MODEL REDUCTION, BAYESIAN & DEEP LEARNING APPROACHES FOR FLOWS IN FRACTURED POROUS MEDIA

A Dissertation

by

SIU WUN CHEUNG

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

DOCTOR OF PHILOSOPHY

Chair of Committee,	Yalchin Efendiev
Co-Chair of Committee,	Tsz Shun Eric Chung
Committee Members,	Eduardo Gildin
	Raytcho Lazarov
	Jianxiin Zhou
Head of Department,	Sarah Witherspoon

May 2020

Major Subject: Mathematics

ABSTRACT

Numerical modelling of flow problems in fractured porous media has important applications in many engineering areas, such as unconventional reservoir simulation and nuclear waste disposal. Simulation of the flow problems in porous media is challenging as numerical discretization results in a very fine mesh for capturing the finest scales and high contrast of the physical properties. On the other hand, the effects of fractures are often modelled by multicontinuum models, resulting coupled systems of equations describing the interactive flow of different continua in heterogenous porous media. While multicontinuum models are widely adopted by different applications, for instance, naturally fractured porous media is modelled by dual porosity approach, shale gas production is modelled by the interactive flow of organic matter, inorganic matter and multiscale fractures in a heterogeneous media, and vuggy carbonate reservoir simulation is characterized by the complex interaction between matrix, fractures and vugs, numerical solutions on the fine grid are often prohibitively expensive in these complex multiscale problems.

Extensive research effort had been devoted to developing efficient methods for solving multiscale problems at reduced expense, for example, numerical homogenization approaches and multiscale methods, including Multiscale Finite Element Methods, Variational Multiscale Methods, Heterogeneous Multiscale Methods. The common goal of these methods is to construct numerical solvers on the coarse grid, which is typically much coarser than the fine grid which captures all the heterogeneities in the medium properties. In numerical homogenization approaches, effective properties are computed and the global problem is formulated and solved on the coarse grid. However, these approaches are limited to the cases when the medium properties possess scale separation. In this dissertation, we discuss and analyze novel multiscale model reduction techniques with different model problems arising from flows in porous media and numerical discretization techniques, which can be used for obtaining accurate coarse-scale approximations, even in the case of absence of scale separation.

On the other end, Bayesian approaches have been developed for forward and inverse problems

to address the uncertainties associated with the solution and the variations of the field parameters, and neural networks approaches are proposed for prediction of flow problems. In the dissertation, we also present methodologies of combining model reduction approaches with Bayesian approaches and deep learning approaches for efficient solution sampling and prediction for flow problems in porous media.

ACKNOWLEDGMENTS

First and foremost, I would like to express my utmost gratitude to my advisor, Professor Yalchin Efendiev, for his excellent guidance and continuous support throughout my Ph. D. study at Texas A&M Univeristy. He has always been inspiring to my research and career development and caring to me during my difficult times. It is my pleasure to work with him. I truly appreciate all his contributions of times and ideas to help me complete this dissertation.

Second, I am really grateful to my co-advisor Professor Eric T. Chung, who introduced me into the world of applied mathematics and scientific computing. He has always been patient and detailed in explaining his research ideas. I really appreciate his generous support for my visits to the Chinese University of Hong Kong during my Ph. D. study.

I would also like to thank Professor Eduardo Gildin, Professor Raytcho Lazarov and Professor Jianxin Zhou for serving as my committee members, guiding my research and providing me priceless advices in the past three years. Their insightful comments are crucial in the completion of this dissertation.

My Ph. D. study was much more enriched by my friends from the Department of Mathematics of Texas A&M University. I would like to thank everyone for their stimulating discussions on research and study. I would also like to thank all my friends for their support and encouragement during my difficult times.

Most importantly, my heartfelt gratitude goes to my family. Their unconditional love and caring provide me with incentives and confidence. Their understanding and support allow me to explore my interests and pursue further studies. Without their love and patience, I would not have been able to complete this dissertation. I warmly appreciate their generosity and encouragement.

CONTRIBUTORS AND FUNDING SOURCES

Contributors

This work was supervised by a dissertation committee consisting of Professor Yalchin Efendiev and Professor Eric T. Chung of the Department of Mathematics.

All other work conducted for the dissertation was completed by the student independently.

Funding Sources

Graduate study was supported by a fellowship from Texas A&M University.

NOMENCLATURE

Ω	Spatial domain
κ	Permeability
$\mathcal{T}^h$	Fine-scale partition
$\mathcal{T}^H$	Coarse-scale partition
h	Fine mesh size
H	Coarse mesh size
K	Coarse grid element
E	Coarse grid edge
ω	Coarse neighborhood
χ	Partition of unity
$K_{i,m}$	Coarse oversampled region
ϕ	Spectral basis function
ψ	Multiscale basis function

TABLE OF CONTENTS

Page

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

1. INTRODUCTION

1.1 Literature

Many engineering applications require numerical simulation in heterogeneous media. For example, Darcy flow equation in heterogeneous media is used to describe fluid flow in porous medium in reservoir simulation, and wave equation in heterogeneous media has been widely used for subsurface modeling. Physical properties in heterogeneous media possess multiple scales and high contrast, while the interactive effects between the microscope and the macroscopic scales have to be taken account in order to obtain accurate solutions. Examples of high-contrast physical properties are shown in Figure 1.1.

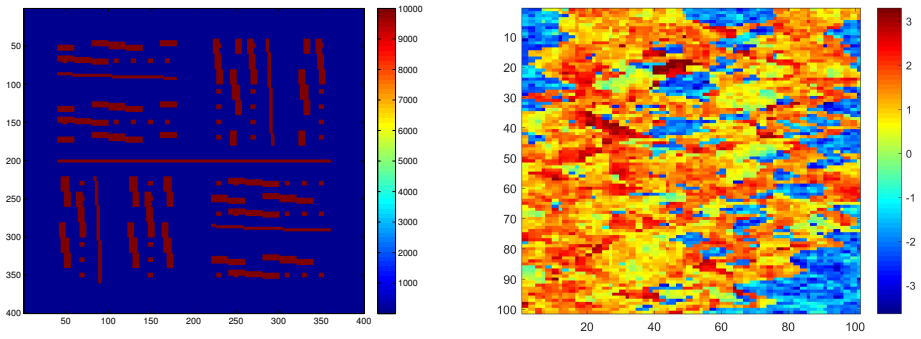

Figure 1.1: Examples of high-contrast permeability fields. Left: a channelized media. Right: SPE10 benchmark in logarithmic scale.

There has been extensive research effort devoted to develop and computational methods for flow simulations, resulting in a class of mature and well studied numerical methods, such as finite element methods [1, 2, 3, 4, 5, 6] and discontinuous Galerkin methods [7, 8, 9, 10, 11]. In order to resolve the multiscale features in numerical approximations, the computational mesh has to be sufficiently fine to capture the variations of the physical properties in the finest scale. As a result, direct numerical simulations on the fine grid are often prohibitively expensive in these

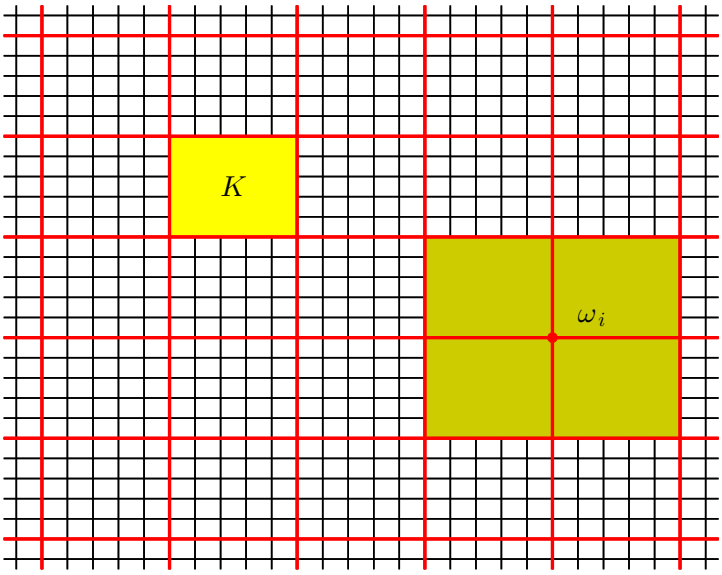

Figure 1.2: Illustration of fine grid, coarse grid and coarse neighborhood.

complex multiscale problems. To this end, extensive research effort had been devoted to developing efficient methods for solving multiscale problems at reduced expense, for example, numerical homogenization approaches [12, 13] and multiscale methods, including Multiscale Finite Element Methods (MsFEM) [14, 15, 16, 17, 18], Variational Multiscale Methods (VMS) [19, 20, 21, 22], Heterogeneous Multiscale Methods (HMM) [23, 24, 25] and and Generalized Multsicale Finite Element Methods (GMsFEM) [26, 27, 28, 29, 30, 31]. While a fine grid is used to capture all the heterogeneities in the medium properties, the objective of these methods is to construct numerical solvers on a coarse grid, which is typically much coarser than the fine grid. An illustration of the fine grid and the coarse grid and a coarse element are shown in Figure 1.2.

In numerical homogenization approaches, coarse-scale effective properties are computed and the global problem is formulated and solved on the coarse grid. However, these approaches are limited to the cases when the medium properties possess scale separation. On the other hand, multiscale methods construct of multiscale basis functions which are responsible for capturing the local oscillatory effects of the solution. Once the multiscale basis functions, coarse-scale equations are formulated. Moreover, fine-scale information can be recovered by the coarse-scale coefficients

and mutliscale basis functions. Meanwhile many existing mutliscale methods, such as MsFEM, VMS and HMM, construct one basis function per local coarse region to handle the effects of local heterogeneities. However, for more complex multiscale problems, each local coarse region contains several high-conductivity regions and multiple multiscale basis functions are required to represent the local solution space. In the cases where there is no scale separation, a systemic approach for adding degrees of freedom that capture the interactive effects between different scales is required.

GMsFEM is developed to allow systematic enrichment of the coarse-scale space with fine-scale information and identify the underlying low-dimensional local structures for solution representation. The main idea of GMsFEM is to extract local dominant modes by carefully designed local spectral problems in coarse regions, and the convergence of the GMsFEM is related to eigenvalue decay of local spectral problems. For a more detailed discussion on GMsFEM, we refer the readers to [32, 26, 33, 27, 34, 29, 35, 36, 37, 38] and the references therein. One of the main key feature of GMsFEM is a relation between the numbers of high-conductivity regions and multiscale basis functions used in local coarse neighborhoods for obtaining good approximations as supported by analysis. In general, an error estimate dependent on the coarse mesh size is non-trivial for multiscale model reduction methods, which is an emerging field in research [39, 40, 35, 41].

In many real-life applications, media properties may contain uncertainties and limited observation data about the flow profile may be available. It is important to take the effects of uncertainties and data into account to obtain quality solutions. Through using a Bayesian framework, one can include uncertainties in the media properties and compute the solution and the uncertainties associated with the solution and the variations of the field parameters. Bayesian approaches have also been widely used for forward and inverse problems [42, 43, 44, 45, 46, 47, 48, 49, 50, 51]. On the other side, there have been many works discovering the expressivity of deep neural nets theoretically [52, 53, 54, 55, 56, 57]. The universal approximation property of neural networks has been investigated in a lot of recent studies. It has been shown that deep networks are powerful and versatile in approximating wide classes of functions. Many researchers are inspired to take ad-

vantages of the multiple-layer structure of the deep neural networks in approximating complicated functions, and utilize it in the area of solving partial differential equations and model reductions. For instance, in [58], the authors propose a deep neural network to express the physical quantity of interest as a function of random input coefficients, and shows this approach can solve parametric PDE problems accurately and efficiently by some numerical tests. There is another work by E et. al [59], which aims to represent the trial functions in the Ritz method by deep neural networks (DNN). Then the DNN surrogate basis functions are utilized to solve the Poisson problem and eigenvalue problems. In [60], the authors build a connection between residual networks (ResNet) and the characteristic transport equation. Increasingly more research efforts have been devoted to build robust neural network techniques for approximations of multiscale problems related to reservoir simulation [61, 62, 63, 64, 65, 66, 67].

1.2 Organization of this dissertation

In this dissertation, we will study the development of a new class of local multiscale model reduction framework, namely Constraint Energy Minimizing Generalized Multiscale Finite Element Methods (CEM-GMsFEM), on flow problems in porous heterogeneous media. The new approach is motivated by GMsFEM and achieves spectral convergence. Through the design of local spectral problems, our method results in the minimal degree of freedom in representing high-contrast features. At the same time, the new multiscale method exhibits convergence on coarse mesh size independent of scales and contrast. Two formulations, namely the symmetric interior penalty discontinuous Galerkin (IPDG) model reduction for Darcy flow and coupled model reduction for multicontinuum flow problems, are considered. The advantages of the method is verified both theoretically and numerically. We establish a criterion for the oversampling size which is sufficient for linear coarse-mesh convergence independent of the contrast. Numerical results are presented to show the performance of the method for simulation on flow problem in high-contrast heterogeneous media.

In the later chapters of this dissertation, we will study the application of the model reduction techniques to solution sampling and prediction in the scenarios with limited observation data

and subject to uncertainties, with the use of probabilistic and machine learning tools. We propose Bayesian and neural network approaches for addressing the difficulties in these problems and highlight the advantages brought by the model reduction techniques, which justify the usefulness and importance for accurate and reliable reduced-order models.

2. CONSTRAINT ENERGY MINIMIZING GENERALIZED MULTISCALE DISCONTINUOUS GALERKIN METHOD

In this chapter, we present Constraint Energy Minimizing Generalized Multiscale Discontinuous Galerkin Method (CEM-GMsDGM). There are two key ingredients of the presented approach. The first main ingredient is the local spectral problems in each coarse block for identification of auxiliary basis functions. The low-energy dominant modes, which are eigenvectors corresponding to small eigenvalues of local spectral problems, are used as auxiliary basis functions for further construction. The auxiliary basis functions possess the information related to high conductivity channels and it suffices to use the same number of auxiliary basis functions as the number of channels in a coarse block. The second ingredient is the constraint energy minimization problems for definition of multiscale basis functions. Each of the auxiliary basis functions sets up an independent constraint and uniquely defines a corresponding multiscale basis function. The multiscale basis functions will then be used to span the multiscale space and used to solve the coarse-scale global problem in IPDG formulation. We remark that the local spectral problems and the constraint energy minimization problems are carefully designed and supported by our analysis. Thanks to the design of local spectral problems, the auxiliary space is of minimal dimension for representing high-contrast features and obtaining a contrast-independent convergence. Due to the fact that the dimensions of the auxiliary space and the multiscale space are identical, the multiscale space is of minimum dimension as well. In the construction of multiscale basis functions, the constraints are responsible for handling non-decaying components represented by the auxiliary basis functions in the high conductivity regions and achieving linear convergence in coarse mesh size. On the other hand, the multiscale basis functions are supported in oversampled coarse regions and allowed to have discontinuity on the coarse grid. Therefore, the IPDG bilinear form is also used to define the energy term in the constraint energy minimization problems. The advantages of the method is verified both theoretically and numerically. We analyze the method for solving Darcy flow problem and establish a criterion for the oversampling size which is sufficient for linear coarse-mesh

convergence independent of the contrast. Numerical results are presented to show the performance of the method for simulation on flow problem in high-contrast heterogeneous media.

The chapter is organized as follows. In Section 2.1, we will introduce the notions of grids, and essential discretization details such as DG finite element spaces and IPDG formulation on the coarse grid. The details of the proposed method will be presented in Section 2.2. The method will be analyzed in Section 2.3. Numerical results will be provided in Section 2.4.

2.1 Preliminaries

We consider the following high-contrast flow problem

$$-\operatorname{div}(\kappa \nabla u) = f \text{ in } \Omega, \tag{2.1}$$

subject to the homogeneous Dirichlet boundary condition $u = 0$ on $\partial\Omega$, where $\Omega \subset \mathbb{R}^2$ is the computational domain and f is a given source term. We assume that the permeability field κ is highly heterogeneous with very high contrast $\kappa_0 \leq \kappa \leq \kappa_1$.

Next, we introduce the notions of coarse and fine meshes. We start with a usual partition $\mathcal{T}^H$ of Ω into finite elements, which does not necessarily resolve any multiscale features. The partition $\mathcal{T}^H$ is called a coarse grid and a generic element K in the partition $\mathcal{T}^H$ is called a coarse element. Moreover, $H > 0$ is called the coarse mesh size. We let N_c be the number of coarse grid nodes and N be the number of coarse elements. We also denote the collection of all coarse grid edges by $\mathcal{E}^H$. We perform a refinement of $\mathcal{T}^H$ to obtain a fine grid $\mathcal{T}^h$, where $h > 0$ is called the fine mesh size. It is assumed that the fine grid is sufficiently fine to resolve the solution. An illustration of the fine grid and the coarse grid and a coarse element are shown in Figure 2.1.

We are now going to discuss the discontinuous Galerkin (DG) discretization and the interior penalty discontinuous Galerkin (IPDG) global formulation. For the i-th coarse block K_i, we denote the restriction of the Sobolev space $H^1(\Omega)$ on K_i by $V(K_i)$. We let $V_h(K_i)$ be the conforming

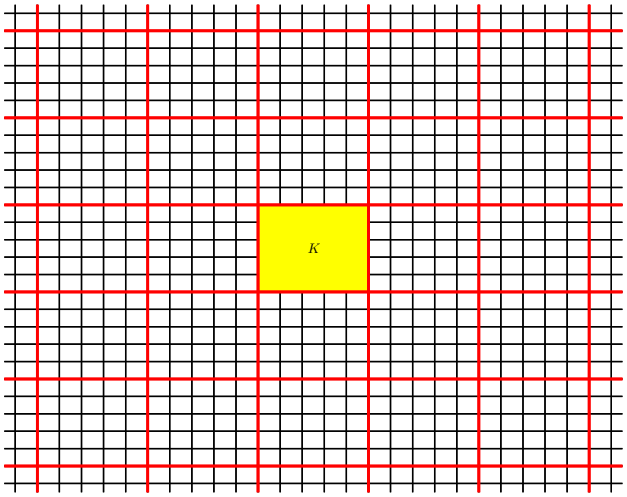

Figure 2.1: An illustration of the fine grid and the coarse grid and a coarse element.

bilinear elements defined on the fine grid $\mathcal{T}^h$ in K_i, i.e.

$$V_h(K_i) = \left\{ v \in V(K_i) : v|_\tau \in \mathbb{Q}^1(\tau) \text{ for all } \tau \in \mathcal{T}^h \text{ and } \tau \subset K_i \right\}, \tag{2.2}$$

where $\mathbb{Q}^1(\tau)$ stands for the bilinear element on the fine grid block τ. The DG approximation space is then given by the space of coarse-scale locally conforming piecewise bilinear fine-grid basis functions, namely

$$V_h = \oplus_{i=1}^N V_h(K_i). \tag{2.3}$$

We remark that functions in V_h are continuous within coarse blocks, but discontinuous across the coarse grid edges in general. The global formulation of IPDG method then reads: find $u_h \in V_h$ such that

$$a_{DG}(u_h, w) = \int_\Omega f w \, dx \text{ for all } w \in V_h, \tag{2.4}$$

where the bilinear form a_{DG} is defined by:

$$\begin{aligned} a_{DG}(v,w) = & \sum_{K\in\mathcal{T}^H}\int_K \kappa\nabla v\cdot\nabla w\,dx - \sum_{E\in\mathcal{E}^H}\int_E \{\kappa\nabla v\cdot n_E\}[\![w]\!]\,d\sigma \\ & - \sum_{E\in\mathcal{E}^H}\int_E \{\kappa\nabla w\cdot n_E\}[\![v]\!]\,d\sigma + \frac{\gamma}{h}\sum_{E\in\mathcal{E}^H}\int_E \overline{\kappa}[\![v]\!][\![w]\!]\,d\sigma, \end{aligned} \tag{2.5}$$

where $\gamma > 0$ is a penalty parameter and n_E is a fixed unit normal vector defined on the coarse edge $E \in \mathcal{E}^H$. Note that, in (2.5), the average and the jump operators are defined in the classical way. Specifically, consider an interior coarse edge $E \in \mathcal{E}^H$ and let K^+ and K^- be the two coarse grid blocks sharing the edge E, where the unit normal vector n_E is pointing from K^+ to K^-. For a piecewise smooth function G with respect to the coarse grid $\mathcal{T}^H$, we define

$$\begin{aligned} \{G\} &= \frac{1}{2}\left(G^+ + G^-\right), \\ [\![G]\!] &= G^+ - G^-, \end{aligned} \tag{2.6}$$

where $G^+ = G|_{K^+}$ and $G^- = G|_{K^-}$. Moreover, on the edge E, we define $\overline{\kappa} = \left(\kappa_{K^+} + \kappa_{K^-}\right)/2$, where $\kappa_{K^\pm}$ is the maximum value of κ over $K^\pm$. For a coarse edge E lying on the boundary $\partial\Omega$, we define $\{G\} = [\![G]\!] = G$, and $\kappa = \kappa_K$ on E, where we always assume that n_E is pointing outside of Ω.

First, we define the energy norm on the space V of coarse-grid piecewise smooth functions by

$$\|w\|_a^2 = a_{DG}(w,w) \text{ for all } w \in V. \tag{2.7}$$

We also define the DG-norm on V by

$$\|w\|_{DG}^2 = \sum_{K\in\mathcal{T}^H}\int_K \kappa|\nabla w|^2\,dx + \frac{\gamma}{h}\sum_{E\in\mathcal{E}^H}\int_E \overline{\kappa}[\![w]\!]^2\,d\sigma \text{ for all } w \in V. \tag{2.8}$$

The two norms are equivalent on the subspace of piecewise bi-cubic polynomials in V: there exists

$C_0 \geq 1$ such that

$$C_0^{-1}\|w\|_a \leq \|w\|_{DG} \leq C_0\|w\|_a. \tag{2.9}$$

The continuity and coercivity results of the bilinear form a_{DG} with respect to the DG-norm is ensured by a sufficiently large penalty parameter γ. While the method works well for general highly heterogeneous field κ, we assume κ is piecewise constant on the fine grid $\mathcal{T}^h$ for the sake of simplicity in our analysis presented in Section 2.3.

2.2 Method description

In this section, we will present the construction of the multiscale basis functions. First, we will use the concept of GMsFEM to construct our auxiliary multiscale basis functions on a generic coarse block K in the coarse grid. We consider $V_h(K_i)$ as the snapshot space in K_i and perform a dimension reduction through a spectral problem, which is to find a real number $\lambda_j^{(i)}$ and a function $\phi_j^{(i)} \in V_h(K_i)$ such that

$$a_i\left(\phi_j^{(i)}, w\right) = \lambda_j^{(i)} s_i\left(\phi_j^{(i)}, w\right) \text{ for all } w \in V_h(K_i), \tag{2.10}$$

where a_i is a symmetric non-negative definite bilinear operator and s_i is a symmetric positive definite bilinear operators defined on $V_h(K_i) \times V_h(K_i)$. We remark that the above problem is solved on the fine mesh in the actual computations. Based on our analysis, we can choose

$$\begin{aligned} a_i(v, w) &= \int_{K_i} \kappa \nabla v \cdot \nabla w \, dx, \\ s_i(v, w) &= \int_{K_i} \widetilde{\kappa} v w \, dx, \end{aligned} \tag{2.11}$$

where $\tilde{\kappa} = \sum_{j=1}^{N_c} \kappa |\nabla \chi_j^{ms}|^2$ and $\{\chi_j^{ms}\}_{j=1}^{N_c}$ are the standard multiscale finite element (MsFEM) basis functions. We let $\lambda_j^{(i)}$ be the eigenvalues of (2.10) arranged in ascending order in j, and use the first L_i eigenfunctions to construct our local auxiliary multiscale space

$$V_{aux}^{(i)} = \text{span}\{\phi_j^{(i)} : 1 \leq j \leq L_i\}. \tag{2.12}$$

The global auxiliary multiscale space V_{aux}^h is then defined as the sum of these local auxiliary multiscale spaces

$$V_{aux} = \oplus_{i=1}^{N} V_{aux}^{(i)}. \tag{2.13}$$

For the local auxiliary multiscale space $V_{aux}^{(i)}$, the bilinear form s_i in (2.11) defines an inner product with norm $\|v\|_{s(K_i)} = s\left(v,v\right)^{\frac{1}{2}}$. These local inner products and norms provide a natural definitions of inner product and norm for the global auxiliary multiscale space V_{aux}, which are defined by

$$\begin{aligned} s\left(v,w\right) &= \sum_{i=1}^{N} s_i\left(v,w\right) \text{ for all } v,w \in V_{aux}, \\ \|v\|_s &= s\left(v,v\right)^{\frac{1}{2}} \text{ for all } v \in V_{aux}. \end{aligned} \tag{2.14}$$

We note that $s\left(v,w\right)$ and $\|v\|_s$ are also an inner product and norm for the space V_h. Before we move on to discuss the construction of multiscale basis functions, we introduce some tools which will be used to describe our method and analyze the convergence. We first introduce the concept of ϕ-orthogonality. For $1 \leq i \leq N$ and $1 \leq j \leq L_i$, in coarse block K_i, given auxiliary basis function $\phi_j^{(i)} \in V_{aux}$, we say that $\psi \in V_h$ is $\phi_j^{(i)}$-orthogonal if

$$s\left(\psi, \phi_{j'}^{(i')}\right) = \delta_{i,i'}\delta_{j,j'} \text{ for all } 1 \leq j' \leq L_{i'} \text{ and } 1 \leq i' \leq N. \tag{2.15}$$

We also introduce a projection operator $\pi : V_h \to V_{aux}$ by $\pi = \sum_{i=1}^{N} \pi_i$, where

$$\pi_i(v) = \sum_{j=1}^{L_i} \frac{s_i\left(v, \phi_j^{(i)}\right)}{s_i\left(\phi_j^{(i)}, \phi_j^{(i)}\right)} \phi_j^{(i)} \text{ for all } v \in V_h, \text{ for all } i = 1, 2, \ldots, N. \tag{2.16}$$

Next, we construct our global multiscale basis functions in V_h. The global multiscale basis function $\psi_j^{(i)} \in V_h$ is defined as the solution of the following constrained energy minimization problem

$$\psi_j^{(i)} = \operatorname{argmin}\left\{a_{DG}\left(\psi,\psi\right) : \psi \in V_h \text{ is } \phi_j^{(i)}\text{-orthogonal}\right\}. \tag{2.17}$$

By introducing a Lagrange multiplier, the minimization problem (2.17) is equivalent to the following variational problem: find $\psi_j^{(i)} \in V_h$ and $\mu_j^{(i)} \in V_{aux}$ such that

$$\begin{aligned} a_{DG}\left(\psi_j^{(i)}, \psi\right) + s\left(\psi, \mu_j^{(i)}\right) &= 0 \text{ for all } \psi \in V_h, \\ s\left(\psi_j^{(i)} - \phi_j^{(i)}, \mu\right) &= 0 \text{ for all } \mu \in V_{aux}. \end{aligned} \tag{2.18}$$

Now we discuss the construction our localized multiscale basis functions. We first denote by $K_{i,m}$ an oversampled domain formed by enlarging the coarse grid block K_i by m coarse grid layers. An illustration of an oversampled domain is shown in Figure 2.2. We introduce the subspace $V_h\left(K_{i,m}\right)$, which contains restriction of fine-scale basis functions in V_h on the oversampled domain $K_{i,m}$. We also define $V_{h,0}\left(K_{i,m}\right) = V_h(K_{i,m}) \cap H_0^1(K_{i,m})$ by the subspace of functions in $V_h\left(K_{i,m}\right)$ vanishing on the boundary of the oversampled domain $K_{i,m}$. Motivated by the construction of our

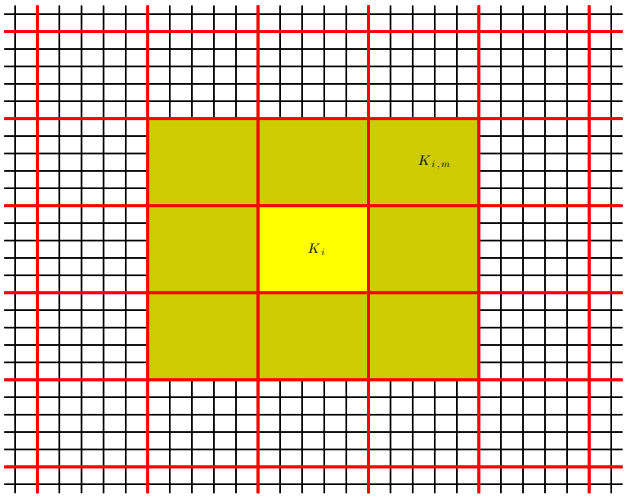

Figure 2.2: An illustration of an oversampled domain formed by enlarging K_i with 1 coarse grid layer.

global multiscale basis functions, the method for construction of the localized multiscale basis functions are as follows: The localized multiscale basis function $\psi_{j,ms}^{(i)} \in V_h\left(K_{i,m}\right)$ is defined as

the solution of the following constrained energy minimization problem

$$\psi_{j,ms}^{(i)} = \operatorname{argmin}\left\{a_{DG}\left(\psi,\psi\right) : \psi \in V_h\left(K_{i,m}\right) \text{ is } \phi_j^{(i)}\text{-orthogonal}\right\}. \tag{2.19}$$

Using the method of Lagrange multiplier, the minimization problem (2.19) is equivalent to the following variational problem: find $\psi_{j,ms}^{(i)} \in V_h\left(K_{i,m}\right)$ and $\mu_{j,ms}^{(i)} \in \oplus_{K_{i'}\subset K_{i,m}} V_{aux}^{(i')}$ such that

$$\begin{aligned} a_{DG}\left(\psi_{j,ms}^{(i)},\psi\right) + s\left(\psi,\mu_{j,ms}^{(i)}\right) &= 0 \text{ for all } \psi \in V_h\left(K_{i,m}\right), \\ s\left(\psi_{j,ms}^{(i)} - \phi_j^{(i)},\mu\right) &= 0 \text{ for all } \mu \in \oplus_{K_{i'}\subset K_{i,m}} V_{aux}^{(i')}. \end{aligned} \tag{2.20}$$

We use the localized multiscale basis functions to construct the multiscale DG finite element space, which is defined as

$$V_{ms} = \text{span}\{\psi_{j,ms}^{(i)} : 1 \le j \le L_i, 1 \le i \le N\}. \tag{2.21}$$

We remark that the multiscale finite element space V_{ms} is a subspace of V_h. After the multiscale DG finite element space is constructed, the multiscale solution u_{ms} is given by: find $u_{ms} \in V_{ms}$ such that

$$a_{DG}\left(u_{ms},w\right) = \int_\Omega f w\, dx \text{ for all } w \in V_{ms}. \tag{2.22}$$

2.3 Convergence analysis

In this section, we will analyze the proposed method. Besides the energy norm and the DG norm, we also define the s-norm on V by

$$\|w\|_s^2 = \sum_{K\in\mathcal{T}^H} \int_K \widetilde{\kappa}|w|^2\, dx. \tag{2.23}$$

Given a subdomain $\Omega' \subseteq \Omega$ formed by a union of coarse blocks $K \in \mathcal{T}^H$, we also define the local s-norm by

$$\|w\|_{s(\Omega')}^2 = \sum_{K\subseteq\Omega'} \int_K \widetilde{\kappa}|w|^2\, dx. \tag{2.24}$$

The flow of our analysis goes as follows. First, we prove the convergence using the global multiscale basis functions. With the global multiscale basis functions constructed, the global multiscale finite element space is defined by

$$V_{glo} = \text{span}\{\psi_j^{(i)} : 1 \leq j \leq L_i, 1 \leq i \leq N\}, \tag{2.25}$$

and an approximated solution $u_{glo} \in V_{glo}$ is given by

$$a_{DG}\left(u_{glo}, w\right) = \int_{\Omega} f w \, dx \text{ for all } w \in V_{glo}. \tag{2.26}$$

We remark that the construction of global multiscale basis functions motivates the construction of localized multiscale basis functions. The approximated solution u_{glo} will also be used in our convergence analysis. Next, we give an estimate of the difference between the global multiscale functions $\psi_j^{(i)}$ and the localized multiscale basis functions $\psi_{j,ms}^{(i)}$, in order to show that using the multiscale solution u_{ms} provide similar convergence results as the global solution u_{glo}. For this purpose, we denote the kernel of the projection operator π by $\widetilde{V}_h$. Then, for any $\psi_j^{(i)} \in V_{glo}$, we have

$$a_{DG}\left(\psi_j^{(i)}, w\right) = 0 \text{ for all } w \in \widetilde{V}_h, \tag{2.27}$$

which implies $\widetilde{V}_h \subseteq V_{glo}^{\perp}$, where $V_{glo}^{\perp}$ is the orthogonal complement of V_{glo} with respect to the inner product $a_{DG}\left(\cdot, \cdot\right)$. Moreover, since $\dim\left(V_{glo}\right) = \dim\left(V_{aux}\right)$, we have $\widetilde{V}_h = V_{glo}^{\perp}$ and $V_h = V_{glo} \oplus \widetilde{V}_h$.

The convergence analysis will start with the following lemma, which concerns about the convergence of the approximated solution by the global multiscale basis functions.

Lemma 2.3.1. *Let $u_h \in V_h$ be the solution of* (2.4) *and $u_{glo} \in V_{glo}$ be the solution of* (2.26) *with the global multiscale basis functions defined by the constrained energy minimization problem* (2.17)*. Then we have $u_h - u_{glo} \in \widetilde{V}_h$ and*

$$\|u_h - u_{glo}\|_a \leq \Lambda^{-\frac{1}{2}} \|\widetilde{\kappa}^{-\frac{1}{2}} f\|_{L^2(\Omega)}, \tag{2.28}$$

where

$$\Lambda = \min_{1\leq i\leq N} \lambda_{L_i+1}^{(i)}. \tag{2.29}$$

Moreover, if we replace the multiscale partition of unity $\{\chi_j^{ms}\}$ *by the bilinear partition of unity, we have*

$$\|u_h - u_{glo}\|_a \leq CH\Lambda^{-\frac{1}{2}} \|\kappa^{-\frac{1}{2}} f\|_{L^2(\Omega)}. \tag{2.30}$$

Proof. By the definitions of u_h in (2.4) and u_{glo} in (2.26), we have

$$\begin{aligned} a_{DG}(u_h, w) &= \int_\Omega fw\, dx \text{ for all } w \in V_h, \\ a_{DG}(u_{glo}, w) &= \int_\Omega fw\, dx \text{ for all } w \in V_{glo}. \end{aligned} \tag{2.31}$$

Since $V_{glo} \subseteq V_h$, this yields the Galerkin orthogonality property.

$$a_{DG}(u_h - u_{glo}, w) = 0 \text{ for all } w \in V_{glo}, \tag{2.32}$$

which implies $u_h - u_{glo} \in V_{glo}^{\perp} = \widetilde{V}_h$. In particular, if we take $w = u_{glo}$ in (2.32), together with (2.4), we have

$$\begin{aligned} \|u_h - u_{glo}\|_a^2 &= a_{DG}(u_h, u_h - u_{glo}) \\ &= (f, u_h - u_{glo})_{0,\Omega} \\ &\leq \|\widetilde{\kappa}^{-\frac{1}{2}} f\|_{L^2(\Omega)} \|u_h - u_{glo}\|_s. \end{aligned} \tag{2.33}$$

Since $u_h - u_{glo} \in \widetilde{V}_h$, we have $\pi(u_h - u_{glo}) = 0$. Furthermore, since K_i are disjoint, we have $\pi_i(u_h - u_{glo}) = 0$ for all $i = 1, 2, \ldots, N$. This implies

$$\begin{aligned} \|u_h - u_{glo}\|_s^2 &= \sum_{i=1}^N \|u_h - u_{glo}\|_{s(K_i)}^2 \\ &= \sum_{i=1}^N \|(I - \pi_i)(u_h - u_{glo})\|_{s(K_i)}^2 \end{aligned} \tag{2.34}$$

By the s_i-orthogonality of the eigenfunctions $\phi_j^{(i)}$, we have

$$\begin{aligned} \| \left(I - \pi_i\right) \left(u_h - u_{glo}\right) \|_{s(K_i)}^2 &\leq \left(\lambda_{L_i+1}^{(i)}\right)^{-1} a_i(u_h - u_{glo}, u_h - u_{glo}) \\ &\leq \Lambda^{-1} a_i(u_h - u_{glo}, u_h - u_{glo}). \end{aligned} \tag{2.35}$$

Therefore, we have

$$\begin{aligned} \|u_h - u_{glo}\|_s^2 &\leq \Lambda^{-1} \sum_{i=1}^{N} a_i(u_h - u_{glo}, u_h - u_{glo}) \\ &\leq \Lambda^{-1} \|u_h - u_{glo}\|_a^2. \end{aligned} \tag{2.36}$$

Using (2.33) and (2.36), we obtain our desired result. The second part of the result follows from the property $|\nabla \chi_j| = O\left(H^{-1}\right)$ of the bilinear partition of unity. □

The next step is to prove the global basis functions are indeed localizable. This makes use of the following lemma, which states some approximation properties of the projection operator π. In the analysis, we will make use of the Lagrange interpolation operator and a bubble function in the coarse grid. We define the Lagrange interpolation operator $I_h : C^0(\Omega) \cap H_0^1(\Omega) \to C^0(\Omega) \cap V_h$ by: for all $u \in C^0(\Omega) \cap H_0^1(\Omega)$, the interpolant $I_h u \in C^0(\Omega) \cap V_h$ is defined piecewise on each fine block $\tau \in \mathcal{T}^h$ by

$$(I_h u)(x) = u(x) \text{ for all vectices } x \text{ of } \tau, \tag{2.37}$$

which satisfies the standard approximation properties: there exists $C_I \geq 1$ such that for every $u \in C^0(\Omega) \cap H_0^1(\Omega)$,

$$\left\| \widetilde{\kappa}^{\frac{1}{2}} (u - I_h u) \right\|_{L^2(\tau)} + h \left\| \kappa^{\frac{1}{2}} \nabla \left(u - I_h u\right) \right\|_{L^2(\tau)} \leq C_I h \left\| \kappa^{\frac{1}{2}} \nabla u \right\|_{L^2(\tau)}, \tag{2.38}$$

on each fine block $\tau \in \mathcal{T}^h$. For any coarse grid block K, we define a bubble function B on K, i.e. $B(x) = 0$ for all $x \in \partial K$ and $B(x) > 0$ for all $x \in \text{int}\,(K)$. More precisely, we take $B = \prod_j \chi_j^{ms}$, where the product is taken over all the coarse grid nodes lying on the boundary ∂K. We can then

define the constant

$$C_\pi = \sup_{K\in\mathcal{T}^H,\mu\in V_{aux}} \frac{\int_K \widetilde{\kappa}\mu^2}{\int_K \widetilde{\kappa}B\mu^2}. \tag{2.39}$$

In our following analysis, we will assume the following smallness criterion on the fine mesh size h:

$$C_\pi C_I(C_{\mathcal{T}}^2+\lambda_{max})\|\Theta\|^{\frac{1}{2}}_{L^\infty(\Omega)}h<1, \tag{2.40}$$

where $C_{\mathcal{T}}$ is the maximum number of vertices over all coarse elements $K\in\mathcal{T}^H$ and

$$\begin{aligned}\lambda_{max} &= \max_{1\le i\le N}\lambda^{(i)}_{L_i},\\ \Theta &= \sum_j |\nabla\chi_j^{ms}|^2.\end{aligned} \tag{2.41}$$

Lemma 2.3.2. *Assume the smallness criterion* (2.40) *on the fine mesh size h. For any $v_{aux}\in V_{aux}$, there exists a function $v\in C^0(\Omega)\cap V_h$ such that*

$$\pi(v)=v_{aux},\quad \|v\|_a^2\le D\|v_{aux}\|_s^2,\quad \textit{supp}(v)\subseteq \textit{supp}(v_{aux}), \tag{2.42}$$

where the constant D is defined by

$$D=\left(\frac{2C_\pi(1+C_I^2)\left(C_{\mathcal{T}}^2+\lambda_{max}\right)}{1-C_\pi C_I\left(C_{\mathcal{T}}^2+\lambda_{max}\right)\|\Theta\|^{\frac{1}{2}}_{L^\infty(K_i)}h}\right)^2. \tag{2.43}$$

Proof. Let $v_{aux}\in V_{aux}^{(i)}$. We consider the following constraint minimization problem on the block K_i:

$$v=\operatorname{argmin}\left\{a_{DG}(v,v):\ v\in V_{h,0}(K_i),\quad s_i(v,\nu)=s_i(v_{aux},\nu)\text{ for all }\nu\in V_{aux}^{(i)}\right\}. \tag{2.44}$$

The minimization problem (2.44) is equivalent to the following variational problem: find $(v,\mu)\in$

$V_{h,0}(K_i) \times V_{aux}^{(i)}$ such that

$$\begin{aligned} a_i(v,w) + s_i(w,\mu) &= 0 \text{ for all } w \in V_{h,0}(K_i), \\ s_i(v - v_{aux}, \nu) &= 0 \text{ for all } \nu \in V_{aux}^{(i)}. \end{aligned} \tag{2.45}$$

The existence of solution of (2.45) is based on an inf-sup condition:

$$\inf_{\nu \in V_{aux}^{(i)}} \sup_{w \in V_{h,0}(K_i)\setminus\{0\}} \frac{s_i(w,\nu)}{a_i(w,w)} \geq \beta, \tag{2.46}$$

where $\beta > 0$ is a constant independent to be determined. Pick any $\nu \in V_{aux}^{(i)}$. We take $w = I_h(B\nu) \in C^0(\Omega) \cap V_h$. Since $\nu \in V_h(K_i)$ and $B(x) = 0$ for all vertices of K_i, we have $w \in V_{h,0}(K_i)$. First, we see that

$$\begin{aligned} s_i(w,\nu) &= \int_{K_i} \widetilde{\kappa} B\nu^2 + \int_{K_i} \widetilde{\kappa}(I_h(B\nu) - B\nu)\nu \\ &\geq C_\pi^{-1} \|\nu\|_{s(K_i)}^2 - \|I_h(B\nu) - B\nu\|_s \|\nu\|_{s(K_i)} \\ &\geq C_\pi^{-1} \|\nu\|_{s(K_i)}^2 - C_I h \left\| \widetilde{\kappa}^{\frac{1}{2}} \nabla(B\nu) \right\|_{L^2(K_i)} \|\nu\|_{s(K_i)} \\ &\geq C_\pi^{-1} \|\nu\|_{s(K_i)}^2 - C_I \|\Theta\|_{L^\infty(K_i)}^{\frac{1}{2}} h \left\| \kappa^{\frac{1}{2}} \nabla(B\nu) \right\|_{L^2(K_i)} \|\nu\|_{s(K_i)}. \end{aligned} \tag{2.47}$$

On the other hand, we observe that

$$\begin{aligned} a_i(w,w) &\leq 2 \left(\left\| \kappa^{\frac{1}{2}} \nabla(B\nu) \right\|_{L^2(K_i)}^2 + \left\| \kappa^{\frac{1}{2}} \nabla(B\nu - I_h(B\nu)) \right\|_{L^2(K_i)}^2 \right) \\ &\leq 2(1 + C_I^2) \left\| \kappa^{\frac{1}{2}} \nabla(B\nu) \right\|_{L^2(K_i)}. \end{aligned} \tag{2.48}$$

It remains to estimate the term $\left\| \kappa^{\frac{1}{2}} \nabla(B\nu) \right\|_{L^2(K_i)}$. Since $0 \leq \chi_j^{ms} \leq 1$, we have $0 \leq B \leq 1$ and $|\nabla B|^2 \leq C_{\mathcal{T}}^2 \Theta$. Using these facts together with $\nabla(B\nu) = (\nabla B)\nu + B(\nabla\nu)$, we imply

$$\begin{aligned} \left\| \kappa^{\frac{1}{2}} \nabla(B\nu) \right\|_{L^2(K_i)}^2 &\leq C_{\mathcal{T}}^2 \|\nu\|_{s(K_i)}^2 + a_i(\nu,\nu) \\ &\leq \left(C_{\mathcal{T}}^2 + \lambda_{max} \right) \|\nu\|_{s(K_i)}^2 \end{aligned} \tag{2.49}$$

By taking the inf-sup constant

$$\beta = \frac{1 - C_\pi C_I \left(C_{\mathcal{T}}^2 + \lambda_{max}\right) \|\Theta\|_{L^\infty(K_i)}^{\frac{1}{2}} h}{2C_\pi(1 + C_I^2)\left(C_{\mathcal{T}}^2 + \lambda_{max}\right)}, \tag{2.50}$$

we prove the inf-sup condition (2.46) and therefore the existence of $(v, \mu) \in V_{h,0}(K_i) \times V_{aux}^{(i)}$ in (2.45). It is then direct to check that the solution $v \in V_{h,0}(K_i)$ satisfies the desired properties. $\square$

We remark that, without loss of generality, we can assume $D \geq C_0^2(1 + C_I^2)$. We are now going to establish an estimate of the difference between the global multiscale basis functions and localized multiscale basis functions. We will see that the global multiscale basis functions have a decay property, and their values are small outside a suitably large oversampled domain. We will make use of a cutoff function in our proof. For each coarse block K_i and $M > m$, the oversampling regions $K_{i,M}$ and $K_{i,m}$ define an outer neighborhood and an inner neighborhood respectively. We define $\chi_i^{M,m} \in \text{span}\{\chi_j^{ms}\}$ such that $0 \leq \chi_i^{M,m} \leq 1$ and

$$\chi_i^{M,m} = 1 \text{ in } K_{i,m} \text{ and } \chi_i^{M,m} = 0 \text{ in } \Omega \setminus K_{i,M}. \tag{2.51}$$

Moreover, we define the following DG norm for $w \in V$ on $K_{i,M} \setminus K_{i,m}$.

$$\|w\|_{DG(K_{i,M}\setminus K_{i,m})}^2 = \sum_{K_k \subset K_{i,M}\setminus K_{i,m}} a_k(w, w) + \frac{\gamma}{h} \sum_{E \in \mathcal{E}^H(K_{i,M}\setminus K_{i,m})} \int_E \overline{\kappa}[\![w]\!]^2 d\sigma, \tag{2.52}$$

where $\mathcal{E}^H(K_{i,M} \setminus K_{i,m})$ denotes the collection of all coarse grid edges in $\mathcal{E}^H$ which lie within in the interior of $K_{i,M} \setminus K_{i,m}$ and the boundary of $K_{i,M}$. We remark that the definition also applies to a region $\Omega \setminus K_{i,m}$ in the case when M is sufficiently large.

Lemma 2.3.3. *Assume the smallness criterion* (2.40) *on the fine mesh size* h*. Suppose* $m > 2$ *is the number of coarse grid layers in the oversampled domain* $K_{i,m}$ *extended from the coarse grid block* K_i*. Let* $\phi_j^{(i)} \in V_{aux}$ *be a given auxiliary multiscale basis function. Let* $\psi_j^{(i)} \in V_{glo}$ *be the global multiscale basis function obtained from* (2.17)*, and* $\psi_{j,ms}^{(i)} \in V_h\left(K_{i,m}\right)$ *be the localized multiscale*

basis function obtained from (2.19)*. Then we have*

$$\|\psi_j^{(i)} - \psi_{j,ms}^{(i)}\|_a^2 \leq E\|\phi_j^{(i)}\|_{s(K_i)}^2, \tag{2.53}$$

where $E = 40D^3(1+\Lambda^{-1})\left(1+6D^{-2}\left(1+\Lambda^{-\frac{1}{2}}\right)^{-1}\right)^{1-m}$.

Proof. By the variational formulations (2.18) and (2.20), we have

$$a_{DG}\left(\psi_j^{(i)} - \psi_{j,ms}^{(i)}, \psi\right) + s_i\left(\psi, \mu_j^{(i)} - \mu_{j,ms}^{(i)}\right) = 0 \text{ for all } \psi \in V_h(K_{i,m}). \tag{2.54}$$

By Lemma 3.3.2, there exists $\widetilde{\phi}_j^{(i)} \in V_h$ such that

$$\pi(\widetilde{\phi}_j^{(i)}) = \phi_j^{(i)}, \quad \|\widetilde{\phi}_j^{(i)}\|_a^2 \leq D\|\phi_j^{(i)}\|_{s(K_i)}^2, \quad \text{supp}\left(\widetilde{\phi}_j^{(i)}\right) \subseteq K_i. \tag{2.55}$$

We take $\eta = \psi_j^{(i)} - \widetilde{\phi}_j^{(i)} \in V_h$ and $\zeta = \widetilde{\phi}_j^{(i)} - \psi_{j,ms}^{(i)} \in V_h(K_{i,m})$. By definition, we have $\pi(\eta) = \pi(\zeta) = 0$ and therefore $\eta, \zeta \in \widetilde{V}_h$. Again, by Lemma 3.3.2, there exists $\rho \in V_h$ such that

$$\pi(\rho) = \pi(I_h(\chi_i^{m,m-1}\eta)), \quad \|\rho\|_a^2 \leq D\|\pi(I_h(\chi_i^{m,m-1}\eta))\|_s^2, \quad \text{supp}\,(\rho) \subseteq K_{i,m} \setminus K_{i,m-1}. \tag{2.56}$$

Take $\tau = \rho - I_h(\chi_i^{m,m-1}\eta) \in V_h$. Again, $\pi(\tau) = 0$ and hence $\tau \in \widetilde{V}_h$. Taking $\psi = \tau - \zeta \in V_h(K_{i,m})$ in (2.54) and making use of the fact $\tau - \zeta \in \widetilde{V}_h$, we have

$$a_{DG}\left(\psi_j^{(i)} - \psi_{j,ms}^{(i)}, \tau - \zeta\right) = 0, \tag{2.57}$$

and therefore

$$\begin{aligned}
\left\|\psi_j^{(i)} - \psi_{j,ms}^{(i)}\right\|_a^2 &= a_{DG}\left(\psi_j^{(i)} - \psi_{j,ms}^{(i)}, \eta + \zeta\right) \\
&= a_{DG}\left(\psi_j^{(i)} - \psi_{j,ms}^{(i)}, \eta + \tau\right) \\
&\leq \left\|\psi_j^{(i)} - \psi_{j,ms}^{(i)}\right\|_a \|\eta + \tau\|_a,
\end{aligned} \tag{2.58}$$

which in turn implies

$$\begin{aligned}\left\|\psi_j^{(i)}-\psi_{j,ms}^{(i)}\right\|_a^2 &\leq \|\eta+\tau\|_a^2\\ &= \left\|I_h((1-\chi_i^{m,m-1})\eta)+\rho\right\|_a^2\\ &\leq 2\left(\left\|I_h((1-\chi_i^{m,m-1})\eta)\right\|_a^2+\|\rho\|_a^2\right)\\ &\leq 2\left(C_0^2\left\|I_h((1-\chi_i^{m,m-1})\eta)\right\|_{DG}^2+\|\rho\|_a^2\right)\\ &\leq 2\left(2C_0^2(1+C_I^2)\left\|(1-\chi_i^{m,m-1})\eta\right\|_{DG}^2+\|\rho\|_a^2\right).\end{aligned} \tag{2.59}$$

For the first term on the right hand side of (2.59), by using $\nabla\left((1-\chi_i^{m,m-1})\eta\right)=-\nabla\chi_i^{m,m-1}\eta+(1-\chi_i^{m,m-1})\nabla\eta$ and $0\leq 1-\chi_i^{m,m-1}\leq 1$, we have

$$\left\|(1-\chi_i^{m,m-1})\eta\right\|_a^2\leq 2\left(\|\eta\|_{DG(\Omega\setminus K_{i,m-1})}^2+\|\eta\|_{s(\Omega\setminus K_{i,m-1})}^2\right). \tag{2.60}$$

For the second term on the right hand side of (2.59), using the definition of ρ in (2.56) and $0\leq 1-\chi_i^{m,m-1}\leq 1$, we obtain

$$\|\rho\|_a^2\leq D\|\pi(\chi_i^{m,m-1}\eta)\|_s^2\leq D\|\chi_i^{m,m-1}\eta\|_s^2\leq D\|\eta\|_{s(\Omega\setminus K_{i,m-1})}^2. \tag{2.61}$$

Moreover, since $\eta\in\widetilde{V}_h$, by the spectral problem (2.10), we have

$$\|\eta\|_{s(\Omega\setminus K_{i,m-1})}^2\leq\Lambda^{-1}\sum_{K_k\subset\Omega\setminus K_{i,m-1}}a_k(\eta,\eta). \tag{2.62}$$

Combining all these estimates, we obtain

$$\left\|\psi_j^{(i)}-\psi_{j,ms}^{(i)}\right\|_a^2\leq 10D(1+\Lambda^{-1})\|\eta\|_{DG(\Omega\setminus K_{i,m-1})}^2. \tag{2.63}$$

Next, we will provide a recursive estimate for η in the number of oversampling layers m. We take $\xi=1-\chi_i^{m-1,m-2}$. Then $0\leq\xi\leq 1$ and $\xi=1$ in $\Omega\setminus K_{i,m-1}$. Using $\nabla(\xi^2\eta)=\xi^2\nabla\eta+2\xi\eta\nabla\xi$,

for every $K \in \mathcal{T}^H$, we have

$$\int_K \kappa \nabla \eta \cdot \nabla(\xi^2 \eta) = \int_K \kappa \nabla \eta \cdot \left(\xi^2 \nabla \eta + 2\xi\eta\nabla\xi\right) = \int_K \kappa |\nabla(\xi\eta)|^2 - \int_K \kappa |\nabla \xi|^2 \eta^2. \tag{2.64}$$

In addition, using $\nabla(\xi\eta) = \xi\nabla\eta + \eta\nabla\xi$, for every $E \in \mathcal{E}^H$, we have

$$\begin{aligned}
&- \int_E \{\kappa \nabla \eta \cdot n_E\} [\![\xi^2 \eta]\!] - \int_E \{\kappa \nabla(\xi^2\eta) \cdot n_E\} [\![\eta]\!] + \frac{\gamma}{h} \int_E \overline{\kappa} [\![\eta]\!] [\![\xi^2 \eta]\!] \\
&= - \int_E \{\kappa \nabla \eta \cdot n_E\} [\![\xi^2 \eta]\!] - \int_E \left\{\kappa(\xi^2 \nabla \eta + 2\xi\eta\nabla\xi) \cdot n_E\right\} [\![\eta]\!] + \frac{\gamma}{h} \int_E \overline{\kappa} [\![\eta]\!] [\![\xi^2 \eta]\!] \\
&= -2 \left(\int_E \{\kappa \xi \nabla \eta \cdot n_E\} [\![\xi\eta]\!] + \int_E \{\kappa \eta \nabla \xi \cdot n_E\} [\![\xi\eta]\!] \right) + \frac{\gamma}{h} \int_E \overline{\kappa} [\![\xi\eta]\!]^2 \\
&= -2 \int_E \{\kappa \nabla(\xi\eta) \cdot n_E\} [\![\xi\eta]\!] + \frac{\gamma}{h} \int_E \overline{\kappa} [\![\xi\eta]\!]^2.
\end{aligned} \tag{2.65}$$

Summing over $K \in \mathcal{T}^H$ and $E \in \mathcal{E}^H$, we obtain

$$\|\xi\eta\|_a^2 \leq a_{DG}(\eta, \xi^2\eta) + \|\eta\|^2_{s(K_{i,m-1} \setminus K_{i,m-2})}, \tag{2.66}$$

where we make use of the fact that $\nabla\xi = 0$ outside $K_{i,m-1} \setminus K_{i,m-2}$. We start with estimating the first term on the right hand side of (2.66). For any coarse element $K_k \in \Omega \setminus K_{i,m-1}$, since $\xi = 1$ in K_k and $\eta \in \widetilde{V}_h$, we have

$$s\left(\xi^2\eta, \phi_j^{(k)}\right) = s\left(\eta, \phi_j^{(k)}\right) = 0 \text{ for all } j = 1, 2, \ldots, L_k. \tag{2.67}$$

On the other hand, for any coarse element $K_k \in K_{i,m-2}$, since $\xi = 0$ in K_k, we have

$$s\left(\xi^2\eta, \phi_j^{(k)}\right) = 0 \text{ for all } j = 1, 2, \ldots, L_k. \tag{2.68}$$

Therefore, $\text{supp}(\pi(I_h(\xi^2\eta))) \subset K_{i,m-1} \setminus K_{i,m-2}$. By Lemma 3.3.2, there exists $\sigma \in V_h$ such that

$$\pi(\sigma) = \pi(I_h(\xi^2\eta)), \quad \|\gamma\|_a^2 \leq D\|\pi(I_h(\xi^2\eta))\|_s^2, \quad \text{supp}(\sigma) \subset K_{i,m-1} \setminus K_{i,m-2}. \tag{2.69}$$

For any coarse element $K_k \subset K_{i,m-1} \setminus K_{i,m-2}$, since $0 \leq \xi \leq 1$ and $\pi(\eta) = 0$, we have

$$\|\pi(I_h(\xi^2\eta))\|^2_{s(K_k)} \leq \|I_h(\xi^2\eta)\|^2_{s(K_k)} \leq \|I_h(\eta)\|^2_{s(K_k)} = \|\eta\|^2_{s(K_k)} \leq \Lambda^{-1} a_k(\eta, \eta). \tag{2.70}$$

Summing over $K_k \subset K_{i,m-1} \setminus K_{i,m-2}$, we obtain

$$\|\pi(I_h(\xi^2\eta))\|^2_s \leq \Lambda^{-1} \sum_{K_k \subset K_{i,m-1} \setminus K_{i,m-2}} a_k(\eta, \eta). \tag{2.71}$$

We take $\theta = I_h(\xi^2\eta) - \sigma$. Again, $\pi(\theta) = 0$ and $\theta \in \widetilde{V}_h$, which yields

$$a_{DG}\left(\psi_j^{(i)}, \theta\right) = 0. \tag{2.72}$$

On the other hand, $\text{supp}(\theta) \subset \Omega \setminus K_{i,m-2}$ and $\text{supp}(\widetilde{\phi}_j^{(i)}) \subset K_i$. Since θ and $\widetilde{\phi}_j^{(i)}$ has disjoint supports, we have

$$a_{DG}\left(\widetilde{\phi}_j^{(i)}, \theta\right) = 0. \tag{2.73}$$

Therefore, we obtain

$$a_{DG}(\eta, \theta) = a_{DG}\left(\psi_j^{(i)} - \widetilde{\phi}_j^{(i)}, \theta\right) = 0. \tag{2.74}$$

Recall from the definition that $I_h(\xi^2\eta) = \theta + \sigma$ and $\text{supp}(\sigma) \subset K_{i,m-1} \setminus K_{i,m-2}$. Hence we have

$$\begin{aligned} a_{DG}(\eta, I_h(\xi^2\eta)) &= a_{DG}(\eta, \sigma) \\ &\leq C_0 \|\eta\|_{DG(K_{i,m-1} \setminus K_{i,m-2})} \|\sigma\|_a \\ &\leq C_0 D^{\frac{1}{2}} \|\eta\|_{DG(K_{i,m-1} \setminus K_{i,m-2})} \|\pi(I_h(\xi^2\eta))\|_s \\ &\leq D\Lambda^{-\frac{1}{2}} \|\eta\|^2_{DG(K_{i,m-1} \setminus K_{i,m-2})}. \end{aligned} \tag{2.75}$$

On the other hand, making use of the fact that $\xi^2 = 0$ in $K_{i,m-2}$ and $\xi^2 = 1$ in $\Omega \setminus K_{i,m-1}$, we observe that $\xi^2\eta = I_h(\xi^2\eta)$ outside $K_{i,m-1} \setminus K_{i,m-2}$. Moreover, $\xi^2\eta - I_h(\xi^2\eta)$ is globally

continuous. Thus, we obtain

$$\begin{aligned} a_{DG}(\eta, \xi^2\eta - I_h(\xi^2\eta)) &\leq C_0^2 \|\eta\|_{DG(K_{i,m-1}\setminus K_{i,m-2})} \left\|\xi^2\eta - I_h(\xi^2\eta)\right\|_{DG(K_{i,m-1}\setminus K_{i,m-2})} \\ &\leq C_0^2 C_I \|\eta\|_{DG(K_{i,m-1}\setminus K_{i,m-2})} \left\|\xi^2\eta\right\|_{DG(K_{i,m-1}\setminus K_{i,m-2})} \\ &\leq \frac{D}{2}\left(\|\eta\|^2_{DG(K_{i,m-1}\setminus K_{i,m-2})} + \left\|\xi^2\eta\right\|^2_{DG(K_{i,m-1}\setminus K_{i,m-2})}\right). \end{aligned} \tag{2.76}$$

Again, using $\nabla(\xi^2\eta) = \xi^2\nabla\eta + 2\xi\eta\nabla\xi$, we have

$$\left\|\xi^2\eta\right\|^2_{DG(K_{i,m-1}\setminus K_{i,m-2})} \leq 2\|\eta\|^2_{DG(K_{i,m-1}\setminus K_{i,m-2})} + 8\|\eta\|^2_{s(K_{i,m-1}\setminus K_{i,m-2})}. \tag{2.77}$$

Combining (2.66), (2.75), (2.76) and (2.77), we arrive at

$$\|\xi\eta\|_a^2 \leq D\left(\left(\frac{3}{2} + \Lambda^{-\frac{1}{2}}\right)\|\eta\|^2_{DG(K_{i,m-1}\setminus K_{i,m-2})} + 5\|\eta\|^2_{s(K_{i,m-1}\setminus K_{i,m-2})}\right). \tag{2.78}$$

Moreover, since $\pi(\eta) = 0$, we have

$$\|\eta\|_{s(K_{i,m-1}\setminus K_{i,m-2})} \leq \Lambda^{-\frac{1}{2}}\|\eta\|_{DG(K_{i,m-1}\setminus K_{i,m-2})}, \tag{2.79}$$

which implies

$$\|\xi\eta\|_a^2 \leq 6D\left(1 + \Lambda^{-\frac{1}{2}}\right)\|\eta\|^2_{DG(K_{i,m-1}\setminus K_{i,m-2})}. \tag{2.80}$$

By the equivalence of norms, we have

$$\|\eta\|^2_{DG(\Omega\setminus K_{i,m-1})} \leq C_0^2\|\xi\eta\|_a^2 \leq 6D^2\left(1 + \Lambda^{-\frac{1}{2}}\right)\|\eta\|^2_{DG(K_{i,m-1}\setminus K_{i,m-2})}. \tag{2.81}$$

We obtain the recurrence estimate

$$\begin{aligned} \|\eta\|^2_{DG(\Omega\setminus K_{i,m-2})} &= \|\eta\|^2_{DG(\Omega\setminus K_{i,m-1})} + \|\eta\|^2_{DG(K_{i,m-1}\setminus K_{i,m-2})} \\ &\geq \left(1 + 6D^{-2}\left(1 + \Lambda^{-\frac{1}{2}}\right)^{-1}\right)\|\eta\|^2_{DG(\Omega\setminus K_{i,m-1})}. \end{aligned} \tag{2.82}$$

Inductively, we have

$$\begin{aligned}\|\eta\|^2_{DG(\Omega\setminus K_{i,m-1})} &\leq \left(1+6D^{-2}\left(1+\Lambda^{-\frac{1}{2}}\right)^{-1}\right)^{1-m}\|\eta\|^2_{DG(\Omega\setminus K_{i,1})}\\ &\leq D\left(1+6D^{-2}\left(1+\Lambda^{-\frac{1}{2}}\right)^{-1}\right)^{1-m}\|\eta\|^2_a.\end{aligned} \tag{2.83}$$

Combining (2.63) and (2.83), we see that

$$\left\|\psi_j^{(i)}-\psi_{j,ms}^{(i)}\right\|_a^2 \leq 10D^2\left(1+\Lambda^{-1}\right)\left(1+6D^{-2}\left(1+\Lambda^{-\frac{1}{2}}\right)^{-1}\right)^{1-m}\|\eta\|_a^2 \tag{2.84}$$

By the energy minimizing property of $\psi_j^{(i)}$, we have

$$\|\eta\|_a \leq \|\psi_j^{(i)}\|_a + \|\widetilde{\phi}_j^{(i)}\|_a \leq 2\|\widetilde{\phi}_j^{(i)}\|_a \leq 2D^{\frac{1}{2}}\|\phi_j^{(i)}\|_{s(K_i)}. \tag{2.85}$$

We obtain the desired result. □

Now, we are ready to establish our main theorem, which estimates the error between the solution u_h and the multiscale solution u_{ms}.

Theorem 2.3.4. *Let $u_h \in V_h$ be the solution of* (2.4)*, $u_{glo} \in V_{glo}$ be the solution of* (2.26) *with the global multiscale basis functions defined by* (2.17)*, and $u_{ms} \in V_{ms}$ be the multiscale solution of* (2.22) *with the localized multiscale basis functions defined on an oversampled domain with $m > 2$ coarse grid layers by* (2.19)*. Then we have*

$$\|u_h - u_{ms}\|_a \leq C\Lambda^{-\frac{1}{2}}\|\widetilde{\kappa}^{-\frac{1}{2}}f\|_{L^2(\Omega)} + Cm^dE^{\frac{1}{2}}\|u_{glo}\|_s, \tag{2.86}$$

Moreover, if we let $k = O\left(\log\left(\frac{\kappa_1}{H}\right)\right)$ and replace the multiscale partition of unity $\{\chi_j^{ms}\}$ by the bilinear partition of unity, we have

$$\|u_h - u_{ms}\|_a \leq CH\Lambda^{-\frac{1}{2}}\|\kappa^{-\frac{1}{2}}f\|_{L^2(\Omega)}. \tag{2.87}$$

Proof. First, we write u_{glo} in the linear combination of the basis $\{\psi_k^{(j)}\}$

$$u_{glo} = \sum_{i=1}^{N}\sum_{j=1}^{L_i} \alpha_j^{(i)} \psi_j^{(i)}. \tag{2.88}$$

and define $\widehat{u}_{ms} \in V_{ms}$ by

$$\widehat{u}_{ms} = \sum_{i=1}^{N}\sum_{j=1}^{L_i} \alpha_j^{(i)} \psi_{j,ms}^{(i)}. \tag{2.89}$$

From (2.4) and (2.22), we obtain the Galerkin orthogonality

$$a_{DG}(u_h - u_{ms}, w) = 0 \text{ for all } w \in V_{ms}, \tag{2.90}$$

which gives

$$\|u_h - u_{ms}\|_a \leq \|u_h - \widehat{u}_{ms}\|_a \leq \|u_h - u_{glo}\|_a + \|u_{glo} - \widehat{u}_{ms}\|_a. \tag{2.91}$$

Using Lemma 3.3.3, we see that

$$\begin{aligned}
\|u_{glo} - \widehat{u}_{ms}\|_a^2 &= \left\| \sum_{i=1}^{N}\sum_{j=1}^{L_i} \alpha_j^{(i)} (\psi_j^{(i)} - \psi_{j,ms}^{(i)}) \right\|_a^2 \\
&\leq Cm^d \sum_{i=1}^{N} \left\| \sum_{j=1}^{L_i} \alpha_j^{(i)} (\psi_j^{(i)} - \psi_{j,ms}^{(i)}) \right\|_a^2 \\
&\leq Cm^d E \sum_{i=1}^{N} \left\| \sum_{j=1}^{L_i} \alpha_j^{(i)} \phi_j^{(i)} \right\|_s^2 \\
&= Cm^d E \|u_{glo}\|_s^2,
\end{aligned} \tag{2.92}$$

where the last equality follows from the orthogonality of the eigenfunctions in (2.10). Using the estimates (2.28) and (2.92) in (2.91), we have

$$\|u_h - u_{ms}\|_a \leq \Lambda^{-\frac{1}{2}} \|\widetilde{\kappa}^{-\frac{1}{2}} f\|_{L^2(\Omega)} + Cm^{\frac{d}{2}} E^{\frac{1}{2}} \|u_{glo}\|_s. \tag{2.93}$$

This completes the first part of the theorem. Next, we assume the partition of unity functions are

bilinear, and we are going to estimate $\|u_{glo}\|_s$. Using the fact that $|\nabla \chi_k| = O(H^{-1})$, we have

$$\|u_{glo}\|_s^2 \leq CH^{-2}\kappa_1 \|u_{glo}\|_{L^2(\Omega)}^2. \tag{2.94}$$

Then, by Poincaré inequality, we have

$$\|u_{glo}\|_{L^2(\Omega)}^2 \leq C\kappa_0^{-1} \|u_{glo}\|_a^2. \tag{2.95}$$

By taking $w = u_{glo} \in V_{glo}$ in (2.26), we obtain

$$\|u_{glo}\|_a^2 = (f, u_{glo})_{0,\Omega} \leq \|\widetilde{\kappa}^{-\frac{1}{2}} f\|_{L^2(\Omega)} \|u_{glo}\|_s. \tag{2.96}$$

Combining these estimates, we have

$$\|u_{glo}\|_s \leq CH^{-2}\kappa_0^{-1}\kappa_1 \|\widetilde{\kappa}^{-\frac{1}{2}} f\|_{L^2(\Omega)}. \tag{2.97}$$

To obtain our desired result, we need

$$H^{-2}\kappa_1 m^{\frac{d}{2}} E^{\frac{1}{2}} = O(1). \tag{2.98}$$

Taking logarithm, we have

$$\log(H^{-2}) + \log(\kappa_1) + \frac{d}{2}\log(m) + \frac{1-m}{2}\log\left(1 + \Lambda^{-\frac{1}{2}}\right) = O(1). \tag{2.99}$$

Thus, taking $m = O\left(\log\left(\frac{\kappa_1}{H}\right)\right)$ completes the proof of the second result. □

2.4 Numerical results

In this section, we will present numerical examples with high contrast media to demonstrate the convergence of our proposed method with respect to the coarse mesh size H and the number of oversampling layers m, and illustrate possible improvements in error robustness with respect

to contrast by employing the idea of constructing multiscale basis function by relaxation method introduced in [39]. In all the experiments, the IPDG penalty parameter in (2.5) is set to be $\gamma = 4$, so as to ensure the coercivity of the bilinear form a_{DG}. We consider a highly heterogeneous permeability field κ in $\Omega = [0.1]^2$ as shown in Figure 2.3, with the background value is $\kappa = 1$ and the value in the channels and inclusions is 10^4. and the resolution is 400×400, i.e. κ is piecewise constant on a fine grid with mesh size $h = 1/400$. The coarse mesh size varies from $H = 1/80$ to $H = 1/10$, and the number of oversampling layers varies from $m = 3$ to $m = 6$. In all these combinations, there are no more than 3 high conductivity channels in a coarse block $K \in \mathcal{T}^H$. As a result, we have 3 small eigenvalues in a local spectral problem (2.10), and it suffices to use 3 auxiliary basis functions per coarse block to construct the correspoding localized multiscale basis functions. The source function is taken as

$$f(x, y) = 2\pi^2 \sin(\pi x) \sin(\pi y) \text{ for all } (x, y) \in \Omega. \tag{2.100}$$

Table 2.1 records the error when we take the number of oversampling layer to be approximately $m \approx 4 \log(1/H)/\log(1/10)$. The results show that the method provides optimal convergence in energy norm, which agrees with our theoretical finding in Section 2.3, and the L^2 error converges with second order. Table 2.2 records the error with various number of oversampling layers and a fixed coarse mesh sizes $H = 1/40$. It can be observed that increasing the number of oversampling layers improves the quality of approximations, but the decay in error is limited when the oversampling region is sufficiently large. This numerically verifies that the multiscale basis functions can indeed be localized.

m	H	Energy error	L^2 error
4	1/10	7.4625%	0.7653%
6	1/20	1.5392%	0.0625%
7	1/40	0.7266%	0.0160%
8	1/80	0.3433%	0.0035%

Table 2.1: History of convergence with different coarse mesh size H for Experiment 1.

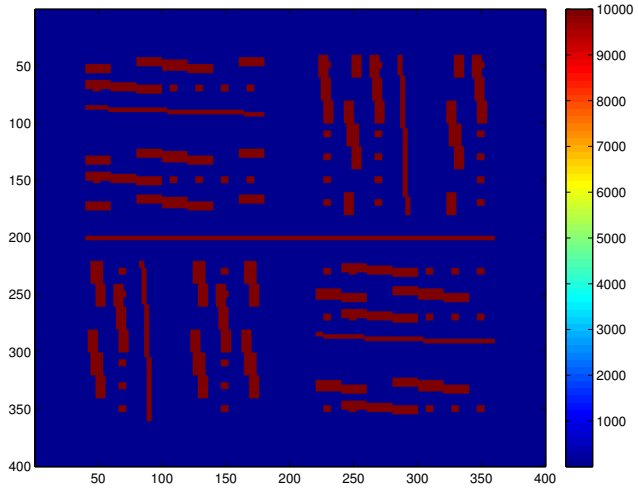

Figure 2.3: The permeability field κ for Experiment 1.

m	Energy error	L^2 error
3	84.7517%	72.3079%
4	19.0936%	3.6716%
5	2.6687%	0.0720%
6	0.7836%	0.0161%
7	0.7266%	0.0160%
8	0.7259%	0.0160%

Table 2.2: Error table with different number of oversampling layers m and a fixed coarse mesh size $H = 1/40$ for Experiment 1.

Next, we present the idea of the relaxed formulation of (2.19). Instead of using the method of Lagrange multiplier as in (2.20), the ϕ-orthogonality is imposed weakly by a penalty formulation. The localized multiscale basis function $\psi_{j,ms}^{(i)} \in V_h\left(K_{i,m}\right)$ is defined as the solution of the following relaxed constrained energy minimization problem

$$\psi_{j,ms}^{(i)} = \operatorname{argmin}\left\{a_{DG}\left(\psi,\psi\right) + s\left(\pi\left(\psi\right) - \phi_j^{(i)}, \pi\left(\psi\right) - \phi_j^{(i)}\right) : \psi \in V_h\left(K_{i,m}\right)\right\}. \tag{2.101}$$

The minimization problem (2.101) is equivalent to the following variational problem: find $\psi_{j,ms}^{(i)} \in$

$V_h(K_{i,m})$ such that

$$a_{DG}\left(\psi_{j,ms}^{(i)},\psi\right)+s\left(\pi\left(\psi_{j,ms}^{(i)}\right),\pi(\psi)\right)=s\left(\phi_j^{(i)},\pi(\psi)\right) \text{ for all } \psi\in V_h(K_{i,m}). \tag{2.102}$$

The construction of multiscale finite element space and coarse-scale model then follow (2.21) and (2.22) respectively. We compare the performance of the multiscale method with multiscale basis functions constructed by method of Lagrange multiplier (2.20) and the relaxation method (2.102) at different contrast values, where the coarse mesh size is taken as $H = 1/10$ and the number of oversampling layers as $m = 4$. In Table 2.3, we record the energy error and L^2 error with different contrast κ_1, where $\kappa_1 \gg 1$ is the value of κ in the high conductivity channels. It can be seen that the relaxation method is more robust with respect to contrast.

	Lagrange multiplier		Relaxation	
κ_1	Energy error	L^2 error	Energy error	L^2 error
10^4	7.4625%	0.7653%	6.3757%	0.6395%
10^5	12.6299%	1.6977%	6.3986%	0.6467%
10^6	32.1465%	10.5146%	6.4020%	0.6478%
10^7	64.1190%	41.8127%	6.4049%	0.6481%
10^8	77.1229%	60.4947%	6.4301%	0.6503%

Table 2.3: Comparison of the method of Lagrange multiplier and the relaxation method with different contrast values for Experiment 1.

3. CONSTRAINT ENERGY MINIMIZING GENERALIZED MULTISCALE FINITE ELEMENT METHOD FOR DUAL CONTINUUM MODEL *

Dual continuum models are used to describe a wide range of scientific and engineering applications, for example, complex processes in shale reservoirs, where such models are used to describe a complex interaction of the organic and inorganic matter. In real world applications, properties of the dual continuum models are highly heterogeneous and leads to the construction of the fine grids to resolve also small scale heterogeneity in level of mesh construction. Direct simulation on the fine grid is computationally expensive. In this chapter, we consider a dual continuum model for describing fluid flow in porous media with highly connected fracture network, where we have coupled system of equations for porous matrix and for fracture network with specific mass transfer between them. We present Constraint Energy Minimizing Generalized Multiscale Finite Element Method (CEM-GMsFEM) as a model reduction technique for the dual continuum model. We establish theoretical results showing that the method provides a convergence depends only on the coarse mesh size and independent of scales and contrast. we will construct a set of local auxiliary multiscale basis functions, as in GMsFEM. These functions are dominant eigenfunctions of local spectral problems, and the number of these functions is the same as the number of high contrast channels. We emphasize that this is the minimal number of degrees of freedoms required to represent channelized effects. We also remark that these eigenfunctions are crucial in the construction of localized basis functions. The second key component is multiscale basis functions. These functions are obtained by minimizing an energy functional subject to certain constraints. These constraints are formulated using the auxiliary functions with the purpose of obtaining localized multiscale basis functions. In particular, for each of the auxiliary function, the constraints require the minimizer of the energy functional is orthogonal, in a weighted L^2 sense, to all other auxiliary functions except the selected one. For the selected auxiliary functions, the constraints require the

* Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

minimizer of the energy functional to satisfy a normalized condition. Combining the effects of auxiliary functions and energy minimization, we show that the minimizer of the energy functional has exponential decay property, and is very small outside an oversampling region obtained by the support of the selected auxiliary function. Moreover, the resulting multiscale method obtained by a Galerkin formulation has a mesh dependent convergence rate.

Similar to Chapter 2, we will first construct a set of local auxiliary multiscale basis functions. These functions are dominant eigenfunctions of local spectral problems, and the number of these functions is the same as the number of high contrast channels. We emphasize that this is the minimal number of degrees of freedoms required to represent channelized effects. We also remark that these eigenfunctions are crucial in the construction of localized basis functions. Using the auxiliary basis functions, we define multiscale basis functions by minimizing an energy functional subject to certain constraints. These constraints are formulated using the auxiliary functions with the purpose of obtaining localized multiscale basis functions. In particular, for each of the auxiliary function, the constraints require the minimizer of the energy functional is orthogonal, in a weighted L^2 sense, to all other auxiliary functions except the selected one. For the selected auxiliary functions, the constraints require the minimizer of the energy functional to satisfy a normalized condition. Combining the effects of auxiliary functions and energy minimization, we show that the minimizer of the energy functional has exponential decay property, and is very small outside an oversampling region obtained by the support of the selected auxiliary function. Moreover, the resulting multiscale method obtained by a Galerkin formulation has a mesh dependent convergence rate.

The chapter is organized as follows. In Section 3.1, we will introduce the dual continuum model. Our multiscale method will be presented in Section 3.2 and analyzed in Section 3.3. Finally, in Section 3.4, we will present some numerical tests.

3.1 Dual continuum Model

We consider the following dual continuum model

$$\begin{aligned} c_1\frac{\partial p_1}{\partial t} - \text{div}(\kappa_1\nabla p_1) + \sigma(p_1 - p_2) &= f_1, \\ c_2\frac{\partial p_2}{\partial t} - \text{div}(\kappa_2\nabla p_2) - \sigma(p_1 - p_2) &= f_2, \end{aligned} \tag{3.1}$$

in a computational domain $\Omega \subset \mathbb{R}^2$. Here, for $i = 1, 2$, c_i is the compressibility, p_i is the pressure, κ_i is the permeability, and f_i is the source function for the i-th continuum. In addition, the continua are coupled through the mass exchange, and σ is a parameter which accounts for the strength of mass transfer between the continua. One particular application of the dual continuum model 3.1 is to represent the global interactive effects of the unresolved fractures and the matrix.

Let Ω be domain with high conductive channels (heterogeneous media)

$$\Omega = D_m^i \cup D_f^i, \quad D_f^i = \cup_{l=1}^{n_f} D_{f,l}^i \tag{3.2}$$

where indices m and f represent the two subdomains with low and high permeability, n_f is the number of high conductive channels, i is the continuum. We prescribe the initial condition $p_i(0, \cdot) = p_i^0$ in Ω and the boundary condition $p_i(t, \cdot) = 0$ on $\partial\Omega$ for $t > 0$. Furthermore, we have

$$\kappa_i(x) = \begin{cases} \kappa_i^m, & x \in D_m^i, \\ \kappa_{l,i}^f, & x \in D_{f,l}^i, \end{cases}, \quad c_i(x) = \begin{cases} c_i^m, & x \in D_m^i, \\ c_{l,i}^f, & x \in D_{f,l}^i, \end{cases}, \quad i = 1, 2, \quad l = 1, ..., n_f,$$

where $\kappa_{l,i}^f$ and $c_{l,i}^f$ are the permeability and compressibility on the l-th channel for the continuum i in subdomain $D_{f,l}^i$; κ_i^m and c_i^m are the permeability and compressibility in subdomain D_m^i. Here, we assume the permeability fields are uniformly bounded, i.e.

$$0 < \underline{\kappa} \leq \kappa_i(x) \leq \overline{\kappa} \quad \text{for } x \in \Omega, \quad \text{for } i = 1, 2. \tag{3.3}$$

Let $V = [H_0^1(\Omega)]^2$. Also, for a subdomain $D \subset \Omega$, we denote the restriction of V on D by $V(D)$, and the subspace of $V(D)$ with zero trace on ∂D by $V_0(D)$. The weak formulation of 3.1 then reads: find $p = (p_1, p_2)$ such that $p(t, \cdot) \in V_0$ and

$$c\left(\frac{\partial p}{\partial t}, v\right) + a_Q(p, v) = (f, v), \tag{3.4}$$

for all $v = (v_1, v_2)$ with $v(t, \cdot) \in V_0$. Here, $(\cdot, \cdot)$ denotes the standard $L^2(\Omega)$ inner product. Moreover, the bilinear forms are defined as:

$$\begin{aligned}
c_i(p_i, v_i) &= \int_{D_m^i} c_i^m p_i v_i \, dx + \sum_l \int_{D_{f,l}^i} c_{l,i}^f p_i v_i \, dx = \int_\Omega c_i(x) p_i v_i \, dx, \\
c(p, v) &= \sum_i c_i(p_i, v_i), \\
a_i(p_i, v_i) &= \int_{D_m^i} \kappa_i^m \nabla p_i \cdot \nabla v_i \, dx + \sum_l \int_{D_{f,l}^i} \kappa_{l,i}^f \nabla p_i \cdot \nabla v_i \, dx = \int_\Omega \kappa_i(x) \nabla p_i \cdot \nabla v_i \, dx, \\
a(p, v) &= \sum_i a_i(p_i, v_i), \\
q(p, v) &= \sum_i \sum_l \int_\Omega \sigma(p_i - p_l) v_i \, dx, \\
a_Q(p, v) &= a(p, v) + q(p, v), \quad (f, v) = \sum_i (f_i, v_i),
\end{aligned} \tag{3.5}$$

3.2 Method description

In this section, we will describe the details of our proposed method. To start with, we introduce the notions of coarse and fine meshes. We start with a usual partition $\mathcal{T}^H$ of Ω into finite elements, which does not necessarily resolve any multiscale features. The partition $\mathcal{T}^H$ is called a coarse grid and a generic element K in the partition $\mathcal{T}^H$ is called a coarse element. Moreover, $H > 0$ is called the coarse mesh size. We let N_c be the number of coarse grid nodes and N be the number of coarse elements. We also denote the collection of all coarse grid edges by $\mathcal{E}^H$. We perform a refinement of $\mathcal{T}^H$ to obtain a fine grid $\mathcal{T}^h$, where $h > 0$ is called the fine mesh size. It is assumed

that the fine grid is sufficiently fine to resolve the solution. An illustration of the fine grid and the coarse grid and a coarse element are shown in Figure 2.1. We remark that the fine grid is only used in solving local problems numerically. In our analysis, the fine grid does not play a role as we assume that all local problems are solved continuously.

We define local bilinear forms on a coarse element K_j by:

$$\begin{aligned} a_i^{(j)}(p_i, v_i) &= \int_{K_j} \kappa_i(x) \nabla p_i \cdot \nabla v_i \, dx, \\ a^{(j)}(p, v) &= \sum_i a_i^{(j)}(p_i, v_i), \\ q^{(j)}(p, v) &= \sum_i \sum_l \int_{K_j} \sigma(p_i - p_l) v_i \, dx, \\ a_Q^{(j)}(p, v) &= a^{(j)}(p, v) + q^{(j)}(p, v), \\ s_i^{(j)}(p_i, v_i) &= \int_{K_j} \widetilde{\kappa}_i(x) p_i v_i \, dx, \\ s^{(j)}(p, v) &= \sum_i s_i^{(j)}(p_i, v_i), \end{aligned} \tag{3.6}$$

where $\widetilde{\kappa}_i = \kappa_i \sum_{k=1}^{N_c} |\nabla \chi_k|^2$ and $\{\chi_k\}$ is a set of bilinear partition of unity functions for the coarse grid partition of the domain Ω. We also define the bilinear form s by:

$$s(p, v) = \sum_j s^{(j)}(p, v). \tag{3.7}$$

Next, we will use the concept of GMsFEM to construct our auxiliary multiscale basis functions. The auxiliary basis functions are coupled, and defined by a spectral problem, which is to find a real number $\lambda_k^{(j)}$ and a function $\phi_k^{(j)} \in V(K_j)$ such that

$$a_Q^{(j)}(\phi_k^{(j)}, v) = \lambda_k^{(j)} s^{(j)}(\phi_k^{(j)}, v) \text{ for all } v \in V(K_j). \tag{3.8}$$

We let $\lambda_k^{(j)}$ be the eigenvalues of 3.8 arranged in ascending order in k, normalize the eigenfunctions in the norm induced by the inner product s, and use the first L_j eigenfunctions to construct our local

auxiliary multiscale space

$$V_{aux}^{(j)} = \text{span}\{\phi_k^{(j)} : 1 \leq k \leq L_j\}. \tag{3.9}$$

The global auxiliary multiscale space V_{aux} is then defined as the sum of these local auxiliary multiscale spaces

$$V_{aux} = \oplus_{j=1}^{N} V_{aux}^{(j)}. \tag{3.10}$$

Before we move on to discuss the construction of multiscale basis functions, we introduce some tools which will be used to describe our method and analyze the convergence. We first introduce the notion of ϕ-orthogonality. In a coarse block K_j, given an auxiliary basis function $\phi_k^{(j)} \in V_{aux}$, we say that $\psi \in V$ is $\phi_k^{(j)}$-orthogonal if

$$s\left(\psi, \phi_{k'}^{(j')}\right) = \delta_{j,j'}\delta_{k,k'} \text{ for } 1 \leq k' \leq L_{j'} \text{ and } 1 \leq j' \leq N. \tag{3.11}$$

We also introduce a projection operator $\pi : [L^2(\Omega)]^2 \to V_{aux}$ by $\pi = \sum_{j=1}^{N} \pi_j$, where $\pi_j : [L^2(K_j)]^2 \to V_{aux}$ is given by

$$\pi_j(v) = \sum_{k=1}^{L_j} \frac{s^{(j)}(v, \phi_k^{(j)})}{s^{(j)}(\phi_k^{(j)}, \phi_k^{(j)})} \phi_k^{(j)} \text{ for all } v \in [L^2(K_j)]^2. \tag{3.12}$$

Next, we construct our global multiscale basis functions. The global multiscale basis function $\psi_j^{(i)} \in V$ is defined as the solution of the following constrained energy minimization problem

$$\psi_k^{(j)} = \text{argmin}\left\{a_Q(\psi, \psi) : \psi \in V \text{ is } \phi_k^{(j)}\text{-orthogonal}\right\}. \tag{3.13}$$

The minimization problem 3.13 is equivalent to the following variational problem: find $\psi_k^{(j)} \in V$ and $\mu_k^{(j)} \in V_{aux}$ such that

$$\begin{aligned} a_Q(\psi_k^{(j)}, w) + s(w, \mu_k^{(j)}) &= 0 \text{ for all } w \in V, \\ s(\psi_k^{(j)} - \phi_k^{(j)}, \nu) &= 0 \text{ for all } \nu \in V_{aux}. \end{aligned} \tag{3.14}$$

Motivated by the construction of global multiscale basis functions, we define our localized multiscale basis functions. For each element K_j, an oversampled domain formed by enlarging the coarse grid block K_j by m coarse grid layers. An illustration of an oversampled domain is shown in Figure 2.2. The localized multiscale basis function $\psi_{k,ms}^{(j)} \in V_0(K_{j,m})$ is defined as the solution of the following constrained energy minimization problem

$$\psi_{k,ms}^{(j)} = \text{argmin}\left\{a_Q(\psi,\psi) : \psi \in V_0(K_{j,m}) \text{ is } \phi_k^{(j)}\text{-orthogonal}\right\}. \tag{3.15}$$

The minimization problem 3.15 is equivalent to the following variational problem: find $\psi_{k,ms}^{(j)} \in V_0(K_{j,m})$ and $\mu_{k,ms}^{(j)} \in \oplus_{K_{j'}\subset K_{j,m}} V_{aux}^{(j')}$ such that

$$\begin{aligned} a_Q(\psi_{k,ms}^{(j)}, w) + s(w, \mu_{k,ms}^{(j)}) &= 0 \text{ for all } w \in V_0(K_{j,m}), \\ s(\psi_{k,ms}^{(j)} - \phi_k^{(j)}, \nu) &= 0 \text{ for all } \nu \in \oplus_{K_{j'}\subset K_{j,m}} V_{aux}^{(j')}. \end{aligned} \tag{3.16}$$

We use the localized multiscale basis functions to construct the multiscale finite element space, which is defined as

$$V_{ms} = \text{span}\{\psi_{k,ms}^{(j)} : 1 \le k \le L_j, 1 \le j \le N\}. \tag{3.17}$$

The multiscale solution is then given by: find $p_{ms} = (p_{ms,1}, p_{ms,2})$ with $p_{ms}(t,\cdot) \in V_{ms}$ such that for all $v = (v_1, v_2)$ with $v(t,\cdot) \in V_{ms}$,

$$c\left(\frac{\partial p_{ms}}{\partial t}, v\right) + a_Q(p_{ms}, v) = (f, v). \tag{3.18}$$

3.3 Convergence Analysis

In this section, we will analyze the proposed method. First, we define the following norms and semi-norms on V:

$$\begin{aligned} \|p\|_c^2 &= c(p,p), \\ \|p\|_a^2 &= a(p,p), \\ |p|_q^2 &= q(p,p), \\ \|p\|_{a_Q}^2 &= a_Q(p,p), \\ \|p\|_s^2 &= s(p,p). \end{aligned} \tag{3.19}$$

For a subdomain $D = \bigcup_{j\in J} K_j$ composed by a union of coarse grid blocks, we also define the following local norms and semi-norms on V:

$$\begin{aligned} \|p\|_{a(D)}^2 &= \sum_{j\in J} a^{(j)}(p,p), \\ |p|_{q(D)}^2 &= \sum_{j\in J} q^{(j)}(p,p), \\ \|p\|_{a_Q(D)}^2 &= \sum_{j\in J} a_Q^{(j)}(p,p), \\ \|p\|_{s(D)}^2 &= \sum_{j\in J} s^{(j)}(p,p). \end{aligned} \tag{3.20}$$

The flow of our analysis goes as follows. First, we prove the convergence using the global multiscale basis functions. With the global multiscale basis functions constructed, the global multiscale finite element space is defined by

$$V_{glo} = \text{span}\{\psi_k^{(j)} : 1 \le k \le L_j, 1 \le j \le N\}, \tag{3.21}$$

and an approximated solution $p_{glo} = (p_{glo,1}, p_{glo,2})$, where $p_{glo}(t, \cdot) \in V_{glo}$, is given by

$$c\left(\frac{\partial p_{glo}}{\partial t}, v\right) + a_Q(p_{glo}, v) = (f, v), \tag{3.22}$$

for all $v = (v_1, v_2)$ with $v(t, \cdot) \in V_{glo}$. Next, we give an estimate of the difference between the global multiscale functions $\psi_k^{(j)}$ and the local multiscale basis functions $\psi_{k,ms}^{(j)}$, in order to show that using the multiscale solution p_{ms} provides similar convergence results as the global solution p_{glo}. For this purpose, we denote the kernel of the projection operator π by $\widetilde{V}$. Then, for any $\psi_k^{(j)} \in V_{glo}$, we have

$$a_Q(\psi_k^{(j)}, w) = 0 \text{ for all } w \in \widetilde{V}, \tag{3.23}$$

which implies $\widetilde{V} \subseteq V_{glo}^{\perp}$, where $V_{glo}^{\perp}$ is the orthogonal complement of V_{glo} with respect to the inner product a_Q. Moreover, since $\dim(V_{glo}) = \dim(V_{aux})$, we have $\widetilde{V} = V_{glo}^{\perp}$ and $V = V_{glo} \oplus \widetilde{V}$.

In addition, we introduce some operators which will be used in our analysis, namely $R_{glo} : V \to V_{glo}$ given by: for any $u \in V$, the image $R_{glo}u \in V_{glo}$ is defined by

$$a_Q(R_{glo}u, v) = a_Q(u, v) \text{ for all } v \in V_{glo}, \tag{3.24}$$

and similarly, $R_{ms} : V \to V_{ms}$ given by: for any $u \in V$, the image $R_{ms}u \in V_{ms}$ is defined by

$$a_Q(R_{ms}u, v) = a_Q(u, v) \text{ for all } v \in V_{ms}. \tag{3.25}$$

We also define $\mathcal{C} : V \to V$ given by: for any $u \in V$, the image $\mathcal{C}u \in V$ is defined by

$$(\mathcal{C}u, v) = c(u, v) \text{ for all } v \in V. \tag{3.26}$$

Moreover, the operator $\mathcal{A} : D(\mathcal{A}) \to [L^2(\Omega)]^2$ is defined on a subspace $D(\mathcal{A}) \subset V$ by: for any

$u \in D(\mathcal{A})$, the image $\mathcal{A}u \in [L^2(\Omega)]^2$ is defined by

$$(\mathcal{A}u, v) = a_Q(u, v) \text{ for all } v \in V. \tag{3.27}$$

We will first show the projection operator R_{glo} onto global multiscale finite element space has a good approximation property with respect to the a_Q-norm and L^2-norm.

Lemma 3.3.1. *Let $u \in D(\mathcal{A})$. Then we have $u - R_{glo}u \in \widetilde{V}$ and*

$$\|u - R_{glo}u\|_{a_Q} \leq CH\underline{\kappa}^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}\|\mathcal{A}u\|_{[L^2(\Omega)]^2}, \tag{3.28}$$

and

$$\|u - R_{glo}u\|_{[L^2(\Omega)]^2} \leq CH^2\underline{\kappa}^{-1}\Lambda^{-1}\|\mathcal{A}u\|_{[L^2(\Omega)]^2}, \tag{3.29}$$

where

$$\Lambda = \min_{1\leq j\leq N} \lambda_{L_j+1}^{(j)}. \tag{3.30}$$

Proof. From 3.24, we see that $u - R_{glo}u \in V_{glo}^{\perp} = \widetilde{V}$. Taking $v = R_{glo}u \in V_{glo}$ in 3.24, we have

$$a_Q(u - R_{glo}u, R_{glo}u) = 0. \tag{3.31}$$

Therefore, we have

$$\begin{aligned}
\|u - R_{glo}u\|_{a_Q}^2 &= a_Q(u - R_{glo}u, u - R_{glo}u) \\
&= a_Q(u - R_{glo}u, u) \\
&= a_Q(u, u - R_{glo}u) \\
&= (\mathcal{A}u, u - R_{glo}u) \\
&\leq \|\widetilde{\kappa}^{-\frac{1}{2}}\mathcal{A}u\|_{[L^2(\Omega)]^2}\|u - R_{glo}u\|_s,
\end{aligned} \tag{3.32}$$

where $\widetilde{\kappa}(x) = \min\{\widetilde{\kappa}_i(x), \widetilde{\kappa}_{l,i}(x)\}$. Since $u - R_{glo}u \in \widetilde{V}$, we have $\pi_j(u - R_{glo}u) = 0$ for all

$j = 1, 2, \ldots, N$ and

$$\begin{aligned} \|u - R_{glo}u\|_s^2 &= \sum_{j=1}^{N} \|u - R_{glo}u\|_{s(K_j)}^2 \\ &= \sum_{j=1}^{N} \|(I - \pi_j)(u - R_{glo}u)\|_{s(K_j)}^2. \end{aligned} \tag{3.33}$$

By the orthogonality of the eigenfunctions $\phi_k^{(j)}$, we have

$$\sum_{j=1}^{N} \|(I - \pi_j)(u - R_{glo}u)\|_{s(K_j)}^2 \leq \frac{1}{\Lambda} \sum_{j=1}^{N} \|u - R_{glo}u\|_{a_Q(K_j)}^2 \leq \frac{1}{\Lambda} \|u - R_{glo}u\|_{a_Q}^2. \tag{3.34}$$

Finally, using the fact that $|\nabla \chi_k| = O(H^{-1})$, we obtain the first estimate 3.28.

For the second estimate 3.29, we use a duality argument. Define $w \in V$ by

$$a_Q(w, v) = (u - R_{glo}u, v) \text{ for all } v \in V. \tag{3.35}$$

Then we have

$$\|u - R_{glo}u\|_{[L^2(\Omega)]^2}^2 = (u - R_{glo}u, u - R_{glo}u) = a_Q(w, u - R_{glo}u). \tag{3.36}$$

Taking $v = R_{glo}w \in V_{glo}$ in 3.24, we have

$$a_Q(u - R_{glo}u, R_{glo}w) = 0. \tag{3.37}$$

Note that $w \in D(\mathcal{A})$ and $\mathcal{A}w = u - R_{glo}u$. Hence

$$\begin{aligned}\|u - R_{glo}u\|^2_{[L^2(\Omega)]^2} &= a_Q(w - R_{glo}w, u - R_{glo}u) \\ &\leq \|w - R_{glo}w\|_{a_Q}\|u - R_{glo}u\|_{a_Q} \\ &\leq \left(CH\underline{\kappa}^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}\|\mathcal{A}w\|_{[L^2(\Omega)]^2}\right)\left(CH\underline{\kappa}^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}\|\mathcal{A}u\|_{[L^2(\Omega)]^2}\right) \\ &\leq CH^2\underline{\kappa}^{-1}\Lambda^{-1}\|u - R_{glo}u\|_{[L^2(\Omega)]^2}\|\mathcal{A}u\|_{[L^2(\Omega)]^2}.\end{aligned} \tag{3.38}$$

□

We remark that the quantity Λ is contrast independent as we include all eigenfunctions corresponding to small contrast dependent eigenvalues in our basis construction.

We are now going to prove the global basis functions are localizable. For each coarse block K, we define B to be a bubble function with $B(x) > 0$ for all $x \in \text{int}(K)$ and $B(x) = 0$ for all $x \in \partial K$. We will take $B = \prod_j \chi_j^{ms}$ where the product is taken over all vertices j on the boundary of K, and $\{\chi_j\}$ is a set of bilinear partition of unity functions for the coarse grid partition of the domain Ω. Using the bubble function, we define the constant

$$C_\pi = \sup_{K\in\mathcal{T}^H,\nu\in V_{aux}} \frac{s(\nu,\nu)}{s(B\nu,\nu)}. \tag{3.39}$$

We also define

$$\lambda_{max} = \max_{1\leq j\leq N}\max_{1\leq k\leq L_j} \lambda_k^{(j)}. \tag{3.40}$$

Lemma 3.3.2. *For all $v_{aux} \in V_{aux}$, there exists a function $v \in V$ such that*

$$\pi(v) = v_{aux}, \quad \|v\|^2_{a_Q} \leq D\|v_{aux}\|^2_s, \quad \textit{supp}(v) \subset \textit{supp}(v_{aux}). \tag{3.41}$$

We write $D = 2(1 + 2C_p^2\sigma\underline{\kappa}^{-1})(C_{\mathcal{T}} + \lambda^2_{max})$, where $C_{\mathcal{T}}$ is the square of the maximum number of vertices over all coarse elements, and C_p is a Poincaré constant.

Proof. Let $v_{aux} \in V_{aux}^{(j)}$ with $\|v_{aux}\|_{s(K_j)} = 1$. We consider the following minimization problem

defined on a coarse block K_j.

$$v = \text{argmin}\left\{a_Q(\psi,\psi)\ :\ \psi \in V_0(K_j), \quad s^{(j)}(\psi,\nu) = s^{(j)}(v_{aux},\nu) \text{ for all } \nu \in V_{aux}^{(j)}\right\}. \tag{3.42}$$

We will show that the minimization problem 3.42 has a unique solution. First, we note that the minimization problem 3.42 is equivalent to the following variational problem: find $v \in V_0(K_j)$ and $\mu \in V_{aux}^{(j)}$ such that

$$\begin{aligned} a_Q^{(j)}(v,w) + s^{(j)}(w,\mu) &= 0 \text{ for all } w \in V_0(K_j), \\ s^{(j)}(v - v_{aux},\nu) &= 0 \text{ for all } \nu \in V_{aux}^{(j)}. \end{aligned} \tag{3.43}$$

The well-posedness of 3.43 is equivalent to the existence of $v \in V_0(K_j)$ such that

$$s^{(j)}(v, v_{aux}) \geq C\|v_{aux}\|^2_{s(K_j)}, \quad \|v\|_{a_Q(K_j)} \leq C\|v_{aux}\|_{s(K_j)}, \tag{3.44}$$

where C is a constant to be determined. Now, we take $v = Bv_{aux} \in V_0(K_j)$. Then we have

$$s^{(j)}(v, v_{aux}) = s^{(j)}(Bv_{aux}, v_{aux}) \geq C_\pi^{-1} s\|v_{aux}\|^2_{s(K_j)}. \tag{3.45}$$

On the other hand, since $\nabla v_i = \nabla(Bv_{aux,i}) = v_{aux,i}\nabla B + B\nabla v_{aux,i}$, $|B| \leq 1$ and $|\nabla B|^2 \leq C_{\mathcal{T}} \sum_k |\nabla \chi_k^{ms}|^2$, we have

$$\|v\|^2_{a(K_j)} \leq 2(C_{\mathcal{T}}\|v_{aux}\|^2_{s(K_j)} + \|v_{aux}\|^2_{a_Q(K_j)}). \tag{3.46}$$

By the spectral problem 3.8, we have

$$\|v_{aux}\|_{a_Q(K_j)} \leq \max_{1\leq k\leq L_j} \lambda_k^{(j)} \|v_{aux}\|_{s(K_j)}. \tag{3.47}$$

Moreover, by Poincaré inequality, we have

$$|v|_q^2 \leq 2\sigma \|v\|_{L^2(K_j)}^2 \leq 2C_p^2 \sigma \underline{\kappa}^{-1} \|v\|_{a(K_j)}^2. \tag{3.48}$$

Combining these estimates, we have

$$\|v\|_{a_Q(K_j)}^2 \leq (1 + 2C_p^2 \sigma \underline{\kappa}^{-1}) \|v\|_{a(K_j)}^2 \leq 2(1 + 2C_p^2 \sigma \underline{\kappa}^{-1})(C_{\mathcal{T}} + \lambda_{max}^2) \|v_{aux}\|_{s(K_j)}^2. \tag{3.49}$$

This shows that the minimization problem 3.42 has a unique solution $v \in V_0(K_j)$, which satisfies our desired properties. □

Here, we make a remark that we can assume $D \geq 1$ without loss of generality.

In order to estimate the difference between the global basis functions and localized basis functions, we need the notion of a cutoff function with respect to the oversampling regions. For each coarse grid K_j and $M > m$, we define $\chi_j^{M,m} \in \text{span}\{\chi_k^{ms}\}$ such that $0 \leq \chi_j^{M,m} \leq 1$ and $\chi_j^{M,m} = 1$ on the inner region $K_{j,m}$ and $\chi_j^{M,m} = 0$ outside the region $K_{j,M}$.

The following lemma shows that our multiscale basis functions have a decay property. In particular, the global basis functions are small outside an oversampled region specified in the lemma, which is important in localizing the multiscale basis functions.

Lemma 3.3.3. *Given $\phi_k^{(j)} \in V_{aux}^{(j)}$ and an oversampling region $K_{j,m}$ with number of layers $m \geq 2$. Let $\psi_{k,ms}^{(j)}$ be a localized multiscale basis function defined on $K_{j,m}$ given by 3.15, and $\psi_k^{(j)}$ be the corresponding global basis function given by 3.13. Then we have*

$$\|\psi_k^{(j)} - \psi_{k,ms}^{(j)}\|_{a_Q}^2 \leq E \|\phi_k^{(j)}\|_{s(K_j)}^2, \tag{3.50}$$

where $E = 24D^2(1 + \Lambda^{-1}) \left(1 + \dfrac{\Lambda^{\frac{1}{2}}}{2D^{\frac{1}{2}}}\right)^{1-m}$.

Proof. By Lemma 3.3.2, there exists $\widetilde{\phi}_k^{(j)} \in V$ such that

$$\pi(\widetilde{\phi}_k^{(j)}) = \phi_k^{(j)}, \quad \|\widetilde{\phi}_k^{(j)}\|_{a_Q}^2 \leq D\|\phi_k^{(j)}\|_s^2, \quad \text{supp}(\widetilde{\phi}_k^{(j)}) \subset K_j. \tag{3.51}$$

We take $\eta = \psi_k^{(j)} - \widetilde{\phi}_k^{(j)} \in V$ and $\zeta = \widetilde{\phi}_k^{(j)} - \psi_{k,ms}^{(j)} \in V_0(K_{j,m})$. Then $\pi(\eta) = \pi(\zeta) = 0$ and hence $\eta, \zeta \in \widetilde{V}$. Again, by Lemma 3.3.2, there exists $\beta \in V$ such that

$$\pi(\beta) = \pi(\chi_j^{m,m-1}\eta), \quad \|\beta\|_{a_Q}^2 \leq D\|\pi(\chi_j^{m,m-1}\eta)\|_s^2, \quad \text{supp}(\beta) \subset K_{j,m} \setminus K_{j,m-1}. \tag{3.52}$$

Take $\tau = \beta - \chi_j^{m,m-1}\eta \in V_0(K_{j,m})$. Again, $\pi(\tau) = 0$ and hence $\tau \in \widetilde{V}$. Now, by the variational problems 3.14 and 3.16, we have

$$a_Q(\psi_k^{(j)} - \psi_{k,ms}^{(j)}, w) + s(w, \mu_k^{(j)} - \mu_{k,ms}^{(j)}) = 0 \text{ for all } w \in V_0(K_{j,m}). \tag{3.53}$$

Taking $w = \tau - \zeta \in V_0(K_{j,m})$ and using the fact that $\tau - \zeta \in \widetilde{V}$, we have

$$a_Q(\psi_k^{(j)} - \psi_{k,ms}^{(j)}, \tau - \zeta) = 0, \tag{3.54}$$

which implies

$$\begin{aligned} \|\psi_k^{(j)} - \psi_{k,ms}^{(j)}\|_{a_Q}^2 &= a_Q(\psi_k^{(j)} - \psi_{k,ms}^{(j)}, \psi_k^{(j)} - \psi_{k,ms}^{(j)}) \\ &= a_Q(\psi_k^{(j)} - \psi_{k,ms}^{(j)}, \eta + \zeta) \\ &= a_Q(\psi_k^{(j)} - \psi_{k,ms}^{(j)}, \eta + \tau) \\ &\leq \|\psi_k^{(j)} - \psi_{k,ms}^{(j)}\|_{a_Q}\|\eta + \tau\|_{a_Q}. \end{aligned} \tag{3.55}$$

Therefore, we have

$$\begin{aligned}\|\psi_k^{(j)} - \psi_{k,ms}^{(j)}\|_{a_Q}^2 &\leq \|\eta + \tau\|_{a_Q}^2 \\ &= \|(1 - \chi_j^{m,m-1})\eta + \beta\|_{a_Q}^2 \\ &\leq 2\left(\|(1 - \chi_j^{m,m-1})\eta\|_{a_Q}^2 + \|\beta\|_{a_Q}^2\right).\end{aligned} \tag{3.56}$$

For the first term on the right hand side of 3.56, since $\nabla\left((1 - \chi_j^{m,m-1})\eta_i\right) = (1 - \chi_j^{m,m-1})\nabla\eta_i - \eta_i\nabla\chi_j^{m,m-1}$ and $|1 - \chi_j^{m,m-1}| \leq 1$, we have

$$\|(1 - \chi_j^{m,m-1})\eta_i\|_{a_i}^2 \leq 2\left(\|\eta_i\|_{a_i(\Omega\setminus K_{j,m-1})}^2 + \|\eta_i\|_{s_i(\Omega\setminus K_{j,m-1})}^2\right). \tag{3.57}$$

On the other hand, we have

$$|(1 - \chi_j^{m,m-1})\eta|_q^2 \leq |\eta|_{q(\Omega\setminus K_{j,m-1})}^2. \tag{3.58}$$

For the second term on the right hand side of 3.56, we first see that for $K_{j'} \subset K_{j,m-1}$,

$$s\left(\chi_j^{m,m-1}\eta, \phi_k^{(j')}\right) = s^{(j')}\left(\chi_j^{m,m-1}\eta, \phi_k^{(j')}\right) = s^{(j')}\left(\eta, \phi_k^{(j')}\right) = 0, \tag{3.59}$$

since $\chi_j^{m,m-1} = 1$ on $K_{j,m-1}$ and $\eta \in \widetilde{V}$. On the other hand, for $K_{j'} \subset \Omega \setminus K_{j,m}$,

$$s\left(\chi_j^{m,m-1}\eta, \phi_k^{(j')}\right) = s^{(j')}\left(\chi_j^{m,m-1}\eta, \phi_k^{(j')}\right) = 0, \tag{3.60}$$

since $\chi_j^{m,m-1} = 0$ on $\Omega \setminus K_{j,m}$. Therefore, we have supp $\left(\pi(\chi_j^{m,m-1}\eta)\right) \subset K_{j,m} \setminus K_{j,m-1}$. Using 3.52 and $|\chi_j^{m,m-1}| \leq 1$, we have

$$\|\beta\|_{a_Q}^2 \leq D\|\pi(\chi_j^{m,m-1}\eta)\|_{s(K_{j,m}\setminus K_{j,m-1})}^2 \leq D\|\chi_j^{m,m-1}\eta\|_{s(K_{j,m}\setminus K_{j,m-1})}^2 \leq D\|\eta\|_{s(K_{j,m}\setminus K_{j,m-1})}^2. \tag{3.61}$$

Since $\eta \in \widetilde{V}$, by the spectral problem 3.8, we obtain

$$\|\eta\|^2_{s(K_{j,m}\setminus K_{j,m-1})} \leq \Lambda^{-1}\|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-1})}. \tag{3.62}$$

Combining these estimates, we have

$$\|\psi_k^{(j)} - \psi_{k,ms}^{(j)}\|^2_{a_Q} \leq (4+4\Lambda^{-1}+2D\Lambda^{-1})\|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-1})} \leq 6D(1+\Lambda^{-1})\|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-1})}. \tag{3.63}$$

Next, we will prove a recursive estimate for $\|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-1})}$. We take $\xi = 1 - \chi_j^{m-1,m-2}$. Then $\xi = 1$ in $\Omega \setminus K_{j,m-1}$ and $0 \leq \xi \leq 1$. Hence, using $\nabla(\xi^2\eta_i) = \xi^2\nabla\eta_i + 2\xi\eta_i\nabla\xi$, we have

$$|\xi\eta|^2_a = a(\eta, \xi^2\eta) + \|\eta\|^2_{s(K_{j,m-1}\setminus K_{j,m-2})}, \tag{3.64}$$

which results in

$$\|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-1})} \leq \|\xi\eta\|^2_{a_Q} \leq a_Q(\eta, \xi^2\eta) + \|\eta\|^2_{s(K_{j,m-1}\setminus K_{j,m-2})}. \tag{3.65}$$

We will estimate the first term on the right hand side of 3.65. First, we note that, for any coarse element $K_{j'} \subset \Omega \setminus K_{j,m-1}$, since $\xi = 1$ in $K_{j'}$ and $\eta \in \widetilde{V}$, we have

$$s\left(\xi^2\eta, \phi_{k'}^{(j')}\right) = s\left(\eta, \phi_{k'}^{(j')}\right) = 0 \text{ for all } k' = 1, 2, \ldots, L_{j'}. \tag{3.66}$$

On the other hand, for any coarse element $K_{j'} \subset K_{j,m-2}$, since $\xi = 0$ in $K_{j,m-2}$, we have

$$s\left(\xi^2\eta, \phi_{k'}^{(j')}\right) = 0 \text{ for all } k' = 1, 2, \ldots, L_{j'}. \tag{3.67}$$

Therefore, $\text{supp}(\pi(\xi^2\eta)) \subset K_{j,m-1} \setminus K_{j,m-2}$. By Lemma 3.3.2, there exists $\gamma \in V$ such that

$$\pi(\gamma) = \pi(\xi^2\eta), \quad \|\gamma\|^2_{a_Q} \leq D\|\pi(\xi^2\eta)\|^2_s, \quad \text{supp}(\gamma) \subset K_{j,m-1} \setminus K_{j,m-2}. \tag{3.68}$$

Take $\theta = \xi^2\eta - \gamma$. Again, $\pi(\theta) = 0$ and hence $\theta \in \widetilde{V}$. Therefore, we have

$$a_Q(\psi_k^{(j)}, \theta) = 0. \tag{3.69}$$

Additionally, supp$(\theta) \subset \Omega \setminus K_{j,m-2}$. Recall that, in 3.51, we have supp$(\widetilde{\phi}_k^{(j)}) \subset K_j$. Hence θ and $\widetilde{\phi}_k^{(j)}$ have disjoint supports, and

$$a_Q(\widetilde{\phi}_k^{(j)}, \theta) = 0. \tag{3.70}$$

Therefore, we obtain

$$a_Q(\eta, \theta) = a_Q(\psi_k^{(j)}, \theta) - a_Q(\widetilde{\phi}_k^{(j)}, \theta) = 0. \tag{3.71}$$

Note that $\xi^2\eta = \theta + \gamma$. Using 3.68, we have

$$\begin{aligned} a_Q(\eta, \xi^2\eta) &= a_Q(\eta, \gamma) \\ &\leq \|\eta\|_{a_Q(K_{j,m-1}\setminus K_{j,m-2})} \|\gamma\|_{a_Q(K_{j,m-1}\setminus K_{j,m-2})} \\ &\leq D^{\frac{1}{2}} \|\eta\|_{a_Q(K_{j,m-1}\setminus K_{j,m-2})} \|\pi(\xi^2\eta)\|_{s(K_{j,m-1}\setminus K_{j,m-2})}. \end{aligned} \tag{3.72}$$

For any coarse element $K_{j'} \subset K_{j,m-1} \setminus K_{j,m-2}$, since $\pi(\eta) = 0$, we have

$$\|\pi(\xi^2\eta)\|_{s(K_{j'})} \leq \|\xi^2\eta\|_{s(K_{j'})} \leq \|\eta\|_{s(K_{j'})} \leq \Lambda^{-\frac{1}{2}} \|\eta\|_{a_Q(K_{j'})}. \tag{3.73}$$

Summing up over all $K_{j'} \subset K_{j,m-1} \setminus K_{j,m-2}$, we obtain

$$\|\pi(\xi^2\eta)\|_{s(K_{j,m-1}\setminus K_{j,m-2})} \leq \Lambda^{-\frac{1}{2}} \|\eta\|_{a_Q(K_{j,m-1}\setminus K_{j,m-2})}. \tag{3.74}$$

Hence, the first term on the right hand side of 3.65 can be estimated by

$$a(\eta, \xi^2\eta) \leq D^{\frac{1}{2}} \Lambda^{-\frac{1}{2}} \|\eta\|^2_{a_Q(K_{j,m-1}\setminus K_{j,m-2})}. \tag{3.75}$$

For the second term on the right hand side of 3.65, a similar argument gives supp$(\xi\eta) \subset K_{j,m-1} \setminus$

$K_{j,m-2}$, and

$$\|\eta\|_{s(K_{j,m-1}\setminus K_{j,m-2})} \leq \Lambda^{-\frac{1}{2}}\|\eta\|_{a_Q(K_{j,m-1}\setminus K_{j,m-2})}. \tag{3.76}$$

Putting 3.65, 3.75 and 3.76 together, we have

$$\|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-1})} \leq (1+D^{\frac{1}{2}})\Lambda^{-\frac{1}{2}}\|\eta\|^2_{a_Q(K_{j,m-1}\setminus K_{j,m-2})} \leq 2D^{\frac{1}{2}}\Lambda^{-\frac{1}{2}}\|\eta\|^2_{a_Q(K_{j,m-1}\setminus K_{j,m-2})}. \tag{3.77}$$

Therefore,

$$\|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-2})} = \|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-1})} + \|\eta\|^2_{a_Q(K_{j,m-1}\setminus K_{j,m-2})} \geq \left(1+\frac{\Lambda^{\frac{1}{2}}}{2D^{\frac{1}{2}}}\right)\|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-1})}. \tag{3.78}$$

Inductively, we have

$$\|\eta\|^2_{a_Q(\Omega\setminus K_{j,m-1})} \leq \left(1+\frac{\Lambda^{\frac{1}{2}}}{2D^{\frac{1}{2}}}\right)^{1-m}\|\eta\|^2_{a_Q(\Omega\setminus K_j)} \leq \left(1+\frac{\Lambda^{\frac{1}{2}}}{2D^{\frac{1}{2}}}\right)^{1-m}\|\eta\|^2_{a_Q}. \tag{3.79}$$

Finally, by the energy minimzing property of $\psi_k^{(j)}$ and 3.51, we have

$$\|\eta\|_{a_Q} = \|\psi_k^{(j)} - \widetilde{\phi}_k^{(j)}\|_{a_Q} \leq 2\|\widetilde{\phi}_k^{(j)}\|_{a_Q} \leq 2D^{\frac{1}{2}}\|\phi_k^{(j)}\|_{s(K_j)}. \tag{3.80}$$

Combining 3.63, 3.79 and 3.80, we obtain our desired result. □

The above lemma motivates us to define localized multiscale basis functions in 3.15. The following lemma suggests that, similar to the projection operator R_{glo} onto the global multiscale finite element space, the projection operator R_{ms} onto our localized multiscale finite element space also has a good approximation property with respect to the a_Q-norm and L^2-norm.

Lemma 3.3.4. *Let $u \in D(\mathcal{A})$. Let $m \geq 2$ be the number of coarse grid layers in the oversampling regions in 3.15. If $m = O\left(\log\left(\frac{\overline{\kappa}}{H}\right)\right)$, then we have*

$$\|u - R_{ms}u\|_{a_Q} \leq CH\underline{\kappa}^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}\|\mathcal{A}u\|_{[L^2(\Omega)]^2}, \tag{3.81}$$

and

$$\|u - R_{ms}u\|_{[L^2(\Omega)]^2} \leq CH^2\underline{\kappa}^{-1}\Lambda^{-1}\|\mathcal{A}u\|_{[L^2(\Omega)]^2}. \tag{3.82}$$

Proof. We write $R_{glo}u = \sum_{j=1}^{N}\sum_{k=1}^{L_j}\alpha_k^{(j)}\psi_k^{(j)}$, and define $w = \sum_{j=1}^{N}\sum_{k=1}^{L_j}\alpha_k^{(j)}\psi_{k,ms}^{(j)} \in V_{ms}$. By the Galerkin orthogonality in 3.25, we have

$$\|u - R_{ms}u\|_{a_Q} \leq \|u - w\|_{a_Q} \leq \|u - R_{glo}u\|_{a_Q} + \|R_{glo}u - w\|_{a_Q}. \tag{3.83}$$

Using Lemma 3.3.3, we see that

$$\begin{aligned}
\|R_{glo}u - w\|_{a_Q}^2 &= \left\|\sum_{j=1}^{N}\sum_{k=1}^{L_j}\alpha_k^{(j)}(\psi_k^{(j)} - \psi_{k,ms}^{(j)})\right\|_{a_Q}^2 \\
&\leq (2m+1)^d\sum_{j=1}^{N}\left\|\sum_{k=1}^{L_j}\alpha_k^{(j)}(\psi_k^{(j)} - \psi_{k,ms}^{(j)})\right\|_{a_Q}^2 \\
&\leq E(2m+1)^d\sum_{j=1}^{N}\left\|\sum_{k=1}^{L_j}\alpha_k^{(j)}\phi_k^{(j)}\right\|_{s}^2 \\
&= E(2m+1)^d\|R_{glo}u\|_s^2,
\end{aligned} \tag{3.84}$$

where the last equality follows from the orthogonality of the eigenfunctions in 3.8. Combining 3.83, 3.84, together with 3.28 in Lemma 3.3.1, we have

$$\|u - R_{ms}u\|_{a_Q} \leq CH\underline{\kappa}^{-\frac{1}{2}}\Lambda^{-\frac{1}{2}}\|\mathcal{A}u\|_{[L^2(\Omega)]^2} + E^{\frac{1}{2}}(2m+1)^{\frac{d}{2}}\|R_{glo}u\|_s. \tag{3.85}$$

Next, we are going to estimate $\|R_{glo}u\|_s$. Using the fact that $|\nabla\chi_k| = O(H^{-1})$, we have

$$\|R_{glo}u\|_s^2 \leq CH^{-2}\overline{\kappa}\|R_{glo}u\|_{[L^2(\Omega)]^2}^2. \tag{3.86}$$

Then, by Poincaré inequality, we have

$$\|R_{glo}u\|^2_{[L^2(\Omega)]^2} \leq C_p \underline{\kappa}^{-1} \|R_{glo}u\|^2_{a_Q}. \tag{3.87}$$

By taking $v = R_{glo}u$ in 3.24, we obtain

$$\|R_{glo}u\|^2_{a_Q} = a_Q(u, R_{glo}u) = (\mathcal{A}u, R_{glo}u) \leq CH\underline{\kappa}^{-\frac{1}{2}} \|\mathcal{A}u\|_{[L^2(\Omega)]^2} \|R_{glo}u\|_s. \tag{3.88}$$

Combining these estimates, we have

$$\|R_{glo}u\|_s \leq CH^{-1}\overline{\kappa}\underline{\kappa}^{-\frac{1}{2}} \|\mathcal{A}u\|_{[L^2(\Omega)]^2}. \tag{3.89}$$

To obtain our desired result, we need

$$H^{-2}\overline{\kappa}(2m+1)^{\frac{d}{2}} E^{\frac{1}{2}} = O(1). \tag{3.90}$$

Taking logarithm, we have

$$\log(H^{-2}) + \log(\overline{\kappa}) + \frac{d}{2}\log(2m+1) + \frac{1-m}{2}\log\left(1 + \frac{\Lambda^{\frac{1}{2}}}{3D^{\frac{1}{2}}}\right) = O(1). \tag{3.91}$$

Thus, taking $m = O\left(\log\left(\dfrac{\overline{\kappa}}{H}\right)\right)$ completes the proof of 3.81. The proof of 3.82 follows from a duality argument as in Lemma 3.3.1. □

We are now ready to establish our main theorem, which estimates the error between the solution p and the multiscale solution p_{ms}.

Theorem 3.3.5. *Suppose $f \in [L^2(\Omega)]^2$. Let $m \geq 2$ be the number of coarse grid layers in the oversampling regions in 3.15. Let p be the solution of 3.4 and p_{ms} be the solution of 3.18. If*

$m = O\left(\log\left(\frac{\overline{\kappa}}{H}\right)\right)$, *then we have*

$$\|p(T,\cdot) - p_{ms}(T,\cdot)\|_c^2 + \int_0^T \|p - p_{ms}\|_{a_Q}^2 dt \leq CH^2\underline{\kappa}^{-1}\Lambda^{-1}\left(\|p^0\|_{a_Q}^2 + \int_0^T \|f\|_{[L^2(\Omega)]^2}^2\, dt\right). \tag{3.92}$$

Proof. Taking $v = \dfrac{\partial p}{\partial t}$ in 3.4, we have

$$\left\|\frac{\partial p}{\partial t}\right\|_c^2 + \frac{1}{2}\frac{d}{dt}\|p\|_{a_Q}^2 = \left(f, \frac{\partial p}{\partial t}\right) \leq C\|f\|_{[L^2(\Omega)]^2}^2 + \frac{1}{2}\left\|\frac{\partial p}{\partial t}\right\|_c^2. \tag{3.93}$$

Integrating over $(0,T)$, we have

$$\frac{1}{2}\int_0^T \left\|\frac{\partial p}{\partial t}\right\|_c^2 dt + \frac{1}{2}\|p(T,\cdot)\|_{a_Q}^2 \leq C\left(\|p^0\|_{a_Q}^2 + \int_0^T \|f\|_{[L^2(\Omega)]^2}^2 dt\right). \tag{3.94}$$

Similarly, taking $v = \dfrac{\partial p_{ms}}{\partial t}$ in 3.18 and integrating over $(0,T)$, we have

$$\frac{1}{2}\int_0^T \left\|\frac{\partial p_{ms}}{\partial t}\right\|_c^2 dt + \frac{1}{2}\|p_{ms}(T,\cdot)\|_{a_Q}^2 \leq C\left(\|p^0\|_{a_Q}^2 + \int_0^T \|f\|_{[L^2(\Omega)]^2}^2 dt\right). \tag{3.95}$$

On the other hand, from 3.4, we see that

$$\mathcal{A}p = f - \mathcal{C}\frac{\partial p}{\partial t}, \tag{3.96}$$

and therefore

$$\|\mathcal{A}p\|_{[L^2(\Omega)]^2} \leq C\left(\|f\|_{[L^2(\Omega)]^2} + \left\|\frac{\partial p}{\partial t}\right\|_c\right). \tag{3.97}$$

By the definition of p in 3.4 and p_{ms} in 3.18, for all $v \in V_{ms}, t \in (0,T)$, we have

$$c\left(\frac{\partial(p - p_{ms})}{\partial t}, v\right) + a_Q(p - p_{ms}, v) = 0. \tag{3.98}$$

Therefore, we have

$$\begin{aligned}
&\frac{1}{2}\frac{d}{dt}\|p-p_{ms}\|_c^2+\|p-p_{ms}\|_{a_Q}^2\\
&=c\left(\frac{\partial(p-p_{ms})}{\partial t},p-p_{ms}\right)+a_Q(p-p_{ms},p-p_{ms})\\
&=c\left(\frac{\partial(p-p_{ms})}{\partial t},p-R_{ms}p\right)+a_Q(p-p_{ms},p-R_{ms}p)\\
&\leq\left\|\frac{\partial(p-p_{ms})}{\partial t}\right\|_c\|p-R_{ms}p\|_c+\|p-p_{ms}\|_{a_Q}\|p-R_{ms}p\|_{a_Q}\\
&\leq\left(\left\|\frac{\partial p}{\partial t}\right\|_c+\left\|\frac{\partial p_{ms}}{\partial t}\right\|_c\right)\|p-R_{ms}p\|_c+\frac{1}{2}\|p-p_{ms}\|_{a_Q}^2+\frac{1}{2}\|p-R_{ms}p\|_{a_Q}^2.
\end{aligned}\tag{3.99}$$

Integrating over $(0,T)$ and using 3.97 with Lemma 3.3.4, we have

$$\begin{aligned}
&\frac{1}{2}\|p(T,\cdot)-p_{ms}(T,\cdot)\|_c^2+\frac{1}{2}\int_0^T\|p-p_{ms}\|_{a_Q}^2dt\\
&\leq\int_0^T\left(\left\|\frac{\partial p}{\partial t}\right\|_c+\left\|\frac{\partial p_{ms}}{\partial t}\right\|_c\right)\|p-R_{ms}p\|_c dt+\frac{1}{2}\int_0^T\|p-R_{ms}p\|_{a_Q}^2dt\\
&\leq\left(\int_0^T\left(\left\|\frac{\partial p}{\partial t}\right\|_c+\left\|\frac{\partial p_{ms}}{\partial t}\right\|_c\right)^2dt\right)^{\frac{1}{2}}\left(\int_0^T\|p-R_{ms}p\|_c^2\,dt\right)^{\frac{1}{2}}+\frac{1}{2}\int_0^T\|p-R_{ms}p\|_{a_Q}^2dt\\
&\leq\left(\int_0^T\left(\left\|\frac{\partial p}{\partial t}\right\|_c+\left\|\frac{\partial p_{ms}}{\partial t}\right\|_c\right)^2dt\right)^{\frac{1}{2}}\left(\int_0^T CH^4\underline{\kappa}^{-2}\Lambda^{-2}\left(\|f\|_{[L^2(\Omega)]^2}+\left\|\frac{\partial p}{\partial t}\right\|_c\right)^2dt\right)^{\frac{1}{2}}+\\
&\qquad\int_0^T CH^2\underline{\kappa}^{-1}\Lambda^{-1}\left(\|f\|_{[L^2(\Omega)]^2}+\left\|\frac{\partial p}{\partial t}\right\|_c\right)^2dt\\
&\leq CH^2\underline{\kappa}^{-1}\Lambda^{-1}\int_0^T\left(\left\|\frac{\partial p}{\partial t}\right\|_c^2+\left\|\frac{\partial p_{ms}}{\partial t}\right\|_c^2+\|f\|_{[L^2(\Omega)]^2}^2\right)dt.
\end{aligned}\tag{3.100}$$

Finally, combining 3.94, 3.95 and 3.100, we obtain our desired result. □

3.4 Numerical Examples

In this section, we present two numerical examples. We perform numerical experiments with high-contrast media to see the orders of convergence of our proposed method in energy norm and L^2 norm. We will also study the effects of the number of oversampling layers m on the quality

of the approximations. In all the experiments, we take the spatial domain to be $\Omega = (0,1)^2$ and the fine mesh size to be $h = 1/256$. An example of the media κ_1 and κ_2 used in the experiments is illustrated in FIgure 3.1. In the figure, the contrast values, i.e. the ratio of the maximum and the minimum in Ω, of the media are $\overline{\kappa}_1 = 10^4$ and $\overline{\kappa}_2 = 10^4$. We will also see the effects of the contrast values of the media on the error, while the configurations of the media remain unchanged.

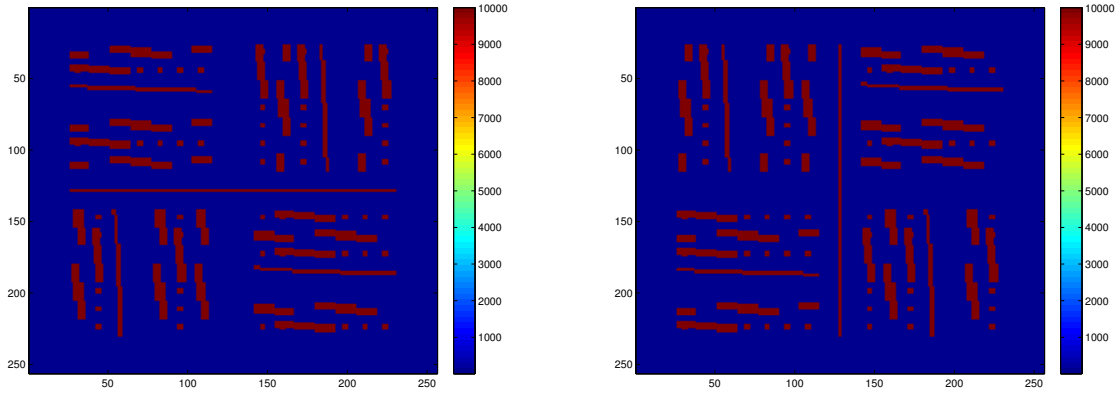

Figure 3.1: Media used in numerical experiments. κ_1 (left) and κ_2 (right). Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

3.4.1 Experiment 1.

In this experiment, we consider the dual continuum model in the steady state, i.e.

$$\begin{aligned} -\text{div}(\kappa_1 \nabla p_1) + \sigma(p_1 - p_2) &= f_1, \\ -\text{div}(\kappa_2 \nabla p_2) - \sigma(p_1 - p_2) &= f_2, \end{aligned} \tag{3.101}$$

where the configuration of the media κ_1 and κ_2 are illustrated in FIgure 3.1. The conductivity values in the background are fixed to be $\kappa_1^m = 1$ and $\kappa_2^m = 1$, while the conductivity values κ_1^f and κ_2^f in the channels are high. The physical parameter for mass transfer is set to be $\sigma = 1$. The

source functions are taken as $f_1(x,y) = 2\pi^2 \sin(\pi x)\sin(\pi y)$ and $f_2(x,y) = 1$ for all $(x,y) \in \Omega$. The steady-state equation 3.101 has a weak formulation: find $p = (p_1, p_2)$ with $p_i \in V$ such that

$$a_Q(p, v) = (f, v), \tag{3.102}$$

for all $v = (v_1, v_2)$ with $v_i \in V$. The numerical solution is then given by: find $p_{ms} = (p_{ms,1}, p_{ms,2})$ with $p_{ms,i} \in V_{ms}$ such that

$$a_Q(p_{ms}, v) = (f, v), \tag{3.103}$$

for all $v = (v_1, v_2)$ with $v_i \in V_{ms}$. In other words, we have $p_{ms} = R_{ms}p$ according to the definition 3.25, and the theoretical orders of convergence follow Lemma 3.3.4.

Figure 3.2 illustrates the numerical solution of the steady-state flow problem. Tables 3.1–3.3 record the error in L^2 norm and a_Q norm with various settings. In Table 3.1, we take the conductivity values in the channels to be $\kappa_1^f = 10^4$ and $\kappa_2^f = 10^6$. We use 6 basis functions per oversampled region since there are 6 small eigenvalues in the spectrum, and according to our analysis, we need to include the first 6 spectral basis functions in the auxiliary space to have good convergence. As we refine coarse mesh size H, we fix the number of oversampling layers to be $m \approx 9\log(1/H)/\log(64)$, which is suggested by our analysis. The results show that the numerical approximations are very accurate, and the errors converge with refinement of the coarse mesh size. Table 3.2 shows the same quantities when the number of basis functions used in each coarse region is reduced to 4. By comparing to Table 3.1, it can be seen that the errors are larger than those when we use 6 basis functions. Figure 3.3 depicts the log-log plot (in exponential base) of L^2 error and energy error against coarse mesh size H. The least-squares fit suggests that we obtain a better convergence order in our numerical experiment compared with the theoretical result. Table 3.3 compares the a_Q error with various combinations of number of layers m and contrast value $\overline{\kappa}$, where the conductivity values in the channels are the same, with 6 basis functions per coarse region and coarse mesh size $H = 1/16$. It can be seen that with a larger oversampled region, the error decreases. On the other hand, the error increases with the contrast value.

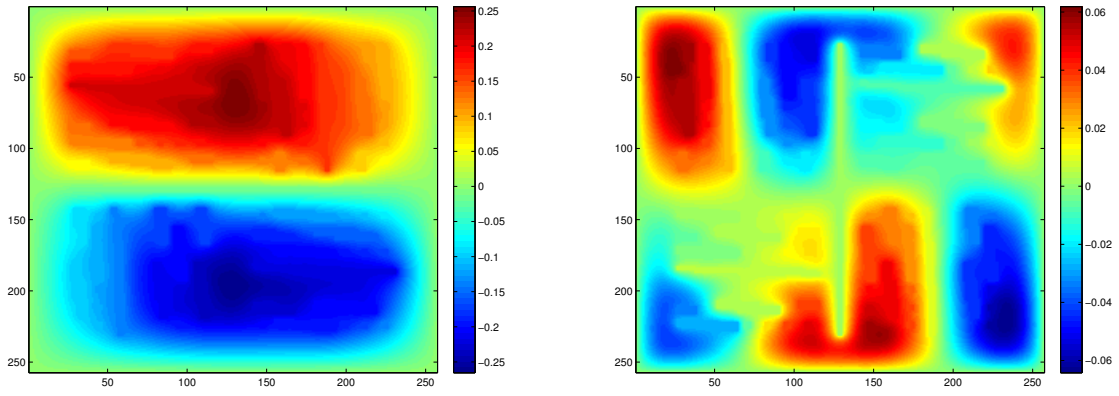

Figure 3.2: Plots of numerical solution: $p_{ms,1}$ (left) and $p_{ms,2}$ (right). Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

H	m	a_Q error	order	L^2 error	order
$1/8$	4	33.4293%	–	15.8783%	–
$1/16$	6	5.7191%	2.55	0.6265%	4.66
$1/32$	7	1.2437%	2.20	0.0504%	3.64
$1/64$	9	0.3585%	1.79	0.0067%	2.91

Table 3.1: History of convergence with 6 basis functions in Experiment 1. Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

3.4.2 Experiment 2.

In this experiment, we consider the time-dependent dual continuum model 3.1. We are interested in finding a numerical approximation in the temporal domain $[0, T]$, where the final time is set to be $T = 5$. The configuration of the media κ_1 and κ_2 are illustrated in FIgure 3.1. The conductivity values in the background are set to be $\kappa_1^m = 10^{-1}$ and $\kappa_2^m = 10^0$, while the values in the channels are taken as $\kappa_1^f = 10^4$ and $\kappa_2^f = 10^6$. The velocities in the background are taken as

H	m	a_Q error	order	L^2 error	order
$1/8$	4	43.9247%	–	34.2923%	–
$1/16$	6	7.7963%	2.49	1.0463%	5.03
$1/32$	7	1.5417%	2.34	0.0709%	3.88
$1/64$	9	0.4993%	1.63	0.0124%	2.52

Table 3.2: History of convergence with 4 basis functions in Experiment 1. Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

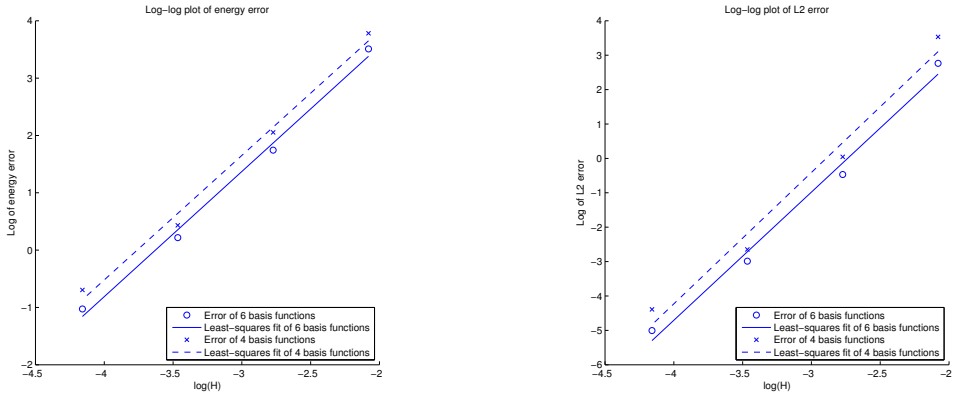

Figure 3.3: Log-Log plot for errors in Experiment 1. Left: energy error; the slope for 6 basis functions is 2.18 and for 4 basis functions is 2.17. Right: L^2 error; the slope for 6 basis functions is 3.73 and for 4 basis functions is 3.82. Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

$c_1^m = 10^1$ and $c_2^m = 10^3$, while the values in the channels are taken as $c_1^f = 10^2$ and $c_2^f = 10^4$. The physical parameter for mass transfer is set to be $\sigma = 25$. The source functions are taken as time-independent, where $f_1(t, x, y) = 0$ for all $(t, x, y) \in [0, T] \times \Omega$ and f_2 is depicted in Figure 3.4. The initial condition is given as $p_1(0, x, y) = 0$ and $p_2(0, x, y) = 0$ for all $(x, y) \in \Omega$.

Figure 3.5 illustrates the numerical solutions at time instants $t = 1.25$, $t = 2.5$ and $t = 5$ respectively. Tables 3.4 records the error in L^2 norm and a_Q norm with 6 basis functions per oversampled region and number of oversampling layers set to be $m \approx 9 \log(1/H)/\log(64)$. Again,

m	$\overline{\kappa} = 10^4$	$\overline{\kappa} = 10^5$	$\overline{\kappa} = 10^6$
3	22.4683%	51.0835%	69.4279%
4	6.3274%	10.1892%	25.6786%
5	5.7205%	5.7978%	6.4329%
6	5.7122%	5.7220%	5.7231%

Table 3.3: Comparison of a_Q error with different number of layers m and contrast value $\overline{\kappa}$ in Experiment 1. Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

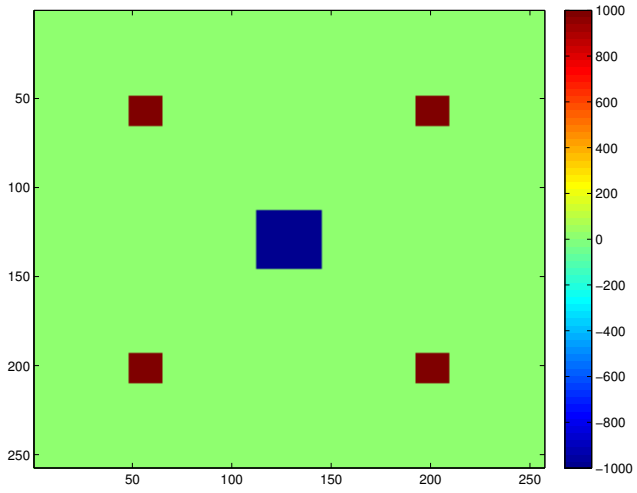

Figure 3.4: Source function f_2 in Experiment 2. Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

the results show that the numerical approximations are very accurate, and the errors converge with with refinement of the coarse mesh size. Figure 3.6 shows the log-log plots of the energy error and L^2 error against coarse mesh size H in exponential base. The least-squares fits again illutstrate our method provides good convergence rates.

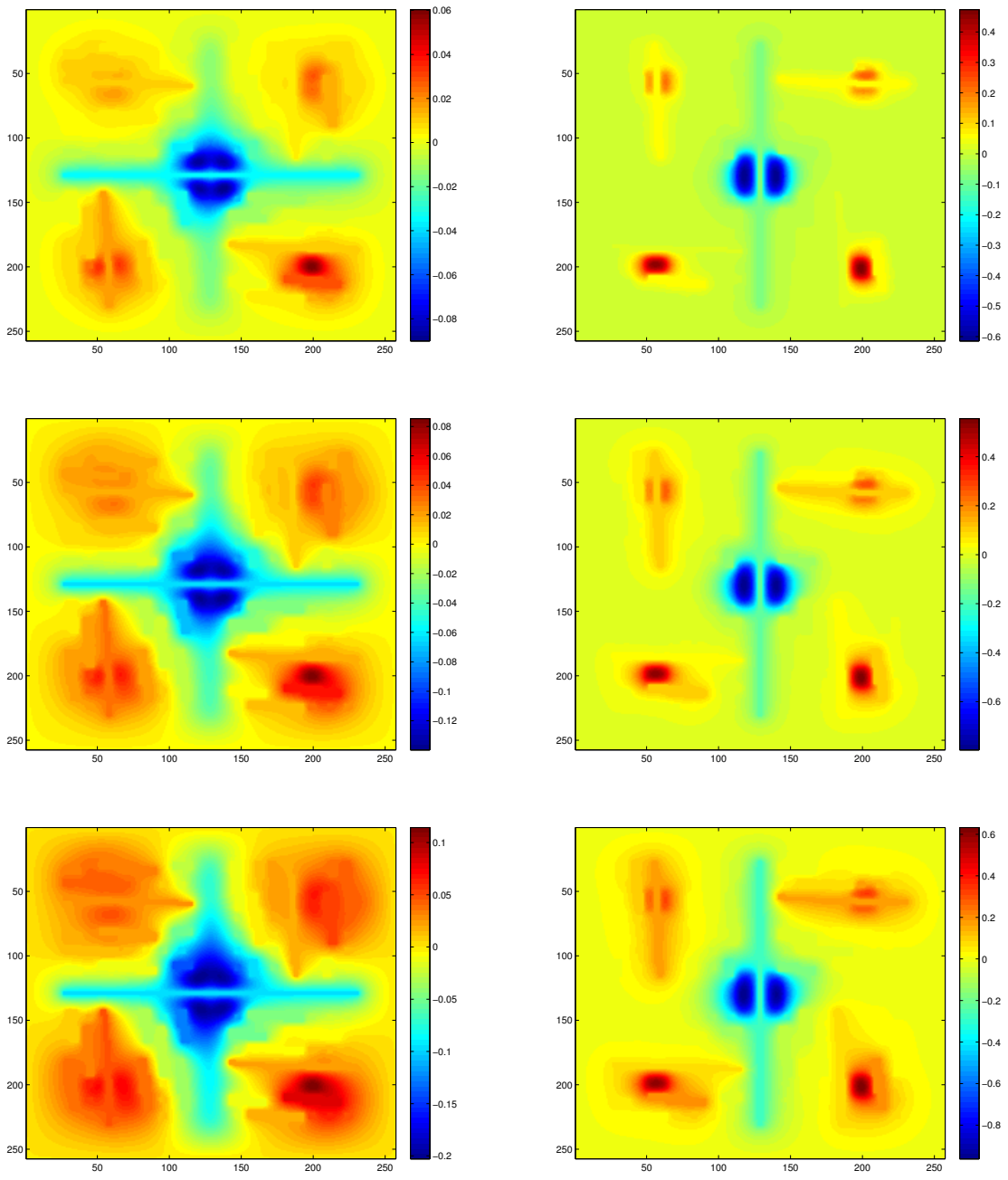

Figure 3.5: Plots of numerical solution at different time instants: $p_{ms,1}$ (left) and $p_{ms,2}$ (right) in Experiment 2. Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

H	m	Δt	a_Q error	order	L^2 error	order
$1/8$	4	1	92.0441%	–	58.6453%	–
$1/16$	6	0.5	20.9725%	2.13	5.2984%	3.47
$1/32$	7	0.25	6.7504%	1.64	0.7718%	2.78
$1/64$	9	0.125	1.9074%	1.82	0.0934%	3.05

Table 3.4: History of convergence with 6 basis functions in Experiment 2. Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

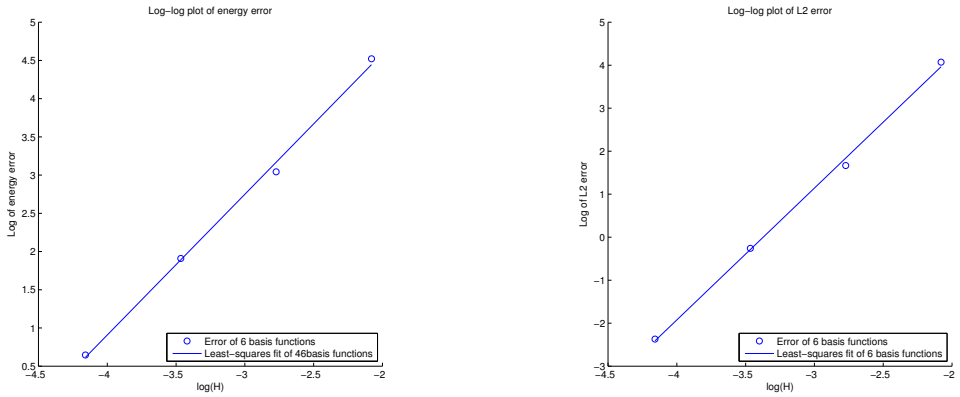

Figure 3.6: Log-Log plot for errors in Experiment 2. Left: energy error; the slope for 6 basis functions is 1.84. Right: L^2 error; the slope for 6 basis functions is 3.07. Reprinted with permission from "Constraint Energy Minimizing Generalized Multiscale Finite Element Method for Dual Continuum Model" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Wing Tat Leung and Maria Vasilyeva. To be published in Communications in Mathematical Sciences by International Press of Boston, Inc.

4. BAYESIAN MUITLSCALE APPROACH FOR MODELING MISSING SUBGRID INFORMATION WITH UNCERTAINTIES AND OBSERVATION DATA *

In many science and engineering applications, such as composite material and porous media, the underlying PDE model may contain high-dimensional coefficient field which varies in multiple scales. Detailed description of the media at the finest scale often comes with uncertainties due to uncertainties. Moreover, limited observational data for the solution may be available. It is therefore desirable to compute realizations of solutions and estimate the associated uncertainties in a probabilistic setting. Through using a Bayesian framework, one can include uncertainties in the media properties and compute the solution and the uncertainties associated with the solution and the variations of the field parameters. An uncertainty band around the solution can be computed. In some applications, there is observational data of the solution available. For example, in reservoir modeling, oil/water pressure data from different well locations can be measured. This observational data can serve as an important information and be used as additional constraints on our solution and basis selection. In practical applications, the accuracy of the data is essential in the quality of the solution. It is therefore desirable to develop methods for regularizing the solution in terms of our quantity of interest.

In our approach, we make use of the advantages of numerical discretization of the underlying PDE by GMsFEM, develop a regression set-up and use Bayesian variable selection techniques to devise a method for posterior modeling and uncertainty quantification. The main ingredients of our method include:

- permanent basis functions – dominant modes in local regions for computing an inexpensive multiscale approximation (called the "fixed" solution),
- additional basis functions – remaining modes in resolving missing subgrid information,

* Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

- prior distribution – residual-based probability distribution for sampling realizations of multiscale solution built around the fixed solution,
- posterior distribution – probability distribution including minimization of residual of the PDE system and mismatch of the dynamic observational data.

We construct local multiscale basis functions using GMsFEM, and use a few basis functions in each local region as permanent basis functions. The remaining multiscale basis functions are categorized as additional basis functions, and are selected stochastically using the residual information. Using the permanent basis functions, a fixed solution is built and the residual is computed, which is used to impose a prior probability on the additional basis functions accordingly. Using a likelihood for penalizing the residual and the mismatch in observational data, we define our posterior probability on the additional basis functions.

The chapter is organized as follows. First, we briefly describe the ideas of GMsFEM in Section 4.1. Next, we discuss our Bayesian formulation in Section 4.2. In Section 4.3, we present numerical examples for our problem.

4.1 Preliminaries

Let Ω be the computational domain. We consider the forward model

$$\frac{\partial u}{\partial t} - \text{div}(\kappa(x,t)\nabla u) = f \quad \text{in } \Omega \times (0,T), \tag{4.1}$$

subject to smooth initial and boundary conditions. Here f is a given source term and $\mathcal{L}$ is a multiscale elliptic differential operator. Using standard numerical discretizations such as finite element or discontinuous Galerkin methods, the fine-scale solution $u_h \in V_h$ can be obtained by solving the variational problem:

$$\int_0^T \int_\Omega \frac{\partial u}{\partial t} v + a(u,v) = \int_0^T \int_\Omega fv \quad \text{for all } v \in V_h, \tag{4.2}$$

in a suitably defined H^1-conforming finite element space V_h depending on the boundary condition. In this work, for the sake of simplicity, we assume homogeneous Dirichlet boundary condition in the numerical examples. The Bayesian approach can be easily extended to other boundary conditions. Here the bilinear form $a(u, v)$ is a symmetric and positive-definite bilinear form defined as

$$a(u, v) = \int_0^T \int_\Omega \kappa \nabla u \cdot \nabla v.$$

However, in practice, the mesh size has to be very small in order to resolve all scales. The resultant linear system is huge and ill-conditioned, and solving such a system is computationally expensive. The objective of GMsFEM is to develop a multiscale model reduction which allows us to seek an inexpensive approximated solution by solving (4.1) on a coarse grid (see Figure 1.2 for an illustration).

We introduce the notation for the coarse and fine grid. The computational domain Ω is partitioned by a coarse grid $\mathcal{T}^H$. The coarse grid contains multiscale features of the problem and require many degrees of freedom for modeling. We denote by the numbers of nodes and edges in the coarse grid by N_c and N_e respectively. We also denote a generic coarse grid element by K and the coarse mesh size by H. Next, we let $\mathcal{T}^h$ be a partition of Ω obtained from a refinement of $\mathcal{T}^H$. We call $\mathcal{T}^h$ the fine grid and $h > 0$ the fine mesh size $h > 0$. The fine mesh size h is sufficiently small such that the fine mesh resolves the multiscale features of the problem.

Using GMsFEM, multiscale basis functions, which capture local information, are constructed on the fine grid $\mathcal{T}^h$. A reduced number of basis functions is used in computations, which are done on the coarse grid $\mathcal{T}^H$. For each coarse region ω_i (or K) and time interval (T_{n-1}, T_n), we identify local multiscale basis functions ϕ_j^{n,ω_i} $(j = 1, ..., N_{\omega_i})$ and seek an approximated solution in the linear span of these basis functions. For problems with scale separation, a small number of basis functions is sufficient. For more complicated heterogeneities in many real-world applications, one needs a systematic approach to seek additional basis functions. Next, we will discuss some basic ingredients in the construction of our multiscale basis functions.

In each coarse region ω_i, the necessary information is contained in a local snapshot space

$V_{\text{snap}}^{n,\omega_i} = \text{span}\{\psi_j^{n,\omega_i}\} \subseteq V_h(\omega_i)$. The choice of the snapshot space depends on the global discretization and the particular application. One can also reduce the computational cost by computing fewer snapshot basis functions using randomized boundary conditions or source terms.

Next, based on our analysis, we design a local spectral problem for our multiscale basis functions ϕ_j^{n,ω_i} from the local snapshot space, and construct the local offline space $V_{H,\text{off}}^{n,\omega_i} = \text{span}\{\phi_j^{n,\omega_i}\} \subset V_{\text{snap}}^{n,\omega_i}$, which is a small-dimensional principal component subspace of the snapshot space. Through the spectral problem, we can select the dominant eigenvectors (corresponding to the smallest eigenvalues) as important degrees of freedom. We will then find an approximated solution in the linear span of multiscale basis functions in the offline space: find $u_H^n \in V_{H,\text{off}}^n$ can be obtained by solving the variational problem:

$$\begin{aligned} &\int_{T_{n-1}}^{T_n} \int_\Omega \frac{\partial u_H^n}{\partial t} v + a_n(u_H^n, v) + \int_\Omega u_H^n(x, T_{n-1}^+) v(x, T_{n-1}^+) \\ &= \int_{T_{n-1}}^{T_n} \int_\Omega f v + \int_\Omega u_H^{n-1}(x, T_{n-1}^-) v(x, T_{n-1}^+) \qquad \text{for all } v \in V_{H,\text{off}}^n, \end{aligned} \tag{4.3}$$

where $V_{H,\text{off}}^n = \oplus_i V_{H,\text{off}}^{n,\omega_i}$ and $a_n(u,v) = \int_{T_{n-1}}^{T_n} \int_\Omega \kappa \nabla u \cdot \nabla v$.

We remark that the choice of the spectral problem is important as the convergence rate of the method is proportional to $1/\Lambda_*$, where Λ_* is the smallest eigenvalue among all coarse blocks whose corresponding eigenvector is not included in the offline space. Therefore, we have to select a good local spectral problem in order to to remove as many small eigenvalues as possible so that we can obtain a reduced dimension coarse space and achieve a high accuracy.

In GMsFEM, the subgrid information is represented in the form of local multiscale basis functions. Local degrees of freedom are added as needed. It results in a set of numerical macroscopic equations for problems without scale separation and identifies important features for multiscale problems. Because of the local nature of proposed multiscale model reduction, the degrees of freedom can be added adaptively based on error estimators. However, due to the computational cost, one often uses fewer basis functions, which leads to discretization errors. Next, we discuss the detailed formulation of our Bayesian approach.

4.2 Bayesian formulation

We propose a Bayesian approach to resolve the missing subgrid information probabilistically in multiscale problems. The method starts with constructing multiscale basis functions and uses a few basis functions as permanent basis functions. Using these basis functions, an approximated solution can be obtained. Using the residual information, we can select additional basis functions stochastically. The construction of prior distribution and likelihood, which consists of residual minimization, is discussed. Such a probabilistic approach is useful for problems with limited additional information about the solution, as the additional information can be included in the likelihood. In this section, using the framework of GMsFEM, we will discuss a Bayesian formulation with measured data taken into account as an information on the solution.

4.2.1 Modeling the solution using GMsFEM multiscale basis functions

First, we select the dominant scale corresponding to the small eigenvalues in GMsFEM spectral problem to form a set of "permanent" basis functions, denoted by $\phi_j^{n,\omega_i}(x,t) \in V_{H,\text{off}}^n$. We can solve the Galerkin projection of (4.3) onto the span of permanent basis functions for an inexpensive fixed solution

$$u_H^{n,\text{fixed}}(x,t) = \sum_{i,j} \beta_{i,j}^n \, \phi_j^{n,\omega_i}(x,t),$$

where $\beta_{i,j}^n$'s are defined in each computational time interval.

The rest of the basis functions from local spectral problems, denoted by $\phi_{j,+}^{n,\omega_i}$, are called additional basis functions and correspond to unresolved scales. Using all the basis functions results a prohibitively large linear system and therefore, a mechanism that can select a small subset of the unused basis can be useful. The selected additional multiscale functions constitutes a linear space and gives a correction to the fixed solution. The coarse-scale solution at n-th time interval can then be written as the sum of the fixed and the additional part:

$$u_H^n = u_H^{n,\text{fixed}} + u_H^{n,+}.$$

Here, the solution of the coarse-scale system is assumed to be normal around the fixed solution with small variance. The solution involving unresolved scales can be expanded as

$$u_H^n(x,t) = \sum_{i,j} \beta_{i,j}^n \, \phi_j^{n,\omega_i}(x,t) + \sum_{i,j} \beta_{i,j,+}^n \, \phi_{j,+}^{n,\omega_i}(x,t),$$

where all but few coefficients $\beta_{i,j,+}$ are expected to be zero. Hence, the problem boils down to a model selection problem involving unused basis functions.

The linearization of a PDE system and the linear form involving additional basis provide a natural framework for Bayesian variable selection [68, 69, 70]. Suppose some observational data of $D^n(u^n)$ depending on the solution u^n are available at some grid points with some measurement error. The objective of our Bayesian formulation is to select and add appropriate additional multiscale functions $\phi_{j,+}^{n,\omega_i}$ in a systematic manner.

4.2.2 Bayesian formulation on variable selection problem

In this section, we discuss all the ingredients in our Bayesian formulation, including the prior and the posterior used in our sampling algorithms. Our proposed algorithm is residual-driven and also takes mismatch in observational data into account. We sample the correction $u_H^{n,+}$ by drawing samples of the indicator functions $\mathcal{I}^n$ and $\mathcal{J}^n$, and the coefficient vector β_+^n. We define suitable probability function for each of these random variables. Finally, this structure enables us to compute the posterior or conditional distribution of the basis selection probability and conditional solution of the system given by the observational data and the coarse-scale model.

We now define the residual and discuss the selection probability on the subregion and additional basis function based on the residual. Building our solution around the fixed solution, the residual

operator of equation (4.3) is defined as

$$\begin{aligned} R^n(u_H^{n,+};v) = &\int_{T_{n-1}}^{T_n}\int_\Omega fv + \int_\Omega u_H^{n-1}(x,T_{n-1}^-)v(T_{n-1}^+) \\ &- \int_{T_{n-1}}^{T_n}\int_\Omega \frac{\partial u_H^{n,\text{fixed}}}{\partial t}\, v + a_n(u_H^{n,\text{fixed}},v) - \int_\Omega u_H^{n,\text{fixed}}(x,T_{n-1}^+)v(T_{n-1}^+) \\ &- \int_{T_{n-1}}^{T_n}\int_\Omega \frac{\partial u_H^{n,+}}{\partial t}\, v + a_n(u_H^{n,+},v) - \int_\Omega u_H^{n,+}(x,T_{n-1}^+)v(T_{n-1}^+). \end{aligned} \tag{4.4}$$

We note that, since the fixed solution is the Galerkin projection onto the linear span of the permanent basis functions, for any permanent basis function ϕ_j^{n,ω_i}, we actually have

$$R^n(0;\phi_j^{n,\omega_i}) = 0.$$

For the additional basis functions $\phi_{j,+}^{n,\omega_i}$, the term $R^n(0;\phi_{j,+}^{n,\omega_i})$ provides a correlation of that basis function. We also denote the fine-scale residual vector by R^n.

Suppose an observational data model $Y^n = D^n(u^n)$ is supplemented to the PDE model. Here, observations Y^n are available in some coarse regions, and D^n is a function which describes the relation between the solution u^n and the the observations Y^n. In general, the function D^n can be nonlinear. In the numerical examples in this paper, D^n is taken to be some linear coarse-scale observations. We denote by E^n the mismatch between the given measurement Y^n and the image of the coarse-scale solution u_H^n under D^n, i.e.

$$E^n = Y^n - D^n(u_H^n).$$

Since we have a linear PDE model and a linear observation function D^n, the fine-scale residual R^n and the measurement mismatch E^n can be written in an affine representation in terms of coefficients β_+^n of the additional basis functions, i.e.

$$R^n = K^n\beta_+^n - b^n \text{ and } E^n = S^n\beta_+^n - g^n.$$

4.2.2.1 *Residual-based Bernoulli prior on indicator functions*

First, we identify some local neighborhoods for which multiscale basis functions should be added. Independent Bernoulli prior can be assumed for each local region being selected for adding basis. Next, for each local region ω_i selected, each multiscale basis function $\phi_{j,+}^{n,\omega_i}$ is selected with another independent Bernoulli prior given that corresponding subregion is selected. The selection probability for the Bernoulli distribution is given by residual in the fine-scale system, where prior favors the scales that have more correlation with the residual.

In the construction of the Bernoulli prior on the local subregions, we consider the 1-norm of the residual vector

$$\alpha(\omega_i) = \sum_j |R^n(0; \phi_{j,+}^{n,\omega_i})|.$$

Let N_ω be the average number of subregions where additional basis functions will be added. Then we rescale the norm by

$$\widehat{\alpha}(\omega_i) = \frac{\alpha(\omega_i)}{\sum_k \alpha(\omega_k)} N_\omega, \tag{4.5}$$

and set the selection probability of the region ω_i as $\min\{\widehat{\alpha}(\omega_i), 1\}$. An indicator function $\mathcal{J}^n$ can then be defined according to the activity of the local neighborhoods. In a sample, we use $\mathcal{J}_i^n = 1$ to denote the region ω_i being selected, and $\mathcal{J}_i^n = 0$ otherwise.

Next, we discuss the prior probability on the additional basis functions. For a selected region ω_i, suppose we would select N_{basis} additional basis functions on average. Then we consider

$$\alpha(\phi_{j,+}^{n,\omega_i}) = |R^n(0; \phi_{j,+}^{n,\omega_i})|,$$

and rescale it by

$$\widehat{\alpha}(\phi_{j,+}^{n,\omega_i}) = \frac{\alpha(\phi_{j,+}^{n,\omega_i})}{\sum_k \alpha(\phi_{k,+}^{n,\omega_i})} N_{\text{basis}}, \tag{4.6}$$

and set the selection probability of the basis function $\phi_{j,+}^{n,\omega_i}$ as $\min\{\widehat{\alpha}(\phi_{j,+}^{n,\omega_i}), 1\}$. Similarly, we define an indicator function $\mathcal{I}^n$ on the basis functions. We write $\mathcal{I}_{i,j}^n = 1$ if the basis function $\phi_{j,+}^{n,\omega_i}$ is active and $\mathcal{I}_{i,j}^n = 0$ otherwise.

4.2.2.2 *Residual-data-based prior on coefficient vector*

Next, using this residual information, a sequential scheme to add coarse regions and additional basis functions for each selected region is introduced. The probability of each coarse region region or additional basis function being selected are proportional to the residual information they contain. Later, using the residual information as prior, a full Bayesian method is developed to select additional basis functions given the observations and the model. The likelihood of Y^n is

$$P(Y^n|\beta_+^n) \sim \exp\left(-\frac{\|E^n\|^2}{2\sigma_d^2}\right). \tag{4.7}$$

Assuming the true solution Gaussian around the fixed model which gives a model based prior of the form for u_H^n:

$$\pi(u_H^n|\beta^n(\mathcal{I}^n,\mathcal{J}^n),u_H^{n-1}) \sim \exp\left(-\frac{\|R^n\|^2}{2\sigma_L^2}\right) \tag{4.8}$$

where R^n is the vector of residual when the test functions are varied over the all fine-scale basis functions. This gives a pseudo-likelihood for the residuals. For the coefficient vector β_+^n independent normal priors are assumed with mean zero and a large prior variance, i.e. a flat normal prior is assumed. The distribution of the new coefficients given the indices corresponding to the basis/ sub-region selection and new observations

$$P(\beta_+^n|Y^n,(\mathcal{I}^n,\mathcal{J}^n),u_H^{n-1}) \propto P(Y^n|\beta_+^n)\pi(u_H^n|\beta^n(\mathcal{I}^n,\mathcal{J}^n),u_H^{n-1}). \tag{4.9}$$

4.2.2.3 *Posterior around fixed solution using residual-data-minimizing likelihood*

Using residual information from the PDE model as prior for basis selection, a Bayesian variable selection method can be devised. Posterior estimates are computed in each time interval sequentially from the estimates of the earlier time intervals. In each time interval, one or more coarse regions are selected by the ad hoc cut off $\min\{\widehat{\alpha}(\omega_i),1\}$ on the rescaled residual norm defined in (4.5). At each selected coarse region, extra useful basis functions are selected from the following

posterior distribution involving the joint prior distribution based on the PDE model and the prior on the coefficient:

$$\begin{aligned}\pi_1(\beta_+^n, (\mathcal{I}^n, \mathcal{J}^n), u_H^n) \sim \pi(u_H^n|\beta^n(\mathcal{I}^n, \mathcal{J}^n), u_H^{n-1}) \\ \pi(\beta_+^n|\mathcal{I}^n, \mathcal{J}^n)\pi(\mathcal{I}^n|\mathcal{J}^n)c_d(\mathcal{I}^n, \mathcal{J}^n),\end{aligned} \tag{4.10}$$

for a model dependent constant $c_d(\mathcal{I}^n, \mathcal{J}^n)$. On β_+^n flat normal priors are used. The model dependent constant $c_d(\mathcal{I}^n, \mathcal{J}^n)$ depends on the PDE model and the design matrix for the observation $D^n(u_H^n)$. The posterior is then given by:

$$P(\beta_+^n, \mathcal{I}^n|Y^n) \sim P(Y^n|\beta_+^n(\mathcal{I}^n, \mathcal{J}^n))\pi_1(\beta_+^n, (\mathcal{I}^n, \mathcal{J}^n), u_H^n). \tag{4.11}$$

Remark 4.2.1. *The term $c_d(\mathcal{I}^n, \mathcal{J}^n)$ is proportional to the square root of the determinant of the information matrix of β_+^n for given $\mathcal{I}^n, \mathcal{J}^n$, in the posterior distribution without the normalizing term c_d, and gives a empirical Bayes type prior for the model probability. This choice is motivated by selecting basis based on only likelihood and the residual information and not penalizing the model size. The term c_d is cancelled in the MCMC step (given later) after integrating out the coefficient β_+^n.*

4.2.3 Sampling algorithms

Based on our Bayesian formulation, we propose two different sampling methods, namely sequential sampling and full posterior MCMC sampling, for modeling unresolved scales. The sequential sampling method uses prior information to directly select additional basis functions and is inexpensive. The MCMC sampling method requires full posterior sampling and is more accurate than the sequential sampling method. A schematic representation of the methods is presented in Figure 4.1.

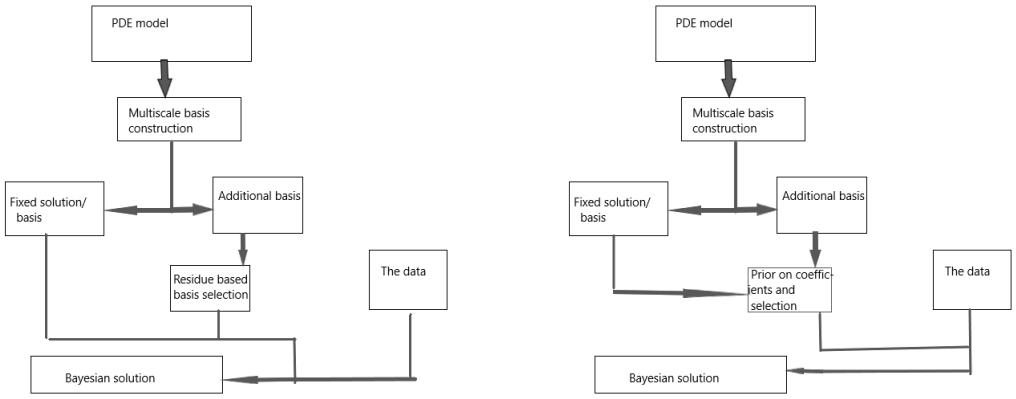

Figure 4.1: A schematic illustration of sequential sampling (left) and MCMC sampling (right). Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

4.2.3.1 Sequential sampling

First, we present a sequential sampling method which uses the prior distributions as discussed in the previous section to generate realizations of the solution.

Algorithm 1 Generation of sequential sample

1: Sample $\mathcal{J}^n$ according to Bernoulli prior

2: Sample $\mathcal{I}^n$ in the regions ω_i for which $\mathcal{J}_i^n = 1$ according to Bernoulli prior

3: Sample β_+^n according to (4.7), (4.8) and (4.9).

4: **return** $\mathcal{I}^n, \mathcal{J}^n, \beta_+^n$

The sequential sampling method directly makes use of the prior information given from the fixed solution. While the sequential sampling method is inexpensive, the usefulness of the selected basis functions in sequential sampling method therefore heavily depends on the quality of the fixed solution. In order to provide a better distribution of the additional basis functions, a full posterior sampling method is proposed to model the resolved scales.

4.2.3.2 *Full posterior MCMC sampling*

Next, we present the details of full posterior MCMC sampling for modeling unresolved scales. More precisely, we discuss the details of the acceptance-rejection mechanism in a Markov-chain Monte Carlo (MCMC) method. In a sampling step for a particular basis function $\phi_{j,+}^{n,\omega_i}$, suppose we have a original configuration $\mathcal{I}^n$ for the indicator function on the additional basis functions. We define two configurations $\mathcal{I}_+^n$ and $\mathcal{I}_-^n$ by setting $\phi_{j,+}^{n,\omega_i}$ active in $\mathcal{I}_+^n$ and inactive in $\mathcal{I}_-^n$, while indicators on all other additional basis functions being the same as $\mathcal{I}^n$. (One of these two configurations should be exactly $\mathcal{I}^n$ itself.) For each configuration, the mode of the posterior distribution is achieved by the solution of their respective linear system

$$\left(\frac{1}{2\sigma_L^2}(K^n)^T K^n + \frac{1}{2\sigma_d^2}(S^n)^T S^n\right)\beta_+^n = \frac{1}{2\sigma_L^2}(K^n)^T b^n + \frac{1}{2\sigma_d^2}(S^n)^T g^n, \tag{4.12}$$

while the solution minimizes a weighted sum of the residual and the mismatch in each system. If we denote the residual and the mismatch by R_+^n and E_+^n for the system for the configuration $\mathcal{I}_+^n$, and R_-^n and E_-^n similarly for $\mathcal{I}_-^n$, then the acceptance-rejection probability ratio is given by

$$\frac{p(\phi_{j,+}^{n,\omega_i})}{1-p(\phi_{j,+}^{n,\omega_i})} = \frac{\widehat{\alpha}(\phi_{j,+}^{n,\omega_i})}{1-\widehat{\alpha}(\phi_{j,+}^{n,\omega_i})} \exp\left(-\frac{\|R_+^n\|^2 - \|R_-^n\|^2}{2\sigma_L^2} - \frac{\|E_+^n\|^2 - \|E_-^n\|^2}{2\sigma_d^2}\right). \tag{4.13}$$

Then we update the configuration with $\mathcal{I}_+^n$ and $\mathcal{I}_-^n$ with probability $p(\phi_{j,+}^{n,\omega_i})$ and $1-p(\phi_{j,+}^{n,\omega_i})$ respectively.

The posterior sampling can be performed by a Gibbs sampling algorithm after marginalizing over β_+^n. Here we present a flow of the MCMC algorithm. The posterior distribution given the index set $\mathcal{I}^n$ follows multivariate normal with mean with $\beta^n(\mathcal{I}^n)_+$. In the generation of a particular example, the MCMC steps go as follows:

Algorithm 2 Generation of MCMC sample

Sample $\mathcal{J}^n$ according to Bernoulli prior
Sample $\mathcal{I}^n$ in the regions ω_i for which $\mathcal{J}_i^n = 1$ according to Bernoulli prior
for all $\phi_{k,+}^{n,\omega_i}$ with $\mathcal{J}_i^n = 1$ **do**
    Generate the linear system (4.12) for each of configurations $\mathcal{I}_+^n$ and $\mathcal{I}_-^n$
    Solve for modes β_+^n of posterior distribution in the two systems (4.12)
    Calculate $p(\phi_{j,+}^{n,\omega_i})$ by (4.13)
    Generate a random number $\xi \sim \mathcal{U}[0,1]$
    if $\xi < p(\phi_{j,+}^{n,\omega_i})$ **then**
        $\mathcal{I}^n \leftarrow \mathcal{I}_+^n$, i.e. $\mathcal{I}_{i,j}^n \leftarrow 1$
    else
        $\mathcal{I}^n \leftarrow \mathcal{I}_-^n$, i.e. $\mathcal{I}_{i,j}^n \leftarrow 0$
    end if
end for
return $\mathcal{I}^n, \mathcal{J}^n, \beta_+^n$

4.3 Numerical results

In this section, we present two numerical examples. In both examples, the computational domain is $\Omega = (0,1)^2$. We consider the parabolic equation

$$\frac{\partial u}{\partial t} - \text{div}(\kappa \nabla u) = f,$$

where f is a given source term, and κ is a space-time permeability field. The initial permeability field $\kappa_0 = \kappa(\cdot, 0)$ are shown in Figure 4.2, and the contrast $\dfrac{\max \kappa}{\min \kappa}$ is increasing over time t as $\dfrac{\max \kappa}{\min \kappa} = 10000e^{250t}$. For simplicity, homogeneous Dirichlet boundary condition is prescribed.

Next, we discuss the discretization used in the examples. We divide the domain Ω into a 10×10

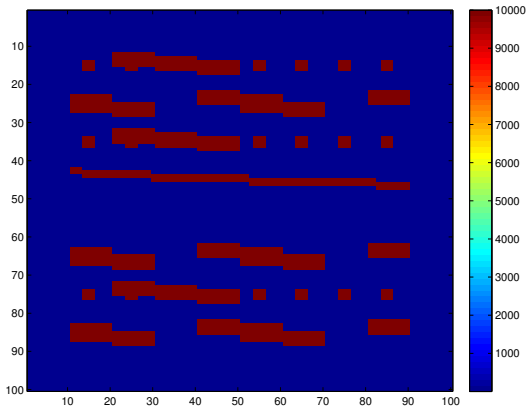

Figure 4.2: The permeability field κ_0. Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

coarse grid and 100×100 fine grid. For the sake of simplicity, we make use of the continuous Galerkin formulation in spatial discretization, use local fine-scale spaces consisting of fine-grid basis functions with a coarse region ω_i as our snapshot basis functions, construct multiscale basis functions independent of time, and employ the implicit Euler formula in temporal discretization. At each time instant t_n, we seek numerical solution u_h^{n+1} in the standard conforming bilinear finite space space V_h on the fine grid $\mathcal{T}^h$, i.e.

$$V_h = \left\{ v \in C_0(\Omega) : v|_\tau \in \mathbb{Q}^1(\tau) \text{ for all } \tau \in \mathcal{T}^h \right\} \subset H_0^1(\Omega).$$

The variational formulation is given by: find $u_h^{n+1} \in V_h$ such that

$$\int_\Omega \frac{u_h^{n+1} - u_h^n}{\Delta t} v + \int_\Omega \kappa \nabla u_h^{n+1} \cdot \nabla v = \int_\Omega f v \text{ for all } v \in V_h.$$

The multiscale basis functions are obtained from eigenfunctions in the local snapshot space with small eigenvalues in the following spectral problem: find $(\phi_j^{\omega_i}, \lambda_j^i) \in V_{\text{snap}}^{\omega_i} \times \mathbb{R}$ such that

$$a_i(\phi_j^{\omega_i}, w) = \lambda_j^i s_i(\phi_j^{\omega_i}, w) \quad \text{for all } w \in V_{\text{snap}}^{\omega_i}.$$

Here the bilinear forms a_i and s_i are defined by

$$a_i(v,w) = \int_{\omega_i} \kappa_0 \nabla v \cdot \nabla w \quad \text{and} \quad s_i(v,w) = \int_{\omega_i} \widetilde{\kappa}_0 vw,$$

where $\widetilde{\kappa}_0 = \sum_{i=1}^{N_c} \kappa_0 |\nabla \chi_i^{ms}|^2$ and χ_i^{ms} are the standard multiscale finite element basis functions. The eigenvalues λ_j^i are arranged in ascending order, and the multiscale basis functions are constructed by multiplying the partition of unity to the eigenfunctions. We will use the first L_i eigenfunctions to construct our offline space $V_{H,\text{off}}^{\omega_i}$. We construct the offline space $V_{H,\text{off}} = \oplus_i V_{H,\text{off}}^{\omega_i}$.

4.3.1 Experiment 1

In the first example, we investigate the performance our proposed method. The source function is taken as $f = 1$. We will compare the solutions at the time instant $T = 0.02$.

We compute 2 permanent basis functions and 18 additional basis functions per coarse neighborhood. The permanent basis functions are used to compute "fixed" solution and use our Bayesian framework to seek additional basis functions by solving small global problems and making use of given dynamic observational data. In this example, we consider four observational data

$$D_i^n = \int_{K_i} u^n, \quad i = 1,2,3,4,$$

where the locations of the centers of the coarse grid elements K_i are shown in Figure 4.3. On average we select 27 local regions at which multiscale basis functions are added. In these coarse blocks, we apply both sequential sampling and full sampling and generate 100 samples.

Figure 4.4 shows the reference solution and the sample mean at $T = 0.02$. The L^2 error for the mean at $T = 0.02$ is 0.63% in the full sampling method, lower than 1.92% in the sequential sampling method.

In Figure 4.5, the residual and L^2 errors are plotted over the sampling process. We observe that the errors and the residual in full sampling decrease and stabilize in a few iterations. Moreover, the full sampling gives more accurate solutions associated with our error threshold in the residual.

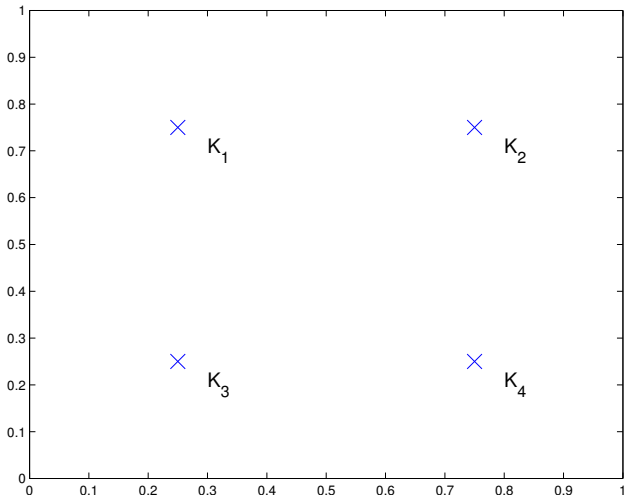

Figure 4.3: Locations of the centers of the coarse grid elements K_i. Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

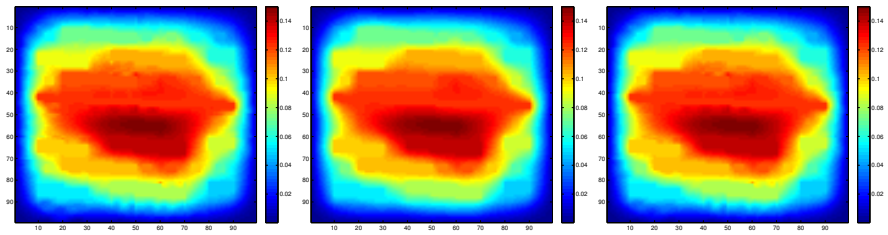

Figure 4.4: Plots of the reference solution (left), sequential sample mean (middle) and full sample mean (right) of numerical solution at $T = 0.02$. Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

In Table 4.1, we compare the percentages of additional basis selected by the full sampling method with different combinations of σ_L and σ_d. Tables 4.2 and 4.3 record the L^2 error of the solution and the maximum observational error, i.e.

$$\max_{1\leq i\leq 4}\left|\int_{K_i}(u^N - u_H^N)\right|,$$

with these combinations of σ_L and σ_d. It can be observed that a smaller σ_L results in a larger

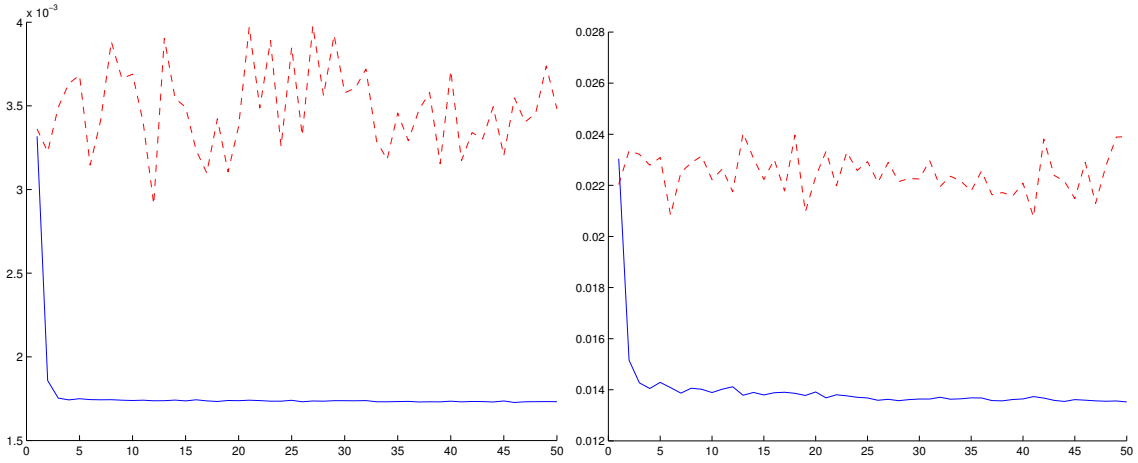

Figure 4.5: Residual (left) and L^2 error (right) vs sample using sequential sampling (red dotted line) and full sampling (blue solid line) at time $T = 0.02$. Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

number of additional basis functions selected and a significant improvements in the L^2 error of the numerical solution. On the other hand, a smaller σ_d does not significantly increase the number of additional basis functions selected, but improves the quality of our solution by greatly reducing the mismatch with observational data. This shows our method is useful when the accuracy of the observational data is important.

	σ_d		
σ_L	1×10^{-6}	1×10^{-3}	1×10^{0}
5×10^{-4}	74.49%	72.22%	73.46%
1×10^{-3}	48.15%	47.94%	48.15%
2×10^{-3}	32.10%	31.07%	32.30%

Table 4.1: Percentage of additional basis selected in the selected subdomains with various σ_L and σ_d. Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

	σ_d		
σ_L	1×10^{-6}	1×10^{-3}	1×10^{0}
5×10^{-4}	0.39%	0.51%	0.63%
1×10^{-3}	1.35%	1.35%	1.07%
2×10^{-3}	1.54%	1.52%	1.29%

Table 4.2: L^2 error in the solution with various σ_L and σ_d. Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

	σ_d		
σ_L	1×10^{-6}	1×10^{-3}	1×10^{0}
5×10^{-4}	2.59×10^{-12}	1.33×10^{-5}	2.98×10^{-2}
1×10^{-3}	1.79×10^{-11}	1.98×10^{-5}	1.33×10^{-2}
2×10^{-3}	9.72×10^{-12}	1.07×10^{-5}	5.61×10^{-2}

Table 4.3: Maximum observational error with various σ_L and σ_d. Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

4.3.2 Experiment 2

As a second example, we employ our method to simulate an inflow-outflow problem. The source function is taken as $f = \chi_{K_1} + \chi_{K_2} - \chi_{K_3} - \chi_{K_4}$. The source term f is shown in Figure 4.6. The dynamic observational data is the average value on the coarse grid regions K_3 and K_4, i.e.

$$D_1^n = \frac{\int_{K_3} u^n}{|K_3|}, \quad D_2^n = \frac{\int_{K_4} u^n}{|K_4|}.$$

In real situations, K_3 and K_4 are the locations of the production wells, while K_1 and K_2 are the locations of the injection wells. In practice, the accuracy of the average value at the production wells are essential.

We compute 2 permanent basis functions and 18 additional basis functions per coarse neighborhood. The permanent basis functions are used to compute "fixed" solution and use our Bayesian

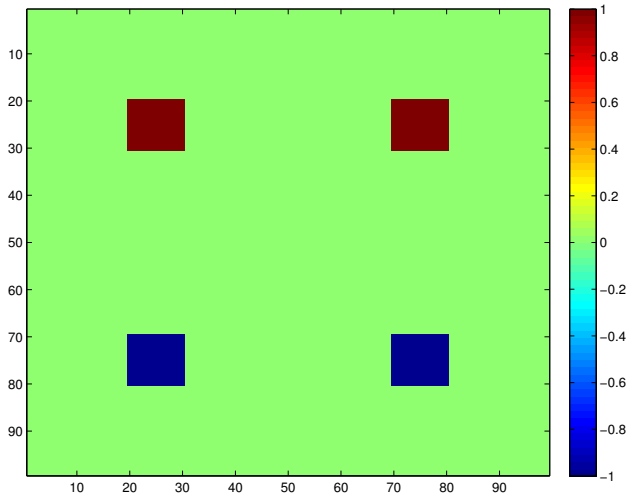

Figure 4.6: Source function f in the inflow-outflow problem. Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

framework to seek additional basis functions by solving small global problems and making use of given observational data. On average we select 27 of local regions at which multiscale basis functions are added. In these coarse blocks, we apply both sequential sampling and full sampling and generate 100 samples. The thresholds are set as $\sigma_L = 9\times10^{-6}$ and $\sigma_d = 1\times10^{-7}$. We also compare our proposed method when there is no available observation data and only a residual-minimizing likelihood is used.

In the numerical simulation, 49.79% of the additional basis functions are selected in the selected subdomains using our proposed method, compared with 49.18% in the absence of observation data. Figure 4.7 shows the reference solution and the sample mean at $T = 0.02$. The L^2 error for the mean at $T = 0.02$ is 2.71% and 2.61% respectively. Moreover, the maximum error in observational data in our proposed method is 1.72×10^{-12}, much lower than 3.54×10^{-4} in the absence of observation data.

These results demonstrate that our proposed Bayesian approach is able to select important basis functions to model the missing subgrid information, both in minimizing the residual of the problem and reducing the error in the targeted observational data.

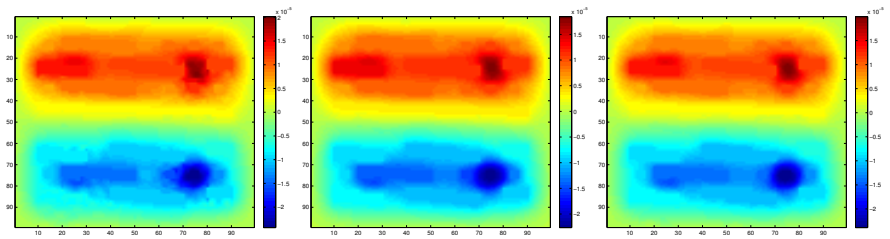

Figure 4.7: Plots of the reference solution (left), sequential sample mean (middle) and full sample mean (right) of numerical solution at $T = 0.02$. Reprinted with permission from "Dynamic Data-driven Bayesian GMsFEM" by Siu Wun Cheung and Nilabja Guha, 2019. Journal of Computational and Applied Mathematics, Volume 353, Pages 72–85, Copyright [2019] by Elsevier.

5. DEEP GLOBAL MODEL REDUCTION LEARNING IN POROUS MEDIA FLOW SIMULATION *

In this chapter, we use deep learning concepts combined with Proper Orthogonal Decomposition (POD) model reduction methodologies constrained at observation locations to predict flow dynamics. We consider a neural network-based approximation of nonlinear flow dynamics. Flow dynamics is regarded as a multi-layer network, where the solution at the current time step depends on the solution at the previous time instant and associated input parameters, such as well rates and permeability fields. This allows us to treat the solution via multi-layer network structures, where each layer is a nonlinear forward map and to design novel multi-layer neural network architectures for simulations using our reduced-order model concepts. The resulting forward model takes into account available data at locations and can be used to reduce the computational cost associated with forward solves in nonlinear problems.

We will rely on rigorous model reduction concepts to define unknowns and connections for each layer. Reduced-order models are important in constructing robust learning algorithms since they can identify the regions of influence and the appropriate number of variables, thus allow using small-dimensional maps. In this work, modified proper orthogonal basis functions will be constructed such that the degrees of freedom have physical meanings (e.g., represent the solution values at selected locations). Since the constructed basis functions have limited support, it will allow localizing the forward dynamics by writing the forward map for the solution values at selected locations with pre-computed neighborhood structure. We use a proper orthogonal decomposition model with these specifically designed basis functions that are constrained at locations. A principal component subspace is constructed by spanning these basis functions and numerical solutions are sought in this subspace. As a result, the neural network is inexpensive to construct.

Our approach combines the available data and physical models, which constitutes a data-driven

* Reprinted with permission from "Deep Global Model Reduction Learning in Porous Media Flow Simulation" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Eduardo Gildin, Yating Wang and Jingyan Zhang, 2020. Computational Geosciences, Volume 24, Pages 261–274, Copyright [2020] by Springer.

modification of the original reduced-order model. To be specific, in the network, our reduced-order models will provide a forward map, and will also be modified (âĂIJtrainedâĂİ) using available observation data. Due to the lack of available observation data, we will use computational data to supplement as needed. The interpolation between data-rich and data-deficient models will also be studied. We will also use deep learning algorithms to train the elements of the reduced model discrete system. In this case, deep learning architectures will be employed to approximate the elements of the discrete system and reduced-order model basis functions.

We will present numerical results using deep learning architectures to predict the solution and reduced-order model variables. In the reduced-order model, designated basis functions allow interpolating the solution between observation points. A multi-layer neural network based is then built to approximate the evolution of the coefficients and, therefore, the flow dynamics. We examine how the network architecture, which includes the number of layers, and neurons, affects the approximation. Our numerical results show that with a fewer number of layers, the flow dynamics can be approximated. Our numerical results also indicate that the data-driven approach improves the quality of approximation.

The chapter is organized as follows. In Section 5.1, we present a general model and some basic concepts of POD. Section 5.2 is devoted to our model learning. In Section 5.3, we present numerical results. We conclude in the last section.

5.1 Preliminaries

In this section, we introduce a general problem setting and review the concept of POD based global model reduction, which is a technique of dimensionality reduction of large-scale system of ordinary differential equations (ODE) and its application to nonlinear partial differential equations (PDE). Consider a time-dependent PDE in the general form

$$\frac{\partial}{\partial t}u = \mathcal{L}(u) + g \quad \text{in } \Omega \times (0, T), \tag{5.1}$$

where Ω is the spatial domain, $(0, T)$ is the temporal domain, $\mathcal{L}$ is a spatio-differential operator on the unknown u and g is a given source function. The flow dynamic is prescribed to some given initial condition and boundary condition. We consider spatial discretization procedure by finite element method on a Eulerian mesh $\mathcal{T}_h$ for the spatial domain Ω. Let V_h be a finite element space spanned by the nodal basis $\{\phi_j\}_{j=1}^n$ on $\mathcal{T}_h$. We seek numerical solution of (5.1) by an expansion

$$u(x,t) = \sum_{j=1}^{n} y_j(t)\phi_j(x), \tag{5.2}$$

which yields a system of ODE in the form

$$\frac{d}{dt}\mathbf{y}(t) = \mathbf{B}\mathbf{y}(t) + \mathbf{f}(\mathbf{y}(t)), \tag{5.3}$$

where $\mathbf{y}(t) \in \mathbb{R}^n$ is the state vector, $\mathbf{B} \in \mathbb{R}^{n\times n}$ is a constant matrix, and $\mathbf{f} : \mathbb{R}^n \to \mathbb{R}^n$ is a nonlinear function. In our applications, the dimension n corresponds to the number of physical grid points in the mesh. In general, the dimension n is huge and model reduction techniques provide efficient reduced-order models and bring computational savings.

5.1.1 Proper Orthogonal Decomposition

Proper Orthogonal Decomposition is a popular mode decomposition method, which aims at reducing the order of the model by extracting important relevant feature representation with a low dimensional space. In this section, we briefly discuss the POD method. For a more detailed discussion of the use of POD on dynamic systems, the reader is referred to [71, 72]. In POD, a low-dimensional set of modes, i.e., important degrees of freedom, are identified based on processing information from a sequence of snapshots, i.e., instantaneous solutions from the dynamic process, and extracting the most energetic structures in terms of the largest singular values. In the statistical point of view, the extracted modes are uncorrelated and form an optimal reduced order model, in the sense that the variance is maximized and the mean squared distance between the snapshots and the POD subspace is minimized.

Proper orthogonal decomposition starts with a collection of $N \ll n$ instantaneous snapshots $\{\mathbf{y}_j\}_{j=1}^N \subset \mathbb{R}^n$, where the snapshot times in the above sequence is assumed to be equidistant. The snapshots span a snapshot space of dimension r and are arranged in a matrix form known as the snapshot matrix

$$\mathbf{Y} = [\mathbf{y}_1 \ \ \mathbf{y}_2 \ \ \cdots \ \ \mathbf{y}_N] \in \mathbb{R}^{n\times N}. \tag{5.4}$$

The idea of POD is to seek the subspace of a certain dimension which best approximates the linear space spanned by the snapshots. Among all subsets of $m < r$ orthonormal vectors in $\mathbb{R}^n$, we seek the POD basis $\{\mathbf{v}_j\}_{j=1}^m$ by solving a minimization problem

$$\underset{\substack{\{\mathbf{v}_j\}_{j=1}^m \subset \mathbb{R}^n \\ \langle \mathbf{v}_i, \mathbf{v}_j\rangle = \delta_{ij}}}{\operatorname{argmin}} \sum_{i=1}^{N} \left\| \mathbf{y}_i - \sum_{j=1}^{m} \langle \mathbf{y}_i, \mathbf{v}_j \rangle \, \mathbf{v}_j \right\|_2^2, \tag{5.5}$$

The minimzation problem is processed by performing a singular value decomposition on the snapshot matrix $\mathbf{Y}$

$$\mathbf{Y} = \mathbf{V}\Lambda\mathbf{W}^T, \tag{5.6}$$

where $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \cdots, \mathbf{v}_r] \in \mathbb{R}^{n\times r}$ and $\mathbf{W} = [\mathbf{w}_1, \mathbf{w}_2, \cdots, \mathbf{w}_r] \in \mathbb{R}^{N\times r}$ consist of the left-singular vectors and right-singular vectors of $\mathbf{Y}$ respectively, and $\Lambda = \operatorname{diag}(\sigma_1, \sigma_2, \cdots, \sigma_r) \in \mathbb{R}^{r\times r}$ is the diagonal matrix consisting of the singular values of $\mathbf{Y}$. Constructively, we denote the correlation matrix from the snapshot sequence by $\mathbf{C} = \mathbf{Y}^T\mathbf{Y}$, and compute the eigenvalue decomposition on $\mathbf{C}$

$$\mathbf{C}\mathbf{q}_j = \lambda_j \mathbf{q}_j, \tag{5.7}$$

and obtain the singular values $\{\sigma_j\}_{j=1}^r$ and singular vectors $\{\mathbf{v}_j\}_{j=1}^r$ by

$$\sigma_j = \sqrt{\lambda_j} \text{ and } \mathbf{v}_j = \frac{1}{\sigma_j}\mathbf{Y}\mathbf{q}_j. \tag{5.8}$$

Here the singular values are arranged in descending energy ranking, i.e., $\sigma_1 \geq \sigma_2 \geq \cdots \geq \sigma_r > 0$, which correspond to the energy content of a mode. The energy ranking provides a measure of

the importance of the mode in capturing the relevant dynamic process. The POD basis, i.e. the solution of the minimization problem (5.5), is then given by selecting the first m singular vectors $\{\mathbf{v}_j\}_{j=1}^m$. In this case, we have

$$\sum_{i=1}^{N} \left\| \mathbf{y}_i - \sum_{j=1}^{m} \langle \mathbf{y}_i, \mathbf{v}_j \rangle \, \mathbf{v}_j \right\|_2^2 = \sum_{j=m+1}^{r} \sigma_j^2. \tag{5.9}$$

The size m of the POD basis has to be sufficiently large to include the first few largest singular values and ensure a good approximation to the snapshot matrix. The number of basis can be pre-defined or determined by means of fractional energy, i.e. fixing a threshold E_0, pick the smallest integer m such that

$$E = \frac{\sum_{j=1}^{m} \sigma_j^2}{\sum_{j=1}^{r} \sigma_j^2} > E_0, \tag{5.10}$$

In general, a few basis is needed if the singular values decay quickly. The rate of decay depends on the intrinsic dynamics of the system and the selection of the snapshots.

5.1.2 Fully discrete reduced-order model

Using the aforementioned POD basis $\{\mathbf{v}_j\}_{j=1}^m$, we can express the solution as

$$\mathbf{y}(t) \approx \sum_{j=1}^{m} \widetilde{c}_j(t) \mathbf{v}_j = \mathbf{V}\widetilde{c}(t), \tag{5.11}$$

where $\widetilde{c}(t) = (\widetilde{c}_1(t), \widetilde{c}_2(t), \ldots, \widetilde{c}_m(t))^T \in \mathbb{R}^m$ is the coordinates of $\mathbf{y}(t)$ with respect to the POD basis. We therefore derive a reduced-order ODE system

$$\frac{d}{dt}\widetilde{c}(t) = \mathbf{V}^T\mathbf{B}\mathbf{V}\widetilde{c}(t) + \mathbf{V}^T\mathbf{f}(\mathbf{V}\widetilde{c}(t)), \tag{5.12}$$

and further reduce it to an algebraic system. We consider a partition $0 = t_0 < t_1 < \ldots < t_s = T$ for the temporal domain $(0, T)$. Using, for example, implicit Euler method for temporal

discretization, we obtain a recurrence relation

$$\widetilde{c}^{n+1} = \widetilde{c}^{n} + (t_{n+1} - t_n)\left(\mathbf{V}^T\mathbf{B}\mathbf{V}\widetilde{c}^{n+1} + \mathbf{V}^T\mathbf{f}(\mathbf{V}\widetilde{c}^{n+1})\right), \tag{5.13}$$

where $\widetilde{c}^n$ denotes the numerical solution of $\widetilde{c}(t)$ at the time instant $t = t_n$. The nonlinear term $\mathbf{f}$ can be handled with different techniques, such as direct linearization method, fixed point iterations and Discrete Empirical Interpolation Method (DEIM), depending on situations and need for accuracy in particular applications.

5.1.3 Construction of nodal basis functions

Next, we present the construction of basis functions in the POD subspace. The basis functions are designed such that the degrees of freedom have physical meanings (e.g., represent the solution values at selected locations). Since the constructed basis functions have limited support, it will allow localizing the forward dynamics by writing the forward map for the solution values at selected locations with pre-computed neighborhood structure.

Given a set of nodes $\{x_k\}_{k=1}^m$ in the mesh $\mathcal{T}_h$, which correspond to particular physical points in the spatial domain Ω, we construct nodal basis functions by linear combinations of POD modes $\{\mathbf{v}_j\}_{j=1}^m$. More precisely, we seek coefficients α_{ij} such that

$$\sum_{j=1}^{m} \alpha_{ij}\mathbf{V}_{kj} = \delta_{ik}. \tag{5.14}$$

We remark that $\mathbf{V}_{kj}$ is the nodal evaluation of the interpolant of $\mathbf{v}_j$ in the finite element space V_h at the node x_k. The nodal basis ψ_k at the node x_k is then defined by

$$\psi_k = \sum_{j=1}^{m} \alpha_{kj}\mathbf{v}_j. \tag{5.15}$$

The set of nodal basis spans exactly the POD subspace. A reduced-order state vector $\mathbf{y}(t)$ in the

POD subspace can then be written in the expansion

$$\mathbf{y}(t) = \sum_{k=1}^{m} c_k(t)\psi_k, \tag{5.16}$$

where the coefficients $c_k(t)$ represent the nodal evaluation of the finite element approximation of $u(x,t)$ at the node x_k. Furthermore, the coefficients $\widetilde{c}^n$ of the original POD basis functions and c^n of the POD nodal basis functions are related by

$$c^n = \mathbf{V}_m \widetilde{c}^n, \tag{5.17}$$

where $\mathbf{V}_m \in \mathbb{R}^{m\times m}$ is the submatrix obtained from $\mathbf{V}$ by taking the rows corresponding to the nodes $\{x_k\}_{k=1}^m$.

5.2 Deep Global Model Reduction and Learning

5.2.1 Main idea

We will make use of the reduced-order model described in Section 5.1 to model the flow dynamics, and a deep neural network to approximate the flow profile. In many cases, the flow profile is dependent on data. The idea of this work is to make use of deep learning to combine the reduced-order model and available data and provide an efficient numerical model for modelling the flow profile.

First, we note that the solution at the time instant $n+1$ depends on the solution at the time instant n and input parameters I^{n+1}, such as permeability field and source terms. Here, we would like to use a neural network to describe the relationship of the solutions between two consecutive time instants. Suppose we have a total m sample realization in the training set. For each realization, given a set of input parameters, we solve the aforementioned reduced-order model and obtain the coefficients at particular points

$$\{c^0, \cdots, c^k\} \tag{5.18}$$

at all time steps. Our goal is to use deep learning techniques to train the trajectories and find a

network $\mathcal{N}$ to describe the pushforward map between c^n and c^{n+1} for any training sample.

$$c^{n+1} \sim \mathcal{N}(c^n, I^{n+1}), \tag{5.19}$$

where I^{n+1} is an input parameter which could vary over time, and $\mathcal{N}$ is a multi-layer network to be trained. The network $\mathcal{N}$ will approximate the discrete flow dynamics.

In our neural network, c^n and I^{n+1} are the inputs, c^{n+1} is the output. One can take the coefficients from time 0 to time $k-1$ as input, and from time 1 to k as output in the training process. In this case, a universal neural net $\mathcal{N}$ is obtained. The solution at time 0 can then be forwarded all the way to time k by repeatedly applying the universal network k times, that is,

$$c^k \sim \mathcal{N}(\mathcal{N} \cdots \mathcal{N}(c^0, I^1) \cdots, I^{k-1}), I^k). \tag{5.20}$$

After a network is trained, it can be used for predicting the trajectory given a new set of input parameters I^{n+1} and realization of coefficients at initial time c^0_{new} by

$$c^k_{\text{new}} \sim \mathcal{N}(\mathcal{N} \cdots \mathcal{N}(c^0_{\text{new}}, I^1) \cdots, I^{k-1}), I^k). \tag{5.21}$$

Alternatively, one can also train each forward map for any two consecutive time instants as needed. That is, we will have $c^{n+1} \sim \mathcal{N}_{n+1}(c^n, I^{n+1})$, for $n = 0, 1, \ldots, k-1$. In this case, to predict the final time solution c^k_{new} given the initial time solution ${c_{\text{new}}}^0$, we use k different networks $\mathcal{N}_1, \cdots, \mathcal{N}_k$

$$c^k_{\text{new}} \sim \mathcal{N}_k(\mathcal{N}_{k-1} \cdots \mathcal{N}_1(c^0_{\text{new}}, I^1) \cdots, I^{k-1}), I^k). \tag{5.22}$$

We remark that, besides the solution u^n at the previous time instant, the other input parameters I^{n+1} such as permeability or source terms can be different when entering the network at different time steps.

In this work, we would like to incorporate available observed data in the neural network. The

observation data will help to supplement the computational data which are obtained from the underlying reduced order model, and improve the performance of the neural network model such that it will take into account real data effects. From now on, we use $\{c_s^0, \cdots, c_s^k\}$ to denote the simulation data, and $\{c_o^0, \cdots, c_o^k\}$ to denote the observation data.

One can get the observation data from real field experiment. However in this work, we generate the observation data by running a new simulation on the "true permeability field" using standard finite element method, and using the results as observed data. For the computational data, we will perturb the "true permeability field", and use the reduced-order model, i.e., POD model for simulation. In the training process, we are interested in investigating the effects of observation data in the output. One can compare the performance of deep neural networks when using different combinations of computation and observation data.

For the comparison, we will consider the following three networks

- Network A: Use all observation data as output,

$$c_o{}^{n+1} \sim \mathcal{N}_o(c_s{}^n, I^{n+1}) \tag{5.23}$$

- Network B: Use a mixture of observation data and simulation data as output,

$$c_m{}^{n+1} \sim \mathcal{N}_m(c_s{}^n, I^{n+1}) \tag{5.24}$$

- Network C: Use all simulation data (no observation data) as output,

$$c_s{}^{n+1} \sim \mathcal{N}_s(c_s{}^n, I^{n+1}) \tag{5.25}$$

where c_m is a mixture of simulation data and observed data.

The first network (Network A) corresponds to the case when the observation data is sufficient. One can merely utilize the observation data in the training process. That is, the observation data at

time $n+1$ can be learnt as a function of the observation data at time n. This map will fit the real data very well given enough training data; however, it will not be able to approximate the reduced-order model. Moreover, in the real application, the observation data are hard to obtain, and in order to make the training effective, deep learning requires a huge amount of data. Thus Network A is not applicable in real case, and we will use the results from Network A as a reference.

The third network (Network C), on the other hand, will simply take all simulation data in the training process. In this case, one will get a network describes the simulation model (in our example, the POD reduced-order model) as best as it can but ignore the observational data effects. This network can serve as an emulator to do a fast simulation. We will also utilize Network C results as a reference.

We are interested in investigating the performance of Network B, where we take a combination of computational data and observational data to train. It will not only take in to account the underlying physics but also use the real data to modify the reduced-order model, thus resulting in a data-driven model.

5.2.2 Network structures

Mathematically, a neural network $\mathcal{N}$ of L layers with input $\mathbf{x}$ and output $\mathbf{y}$ is a function in the form

$$\mathcal{N}(\mathbf{x};\theta) = \sigma(W_L\sigma(\cdots\sigma(W_2\sigma(W_1\mathbf{x}+b_1)+b_2)\cdots)+b_L),$$

where $\theta := (W_1, \cdots, W_L, b_1, \cdots, b_L)$ is a set of network parameters, W's are the weight matrices and b's are the bias vectors. The activation function σ acts as entry-wise evaluation. A neural network describes the connection of a collection of nodes (neurons) sit in successive layers. The output neurons in each layer is simultaneously the input neurons in the next layer. The data propagate from the input layer to the output layer through hidden layers. The neurons can be switched on or off as the input is propagated forward through the network. The weight matrices W's control the connectivity of the neurons. The number of layers L describes the depth of the neural network. Figure 5.1 depicts a deep neural network in out setting, in which each circular node represents a

neuron and each line represents a connection from one neuron to another. The input layer of the neural network consists of the coefficients c^n and the input parameters I^n.

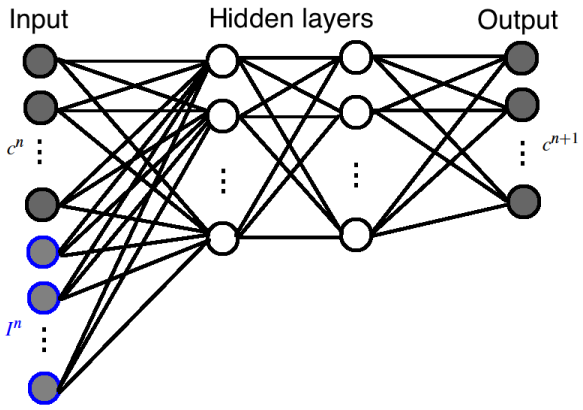

Figure 5.1: An illustration of deep neural network. Reprinted with permission from "Deep Global Model Reduction Learning in Porous Media Flow Simulation" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Eduardo Gildin, Yating Wang and Jingyan Zhang, 2020. Computational Geosciences, Volume 24, Pages 261–274, Copyright [2020] by Springer.

Given a set of data $(\mathbf{x}_j, \mathbf{y}_j)$, the deep neural network aims to find the parameters θ^* by solving an optimization problem

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \frac{1}{N} \sum_{j=1}^{N} ||\mathbf{y}_j - \mathcal{N}(\mathbf{x}_j; \theta)||_2^2,$$

where N is the number of the samples. Here, the function $L(\theta) = \frac{1}{N} \sum_{j=1}^{N} ||\mathbf{y}_j - \mathcal{N}(\mathbf{x}_j; \theta)||_2^2$ is known as the loss function. One needs to select suitable number of layers, number of neurons in each layer, the activation function, the loss function and the optimizers for the network.

As discussed in the previous section, we consider three different networks, namely $\mathcal{N}_o$, $\mathcal{N}_m$ and $\mathcal{N}_s$. For each of these networks, we take the vector $\mathbf{x} = ({c_s}^n, I^{n+1})$ containing the numerical solution vectors and the data at a particular time step as the input. In our setting, the input parameter I^{n+1}, if present, could be the static permeability field or the source function. Based on the avail-

ability of the observational data in the sample pairs, we will select an appropriate network among (5.23), (5.24) and (5.25) accordingly. The output $\mathbf{y} = {c_\alpha}^{n+1}$ is taken as the numerical solution at the next time instant, where $\alpha = o, m, s$ corresponds to the network.

Here, we briefly summarize the architecture of the network $\mathcal{N}_\alpha$, where $\alpha = o, m, s$ for three networks we defined in (5.23), (5.24) and (5.25) respectively.

As for the input of the network, we use $\mathbf{x} = ({c_s}^n, I^{n+1})$, which are the vectors containing the numerical solution vectors and the input parameters in a particular time step. The corresponding output data are $\mathbf{y} = {c_\alpha}^{n+1}$, which contains the numerical solution in the next time step. In between the input and output layer, we test on 3–10 hidden layers with 20-400 neurons in each hidden layer. In the training, there are $N = mk$ sample pairs of $(\mathbf{x}_j, \mathbf{y}_j)$ collected, where m is the number of realizations of flow dynamics and k is the number of time steps.

In between layers, we need the activation function. The ReLU function (rectified linear unit activation function) is a popular choice for activation function in training deep neural network architectures [73]. However, in optimizing a neural network with ReLU as activation function, weights on neurons which do not activate initially will not be adjusted, resulting in slow convergence. Alternatively, leaky ReLU can be employed to avoid such scenarios [74]. We choose leaky ReLU in our network structure. As for the training optimizer, we use AdaMax [75], which is a stochastic gradient descent (SGD) type algorithm well-suited for high-dimensional parameter space, in minimizing the loss function.

5.3 Numerical examples

In this section, we present numerical examples. We apply our method to predict the evolution of the pressure in a nonlinear single-phase flow problem. Using POD global model reduction technique, we obtain coefficients of numerical solutions in the reduced-order model and use as training samples to construct neural network approximations of the corresponding nonlinear flow dynamics. All the network training are performed using the Python deep learning API Keras [76].

As a first example, we consider a simple nonlinear single-phase flow in the spatial domain

$\Omega = [0,1] \times [0,1]$:

$$\frac{\partial u}{\partial t} - \text{div}(\kappa(x,u)\nabla u) = g \quad \text{in} \quad \Omega, \tag{5.26}$$

subject to homogeneous Dirichlet boundary condition $u|_{\partial\Omega} = 0$. This equation describes unsaturated flow in heterogeneous media, which are widely used [77, 78, 79, 80, 81]. In our simulations, we will use an exponential model $\kappa(x,u) = \kappa(x)\exp(\alpha u)$. Here, u is the pressure of flow, g is a time-dependent source term and α is a nonlinearity parameter. The function $\kappa(x)$ is a stationary heterogeneous permeability field of high contrast, i.e., with large variations within the domain Ω. In this example, we focus on permeability fields that contain wavelet-like channels as shown in Figure 5.2. In each realization of the permeability field, there are two non-overlapping channels with high conductivity values in the domain Ω, while the conductivity value in the background is 1. Channelized permeability fields are challenging for model reduction and prediction and, thus, we focus on flows corresponding to these permeability fields. The numerical tests for Gaussian permeability fields ([82, 83]) show a good accuracy because of the smoothness of the solution with respect to the parameters.

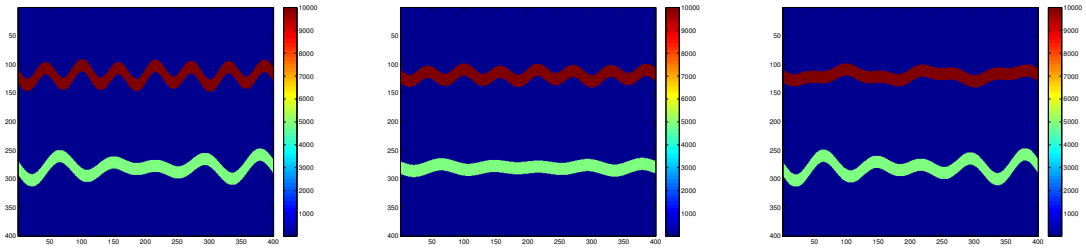

Figure 5.2: Samples of static permeability field used in single-phase flow. Reprinted with permission from "Deep Global Model Reduction Learning in Porous Media Flow Simulation" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Eduardo Gildin, Yating Wang and Jingyan Zhang, 2020. Computational Geosciences, Volume 24, Pages 261–274, Copyright [2020] by Springer.

Next, we present the details of numerical discretization of the problem. Suppose the spatial domain Ω is partitioned into a rectangular mesh $\mathcal{T}_h$, and a set of piecewise bilinear conforming

finite element basis functions $\{v_j\}$ is constructed on the mesh. We denote the finite element space by $V_h = \mathbb{Q}^1(\mathcal{T}_h)$. Using direct linearization for the nonlinear term, implicit Euler method for temporal discretization and a Galerkin finite element method for spatial discretization, the numerical solution u_h^{n+1} at the time instant $n+1$ is obtained by solving the following variational formulation: find $u_h^{n+1} \in V_h$ such that

$$\int_\Omega \frac{u_h^{n+1} - u_h^n}{\Delta t} v + \int_\Omega \kappa \exp(\alpha u_h^n) \nabla u_h^{n+1} \cdot \nabla v = \int_\Omega g^{n+1} v \quad \text{for all } v \in V_h. \tag{5.27}$$

Here Δt is the time step and h is the mesh size. With a slight abuse of notation, we again denote the coefficients of the numerical solution with the piecewise bilinear basis functions by u_h^{n+1}. Then, the variational formulation can be written in the matrix form

$$u_h^{n+1} = (M + \Delta t A(u_h^n))^{-1}(M u_h^n + \Delta t b^{n+1}), \tag{5.28}$$

where M, $A(u_h^n)$ and b^{n+1} are the mass matrix, the stiffness matrix and the load vector with respect to the bilinear basis functions v_j, i.e.,

$$\begin{aligned} M_{ij} &= \int_\Omega v_i v_j, \\ [A(u_h^n)]_{ij} &= \int_\Omega \kappa \exp(\alpha u_h^n) \nabla v_i \cdot \nabla v_j, \\ b_i^{n+1} &= \int_\Omega g^{n+1} v_i. \end{aligned} \tag{5.29}$$

In our simulation, the flow is simulated from an initial time $t = 0$ to a final time $t = 0.01$ in 10 time steps. Realizations of flow dynamics are computed using independent and uniformly distributed initial conditions. We use POD to extract dominant modes from snapshot solutions and construct POD nodal basis functions. Examples of POD nodal basis functions are shown in Figure 5.3. Simulation data of the dynamic process under the reduced-order model are then obtained and used in the training set. Different forms of inputs, depending on situations, are investigated. Using these data as samples, universal multi-layer networks are trained to approximate the flow

dynamics. We use the trained networks to predict the output with some new unseen inputs, and reconstruct the numerical solution using the predicted coefficients. We examine the quality of our networks by computing the L^2 error between our predicted solution u^n_{pred} and the reference solution u^n_{ref}, i.e.,

$$\begin{aligned} \|u^n_{\text{ref}} - u^n_{\text{pred}}\|_{L^2(\Omega)} &= \left(\int_\Omega |u^n_{\text{ref}} - u^n_{\text{pred}}|^2 \, dx \right)^{\frac{1}{2}}, \\ \|u^n_{\text{ref}} - u^n_{\text{pred}}\|_{H^1(\Omega)} &= \left(\int_\Omega |\nabla(u^n_{\text{ref}} - u^n_{\text{pred}})|^2 \, dx \right)^{\frac{1}{2}}, \end{aligned} \tag{5.30}$$

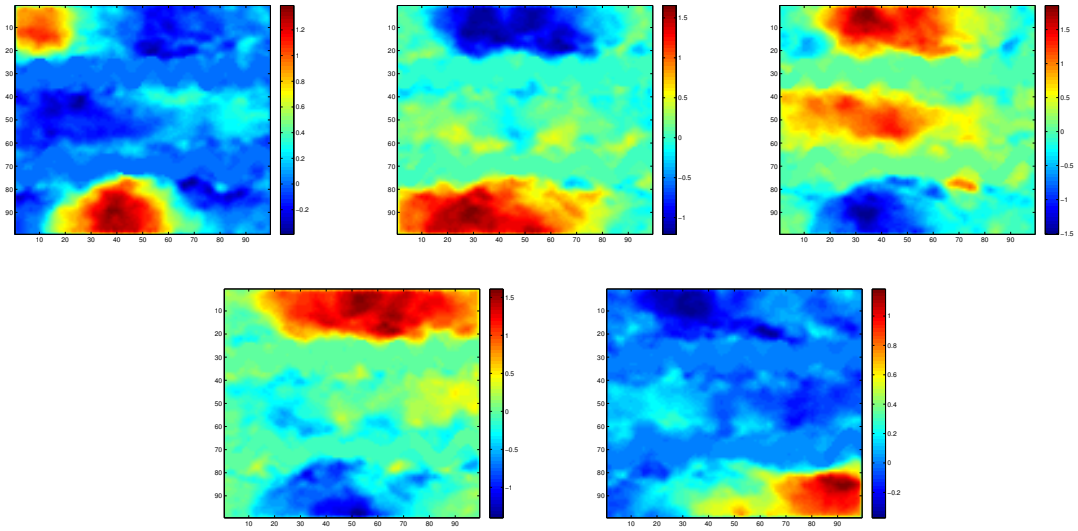

Figure 5.3: Illustration of nodal basis functions. Reprinted with permission from "Deep Global Model Reduction Learning in Porous Media Flow Simulation" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Eduardo Gildin, Yating Wang and Jingyan Zhang, 2020. Computational Geosciences, Volume 24, Pages 261–274, Copyright [2020] by Springer.

5.3.1 Experiment 1

In this experiment, we consider flow in a fixed static channelized field κ and a time-independent source g fixed among all the samples. The nonlinearity constant is chosen as $\alpha = 20$. We use

POD to extract 10 dominant modes from 1000 snapshot solutions and construct POD nodal basis functions. In the neural network, we simply take the input and output as

$$\mathbf{x} = {c_s}^n \quad \text{and} \quad \mathbf{y} = {c_s}^{n+1}. \tag{5.31}$$

In the generation of samples, we consider independent and uniformly distributed initial conditions ${c_s}^0$. We generate 100 realizations of initial conditions ${c_s}^0$, and evolve the reduced-order dynamic process to obtain ${c_s}^n$ for $n = 1, 2, \ldots, 10$. We remark that these simulation data provide a total of 1000 samples of the pushforward map.

We use the 900 samples given by 90 realizations as training set and the 100 samples given by 10 remaining realizations as testing samples. Using the training data and a given network architecture, we find a set of optimized parameter θ^* which minimizes the loss function, and obtain optimized network parameters θ^*. The network $\mathcal{N}$ is then used to predict the 1-step dynamic, i.e.,

$${c_s}^{n+1} \approx \mathcal{N}({c_s}^n; \theta^*). \tag{5.32}$$

We also use the composition of the network $\mathcal{N}$ to predict the final-time solution, i.e.,

$${c_s}^{10} \approx \mathcal{N}(\mathcal{N}(\cdots \mathcal{N}({c_s}^0; \theta^*) \cdots ; \theta^*); \theta^*). \tag{5.33}$$

We use the same set of training data and testing data and compare the performance of different network architectures. We examine the performance of the networks by the mean of L^2 percentage error of the 1-step prediction and the final-time prediction in the testing samples. The error is computed by comparing to the solution formed by the simulation data ${c_s}^n$.

The results are summarized in Table 5.1. It can be observed that if thee network architecture is too simple, i.e. contains too few layers or neurons, the neural network built may become useless in prediction.

Layer	Neuron	1-step	Final-time
3	20	0.1776	3.4501e+09
	100	0.0798	6.3276
	400	0.0613	4.9855
5	20	0.1499	6.5970e+06
	100	0.0753	6.3101
	400	0.0602	4.8137
10	20	0.1024	5.4183
	100	0.0750	4.3271
	400	0.0609	1.8834

Table 5.1: Mean of L^2 percentage error with different network architectures in Experiment 1. Reprinted with permission from "Deep Global Model Reduction Learning in Porous Media Flow Simulation" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Eduardo Gildin, Yating Wang and Jingyan Zhang, 2020. Computational Geosciences, Volume 24, Pages 261–274, Copyright [2020] by Springer.

5.3.2 Experiment 2

In the second experiment, we consider flow in static channelized fields κ and a time-independent source g fixed among all the samples. The nonlinearity constant is chosen as $\alpha = 10$. The coefficient fields κ differ in the conductivity value in channels. The high conductivity values in the two channels are parametrized by

$$\begin{aligned} \kappa_1 &= 10000e^{\eta_1}, \\ \kappa_2 &= 5000e^{\eta_2}, \end{aligned} \tag{5.34}$$

where $\eta = (\eta_1, \eta_2)$ is taken from a uniform distribution in $[-0.5, 0.5]^2$. We use POD to extract 5 dominant modes from 1000 snapshot solutions and construct POD nodal basis functions. In the neural network, we simply take the input and output as

$$\mathbf{x} = ({c_s}^n, \eta) \quad \text{and} \quad \mathbf{y} = {c_s}^{n+1}. \tag{5.35}$$

We generate 100 realizations of initial conditions ${c_s}^0$ and parameters η, and evolve the reduced-order dynamic process to obtain ${c_s}^n$ for $n = 1, 2, \ldots, 10$. We remark that these simulation data provide a total of 1000 samples of the pushforward map. We use the 900 samples given by 90 realizations as training set and the 100 samples given by 10 remaining realizations as testing samples. Using the training data and a given network architecture, we find a set of optimized parameter θ^* which minimizes the loss function, and obtain optimized network parameters θ^*. The network $\mathcal{N}$ is then used to predict the 1-step dynamic, i.e.,

$$ {c_s}^{n+1} \approx \mathcal{N}({c_s}^n, \eta; \theta^*). \tag{5.36} $$

We also use the composition of the network $\mathcal{N}$ to predict the final-time solution, i.e.,

$$ {c_s}^{10} \approx \mathcal{N}(\mathcal{N}(\cdots \mathcal{N}({c_s}^0, \eta; \theta^*) \cdots, \eta; \theta^*), \eta; \theta^*). \tag{5.37} $$

In this example, we investigate the advantage of our approach of combining deep learning with POD nodal basis functions. Instead of using the coefficients of the solution with respect to POD nodal basis functions $\{\psi_k\}_{k=1}^m$ for representing the flow dynamics, one can also use other discretizations, for example, the standard bilinear elements nodal functions or the POD basis functions $\{\mathbf{v}_j\}_{j=1}^m$ Using the same idea as in Section 5.2, we can learn from the respective data and construct corresponding neural networks for approximations. In this experiment, we compare the training cost and the performance of the neural networks using different underlying discretizations, by using the same set of training data and testing data. All networks consist of 3 hidden layers of 20 neurons and are trained in 500 epochs. We examine the performance of the networks by comparing the 1-step prediction and the final-time prediction to the corresponding numerical method.

A comparison of discretizations is presented in Table 5.2, which suggest that the model reduction technique brings several advantages to neural network approximation of flow dynamics. First, the use of POD reduces the number of trainable parameters in the network and thus short-

ening the elapsed time for network training. In our simple experiment, as shown in Table 5.2, the elapsed time for training the networks in POD reduced-order models is around 1/10 of elapsed time for training the networks in the standard nodal coordinates. Second, instead of extracting features solely in the learning process, the reduced order model predefines some features which are important in representing the flow and facilitates the learning process. This allows the information propagates more easily through the multi-layer networks and provides a smaller prediction error. Lastly, learning the evolution in the standard nodal coordinates becomes infeasible in large-scale computation. Both elapsed runtime for sample generation and memory required for sample storage grow dramatically with increased number of degree of freedom. The reduced-order model provides a cheap alternative for learning the flow dynamics in this scenario. As shown in Table 5.2, the CPU time for one forward run in the full model is 0.4499 seconds, which is short due to the simplicity of the linearization scheme in the simple experiment. However, with the reduced order model, the CPU time for a single forward run is reduced to 0.0003 seconds. We remark that the use of reduced-order models will be even more advantageous in complicated problems. For example, for repeatedly modelling highly nonlinear flows in highly heterogeneous flows, the nonlinear solver in the high-fidelity space will be computationally expensive. Moreover, the prediction error using POD nodal basis functions $\{\psi_k\}_{k=1}^m$ is smaller than using the original POD basis functions $\{\mathbf{v}_j\}_{j=1}^m$. This suggests that nodal values provide a more stable and well-conditioned coordinate system.

Remark 5.3.1. *The wide neural network using standard nodal coordinates can be viewed as a generalization of dynamic mode decomposition (DMD) [84]. DMD is a dimensionality reduction technique which extracting dynamical features from flow data. Given a sequence of snapshots* $\{u_h^0, u_h^1, \ldots, u_h^K\}$*, DMD seeks a linear mapping* A *which fits the snapshots by* $u_h^{n+1} = Au_h^n$*, which can be seen as the simplest neural network with linear activation function and without bias and hidden layers that maps* u_h^n *to* u_h^{n+1}*. Optimal mode decomposition (OMD) [85], a variant of DMD, seeks a linear mapping* A *with a user-defined rank* k*, which is equivalent to seek a wide neural*

Coordinates	Standard nodal	POD	POD nodal
Dimension	9801	5	5
Forward runtime (seconds)	0.4499	0.0003	0.0003
# trainable parameters	403161	1525	1525
Training time (seconds)	587.14	62.60	57.56
L^2 error for 1-step	0.9529%	0.5751%	0.3957%
H^1 error for 1-step	2.9588%	0.6020%	0.4395%
L^2 error for final time	4.8563%	3.4943%	3.0266%
H^1 error for final time	5.9270%	3.7307%	3.3762%

Table 5.2: History of training cost and prediction error with different discretization in Experiment 2. Reprinted with permission from "Deep Global Model Reduction Learning in Porous Media Flow Simulation" by Siu Wun Cheung, Eric T. Chung, Yalchin Efendiev, Eduardo Gildin, Yating Wang and Jingyan Zhang, 2020. Computational Geosciences, Volume 24, Pages 261–274, Copyright [2020] by Springer.

network in the form

$$u_h^{n+1} = W_2 W_1 u_h^n, \tag{5.38}$$

i.e. a 2-layer network with linear activation and no bias, and with k neurons in the immediate hidden layer. In this sense, we can build more general neural networks than DMD or OMD, which provides higher interpretability for more complex and nonlinear flow dynamics.

6. SUMMARY AND CONCLUSIONS

Lastly, we conclude this dissertation with a brief summary. Flow problems in porous heterogeneous media give rise to high-dimensional fine scale systems. In order to reduce the computational expense, we make use of rigorous mathematical tools to develop model reduction, statistical and machine learning approaches for efficient numerical solvers.

In Chapter 2, we present CEM-GMsDGM, a local multiscale model reduction approach in the discontinuous Galerkin framework. The multiscale basis functions are defined in coarse oversampled regions by a constraint energy minimization problem, which are in general discontinuous on the coarse grid, and coupled by the IPDG formulation. Thanks to the definition of local spectral problems, the dimension of auxiliary space is minimal for sufficiently representing the high conductivity regions, and provides the most locally compressed multiscale space. In our analysis for the Darcy flow problem, we show that the method provides optimal convergence in the coarse mesh size, which is independent of the contrast, provided that the oversampling size is appropriately chosen. The convergence of the method for solving Darcy flow is theoretically analyzed and numerically verified.

In Chapter 3, we present the CEM-GMsFEM for a dual continuum model. Auxiliary basis functions, obtained from local coupled spectral problems, are used to identify high contrast channels and fracture networks. Then, we solve an energy minimization with some constraints related to the auxiliary functions. We show that the basis functions are localized and that the resulting method has a mesh dependent convergence. Numerical results are presented to confirm the theory.

In Chapter 4, we propose a dynamic data-driven Bayesian approach for basis selection in multiscale problems, in the Generalized multiscale finite element method framework. The method is used to solve time-dependent problems in heterogeneous media with available dynamic observational data on the solution. Our method selects important degrees of freedom probabilistically. Using the construction of offline basis functions in GMsFEM, we choose the first few eigenfunctions with smallest eigenvalues as permanent basis functions and compute the fixed solution. The

fixed solution is used to compute the residual information, and impose a prior probability distribution on the rest of basis functions. The likelihood involves a residual and observational error minimization. The resultant posterior distribution allows us to compute multiple realizations of the solution, providing a probabilistic description for the un-resolved scales as well as regularizing the solution by the dynamic observational data. In our numerical experiments, we see that our sampling process quickly stabilizes at a steady state. We also see that the design of our likelihood and posterior is useful in reducing the error in observational data.

In Chapter 5, we combine some POD techniques with deep learning concepts in the simulations for flows in porous media. The observation data is given at some locations. We construct POD modes such that the degrees of freedom represent the values of the solution at certain locations. Furthermore, we write the solution at the current time as a multi-layer network that depends on the solution at the initial time and input parameters, such as well rates and permeability fields. This provides a natural framework for applying deep learning techniques for flows in channelized media. We provide the details of our method and present numerical results. In all numerical results, we study nonlinear flow equation in channelized media and consider various channel configurations. Our results show that multi-layer network provides an accurate approximation of the forward map and can incorporate the observed data. Moreover, by incorporating some observed data (from true model) and some computational data, we modify the reduced-order model. This way, one can use the observed data to modify reduced-order models which honor the observed data.

REFERENCES

[1] S. Brenner and L. Scott, *The Mathematical Theory of Finite Element Methods*. New York: Springer-Verlag, 2007.

[2] V. Girault and P.-A. Raviart, *Finite element approximation of the Navier-Stokes equations*, vol. 749 of *Lecture Notes in Mathematics*. Berlin: Springer-Verlag, 1979.

[3] D. Boffi, F. Brezzi, and M. Fortin, *Mixed Finite Element Methods and Applications*. Heidelberg: Springer-Verlag, 2013.

[4] I. Babuška, V. Nistor, and N. Tarfulea, "Generalized finite element method for second-order elliptic operators with Dirichlet boundary conditions," *J. Comput. Appl. Math.*, vol. 218, pp. 175–183, 2008.

[5] P. Bochev and M. Gunzburger, *Least-squares finite element methods*, vol. 166. Springer Science & Business Media, 2009.

[6] D. Arnold, R. Falk, and R. Winther, "Finite element exterior calculus, homological techniques, and applications," *Acta Numer.*, vol. 15, pp. 1–155, 2006.

[7] B. Rivière, *Discontinuous Galerkin methods for solving elliptic and parabolic equations: theory and implementation*. Society for Industrial and Applied Mathematics, 2008.

[8] D. Arnold, F. Brezzi, B. Cockburn, and L. Marini, "Unified analysis of discontinuous Galerkin methods for elliptic problems," *SIAM J. Numer. Anal.*, vol. 39, no. 5, pp. 1749–1779, 2001/02.

[9] B. Cockburn, J. Gopalakrishnan, and R. Lazarov, "Unified hybridization of discontinuous Galerkin, mixed, and continuous Galerkin methods for second order elliptic problems," *SIAM J. Numerical Analysis*, vol. 47, no. 2, pp. 1319–1365, 2009.

[10] L. Demkowicz and J. Gopalakrishnan, "An overview of the discontinuous Petrov Galerkin method," in *Recent Developments in Discontinuous Galerkin Finite Element Methods for Partial Differential Equations*, pp. 149–180, Springer International Publishing, 2014.

[11] D. Elfverson, E. Georgoulis, A. MÃělqvist, and D. Peterseim, "Convergence of a discontinuous galerkin multiscale method," *SIAM Journal on Numerical Analysis*, vol. 51, no. 6, pp. 3351–3372, 2013.

[12] G. Papanicolau, A. Bensoussan, and J.-L. Lions, *Asymptotic analysis for periodic structures*. Elsevier, 1978.

[13] X. Wu, Y. Efendiev, and T. Hou, "Analysis of upscaling absolute permeability," *Discrete and Continuous Dynamical Systems, Series B.*, vol. 2, pp. 158–204, 2002.

[14] T. Hou and X. Wu, "A multiscale finite element method for elliptic problems in composite materials and porous media," *J. Comput. Phys.*, vol. 134, pp. 169–189, 1997.

[15] Y. Efendiev, T. Hou, and X. Wu, "Convergence of a nonconforming multiscale finite element method," *SIAM J. Numer. Anal.*, vol. 37, pp. 888–910, 2000.

[16] Z. Chen and T. Y. Hou, "A mixed multiscale finite element method for elliptic problems with oscillating coefficients," *Mathematics of Computation*, vol. 72, no. 242, pp. 541–576, 2003.

[17] Y. Efendiev and T. Hou, *Multiscale Finite Element Methods: Theory and Applications*. Springer, 2009.

[18] C. Chu, I. Graham, and T. Hou, "A new multiscale finite element methods for high-contrast elliptic interface problem," *Mathematics of Computation*, vol. 79, pp. 1915–1955, 2010.

[19] T. Hughes, G. Feijóo, L. Mazzei, and J.-B. Quincy, "The variational multiscale method - a paradigm for computational mechanics," *Comput. Methods Appl. Mech Engrg.*, vol. 127, pp. 3–24, 1998.

[20] T. Hughes and G. Sangalli, "Variational multiscale analysis: the fine-scale Green's function, projection, optimization, localization, and stabilized methods," *SIAM Journal on Numerical Analysis*, vol. 45, no. 2, pp. 539–557, 2007.

[21] O. Iliev, R. Lazarov, and J. Willems, "Variational multiscale finite element method for flows in highly porous media," *Multiscale Model. Simul.*, vol. 9, no. 4, pp. 1350–1372, 2011.

[22] V. Calo, Y. Efendiev, and J. Galvis, "A note on variational multiscale methods for high-contrast heterogeneous porous media flows with rough source terms," *Advances in water resources*, vol. 34, no. 9, pp. 1177–1185, 2011.

[23] W. E and B. Engquist, "Heterogeneous multiscale methods," *Comm. Math. Sci.*, vol. 1, no. 1, pp. 87–132, 2003.

[24] A. Abdulle, "On a priori error analysis of fully discrete heterogeneous multiscale fem," *SIAM J. Multiscale Modeling and Simulation*, vol. 4, no. 2, pp. 447–459, 2005.

[25] W. E, P. Ming, and P. Zhang, "Analysis of the heterogeneous multiscale method for elliptic homogenization problems," *J. Amer. Math. Soc.*, vol. 18, no. 1, pp. 121–156, 2005.

[26] Y. Efendiev, J. Galvis, and T. Hou, "Generalized multiscale finite element methods," *Journal of Computational Physics*, vol. 251, pp. 116–135, 2013.

[27] E. Chung, Y. Efendiev, and G. Li, "An adaptive GMsFEM for high-contrast flow problems," *Journal of Computational Physics*, vol. 273, pp. 54–76, 2014.

[28] E. T. Chung, Y. Efendiev, R. L. Gibson Jr, and M. Vasilyeva, "A generalized multiscale finite element method for elastic wave propagation in fractured media," *GEM-International Journal on Geomathematics*, pp. 1–20, 2015.

[29] E. Chung, Y. Efendiev, and T. Y. Hou, "Adaptive multiscale model reduction with generalized multiscale finite element methods," *Journal of Computational Physics*, vol. 320, pp. 69–95, 2016.

[30] E. T. Chung, Y. Efendiev, and W. T. Leung, "An adaptive generalized multiscale discontinuous galerkin method (GMsDGM) for high-contrast flow problems," *arXiv preprint arXiv:1409.3474*, 2014.

[31] E. Chung, Y. Efendiev, and W. Leung, "An adaptive generalized multiscale discontinuous galerkin method for high-contrast flow problems," *SIAM Multiscale Modeling and Simulation*, vol. 16(3), pp. 1227–1257, 2018.

[32] Y. Efendiev, J. Galvis, and X. Wu, "Multiscale finite element methods for high-contrast problems using local spectral basis functions," *Journal of Computational Physics*, vol. 230, pp. 937–955, 2011.

[33] Y. Efendiev, J. Galvis, G. Li, and M. Presho, "Generalized multiscale finite element methods. Oversampling strategies," *International Journal for Multiscale Computational Engineering, accepted*, 2013.

[34] E. T. Chung, Y. Efendiev, and W. T. Leung, "Residual-driven online generalized multiscale finite element methods," *Journal of Computational Physics*, vol. 302, pp. 176–190, 2015.

[35] S. W. Cheung, E. T. Chung, Y. Efendiev, W. T. Leung, and M. Vasilyeva, "Constraint energy minimizing generalized multiscale finite element method for dual continuum model," *arXiv preprint arXiv:1807.10955*, 2018.

[36] J. S. R. Park, S. W. Cheung, T. Mai, and V. H. Hoang, "Multiscale simulations for upscaled multi-continuum flows," *arXiv preprint arXiv:1909.04722*, 2019.

[37] M. Wang, S. W. Cheung, E. T. Chung, M. Vasilyeva, and Y. Wang, "Generalized multiscale multicontinuum model for fractured vuggy carbonate reservoirs," *Journal of Computational and Applied Mathematics*, vol. 366, p. 112370, 2020.

[38] M. Vasilyeva, E. T. Chung, S. W. Cheung, Y. Wang, and G. Prokopev, "Nonlocal multicontinua upscaling for multicontinua flow problems in fractured porous media," *Journal of Computational & Applied Mathematics*, vol. 355, pp. 258–267, 2019.

[39] E. T. Chung, Y. Efendiev, and W. T. Leung, "Constraint energy minimizing generalized multiscale finite element method," *Computer Methods in Applied Mechanics and Engineering*, vol. 339, pp. 298–319, 2018.

[40] E. Chung, Y. Efendiev, and W. T. Leung, "Constraint energy minimizing generalized multiscale finite element method in the mixed formulation," *Computational Geosciences*, vol. 22, no. 3, pp. 677–693, 2018.

[41] S. W. Cheung, E. T. Chung, and W. T. Leung, "Constraint energy minimizing generalized multiscale discontinuous galerkin method," *arXiv preprint arXiv:1909.12461*, 2019.

[42] H. Owhadi, "Bayesian numerical homogenization," *Multiscale Modeling & Simulation*, vol. 13, no. 3, pp. 812–828, 2015.

[43] I. Bilionis, N. Zabaras, B. A. Konomi, and G. Lin, "Multi-output separable gaussian process: towards an efficient, fully bayesian paradigm for uncertainty quantification," *Journal of Computational Physics*, vol. 241, pp. 212–239, 2013.

[44] I. Bilionis and N. Zabaras, "Solution of inverse problems with limited forward solver evaluations: a bayesian perspective," *Inverse Problems*, vol. 30, no. 1, p. 015004, 2013.

[45] Y. Marzouk and D. Xiu, "A stochastic collocation approach to bayesian inference in inverse problems," *Communications in Computational Physics*, vol. 6, no. 4, pp. 826–847, 2009.

[46] M. Arnst, R. Ghanem, and C. Soize, "Identification of bayesian posteriors for coefficients of chaos expansions," *Journal of Computational Physics*, vol. 229, no. 9, pp. 3134–3154, 2010.

[47] A. M. Stuart, "Inverse problems: a bayesian perspective," *Acta Numerica*, vol. 19, pp. 451–559, 2010.

[48] N. Guha and X. Tan, "Multilevel approximate bayesian approaches for flows in highly heterogeneous porous media and their applications.," *Journal of Computational and Applied Mathematics*, vol. 317, pp. 700–717, 2017.

[49] K. Yang, N. Guha, Y. Efendiev, and B. Mallick, "Bayesian and variational bayesian approaches for flows in heterogeneous random media.," *Journal of Computational Physics*, vol. 345, pp. 275–293, 2017.

[50] Y. Efendiev, W. T. Leung, S. W. Cheung, N. Guha, V. H. Hoang, and B. Mallick, "Bayesian multiscale finite element methods. modeling missing subgrid information probabilistically," *International Journal for Multiscale Computational Engineering*, vol. 15, no. 2, pp. 175–197, 2017.

[51] S. W. Cheung and N. Guha, "Dynamic data-driven bayesian gmsfem," *Journal of Computational and Applied Mathematics*, vol. 353, pp. 72 – 85, 2019.

[52] G. Cybenko, "Approximations by superpositions of sigmoidal functions," *Mathematics of Control, Signals, and Systems*, vol. 2, no. 4, pp. 303–314, 1989.

[53] K. Hornik, "Approximation capabilities of multilayer feedforward networks," *Neural Networks*, vol. 4, no. 2, p. 251âĂŞ257, 1991.

[54] B. C. CsÃąji, "Approximation with artificial neural networks," *Faculty of Sciences, Etvs Lornd University*, vol. 24, no. 48, 2001.

[55] M. Telgrasky, "Benefits of depth in neural nets," *JMLR: Workshop and Conference Proceedings*, vol. 49, no. 123, 2016.

[56] H. M. Q. Liao and T. Poggio., "Learning functions: when is deep better than shallow," *arXiv:1603.00988v4*, 2016.

[57] B. Hanin, "Universal function approximation by deep neural nets with bounded width and relu activations," *arXiv:1708.02691*, 2017.

[58] Y. Khoo, J. Lu, and L. Ying, "Solving parametric pde problems with artificial neural networks," *arXiv:1707.03351*, 2017.

[59] E. Weinan and B. Yu, "The deep Ritz method: A deep learning-based numerical algorithm for solving variational problems," *Communications in Mathematics and Statistics*, vol. 6, no. 1, pp. 1–12, 2018.

[60] Z. Li and Z. Shi, "Deep residual learning and pdes on manifold," *arXiv:1708.05115.*, 2017.

[61] K. Wang and W. Sun, "A multiscale multi-permeability poroplasticity model linked by recursive homogenizations and deep learning," *Computer Methods in Applied Mechanics and Engineering*, vol. 334, pp. 337–380, 2018.

[62] Y. Wang and G. Lin, "Efficient deep learning techniques for multiphase flow simulation in heterogeneous porous media," *arXiv:1907.09571*, 2019.

[63] S. W. Cheung, E. T. Chung, Y. Efendiev, E. Gildin, Y. Wang, and J. Zhang, "Deep global model reduction learning in porous media flow simulation," *Computational Geosciences*, 2019.

[64] Y. Wang, S. W. Cheung, E. T. Chung, Y. Efendiev, and M. Wang, "Deep multiscale model learning," *Journal of Computational Physics*, vol. 406, no. 109071, 2020.

[65] M. Wang, S. W. Cheung, W. T. Leung, E. T. Chung, Y. Efendiev, and M. Wheeler, "Reduced-order deep learning for flow dynamics. the interplay between deep learning and model reduction," *Journal of Computational Physics*, vol. 401, no. 108939, 2020.

[66] M. Wang, S. W. Cheung, W. T. Leung, E. T. Chung, Y. Efendiev, and Y. Wang, "Prediction of discretization of gmsfem using deep learning," *Mathematics*, vol. 7, no. 5, 2019.

[67] J. Zhang, S. W. Cheung, Y. Efendiev, E. Gildin, and E. T. Chung, "Deep model reduction-model learning for reservoir simulation.," *SPE-193912-MS*.

[68] L. Kuo and B. Mallick, "Variable selection for regression models.," *Sankhya: The Indian Journal of Statistics, Series B*, pp. 65–81, 1998.

[69] E. I. George and R. McCulloch, "Variable selection via gibbs sampling.," *Journal of the American Statistical Association*, vol. 88, no. 423, pp. 881–889, 1993.

[70] J. G. Scott and J. Berger, "Bayes and empirical-bayes multiplicity adjustment in the variable-selection problem.," *The Annals of Statistics*, vol. 38, no. 5, pp. 2587–2619, 2010.

[71] M. Hinze and S. Volkwein, "Proper orthogonal decomposition surrogate models for nonlinear dynamical systems: error estimates and suboptimal control," in *Dimension Reduction of Large-Scale Systems* (P. Benner, V. Mehrmann, and D. Sorensen, eds.), vol. 45 of *Lecture Notes in Computational Science and Engineering*, pp. 261–306, Springer Berlin Heidelberg, 2005.

[72] G. Kerschen, J.-c. Golinval, A. F. Vakakis, and L. A. Bergman, "The method of proper orthogonal decomposition for dynamical characterization and order reduction of mechanical systems: An overview," *Nonlinear Dynamics*, vol. 41, no. 1, pp. 147–169, 2005.

[73] X. Glorot, A. Bordes, and Y. Bengio, "Deep sparse rectifier neural networks," in *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pp. 315–323, PMLR, 2011.

[74] A. Maas, A. Hannun, and A. Ng, "Rectifier nonlinearities improve neural network acoustic models," *Proc. icml*, vol. 30, no. 1, 2013.

[75] D. P. Kingma and J. Ba., "Adam: A method for stochastic optimization.," *arXiv preprint arXiv:1412.6980*, 2014.

[76] F. Chollet *et al.*, "Keras." `https://keras.io`, 2015.

[77] L. A. Richards, "Capillary conduction of liquids through porous mediums," *physics*, vol. 1, no. 5, pp. 318–333, 1931.

[78] W. Gardner, "Some steady-state solutions of the unsaturated moisture flow equation with application to evaporation from a water table," *Soil science*, vol. 85, no. 4, pp. 228–232, 1958.

[79] M. T. Van Genuchten, "A closed-form equation for predicting the hydraulic conductivity of unsaturated soils 1," *Soil science society of America journal*, vol. 44, no. 5, pp. 892–898, 1980.

[80] P. Dostert, Y. Efendiev, and B. Mohanty, "Efficient uncertainty quantification techniques in inverse problems for richardsâĂŹ equation using coarse-scale simulation models," *Advances in water resources*, vol. 32, no. 3, pp. 329–339, 2009.

[81] M. A. Celia, E. T. Bouloutas, and R. L. Zarba, "A general mass-conservative numerical solution for the unsaturated flow equation," *Water resources research*, vol. 26, no. 7, pp. 1483–1496, 1990.

[82] Y. Efendiev, A. Datta-Gupta, V. Ginting, X. Ma, and B. Mallick, "An efficient two-stage markov chain monte carlo method for dynamic data integration," *Water Resources Research*, vol. 41, no. 12, 2005.

[83] H. X. Vo and L. J. Durlofsky, "A new differentiable parameterization based on principal component analysis for the low-dimensional representation of complex geological models," *Mathematical Geosciences*, vol. 46, no. 7, pp. 775–813, 2014.

[84] P. J. Schmid, "Dynamic mode decomposition of numerical and experimental data," *Journal of Fluid Mechanics*, vol. 656, pp. 5–28, 2010.

[85] A. Wynn, D. S. Pearson, B. Ganapathisubramani, and P. J. Goulart, "Optimal mode decomposition for unsteady flows," *Journal of Fluid Mechanics*, vol. 733, pp. 473–503, 2013.