatistical software library and user's guide*, 2nd edition, Oxford University Press, New York, 1998.

[17] P. Drábek, A. Kufner, and F. Nicolosi, *Quasilinear elliptic equations with degenerations and singularities*, volume 5 of de Gruyter Series in Nonlinear Analysis and Applications, Walter de Gruyter & Co., Berlin, 1997.

[18] L. J. Durlofsky, *Numerical calculation of equivalent grid block permeability tensors for heterogeneous porous media*, Water Resour. Res., 27(1991), pp. 699–708.

[19] L. J. Durlofsky, *Representation of grid block permeability in coarse scale models of randomly heterogeneous porous-media*, Water Resour. Res., 28(1992), pp. 1791–1800.

[20] 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, 2(1998), pp. 73–92.

[21] L. J. Durlofsky, R. A. Behrens, R. C. Jones, and A. Bernath, *Scale up of heterogeneous three dimensional reservoir descriptions*, SPE paper 30709, 1996.

[22] 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(1997), pp. 335–347.

[23] W. E and E. Engquist, *The heterogeneous multi-scale methods*, Comm. Math. Sci., 1, 2003.

[24] Y. Efendiev and A. Pankov, *Meyers type estimates for approximate solutions of nonlinear elliptic equations and their applications* (submitted to Num. Math).

[25] Y. R. Efendiev, *Exact upscaling of transport in porous media and its applications*, Institute for Mathematics and Applications (IMA) Preprint Series, 1724, October, 2000, http://www.ima.umn.edu/preprints/oct2000/oct2000.html.

[26] Y. R. Efendiev and L. J. Durlofsky, *Numerical modeling of subgrid heterogeneity in two phase flow simulations*, Water Resour. Res., 38(2002), pp. 1128.

[27] Y. R. Efendiev, L. J. Durlofsky, and S. H. Lee, *Modeling of subgrid effects in coarse scale simulations of transport in heterogeneous porous media*, Water Resour. Res., 36(2000), pp. 2031–2041.

[28] Y. R. Efendiev, T. Y. Hou, and X. H. Wu, *Convergence of a nonconforming multiscale finite element method*, SIAM J. Num. Anal., 37(2000), pp. 888–910.

[29] R. E. Ewing, T. Lin, and Y. Lin, *On the accuracy of the finite volume element method based on piecewise linear polynomials*, SIAM J. Numer. Anal., 39(2002), pp. 1865–1888.

[30] R. E. Ewing, ed., *The mathematics of reservoir simulation*, volume 1 of *Frontiers in Applied Mathematics*, Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 1983.

[31] R. E. Ewing, *Mathematical modeling and large-scale computing in energy and environmental research*, in The Merging of Disciplines: New Directions in Pure, Applied, and Computational Mathematics (Laramie, Wyo., 1985), Springer, New York, 1986, pp. 45–59.

[32] R. E. Ewing, *Upscaling of biological processes and multiphase flow in porous media*, in Fluid Flow and Transport in Porous Media: Mathematical and Numerical Treatment (South Hadley, MA, 2001), volume 295 of Contemp. Math., Amer. Math. Soc., Providence, RI, 2002, pp. 195–215.

[33] R. Eymard, T. Gallouët, and R. Herbin, *Finite volume methods*, in Handbook of Numerical Analysis, Vol. VII, North-Holland, Amsterdam, 2000, pp. 713–1020.

[34] N. Fusco and G. Moscariello, *On the homogenization of quasilinear divergence structure operators*, Ann. Mat. Pura Appl. (4), 146(1987), pp. 1–13.

[35] N. Fusco and G. Moscariello, *Further results on the homogenization of quasilinear operators*, Ricerche Mat., 35(1986), pp. 231–246.

[36] W. R. Gardner, *Some steady state solutions of the unsaturated moisture flow equation with application to evaporation from a water table*, Soil. Sci., 85(1958), pp. 228–232.

[37] D. Gilbarg and N. S. Trudinger, *Elliptic partial differential equations of second order*, volume 224 of Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences], 2nd edition, Springer-Verlag, Berlin, 1983.

[38] B. H. Gilding, *Properties of solutions of an equation in the theory of infiltration*, Arch. Rational Mech. Anal., 65(1997), pp. 203–225.

[39] B. H. Gilding and L. A. Peletier, *The Cauchy problem for an equation in the theory of infiltration*, Arch. Rational Mech. Anal., 61(1976), pp. 127–140.

[40] V. Ginting, *Analysis of two-scale finite volume element for elliptic problem* (to appear in Journal of Numerical Mathematics).

[41] R. Haverkamp, M. Vauclin, J. Touma, P. J. Wierenga, and G. Vachaud, *A comparison of numerical solution models for one-dimensional infiltration*, Soil. Sci. Soc. Am. J., 41(1977), pp. 285–294.

[42] 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(1997), pp. 169–189.

[43] T. Y. Hou, X. H. Wu, and Z. Cai, *Convergence of a multiscale finite element method for elliptic problems with rapidly oscillating coefficients*, Math. Comp.,

68(1999), pp. 913–943.

[44] T. Hughes, G. Feijoo, L. Mazzei, and J. Quincy, *The variational multiscale method - a paradigm for computational mechanics*, Comput. Methods Appl. Mech. Engrg, 166(1998), pp. 3–24.

[45] V. V. Jikov, S. M. Kozlov, and O. A. Oleinik, *Homogenization of differential operators and integral functionals*. Springer-Verlag, New York, 1994.

[46] P. Langlo and M. S. Espedal, *Macrodispersion for two-phase, immisible flow in porous media*, Advances in Water Resources, 17(1994), pp. 297–316.

[47] F. Lehmann and P. Ackerer, *Comparison of iterative methods for improved solutions of the fluid flow equation in partially saturated porous media*, Transport in Porous Media, 31(1998), pp. 275–291.

[48] R. Li, Z. Chen, and W. Wu, *Generalized difference methods for differential equations*, volume 226 of Monographs and Textbooks in Pure and Applied Mathematics, Numerical analysis of finite volume methods, Marcel Dekker Inc., New York, 2000.

[49] N. G. Meyers and A. Elcrat, *Some results on regularity for solutions of nonlinear elliptic systems and quasi-regular functions*, Duke Math. J., 42(1975), pp. 121–136.

[50] I. D. Mishev, *Finite volume element methods for non-definite problems*, Numer. Math., 83(1999), pp. 161–175.

[51] Y. Nakano, *Application of recent results in functional analysis to the problem of wetting fronts*, Water Resour. Res., 1980, pp. 314–318.

[52] A. Pankov, *G-convergence and homogenization of nonlinear partial differential operators*, Kluwer Academic Publishers, Dordrecht, 1997.

[53] K. Rathfelder and L. M. Abriola, *Mass conservative numerical solutions of the head-based Richards' equation*, Water Resour. Res., 1994, pp. 2579–2586.

[54] L. A. Richards, Capillary conduction of liquids through porous mediums. *Physics*, 1:318–333, 1931.

[55] I. V. Skrypnik, *Methods for analysis of nonlinear elliptic boundary value problems*, volume 139 of *Translations of Mathematical Monographs*, American Mathematical Society, Providence, RI, 1994, translated from the 1990 Russian original by Dan D. Pascali.

[56] R. Srivastava and T.-C. J. Yeh, *Analytical solutions for one-dimensional, transient, infiltration toward the water table in homogeneous and layered soils*, Water Resour. Res., 1991, pp. 753–762.

[57] M. Th. van Genuchten, *A closed-form equation for predicting the hydraulic conductivity of unsaturated soils*, Soil. Sci. Soc. Am. J., 44(1980), pp. 892–898.

[58] H. Wackernagle, *Multivariate geostatistics: an introduction with applications*, Springer, New York, 1998.

[59] A. W. Warrick, *Time-dependent linearized infiltration: III. strip and disc sources*, Soil. Sci. Soc. Am. J., 40(1976), pp. 639–643.

[60] A. W. Warrick, A. Islas, and D. O. Lomen, *An analytical solution to richards' equation for time-varying infiltration*, Water Resour. Res., 1991, pp. 736–766.

[61] X. H. Wen and J. J. Gomez-Hernandez, *Upscaling hydraulic conductivities in heterogeneous media: an overview*, Journal of Hydrology, 183(1996), pp. ix–xxxii.

[62] Y. Zhang, T. Y. Hou, and X. H. Wu, *Convergence of a nonconforming multiscale finite element method*, preprint, 1998.

[63] W. Zijl and A. Trykozko, *Numerical homogenization of two-phase flow in porous media*, Comput. Geosci., 6(2002), pp. 49–71.

APPENDIX A

SEVERAL INEQUALITIES

- **Young's inequality**

  Let a and b be two real numbers. Then

  $$|a\,b| \leq \frac{1}{p}\,|a|^p + \frac{1}{q}\,|b|^q,$$

  where $1 < p, q < \infty$ with $1/p + 1/q = 1$.

- **Holder's inequality**

  If $u \in L_p(\Omega)$ and $v \in L_q(\Omega)$, with $1 \leq p, q \leq \infty$, $1/p + 1/q = 1$ then

  $$\|u\,v\|_{L_1(\Omega)} \leq \|u\|_{L_p(\Omega)}\,\|v\|_{L_q(\Omega)}.$$

- **Jensen's inequality**

  Let φ be a convex function on $(-\infty, \infty)$ and f an integrable function on $[0, 1]$. Then

  $$\varphi\left(\int_0^1 f(x)\,dx\right) \leq \int_0^1 \varphi(f(x))\,dx.$$

VITA

Victor Eralingga Ginting was born in Medan, Indonesia, April 25, 1972. He earned a B.S. degree in civil engineering from the Institute of Technology Bandung, Indonesia in April 1995. Then he worked for a year as a research assistant in the Dynamics Laboratory of Inter University Center, Institute of Technology Bandung. In the Fall 1996 he became a graduate student specializing in ocean engineering, Department of Civil Engineering at Texas A&M University. He was awarded an M.S. degree in Summer 1998. After graduation, he joined the Department of Mathematics, Texas A&M University to pursue a doctoral program. In 2002, he served a summer internship at ChevronTexaco, San Ramon, CA. His permanent address is Jl. Sunggal No. 61, Medan - 20122, Sumatera Utara, Indonesia.