REDUCED ORDER MODEL AND UNCERTAINTY QUANTIFICATION FOR STOCHASTIC POROUS MEDIA FLOWS

A Dissertation

by

JIA WEI

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

DOCTOR OF PHILOSOPHY

August 2012

Major Subject: Mathematics

REDUCED ORDER MODEL AND UNCERTAINTY QUANTIFICATION FOR STOCHASTIC POROUS MEDIA FLOWS

A Dissertation

by

JIA WEI

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

DOCTOR OF PHILOSOPHY

Approved by:

Co-Chairs of Committee,	Yalchin Efendiev
	Akhil Datta-Gupta
Committee Members,	Raytcho Lazarov
	Jay Walton
Head of Department,	Emil Straube

August 2012

Major Subject: Mathematics

ABSTRACT

Reduced Order Model and Uncertainty Quantification for

Stochastic Porous Media Flows. (August 2012)

Jia Wei, B.S., Yunnan University, China;

M.S., Nankai University, China

Co-Chairs of Advisory Committee: Dr. Yalchin Efendiev
Dr. Akhil Datta-Gupta

In this dissertation, we focus on the uncertainty quantification problems where the goal is to sample the porous media properties given integrated responses. We first introduce a reduced order model using the level set method to characterize the channelized features of permeability fields. The sampling process is completed under Bayesian framework. We hence study the regularity of posterior distributions with respect to the prior measures.

The stochastic flow equations that contain both spatial and random components must be resolved in order to sample the porous media properties. Some type of upscaling or multiscale technique is needed when solving the flow and transport through heterogeneous porous media. We propose ensemble-level multiscale finite element method and ensemble-level preconditioner technique for solving the stochastic flow equations, when the permeability fields have certain topology features. These methods can be used to accelerate the forward computations in the sampling processes.

Additionally, we develop analysis-of-variance-based mixed multiscale finite element method as well as a novel adaptive version. These methods are used to study the forward uncertainty propagation of input random fields. The computational cost is saved since the high dimensional problem is decomposed into lower dimensional problems.

We also work on developing efficient advanced Markov Chain Monte Carlo methods. Algorithms are proposed based on the multi-stage Markov Chain Monte Carlo and Stochastic Approximation Monte Carlo methods. The new methods have the ability to search the whole sample space for optimizations. Analysis and detailed numerical results are presented for applications of all the above methods.

ACKNOWLEDGMENTS

First and foremost I would like to express deep gratitude to my advisor, Dr. Yalchin Efendiev, for all his guidance and support throughout my graduate study. His many insights and ideas have been invaluable to the successful completion of my graduate research. I would like to thank Dr. Akhil Datta-Gupta, Dr. Raytcho Lazarov and Dr. Jay Walton for not only serving on my committee, but also for providing many insightful comments and teaching me knowledge of numerical methods, mathematical modeling and petroleum engineering.

I would like to thank Dr. Guang Lin, for his helpful discussions on high dimensional representative techniques and sampling schemes. Thanks for his thoughtful guidance and helps during my internship at Pacific Northwest National Lab and consistent encouragement since then. I would like to thank Dr. Lijian Jiang for his patient help and discussions in our collaboration. I am also grateful to my undergraduate advisor Dr. Xiaohua Zhao and my master advisor Dr. Jishou Ruan for their everlasting support and encouragement.

I would like to thank the professors and staff of the Department of Mathematics at Texas A&M University. In addition, I wish to thank the entire numerical analysis group for providing a great environment for research. I want to thank Ms. Monique Stewart for her incredible patience and knowledge with the university administration.

I would like to also thank my fellow graduate students and friends, Yankun Gong, Mei Yang, Yan Li, Yi-Ching Wang, Yang Qi, Changchun Wang. Their great help both inside and outside my academic life and their friendship are very much appreciated. I also wish to thank everybody else ever helped me during my Ph.D. study.

Last, but certainly not least, I would like to thank my husband, Tao Wang, and

my parents, Xinrong Li and Hongbin Wei, for their incredible patience and encouragement throughout my graduate career.

TABLE OF CONTENTS

Page

CHAPTER Page

CHAPTER Page

LIST OF TABLES

TABLE Page

LIST OF FIGURES

FIGURE Page

FIGURE Page

FIGURE Page

FIGURE Page

CHAPTER I

INTRODUCTION

Uncertainties exist inherently in reservoir modeling and flow problems in heterogeneous porous media. There are uncertainties coming from modeling error, which reflects the differences between the mathematical models and physical world. This kind of uncertainty exists after certain models are chosen. We focus more on the uncertainties that are inherited from the description of subsurface characteristics, for example, porosity and permeability. Large uncertainties in reservoirs can greatly affect the prediction of production and the reservoir management. To better predict the performance of reservoir, these uncertainties need to be understood and quantified. Better decisions can be obtained based on a model with fewer uncertainties in the parameters. Our goal here is to quantify and reduce these uncertainties.

The uncertainty quantification of models including utilizing two types of reservoir data: static and dynamic data. The static data are time independent, such as the measurements of permeability and porosity. The dynamic data, for instance saturations, oil and water productions, are functions of time coming from production process. Different approaches are proposed to integrate these two kinds of data [20, 21, 79, 80, 98]. Among them, Bayesian framework [29, 34, 61, 67, 87] provides a good way to connect static and dynamic data to quantify the uncertainty in reservoir parameters. The quantification results of the parameters are described by a probability distribution function, called posterior. This distribution can be obtained through Bayes' rule from a combination of the prior and the likelihood distribution. The prior distribution of parameters is knowledge before any production data are observed.

This dissertation follows the style of the Journal of Computational Physics.

The likelihood function reveals the relations between the dynamic data and reservoir parameters. Markov chain Monte Carlo (MCMC) method and its modifications have been used to sample the posterior distribution of the permeability fields.

While Bayesian framework provides a natural way to combine static and dynamic data for uncertainty quantification, there exist several difficulties. The reservoir parameters are typically defined on a large number of grid blocks. This leads to high computational cost when we try to resolve the system. For MCMC method, forward simulation is needed for each sample. Since a large number of samples are necessary for the convergence of MCMC, the computational cost in forward model is the most expensive part in the whole algorithm. To accelerate the forward models, different methods are developed such as building surrogates based on general polynomial chaos [70, 104], using radial basis functions [72, 73], etc. On the other hand, MCMC method have low acceptance rate and local-trap problem [18, 54, 55, 56]. In both cases, more forward simulations are needed to overcome the problems. To improve, modifications of MCMC [34, 37, 61, 55, 56] are proposed.

In this dissertation, we try to make the uncertainty quantification of reservoir models more efficient in both ways. We start with building reduced order model to express the permeability fields of interest. The dimension of the problem gets reduced through proper parameterization. Then to accelerate the forward simulation, we propose a special ensemble-level mixed multiscale finite element method (MsFEM) and an ensemble-level preconditioner. We also propose algorithms to study the forward uncertainty propagation. These algorithms can serve as surrogates with less computational expense for forward simulations. For the sampling process itself, we combine the multi-stage MCMC [29, 34] with the Stochastic Approximation Monte Carlo (SAMC) [56] and Double Annealing SAMC (DASAMC) [18, 54, 55] to overcome local-trap problem.

In Chapter II, we cover the preliminary background material. We introduce the porous media flow equations that will be studied throughout the dissertation. We then introduce the Karhunen-Loève (K-L) expansion techniques for representing the random field information required in the two-phase immiscible flow equations. Then we discuss the multiscale methods involved in solving the porous media flow equations. Lastly, we present the Bayesian framework for inverse problems.

In Chapter III, we study permeability fields with channelized structures. The permeability facies are described by two-point correlation functions; while interfaces are represented via smooth pseudo-velocity fields in a level set formulation. Then one can reduce the dimension of the parameterization space by selecting dominant modes in K-L expansion. We study errors introduced in such truncations by estimating the difference in the expectation of a function with respect to full and truncated posterior. The estimation shows that this error can be bounded by the tail of K-L eigenvalues with a constant independent of the dimension of the space. To speed up Bayesian computations, we use an efficient two-stage MCMC that utilizes mixed MsFEM to screen the proposals. The numerical results show the validity of the proposed parameterization to channel geometry and the error estimations.

In Chapter IV, we consider ensembles of permeability fields with high contrast channels and inclusions. Our objective here is to construct special multiscale basis functions for the whole ensemble. By constructing the common coarse basis functions for the permeability fields with certain topological properties, the solution of elliptic equation can be obtained by projecting to the space spanned by these pre-computed basis functions. As the expensive part of the multiscle method is simplified, the forward simulation can be completed with less computational cost. We also apply this coarse multiscale approximation to the design of the two-level domain decomposition preconditioner. Numerical experiments show that the ensemble-level multiscale finite

element method converges to the fine scale solution, and the ensemble-level domain decomposition preconditioner condition number is independent of the high contrast in the coefficient. We also present numerical results of using the ensemble methods to inverse problem.

In Chapter V, we do model reduction in both stochastic space and physical space to treat the high dimensionality and heterogeneity efficiently in stochastic two-phase flows. We use an analysis-of-variance (ANOVA)-based mixed multiscale finite element method (MsFEM) to decompose the high dimensional problem into a set of lower dimensional problems, which requires less computational cost, and the mixed MsFEM can capture the heterogeneities on a coarse grid. To enhance the efficiency of traditional ANOVA, we develop a new adaptive ANOVA method, where the most active dimensions can be selected before conducting ANOVA decomposition. A number of numerical examples in two-phase stochastic flows are presented and showed the performance of the ANOVA-based mixed MsFEM.

In Chapter VI, we combine the multi-stage MCMC with SAMC and DASAMC methods to overcome the local-trap problem in uncertainty quantification. The multi-stage MCMC screens out the bad proposals by simulations in coarse scale. The computational expense can be saved since the acceptance rate is increased. SAMC and DASAMC have the ability to automatically lift the low permeability part of the posterior distribution and allow the samples to travel over more parts in the sample space. Our proposed multi-stage SAMC and multi-stage DASAMC combine both merits. Numerical results show the effective of our algorithms.

Lastly, in Chapter VII, we summarize our findings and present possibilities for future research.

CHAPTER II

BACKGROUND MATERIALS

In this chapter, the background materials and notations, which are necessary for the later chapters, are introduced. We first introduce the porous media flow equations as our geological model. Secondly, we introduce the parameterization of permeability fields, i.e., the Karhunen-Loève (K-L) expansion is discussed, since it will be used extensively throughout the later discussions. Then we discuss the multiscale methods involved in solving the porous media flow equations. Lastly, we present the Bayesian framework for inverse problems.

2.1. Geological model

A significant part of the computational expense in any dynamic data integration method is the modeling of flow and transport through high-resolution geologic models. We consider two-phase flow in a subsurface formation (denoted by D) under the assumption that the displacement is dominated by viscous effect. For clarity of exposition, we neglect the effects of gravity, compressibility, and capillary pressure, although our proposed approach is independent of the choice of physical mechanisms. Also, the porosity will be considered to be constant. The two phases will be referred to water and oil (or a non-aqueous phase liquid), designated by subscripts w and o, respectively. We write Darcy's law for each phase as

$$v_j = -\frac{k_{rj}(S)}{\mu_j} k(x,\omega)\nabla p, \tag{2.1}$$

where v_j is the phase velocity, $k(x,\omega)$ is the permeability tensor, k_{rj} is the relative permeability to phase j ($j = o, w$), S is the water saturation (volume fraction) and

p is the pressure. The ω denotes that $k(x,\omega)$ is a realization of a random space Ω. Combining Darcy's law with a statement of conservation of mass allows us to express the governing equations in terms of pressure and saturation equations

$$\nabla \cdot (\lambda(S)k(x,\omega)\nabla p) = Q_s, \tag{2.2}$$

$$\frac{\partial S}{\partial t} + \ v \cdot \nabla f(S) = 0, \tag{2.3}$$

where λ is the total mobility, Q_s is a source term, f is the fractional flux of water, and v is the total velocity, which are respectively given by

$$\lambda(S) = \frac{k_{rw}(S)}{\mu_w} + \frac{k_{ro}(S)}{\mu_o},$$

$$f(S) = \frac{k_{rw}(S)/\mu_w}{k_{rw}(S)/\mu_w + k_{ro}(S)/\mu_o},$$

$$v = v_w + v_o = -\lambda(S)k \cdot \nabla p.$$

The above descriptions are referred to as the geological model of the two-phase flow problem.

We use various production characteristics solving from the governing system (2.1)-(2.3) for different reasons in the later chapters. We compute the saturation S at different pore volume injected (PVI). PVI represents dimensionless time and is computed via

$$\text{PVI} = \int \frac{Q}{V_p} dt,$$

where V_p is the total pore volume of the system, $Q = \int_{\partial D^{out}} v \cdot n ds$ is the total flow rate and ∂D^{out} is the outflow boundary. The fraction of water produced in relation to the total production rate, denoted by $F(t)$ (or F in future discussion). $F(t)$ for a two-phase water-oil flow, is defined as the fraction of water in the produced fluid, i.e.,

$$F(t) = \frac{q_w(t)}{q_w(t) + q_o(t)},$$

where q_w and q_o are the flow rates of water and oil at the production edge of the model at time t. Then,

$$F(t) = \frac{\int_{\partial D^{out}} v_n f(S) dl^F}{\int_{\partial D^{out}} v_n dl^F}, \tag{2.4}$$

where v_n is normal velocity field. We also monitor the breakthrough time T_w defined as $F^{-1}(t)$ at the producer and the cumulative oil production Q_o at t PVI, i.e.,

$$Q_o = -\frac{1}{\int_D \varphi dx} \int_0^t \left(\int_D \min(q_o(x,\tau), 0) dx \right) d\tau.$$

By analyzing these production characteristics, we can understand the behavior of flows in the subsurface, and make predictions and decisions based on this information.

2.2. Karhunen-Loève expansion

In order to model the flow in heterogeneous porous media accurately through the system (2.1)-(2.3), a good parameterization of the permeability fields $k(x,\omega)$ is important. In this section, we introduce the K-L expansion we used to parameterize permeability fields.

Traditional geostatistical techniques for subsurface characterization rely on two-point correlation functions to describe the spatial variability. We assume that the permeability field $k(x,\omega)$ follows a log-Gaussian distribution. We then consider the function $a(x,\omega) = \log(k(x,\omega))$, instead of $k(x,\omega)$ for convenience. Our parametrization of permeability field starts from the two-point correlation function of $a(x,\omega)$, i.e.

$$R(x,y) = \mathrm{cov}[a](x,y) = E[(a(x,\omega) - E[a(x,\omega)])(a(y,\omega) - E[a(y,\omega)])],$$

where $E[\cdot]$ refers to the expectation and $x, y \in D$ are points in the spatial domain. In this work, $R(x,y)$ is assumed to be known. If R is unknown, with other information

sufficient to define permeability fields, our proposed methods still work. For permeability field given by a two-point correlation function, the Karhunen-Loève (K-L) expansion [42, 102] can be used to get an expression for $k(x,\omega)$ or $a(x,\omega)$. Further, a description with possibly fewer degrees of freedom can be obtained based on the expansion. More specifically, the expansion is done by representing the permeability field in terms of an optimal L^2 basis. By truncating the expansion, we can represent the permeability matrix by a small number of random parameters.

We briefly recall some properties of the K-L expansion. For simplicity, we assume that $E[a(x,\omega)] = 0$. Suppose $a(x,\omega)$ is a second order stochastic process with $E[\int_D a^2(x,\omega)dx] < \infty$, given an orthonormal basis $\{\psi_i\}$ in $L^2(D)$, we can expand $a(x,\omega)$ as a general Fourier series

$$a(x,\omega) = \sum_i a_i(\omega)\psi_i(x), \quad a_i(\omega) = \int_D a(x,\omega)\psi_i(x)dx.$$

The special L^2 basis $\{\psi_i\}$ that makes the random variables a_i uncorrelated is of interest here. Namely, $E(a_i a_j) = 0$ for all $i \neq j$. The basis functions $\{\psi_i\}$ satisfy

$$E(a_i a_j) = \int_D \psi_i(x)dx \int_D R(x,y)\psi_j(y)dy = 0, \quad i \neq j.$$

Since $\{\psi_i\}$ is a complete basis in $L^2(D)$, it follows that $\psi_i(x)$ are eigenfunctions of $R(x,y)$,

$$\int_D R(x,y)\psi_i(y)dy = \lambda_i\psi_i(x), \quad i = 1, 2, \cdots, \tag{2.5}$$

where $\lambda_i = E[a_i^2] > 0$. Furthermore, we have

$$R(x,y) = \sum_i \lambda_i\psi_i(x)\psi_i(y).$$

This is saying that the function $R(x,y)$ induces an integral operator $T_a : L^2(D) \longrightarrow$

$L^2(D)$ by

$$T_a g(\cdot) = \int_D R(x,\cdot) g(x) dx \quad \forall g \in L^2(D).$$

The operator T_a is compact and self-adjoint. The eigenpairs $(\lambda_i, \psi_i(x))_{i\geq 1}$ of T_a satisfy

$$(\psi_i, \psi_j)_{L^2(D)} = \delta_{ij}, \quad \lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_i \cdots, \quad \lim_{i\longrightarrow\infty} \lambda_i = 0,$$

where $(\cdot,\cdot)_{L^2}$ is the usual L^2 inner product. Define the mutually uncorrelated random variables by

$$\theta_i(\omega) := \frac{1}{\sqrt{\lambda_i}} \int_D a(x,\omega)\psi_i(x) dx, \quad i = 1, 2, \cdots,$$

i.e., denote $\theta_i = a_i/\sqrt{\lambda_i}$, then θ_i satisfies $E(\theta_i) = 0$ and $E(\theta_i\theta_j) = \delta_{ij}$. It follows that

$$a(x,\omega) = \sum_{i=1}^{\infty} \sqrt{\lambda_i}\theta_i(\omega)\psi_i(x), \tag{2.6}$$

where ψ_i and λ_i satisfy (2.5). The L^2 basis functions $\psi_i(x)$ are deterministic and resolve the spatial dependence of the permeability field. The randomness is represented by the scalar random variables θ_i. The expansion (2.6) is called the K-L expansion.

If we discretize the domain D by a rectangular mesh, the continuous K-L expansion (2.6) is reduced to finite terms and $\psi_i(x)$ are discrete fields. The discretized K-L expansion is given by

$$a_N = \sum_{i=1}^{N} \sqrt{\lambda_i}\theta_i\psi_i. \tag{2.7}$$

2.3. Multiscale methods

With the parameterization of $k(x,\omega)$ introduced in Section 2.2, the geological model discussed in Section 2.1 is completed. Solving this model for the quantities of interest usually requires large computational cost, because of the high dimensionality in spatial space. The system (2.1)-(2.3) is called fine-scale model. To reduce the

computational expense of the fine model, the mixed multiscale Finite Element Method (MsFEM) is utilized to solve the flow equation on a coarse grid and further use the velocity field on a coarse grid to compute the fractional flow that is the quantity of interest in our simulations.

As for the coarse-scale model, we will consider single-phase flow based multiscale simulation methods. This technique is similar to upscaling methods [23], except that instead of computing effective properties, multiscale basis functions are calculated. These basis functions are coupled through a variational formulation of the problem. For multi-phase flow and transport simulations, the conservative fine-scale velocity is often needed. For this reason, the mixed MsFEM is used. We refer to [1, 17] for mixed multiscale finite element and its use in two-phase flow and transport. In our simulations, the multiscale basis functions are computed for the velocity once with $\lambda = 1$. These basis functions are used later without any update for solving two-phase flow equations. As a result, we obtain a coarse-scale velocity field that is used for solving the transport equation on the coarse grid. We note that mixed MsFEM can be implemented on unstructured grids [32].

We present MsFEM used for spatial discretization here. To this end, we consider a second-order elliptic equation,

$$\begin{cases} -\mathrm{div}(k\nabla p) = f \quad \text{in} \quad D \\ -k\nabla p \cdot n = g \quad \text{on} \quad \partial D. \end{cases} \tag{2.8}$$

Eq. (2.8) describes the single-phase flow equation in porous media. The p refers to pressure, f refers to source (well or sink) and velocity $v = -k\nabla p$. For mixed

formulation, we rewrite Eq. (2.8) as

$$\begin{cases} k^{-1}v + \nabla p = 0 \ \text{ in } D \\ \operatorname{div}(v) = f \ \text{ in } D \\ v \cdot n = g(x) \ \text{ on } \partial D. \end{cases} \tag{2.9}$$

The weak formulation of (2.9) reads: seek $(v, p) \in H(\operatorname{div}, D) \times L^2(D)/R$ such that $v \cdot n = g$ and

$$\begin{cases} \left\langle k^{-1}v, u \right\rangle - \left\langle \operatorname{div}(u), p \right\rangle = 0 \quad \forall u \in H_0(\operatorname{div}, D) \\ \left\langle \operatorname{div}(v), q \right\rangle = \left\langle f, q \right\rangle \quad \forall q \in L^2(D). \end{cases} \tag{2.10}$$

Let $V_h \subset H(\operatorname{div}, D)$ and $Q_h \subset L^2(D)/R$ be the finite element spaces for velocity and pressure, respectively. We use the mixed MsFEM [17, 32] for (2.10). It means that mixed finite element approximation is performed on coarse grid, where the finite element basis functions are defined. In the mixed MsFEM, we use piecewise constant basis functions on a coarse grid for pressure. For the velocity, we define multiscale velocity basis functions. The degree of freedom of the multiscale velocity basis function is defined on the interface of the coarse grid. Let e_i^K be a generic edge or face of the coarse block K. The multiscale basis equation associated with e_i^K is defined by

$$\begin{cases} -\operatorname{div}(k \nabla \mathrm{w}_i^K) = \dfrac{1}{|K|} \quad \text{in } K \\ -k \nabla \mathrm{w}_i^K \cdot n = \begin{cases} b_i^K \ \text{ on } e_i^K \\ 0 \ \text{ else.} \end{cases} \end{cases} \tag{2.11}$$

For local mixed MsFEM [17], $b_i^K = \frac{1}{|e_i^K|}$. If the media demonstrate strong non-local features including channels, fracture and shale barriers, some global information is needed to define the boundary condition b_i^K for better accuracy of approximation

[1, 50]. Then,

$$\psi_i^K = -k\nabla \mathrm{w}_i^K, \tag{2.12}$$

defines the multiscale velocity basis function associated to e_i^K, and the multiscale finite dimensional space for velocity is defined by

$$V_h = \bigoplus_{K,i} \psi_i^K.$$

For each edge e_i, the basis functions can be combined in adjacent coarse-grid blocks. We denote the basis function for the edge e_i by ψ_i. Let K_1 and K_2 be two adjacent coarse-grid blocks, then ψ_i solves (2.11) in K_1 and solves $\mathrm{div}(\psi_i) = -1/|K_2|$ in K_2, and $b_i^{K_2} = -1/|e_i^{K_2}|$ on $e_i^{K_2}$ and 0 otherwise. More precisely, $\psi_i = \psi_i^{K_1}$ in K_1 and $\psi_i = -\psi_i^{K_2}$ in K_2, where ψ_i^K is defined in (2.12) (see Figure 2.1 for illustration).

Define $V_h^0 = V_h \cap H_0(\mathrm{div}, D)$. The numerical mixed formulation is to find $(v_h, p_h) \in V_h \times Q_h$ such that $v_h \cdot n = g_h$ on ∂D and

$$\begin{cases} \left\langle k^{-1} v_h, u_h \right\rangle + \left\langle \mathrm{div}(u_h), p_h \right\rangle = 0 \quad \forall u_h \in V_h^0 \\ \left\langle \mathrm{div}(v_h), q_h \right\rangle = \left\langle f, q_h \right\rangle \quad \forall q_h \in Q_h. \end{cases}$$

2.4. Bayesian framework

Our objective here is to sample the permeability fields, i.e., reservoir parameters, given flow and transport data and the known permeabilities at some spatial locations corresponding to wells. For a given permeability field $k(x, \omega)$, we denote the observed flow and transport data as F_{obs}. The Bayesian approach is based on Bayes' rule, i.e.,

$$\pi(k|F_{obs}) = \frac{L(F_{obs}|k)\pi_0(k)}{\pi(F_{obs})},$$

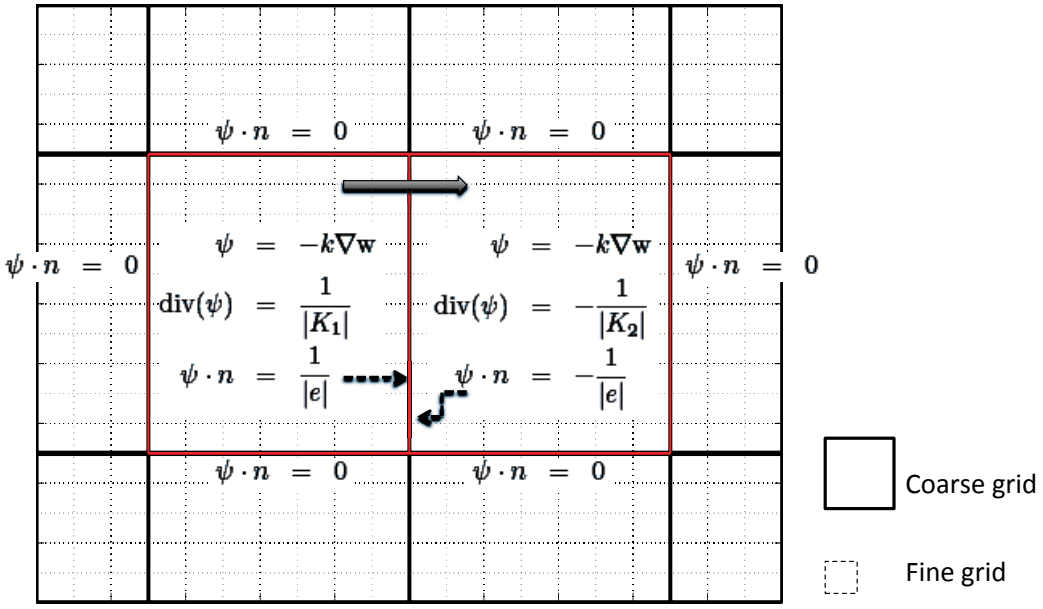

Fig. 2.1. Schematic description of a velocity basis function for an edge combining adjacent basis functions.

where k is the parameter, L is the likelihood function and π_0 is the prior density. $\pi(k|F_{obs})$ is the desired posterior distribution, denoted just as $\pi(k)$ in some situations.

Bayesian method has the ability to incorporate prior knowledge. A combination of knowledge from both prior and likelihood functions (Figure 2.2) gives the posterior distribution (Figure 2.3) is an illustration for this property. The results of Bayesian method approximate the maximum likelihood estimation results as the sample size increases. Instead of getting estimation with a confidence interval, it provides more quantities of interest. It is also easy to set up and estimate difficult models.

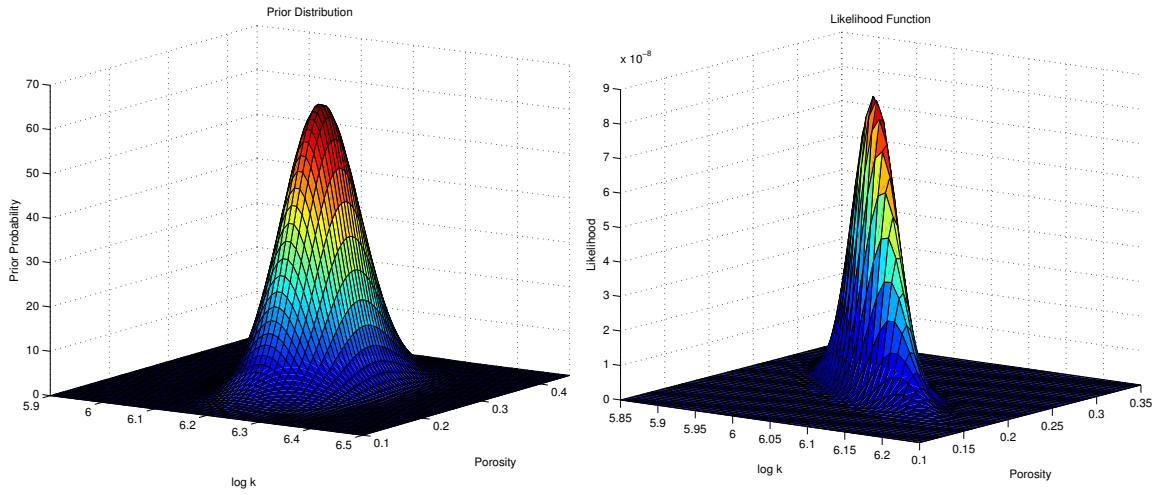

Fig. 2.2. Left: Prior distribution. Right: Likelihood function.

In our case, the prior information $\pi_0(k)$ is given by geologist as known. To find the posterior distribution, so that the uncertainty of the reservoir model can be quantified, we need a proper likelihood function. We denote F_k to be the integrated response corresponding to a given permeability field $k(x, \omega)$. There is a non-linear mapping from permeability field k to F_k, which is not one-to-one. As a consequence, many different permeability realizations may exist for a given set of production data, so random noise should be added to the generated flow and transport data from the simulator. In addition to that, F_k also contains measurement errors. We define the combined model error and measurement error as a random error ϵ. The model can be written as

$$F_{obs} = F_k + \epsilon,$$

where ϵ is distributed as $N(0, \sigma_f^2 I)$, i.e., $L(F_{obs}|k)$ is assumed to be $N(F_k, \sigma_f^2 I)$. So the likelihood function has the form

$$L(F_{obs}|k) \propto \exp\left(-\frac{\|F_{obs} - F_k\|^2}{\sigma_f^2}\right).$$

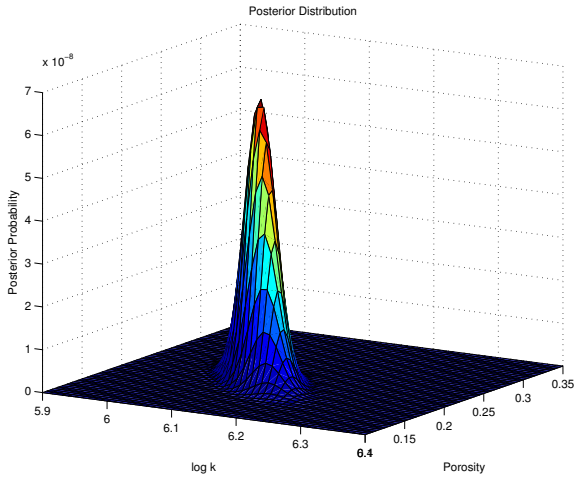

Fig. 2.3. Posterior distribution.

Under these assumptions, the posterior distribution we are interested in is

$$\pi(k) := \pi(k|F_{obs}) \propto \exp\left(-\frac{\|F_{obs} - F_k\|^2}{\sigma_f^2}\right)\pi_0(k).$$

The rest of the dissertation will focus on solving problems in finding this posterior distribution.

CHAPTER III

REDUCED ORDER MODEL AND BAYESIAN POSTERIOR REGULARIZATION

In this chapter, we study uncertainty quantification for flows in heterogeneous porous media. Permeability fields within facies are assumed to be described by two-point correlation functions, while interfaces that separate facies are represented via smooth pseudo-velocity fields in a level set formulation to get reduced dimensional parameterization. The permeability fields within facies and velocity fields representing interfaces can be described using Karhunen-Loève (K-L) expansion, where one can reduce the dimension of the parameterization space by selecting dominant modes in K-L expansion. We study errors introduced in such truncations by estimating the difference in the expectation of a function with respect to full and truncated posterior. The estimation shows that this error can be bounded by the tail of K-L eigenvalues with a constant independent of the dimension of the space. The fact that the constants are independent of the dimension is important to guarantee the feasibility of truncations with respect to posterior distributions. To speed up Bayesian computations, we use an efficient two-stage Markov chain Monte Carlo (MCMC) that utilizes mixed MsFEM to screen the proposals. The numerical results show the validity of the proposed parameterization to channel geometry and the error estimations.

This chapter is organized with first preliminary and motivation, followed with the parameterization of channelized permeability fields. Then the next section is devoted to the estimation of the posterior error due to the truncation in the prior parameterization. After that we briefly describe the sampling algorithms. Numerical results are presented in the last section.

3.1. Introduction

The distribution of subsurface properties is mainly controlled by the location of distinct geologic facies with sharp contrasts in properties, such as permeability and porosity, across facies boundaries [100]. E.g., in a fluvial setting, high permeability channel sands are often embedded in a nearly impermeable background causing the dominant fluid movement to be restricted within these channels. Under such conditions, the channel geometry plays an important role in determining the flow behavior in the subsurface. Consequently, in predicting the flow through highly heterogeneous porous formations, it is important to model facies boundaries accurately and to properly account for the uncertainties in these models.

Traditional geostatistical techniques for subsurface characterization have relied on two-point correlation functions to describe the spatial variability. Such spatial fields do not reproduce discrete and irregular geologic features, such as fluvial channels [22, 43, 51]. The success of object-based models, such as discrete Boolean or object-based models [35], is heavily dependent on the parameters to specify the object size, shapes, proportion and orientation. Typically, these parameters are highly uncertain, particularly in the early stages of subsurface characterization [16, 22]. For example, in a channel type environment, the channel sands may be observed at only a few well locations. There are many plausible channel geometries that will satisfy the channel sand and well intersections. Thus, the stochastic models for channels will require specification of random variables that govern the channel boundaries. All the parameters have considerable uncertainty associated with them and will impact fluid flow in the subsurface.

A considerable amount of prior information is typically available for building the facies models for fluid flow simulation [100]. These include well logs and cores,

seismic data and geologic conceptualization based on outcrops and analogues. Although the prior information plays a vital role in reducing uncertainty and preserving geologic realism, it is imperative that the geologic models reproduce the dynamic response based on flow and transport data. In the last decade, significant progress has been made in conditioning pixel-based geologic models to flow and transport data [14, 47, 51, 52, 80, 97, 101]. The approach typically involves the solution of an inverse problem requiring the minimization of a suitably defined objective function. Both gradient-based methods and combinatorial optimization methods have been used for this purpose. The existing approaches are not readily applicable to facies-based models where the primary goal is to locate the facies boundaries and preserve the contrast in facies properties.

In this chapter, we consider Bayesian hierarchical models that will preserve the facies architecture and at the same time populate the petrophysical properties within the facies in a geologically consistent manner by incorporating available static and dynamic information. To maintain the contrast in facies properties, we represent the facies boundaries using level sets that provide a systematic method for morphing the facies shapes to reconstruct a wide variety of facies geometries [81, 82, 90, 94]. Although level sets have recently been used to represent facies boundaries [99], the novelty of our proposed approach is in the efficient Bayesian hierarchical uncertainty quantification technique that we employ to perturb the facies boundaries and properties to match the dynamic response such as multiphase production history. The description of the facies boundaries in our level set approach will be based on parameterization of the pseudo-velocity fields that deform the interfaces. We will mostly focus on smooth interfaces that will require smooth velocity fields in the level set methods. The space of smooth velocity fields can be parameterized with fewer parameters, thus providing us with a small dimensional uncertainty space to explore.

One section of this chapter is devoted to studying the regularity of the posterior measure with respect to the prior measure. In particular, we estimate the difference in the expectation of a function with respect to full and truncated posterior distributions. Here, the full posterior distribution refers to the posterior computed using all parameter space, while the truncated posterior distribution refers to the posterior computed using truncated parameter space. The error in the fractional flow (the quantity that is often measured) is obtained in terms of the truncation error in K-L expansion. In particular, we show that the error is proportional to the truncation error in K-L expansion. Moreover, we show that the constants in these error estimates are independent of the dimension of the parameter space. The latter is important in our application, as the dimension of the parameter space which is decided by the dimension of discretization of the domain can be large. We note that some general principles for the regularity of the posterior measure with respect to the prior measure are introduced in [92].

A significant part of the computational expense in any dynamic data integration method is the modeling of flow and transport through high-resolution geologic models. To precondition these simulations, we adopt multi-stage MCMC approaches to minimize the number of fine-scale flow simulations during the MCMC sampling. In these approaches, simplified models using mixed MsFEM are used to screen the proposals before running detailed fine-scale simulations. Note that our forward model consists of coupled flow and transport equations and mixed MsFEM is used to solve the flow equation on a coarse grid and further use the velocity field on a coarse grid to compute the fractional flow that is the quantity of interest in our simulations. We note that there are a number of other methods developed for sampling the posterior based on emulators [64, 68, 69, 70].

In the last section, the numerical results are presented to investigate the theo-

retical bounds of the posterior error due to the truncation. We show some sampling results using the two-stage MCMC algorithm as well.

3.2. Parameterization of permeability fields

In this section, we introduce parameterization of the permeability fields. First, a heterogeneous permeability field is decomposed into several high and low permeable subregions, where each region represents a facies (see Figure 3.1 for illustration). The permeability field within each facies is assumed to follow a log-Gaussian distribution with a known spatial covariance. This type of hierarchical representation allows us to write of the permeability field as

$$k(x,\omega) = \sum_i k_i(x,\omega) I_{D_i}(x), \tag{3.1}$$

where I_{D_i} is an indicator function of region D_i (i.e., $I(x) = 1$ if $x \in D_i$ and $I(x) = 0$ otherwise). It is to be noted that in our approach the permeability field description is defined on a finite dimension whereas the partial differential equations (PDEs) to solve the forward problem are defined on an infinite dimensional setting.

3.2.1. Parameterization of interfaces

Suppose that any interface is a zero level set function $\varphi(x,\tau) = 0$. The evolution equation for an interface is given by

$$\frac{\partial \varphi}{\partial \tau} + w \cdot \nabla \varphi = 0, \tag{3.2}$$

where w is a pseudo-velocity field and τ is a pseudo-time. We denote φ_i as the i^{th} interface if there are more than one, then φ can be written as $\varphi = \varphi_i$ for different interfaces. More about level set method can be found in [81, 90]. A key is to specify

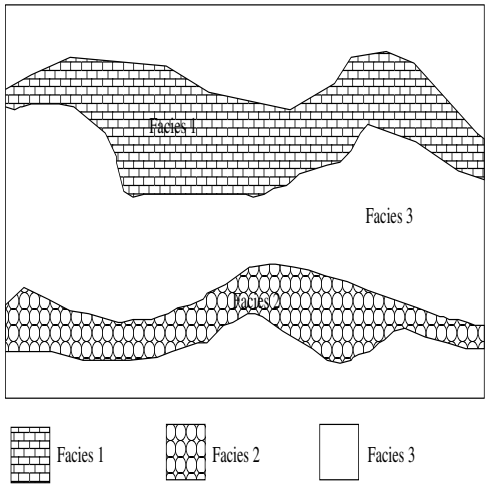

Fig. 3.1. Illustration of the permeability field with facies.

w for (3.2) to describe and update the boundaries of facies.

Now we consider a set of pseudo-velocity fields W, where $W = \{w|\ w$ admits fixed streamlines, and $|w|$ is constant along streamlines on $D\}$. Another words, we will assume that the streamlines are fixed and only magnitude of the velocity along different streamline changes. Note that streamlines, the integrated curves that are locally tangential to the velocity, are determined by the direction of a velocity field at each location. In general, one can also take streamlines to be random. To keep the dimension of the parameter space small, we will take streamlines to be fixed. For example, in our numerical experiment, vertical streamlines are used. We assume further that the magnitude of velocity field $w \in W$ follows the expansion,

$$|w| = \sum_i \alpha_i \phi_i(z), \ \ \alpha_i \sim N(0,1), \ z \in D'. \tag{3.3}$$

The functions $\phi_i(z)$'s are spatial basis for the magnitude of the velocity field and defined on the lower dimensional space of the interface, i.e., $\phi_i(z)$ lives in $D' \subset D$, where $\dim(D') = \dim(D) - 1$. For example, assume that $|w|$ is a second order stochastic process on D' with a given covariance structure, then $\phi_i = \sqrt{\lambda_i}\psi_i$ in (3.3)

is $L_2(D')$ basis. In this case, $|w|$ is expressed as a K-L expansion.

Now, if the initial interface is $\varphi(x_0, \tau_0) = 0$ at τ_0, the interface at $\tau_0 + \tau$ can be written as $\varphi(x_0 + \int_{\tau_0}^{\tau} w(\tau)d\tau, \tau_0 + \tau) = 0$. Any interface is corresponding to a pseudo-velocity field $w \in W$ and a time τ. Therefore, all interfaces of interest can be generated through the evolution equation (3.2), with a pair (w, τ). The following lemma proves that the set of interfaces generated through this one step movement is well defined. Otherwise, the map between the interface set and the pseudo-velocity field W is not one to one.

Lemma 3.2.1. *For any $\varphi(x, \tau) = 0$, $\exists\ \tilde{w} \in W$ with the expansion (3.3), such that $\varphi(x, \tau) = 0$ can be obtained from $\varphi(x_0, \tau_0) = 0$ through the evolution equation (3.2).*

Proof. For any $w_1, w_2 \in W$ with $|w_1| = \sum_i \alpha_{1i}\phi_i(z)$ and $|w_2| = \sum_i \alpha_{2i}\phi_i(z)$, $z \in D'$, the new interface formed by moving the initial $\varphi(x_0, \tau_0) = 0$ with w_1 and w_2 consecutively in time τ_1 and τ_2 is

$$\varphi(x_0 + \int_{\tau_0}^{\tau_1} w_1 d\tau + \int_{\tau_1}^{\tau_2} w_2 d\tau, \tau_0 + \tau_1 + \tau_2) = 0.$$

Assuming $\tau_0 = 0$, we can choose $\tau = \sqrt{\tau_1^2 + \tau_2^2}$, and let $\tilde{\alpha}_i = (\alpha_{1i}\tau_1 + \alpha_{2i}\tau_2)/\tau$, so $\tilde{\alpha}_i \sim N(0, 1)$. For $\tilde{w} \in W$ with $|\tilde{w}| = \sum_i \tilde{\alpha}_i\phi_i(z)$, we have the distance of any particle in an interface moved by (3.2) in a time interval τ is the arc length

$$\begin{aligned} |\tilde{w}|\tau &= \sum_i \tilde{\alpha}_i\phi_i(z)\tau = \sum_i \frac{\alpha_{1i}\tau_1 + \alpha_{2i}\tau_2}{\tau}\tau\phi_i(z) \\ &= \sum_i \alpha_{1i}\tau_1\phi_i(z) + \sum_i \alpha_{2i}\tau_2\phi_i(z) = |w_1|\tau_1 + |w_2|\tau_2. \end{aligned}$$

Since w_1, w_2 and $\tilde{w}$ have the same direction at any location, this implies that

$$\int_0^{\tau} \tilde{w} d\tau = \int_0^{\tau_1} w_1 d\tau + \int_{\tau_1}^{\tau_2} w_2 d\tau.$$

Therefore, the new interface

$$\varphi(x,\tau) = \varphi(x_0 + \int_0^{\tau} \tilde{w} d\tau, \tau) = \varphi(x_0 + \int_0^{\tau_1} w_1 d\tau + \int_{\tau_1}^{\tau_2} w_2 d\tau, \tau_0 + \tau_1 + \tau_2) = 0.$$

Namely, any interface can be obtained by moving the initial interface in a certain time period once by a $\tilde{w} \in W$, a Gaussian random field with deterministic direction. □

In our numerical experiments, we consider vertical streamlines in D. The pseudo-velocity is then $w = (w_x, w_y) = (0, w_y)$ and the magnitude along streamlines is assumed to be $w_y = |w| = \sum_i \alpha_i \phi_i(x)$, $\alpha_i \sim N(0,1)$, $x \in D'$. The Lemma holds for this case, and Figure 3.2 illustrates the process. To get simpler model, we also have numerical examples determining the velocity field via its values at certain discrete locations. The velocity values at these locations are updated as shown in Figure 3.3. In this case, the basis functions ϕ_i's in (3.3) are taken to be the indicator functions for each location, and α_i's are assumed to be linear between nodes.

In Section 3.3, we will show that one can use a truncated series for velocity $|w|$ to perform parameter estimation. Similar procedure can be performed for permeability field within facies. We will estimate associated errors in the resulting posterior distribution.

3.2.2. Parameterization within facies

Now we describe the parameterization of the permeability field within the facies. Each permeability field $k_i(x,\omega)$ follows a log-Gaussian distribution as assumed. We will then consider the function $a_i(x,\omega) = \log(k_i(x,\omega))$. We assume that the two-point correlation function of the log-permeability is known, i.e.,

$$R_i(x,y) = E[(a_i(x,\omega) - E[a_i(x,\omega)])(a_i(y,\omega) - E[a_i(y,\omega)])],$$

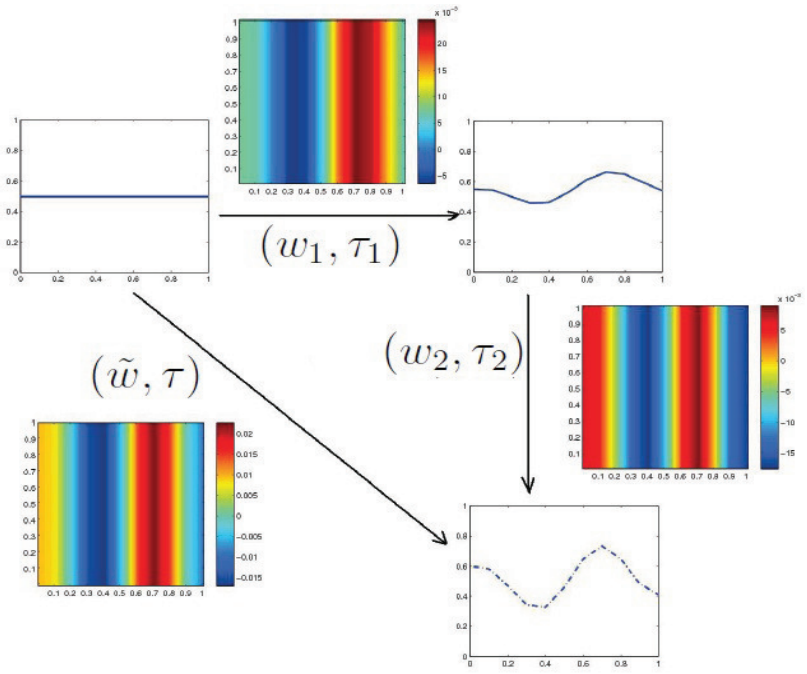

Fig. 3.2. Interface evolution by moving initial interface with different vertical velocity fields.

then the K-L expansion introduced in Section 2.2 is used. After discretizing the domain D and truncating the K-L expansion (2.7) with certain criteria, we can have the expansion

$$a_{N_i} = \sum_{j=1}^{N_i} \sqrt{\lambda_{ij}}\theta_{ij}\psi_{ij}.$$

3.3. Posterior error introduced by truncation

Our goal is to estimate the difference in the expected value of a function with respect to two different posteriors, where one of them is a truncation of the other. We discuss the regulation of posterior distribution first in permeability fields with single facies, and then generalize the theoretical result to permeability fields with channelized structures.

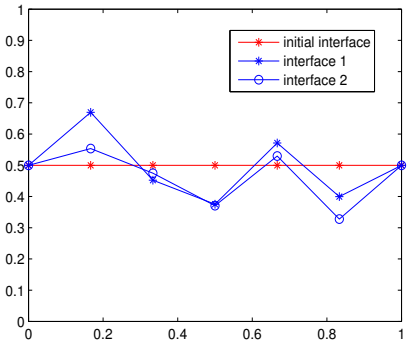

Fig. 3.3. Interface updates using velocity representation at some fixed points.

3.3.1. Single facies

We consider the domain $D = [0,1] \times [0,1]$ and assume that $\nabla p \in L_\infty(D)$, $k \in L_\infty(D)$ and $v \in L_\infty(D)$, where p is pressure, k is permeability field and v is velocity. The lemmas and theorems in this section are obtained under assumptions described in the following paragraph.

Assumptions: (i) $p = 1$ and $S = 1$ on $x = 0$; $p = 0$ on $x = 1$; and no flow boundary conditions on the lateral boundaries $y = 0$ and $y = 1$. (ii) The saturation is a smooth field. Note that if the velocity and initial conditions are smooth functions, then the saturation will be a smooth spatial field. (iii) The permeability field k is a stationary spatial process. (iv) The prior distribution is multivariate Gaussian distribution with identity covariance matrix.

Assume these assumptions hold, and then we first bound the difference between two saturation fields via the difference of the permeability fields in an appropriate norm.

Lemma 3.3.1. *$\|S_1 - S_2\|_{L_2(D)} \leq C\|k_1 - k_2\|_{L_2(D)}$, where S_1 and S_2 are water saturations.*

Proof. In order to get the estimation of saturations, we need the concept of time of

flight. For a particle that starts at a point $\wp$ at $t = 0$ and moves with velocity v, the flow map $P(\wp, T)$ is its position at time $t = T$, i.e.,

$$\frac{dP}{dT} = v(P), \quad P(\wp, 0) = \wp.$$

Time of flight T characterizes particle motion under the velocity field, since velocity is a function of the spatial variable

$$\frac{dT}{dP} = \frac{1}{v(P)}, \quad T = \int_{\wp}^{P} \frac{dr}{v(r)}.$$

Then, by [91] we have

$$\begin{aligned} \|S_1 - S_2\|_{L_2(D)} &\leq C\|T_1 - T_2\|_{L_2(D)} \leq C \left\| \int_{\wp}^{P} \frac{dr}{v_1(r)} - \int_{\wp}^{P} \frac{dr}{v_2(r)} \right\|_{L_2(D)} \\ &\leq C \left\| \int_{\wp}^{P} \frac{v_2(r) - v_1(r)}{v_1(r)v_2(r)} dr \right\|_{L_2(D)} \leq C\|v_2 - v_1\|_{L_2(D)}, \end{aligned} \tag{3.4}$$

since $v_1, v_2 \in L_\infty(D)$.

On the other hand, $v(x) = -k(x)\nabla p$, therefore,

$$\begin{aligned} \|v_1 - v_2\|_{L_2(D)} &= \|k_1\nabla p_1 - k_2\nabla p_2\|_{L_2(D)} \\ &\leq \|k_1\nabla(p_1 - p_2)\|_{L_2(D)} + \|k_1 - k_2\|_{L_2(D)}\|\nabla p_2\|_{L_\infty(D)} \\ &\leq \|k_1\nabla(p_1 - p_2)\|_{L_2(D)} + C\|k_1 - k_2\|_{L_2(D)}. \end{aligned}$$

Also, since $\mathrm{div}(k_1\nabla p_1) = 0, \mathrm{div}(k_2\nabla p_2) = 0$, then $\mathrm{div}(k_1\nabla p_1) - \mathrm{div}(k_2\nabla p_2) = 0$, and further $\mathrm{div}(k_1\nabla(p_1 - p_2)) = \mathrm{div}((k_2 - k_1)\nabla p_2)$, so

$$\begin{aligned} \|k_1\nabla(p_1 - p_2)\|_{L_2(D)} &= \|(k_2 - k_1)\nabla p_2\|_{L_2(D)} \\ &\leq \|k_1 - k_2\|_{L_2(D)}\|\nabla p_2\|_{L_\infty(D)} \leq C\|k_1 - k_2\|_{L_2(D)}. \end{aligned}$$

Therefore,

$$\|v_1 - v_2\|_{L_2(D)} \leq C\|k_1 - k_2\|_{L_2(D)}. \tag{3.5}$$

Then, from (3.4) and (3.5), we have

$$\|S_1 - S_2\|_{L_2(D)} \leq C\|k_1 - k_2\|_{L_2(D)}.$$

□

In Bayesian framework, the reference fractional flow or water-cut $F(k;t) = \int_0^t \int_0^1 v(1,y)S(1,\ y,t)dydt$ is matched to get the target posterior distribution. Next, we will estimate the difference between two water-cut responses via the corresponding permeability fields.

Lemma 3.3.2. *$|F(k_1;t) - F(k_2;t)|^2 \leq C\|k_1 - k_2\|_{L_2(D)}$, where k_1 and k_2 are permeabilities, $F(k_1;t)$ and $F(k_2;t)$ are water-cut functions.*

Proof. Note that

$$\begin{aligned} F(k;t) &= \int_0^t \int_0^1 v(1,y)S(1,y,t)dydt \\ &= \int_0^t \Big[\int_0^1 v(1,y)S(1,y,t)dy - \int_0^1 v(0,y)S(0,y,t)dy \Big] dt \\ &\quad + \int_0^t \int_0^1 v(0,y)S(0,y,t)dydt. \end{aligned}$$

Using $S(0,y,t) = 1$ and $S_t + v \cdot \nabla S = 0$, it follows that

$$\int_0^1 v(0,y)S(0,y,t)dy = \int_0^1 v(0,y)dy = \int_0^1 v(s,y)dy,$$

for any $s \in [0,1]$, since v is divergence free. Then,

$$\begin{aligned}
F(k_1;t) &= \int_0^t \Big[\int_{\partial D} v_1(x,y)S_1(x,y,t)dy\Big]dt + \int_0^t\int_0^1 v_1(0,y)dydt \\
&= \int_0^t \Big[\int_D div\{v_1(x,y)S_1(x,y,t)\}dxdy\Big]dt + \int_0^t\int_0^1 v_1(s,y)dydt \\
&= \int_0^t \Big[\int_D v_1(x,y)\cdot\nabla S_1(x,y,t)dxdy\Big]dt + \int_0^t\int_0^1 v_1(s,y)dydt \\
&= \int_0^t \Big[-\int_D (S_1)_t dxdy\Big]dt + \int_0^t\int_0^1 v_1(s,y)dydt \\
&= -\int_D S_1(x,y,t)dxdy + \int_D S_1(x,y,0)dxdy + \int_0^t\int_0^1 v_1(s,y)dydt.
\end{aligned}$$

There is a similar result for $F(k_2;t)$. Then,

$$\begin{aligned}
|F(k_1;t) - F(k_2;t)|^2 &= \Big|\int_D (S_2(x,y,t) - S_1(x,y,t))dxdy \\
&\quad + \int_D (S_1(x,y,0) - S_2(x,y,0))dxdy \\
&\quad + \int_0^t\int_0^1 (v_1(s,y) - v_2(s,y))dydt\Big|^2 \\
&\le C\Big(\int_D |(S_2(x,y,t) - S_1(x,y,t))|^2 dxdy \\
&\quad + \int_D |S_1(x,y,0) - S_2(x,y,0)|^2 dxdy \\
&\quad + \int_0^t\int_D |v_1(x,y) - v_2(x,y)|^2 dxdydt\Big) \\
&\le C\|k_1 - k_2\|^2_{L_2(D)},
\end{aligned}$$

by Lemma 3.3.1. □

Next, we consider the case with single facies and the permeability that is described via K-L expansion. In particular, we assume $k(x,\omega) = \exp(\sum_{i=1}^N \theta_i\psi_i(x))$ and consider the truncated expansion $k(x,\omega) = \exp(\sum_{i=1}^M \theta_i\psi_i(x))$. Then, the posterior distributions can be written as

$$\pi(\theta) \propto G(\theta_1,\cdots,\theta_N)\pi_0(\theta), \quad \widetilde{\pi}(\theta) \propto \widetilde{G}(\theta_1,\cdots,\theta_M)\pi_0(\theta),$$

where $\pi(\theta)$ is the posterior needed to be sampled, $\widetilde{\pi}(\theta)$ is an approximation of $\pi(\theta)$ and $\pi_0(\theta)$ is the prior distribution. $G(\theta_1, \cdots, \theta_N)$ and $\widetilde{G}(\theta_1, \cdots, \theta_N)$ are likelihoods, where

$$\begin{aligned} G(\theta_1, \cdots, \theta_N) &= \exp\Big(-\frac{\int_0^T |F_{obs} - F(k_1(\theta_1, \cdots, \theta_N); t)|^2 dt}{\sigma_f^2}\Big), \\ \widetilde{G}(\theta_1, \cdots, \theta_M) &= \exp\Big(-\frac{\int_0^T |F_{obs} - F(k_2(\theta_1, \cdots, \theta_M); t)|^2 dt}{\sigma_f^2}\Big). \end{aligned}$$

Next, we estimate the difference between G and $\widetilde{G}$.

Lemma 3.3.3. $|G(\theta_1, \cdots, \theta_N) - \widetilde{G}(\theta_1, \cdots, \theta_M)| \leq \dfrac{C}{\sigma_f^2} \|k_1 - k_2\|_{L_2(D)}$.

Proof. The permeability fields k_1 and k_2 we considered are coming from the posterior distributions, so $|F(k_1; t) - F_{obs}|$ and $|F(k_2; t) - F_{obs}|$ are bounded almost everywhere. Then, by Lemma 3.3.2

$$\begin{aligned} & |G(\theta_1, \cdots, \theta_N) - \widetilde{G}(\theta_1, \cdots, \theta_M)| \\ \leq\ & \frac{C}{\sigma_f^2} \left| \int_0^T |F_{obs} - F(k_1; t)|^2 dt - \int_0^T |F_{obs} - F(k_2; t)|^2 dt \right| \\ \leq\ & \frac{C}{\sigma_f^2} \Big(\int_0^T |2F_{obs} - F(k_2; t) - F(k_1; t)|^2 dt \Big)^{\frac{1}{2}} \cdot \Big(\int_0^T |F(k_1; t) - F(k_2; t)|^2 dt \Big)^{\frac{1}{2}} \\ \leq\ & \frac{C}{\sigma_f^2} \Big(\int_0^T |F(k_1; t) - F(k_2; t)|^2 dt \Big)^{\frac{1}{2}} \\ \leq\ & \frac{C}{\sigma_f^2} \|k_1 - k_2\|_{L_2(D)}. \end{aligned}$$

□

Theorem 3.3.1. *Suppose that the permeability field k is a stationary spatial process on a bounded region and $f(\theta)$ is square integrable with respect to a Gaussian measure, i.e., $\int |f(\theta)|^2 \pi_0(\theta) d\theta < \infty$, then*

$$\big| E_{\pi(\theta)}[f(\theta)] - E_{\widetilde{\pi}(\theta)}[f(\theta)] \big| \leq C \left\{ \sum_{i=M+1}^{N} \lambda_i \right\}^{\frac{1}{2}}, \tag{3.6}$$

where C is independent of dimension N.

Proof. If $f(\theta)$ is square integrable with respect to Gaussian measure (e.g., a polynomial function), we can show that

$$\begin{aligned}
&\big|E_{\pi(\theta)}[f(\theta)] - E_{\tilde{\pi}(\theta)}[f(\theta)]\big| \\
\leq\ & C\int |f(\theta)||G(\theta_1,\cdots,\theta_N) - \widetilde{G}(\theta_1,\cdots,\theta_M)|\pi_0(\theta)d\theta \\
\leq\ & \frac{C}{\sigma_f^2}\int |f(\theta)|\|k_1 - k_2\|_{L_2}\pi_0(\theta)d\theta \\
\leq\ & \frac{C}{\sigma_f^2}\Big(\int |f(\theta)|^2\pi_0(\theta)d\theta\Big)^{\frac{1}{2}}\Big(\int \|k_1 - k_2\|_{L_2}^2\pi_0(\theta)d\theta\Big)^{\frac{1}{2}} \\
\leq\ & \frac{C}{\sigma_f^2}\Big(\int \|k_1 - k_2\|_{L_2}^2\pi_0(\theta)d\theta\Big)^{\frac{1}{2}}.
\end{aligned}$$

To estimate the error of truncation of K-L expansion, let $k_1 = \exp(\sum_{i=1}^{N}\theta_i\sqrt{\lambda_i}\psi_i)$ and $k_2 = \exp(\sum_{i=1}^{M}\theta_i\sqrt{\lambda_i}\psi_i)$. We assume $\theta_i \sim N(0,1)$ for simplicity, then

$$\begin{aligned}
&\left|\int f(\theta)\pi(\theta)d\theta - \int f(\theta)\widetilde{\pi}(\theta)d\theta\right|^2 \\
\leq\ & \frac{C}{\sigma_f^4}\int \left\|\exp\Big(\sum_{i=1}^{N}\theta_i\sqrt{\lambda_i}\psi_i\Big) - \exp\Big(\sum_{i=1}^{M}\theta_i\sqrt{\lambda_i}\psi_i\Big)\right\|_{L_2}^2 \pi_0(\theta)d\theta \\
\leq\ & \frac{C}{\sigma_f^4}\int_D\int \exp\Big(2\sum_{i=1}^{M}\theta_i\sqrt{\lambda_i}\psi_i\Big)\Big[1 - \exp\Big(\sum_{i=M+1}^{N}\theta_i\sqrt{\lambda_i}\psi_i\Big)\Big]^2\pi_0(\theta)d\theta dxdy \\
\leq\ & \frac{C}{\sigma_f^4}\int_D I_1 I_2 dxdy,
\end{aligned}$$

where

$$\begin{aligned}
I_1 &= \int\cdots\int \exp\Big(2\sum_{i=1}^{M}\theta_i\sqrt{\lambda_i}\psi_i\Big)\pi_0(\theta_1,\cdots,\theta_M)d\theta_1\cdots d\theta_M \\
&= \prod_{i=1}^{M}\frac{1}{\sqrt{2\pi}}\int \exp\Big(-\frac{1}{2}(\theta_i^2 - 2\sqrt{\lambda_i}\psi_i)^2 + 2\lambda_i\psi_i^2\Big)d\theta_i = \exp\Big(2\sum_{i=1}^{M}\lambda_i\psi_i^2\Big),
\end{aligned}$$

because ψ_i's are bounded, and

$$\begin{aligned}
I_2 &= \int\cdots\int \Big[1-\exp\Big(\sum_{i=M+1}^{N}\theta_i\sqrt{\lambda_i}\psi_i\Big)\Big]^2\pi_0(\theta_{M+1},\cdots,\theta_N)d\theta_{M+1}\cdots d\theta_N \\
&= \int\cdots\int \Big\{1-2\exp\Big(\sum_{i=M+1}^{N}\theta_i\sqrt{\lambda_i}\psi_i\Big) \\
&\quad +\exp\Big(2\sum_{i=M+1}^{N}\theta_i\sqrt{\lambda_i}\psi_i\Big)\Big\}\prod_{i=M+1}^{N}\frac{1}{\sqrt{2\pi}}\exp\Big(-\frac{\theta_i^2}{2}\Big)d\theta_i \\
&\le 1-2\Big(1+\frac{1}{2}\sum_{i=M+1}^{N}\lambda_i\psi_i^2\Big)+1+2\sum_{i=M+1}^{N}\lambda_i\psi_i^2\Big(\exp\Big(2\sum_{i=M+1}^{N}\lambda_i\psi_i^2\Big)+\frac{1}{2}\Big) \\
&\le C\exp\Big(2\sum_{i=M+1}^{N}\lambda_i\psi_i^2\Big)\sum_{i=M+1}^{N}\lambda\psi_i^2.
\end{aligned}$$

Since k is a stationary spatial process on a bounded region, i.e., for a spatial process where the covariance function depends only on the distance not on the spatial location, then by [89], $\{\psi_i\}$ is uniform $L_\infty(D)$ bounded. So,

$$\begin{aligned}
&\Big|\int f(\theta)\pi(\theta_1,\cdots,\theta_N)d\theta-\int f(\theta)\widetilde{\pi}(\theta_1,\cdots,\theta_N)d\theta\Big| \le \frac{C}{\sigma_f^2}\left\{\int_D I_1I_2dxdy\right\}^{\frac{1}{2}} \\
\le\ & C\left\{\int_D\exp\Big(2\sum_{i=1}^{N}\lambda_i\psi_i^2\Big)\sum_{i=M+1}^{N}\lambda_i\psi_i^2dxdy\right\}^{\frac{1}{2}} \le C\left\{\sum_{i=M+1}^{N}\lambda_i\right\}^{\frac{1}{2}}.
\end{aligned}$$

□

3.3.2. Channelized cases

Consider a permeability field $k(x,\omega)$ in D, see Figure 3.1, which has s facies $\{D_i\}_{i=1}^{s}$ and $\tilde{s}$ interfaces $\{\varphi_i\}_{i=1}^{\tilde{s}}$. Each facies is described by a covariance matrix $R_i(x,y)$ as in Section 3.2.2. Then, the permeability field $k(x,\omega)$ is a function given

by

$$k(x,\omega) = \sum_{i=1}^{s} k_i(x,\omega) I_{D_i}(x),$$

where I_{D_i} is an indicator function on D_i. The permeability of each facies D_i is $k_i(x,\omega) = \exp\{a_i(x,\omega)\} = \exp\{\sum_{j=1}^{\infty} \sqrt{\lambda_{ij}}\theta_{ij}\psi_{ij}(x)\}$ by assumption and each interface is formed by moving the initial interface $\varphi_i(x,t_0) = 0$ by a velocity field with deterministic direction $w_i = \sum_{j=1}^{\infty} \alpha_{ij}\phi_{ij}(z)$ as in Section 3.2 (w_i is used to denote magnitude $|w_i|$ for simplicity). Then, the permeability field $k(x,\omega)$ can be written as

$$k(x,\theta,\alpha) = \sum_{i=1}^{s} \exp(a_i) I_{D_i(\alpha)}(x).$$

Considering the finite discretized case allows us to write a_i and w_i in each D_i as $a_{N_i} = \sum_{j=1}^{N_i} \sqrt{\lambda_{ij}^{(\theta)}}\ \theta_{ij}\psi_{ij}(x)$, $i = 1,\cdots,s$, $x \in D$ and $w_{\tilde{N}_i} = \sum_{j=1}^{\tilde{N}_i} \sqrt{\lambda_{ij}^{(\alpha)}}\alpha_{ij}\phi_{ij}(z)$, $i = 1,\cdots,\tilde{s}$, $z \in D'$ with $\dim D = \dim D' + 1$. Note that $\lambda_{ij}^{(\theta)}$ and $\lambda_{ij}^{(\alpha)}$ usually decrease to 0 fast, the truncated K-L expansions, i.e., $a_{M_i} = \sum_{i=1}^{M_i} \sqrt{\lambda_{ij}^{(\theta)}}\theta_{ij}\psi_{ij}$ and $w_{\tilde{M}_i} = \sum_{j=1}^{\tilde{M}_i} \sqrt{\lambda_{ij}^{(\alpha)}}\alpha_{ij}\phi_{ij}$ can be used to reduce the dimension of the parameter space, which in turn would save CPU time while sampling from the posterior distribution. We denote

$$\begin{aligned} \theta &= (\theta_{11},\cdots,\theta_{1N_1},\cdots,\theta_{s1},\cdots,\theta_{sN_s}), \\ \alpha &= (\alpha_{11},\cdots,\alpha_{1\tilde{N}_1},\cdots,\alpha_{(s-1)1},\cdots,\alpha_{\tilde{s}\tilde{N}_{\tilde{s}}}), \end{aligned}$$

where $(\theta_{i1},\cdots,\theta_{iN_i})$ describe the permeability field $k(\theta,\alpha)$ within the i^{th} facies and $(\alpha_{j1},\cdots,\alpha_{j\tilde{N}_j})$ describe the j^{th} interface. θ_M and $\alpha_{\tilde{M}}$ are truncations of θ and α respectively. Then, the corresponding representations of the permeability field in full

and truncated case are given by

$$\begin{aligned} k(\theta,\alpha) &= \sum_{i=1}^{s}\exp\left(\sum_{j=1}^{N_i}\sqrt{\lambda_{ij}^{(\theta)}}\theta_{ij}\psi_{ij}\right)I_{\{D_i(\alpha_{i1},\cdots,\alpha_{i\tilde{N}_i})\}}, \\ k(\theta_M,\alpha_{\tilde{M}}) &= \sum_{i=1}^{s}\exp\left(\sum_{j=1}^{M_i}\sqrt{\lambda_{ij}^{(\theta)}}\theta_{ij}\psi_{ij}\right)I_{\{D_i(\alpha_{i1},\cdots,\alpha_{i\tilde{M}_i})\}}. \end{aligned}$$

Correspondingly, the two posterior distributions of the permeability field in Bayesian framework are given by

$$\begin{aligned} \pi(\theta,\alpha) &\propto G(\theta,\alpha)\prod_{i=1}^{s}\pi_0(\theta_{i1},\cdots,\theta_{iN_i})\prod_{j=1}^{\tilde{s}}\pi_0(\alpha_{j1},\cdots,\alpha_{j\tilde{N}_j}), \\ \tilde{\pi}(\theta,\alpha) &\propto \tilde{G}(\theta_M,\alpha_{\tilde{M}})\prod_{i=1}^{s}\pi_0(\theta_{i1},\cdots,\theta_{iN_i})\prod_{j=1}^{\tilde{s}}\pi_0(\alpha_{j1},\cdots,\alpha_{j\tilde{N}_j}), \end{aligned}$$

where

$$\begin{aligned} G(\theta,\alpha) &= \frac{1}{\sigma_f^2}\exp\Big(-\int_0^T|F_{obs}-F(k(\theta,\alpha);t)|^2dt\Big), \\ \tilde{G}(\theta_M,\alpha_{\tilde{M}}) &= \frac{1}{\sigma_f^2}\exp\Big(-\int_0^T|F_{obs}-F(k(\theta_M,\alpha_{\tilde{M}});t)|^2dt\Big), \end{aligned}$$

and F_{obs} is the observed fractional flow data. The priors $\pi_0(\theta,\alpha)$ is assumed to be product Gaussian measure.

It is clear that the truncation affects the matching process. Our goal here is to find an estimation of error introduced by this truncation, which also provides a way to choose M_i and $\tilde{M}_j$ for specified requirements.

Theorem 3.3.2. *Suppose the discretized K-L expansion of the log permeability field and the random velocity field are given by* $a_{N_i}=\sum_{j=1}^{N_i}\sqrt{\lambda_{ij}^{(\theta)}}\theta_{ij}\psi_i(x)$ *and* $w_{\tilde{N}_i}=\sum_{j=1}^{\tilde{N}_i}\sqrt{\lambda_{ij}^{(\alpha)}}\alpha_{ij}\phi_{ij}(z)$, *where all* a_{N_i} *and* $w_{\tilde{N}_i}$ *are stationary spatial processes on a bounded region, and the truncated expansions are* $a_{M_i}=\sum_{j=1}^{M_i}\sqrt{\lambda_{ij}^{(\theta)}}\theta_{ij}\psi_{ij}$ *and* $w_{\tilde{M}_i}=\sum_{j=1}^{\tilde{M}_i}\sqrt{\lambda_{ij}^{(\alpha)}}\alpha_{ij}\phi_{ij}$ *respectively. Assume that* $f(\theta,\alpha)$ *is a square integrable function*

with respect to a Gaussian measure, i.e., $\int |f(\theta,\alpha)|^2\pi_0(\theta,\alpha)d\theta d\alpha < \infty$*, then*

$$\begin{aligned} &\left|E_{\pi(\theta,\alpha)}[f(\theta,\alpha)] - E_{\tilde{\pi}(\theta,\alpha)}[f(\theta,\alpha)]\right| \\ \leq\ & C_1 \max_{1\leq i\leq s}\left\{\sum_{j=M_i+1}^{N_i}\lambda_{ij}^{(\theta)}\right\}^{\frac{1}{2}} + C_2 \max_{1\leq i\leq \tilde{s}}\left\{\sum_{j=\tilde{M}_i+1}^{\tilde{N}_i}\lambda_{ij}^{(\alpha)}\right\}^{\frac{1}{2}}, \end{aligned} \tag{3.7}$$

where C_1 *and* C_2 *are independent of dimension* N_i *and* $\tilde{N}_i$*.*

Proof. Note that

$$\left|E_{\pi(\theta,\alpha)}[f(\theta,\alpha)] - E_{\tilde{\pi}(\theta,\alpha)}[f(\theta,\alpha)]\right| \leq C(E_1+E_2),$$

where

$$\begin{aligned} E_1 &= \int |f(\theta,\alpha)|\cdot|\tilde{G}(\theta,\alpha_{\tilde{M}}) - \tilde{G}(\theta_M,\alpha_{\tilde{M}})|\pi_0(\theta,\alpha)d(\theta,\alpha), \\ E_2 &= \int |f(\theta,\alpha)|\cdot|G(\theta,\alpha) - \tilde{G}(\theta,\alpha_{\tilde{M}})|\pi_0(\theta,\alpha)d(\theta,\alpha). \end{aligned}$$

It is clear that Lemma 3.3.3 can be generalized to the multi-facies case to get

$$|\tilde{G}(\theta,\alpha_{\tilde{M}}) - \tilde{G}(\theta_M,\alpha_{\tilde{M}})| \leq \frac{C}{\sigma_f^2}\sum_{i=1}^{s}\|k(\theta_{i1},\cdots,\theta_{iN_i}) - k(\theta_{i1},\cdots,\theta_{iM_i})\|_{L_2(D_i(\alpha_{\tilde{M}}))}.$$

Then,

$$
\begin{aligned}
E_1 &\leq \frac{C}{\sigma_f^2}\int |f(\theta,\alpha)| \sum_{i=1}^{s} \|k(\theta_{i1},\cdots,\theta_{iN_i}) \\
&\quad -k(\theta_{i1},\cdots,\theta_{iM_i})\|_{L_2(D_i(\alpha_{\tilde{M}}))}\pi_0(\theta,\alpha)d(\theta,\alpha) \\
&\leq \frac{C}{\sigma_f^2}\int \sum_{i=1}^{s}\int |f(\theta,\alpha)|\cdot||k(\theta_{i1},\cdots,\theta_{iN_i}) \\
&\quad -k(\theta_{i1},\cdots,\theta_{iM_i})\|_{L_2(D_i(\alpha_{\tilde{M}}))}\pi_0(\theta)d\theta\pi_0(\alpha)d\alpha \\
&\leq \frac{C}{\sigma_f^2}\int \sum_{i=1}^{s}\Big\{\int |f(\theta,\alpha)|^2\pi_0(\theta)d\theta\Big\}^{\frac{1}{2}} \\
&\quad \cdot\Big\{\int \|k(\theta_{i1},\cdots,\theta_{iN_i})-k(\theta_{i1},\cdots,\theta_{iM_i})\|^2_{L_2(D_i(\alpha_{\tilde{M}}))}\pi_0(\theta)d\theta\Big\}^{\frac{1}{2}}\pi_0(\alpha)d\alpha \\
&\leq C\int \sum_{i=1}^{s}\Big\{\sum_{j=M_i+1}^{N_i}\lambda_{ij}^{(\theta)}\Big\}^{\frac{1}{2}}\pi_0(\alpha)d\alpha \leq C\max_{1\leq i\leq s}\left\{\sum_{j=M_i+1}^{N_i}\lambda_{ij}^{(\theta)}\right\}^{\frac{1}{2}},
\end{aligned}
$$

by Theorem 3.3.1. To estimate E_2, the estimation for permeability fields is also needed, i.e.,

$$
\begin{aligned}
\|k(\theta,\alpha)-k(\theta,\alpha_{\tilde{M}})\|^2_{L_2(D)} &= \int_D \left|\sum_{i=1}^{s}k_iI_{D_i(\alpha)}-\sum_{i=1}^{s}k_iI_{D_i(\alpha_{\tilde{M}})}\right|^2 dx \\
&\leq C\sum_{i=1}^{\tilde{s}}\int_{D'}k_i^2|w_{\tilde{N}_i}-w_{\tilde{M}_i}|^2dz \\
&\leq C\sum_{i=1}^{\tilde{s}}\int_{D'}k_i^2\left|\sum_{j=\tilde{M}_i+1}^{\tilde{N}_i}\sqrt{\lambda_{ij}^{(\alpha)}}\alpha_{ij}\phi_{ij}\right|^2 dz \\
&\leq C\sum_{i=1}^{\tilde{s}}\int_{D'}k_i^2\exp\left(2\sum_{j=\tilde{M}_i+1}^{\tilde{N}_i}\sqrt{\lambda_{ij}^{(\alpha)}}\alpha_{ij}\phi_{ij}\right)dz.
\end{aligned}
$$

Then we can get the estimation as

$$
\begin{aligned}
E_2 &\leq \frac{C}{\sigma_f^2}\Big\{\int |f(\theta,\alpha)|^2\pi_0(\theta,\alpha)d(\theta,\alpha)\Big\}^{\frac{1}{2}} \\
&\quad \cdot\Big\{\int ||k(\theta,\alpha)-k(\theta,\alpha_{_{\tilde{M}}})||^2_{L_2(D)}\pi_0(\theta,\alpha)d(\theta,\alpha)\Big\}^{\frac{1}{2}} \\
&\leq \frac{C}{\sigma_f^2}\Big\{\int \sum_{i=1}^{\tilde{s}}\int_{D'} k_i^2 \exp\left(2\sum_{j=\tilde{M}_i+1}^{\tilde{N}_i}\sqrt{\lambda_{ij}^{(\alpha)}}\alpha_{ij}\phi_{ij}\right)dz\pi_0(\theta,\alpha)d(\theta,\alpha)\Big\}^{\frac{1}{2}} \\
&\leq \frac{C}{\sigma_f^2}\Big\{\int_{D'}\Big[\sum_{i=1}^{\tilde{s}}\int k_i^2\pi_0(\theta_i)d\theta_i\int \exp\left(2\sum_{j=\tilde{M}_i+1}^{\tilde{N}_i}\sqrt{\lambda_{ij}^{(\alpha)}}\alpha_{ij}\phi_{ij}\right)\pi_0(\alpha_i)d\alpha_i\Big]dz\Big\}^{\frac{1}{2}} \\
&\leq \frac{C}{\sigma_f^2}\Big\{\sum_{i=1}^{\tilde{s}}\int_{D'}\exp\left(2\sum_{j=1}^{N}\lambda_{ij}^{(\theta)}\psi_{ij}^2\right)\exp\left(2\sum_{j=\tilde{M}_i+1}^{\tilde{N}_i}\lambda_{ij}^{(\alpha)}\phi_{ij}^2\right)dz\Big\}^{\frac{1}{2}} \\
&\leq C\max_{1\leq i\leq\tilde{s}}\left\{\sum_{j=\tilde{M}_i+1}^{\tilde{N}_i}\lambda_{ij}^{(\alpha)}\right\}^{\frac{1}{2}}.
\end{aligned}
$$

Since all a_{N_i} and $w_{\tilde{N}_i}$ are stationary spatial processes on a bounded region, i.e., for a spatial process where the covariance function depends only on the distance not on the spatial location, then [89] tells that the eigenfunctions $\{\psi_{ij}\}$ and $\{\phi_{ij}\}$ are uniform $L_\infty(D)$ bounded. Thus

$$
\begin{aligned}
&\big|E_{\pi(\theta,\alpha)}[f(\theta,\alpha)]-E_{\tilde{\pi}(\theta,\alpha)}[f(\theta,\alpha)]\big| \\
\leq\ & C_1\max_{1\leq i\leq s}\left\{\sum_{j=M_i+1}^{N_i}\lambda_{ij}^{(\theta)}\right\}^{\frac{1}{2}}+C_2\max_{1\leq i\leq\tilde{s}}\left\{\sum_{j=\tilde{M}_i+1}^{\tilde{N}_i}\lambda_{ij}^{(\alpha)}\right\}^{\frac{1}{2}}.
\end{aligned}
$$

Notice that when $M_i\to N_i$ and $\tilde{M}_i\to\tilde{N}_i$, we have

$$
\big|E_{\pi(\theta,\alpha)}(f(\theta,\alpha))-E_{\tilde{\pi}(\theta,\alpha)}(f(\theta,\alpha))\big|\to 0.
$$

□

The dimension of K-L expansion of permeability field and velocity field can be

large and make the computation of forward problem time-consuming. Theorem 3.3.2 shows the error bound of two posteriors when the truncations are introduced in K-L expansions. Based on this bound, the computation can be simplified by choosing less number of terms in truncation, while the error is in a reasonable range.

3.4. Sampling schemes

In this section we introduce the sampling scheme we use in the numerical examples in Section 3.5.2 for cases with channelized permeability field. The sampling scheme can be simplified for cases without channels correspondingly. To see the advantages of this scheme, we first state the standard Metropolis-Hastings algorithm and then point out the motivation of using our scheme.

For channelized permeability field, the standard Metropolis-Hastings algorithm can be formed in the following way to sample from the truncated posterior distribution $P(k|F_{obs})$.

Algorithm (Metropolis-Hastings MCMC [87]) Suppose at the n^{th} step, we have permeability field $k_n(\alpha_n, \theta_n)$.

Step 1. Generate α from a distribution $q_\alpha(\alpha|\alpha_n)$ and θ from a distribution $q_\theta(\theta|\theta_n)$. Then the entire permeability field $k(\alpha, \theta)$ is proposed using (3.1).

Step 2. Accept k as a sample with probability

$$\begin{aligned}\gamma(k_n, k) &= \min\left\{1, \frac{\pi(k)q(k_n|k)}{\pi(k_n)q(k|k_n)}\right\} \\ &= \min\left\{1, \frac{L(F_{obs}|k)}{L(F_{obs}|k_n)} \cdot \frac{\pi_0(\alpha)\pi_0(\theta)}{\pi_0(\alpha_n)\pi_0(\theta_n)} \cdot \frac{q_\alpha(\alpha_n|\alpha)q_\theta(\theta_n|\theta)}{q_\alpha(\alpha|\alpha_n)q_\theta(\theta|\theta_n)}\right\}, \qquad (3.8)\end{aligned}$$

i.e., take $k_{n+1} = k$ with probability $\gamma(k_n, k)$, and $k_{n+1} = k_n$ with probability $1 - \gamma(k_n, k)$. □

Starting with an initial permeability sample k_0, the MCMC algorithm generates

a Markov chain $\{k_n\}$ with the transition kernel as

$$\begin{aligned} K_r(k_n,k) &= \gamma(k_n,k)q(k|k_n) + \Big(1 - \int \gamma(k_n,k)q(k|k_n)dk\Big)\delta_{k_n}(k) \\ &= \gamma(k_n,k)q_\alpha(\alpha|\alpha_n)q_\theta(\theta|\theta_n) \\ &\quad + \Big(1 - \int \gamma(k_n,k)q_\alpha(\alpha|\alpha_n)q_\theta(\theta|\theta_n)d\alpha d\theta\Big)\delta_{\alpha_n}(\alpha)\delta_{\theta_n}(\theta). \end{aligned}$$

The target distribution $\pi(k)$ is the stationary distribution of the Markov chain k_n, so k_n represents the sample generated from $\pi(k)$ after the chain converges and reaches a steady state.

The main disadvantage of MCMC algorithm is the high computational cost in solving the coupled nonlinear PDE system (2.1)-(2.3) on the fine-grid in order to compute F_k in the target distribution $\pi(k)$. Typically, MCMC method in our simulations converges to the steady state after thousands of iterations and the acceptance rate is also very low. A large amount of CPU time is spent on simulating the rejected samples.

The MCMC method can be improved by adapting the proposal distribution $q(k|k_n)$ to the target distribution using a coarse-scale model. The process essentially modifies the proposal distribution $q(k|k_n)$ by incorporating the coarse-scale information. The algorithm for a general two-stage MCMC method with upscaling was introduced in [34].

Let F_k^* be the fractional flow computed by solving the coarse-scale model of (2.1)-(2.3) for the given k. This is done with mixed MsFEM [32]. Mixed MsFEM is used to solve pressure, and saturation is solved on coarse grid. The fine-scale target distribution $\pi(k)$ is approximated on the coarse scale by $\pi^*(k)$. Here, we have

$$\pi(k) \propto \exp\Big(-\frac{\|F_{obs} - F_k\|^2}{\sigma_f^2}\Big)\pi_0(k),$$

$$\pi^*(k) \propto \exp\Big(-\frac{(\mathfrak{G}(\|F_{obs} - F_k^*\|))^2}{\sigma_c^2}\Big)\pi_0(k),$$

where the function $\mathfrak{G}$ is estimated based on offline computations using independent samples from the prior. Using the coarse-scale distribution $\pi^*(k)$ as a filter, the two-stage MCMC can be described as follows.

Algorithm (Two-stage MCMC [34]) Suppose at the n^{th} step, we have α_n, θ_n and permeability field $k_n(\alpha_n, \theta_n)$.

Coarse stage: This step is the same as the MCMC method described earlier.

Step 1. At k_n, generate a trial proposal $\tilde{k}$ from distribution $q_\alpha(\alpha|\alpha_n)$ and $q_\theta(\theta|\theta_n)$. The only difference is the fractional flow F_k^* is computed by solving the coarse-scale model.

Step 2. Take the proposal as

$$k = \begin{cases} \tilde{k} & \text{with probability } \gamma_p(k_n, \tilde{k}), \\ k_n & \text{with probability } 1 - \gamma_p(k_n, \tilde{k}). \end{cases}$$

The acceptance probability is given by

$$\gamma_c(k_n, \tilde{k}) = \min\Big\{1, \frac{L^*(F_{obs}|\tilde{k})}{L^*(F_{obs}|k_n)} \cdot \frac{\pi_0(\alpha)\pi_0(\theta)}{\pi_0(\alpha_n)\pi_0(\theta_n)} \cdot \frac{q_\alpha(\alpha_n|\alpha)q_\theta(\theta_n|\theta)}{q_\alpha(\alpha|\alpha_n)q_\theta(\theta|\theta_n)}\Big\}. \tag{3.9}$$

Therefore, the final proposal k is generated from the effective instrumental distribution

$$Q(k|k_n) = \gamma_c(k_n, k)q(k|k_n) + \Big(1 - \int \gamma_c(k_n, k)q(k|k_n)dk\Big)\delta_{k_n}(k).$$

If $k = \tilde{k}$, go to the Step 3. Otherwise, i.e., $k = k_n$, return to Step 1.

Fine stage:

Step 3. Accept k as a sample with probability

$$\gamma_f(k_n, k) = \min\left(1, \frac{Q(k_n|k)\pi(k)}{Q(k|k_n)\pi(k_n)}\right), \tag{3.10}$$

i.e., $k_{n+1} = k$ with probability $\gamma_f(k_n, k)$, and $k_{n+1} = k_n$ with probability $1-\gamma_f(k_n, k)$. $\square$

Using the argument as in [34], the acceptance probability (3.10) can be simplified as

$$\gamma_f(k_n, k) = \min\left(1, \frac{\pi(k)\pi^*(k_n)}{\pi(k_n)\pi^*(k)}\right).$$

In our numerical example, we use a random walk to generate proposals for the posterior distribution, i.e., at the n^{th} step, we propose $\alpha = \alpha_n + h_\alpha u_\alpha$, where u_α is generated from a $N(0, I)$ distribution. Similarly, we propose $\theta = \theta_n + h_\theta u_\theta$, where u_θ is also generated from a $N(0, I)$ distribution. Here h_α and h_θ represent the step size of the jump in each step of the Metropolis-Hastings algorithm. The values of h_α and h_θ affect the convergence of the MCMC algorithm. The prior distribution of α can be taken to be $N(\alpha_o, \sigma_\alpha^2 I)$. Similarly, the prior distribution of θ can be taken to be $N(\theta_o, \sigma_\theta^2 I)$.

Also, we use a simple relation for modeling coarse- and fine-scale errors. In particular, $\mathfrak{G}$ is taken to be a linear function with the condition $\mathfrak{G}(0) = 0$. Then, $\pi^*(k)$ becomes

$$\pi^*(k) \propto \exp\Big(-\frac{\|F_{obs} - F_k^*\|^2}{\sigma_c^2}\Big)\pi_0(k),$$

i.e., on the coarse-scale $F_{obs}|k$ is assumed to follow $N(F_k^*, \sigma_c^2 I)$ distribution,

$$L^*(F_{obs}|k) \propto \exp\Big(-\frac{\|F_{obs} - F_k^*\|^2}{\sigma_c^2}\Big),$$

where σ_c is the precision associated with the coarse-scale model. The parameter σ_c plays an important role in improving the acceptance rate of the preconditioned MCMC method. The optimal value of σ_c depends on the correlation between $\|F - F_k\|$ and $\|F - F_k^*\|$, which can be estimated by offline computations. Assuming that on

the fine scale $F_{obs}|k$ follows a $N(F_k, \sigma_f^2 I)$ distribution, i.e.,

$$L(F_{obs}|k) \propto \exp\Big(-\frac{\|F_{obs} - F_k\|^2}{\sigma_f^2}\Big),$$

the acceptance probability (3.8) is given by

$$\gamma(k_n, k) = \min\left\{1, \frac{\exp\left(\frac{-\|F_{obs}-F_k\|^2}{\sigma_f^2}\right)}{\exp\left(\frac{-\|F_{obs}-F_{k_n}\|^2}{\sigma_f^2}\right)} \frac{\exp\left(\frac{-\|\theta-\theta_o\|^2}{2\sigma_\theta^2} + \frac{-\|\alpha-\alpha_o\|^2}{2\sigma_\alpha^2}\right)}{\exp\left(\frac{-\|\theta_n-\theta_o\|^2}{2\sigma_\theta^2} + \frac{-\|\alpha_n-\alpha_o\|^2}{2\sigma_\alpha^2}\right)}\right\}.$$

The acceptance probability (3.9) in two-stage MCMC algorithm is similar, i.e.,

$$\gamma(k_n, \tilde{k}) = \min\left\{1, \frac{\exp\left(\frac{-\|F_{obs}-F^*_{\tilde{k}}\|^2}{\sigma_f^2}\right)}{\exp\left(\frac{-\|F_{obs}-F^*_{k_n}\|^2}{\sigma_f^2}\right)} \frac{\exp\left(\frac{-\|\theta-\theta_o\|^2}{2\sigma_\theta^2} + \frac{-\|\alpha-\alpha_o\|^2}{2\sigma_\alpha^2}\right)}{\exp\left(\frac{-\|\theta_n-\theta_o\|^2}{2\sigma_\theta^2} + \frac{-\|\alpha_n-\alpha_o\|^2}{2\sigma_\alpha^2}\right)}\right\},$$

and the acceptance probability (3.11) becomes

$$\gamma_f(k_n, k) = \min\left(1, \frac{\exp\Big(-\frac{\|F_{obs}-F_k\|^2}{\sigma_f^2}\Big)\exp\Big(-\frac{\|F_{obs}-F^*_{k_n}\|^2}{\sigma_c^2}\Big)}{\exp\Big(-\frac{\|F_{obs}-F_{k_n}\|^2}{\sigma_f^2}\Big)\exp\Big(-\frac{\|F_{obs}-F^*_{k}\|^2}{\sigma_c^2}\Big)}\right).$$

3.5. Numerical results

3.5.1. Convergence estimation

The numerical results are presented here to show the validity of error estimation and the sampling method. We consider the expected value of the water-cut function (2.4) because it is one of the important properties for reservoirs. First, the tests are completed for single facies permeability field, and then channelized cases are considered.

3.5.1.1. Single facies

In the first simulation example a permeability field without any channelized structure is considered. To describe the permeability field, a two-point correlation function is defined as

$$R(x,y) = \sigma^2 \exp\Big(-\frac{|x_1 - y_1|^2}{2l_1^2} - \frac{|x_2 - y_2|^2}{2l_2^2}\Big). \tag{3.11}$$

K-L expansion is then used to describe the permeability field. $f(\theta)$ (the quantity of interest) is taken to be water-cut function F. One injector at $(0, 0.5)$ and one producer at $(1, 0.5)$ are considered when we run the forward model in the reference permeability field to get the fractional flow as discussed in Section 2.1 Eq. (2.4). The numerical estimation results are shown in Table 3.1.

We consider two sets of correlation lengths in our numerical examples. In the first example, we take $l_1 = 0.1$, $l_2 = 0.4$ and $\sigma^2 = 2$ with grid number 50×50 and θ's are taken from a log-normal distribution. Eigenvalues decrease rapidly as show in Figure 3.4, and we can truncate when it is below a threshold. The eigenfunctions corresponding to the leading eigenvalues in this case is shown in Figure 3.5.

In our numerical examples, we collect samples for $\pi(\theta)$. The MCMC process with random walk proposal step size 0.3 is used to get samples of $F(\theta)$ as a different number of K-L terms are taken into account. In this case, the proposal $\theta = \theta_n + 0.3 * \varepsilon$, where θ_n represents current value on MCMC chain and $\varepsilon \sim N(0, 1)$. The chain has 10000 iterations with the first 500 samples as burn-in period. The Monte Carlo integration retaining all the terms in the discrete K-L expansion is considered to be true value of $E_{\pi(\theta)}F(\theta)$. Samples with different number of truncated terms are taken to compute $E_{\tilde{\pi}(\theta)}F(\theta)$ in different cases to compare with the true one.

In our second example, we take the case with $l_1 = l_2 = 0.2$ and $\sigma^2 = 2$. Table

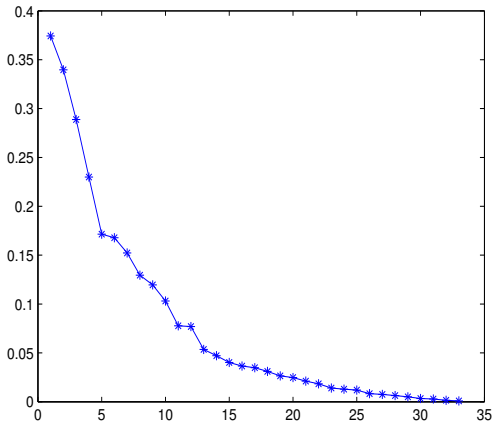

Fig. 3.4. Descending eigenvalues for $l_1 = 0.1$, $l_2 = 0.4$ and $\sigma^2 = 2$.

3.1 shows the results when the number of truncated terms M varies. In both cases, the errors decrease with the same convergence rate related to the sum of eigenvalue remainders of $R(x)$. This can be observed more clearly from Figure 3.6, while the data sets $\{(\sum_{i=M+1}^{N} \lambda_i^{(\theta)})^{\frac{1}{2}}, |E_\pi F - E_{\tilde{\pi}} F|\}$ can be fitted as a line. Namely, the relationship between $|E_\pi F - E_{\tilde{\pi}} F|$ and $(\sum_{i=M+1}^{N} \lambda_i^{(\theta)})^{\frac{1}{2}}$ is linear as shown in Theorem 3.3.1, while ignoring the errors in computing $F(\theta)$.

3.5.1.2. Channelized reservoirs

In our next example, we consider a permeability field with three facies. It is assumed that there is a high permeability layer in the middle and a low permeability layer in the two ends. The corresponding two interfaces are chosen randomly with the condition that the upper facies boundary is always above the lower facies boundary. The two different channels are populated using two log-Gaussian random fields from the truncated K-L expansions with two-point correlation function (3.11). The high permeable layer has correlation lengths $l_1 = 0.1$, $l_2 = 0.4$ and $\sigma = 1$, and the low

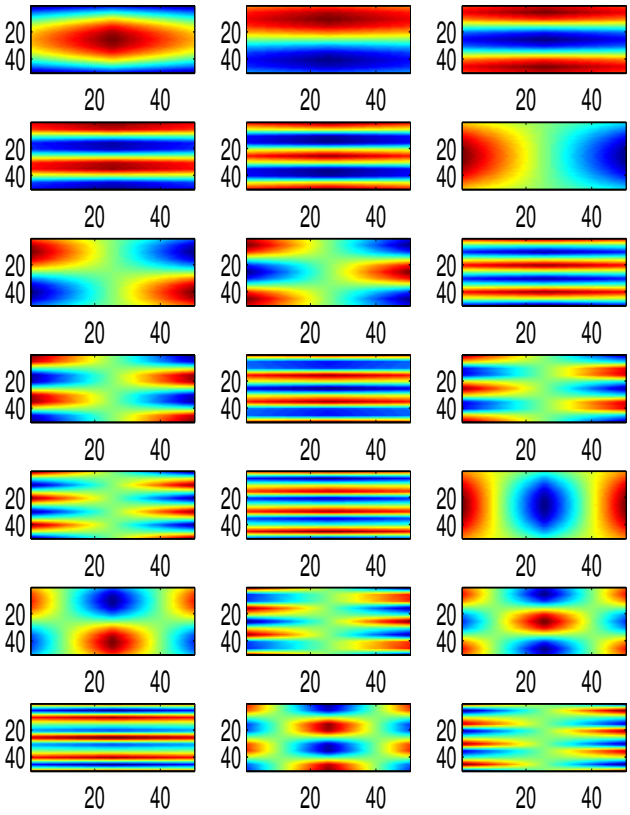

Fig. 3.5. Top 21 eigenvectors.

permeable layer has correlation lengths $l_1 = l_2 = 0.2$ and $\sigma = 0.4$. For both interfaces, a 1-d version of Eq. (3.11) is used with correlation length $l = 0.05$ and $\sigma = 1.5$.

We take a generated permeability field as reference, and run the forward model with one injector at $(0, 0.5)$ and a producer at $(1, 0.5)$ in this reference permeability field to get the fractional flow data F_{obs}. A MCMC chain for 10000 iterations is run to get the posterior of permeability field, with the first 500 samples as burn-in period.

As the number of terms in the truncations of K-L expansions varies, the estima-

Table 3.1. Posterior errors $|E_\pi F - E_{\tilde{\pi}} F|$ when the K-L expansion is truncated to M terms. Left: $l_1 = 0.1$, $l_2 = 0.4$, $\sigma^2 = 2$ and $\sigma_f^2 = 0.001$; Right: $l_1 = 0.2$, $l_2 = 0.2$, $\sigma^2 = 2$ and $\sigma_f^2 = 0.005$.

M	$(\sum_{i=M+1}^{N} \lambda_i^{(\theta)})^{\frac{1}{2}}$	$\lvert E_\pi F - E_{\tilde{\pi}} F\rvert$	M	$(\sum_{i=M+1}^{N} \lambda_i^{(\theta)})^{\frac{1}{2}}$	$\lvert E_\pi F - E_{\tilde{\pi}} F\rvert$
5	1.111681	0.081809	5	1.176697	0.308118
10	0.750662	0.106264	10	0.820661	0.191601
15	0.517555	0.063635	15	0.566938	0.119590
20	0.337901	0.030207	20	0.378454	0.059173
25	0.189272	0.017931	25	0.248267	0.033023
30	0.071924	0.011225	30	0.123347	0.014965

tions of posterior errors, similar to Table 3.1, are reported in Table 3.2. The Monte Carlo integration retaining all the terms in the discrete K-L expansions is considered to be true value. In Table 3.2, we can see that the error between the true value and the estimated value from the truncated posterior decreases consistently as we increase the number of the terms retained in K-L expansion. If we further plot the errors, we can find that the errors lie on a plane (see Figure 3.7) as indicated in Theorem 3.3.2 in Section 3.3.2.

3.5.2. Matching permeability with reduced parameters

In this example we will show that the reference permeability field can be recovered from matching the observations and the quality of matches is certainly affected by the truncation of expansions. There is a high permeable layer in the middle and low permeable layers in the two ends with the same correlation lengths as in Section 3.5.1.2. The interfaces are taken as a linear interpolation of independent points.

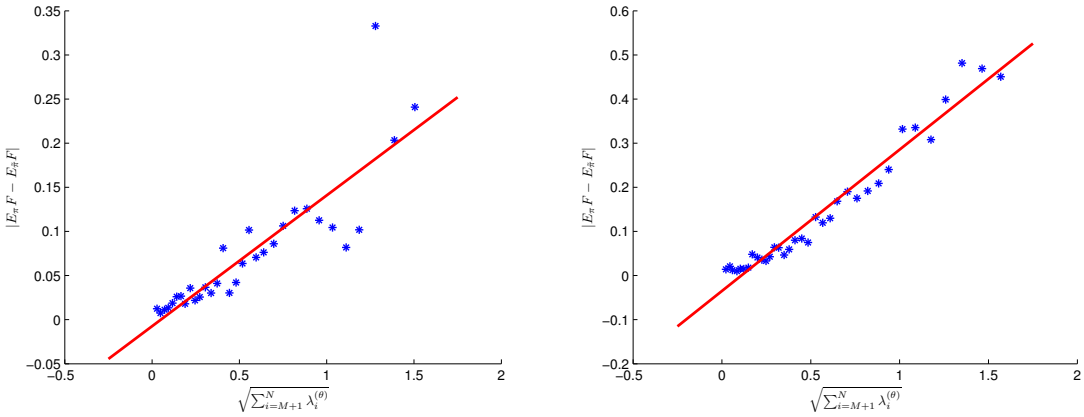

Fig. 3.6. Linear fit of $\{(\sum_{i=M+1}^{N}\lambda_i^{(\theta)})^{\frac{1}{2}}, |E_\pi F - E_{\tilde{\pi}}F|\}$. Left: $l_1 = 0.1$, $l_2 = 0.4$, $\sigma^2 = 2$ and $\sigma_f^2 = 0.001$; Right: $l_1 = 0.2$, $l_2 = 0.2$, $\sigma^2 = 2$ and $\sigma_f^2 = 0.005$.

In the first part, we truncate the K-L expansion and retain only the first 20 terms. We consider 25 points on the facies. So the dimension of θ is 40 and the dimension of α is 25. The two-stage MCMC method is used to sample from the posterior. The initial facies boundaries are taken to be straight lines joining the two ends of the known facies boundaries. We use random walk to perturb θ and α with the step size 0.25 and 0.05, respectively, and with independent Gaussian priors for θ and α. We run the MCMC chain for 10000 iterations and leave out the first 500 samples as burn-in period.

In Figure 3.8, the reference permeability field, the initial permeability field and the mean of the posterior permeability field are shown. We can see that the sample mean is very close to the reference field. On the left plot of Figure 3.9 we can see that the sample estimate of the fractional flow is very close to the observed data. From the right plot of Figure 3.9 we can see that combined error decreases nearly to zero and stays there, so the Markov Chain has converged. The two-stage MCMC has higher acceptance rate [34] (four times in these calculations), because it rejects the bad proposal fast in the first stage, which is inexpensive. Next, we repeat the same

Table 3.2. Posterior errors $|E_\pi F - E_{\tilde{\pi}} F|$ when the K-L expansion is truncated to M terms for different facies.

M_1	M_2	$\tilde{M}$	$(\sum_{j=M_1+1}^{N_1} \lambda_{1j}^{(\theta)})^{\frac{1}{2}}$	$(\sum_{j=M_2+1}^{N_2} \lambda_{2j}^{(\theta)})^{\frac{1}{2}}$	$(\sum_{i=\tilde{M}+1}^{\tilde{N}} \lambda_i^{(\alpha)})^{\frac{1}{2}}$	$\lvert E_\pi F - E_{\tilde{\pi}} F \rvert$
5	5	5	0.526235	0.786077	0.853727	0.109464
10	5	5	0.367011	0.786077	0.853727	0.116172
10	10	10	0.367011	0.530798	0.477141	0.051925
15	10	10	0.253542	0.530798	0.477141	0.093109
15	15	10	0.253542	0.365967	0.477141	0.053869
20	15	15	0.169250	0.365967	0.210844	0.047356
20	20	15	0.169250	0.238932	0.210844	0.019996

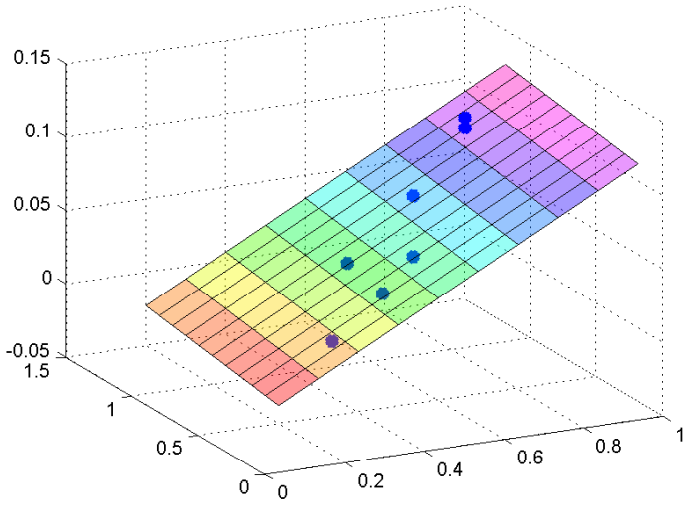

Fig. 3.7. Plots of $(\max\{(\sum_{j=M_1+1}^{N_1} \lambda_{1j}^{(\theta)})^{\frac{1}{2}}, (\sum_{j=M_2+1}^{N_2} \lambda_{2j}^{(\theta)})^{\frac{1}{2}}\}, (\sum_{i=\tilde{M}+1}^{\tilde{N}} \lambda_i^{(\alpha)})^{\frac{1}{2}}, |E_\pi F - E_{\tilde{\pi}} F|)$.

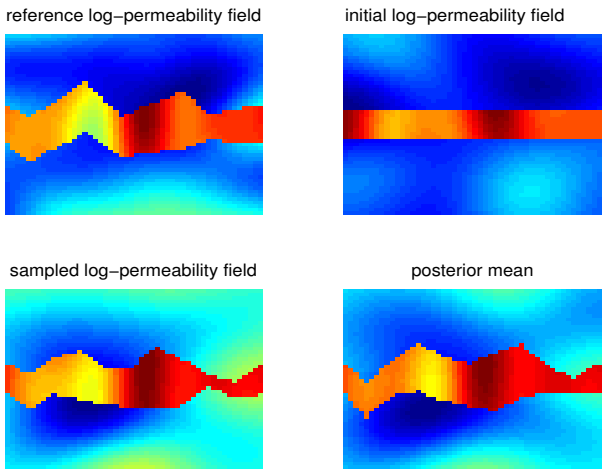

Fig. 3.8. Top left: The true log-permeability field. Top right: Initial log-permeability field. Bottom left: One of the sampled log-permeability fields. Bottom Right: The mean of the sampled log-permeability fields from two-stage MCMC using 20 K-L terms.

procedure of sampling the posterior but we retain 25 terms in the K-L expansion in this case. We use the same reference permeability field and the fractional flow data. The numerical results are shown in Figures 3.10 and 3.11. We can see the sampled mean of the permeability field is more accurate than the previous example with 20 K-L coefficients.

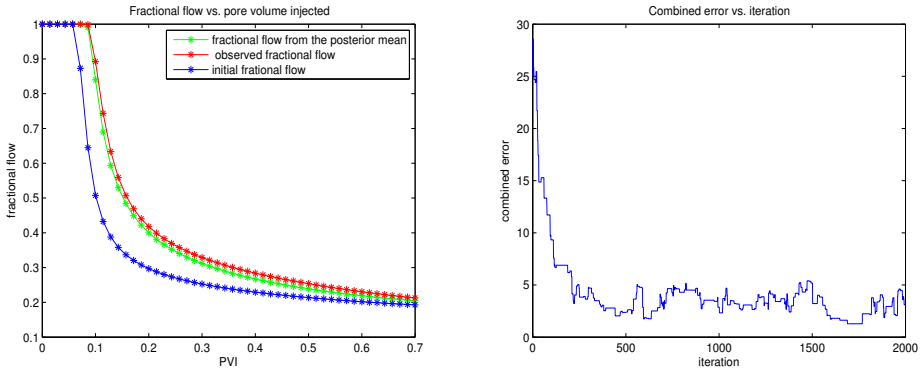

Fig. 3.9. Left: Red line designates the fine-scale reference fractional flow, the blue line designates the initial fractional flow and the green line designates fractional flow corresponding to mean of the sampled permeability field from two-stage MCMC. Right: Fractional flow errors vs. accepted iterations when sampled from the posterior distribution retaining 20 terms in K-L expansion.

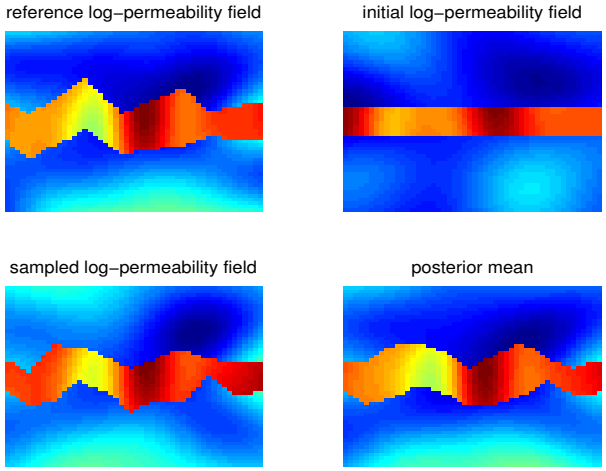

Fig. 3.10. Top left: The true log-permeability field. Top right: Initial log-permeability field. Bottom left: One of the sampled log-permeability fields. Bottom Right: The mean of the sampled log-permeability fields from two-stage MCMC using 25 K-L terms.

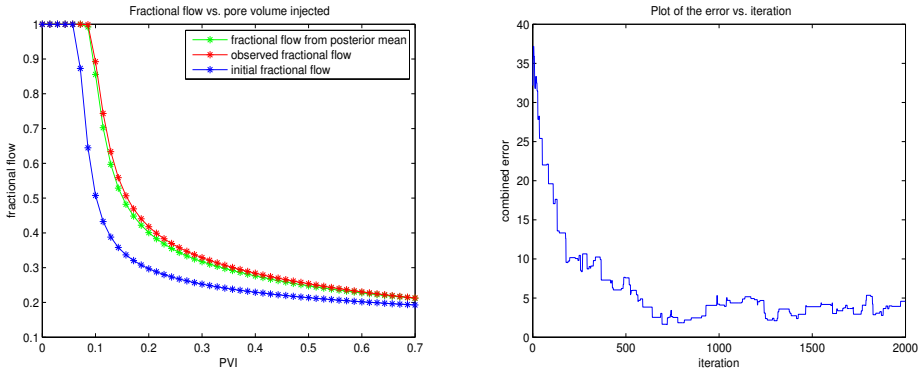

Fig. 3.11. Left: Red line designates the fine-scale reference fractional flow, the blue line designates the initial fractional flow and the green line designates fractional flow corresponding to mean of the sampled permeability field from two-stage MCMC. Right: Fractional flow errors vs. accepted iterations when sampled from the posterior distribution retaining 25 terms in K-L expansion.

CHAPTER IV

ENSEMBLE-LEVEL MULTISCALE FINITE ELEMENT AND PRECONDITIONER FOR CHANNELIZED SYSTEMS AND APPLICATIONS

When we consider multiscale flow in porous media, it is assumed that we can characterize the ensemble of all possible flow scenarios, that is, we can describe all possible permeability configurations needed for the simulations. We construct coarse basis functions that can provide inexpensive coarse approximations that are: (1) adequate for all possible flow scenarios in the given ensemble; (2) robust with respect to the small scales and high variations in each flow scenario. The coarse approximations developed here can be used as a multiscale finite element, or as the coarse solver in a two-level domain decomposition iterative method. The methods presented here extend, to the ensemble case, some of the results in [27] and [40]. Specifically, ensembles of permeability fields with high contrast channels and inclusions are considered. Our objective here is to construct special multiscale basis functions for the whole ensemble of flow scenarios. The coarse basis functions are pre-computed for permeability fields with certain topological properties. This is a preprocessing step. This procedure avoids constructing basis functions or computing upscaled parameters for each permeability realization. Then, for any permeability, the solution of elliptic equation can be projected to the space spanned by these pre-computed basis functions. We apply this coarse multiscale approximation to the design of two-level domain decomposition preconditioner. Numerical experiments show that the ensemble-level multiscale finite element method provides a good approximation to the fine-scale solution. Numerical experiments also show that the ensemble-level domain decomposition preconditioner condition number is independent of the high contrast in the coefficient.

This chapter is organized in the following way: in the first two sections, we present the preliminary and motivation of the ensemble-level methods. Then, we first present the methods we use to complete coarse space and choose a preconditioner. Based on these discussions, we provide more details of the ensemble-level cases for these methods. In the following section, we present numerical results of the ensemble-level methods by comparing the solutions with fine-scale solution. We also explore the behavior of the proposed method in sampling processes.

4.1. Introduction

Many problems in applications occur in media that contains multiple scales and has a high contrast in the properties. For example, the second order elliptic equation with heterogeneous coefficients

$$-\text{div}(k(x)\nabla u) = f \quad \text{in } D \tag{4.1}$$

subject to some boundary conditions is used to model fractured reservoirs. Typically, as we mentioned, the permeability field has large variations. We use $k(x)$ to denote a special set of permeability fields we are interested in. These large variations of k bring additional small-scale parameters into the problem. Also the unknown k can bring other difficulties in modeling and computing the flow. In these cases, the system (4.1) and the numerical method applied need special treatments [23, 24, 86]. In our case, we will only study the possible flow scenarios modeled by (4.1).

Roughly speaking, when only few flow scenarios are considered, we can divide the approaches proposed to solve such systems in two groups, depending if they target to compute a coarse-grid approximation, or a fine-grid (as fine as resolution of $k(x)$) approximation.

For approaches that involve the solution on a coarse grid we refer to [1, 3, 4, 5, 6, 7, 10, 12, 17, 24, 25, 27, 31, 32, 33, 28, 44, 45, 48, 49, 62, 76, 77, 78, 83, 85, 93, 103]. Coarse-grid properties, such as upscaled conductivities or multiscale basis functions, are constructed that represent the media or the solution on the coarse grid in these approaches. We note that, a naive extension of these methods for the case of many possible flow scenarios would require the computations of basis functions or upscaling parameters for each possible scenario in the ensemble. In general, computations of optimal basis functions is expensive in many practical cases. In this chapter, we will use an ensemble-level approach, meaning that we will construct and save coarse basis functions in a *pre-processing step*. This pre-processing involves the identifications of one or few permeability field(s) that characterize the high-contrast structure of the whole ensemble. Then, we use this especial field and apply ideas from [26, 27, 40] to construct the basis functions. The resulting spaces can be used in computations of any flow scenario in the ensemble.

From the approaches that compute fine grid solutions, we only mention iterative methods of domain decomposition type. In general, numerical discretization of flow problems in such heterogeneous media results to very large ill-conditioned systems of linear equations. Therefore, robust iterative methods that converge independent of the contrast and multiple scales are needed (e.g., [9, 19, 41, 71, 84, 96] and references therein). A domain decomposition method uses local subdomain solvers and a coarse global solver to iterate until some convergence criteria is reached. In order to obtain robust domain decomposition, the coarse solver needs to be selected carefully. Using an inadequate coarse solver may lead to contrast-dependent number of iterations required for convergence. In this paper we construct a two-level domain decomposition for a whole ensemble of flow scenarios. The idea is again, to use a coarse solver that uses multiscale basis functions constructed (as in [26, 39, 40]) for special permeability

fields that characterize the high-contrast structure of the ensemble of flow scenarios.

We note that, in many applications, we have some prior information, for example the existence and approximate positions of channels. In these cases, it is possible to characterize the interesting flow scenarios and permeability fields, denoted here by $\{k_m\}_{m\in I}$, where I is an index set. According to the prior information, the permeability fields $\{k_m\}_{m\in I}$ can share some topological similarities. We propose ensemble-level methods in this paper to take advantage of these topological similarities to save computational time on constructing multiscale basis functions and domain decomposition preconditioner.

The idea is to construct basis functions by selecting a few realizations of permeability fields $\{k_m\}_{m\in I}$. Let us assume that N permeability fields $\{k_m\}_{m\in\{i_1,\cdots,i_N\}}$ are selected. In practice, N is small and will correspond to the number of topologically equivalence (as defined later) classes of permeability fields in the ensemble. After a coarse grid of the domain is introduced, the multiscale basis functions can be constructed in each coarse grid block for all scenario in the subset $\{k_m\}_{m\in\{i_1,\cdots,i_N\}}$. Then, these pre-computed basis functions span a finite dimensional space. Once this special space is constructed and saved, for any permeability field coming from $\{k_m\}_{m\in I}$, we can project the solution onto this space. By pre-computing the basis functions on a selected set rather than on the whole index set, the computational expense can be reduced, especially when the size of I is large. The same idea is used in the constructing of a two-level domain decomposition preconditioner to guarantee the robustness of the system.

Our ensemble-level multiscale finite element methods (MsFEMs) for solving the problem on a coarse grid, as well as, in two-level preconditioners, are based on the methods designed in [27, 39, 40]. The construction of coarse spaces starts with an initial choice of multiscale basis functions that are supported in coarse regions sharing a

common node. These basis functions are complemented using weighted local spectral problems defined in coarse blocks sharing a common node. The initial choice of multiscale basis functions is important for the computation of the weight function in the local spectral problem. Furthermore, we identify small eigenvalues and corresponding eigenvectors that represent important features of the solution. Coarse spaces are constructed by multiplying the selected eigenvectors with the initial multiscale basis functions. The estimate convergence of MsFEM and condition number of two-level preconditioners depends on the maximum of the inverse of the eigenvalues whose eigenvectors are not included in the coarse space. The maximum is taken over all coarse nodes. Details will be provided in Section 4.3.1. It is known that the number of iterations required by iterative methods, such as domain decomposition methods, is affected by the contrast in the media properties that are within each coarse-grid block [41]. We choose the initial multiscale space such that the eigenvalues of the local spectral problems increase rapidly.

4.2. Motivation of the ensemble-level idea

The distribution of subsurface properties is mainly controlled by the locations of distinct geologic facies with sharp contrasts in properties, such as permeability and porosity, across facies boundaries [100]. Channelized permeability fields are high contrast media with certain topological similarity. We use multiscale methods to compute fluid behavior. The computational time in this case is mostly spent on constructing multiscale basis functions for each permeability field. In order to use the topological similarity of the permeabilities in the ensemble and reduce computational expense, we come to the idea of ensemble-level methods, which only construct multiscale basis functions based on a subset of permeability fields. As the basis functions

are pre-computed, for the other permeability field, the solution can be projected directly to the space spanned by these functions, instead of building their own basis functions. Therefore, the computational time could be saved. In our discussion below, we assume that some information about the channelized permeability ensemble is known, for example, the number of channels and the existence of topological equivalent among corresponding channels. Especially, we propose to use the permeability field with all the members of the ensemble as its inclusion to construct multiscale basis functions.

These special basis functions can be typically constructed by solving the local flow equation on a coarse grid subject to some boundary conditions. To design robust iterative methods for the solution of the fine-scale problem, one can use coarse-grid multiscale solutions and additional local subdomain corrections (as in domain decomposition methods) for faster convergence. Domain decomposition methods use the solutions of local problems and a coarse problem in constructing preconditioners for the fine-scale system. The number of iterations required by domain decomposition preconditioners is adversely affected by the contrast in the media properties [41]. Because of the complex geometry of fine-scale features (e.g., complex fracture geometry), it is very often *impossible* to separate low and high conductivity regions into different coarse grid blocks. Thus, it may require many iterations for iterative methods to converge. It is important to build a preconditioner, which gives condition number independent of high contrast. Our ensemble approach can give an effective preconditioner based only on pre-selected permeability fields.

As our ensemble-level methods decrease computational expense in the modeling of flow and transport, the dynamic data integration processes can then take advantages from these methods. The problem of finding corresponding parameter fields given observation data has been explored. Bayesian approach is usually used in find-

ing the distribution of the parameter fields. This often requires large number of forward model runs to get convergence. As the computation ability is limited, a method requires less computational time in each run is preferred. In this case, more samples can be collected and the conclusion based on large number of samples is more reliable. Methods have been developed [34, 68, 69, 70, 64] to build approximation of the fine-scale solution, while the approximation requires less computational expense. Our ensemble-level methods can serve the same purpose in a different way for the sampling processes.

4.3. Ensemble-level MsFEM and ensemble-level preconditioner

4.3.1. MsFEM and preconditioner

In this section we review the constructions of coarse scale approximation and preconditioners for a given permeability field $k(x)$. In Section 4.3.2 we will apply these procedures to a special permeability field that characterizes a whole ensemble.

We consider the second order elliptic equation with heterogeneous coefficients

$$-\mathrm{div}(k(x)\nabla u) = f \quad \text{in } D \tag{4.2}$$

subject to some boundary conditions, for example, linear or Neumann boundary conditions (see [32]). $k(x)$ is a heterogeneous spatial permeability field with multiple scales and high contrast, and D is a union of disjoint polygonal subdomains $\{D_i\}_{i=1}^N$. The notation introduced in this section will be used later on.

Let $\mathcal{T}^H$ be a usual conforming partition of D into finite elements (triangles, quadrilaterals and etc.). We call this partition the coarse grid and assume that the coarse grid can be resolved via a finer grid $\mathcal{T}^h$, which is a refinement of $\mathcal{T}^H$. The fine grid is fine enough to describe the coefficient k, but too fine to actually do practical

computations. And the coarse grid is sufficiently coarse to do practical computations, but too coarse to accurately describe the variations of k.

Our objective is to seek multiscale basis functions. Using the coarse triangulation $\mathcal{T}^H$, we introduce basis functions $\{\Phi_i\}_{i=1}^{N_c}$, where N_c is the number of coarse basis functions. Let $\{y_i\}_{i=1}^{N_v}$ be the vertices of the coarse mesh $\mathcal{T}^H$, N_v be the number of coarse nodes, and define the neighborhood of the node y_i by

$$\varpi_i = \bigcup \{K_j \in \mathcal{T}^H; \quad y_i \in \overline{K}_j\}$$

(see Figure 4.1) and the neighborhood of the coarse element K by

$$\varpi_K = \bigcup \{\varpi_j \in \mathcal{T}^H; \quad y_j \in \overline{K}\}.$$

To capture the fine scale features of the solution, the basis functions can be constructed in different ways. For example, the nodal basis of the standard finite element space, denoted as χ_i^0, can be used. The other choice is to find multiscale finite element basis functions χ_i^{ms}, that coincide with χ_i^0 on the boundaries of the coarse partition and satisfy

$$\text{div}(k\nabla\chi_i^{ms}) = 0 \ \text{ in } K \in \varpi_i, \quad \chi_i^{ms} = \chi_i^0 \ \text{ in } \partial K, \ \ \forall\ K \in \varpi_i,$$

where K is a coarse grid block within ϖ_i (see [32] for more details). Also, we can take energy minimizing basis functions χ_i^{emf} (see [106]), where basis functions are obtained by minimizing the energy of the basis functions subject to a global constraint. More precisely, one can use the partition of unity functions $\{\chi_i^{emf}\}_{i=1}^{N_v}$, with N_v being the number of coarse nodes, that provide the least energy. This can be accomplished by solving

$$\min \sum_i \int_{\varpi_i} k|\nabla\chi_i^{emf}|^2,$$

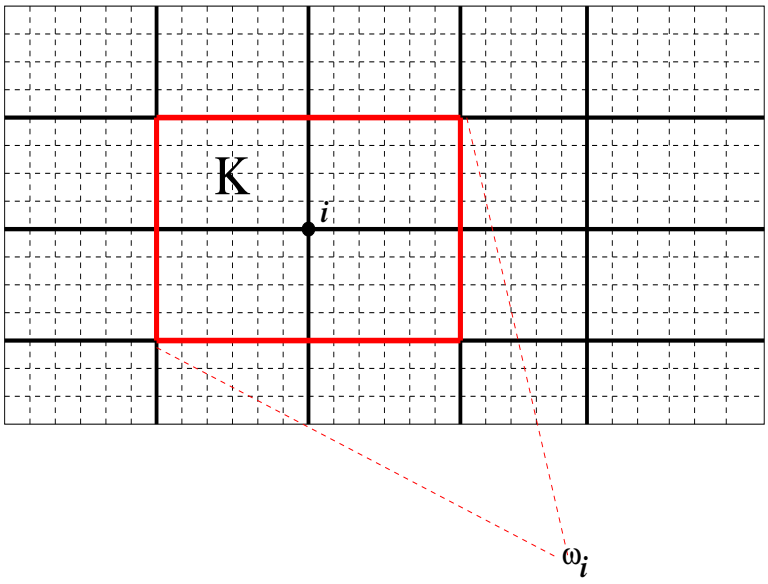

Fig. 4.1. Schematic description of coarse regions.

subject to $\sum_i \chi_i^{emf} = 1$ with $\text{Supp}(\chi_i) \subset \varpi_i$, $i = 1, \cdots, N_v$. Methods using limited global information to construct coarse basis can be found in [31, 83, 106]. Over-sampling techniques [32, 45, 49] are also proposed, since the initial choice of basis function is crucial for determining the dimension of the coarse space needed to obtain an accurate coarse-scale approximation and robust preconditioners.

These coarse spaces often need to be complemented if more accurate coarse-scale solutions or more robust preconditioners are sought. For this reason, we seek additional basis functions that improve the accuracy of the approximation. We will consider complementing the coarse spaces described above by finding appropriate local fields in ϖ_i and by multiplying them with our multiscale functions. We start with the coarse space generated by one basis function per node χ_i^0, χ_i^{ms}, χ_i^{emf} or so on. And further we complement this space by adding basis functions in each ϖ_i.

Consider the eigenvalue problem

$$-\text{div}(k\nabla\psi_l^{\varpi_i}) = \lambda_l^{\varpi_i}\widetilde{k}\psi_l^{\varpi_i},$$

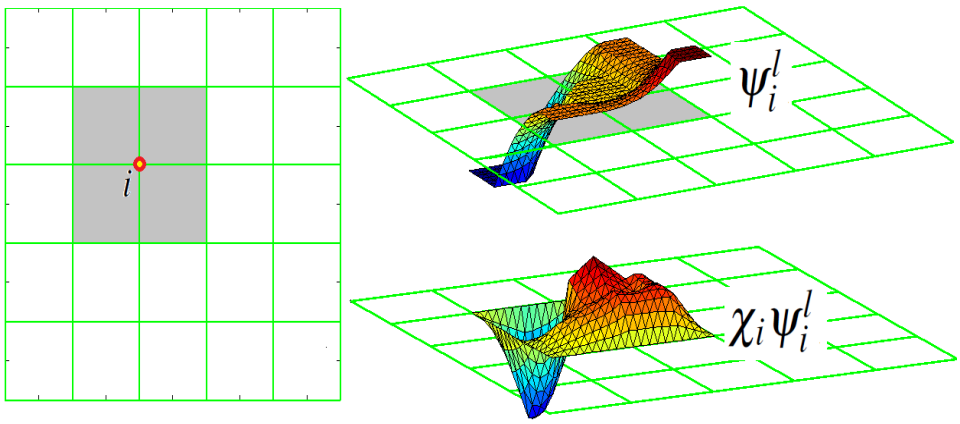

Fig. 4.2. Schematic description of basis function construction. Left: subdomain ϖ_i. Right-Top: Selected eigenvector ψ_i^ℓ with small eigenvalue. Right-Bottom: product $\chi_i\psi_i^\ell$ where χ_i is the initial basis function of node i.

where $\lambda_l^{\varpi_i}$ and $\psi_l^{\varpi_i}$ are eigenvalues and eigenvectors in ϖ_i and $\widetilde{k}$ is defined by

$$\widetilde{k} = \frac{1}{H^2} k \sum_{j=1}^{N_v} |\nabla \chi_j|^2. \tag{4.3}$$

We recall that χ_i are initial multiscale basis functions, e.g., $\chi_i = \chi_i^0$ or $\chi_i = \chi_i^{ms}$ or $\chi_i = \chi_i^{emf}$, and N_v is the number of the coarse nodes. One can choose other multiscale basis functions. Then, multiscale basis functions will be constructed as $\chi_i\psi_i^l$ (see Figure 4.2 for illustration), and the space will be span$\{\chi_i\psi_i^l\}$. Furthermore, the coarse-scale solution is sought based on (4.5) with the coarse spaces defined by (see [27]),

$$V_0^k = \text{span}\{\Phi_{i,\ell}^k = \chi_i\psi_i^l\}. \tag{4.4}$$

We seek $u_0 = \sum_i c_i\Phi_i^k$, where c_i are determined from

$$a^k(u_0, v) = f(v), \quad \text{for all } v \in V_0^k, \tag{4.5}$$

where

$$a^k(u,v) = \int_D k(x)\nabla u(x)\nabla v(x)dx \quad \text{for all } u, v \in H_0^1(D) \tag{4.6}$$

and

$$f(v) = \int_D f(x)v(x)dx \quad \text{for all } v \in H_0^1(D).$$

The coarse problem (4.5) is equivalent to a coarse linear system. Let the coarse matrix be given by $A_0^k = R_0^k A^k (R_0^k)^T$, where A^k is the fine-scale finite element matrix representation of the bi-linear form a^k in (4.6) and

$$(R_0^k)^T = [\Phi_1^k, \ldots, \Phi_{N_c}^k].$$

Here Φ_i's are discrete coarse-scale basis functions defined on a fine grid (i.e., vectors). Multiscale solution u_0 is given by

$$A_0^k u_0 = f_0,$$

where $f_0 = (R_0^k)^T b$.

A lot of iterations are needed to solve this linear system, when the permeability varies largely over spatial domain. A preconditioner can help in building a robust system. The coarse basis function R_0^k can be used in designing two-level domain decomposition preconditioners as well. We denote by $\{D_i'\}_{i=1}^N$ the overlapping decomposition obtained from the original non-overlapping decomposition $\{D_i\}_{i=1}^N$ by enlarging each subdomain D_i to

$$D_i' = D_i \cup \{x \in D, \text{dist}(x, D_i) < \delta_i\}, \quad i = 1, \ldots, N,$$

where dist is some distance function and let $V_0^i(D_i')$ be the set of finite element functions with support in D_i'. We also denote by $R_i^T : V_0^i(D_i') \to V^h$ the extension by zero operator.

We can solve the fine-scale linear system iteratively with the preconditioned conjugate gradient (PCG) method. Any other suitable iterative scheme can be used as well. We use the two-level additive preconditioner of the form

$$(B^k)^{-1} = (R_0^k)^T (A_0^k)^{-1} R_0^k + \sum_{i=1}^{N} R_i^T (A_i^k)^{-1} R_i,$$

where the local matrices are defined by

$$v A_i^k w = a^k(v, w) \quad \text{for all } v, w \in V_0^i(D_i'), \tag{4.7}$$

$i = 1, \cdots, N$. See [71, 96] and references therein. The application of the preconditioner involves solving a coarse-scale system and solving local problems in each iteration. In domain decomposition methods, our main goal is to reduce the number of iterations in the iterative procedure (see [26, 39, 40] for more details).

4.3.2. Ensemble-level MsFEM and preconditioner

Next we consider the ensemble-level construction of MsFEM basis for multiscale approximation and preconditioning. Let $\{k_m\}_{m \in I}$ be given as a set of channelized permeability fields, where I is an index set. We assume that every permeability field k_m, consists of N^{ch} high-conductivity channels and N^{in} high-conductivity inclusions. For simplicity we assume that the high conductivity channels and inclusions are surrounded by a background of conductivity one. To explain our assumptions more effectively, we denote by $C_s(k_m)$ the $s-$th channel of k_m, $s = 1, \cdots, N^{ch}$. Analogously, $I_s(k_m)$ denotes the $s-$th inclusion of k_m, $s = 1, \cdots, N^{in}$.

Here we assume that all permeability fields in the ensemble $\{k_m\}_{m \in I}$ are topologically similar in the sense that:

1. Every k_m has the same number of channels N^{ch} and inclusions N^{in}, and

2. Given any two permeabilities k_m and $k_{m'}$, each channel $C_i(k_m)$ of k_m is topologically equivalent to the channel $C_i(k_{m'})$ presenting in $k_{m'}$.

Furthermore, for simplicity of the presentation, we focus on the case of high contrast binary media, i.e., we are assuming that the background has conductivity value 1 and the channels and inclusions have high permeability value η. See Figure 4.3 for illustration.

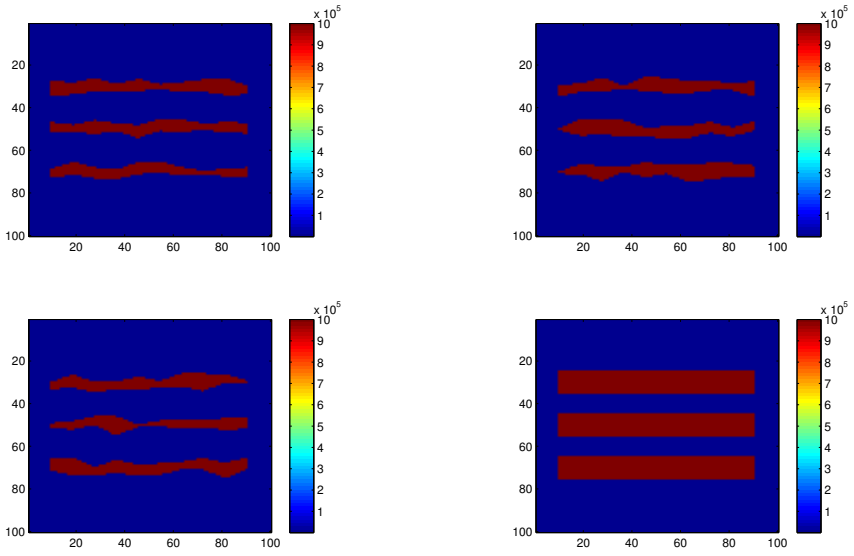

Fig. 4.3. Topologically similar high contrast binary permeability fields with $\eta = 10^6$. The right bottom one serves as k^W.

For the set $\{k_m\}_{m\in I}$ described here, let

$$k^W = \max_m k_m. \tag{4.8}$$

We assume that k^W contains N^{ch} high-conductivity channels and N^{in} high-conductivity inclusions and that k^W 's geometrical configuration is topologically equivalent to any permeability k_m. We observe that the channels and inclusions of k^W are wider/longer than the corresponding channels and inclusions of any permeability k_m in the ensem-

ble.

Here we propose to use the coarse space constructed for k^W to obtain multiscale basis functions for the whole ensemble of fields $\{k_m\}_{m\in I}$. More precisely, we consider the problem of constructing coarse approximations for the second order elliptic equations

$$-\mathrm{div}(k_m(x)\nabla u) = f \ \text{ in } D, \tag{4.9}$$

where $m \in I$. We construct one coarse space (depending only on the coarse partition and k^W) that is suitable for the coarse scale approximation of these equations.

We use the notation introduced in the previous Section 4.3.1. Denote the multiscale basis constructed using the field k^W by $\{\Phi_i^{k^W}\}_{i=1}^{N_c}$. Then, the coarse space we are going to use for the whole ensemble is

$$V_0^{k^W} = \mathrm{span}\{\Phi_i^{k^W}\}_{i=1}^{N_c}. \tag{4.10}$$

For any k_m, Eq. (4.9) can be posed in $V_0^{k^W}$ as: find $u_0 \in V_0^{k^W}$ such that

$$a^{k_m}(u_0, v) = f(v), \quad \text{for all } v \in V_0^{k^W} \tag{4.11}$$

where the bilinear form a^{k_m} is defined as in (4.6) by

$$a^{k_m}(u, v) = \int_D k_m(x)\nabla u(x)\nabla v(x)dx \quad \text{for all } u, v \in H_0^1(D). \tag{4.12}$$

We can also write (4.11) in discrete setting as

$$\widetilde{A}_0^{k_m} u_0 = \widetilde{f}_0^{k_m}, \tag{4.13}$$

with

$$\widetilde{A}_0^{k_m} = \int_D k_m(x)\nabla\Phi_i^{k^W}\nabla\Phi_j^{k^W}dx \quad i.e. \quad \widetilde{A}_0^{k_m} = R_0^{k^W} A^{k_m}(R_0^{k^W})^T, \tag{4.14}$$

where A^{k_m} is the fine-scale finite element matrix representation of the bi-linear form a^{k_m} in (4.12) and, as before,

$$(R_0^{k^W})^T = [\Phi_1^{k^W}, \cdots, \Phi_{N_c}^{k^W}],$$

and $\widetilde{f}_0^{k_m} = (R_0^{k^W})^T b$.

With respect to the equations in (4.9), we note that, if we use the naive approach of constructing basis functions for each permeability k_m, most of the computational time is spent on constructing the coarse basis functions. The ensemble-level coarse space idea helps to reduce the computational complexity of obtaining multiscale coarse approximations of the equations in (4.9). Here, the basis constructions are pre-computed using only the permeability k^W. Our numerical examples show that, under our assumptions on the ensemble $\{k_m\}$, the set of basis functions captures accurately the fine scale solution of each one of the equations in (4.9).

The multiscale space $V_0^{k^W} = \text{span}\{\Phi_i^{k^W}\}_{i=1}^{N_c}$ can also be used for preconditioning. In this case, we avoid constructing expensive (coarse solvers for the) preconditioner for each proposal. Having the expensive part of the preconditioner to be pre-computed only once (constructed using k^W) can save computational time when we need to solve the equations in (4.9) for different k's.

The fine scale matrix form of Eq. (4.9) is

$$A_0^{k_m} u_0 = f_0.$$

We propose to use the following preconditioner,

$$(\widetilde{B}^{k^W})^{-1} = (R_0^{k^W})^T (\widetilde{A}_0^{k_m})^{-1} R_0^{k^W} + \sum_{i=1}^{N} R_i^T (A_i^{k_m})^{-1} R_i, \tag{4.15}$$

where the local matrices are defined in (4.7).

The ensemble-level preconditioner here involves solving local problems using the

current permeability k_m. It also involves solving a coarse-scale system with fixed multiscale basis $\{\Phi_i^{k^W}\}_{i=1}^{N_c}$ coming from k^W, i.e. the coarse matrix will be assembled each time as shown in (4.14). Similar to the idea of the ensemble MsFEM, since the multiscale basis is pre-computed, computational time is saved, i.e. the methods is more efficient than in the case of constructing a new coarse problem every time for different k_m.

4.4. Numerical results

In this section, we present representative numerical results for ensemble coarse-scale approximation (4.13), and for the two level ensemble additive preconditioner (4.15) with the local spectral multiscale coarse space as discussed above. We also present a sampling example using both ensemble coarse-scale approximation and ensemble preconditioner to accelerate the forward computation, and consequently make the sampling process more efficient.

4.4.1. Ensemble-level multiscale and preconditioner

The equation $-\text{div}(k\nabla u) = 1$ is solved with boundary conditions $u = x + y$ on ∂D, while the multiscale basis functions are constructed through solving $-\text{div}(k^W \nabla u) = 1$. We take $D = [0, 1] \times [0, 1]$ that is divided into 8×8 equal square subdomains. Inside each subdomain we use a fine-scale triangulation, where triangular elements constructed from 10×10 squares are used.

We consider the scalar coefficients $k(x)$ and $k^W(x)$ as in (4.8), depicted in Figure 4.4 that corresponds to a background 1 and high conductivity channels $\eta = 10^4$.

First, the accuracy of MsFEMs is investigated (see also [27] for more details). In Figure 4.5, we compare coarse-scale approximations of the solution for various spaces

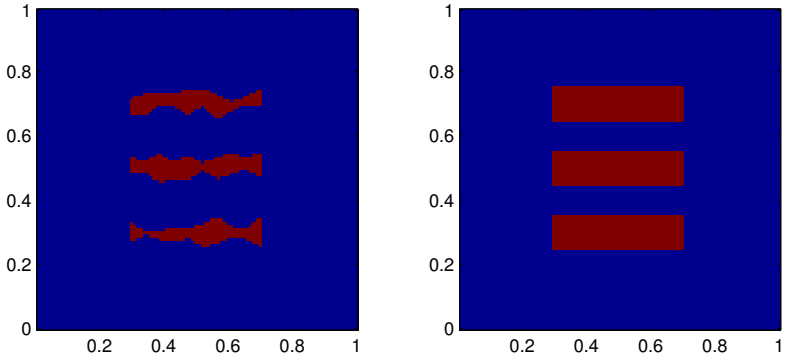

Fig. 4.4. Coefficients k (left) and k^W (right).

on 8×8 coarse grid. In the top left figure, the fine-scale solution is depicted. In the top middle figure, the solution computed with multiscale basis functions with linear boundary conditions (χ^{ms}) is plotted. This multiscale coarse space is denoted as MS. In the top right, we depict the solution computed with energy minimizing basis functions (χ^{emf}) and denote the coarse space to be EMF. The figures in the second row correspond to coarse-scale approximations computed using spectral basis functions, where we use the eigenvectors that correspond to asymptotically small eigenvalues as the contrast increases. Here, the bottom left figure corresponds to the case where the initial space consists of piecewise linear functions, the bottom middle figure corresponds to the case where the initial space consists of multiscale basis functions with linear boundary conditions, and the bottom right figure corresponds to the case where the initial space consists of energy minimizing multiscale basis functions. The multiscale coarse spaces in these cases are denoted as LSM_1, LSM-RE and RLSM, respectively. From the results, we can conclude that the ensemble-level MsFEM does provide good approximations to fine scale solution when proper local

boundary conditions are applied.

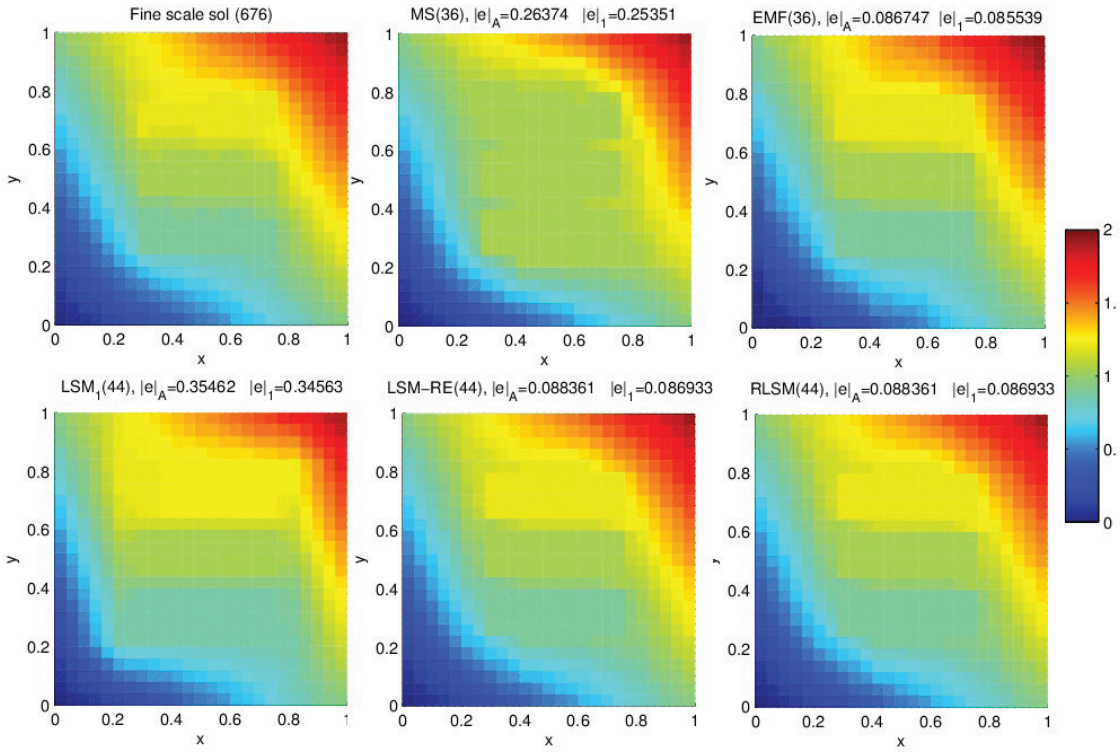

Fig. 4.5. Coarse-scale approximations when $\eta = 10^4$. The errors are in the energy norm ($|\cdot|_A$) and H^1 norm ($|\cdot|_1$). The results are MS(36), $|e|_A = 0.26374$, $|e|_1 = 0.25351$; EMF(36), $|e|_A$=0.086747, $|e|_1 = 0.085539$; $\text{LSM}_1(44)$, $|e|_A = 0.35462$, $|e|_1 = 0.34563$; LSM-RE(44), $|e|_A = 0.088361$, $|e|_1 = 0.086933$; RLSM(44), $|e|_A = 0.088361$, $|e|_1 = 0.086933$.

For preconditioning results, we build a precondtioner for $-\text{div}(k\nabla u) = 1$ based on k^W by (4.15). We will investigate the behavior of the condition number as we increase the contrast for various choices of coarse spaces. We will show that, with the ensemble preconditioner (4.15), one can achieve contrast-independent results with a small dimensional coarse space. In our simulations, we run the preconditioned conjugate gradient (PCG) until the ℓ_2 norm of the residual is reduced by a factor of 10^{10}. In Table 4.1, we show the number of PCG iterations and estimated condition

numbers. We also show the dimensions of the coarse spaces. We use $\widetilde{\text{LSM}}$ to denote the local spectral multiscale coarse space dened in (4.4) where $\tilde{k}$ in (4.3) is used in computing local eigenvalues with $\chi_i = \chi_i^{\text{MS}}$. When energy minimizing is used, the space is denoted as $\widetilde{\text{LSM}}$(EMF). The From the table we can observe that the number of iterations and condition numbers do not change as the contrast increases for the last three coarse spaces, when the ensemble-level preconditioners are used. On the contrary, the ensemble precondtioner methods fail for the first three coarse spaces.

Table 4.1. Number of iterations until convergence and estimated condition number for the PCG and different values of the contrast η with the coefficient depicted in Figure 4.4. We set the tolerance to 1e-10.

η	Linear	MS	EMF
10^3	51(5.05e+002)	52(2.76e+002)	46(1.23e+002)
10^6	80(5.12e+005)	77(2.61e+005)	70(1.05e+005)
10^9	152(5.12e+008)	170(2.61e+008)	147(1.05e+008)
Dim	49	49	49
η	LSM_1	$\widetilde{\text{LSM}}$	$\widetilde{\text{LSM}}$(EMF)
10^3	35(1.22e+001)	33(1.06e+001)	35(1.13e+001)
10^6	40(1.52e+001)	38(1.24e+001)	39(1.22e+001)
10^9	46(1.52e+001)	41(1.24e+001)	42(1.22e+001)
Dim	61	61	61

4.4.2. Application in permeability sampling

The inverse problem of finding the permeability field given certain observations (Section 2.4) is also widely studied [34, 92]. For the inverse problem, a sampling process is often involved. This sampling process requires a large number of forward solutions of the governing equation (4.2), which is usually expensive even for a single solution. The inversion ability of sampling can be restricted by the number of forward solutions. The accuracy and efficiency of the methods used to solve the governing equations are then important. The speed up in forward solving will allow the sam-

pling process to accommodate more realizations. This will lead to reliable results by convergence theorem. As we discuss above, both ensemble MsFEM and ensemble preconditioner are designed for saving computational time, so they can improve the efficiency of a sampling process.

In our numerical example, we take a reference permeability, and use a two-stage MCMC (see [34]), to recover this reference permeability. The ensemble MsFEM results (4.13) will be used as an approximation on coarse stage, and the fine stage solution is obtained by using the ensemble preconditioner (4.15) constructed in Section 4.3.2.

Now the forward equation $-\text{div}(k\nabla u) = 1$ is solved with boundary conditions $u = x + y$ on ∂D. The domain $D = [0, 1] \times [0, 1]$ is divided into 8×8 equal square subdomains, with a fine-scale triangulation consisting 10×10 squares as in Section 4.4.1.

The coarse stage MCMC chain has 1000 iterations, with all the proposals coming from a library of possible permeabilities. The acceptance rate is 0.3191 for the two-stage MCMC, while a standard MCMC has acceptance rate at 0.1419. More discussions about two-stage MCMC can be found in [34]. By plotting the solution errors versus iterations in Figure 4.6, we can observe that the MCMC chain converges. Figure 4.7 shows that the desired reference permeability field could be recovered through our sampling process by taking advantages of the ensemble methods.

4.5. Remarks

The success of ensemble-level MsFEM and ensemble-level preconditioner requires good pre-knowledge of the topological structure in the ensemble permeability fields. The possible lack of good understanding of the geometric structure brings limitations

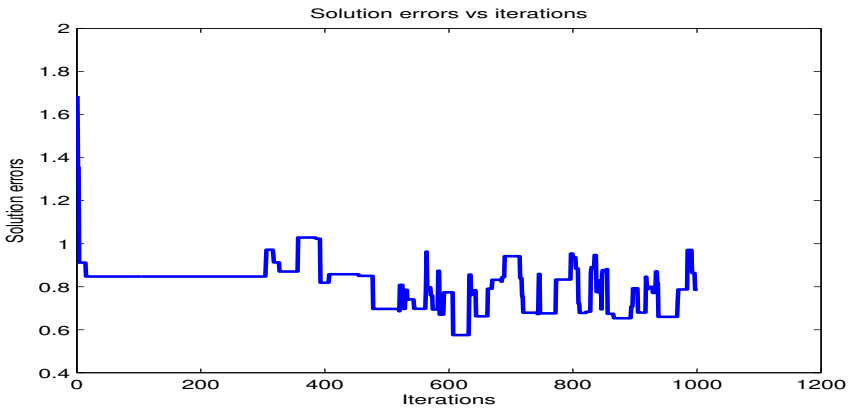

Fig. 4.6. Solution errors vs iterations.

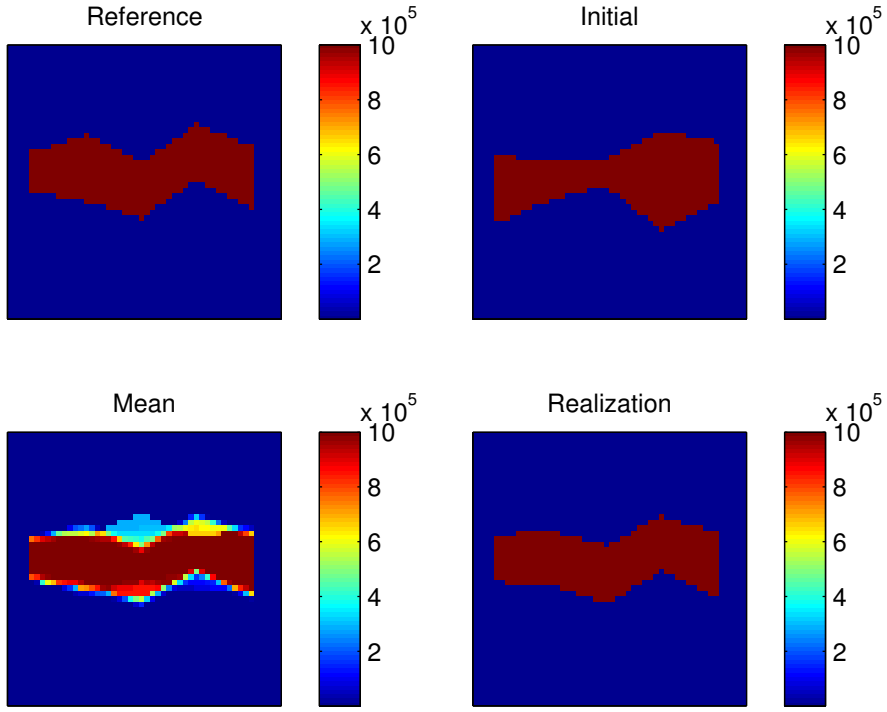

Fig. 4.7. Top left: The true permeability field. Top right: Initial permeability field. Bottom left: The mean of the sampled permeability fields from two-stage MCMC. Bottom Right: One of the sampled permeability fields.

to these methods.

Currently, we assume that the ensemble set $\{k_m\}_{m\in I}$ is topologically similar. The coarse basis functions are built in set $\{k_m\}_{m\in\{i_1,\cdots,i_N\}}$ ($k^W = \max_m(k_m)$ in the description of the methods for simplicity). We also assume that the set $\{i_1,\cdots,i_N\}$ can be properly chosen. But in practice, it is possible that the understanding to the ensemble's topological structure is incomplete, and so the chosen of $\{k_m\}_{m\in\{i_1,\cdots,i_N\}}$ is not suitable. For example, we now assume that the number of high permeability channels and inclusions in the ensemble members are the same and known. But in practice it is hard to get full knowledge of this number. It is possible that channels break into pieces, cross each other and so on. We also assume that the orientation of the channels in the ensemble set are the same and known. This is also a simplified assumption. It is worth considering the situations that the orientation of channels varies in the ensemble members.

The above situations can make the choice of $\{k_m\}_{m\in\{i_1,\cdots,i_N\}}$ difficult. An appropriate set needs to be select to represent structures of all members in the ensemble well. We have some numerical examples showing that that the basis functions built for k^W can fail to work for k_m, if k^W has high permeability channels smaller than k_m. Including more permeability fields in the chosen set to build coarse basis functions can give better results, when the geometrical configurations get complicated in the ensemble. One possible way to extend the above approaches is using the reduced basis method to construct a reduced dimensional local approximation that allows quickly computing the local spectral problem [30].

CHAPTER V

ANOVA-BASED MIXED MSFEM AND APPLICATIONS IN STOCHASTIC TWO-PHASE FLOWS

The stochastic partial differential system has been widely used to model physical processes, where the inputs involve large uncertainties. Flows in random and heterogeneous porous media are one of the cases where the random inputs (e.g., permeability) are often modeled as a stochastic field with a high-dimensional random parameter. To treat the high dimensionality and heterogeneity efficiently, we do model reduction in both stochastic space and physical space. We use an analysis of variance (ANOVA)-based mixed multiscale finite element method (MsFEM) to decompose the high-dimensional problem into a set of lower dimensional problems, which require less computational complexity, and the mixed MsFEM can capture the heterogeneities on a coarse grid. To enhance the efficiency of traditional ANOVA, we develop a new adaptive ANOVA method, where the most active dimensions can be selected before conducting ANOVA decomposition. A number of numerical examples in two-phase stochastic flows are presented and showed the performance of the ANOVA-based mixed MsFEM.

The rest of the chapter is organized as follows. We first present the introduction and background for two-phase flows in stochastic fields. The following section is devoted to the description of mixed MsFEM and probability collocation method (PCM). We then introduce ANOVA-based and adaptive ANOVA-based mixed MsFEM methods. After that, we present the numerical results using the methods introduced for flows in different random porous media.

5.1. Introduction

An important challenge in modeling flows in porous media is the treatment of complex heterogeneities and uncertainties in permeability fields. The high and low permeability may be connected at different scales. The uncertainty may arise from measurement corruption and incomplete knowledge of the physical properties. One way to describe the uncertainty is to model the permeability as a random field, which is often experimentally determined by a covariance function. Stochastic partial differential equations (SPDEs) are often used in modeling complex physical and engineering systems with uncertainties, which are usually characterized by a random field with high-dimensional parameters. To simultaneously tackle the high dimensionality and the heterogeneities, we use the analysis of variance (ANOVA)-based mixed multiscale finite element method (MsFEM) for simulation.

Sampling in high-dimensional random space is very difficult. If the sampling of random space is conducted in full random space, then the number of samples increases drastically with respect to the dimension of the random space. This is the notorious *curse of dimensionality*, which poses great difficulties for the stochastic approximation in a high-dimensional stochastic space. Instead of dealing with the full high-dimensional random space, the ANOVA representation can decompose a high-dimensional model into a set of low-dimensional models [36, 38, 65, 66, 107, 108]. It was first introduced by Fisher in [36]. The decomposition is motivated by observing that there are dominant dimensions and interactions among dimensions in many practical physical systems. In this case, the system accuracy will not be harmed too much if only relatively low dimensions and the interactions among the low dimensions are taken into account. The ANOVA decomposition then splits a high-dimensional stochastic model into many low-dimensional stochastic models, which need much less

computational efforts. The curse of dimensionality can be considerably suppressed using the ANOVA.

For each low-dimensional problem, a sparse grid polynomial collocation method (PCM) can be used. The PCM was first introduced in [95] and has been studied extensively over the years [53]. The efficiency of the Clenshaw-Curtis-based sparse grid stochastic collocation was demonstrated by comparing it with other stochastic methods on an elliptic problem in [105]. In [63], an adaptive hierarchical sparse grid collocation algorithm was developed. In [58, 59, 60], a multi-element PCM was employed to study the random roughness problem, stochastic compressible flow, and plasma flow problems.

At each collocation point, we solve deterministic flow equations in porous media. To treat the heterogeneity of porous media and recover the mass conservative velocity field, we use mixed MsFEM [2, 17, 32]. The main idea of mixed MsFEMs is to incorporate the small-scale information into finite element basis functions and couple them through a global mixed formulation of the problem. The mixed MsFEMs share some similarities with a number of multiscale numerical methods such as the multiscale finite volume method [49], residual-free bubbles [13, 88], two-scale conservation subgrid method [8], variational multiscale method [48], and multiscale mortar method [9].

In this chapter, we combine ANOVA decomposition and mixed MsFEM for stochastic two-phase flow equations. The combination of these two model reduction techniques serves as a remedy to handle the large-scale problems in both stochastic and spatial spaces. Notice that because the ANOVA decomposition has a large number of terms if the dimension of random parameter is high, the total computational cost can still be prohibitive. To improve the efficiency and reduce computational efforts, we propose a novel adaptive ANOVA technique. The advantage of the pro-

posed adaptive technique is to obtain sensitivity with regard to every dimension of the stochastic space, through building a response surface of the function of interest with small computational efforts. The proposed adaptive ANOVA technique is different from the previous adaptive ANOVA methods (ref. [66, 107]). We note that some adaptive ANOVA, techniques in both spatial and random spaces, were developed in [66] in the framework of the heterogeneous multiscale method. In this work, we focus on analyzing the errors introduced by ANOVA decomposition and mixed MsFEM, respectively. It is important to understand the behavior of the two error contributions and find a good trade-off between them.

In our numerical examples, we consider the permeability fields with different statistical properties and heterogeneous structure. We compute the reference solution on the fine grid and use the Monte Carlo method to sample the random space, and compare the solutions using ANOVA-based mixed finite element method (FEM) and ANOVA-based mixed MsFEM with the reference solution. From these computations, we report errors from both ANOVA truncation and mixed MsFEM discretization, separately. We observe that, in our numerical examples, the error introduced by the mixed MsFEM method is the dominant one. By analyzing different functions, we also better understand the effectiveness of ANOVA-based mixed MsFEM for different quantities of interest in oil reservoir simulations. The novel adaptive ANOVA-based MsFEM method is numerically comparable with existing adaptive ANOVA techniques with less online computations.

5.2. Background and notations

5.2.1. Two-phase flow problem

We briefly present the two-phase flow problem introduced in Section 2.1 again in this section. D is a convex bounded domain in $\mathbb{R}^d$ $(d = 2, 3)$ and $(\Omega, \mathcal{F}, P)$ is a probability space, where Ω is the set of outcomes, $\mathcal{F}$ is the σ-algebra generated by Ω, and P is a probability measure. The two-phase flow (referred as water and oil, designated by subscripts w and o, respectively) in random porous media under the assumption that the displacement is dominated by viscous effects, with the effects of gravity, compressibility, and capillary pressure neglected. Porosity is considered to be constant. We formulate the stochastic two-phase flow system (see Eqs. (2.1)-(2.3)): find random fields $p(x,\omega) : \bar{D} \times \Omega \longrightarrow \mathbb{R}$, $v(x,\omega) : \bar{D} \times \Omega \longrightarrow \mathbb{R}$ and $S(x,\omega,t) : \bar{D} \times \Omega \times [0,T] \longrightarrow \mathbb{R}$ such that they almost surely (a.s) satisfy the following equations

$$\begin{cases} \operatorname{div}\big(v(x,\omega)\big) = Q_s \\ v(x,\omega) = -\lambda(S)k(x,\omega)\nabla p(x,\omega) \\ \dfrac{\partial S(x,\omega,t)}{\partial t} + \ v(x,\omega)\cdot\nabla f(S(x,\omega,t)) = 0. \end{cases} \tag{5.1}$$

Let the coefficient $k(x,\omega)$ of Eq. (5.1) be a stochastic field with second moment. To make $k(x,\omega)$ positive, we consider $k(x,\omega)$ to be a logarithmic stochastic field, i.e., $k(x,\omega) := \exp(a(x,\omega))$. Here $a(x,\omega)$ is a stochastic field and its covariance function $\operatorname{cov}[a]$ known. Then by the K-L expansion introduced in Section 2.2, it follows that $a(x,\omega)$ admits the following truncated K-L expansion, i.e.,

$$a(x,\Theta) = E[a] + \sum_{i=1}^{n} \sqrt{\lambda_i}\theta_i(\omega)\psi_i(x),$$

where $\Theta := (\theta_1, \theta_2, \cdots, \theta_n) \in \mathbb{R}^n$. By the truncated K-L expansion, $k(x,\omega) \approx$

$k(x,\Theta) = \exp\big(a(x,\Theta)\big)$. In the chapter, we assume that the stochastic field $k(x.\omega)$ can be accurately parameterized by $k(x,\Theta)$.

5.2.2. Sparse grid collocation

For stochastic systems (5.1), we are interested in the statistics (e.g., mean and variance) of solutions. These properties could be obtained by first sampling randomly from the parameter space using, for example, Monte Carlo method, then solving the deterministic problems on the samples and analyzing the corresponding results. The number of samples required in the Monte Carlo method is large, which leads to high computational cost. Instead, we use a sparse grid based probabilistic collocation method. Sparse grid collocation is known to have the same asymptotic accuracy as tensor product collocation, while requiring many fewer collocation points as the parameter dimension increases. Sparse grids have been successfully applied to stochastic collocation in many recent works (e.g., [74, 75]).

Based on the Smolyak formula [11], a set of collocation points is chosen. With these chosen collocation points and corresponding weights, the statistic properties of the solutions can be obtained. For instance, assume that $\{\theta^{(j)}\}$ is the set of collocation points and $\{w^{(j)}\}$ is the corresponding weights, $j = 1, \cdots, N_c$. At each of the collocation points, the deterministic system is solved and the output $S(x,\theta^{(j)})$ is obtained. Then the moments of $S(x,\Theta)$ can be estimated, e.g.,

$$E[S(x,\Theta)] = \int_\Omega S(x,\xi)dF(\xi) \approx \sum_{j=1}^{N_c} S(x,\theta^{(j)})w^{(j)},$$

$$\sigma^2\big(S(x,\Theta)\big) = \int_\Omega \big(S(x,\xi) - E[S(x,\Theta)]\big)^2 dF(\xi) \approx \sum_{j=1}^{N_c} S^2(x,\theta^{(j)})w^{(j)} - E^2[S(x,\Theta)].$$

5.3. Adaptive ANOVA-based mixed MsFEMs

5.3.1. ANOVA-based mixed MsFEM

ANOVA is a general set of quantitative assessment and analysis tools for capturing the high-dimensional relationships between model inputs and model outputs. ANOVA has been used for improving the efficiency of deducing high-dimensional input-output system behavior, and can be employed to relieve the computational efforts. The ANOVA method has been used in high-dimensional stochastic systems [38, 65, 107, 108]. ANOVA is based on the assumption that only relatively low order correlations of the input variables are important, which is valid in many systems. ANOVA splits a high-dimensional system into a set of low-dimensional systems to reduce the computation in high dimension.

We consider a multivariate output function $S(\Theta) : \mathbb{R}^n \longrightarrow \mathbb{R}$. $S(\Theta)$ is taken to be water saturation or other functions of interest discussed in Section 2.1, in the two-phase flow system (5.1). The statistic properties of $S(\Theta)$ can be obtained by solving the system (5.1) using the mixed MsFEM (see Section 2.3 for more details) and PCM (see Section 5.2.2). For simplicity, we use $S(\Theta)$ in this section, instead of $S_{\text{MsFEM}}(\Theta)$, to denote the solutions obtained in the multiscale framework.

Instead of solving the two-phase flow system (5.1) for $S(\Theta)$ directly, ANOVA represents $S(\Theta)$ as finite hierarchical correlated functions of input variables in the form

$$S(\Theta) = S_0 + \sum_{1\le j_1\le n} S_{j_1}(\theta_{j_1}) + \sum_{1\le j_1<j_2\le n} S_{j_1,j_2}(\theta_{j_1},\theta_{j_2}) + \cdots + S_{1,2,\cdots,n}(\theta_1,\theta_2,\cdots,\theta_n), \tag{5.2}$$

where $S_{j_k}(\theta_{j_k})$ is the first order term, $S_{j_k,j_l}(\theta_{j_k},\theta_{j_l})$ is the second order term, etc. Each of these terms is solved by the mixed MsFEM method.

In the standard ANOVA (5.2), the constant term is taken to be the mean of function $S(\Theta)$, i.e.,

$$S_0 = \int_{\Gamma^n} S(\Theta) d\mu(\Theta).$$

This gives that all higher order terms are mean zero, i.e.,

$$\int_{\Gamma^n} S_{j_1,\cdots,j_s} d\mu(\Theta) = 0,$$

which leads to orthogonality among all the terms, and the variance of $S(\Theta)$ is the sum of variances of all terms, i.e.,

$$\int_{\Gamma^n} S_{j_1,\cdots,j_s} S_{k_1,\cdots,k_l} d\mu(\Theta) = 0, \quad \text{for} \quad (j_1,\cdots,j_s) \neq (k_1,\cdots,k_l),$$

$$\sigma^2(S) = \sum_{j=1}^{n} \sum_{|\mathbf{J}|=j} \sigma^2(S_{\mathbf{J}}).$$

To avoid the computation of high-dimension integration, the Dirac measure is often used instead of the Lebesgue measure. The Dirac measure is defined as $d\mu(\Theta) = \delta(\Theta - \mathbf{c})d\Theta$, where $\mathbf{c}$ is called the anchor point. If $\mathbf{c}$ satisfies that $S(\mathbf{c}) = S_0$, then the ANOVA representation is the same as (5.2), otherwise it becomes an approximation of $S(\Theta)$, i.e.,

$$S(\Theta) \approx S(\mathbf{c}) + \sum_{1 \leq j_1 \leq n} S_{j_1}(\theta_{j_1}) + \sum_{1 \leq j_1 < j_2 \leq n} S_{j_1,j_2}(\theta_{j_1},\theta_{j_2}) + \cdots + S_{1,2,\cdots,n}(\theta_1,\theta_2,\cdots,\theta_n), \tag{5.3}$$

where

$$S_j(\theta_j) = S(\Theta)|_{\Theta=\mathbf{c}\setminus\theta_j} - S(\mathbf{c}),$$

$$S_{j,k}(\theta_j,\theta_k) = S(\Theta)|_{\Theta=\mathbf{c}\setminus(\theta_j,\theta_k)} - S_j(\theta_j) - S_k(\theta_k) - S(\mathbf{c}).$$

The accuracy of this anchored-ANOVA depends on the choice of anchor point $\mathbf{c}$. Discussions about this can be found in [108]. In this chapter, we choose the anchor

point to be the mean of random variable Θ as discussed in [65]. Because low-order terms in ANOVA expansion usually have the dominant contribution, we truncate the ANOVA to the low-order (e.g., second- or third-order) terms and use the truncated ANOVA for approximation. This can significantly reduce the dimensionality of the random inputs.

5.3.2. Adaptive ANOVA-based mixed MsFEM

The ANOVA-based mixed MsFEM reduces the computational complexity in stochastic space by dividing a high-dimensional stochastic problem into a set of lower dimensional problems, while the mixed MsFEM reduces the computational cost in spatial space. The computation in a low-dimensional system is easier, but a large number of such systems can keep the computational cost still high. For example, if the dimension of input parameter space $n = 100$, and we truncate the ANOVA up to second-order, then we have a total of $1 + \begin{pmatrix} 100 \\ 1 \end{pmatrix} + \begin{pmatrix} 100 \\ 2 \end{pmatrix} = 5051$ components (terms) in the truncated ANOVA. This computation is still not cheap. To reduce the total number of terms, the adaptive ANOVA method is often used. The dimensions with dominant interactions are called active or important dimensions. The idea of adaptive ANOVA is to retain the interactions from the important dimensions and neglect the contributions from the unimportant dimensions. The following equation describes the adaptive ANOVA representation,

$$\begin{aligned} S(\Theta) \approx S_0 + \sum_{j_1 \in D_1} S_{j_1}(\theta_{j_1}) + \sum_{(j_1 < j_2) \in D_2} S_{j_1,j_2}(\theta_{j_1}, \theta_{j_2}) \\ + \cdots + \sum_{(j_1 < j_2 < \cdots < j_\nu) \in D_\nu} S_{j_1,j_2,\cdots,j_\nu}(\theta_{j_1}, \cdots, \theta_{j_\nu}). \end{aligned} \tag{5.4}$$

In [65, 107, 108], $\nu = 2$ and $D_1 = \{1, \cdots, n\}$; then D_i , $2 \leq i \leq \nu$ are selected according to the statistical properties of the computed expansion terms. There are two criteria used to find the active dimensions based on the first-order terms in the ANOVA decomposition.

Criterion 1: Use the mean of component function S_j as the indicator to decide the active ANOVA terms (see [65]). Let

$$\eta_j^{(1)} = \frac{E[S_j]}{\sum_{j \in D_1} E[S_j]}, \quad j \in D_1;$$

then, the active dimensions D_2 can be chosen by

$$\sum_{j \in D_2} \eta_j^{(1)} \geq p, \tag{5.5}$$

where p is a proportionality constant with $0 < p < 1$ and close to 1.

Criterion 2: Use the variance of component function S_j as the indicator to decide the active ANOVA terms. Define

$$\eta_j^{(2)} = \frac{\sigma^2(S_j)}{\sum_{j \in D_1} \sigma^2(S_j)}, \quad j \in D_1.$$

The active dimensions D_2 should satisfy

$$\sum_{j \in D_2} \eta_j^{(2)} \geq p, \tag{5.6}$$

where p is a proportionality constant with $0 < p < 1$ and close to 1. This criterion is similar to the criterion used in [15] where $\sigma^2(S)$ instead of $\sum_{j \in D_1} \sigma^2(S_j)$ is used.

The active dimensions for second-order ANOVA terms can be found by the above criteria. If the active dimensions are needed for higher-order terms, we can use similar criterion,

$$\eta_{j_1,j_2}^{(1)} = \frac{E[S_{j_1,j_2}]}{\sum_{j \in D_1} E[S_j]}, \quad \eta_{j_1,j_2}^{(2)} = \frac{\sigma^2(S_{j_1,j_2})}{\sum_{j \in D_1} \sigma^2(S_j)}. \tag{5.7}$$

5.3.3. Novel adaptive ANOVA-based mixed MsFEM

In the above criteria, D_1 is taken to be $\{1, \cdots, n\}$, i.e., all the dimensions are always considered to be important in the ANOVA first-order term computation. The selection of active dimensions is then conducted for second-order terms based on the ratio (e.g., $\eta_j^{(1)}, \eta_j^{(2)}$) associated to each dimension. It is worth noticing that the dimensions with small ratio can be neglected in not only the computation of second-order terms, but also first-order terms. We propose here a variance decomposition-based method to pre-select active dimensions and simplify the computation of ANOVA decomposition starting from the first-order terms.

To pre-select the important dimensions, we solve the original high-dimensional system on the sparse grid collocation points of Ω and build a polynomial chaos approximation for $S(\Theta)$. The variance of $S(\Theta)$ is carried by the coefficients in front of basis functions in the approximation. Since the basis functions in Ω are multiplications of basis functions of dimension one, we can view basis functions to be related to interactions of certain dimensions. In this case, the coefficients are the corresponding variance coming from interactions of certain dimensions. By doing this, we can decompose the variance of $S(\Theta)$ with respect to the set of dimensions. Especially, we can obtain the variance of each dimension itself. So the importance can be estimated before ANOVA decomposition. In this case, D_1 can be selected to be a subset of $\{1, \cdots, n\}$. The sparse grid collocation method is applied in a low level, so the computation is cheap. If a higher-level sparse grid collocation is affordable, more information about the interactions between dimensions can be obtained, and therefore another criterion of selection of active dimensions for higher-order terms is provided.

To be specific about the variance decomposition-based adaptive ANOVA method,

let us consider a scalar function

$$S(x,\Theta):\bar{D}\times\mathbb{R}^n\longrightarrow\mathbb{R}. \tag{5.8}$$

Let $\{\phi_k(\Theta)\}$ be the generalized polynomial chaos (gPC) basis function satisfying $E[\phi_i(\Theta)\ \phi_j(\Theta)]=\delta_{ij}r_i$ in this section and let $\mathbb{P}_N^n(\Theta)$ be the space of all polynomials of $\Theta\in\mathbb{R}^n$ of degree up to N. Then the orthogonal gPC projection of (5.8), for any fixed x, is

$$S_N(\Theta)=\mathbb{P}_N S=\sum_{|\mathbf{k}|=0}^{N}s_{\mathbf{k}}\phi_{\mathbf{k}}(\Theta),$$

where the expansion coefficients are obtained as

$$s_{\mathbf{k}}=\frac{1}{r_{\mathbf{k}}}E[S(\Theta)\phi_{\mathbf{k}}(\Theta)]=\frac{1}{r_{\mathbf{k}}}\int S(\theta)\phi_{\mathbf{k}}(\theta)dF_\Theta(\theta),\quad\forall|\mathbf{k}|\le N,$$

where $r_{\mathbf{k}}=E[\phi_{\mathbf{k}}^2]$ is the normalization constant of the basis ϕ_k, and $F_\Theta(\theta)=P(\theta\le\Theta)$ is the probability distribution of Θ.

Integration rules can be used to approximate the integrals in the expansion coefficients of the continuous gPC. Let $\{\theta^{(1)},\cdots,\theta^{(m)}\}$ be the sparse grid collocation points, then discrete projection of the solution is

$$S_N(\Theta)=\sum_{|\mathbf{k}|=0}^{N}\hat{s}_{\mathbf{k}}\phi_{\mathbf{k}}(\Theta),$$

where the expansion coefficients are $\hat{s}_{\mathbf{k}}=\frac{1}{r_{\mathbf{k}}}\sum_{j=1}S(\theta^{(j)})\phi_{\mathbf{k}}(\theta^{(j)})w^{(j)}$, and $\{w^{(j)}\}$ are the collocation weights. The coefficients $\{\hat{s}_{\mathbf{k}}\}$ are approximations to the exact projection coefficients $\{s_{\mathbf{k}}\}$. The moments of $S(x,\Theta)$ can be approximated by the moments of the approximation $S_N(\Theta)=\sum_{|\mathbf{k}|=0}^{N}\hat{s}_{\mathbf{k}}\phi_{\mathbf{k}}(\Theta)$. The mean

$$\mu=E[S(\Theta)]\approx E[S_N(\Theta)]=\sum_{|\mathbf{k}|=0}^{N}\hat{s}_{\mathbf{k}}E[\phi_{\mathbf{k}}(\Theta)]=\hat{s}_{\mathbf{0}},$$

by orthogonality of $\{\phi_k(\Theta)\}$. Then the variance

$$\sigma^2(S) \;=\; E[S(\Theta)-\mu]^2 \approx E[S_N-\hat{s}_{\mathbf{0}}]^2 = \sum_{|\mathbf{k}|=1}^{N} \hat{s}_{\mathbf{k}}^2 r_{\mathbf{k}}.$$

The variance can be further written as

$$\sigma^2(S) = \sum_{|\mathbf{k}|=1} \hat{s}_{\mathbf{k}}^2 r_{\mathbf{k}} + \cdots + \sum_{|\mathbf{k}|=N} \hat{s}_{\mathbf{k}}^2 r_{\mathbf{k}}. \tag{5.9}$$

Denote $\upsilon = \sigma^2(S)$, $\upsilon_i = \hat{s}_{\mathbf{k}}^2 r_{\mathbf{k}}$, where $\mathbf{k} = (0,\cdots,0,1,0,\cdots,0)$, i.e., $k_i = 1$ and 0 anywhere else. Similarly, $\upsilon_{ij} = \hat{s}_{\mathbf{k}}^2 r_{\mathbf{k}}$, where $k_i = k_j = 1$ or $k_i = 2$ and 0 anywhere else. Then the expansion (5.9) can be rearranged with respect to dimensions, i.e.,

$$\upsilon = \sum_{1\le i\le n} \upsilon_i + \sum_{1\le i\le j\le n} \upsilon_{ij} + \cdots. \tag{5.10}$$

Remark: Since υ_i, $1 \le i \le n$ is the coefficient of basis function $\phi_{\mathbf{k}}(\Theta) = \sum_{j=1}^{n} \phi_{k_j}(\theta_j) = \phi_{k_i}(\theta_i)$, it is considered as the variance associated with the ith dimension. Similarly, as υ_{ij} corresponds to $\phi_{\mathbf{k}}(\Theta) = \phi_{k_i}(\theta_i)\phi_{k_j}(\theta_j)$, it is viewed as the variance associated with interaction of ith and jth dimensions. The other terms can be explained in the same way.

Based on the variance decomposition (5.10), the important dimensions can be selected by sensitivity analysis. Define

$$\eta_j^{(3)} = \frac{\upsilon_j}{\upsilon} \text{ and } \eta_{ij}^{(3)} = \frac{\upsilon_{ij}}{\upsilon}.$$

Then, the important dimensions can be selected similarly to (5.5) and (5.6).

Criterion 3: The active dimension D_1 should satisfy

$$\sum_{j\in D_1} \eta_j^{(3)} \ge p, \tag{5.11}$$

where p is a proportionality constant with $0 < p < 1$ and close to 1.

The advantage of the proposed method here is that, by building a response surface of the function of interest $S(x, \Theta)$ with small cost, we can get the information of sensitivities of each dimension. Then, the adaptivity can start from the first-order term in ANOVA based on (5.11), instead of higher-order terms based on (5.5) and (5.6). Further information about the interactions among dimensions can be obtained if more computational cost is affordable, i.e., when the response surface is built in a higher level of sparse grid points. In this case, the second-order terms can only be computed in the preselected pairs, instead of every pair of important dimension as in Section 5.3.2. Also, Criteria 1 and 2 (see equations (5.5) and (5.6)) assume that the important dimension interactions only happen between active dimensions. But, it is possible that the less important dimension has large interactions with other dimensions. Criteria 1 and 2 will ignore these interactions. The variance decomposition-based adaptive method can provide additional information on the interaction among dimensions to avoid unnecessary over adaptivity.

5.4. Numerical results

In this section, we assume that the random permeability field $k(x, \omega) = \exp\big(a(x, \omega)\big)$ is a log-normal stochastic process as before. Here, the covariance function of $a(x, \omega)$ has the form

$$\mathrm{cov}[a](x, y) = \sigma^2 \exp\Big(-\frac{|x_1 - y_1|^2}{2l_1^2} - \frac{|x_2 - y_2|^2}{2l_2^2}\Big). \tag{5.12}$$

l_1 and l_2 are the correlation lengths in the horizontal and vertical dimension. The stochastic field $a(x, \omega)$ can be approximated by a truncated K-L approximation. In practice, the K-L expansion of $a(x, \omega)$ can be written as

$$a(x, \omega) = E[a] + \alpha \sum_{i=1}^{N} \sqrt{\lambda_i}\theta_i(\omega)\psi_i(x), \tag{5.13}$$

where α is a constant. In our numerical examples, we sample $a(x,\omega)$ in (5.13) by generating random variables θ_i from a uniform distribution on $[-1,1]$.

We take $k_{rw}(S) = S^2$, $\mu_w = 0.1$, $k_{ro}(S) = (1-S)^2$, and $\mu_o = 1$ in simulations. The permeability field $a(x,\Theta)$ is given on a fine grid. The water is injected at the lower-left corner, and the producer is at the upper-right corner. To validate ANOVA-based mixed MsFEM, we compare the results of solving stochastic two-phase flows (5.1) by the following methods:

- Monte Carlo method associated with mixed FEM on the fine grid (MC-FEM);
- ANOVA-based mixed FEM on fine grid (ANOVA-FEM);
- ANOVA-based mixed MsFEM on coarse grid (ANOVA-MsFEM).

When mixed MsFEM is used, the fine grid is coarsened to form a uniform coarse grid. We solve the pressure equation on the coarse grid using the mixed MsFEM and then reconstruct the fine-scale velocity field as a superposition of the multiscale basis functions. The reconstructed field is used to solve the saturation equation by the finite volume method on the fine grid. We solve the two-phase flow system by the classical IMPES (Implicit Pressure Explicit Saturation).

Monte Carlo results are obtained from 10^4 simulations and ANOVA is based on Smolyak sparse grid collocation points with level 2. Various production characteristics introduced in Section 2.1 are compared. We compute the saturation S at 0.2PVI and 0.6PVI and the water-cut curve $F(t)$ defining the fraction of water in the produced fluid as a function of PVI. We monitor the breakthrough time T_w defined as $F^{-1}(10^{-5})$ at the producer and the cumulative oil production at 0.6PVI, i.e.,

$$Q_o = -\frac{1}{\int_D \varphi dx}\int_0^{0.6PVI}\left(\int_D \min(q_o(x,\tau),0)dx\right)d\tau.$$

We consider all the results computed by MC-FEM to be the reference solutions, and we measure errors defined as $e(S) = |\|E[S]\|_{L^1} - \|E[S_{ref}]\|_{L^1}|/\|E[S_{ref}]\|_{L^1}$, $e(F) = \|E[F] - E[F_{ref}]\|_{L^2}/\|E[w_{ref}]\|_{L^2}$, $e(T_w) = |E[T_w] - E[(T_w)_{ref}]|$ and $e(Q_o) = |E[Q_o] - E[(Q_o)_{ref}]|$, where $E[\cdot]$ stands for mean solutions from either ANOVA-FEM or ANOVA-MsFEM. The standard deviation errors are defined in the same manner.

There are different kinds of errors when we use the listed three methods to solve the governing equations. We define e_{total} to be the total error, which is the expectation of absolute error between MC-FEM and ANOVA-MsFEM. The e_{stoch} is defined as the error from dimension reduction and collocation, i.e., the error between MC-FEM and ANOVA-FEM, and the e_{ms} is defined to be error from mixed MsFEM discretization. For water-cut function w, for instance, we have definitions of these errors as follows: $e_{total} = \|E[F_{\text{MC-FEM}}(t)] - E[F_{\text{ANOVA-MsFEM}}(t)]\|_{L^2}$, $e_{stoch} = \|E[F_{\text{MC-FEM}}(t)] - E[F_{\text{ANOVA-FEM}}(t)]\|_{L^2}$, and $e_{ms} = \|E[F_{\text{ANOVA-FEM}}(t)] - E[F_{\text{ANOVA-MsFEM}}(t)]\|_{L^2}$. Errors for the other functions are defined in the same way.

5.4.1. Random permeability field

In this example, we take $l_1 = 0.2$, $l_2 = 0.05$ and $\sigma^2 = 1$ in (5.12), $E[a] = 1$ and $\alpha = 0.05$ in (5.13). The permeability field $a(x, \Theta)$ is given on a 80×80 fine grid. The fine grid is coarsened to form a uniform 8×8 coarse grid, so that each block in the coarse grid contains a 10×10 cell partition from the fine grid. We truncate K-L expansion to be $N = 20$ terms, so that the dimension of this stochastic system is 20.

Saturations at different times and water-cut results are shown in Figures 5.1, 5.2 and 5.3. The quantitative errors are reported in Tables 5.1 and 5.2. For saturation function S at different times, ANOVA-FEM behaves better than ANOVA-MsFEM comparing with the reference ones. The relative large variance of the saturation function is in the flow front as in Figures 5.1 and 5.2. The variance behaves differently

Table 5.1. Relative errors of saturation S at 0.2, 0.6PVI and water-cut F.

	ANOVA-FEM		ANOVA-MsFEM	
	Mean error	Std error	Mean error	Std error
$e(S)$ at 0.2PVI	3.422170e-09	2.184847e-03	3.422172e-09	6.116023e-03
$e(S)$ at 0.6PVI	3.135168e-07	2.745868e-03	3.616744e-03	5.136930e-03
$e(F)$	8.544170e-05	2.620252e-02	1.571879e-02	1.050685e-01

on horizontal and vertical directions too, since the correlation lengths are different in these two directions. It is also interesting to notice that when the time is increasing from 0.2PVI to 0.6PVI, the accuracy of solutions is decreasing. For both ANOVA-FEM and ANOVA-MsFEM, the relative errors increase. This comes from the uncertainty accumulation in time. At 0.2PVI, ANOVA-MsFEM is as good as ANOVA-FEM in mean value, with the same magnitude of standard deviation. When the time increases, ANOVA-MsFEM results are not comparable to ANOVA-FEM. The same situation can be discovered by looking at the standard deviations at 0.2PVI and 0.6PVI separately in Figures 5.1 and 5.2. The standard deviations have the same pattern at 0.2PVI, while at 0.6PVI, the standard deviation from ANOVA-MsFEM is quite different from the above two cases. Water-cut results are the same as saturation at 0.6PVI in the sense of the behavior of these two methods. The standard deviation of water-cut has a large value around water breakthrough time for all three cases in Figure 5.3. From Table 5.2, the absolute errors of water breakthrough time T_w and cumulative oil production Q_o, we can see that ANOVA-MsFEM provides good approximations to these function values, while ANOVA-FEM behaves better than ANOVA-MsFEM. However, the computation of the mixed FEM is more expensive than the mixed MsFEM.

Table 5.2. Absolute errors of water breakthrough time T_w and cumulative oil-production Q_o.

	ANOVA-FEM		ANOVA-MsFEM	
	Mean error	Std error	Mean error	Std error
$e(T_w)$	7.408498e-07	2.345094e-11	2.290520e-06	2.629305e-11
$e(Q_o)$	1.592182e-07	6.972235e-09	1.662888e-03	4.961346e-10

Table 5.3. Different errors of water-cut F, water breakthrough time T_w and cumulative oil-production Q_o.

	e_{total}	e_{stoch}	e_{ms}
F	1.527618e-02	1.147299e-03	1.412888e-02
T_w	2.290520e-06	7.408498e-07	3.031370e-06
Q_o	1.662888e-03	1.592182e-07	1.662729e-03

There are errors introduced by ANOVA-MsFEM as we see from the above analysis (Tables 5.1 and 5.2). It is then important to find out which error is dominant. After fully understanding the structure of errors, improved methods can be developed to target at the dominant error. The error introduced by MsFEM, e_{ms}, has the same magnitude as the total one e_{total} in Table 5.3 . We conclude here that the error introduced by the multiscale method is the dominant one compared with the error introduced by ANOVA representation and collocation methods. Seeking better multiscale methods for better approximations is one of the future directions for ANOVA-based methods.

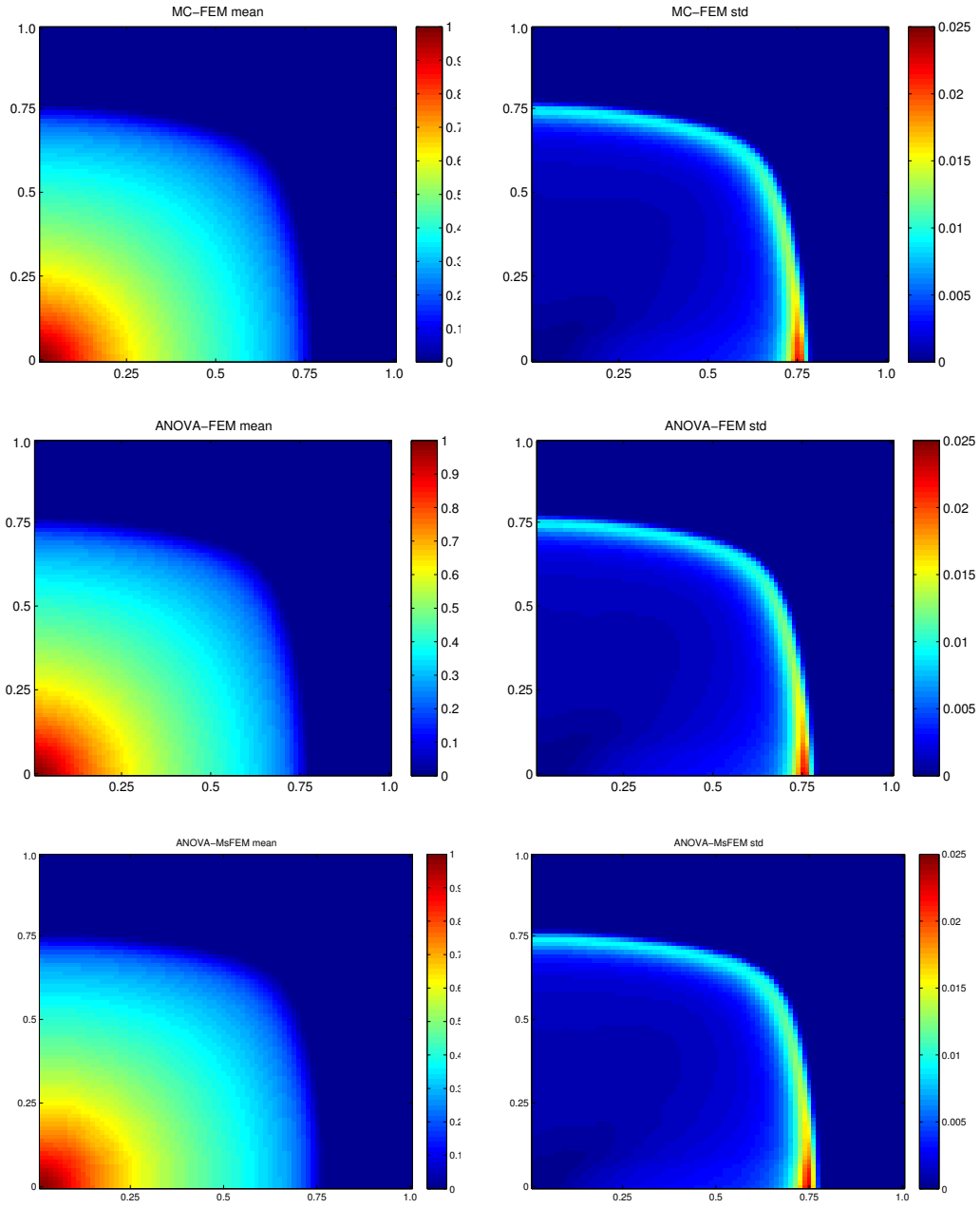

Fig. 5.1. Mean of saturation S and standard deviation of saturation S at 0.2PVI. Left column: mean; Right column: standard deviation. Top row: MC-FEM; Middle row: ANOVA-FEM; Bottom row: ANOVA-MsFEM.

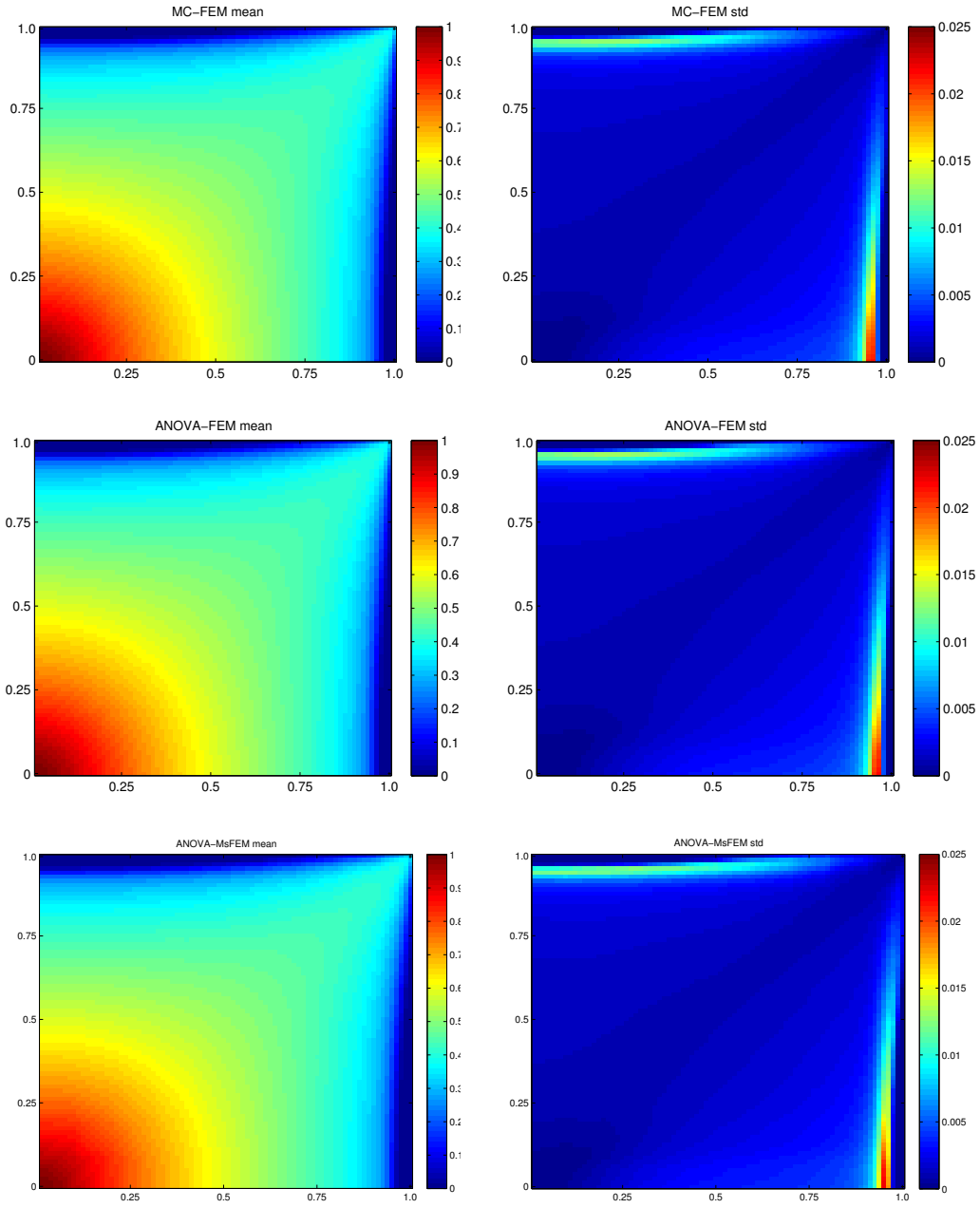

Fig. 5.2. Mean of saturation S and standard deviation of saturation S at 0.6PVI. Left column: mean; Right column: standard deviation. Top row: MC-FEM; Middle row: ANOVA-FEM; Bottom row: ANOVA-MsFEM.

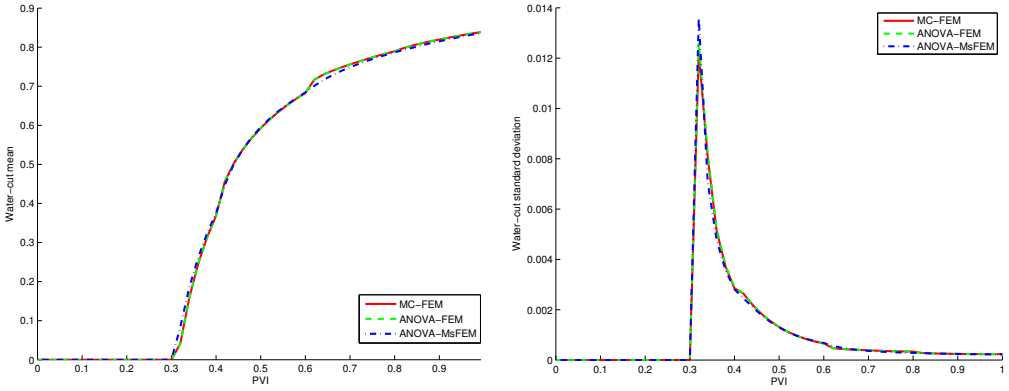

Fig. 5.3. Comparison of mean and standard deviation of water-cut F.

5.4.2. Random permeability field with channelized structure

For our second numerical example, we take $l_1 = 0.2$, $l_2 = 0.05$, and $\sigma^2 = 1$ in (5.12), the same as in the first example. In (5.13), we choose $\alpha = 1$, and $E[a]$ is chosen to have a channelized feature and shows the dominant feature of the permeability $a(x, \omega)$. We choose $N = 20$ terms in the truncated K-L expansion, i.e., $a(x, \omega) = E[a] + \sum_{i=1}^{20} \sqrt{\lambda_i}\theta_i\psi_i$. Consequently,

$$k(x, \omega) = \exp(a(x, \omega)) = \exp(E[a]) \exp\left(\sum_{i=1}^{20} \sqrt{\lambda_i}\Theta_i\psi_i\right) := k_1(x)k_2(x, \omega).$$

For each realization, the permeability $k(x, \omega)$ is defined on a 60×60 fine grid. Figure 5.4 depicts the logarithm of $k_1(x)$ (left) and an arbitrary realization of logarithm of $k_2(x, \omega)$ (right). From the permeability, we can see that $k_1(x)$ represents a main feature of the random permeability $k(x, \omega)$. Here the mixed MsFEM is performed on a 6×6 uniform coarse grid.

Now the variance in this example is larger than the previous example. The results in Tables 5.4 and 5.5 are larger in magnitude than in Tables 5.1 and 5.2. However, the trends are the same as in the first example. The ANOVA-FEM is generally better than ANOVA-MsFEM. The accuracy is decreasing with respect to the increasing time. But

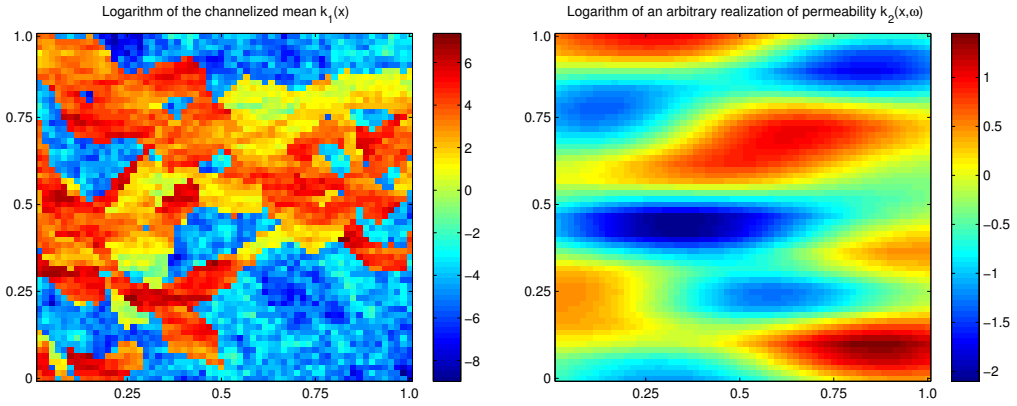

Fig. 5.4. The logarithm of $k_1(x)$ (left) and an arbitrary realization of logarithm of $k_2(x, \omega)$ (right).

the standard deviation of saturations computed by ANOVA-MsFEM is better than the result computed by ANOVA-FEM. The natural inference is that ANOVA-FEM is better than ANOVA-MsFEM, as the mixed MsFEM results are approximations for FEM solutions with less computational cost. But since ANOVA is applied here only up to second-order terms, the process of approximating a nonlinear operator by finite linear operations introduces extra errors. If ANOVA is expanded to high-order terms, the results will converge to the true value, and ANOVA-FEM should be better than ANOVA-MsFEM.

We also notice that, ANOVA approximation has better results for the mean of functions than the standard deviation. The uncertainty of saturations S are coming from where the flow front is (Figures 5.5 and 5.6). The large standard deviation of water-cut w is again around water breakthrough time (Figure 5.7), while the magnitude of standard deviation in this example is larger than the previous example (Figure 5.3), as the variance of the parameter is larger. For this example, we can also see that MsFEM still is the dominant error in the computation from Table 5.6.

Table 5.4. Relative errors of saturation S at 0.2, 0.6PVI and water-cut F.

	ANOVA-FEM		ANOVA-MsFEM	
	Mean error	Std error	Mean error	Std error
$e(S)$ at 0.2PVI	9.472140e-03	3.138906e-01	7.615875e-03	2.625666e-01
$e(S)$ at 0.6PVI	4.296857e-03	3.612245e-01	9.758884e-03	2.605877e-01
$e(F)$	1.515789e-02	5.825024e-01	2.001007e-02	3.293715e-01

Table 5.5. Absolute errors of water breakthrough time T_w and cumulative oil-production Q_o.

	ANOVA-FEM		ANOVA-MsFEM	
	Mean error	Std error	Mean error	Std error
$e(T_w)$	6.797262e-02	4.350505e-03	6.884711e-02	4.416420e-03
$e(Q_o)$	3.126008e-04	2.948469e-05	1.858780e-03	6.022319e-06

Table 5.6. Different errors of water-cut F, water breakthrough time T_w and cumulative oil-production Q_o.

	e_{total}	e_{stoch}	e_{ms}
F	9.937488e-02	7.527776e-02	1.200735e-01
T_w	6.884711e-02	6.797262e-02	1.368197e-01
Q_o	1.858780e-03	3.126008e-04	1.546179e-03

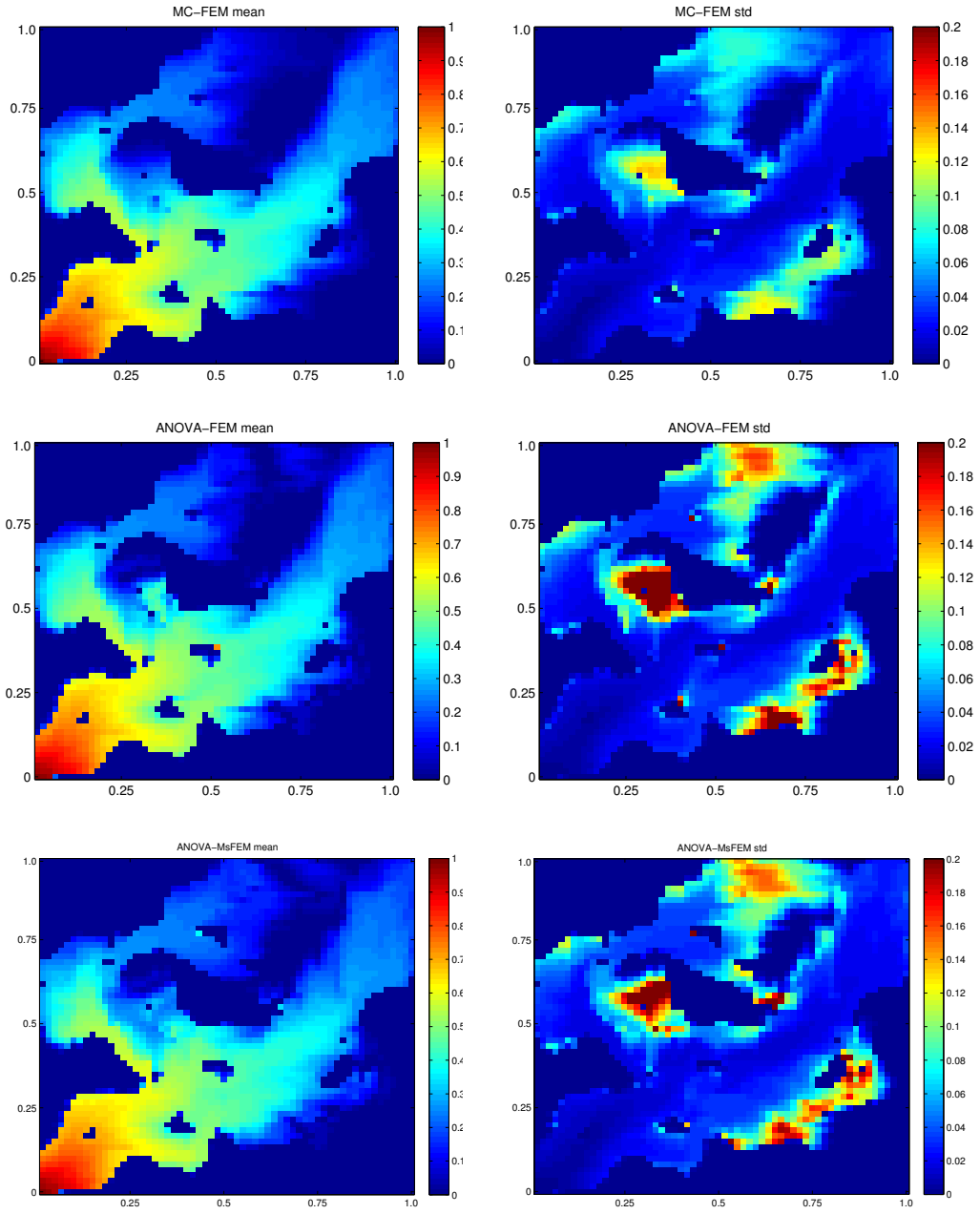

Fig. 5.5. Mean of saturation S and standard deviation of saturation S at 0.2PVI. Left column: mean; Right column: standard deviation. Top row: MC-FEM; Middle row: ANOVA-FEM; Bottom row: ANOVA-MsFEM.

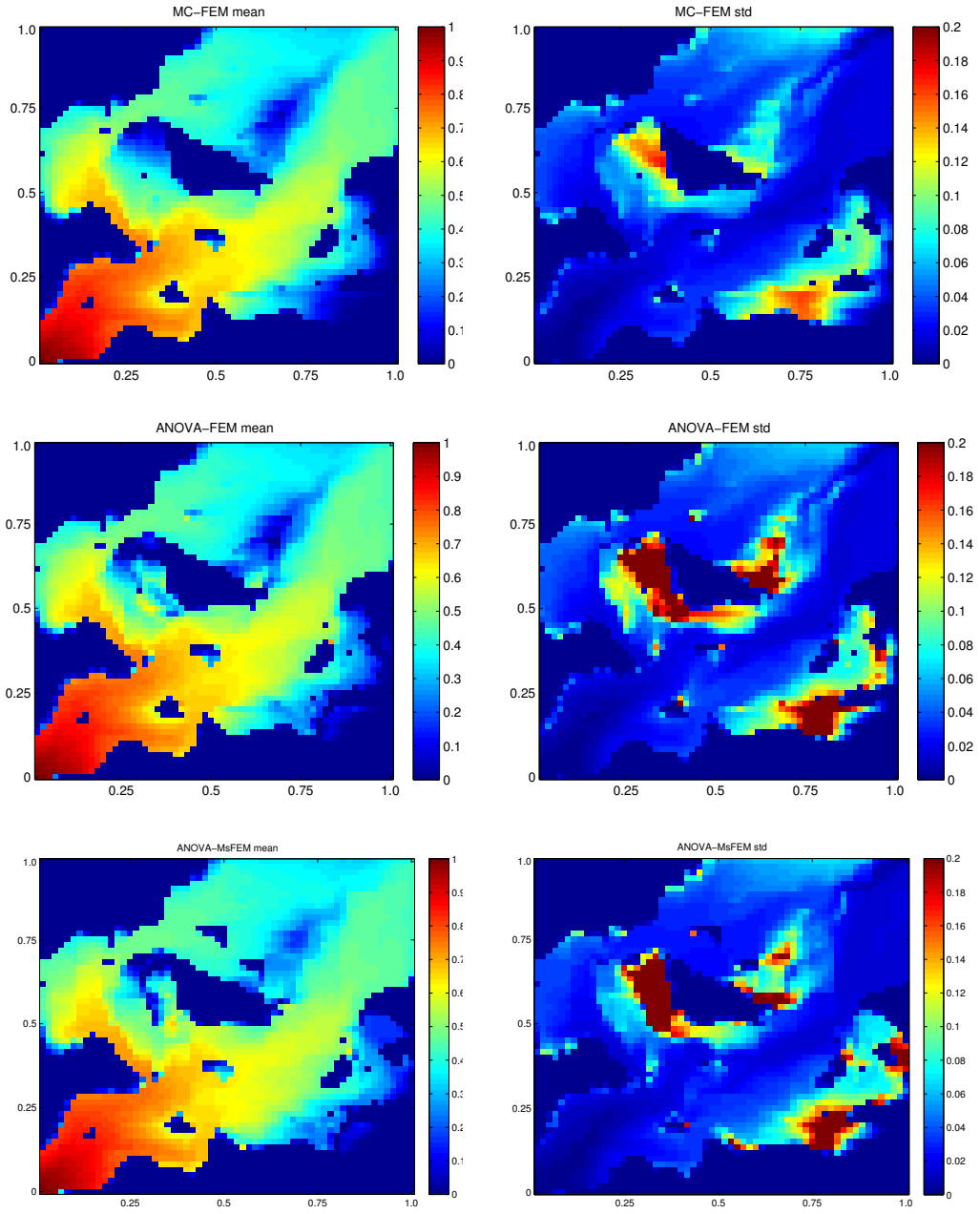

Fig. 5.6. Mean of saturation S and standard deviation of saturation S at 0.6PVI. Left column: mean; Right column: standard deviation. Top row: MC-FEM; Middle row: ANOVA-FEM; Bottom row: ANOVA-MsFEM.

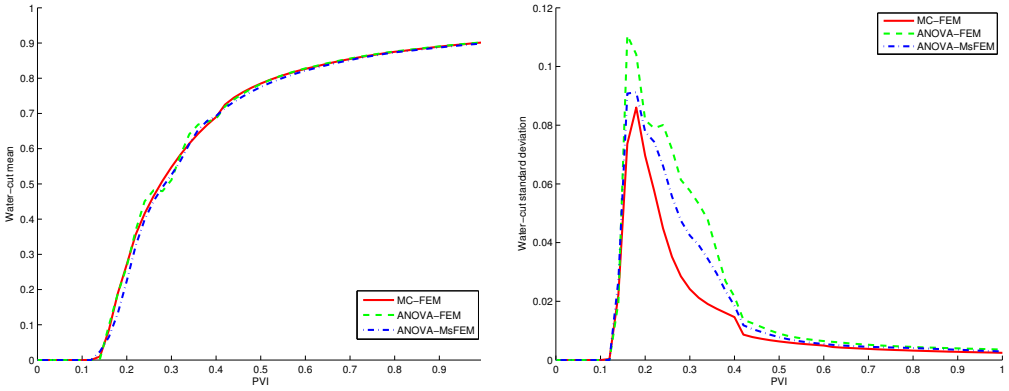

Fig. 5.7. Comparison of mean and variance of water-cut F.

5.4.3. Adaptive ANOVA example

In this simulation, we focus on analysis of the adaptive ANOVA based on the variance decomposition we proposed in Section 5.3.3. Before we show the numerical results, we would like to state several observations and supportive numerical results.

First, when the parameter space is represented through a K-L expansion type expansion (5.13), the active dimensions are determined by the corresponding eigenvalues λ_i. If λ_i decreases to 0 fast, then the dimensions corresponding to small eigenvalues are not important. Second, the adaptive Criteria 1 and 2 (see Eqs. (5.5) and (5.6)) are to some extent comparable [46, 107]. We can find examples that the active dimensions are the same for both criteria. But for special functions, for example water-cut F, Criterion 2 based on variance usually gives less number of active dimensions. Third, when there are no dominant eigenvalues, the important dimensions are usually different for different functions. To obtain a better approximation through any kind of adaptive ANOVA methods, the criterion has to be applied to the functions of interest.

To see that the active dimensions are decided by the corresponding eigenvalues λ_i, we take $l_1 = 0.25$, $l_2 = 0.1$ and $\sigma^2 = 1$ in (5.12), $E[a] = 0$ and $\alpha = 1$ in (5.13). We

Table 5.7. Comparison of adaptive ANOVA for saturation S at 0.2PVI.

	p	# of active dim	relative error of mean	relative error of std
Criterion 1	0.90	21	9.314872e-02	5.618844e-02
Criterion 2	0.90	22	8.425106e-02	4.924464e-02

truncate the K-L approximation to be $N = 50$ terms. Since the correlation lengths in horizontal and vertical directions are larger than previous examples, the covariance function is more smooth and the eigenvalues decay fast. The fine grid is 80×80 and the coarse grid is 8×8. Table 5.7 shows that when $p = 0.9$, for saturation S at 0.2PVI, Criteria 1 and 2 give almost the same active dimensions. In fact, the active dimensions are the same when $p = 0.85,\ 0.9$. And the important dimensions are the dimensions corresponding to the largest 21 or 22 eigenvalues. While if we consider water-cut function w, Criteria 1 and 2 give 17 and 10 active dimensions respectively.

Now we switch back to the example with $l_1 = 0.2$, $l_2 = 0.05$, and $\sigma^2 = 1$ in (5.12), $E[a] = 0$ and $\alpha = 1$ in (5.13), with 20 as the random dimension of the problem. In this case, the 20 eigenvalues have the same magnitude to some extent. Since Criteria 1 and 2 are comparable, and our adaptive criterion is based on variance decomposition, we make comparison only to Criterion 2, which is also based on variance.

We take into account the saturation S at 0.2PVI. To apply variance decomposition-based analysis, we use level 1 Smolyak sparse grid collocation points to build variance decomposition (5.10). 41 points in level 1 are used here. We can get variance up to first order, i.e., $\upsilon = \sum_{1 \le i \le 20} \upsilon_i$. Then each dimension corresponds to one υ_i. The same variance analysis through response surface can be done to the Monte Carlo results based on 10^4 different samples. At the same time, the anchored-ANOVA discussed in previous sections gives a variance for each dimension We treat the variance

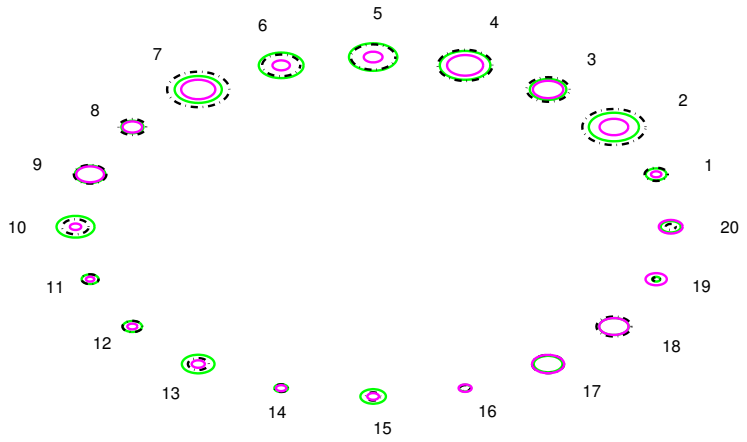

Fig. 5.8. Net graph of importance of dimensions. The radius is proportional to the magnitude of variance. Magenta: Monte Carlo results; Green: adaptive ANOVA approximation (Criterion 2); Black: variance decomposition-based adaptive ANOVA (Criterion 3).

computed by Monte Carlo as the true value as before, and compare these three sets of variance. A net graph plot is adopted to show the results. As in Figure 5.8, the 20 dimensions are labeled, with the radius proportional to the magnitude of variance respectively.

Both adaptive ANOVA and variance decomposition-based adaptive ANOVA results are different from the true ones with regard to matching radius. To further explore the behavior of the two methods, we make two different comparisons. The advantage of our proposed variance decomposition-based adaptive method is that the active dimensions can be found before ANOVA approximation by only a small amount of computations, while the adaptive ANOVA can only find the important dimensions after finishing the computation of all first-order terms.

Table 5.8. A "fair" comparison of adaptive ANOVAs up to first-order terms for saturation S at 0.2PVI.

	p	# of active dim	relative error of mean	relative error of std
Criterion 2	0.90	16	4.489370e-02	3.101351e-01
Criterion 3	0.90	15	4.519126e-02	3.079101e-01

To make a "fair" comparison, we assume that adaptive ANOVA is expanded to first-order terms in these two methods. In the first case, the active dimensions are chosen by Criterion 2, and the mean and variance of saturation S at 0.2PVI are formed by zero-order terms and first-order terms of active dimensions. In the second case, variance decomposition is conducted to choose important dimensions (Criterion 3 Eq. (5.11)), then the adaptive ANOVA includes zero-order and first-order term of important dimensions are considered. The results are in Table 5.8. When $p = 0.90$, Criterion 3 gives one active dimension less than Criterion 2. These dimensions are not exactly the same. For Criterion 2, dimensions $11, 14, 16, 19$ are not included, and for Criterion 3, dimensions $14, 15, 16, 19, 20$ are not included. Criterion 3 gives results as good as Criterion 2 as we can see in Table 5.8. The advantages of our proposed method will be more obvious, when the dimension of the problem is higher. In that case, more computations for the first-order terms can be saved.

Further, we make an "unfair" comparison. For Criterion 2, we keep all the first-order terms, because they have been obtained, and compute second-order terms between active dimensions. For Criterion 3, we only use the first-order terms of pre-fixed active dimensions and compute second-order terms based on these active dimensions. This is "unfair" because Criterion 3 does not include all first-order information. Since there are less first-order terms included in Criterion 3, the results

Table 5.9. An "unfair" comparison of adaptive ANOVAs up to second-order terms for saturation S at 0.2PVI.

	p	# of active dim	relative error of mean	relative error of std
Criterion 2	0.90	16	3.054183e-002	3.959672e-002
Criterion 3	0.90	15	2.065996e-002	7.027969e-002

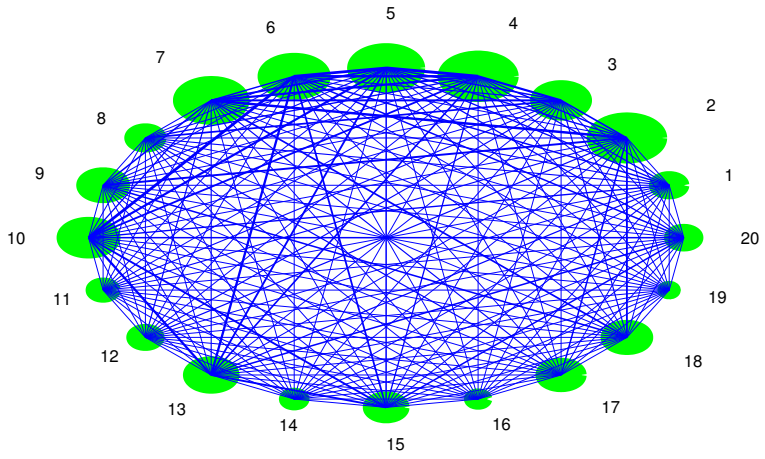

Fig. 5.9. Net graph of importance of dimensions and the interaction between pairs.

are not as good as Criterion 2. But from Table 5.9, we can see the difference between these two methods are comparable.

In fact, if we use level 2 sparse grid collocation points, the variance function can be approximated by higher-order terms in (5.10). For example, we take 841 collocation points in the random space and compute second-order in (5.10). Then the information between dimensions can be obtained before ANOVA expansion. In Figure 5.9, the radius of circles corresponding to each dimension depicts the variance

values associating with certain dimension, and the width of the lines between any pair of dimensions depicts the correlation between that pair. The thicker the line is, the larger the correlation is. By similar criteria to (5.7), the important interactions between pairs of dimensions can be found. Further, there is a chance to avoid the mistake of deleting dimensions which have small variances associated with them, but large correlation with other dimensions. This requires more computations. In our numerical example here, it is as expensive as to compute an ANOVA, so we do not use this method. When the dimension gets larger, this provides a direction for further adaptive ANOVA methods.

CHAPTER VI

ADVANCED SAMPLING METHODS USING STOCHASTIC APPROXIMATION MONTE CARLO

Markov Chain Monte Carlo (MCMC) methods are used in sampling posterior distributions. Based on the basic Metropolis-Hastings (MH) algorithm, many variations are developed to overcome the local-trap problem of the original algorithm. For example, auxiliary variable, population-based and dynamic weighting methods [56] have been developed for different kinds of situations. The other difficulty of MCMC is that the complexity of forward models usually prevents a sufficient number of runs required by the algorithm. We propose an advanced MCMC method combining Stochastic Approximation Monte Carlo (SAMC) [18, 55] and multi-stage MCMC [29, 34]. The new method adopts the merits of both algorithms. The global optimization can be found faster with reduced computational cost.

6.1. Introduction

As we mentioned in the previous chapters, MCMC is a powerful uncertainty quantification algorithm under the Bayesian framework. However, there are two main disadvantages of MCMC. First, the acceptance rate is usually low, so a large number of iterations are needed for convergence. Second, the MCMC chain can get trapped in local minimum, so that again a long chain is required for correctly sampling the desired posterior distribution. For the uncertainty quantification in subsurface problems, both disadvantages lead to the same difficulty. Since high computational cost is required in solving the coupled nonlinear PDE system (2.1)-(2.3) on the fine-grid in order to compute F_k (see (2.4)) in the target distribution $\pi(k)$, a large amount of

computational time is spent to get enough number of samples to overcome local-trap and converge to the steady state. Because of the low acceptance rate, these expensive computations are mostly spent on rejected samples.

As we discussed in Section 3.4, the standard MCMC method can be improved by adapting the proposal distribution $q(k|k_n)$ to the target distribution using a coarse-scale model [29, 34]. The process essentially modifies the proposal distribution $q(k|k_n)$ by incorporating the coarse-scale information. The multi-stage MCMC utilizes mixed MsFEM to screen the proposals. Better proposals lead to high acceptance rate in the fine-stage MCMC, which requires less forward computations on the fine model. Since most computational cost comes from forward runs in Bayesian computations, the whole process can be largely speeded up. The choice of σ_f and σ_c is important in determining the final acceptance rate of multistage MCMC as discussed in Section 3.4, i.e., the optimal value of σ_c can be estimated based on the correlation between $\|F_{obs} - F_k\|$ and $\|F_{obs} - F_k^*\|$. More quantitative discussion about the computational saving can be found in [29, 34].

On the other hand, Stochastic Approximate Monte Carlo (SAMC) [55] has the ability to find the optimization fast by including an automatic mechanism to learn from the previous accepted samples. This learning process is completed by changing the target distribution dynamically. A dynamic weight vector is introduced in the algorithm for this purpose. The target distribution function is redefined at each iteration with the biased weights associated with different subregions, so that the acceptance rate can be biased to the part of sample space with less samples. In this case, samples throughout the whole space can be obtained at the end. The Double Annealing Stochastic Approximate Monte Carlo (DASAMC) [18] is based on SAMC and Annealing SAMC (ASAMC) [54, 55, 57]. The annealing processes in both temperature and spatial serve as composition to the mechanism of SAMC, so that

the global optimization can be found faster while SAMC prevents trapping into local optimizations.

Our goal here is to combine these two methods to be the multi-stage SAMC and multi-stage DASAMC algorithms. The chapter is arranged as follows. First, we state SAMC and DASAMC algorithms. Then we propose the multi-stage SAMC and the multi-stage DASAMC. At the end, we present numerical examples of SAMC and its multi-stage version for integration problems, multimode case and the uncertainty quantification problems.

6.2. SAMC and DASAMC

In this section, we introduce the algorithms of Stochastic Approximate Monte Carlo (SAMC) and Double Annealing Stochastic Approximate Monte Carlo (DASAMC) algorithm.

As we mentioned throughout the thesis, our goal is to find the permeability fields given dynamic production information $F_{obs}(t)$. Under Bayesian framework, we achieve this by sampling from the conditional distribution $\pi(k) := \pi(k|F_{obs})$, i.e.,

$$\pi(k) \propto L(F_{obs}|k)\pi_0(k),$$

where $L(F_{obs}|k)$ represents the likelihood function and requires the forward solution of flow and transport. As in the previous chapters, we assume that the combined errors from the measurement, modeling and numeric satisfy a Gaussian distribution. That is, the likelihood function $L(F_{obs}|k)$ takes the form

$$L(F_{obs}|k) \propto \exp\Big(-\frac{||F_{obs} - F_k||^2}{\sigma_f^2}\Big),$$

where F_k is the fractional flow computed by solving model equations on the fine-grid

for a given k, and σ_f is the precision associated with the measurement F_{obs} and the numerical solution F_k. Then the posterior distribution is

$$\pi(k) \propto \exp\Big(-\frac{||F_{obs} - F_k||^2}{\sigma_f^2}\Big)\pi_0(k). \tag{6.1}$$

In practice, we often take the prior distribution $\pi_0(k)$ to be either non-informative or Gaussian distribution with given parameters, or we can take log to the prior distribution, then our target distribution $\pi(k)$ in (6.1) can be written as

$$\pi(k) \propto \exp\{-U(k)\}, \quad k \in \mathcal{K}, \tag{6.2}$$

where $U(k) = \|F_{obs} - F_k\|^2/\sigma_f^2$ when $\pi_0(k)$ is non-informative, or corresponding forms in other cases. $U(k)$ is called the energy function. $\mathcal{K}$ is the sample space. Now the problem is to sample from the distribution (6.2),

SAMC and DASAMC have similar structures; with DASAMC can be obtained from changing certain parts of SAMC framework. We start with introducing SAMC method.

To achieve the goal of avoiding local-trap problem, SAMC algorithm first partitions the energy space into subspaces. We denote the subregions by $E_1, E_2, \cdots, E_m$, where $E_1 = \{k : U(k) \le u_1\}$, $\cdots$, $E_i = \{k : u_{i-1} < U(k) \le u_i\}$, $\cdots$, $E_m = \{k : U(k) > u_{m-1}\}$; $u_1 < u_2 < \cdots < u_{m-1}$ are numbers specified by the user. Let

$$\varrho_i = \log(\int_{E_i} \exp\{-U(k)\}),$$

then (6.2) becomes

$$\pi(k) \propto \sum_{i=1}^{m} \frac{\exp\{-U(k)\}}{e^{\varrho_i}} I(k \in E_i). \tag{6.3}$$

SAMC seeks to draw samples from each of the subregions with a pre-specified frequency by defining the stationary distribution dynamically at different iterations.

Let k_n be the sample at n^{th} iteration, then the target distribution at this iteration is

$$\pi_n(k) \propto \sum_{i=1}^{m} \frac{\exp\{-U(k)\}}{e^{\varrho_i^n}} I(k \in E_i).$$

We denote $\kappa(u)$ to be the index of the subregion that a sample x with energy u belongs to. For instance, if $x \in E_j$, then $\kappa(U(x)) = j$. Let $\wp = (\wp_1, \cdots, \wp_m)$ be an m-vector with $0 < \wp_i < 1$ and $\sum_{i=1}^{m} \wp_i = 1$, which defines a desired sampling frequency for the subregions. $\wp$ is called the desired sampling distribution. Define $H(\varrho^n, x_{n+1}) = e^{n+1} - \wp$, where $e^{n+1} = (e_1^{n+1}, \cdots, e_m^{n+1})$ and $e_i^{n+1} = 1$ of $k_{n+1} \in E_i$ and 0 otherwise. Let $\{\iota_n\}$ be a positive non-decreasing sequence satisfying the conditions,

$$(i) \ \sum_{n=0}^{\infty} \iota_n = \infty, \quad (ii) \ \sum_{n=0}^{\infty} \iota_n^{\delta} < \infty,$$

for some $\delta \in (1, 2)$. We set

$$\iota_n = \left(\frac{n_0}{\max(n_0, n)} \right)^{\eta},$$

following [55], for some specified values of $n_0 > 0$ and $\eta \in (0.5, 1]$. The constants or constant vectors, such as $m, n_0, \eta, \wp$, are important to the success of this algorithm. Details about how to choose these parameters are discussed in [55]. In order to force the samples traveling around the sample space, $\wp$ is usually taken to be with equal subelements, i.e., $\wp_i = 1/m$ for $i = 1, \cdots, m$. A large number of m is often needed for complicated problems. A large value of n_0 will allow the sampler to reach all subregions fast. With all the notations we introduced, the SAMC algorithm can be describe as follow.

Algorithm (SAMC) Suppose at the n^{th} step, we have k_n.

Step 1. At k_n, generate k_{n+1}.

(1). Generate k according to the proposal distribution $q(k|k_n)$.

(2). Calculate the ratio

$$\gamma(k_n, k) = \exp(\varrho^n_{\kappa(U(k_n))} - \varrho^n_{\kappa(U(k))}) \exp\{-(U(k_n) - U(k))\} \frac{q(k_n|k)}{q(k|k_n)}.$$

(3). Accept the proposal with probability $\min(1, \gamma)$. If it is accepted, set $k_{n+1} = k$; otherwise, set $k_{n+1} = k_n$.

Step 2. Set $\varrho^* = \varrho^n + \iota_n H(\varrho^n, k_{n+1})$, where ι_n is called the gain factor.

Step 3. If $\varrho^* \in \Xi$, set $\varrho^{n+1} = \varrho^*$; otherwise, set $\varrho^{n+1} = \varrho^* + c^*$, where c^* is chosen that $\varrho^* + c^* \in \Xi$. Here Ξ is a vector space. □

The key point of SAMC algorithm is the dynamically changed weights ϱ associated with subregions. With these weights, the target distribution π_n in the n^{th} iteration can be biased to the regions with fewer samples in the chain. In this situation, the possibility for the sampler to travel the whole sample space is high in relatively short chains. And then it has better behavior in elimination local-trap situations.

The differences between DASAMC and SAMC are the annealing procedures in both spatial and temperature. There is a temperature changing over iterations in DASAMC, so that the global optimization can be found faster; while the spatial annealing process restricts the searching of optimization to be around current position, instead of the whole spatial space as in SAMC. This also helps the SAMC algorithm find global minimum with less number of iterations. The double annealing part severs as a composite for SAMC, so that the optimization can be found without wondering around unimportant subregions. The trade-off between these two processes is problem dependent.

Next, we state the DASAMC algorithm. Let $\mathcal{K}_n$ denote the sample space at iteration n. DASAMC first searches in the entire sample space $\mathcal{K}_0 = \cup_{i=1}^m E_i$, and

iteratively searches in the set

$$\mathcal{K}_n = \bigcup_{i=1}^{\kappa(U^n_{min}+\mathcal{N})} E_i, \quad t = 1, 2, \cdots,$$

where U^n_{min} is the best energy value obtained until iteration n, and $\mathcal{N} > 0$ is a user specified parameter which determines the broadness of the sample space at each iteration.

Algorithm (DASAMC) Suppose at the n^{th} step, we have k_n.

Step 1. At k_n, generate k_{n+1}.

(1). Generate k according to the proposal distribution $q(k|k_n)$.

(2). Calculate the ratio

$$\gamma(k_n, k) = \exp(\varrho^n_{\bar{\kappa}(U(k_n))} - \varrho^n_{\bar{\kappa}(U(k))}) \exp\{-(U(k_n) - U(k))/T_n\} \frac{q(k_n|k)}{q(k|k_n)},$$

where $T_n = T_0\rho^{n-1}$ is the cooling temperature. T_0 is a user defined constant. ρ is a preselected decay rate and

$$\bar{\kappa}(U(k)) = \begin{cases} \kappa(U(k)), & \text{if } U(k) \in \mathcal{K}_n, \\ 0, & \text{otherwise .} \end{cases}$$

(3). Accept the proposal with probability $\min(1, \gamma)$. If it is accepted, set $k_{n+1} = k$; otherwise, set $k_{n+1} = k_n$.

Step 2. Set $\varrho^* = \varrho^n + \iota_n(I(k_{n+1} \in E_i) - \wp_i)$, where $i = 1, \cdots, \kappa(U^n_{min} + \mathcal{N})$.

Step 3. If $\varrho^* \in \Xi$, set $\varrho^{n+1} = \varrho^*$; otherwise, set $\varrho^{n+1} = \varrho^* + c^*$, where c^* is chosen that $\varrho^* + c^* \in \Xi$. □

Figure 6.1 illustrate the ideas of SAMC and DASAMC methods. The key points of DASAMC algorithm are the dynamic weights $\exp\{\theta^n_{\kappa(U(k))}\}$, spatial annealing parameter $\kappa(U^n_{min} + \mathcal{N})$ and temperature annealing parameter T_n. Similar to SAMC, the dynamic weights automatically lift the low energy parts of the energy function

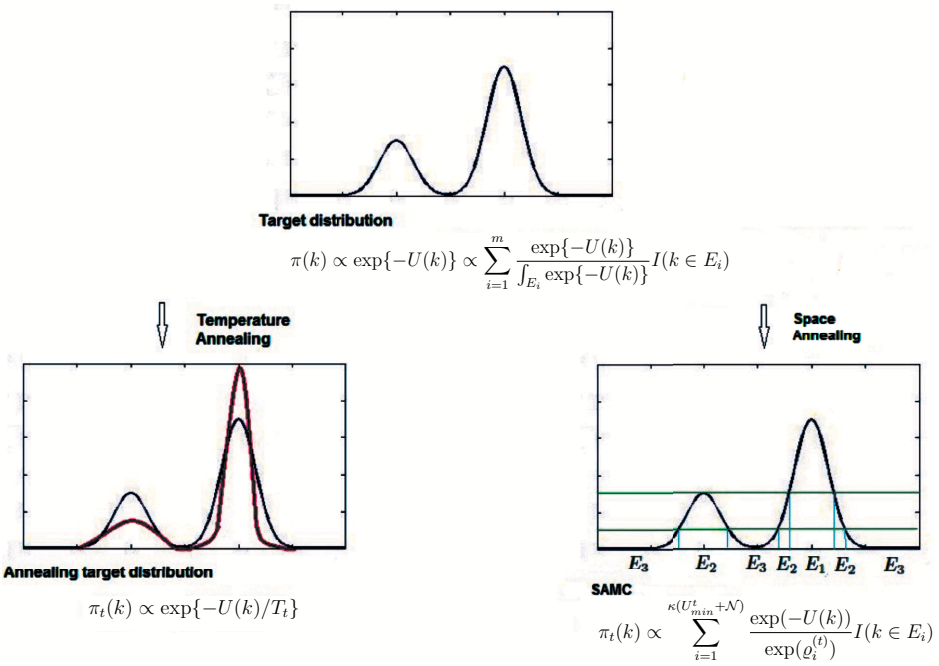

Fig. 6.1. Illustration of SAMC and DASAMC algorithms.

$U(k)$, so that the accepted samples can travel over the whole sample space. At the same time, the spatial annealing parameter restricts the search of the lowest energy to be in the neighborhood around current U^n_{min}, so that the convergence to the optimal can be faster. Our choice of the temperature annealing parameter T_n also helps in finding optimal more efficiently. $T_0\rho^{n-1}$ specifies a desired decreasing rate of the energy function. At the same time, by taking maximum of $T_0\rho^{n-1}$ and U^n_{min}, the temperature annealing rate is controlled by current best energy as a bound, so that samples will be accepted with a reasonable rate. When $\rho = 1$, then the algorithm becomes ASAMC with a constant temperature annealing. If $\rho = 1$ and $T_0 = 1$, then it becomes just SAMC algorithm.

As [18] points out that samples collected from DASAMC algorithm are highly

biased towards the mode. So the samples cannot be directly used to calculate function values, for example, moments. When the mean and variance, or other functions of the posterior distribution are needed, we follow the estimation way proposed in [18]. The important sampling idea is used here. Suppose that $E[h(k)] = \int h(k)\pi(k)dk$ is of interest here, and samples are draw from the proposal distribution $q(k)$, then

$$E[h(k)] = \int h(k)\frac{\pi(k)}{q(k)}q(k)dk \approx \frac{\sum h(k_n)\pi(k_n)/q(k_n)}{\sum \pi(k_n)/q(k_n)},$$

where k_n are proposed sample at n^{th} iteration. This also provides a way to get the unbiased posterior distribution.

6.3. Multi-stage SAMC and multi-stage DASAMC

As we mentioned in Chapter II and the introduction of this chapter, the multi-stage MCMC method can reduce the computational cost of forward simulations by screening the bad proposals out through an approximation of the desired posterior distribution. Next, we combine the multi-stage method with SAMC and DASAMC to develop new sampling schemes for the uncertainty quantification problems in porous media flow problems.

The target distribution of interest is

$$\pi(k) \propto \exp(-\frac{||F_{obs} - F_k||^2}{\sigma_f^2})\pi_0(k), \tag{6.4}$$

where F_k is the fractional flow computed by solving the system equations (2.1)-(2.3) on fine-scale grid. σ_f is the precision associated with the fine model. We approximate the target posterior distribution $\pi(k)$ on the coarse-scale by

$$\pi^*(k) \propto \exp(-\frac{||F_{obs} - F_k^*||^2}{\sigma_c^2})\pi_0(k),$$

where F_k^* is the fractional flow computed by solving the coarse-scale model for give k, and σ_c^2 which is precision associated with the coarse-scale model.

Suppose $E_1, \cdots, E_m$ are m disjoint regions that form a partition of sample space, where $E_i = \{k : u_{i-1} \leq \frac{||F_{obs} - F_k^*||^2}{\sigma_c^2} \leq c_{i+1}\}$. The weights ϱ_n will be updated only when the proposed sample has been accepted by both stages. Without loss of generality, we state here the two-stage SAMC algorithm, while the multi-stage version can be easily generalized.

Algorithm (Two-stage SAMC)

Coarse - stage

Step 1. At k_n generate $\tilde{k}$ by a proposal distribution $q(\tilde{k}|k_n)$.

Step 2. Accept $\tilde{k}$ with probability

$$\gamma_c(k_n, \tilde{k}) = \min\left(1, \frac{\pi_n^*(\tilde{k})q(k_n|\tilde{k})}{\pi_n^*(k_n)q(\tilde{k}|k_n)}\right) = \min\left(1, \frac{\pi_n^*(\tilde{k})}{\pi_n^*(k_n)}\right),$$

i.e., $k = \tilde{k}$ with probability $\gamma_c(k_n, \tilde{k})$, and $k = k_n$ with probability $1 - \gamma_c(k_n, \tilde{k})$. The last equality holds when the proposal distribution q is symmetric. If rejected go to step 1.

Therefore, the proposal k for fine-stage is generated from the effective instrumental distribution

$$Q(k|k_n) = \gamma_c(k_n, k)q(k|k_n) + \Big(1 - \int \gamma_c(k_n, k)q(k|k_n)dk\Big)\delta_{k_n}(k).$$

Here,

$$\pi_n^*(\tilde{k}) = \sum_{i=1}^{m} \frac{\pi^*(\tilde{k})}{e^{\varrho_i^n}} I(\tilde{k} \in E_i), \quad \pi_n^*(k_n) = \sum_{i=1}^{m} \frac{\pi^*(k_n)}{e^{\varrho_i^n}} I(k_n \in E_i).$$

So

$$\gamma_c(k_n,\tilde{k}) \;=\; \min\left(1,\frac{\pi_n^*(\tilde{k})}{\pi_n^*(k_n)}\right)=\min\left(1,\frac{\sum_{i=1}^m \frac{\pi^*(\tilde{k})}{e^{\varrho_i^n}}I(\tilde{k}\in E_i)}{\sum_{i=1}^m \frac{\pi^*(k_n)}{e^{\varrho_i^n}}I(k_n\in E_i)}\right).$$

Fine - stage

Step 3. Accept k with probability

$$\gamma_f(k_n,k)=\min\left(1,\frac{\pi_n(k)Q(k_n|k)}{\pi_n(k_n)Q(k|k_n)}\right),$$

i.e., $k_{n+1}=k$ with probability $\gamma_f(k_n,k)$, and $k_{n+1}=k_n$ with probability $1-\gamma_f(k_n,k)$. If rejected go to step 1.

Using the argument as in [34], the acceptance probability can be simplified as

$$\begin{aligned}\gamma_f(k_n,k) &= \min\left(1,\frac{\pi_n(k)\pi_n^*(k_n)}{\pi_n(k_n)\pi_n^*(k)}\right)\\ &= \min\left(1,\frac{\sum_{i=1}^m \frac{\pi(k)}{e^{\varrho_i^n}}I(k\in E_i)}{\sum_{i=1}^m \frac{\pi(k_n)}{e^{\varrho_i^n}}I(k_n\in E_i)}\frac{\sum_{i=1}^m \frac{\pi^*(k_n)}{e^{\varrho_i^n}}I(k_n\in E_i)}{\sum_{i=1}^m \frac{\pi^*(k)}{e^{\varrho_i^n}}I(k\in E_i)}\right),\end{aligned}$$

where

$$\pi_n(k)=\sum_{i=1}^m \frac{\pi(k)}{e^{\varrho_i^n}}I(k\in E_i)\quad \pi_n(k_n)=\sum_{i=1}^m \frac{\pi(k_n)}{e^{\varrho_i^n}}I(k_n^f\in E_i).$$

Step 4. Set $\varrho^*=\varrho^n+\varrho_n H(\varrho^n,k_{n+1})$. If $\varrho^*\in\Xi$, set $\varrho^{n+1}=\varrho^*$; otherwise, set $\varrho^{n+1}=\varrho^*+c^*$, where c^* is chosen that $\varrho^*+c^*\in\Xi$. Here Ξ is a vector space. □

The multi-stage DASAMC algorithm has similar form as multistage SAMC algorithm except that there are both spatial and temperature annealing processes. We briefly describe the two-stage DASAMC as follows. The weights and temperature are also only updated when the sample gets accepted in both stages.

Algorithm (Two-stage DASAMC)

Coarse - stage

Step 1. At k_n generate $\tilde{k}$ by a proposal distribution $q(\tilde{k}|k_n)$.

Step 2. Accept $\tilde{k}$ with probability

$$\gamma_c(k_n, \tilde{k}) = \min\left(1, \frac{\pi_n^*(\tilde{k})q(k_n|\tilde{k})}{\pi_n^*(k_n)q(\tilde{k}|k_n)}\right),$$

with

$$\pi_n^*(k) = \sum_{i=1}^{\bar{\kappa}(U_{min}^n+\mathcal{N})} \frac{[\pi^*(k)]^{1/T_n}}{e^{\varrho_i^n}} I(k \in E_i).$$

If rejected go to step 1. Therefore, the proposal k for fine-stage is generated from the effective instrumental distribution

$$Q(k|k_n) = \gamma_c(k_n, k)q(k|k_n) + \Big(1 - \int \gamma_c(k_n, k)q(k|k_n)dk\Big)\delta_{k_n}(k).$$

Fine - stage

Step 3. Accept k with probability

$$\gamma_f(k_n, k) = \min\left(1, \frac{\pi_n(k)Q(k_n|k)}{\pi_n(k_n)Q(k|k_n)}\right) = \min\left(1, \frac{\pi_n(k)\pi_n^*(k_n)}{\pi_n(k_n)\pi_n^*(k)}\right),$$

with

$$\pi_n(k) = \sum_{i=1}^{\bar{\kappa}(U_{min}^n+\mathcal{N})} \frac{[\pi(k)]^{1/T_n}}{e^{\varrho_i^n}} I(k \in E_i),$$

where $T_n = T_0\rho^{n-1}$. If rejected go to step 1.

Step 4. Set $\varrho^* = \varrho^n + \varrho_n H(\varrho^n, k_{n+1})$. If $\varrho^* \in \Xi$, set $\varrho^{n+1} = \varrho^*$; otherwise, set $\varrho^{n+1} = \varrho^* + c^*$, where c^* is chosen that $\varrho^* + c^* \in \Xi$. □

6.4. Numerical results

In this section, we design several numerical examples to show the mechanism of SAMC and the uncertainty quantification for reservoir model calibration using multistage SAMC and DASAMC.

6.4.1. Comparison of SAMC and MCMC

To see the differences between SAMC and MCMC through some numerical results, we consider a horizontal layered reservoir in $[0,1] \times [0,1]$. We assume the permeability field has the correlation function as

$$R(y_1, y_2) = \sigma^2 \exp\Big(-\frac{|y_1 - y_2|}{l_2}\Big). \tag{6.5}$$

This is a 1-d exponential correlation function. Similarly, the 1-d Gaussian correlation function can be obtained as a simplification of the 2-d Gaussian correlation function (3.11). In our numerical experiment, we take $l_2 = 0.2$ and $\sigma^2 = 2$. To get the K-L expansion for this random field, the correlation function (6.5) is discretized in a grid with 80 subgrids in vertical direction. The K-L expansion is truncated to be with 20 terms. The random part of K-L expansion is assume to be independent Gaussian $N(0,1)$. The posterior distribution of interest is (6.4) with $\sigma_f^2 = 0.5$, F_{obs} and F_k are water-cut functions. The proposal distribution we use here is a random walk with step size 0.3.

For SAMC method, the energy space needs to be partitioned. We take the energy function to be $U(k) = \|F_{obs} - F_k\|^2/\sigma_f^2$ and divide the energy space to be 6 equal subregions. Each subregion has equal sample frequency $\wp_i = 1/6$. The behavior of SAMC is also affected by the choice of other parameters. In this experiment we take $n_0 = 30$ and $\eta = 1$. 10^4 samples are desired for MCMC while only 10^3 samples are taken in SAMC.

To compare the sampling results from both SAMC and MCMC methods, we compute $\exp(-||F^{ref} - F_k||^2/\sigma_f^2)$ for all samples and plot the values after sorting them. The range of this function is $[0,1]$. Since $U(k)$ is divided to be 6 subregions, so is the region $[0,1]$. Therefore, by looking at the number of samples in each sub-

region of $[0, 1]$, we can see the ability of each method to travel in the sample space. With more subregions reached, the method has sampled the posterior distribution better and is less possible to be trapped in local minimum. Note when the function $\exp(-\frac{||F^{ref}-F_k||^2}{\sigma_f^2})$ is close to 1, the sampled field is close to reference one.

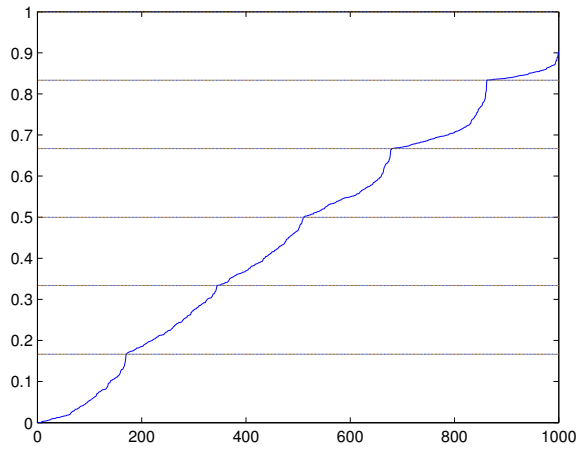

Fig. 6.2. Plot of sorted $\exp\left(-\frac{||F_{obs}-F_k||^2}{\sigma_f^2}\right)$ of SAMC with 10^3 samples.

From Figures 6.2 and 6.3, we can conclude that SAMC has better ability to travel in the sample space than MCMC for the subsurface problems we are interested in. After 10^4 iterations, MCMC could not reach 2 out of 6 subregions, while SAMC reaches all of the subregions with 10^3 samples. In the algorithm of SAMC, the acceptance probability is

$$\gamma(k_n, k) = \min\left(1, \exp(-\frac{||F_{obs} - F_k||^2}{\sigma_f^2} + \frac{||F_{obs} - F_{k_n}||^2}{\sigma_f^2})\frac{\sum_{i=1}^m \frac{1}{e^{\varrho_i^n}} I(k \in E_i)}{\sum_{i=1}^m \frac{1}{e^{\varrho_i^n}} I(k_n \in E_i)}\right).$$

Without the term $\sum_{i=1}^m \frac{1}{e^{\varrho_i^n}} I(k \in E_i) / \sum_{i=1}^m \frac{1}{e^{\varrho_i^n}} I(k_n \in E_i)$, it is just MCMC. This γ makes SAMC accepting samples coming from different subregions with respect to the previous accepted ones with high probability. Also, as the designed sampling

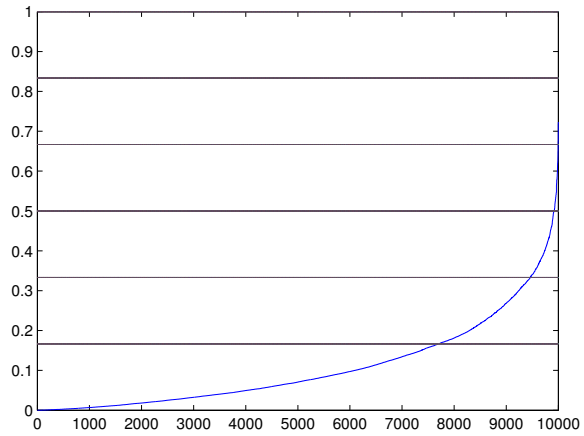

Fig. 6.3. Plot of sorted $\exp\left(-\frac{||F_{obs}-F_k||^2}{\sigma_f^2}\right)$ of MCMC with 10^4 samples.

frequency is 1/6, we can see from the Table 6.1 that the absolute errors of sampling frequency and the number of samples in the corresponding subregions at different iterations are decreasing to 0 as the iterations increase. Let $\mathfrak{n}_\mathfrak{i}$ be the number of samples in the i^{th} subregions.

6.4.2. Monte Carlo integration

In many cases, the samples collected from Bayesian methods are used to compute some kind of integration. For example the moments of some functions of random variables are often interested in. In this section, we use the samples getting from both two-stage SAMC and MCMC to compute the integration of certain functions.

We take the same example as in the previous section, i.e., the correlation function is (6.5), $l_2 = 0.2$, $\sigma^2 = 2$, $\sigma_f^2 = 0.5$ and $\sigma_c^2 = 2\sigma_f^2$. The field is discretized to be 80 in the vertical direction. There are 20 terms in the K-L expansion. Random walk with step size 0.3 is used as the proposal distribution. $n_0 = 300$, $\eta = 1$ and 6 subregions

Table 6.1. Absolute errors of sampling frequency and $\mathfrak{n}_i$ at different iterations.

	Number of iterations				
	200	400	600	800	1000
$\lvert\mathfrak{n}_1 - 1/6\rvert$	0.166667	0.166667	0.025000	0.007083	0.027667
$\lvert\mathfrak{n}_2 - 1/6\rvert$	0.041667	0.015833	0.026667	0.025417	0.015333
$\lvert\mathfrak{n}_3 - 1/6\rvert$	0.013333	0.020833	0.013333	0.002917	0.002333
$\lvert\mathfrak{n}_4 - 1/6\rvert$	0.048333	0.038333	0.013333	0.002083	0.000667
$\lvert\mathfrak{n}_5 - 1/6\rvert$	0.043333	0.035833	0.010000	0.003333	0.007333
$\lvert\mathfrak{n}_6 - 1/6\rvert$	0.103333	0.055833	0.015000	0.015833	0.003333

with equal sampling frequency 1/6 are taken for the two-stage SAMC. A chain with length 10^5 has been taken to compute the true values of integrations, while short chains of length 2000 for both MCMC and two-stage SAMC are gotten for integral computation also.

We can take the integral function of interest to be

$$\hbar(k) = \begin{cases} c_1, \text{ if } 0 \leq \exp(-\frac{||F_{obs}-F_k||^2}{\sigma_f^2}) < \frac{1}{2} \\ c_2, \text{ if } \frac{1}{2} \leq \exp(-\frac{||F_{obs}-F_k||^2}{\sigma_f^2}) < 1. \end{cases} \tag{6.6}$$

As in the previous section, we plot the sorted exponential of the energy function to see how the samples distribute in the whole sample space. From Figures 6.4, 6.5 and 6.6, we see that SAMC reaches subregions more quickly than MCMC. The MCMC travels to more subregions when the length of the chain increases to 10^5, which is usually be considered as a convergent case. If we compute the integration of $\hbar(k)$ based on these samples, the accuracy of integration can be greatly improved by using SAMC, when the integral function has large weight on those regions where

MCMC cannot achieve. The function (6.6) is this kind. Table 6.2 shows that the integration results. The results obtained in MCMC long chain case are considered as the true values. SAMC provides better results in this case.

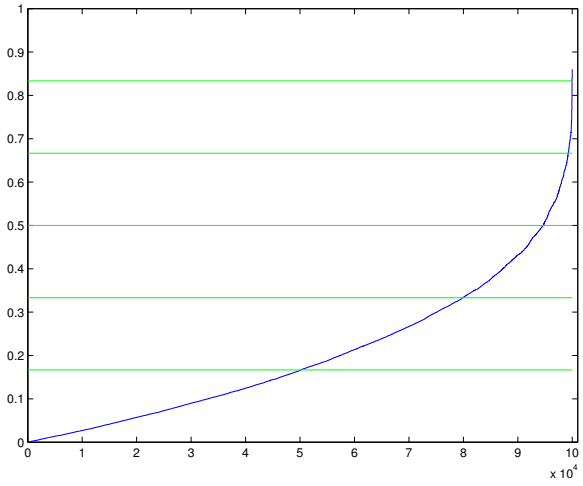

Fig. 6.4. Plot of sorted $\exp\left(-\frac{||F_{obs}-F_k||^2}{\sigma_f^2}\right)$ of MCMC with 10^5 samples.

Table 6.2. Monte Carlo integration $\int \hbar(k)d\pi(k)$ values in different cases.

c_1	c_2	MCMC long chain	MCMC	SAMC
1	5	12.1404	1.2060	3.3440
1	10	14.8159	1.4635	6.2740
1	100	62.9749	6.0985	59.0140

The examples are specially chosen here. We try to show the mechanism of the two-stage SAMC to avoid being trapped in local minimums. The samples are spread out as designed. If the integral functions do not have special properties, these samples getting from SAMC will not give any better results of integration comparing with

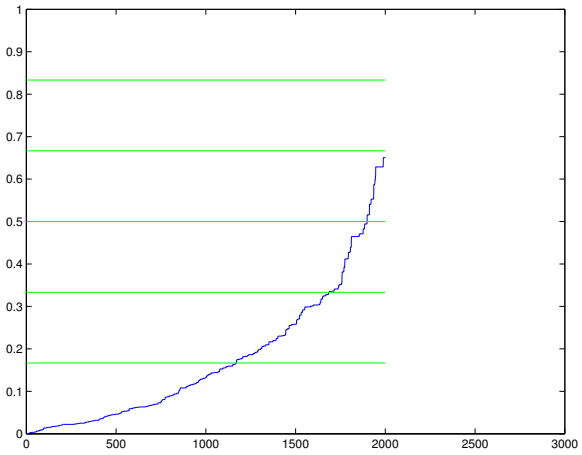

Fig. 6.5. Plot of sorted $\exp\left(-\frac{||F_{obs}-F_k||^2}{\sigma_f^2}\right)$ of MCMC with 2000 samples.

MCMC samples. It is obvious that the weight of each subregion need to be obtained before SAMC samples are used into computation of integrations, or the biased sample getting from SAMC needs to be modified first.

6.4.3. Multimode case

To further show the importance of the two-stage MCMC, we design an example where the posterior distribution is multimode. Assume that the permeability field is

$$\log(k) = k_1\psi + k_2\psi, \quad k_1, k_2 \sim N(0,1), \tag{6.7}$$

where ψ is the eigenfunction getting from K-L decomposition. We also assume that the water-cut response can be separated to be two parts, with each is the response coming from one term of the right hard side of (6.7). So the posterior is

$$\pi(k_1,k_2) \propto \exp(-\frac{||F_{obs_1} - F_{k_1}||^2}{\sigma_{f_1}^2} - \frac{||F_{obs_2} - F_{k_2}||^2}{\sigma_{f_2}^2})\pi_0(k)$$

with the water-cut F_{k_1} and F_{k_2} to be the water-cut measured.

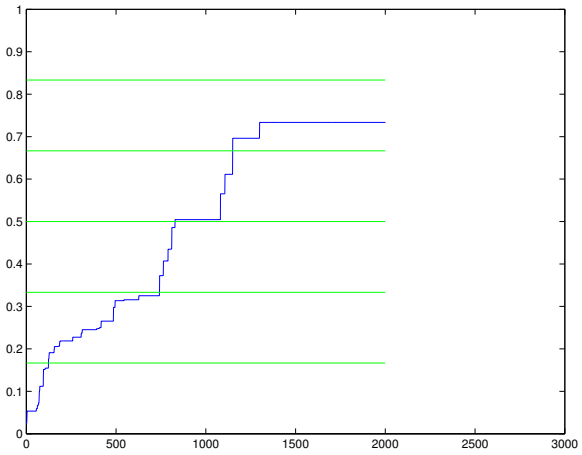

Fig. 6.6. Plot of sorted $\exp\left(-\frac{||F_{obs}-F_k||^2}{\sigma_f^2}\right)$ of SAMC with 2000 samples.

The corresponding parameters in our example are $l_2 = 0.2$, $\sigma^2 = 2$ in (6.5). $\sigma_{f_2}^2 = 0.1$ and $\sigma_{f_1}^2 = \sigma_{f_1}^2/2$. The field is discretized to be 80 in the vertical direction, with 20 terms in the K-L expansion. Random walk with step size 0.3 is used as the proposal distribution. $n_0 = 300$, $\eta = 1$ and 6 subregions with equal sampling frequency 1/6 are taken for the two-stage SAMC. Water-cut has been measured at the position $(1, 0.5)$ and $(1, 1)$. A chain of length 2000 for both MCMC and two-stage SAMC are sampled.

From Figures 6.7 and 6.8, we can see that the two-stage SAMC is able to find the two modes of samples, while MCMC cannot in the same length of chains.

6.4.4. Multi-stage DASAMC

In this section, we present representative simulation results for two-phase flow using two-stage DASAMC method. We consider a two-dimensional system in a square domain $[0, 1] \times [0, 1]$ as the previous examples, where the injection well is placed at the left bottom vertex of the square domain $(0, 0)$ and the production well is placed

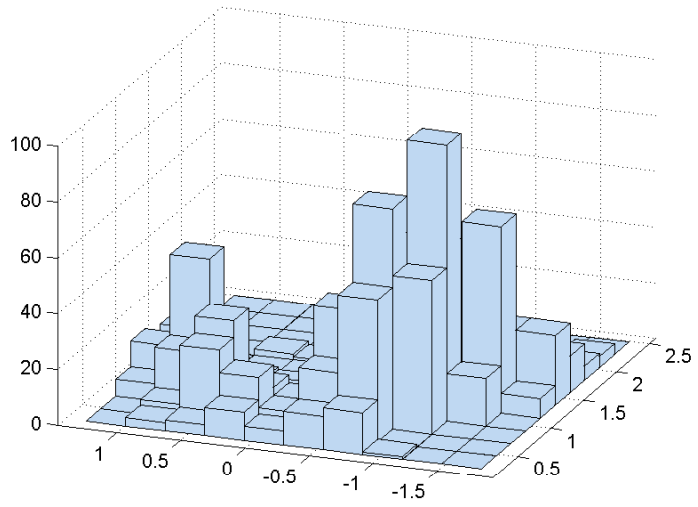

Fig. 6.7. Histogram of SAMC samples.

at the right top vertex $(1, 1)$. As for prior distribution, we assume the permeability is log-normal, namely, $\log(k)$ has Gaussian distribution with correlation function (3.11), i.e.,

$$R(x, y) = \sigma^2 \exp\Big(-\frac{|x_1 - y_1|^2}{2l_1^2} - \frac{|x_2 - y_2|^2}{2l_2^2}\Big),$$

with correlation lengths $l_1 = l_2 = 3$ and variance $\sigma^2 = 2$. The correlation lengths are taken to be large here in order to get fewer terms in K-L expansion. The fine-scale models are of dimension 60×60. Since the first 2 eigenvalues are sufficient to give less than 5 percent errors for the solution, we truncate K-L expansion to be

$$\log k(x, \theta_1, \theta_2) = \sqrt{\lambda_1}\theta_1\psi_1(x) + \sqrt{\lambda_2}\theta_2\psi_2(x). \tag{6.8}$$

This is a significant dimension reduction, since originally the permeability field has $60 \times 60 = 3600$ dimension.

We take the true permeability field to be with reference value $(\theta_1, \theta_2)_{ref} = (0.5, 1)$,

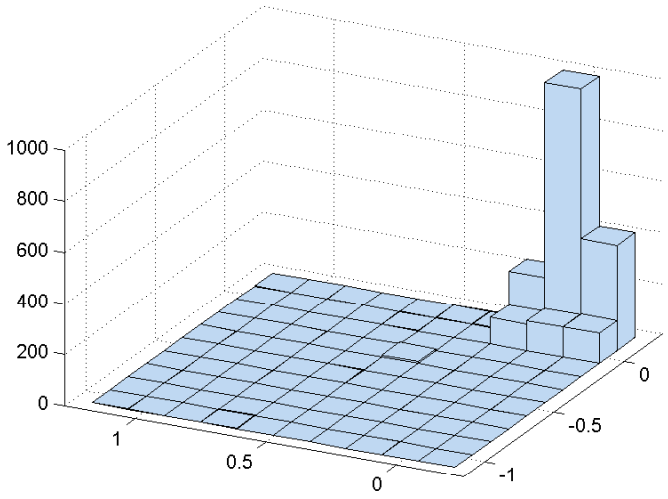

Fig. 6.8. Histogram of MCMC samples.

and initial permeability field with $(\theta_1, \theta_2)_{initial} = (1.438380, 0.325191)$ as shown in Figure 6.9. F_{obs} is computed by solving the governing equations in fine grid at reference permeability field. In order to recover $(\theta_1, \theta_2)_{ref}$ from F_{obs}, we apply MCMC, two-stage MCMC and two-stage DASAMC methods.

First, we run MCMC for 10^4 iterations on the fine grid. The likelihood function $L(F_{obs}|k)$ has $\sigma_f^2 = 0.0001$. The proposals are generated by random walk with step size 0.3. For the two-stage MCMC, we collect a chain with 500 iterations on fine-stage. $\sigma_c^2 = 0.1$ and $\sigma_f^2 = 0.0001$ on coarse step and fine step, respectively. The proposals are generated by random walk with step size 0.3 too. Then, the two-stage DASAMC is used on the coarse grid for 500 iterations, and the best approximation on this step is passed to be the initial for second DASAMC process. The settings of σ_c^2, σ_f^2 the proposal distribution are the same as two-stage MCMC. Coarse grid for both two-stage DASAMC and two-stage MCMC is 6×6.

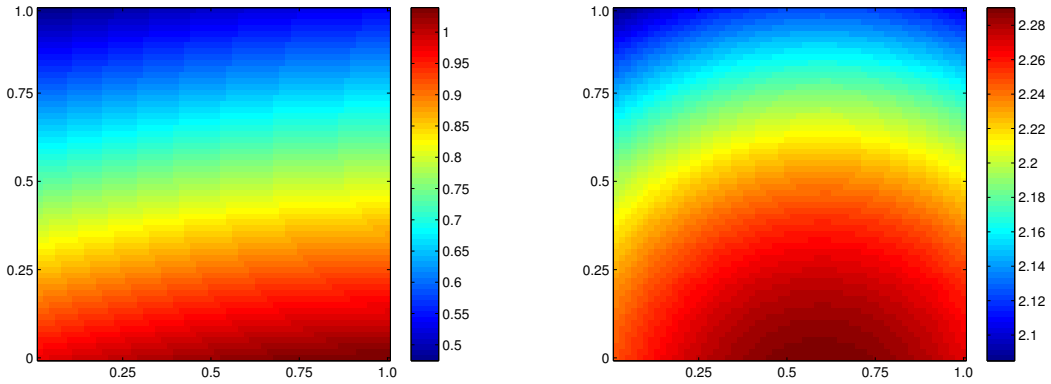

Fig. 6.9. Reference and initial permeability fields. Left: reference field; Right: initial field.

We compare the marginal posteriors of θ_1 and θ_2 for these three cases. The results are shown in Figure 6.10. The posterior distribution of two-stage DASAMC is reconstructed by importance sampling. *Maximum a posteriori estimations* (MAP) of all the cases are reported in Table 6.3. We can conclude from the figure and the table that these methods are able to recover the reference permeability field, with the same posterior distributions. But only 500 fine-scale calls in two-stage MCMC and two-stage DASAMC to get a good approximation of posterior distribution.

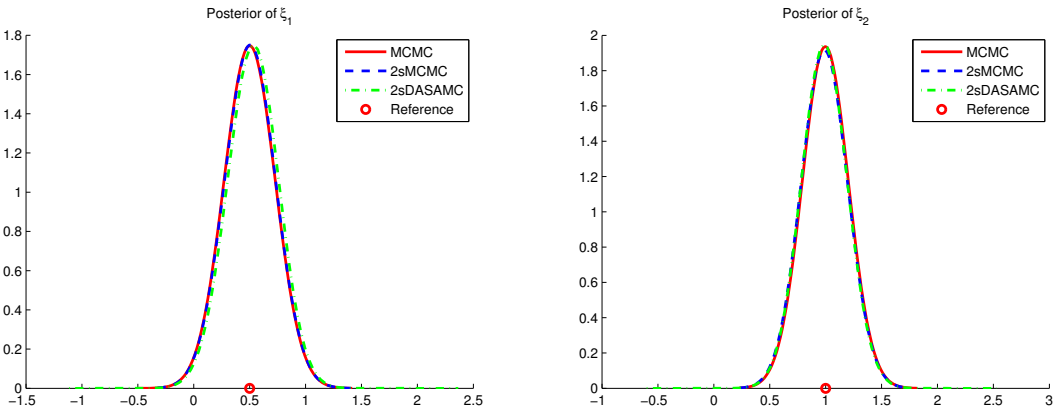

Fig. 6.10. Marginal posteriors of θ_1 and θ_2.

Table 6.3. MAP estimations of different methods.

Reference	MCMC	2sMCMC	2sDASAMC
0.500000	0.502322	0.515817	0.517992
1.000000	1.000062	0.986909	1.005900

CHAPTER VII

SUMMARY

In this dissertation, the subsurface characterization for flows in highly heterogeneous porous media has been studied. We have used channelized spatial fields via a level set approach to describe the permeability field where channel boundaries are assumed to have random locations. We use smooth velocity fields to change the channel boundaries within the level set framework and, thus, the parameterization of channel boundaries can be mapped to that of smooth velocity fields. One of our main contributions is the study of the regularity of posterior distribution. In particular, we study errors introduced in the posterior measure by truncating the prior distribution. The estimation is carried out using finite dimensional uncertainty space and working with infinite dimensional PDEs. This makes the analysis easy and avoids involving "infinite" dimensional probabilistic spaces. We have shown that the truncation error is independent of the dimension of the stochastic space. This is important because the parameter space can have a large dimension, in general. The subsurface characterization is carried out within Bayesian framework where the posterior distribution is sampled. The numerical results show the validity of the proposed parameterization to interfaces and the error estimations.

We have proposed the ensemble-level MsFEM and preconditioner methods to solve the two-phase flow equation with heterogeneous coefficients. The ensemble idea is inspired by trying to reduce the expense in constructing multiscale basis functions. As the expensive construction is required for multiscale basis of new permeability fields, the best situation is to be able to reuse some basis functions with certain restrictions. We apply this idea numerically to channelized permeability fields in our

numerical experiments. The results show that the ensemble-level MsFEM gives good approximation to fine-scale solution, and the ensemble-level preconditioner successes in giving a contrast independent condition number for the preconditioning system. When the sampling process adopts these methods, the sampling efficiency can be achieved.

We have also proposed an ANOVA-based and a novel adaptive ANOVA-based mixed MsFEM for the stochastic two-phase flow problem. The framework has been described. The properties of the methods have been further studied with numerical examples on different random permeability fields. We have investigated the structure of errors without applying other adaptivity techniques. Note that the full ANOVA decomposition will contain more and more subproblems in the stochastic space, when the dimension increases. This fact can make the total computations as expensive as solving a high-dimensional problem directly when the number of the dimension exceeds certain point. The adaptive version is a remedy. A new adaptive ANOVA-based on variance decomposition mixed MsFEM has been proposed and compared with existing adaptive criterion. Our proposed adaptive ANOVA method can decide the active dimensions and interactions among dimensions before computing the decomposition itself. The numerical results show that this novel adaptive method can achieve similar accuracy as other adaptive strategies but with lower computational cost. The advantage in saving computational time will be more obvious when the dimension of the problem becomes higher.

We have combined the multi-stage MCMC algorithm with SAMC and DASAMC to develop multi-stage SAMC and multi-stage DASAMC methods for the subsurface inverse problem. The inversion problems are difficult usually due to the non-uniqueness of the solutions. Our problem has another difficulty that it is nonlinear in forward computation. The multi-stage versions of SAMC and DASAMC inherit

properties from both methods. At one hand, they screen out bad proposals by using the computations on coarse grid as a filter. At the other hand, the dynamic weights associated to each iteration helps the samples travel though the whole sample space, with acceleration to the global optimization by annealing. The numerical results have validated the proposed methods.

In the future, the definition of topological similarity needs to be further explored. For the ANOVA-based MsFEM, the numerical experiments show that the dominant errors are introduced by the mixed multiscale methods compared with the ANOVA decomposition errors. This gives motivation to develop multiscale methods with better accuracy to improve this method. In particular, we plan to consider a systematic enrichment technique developed in [27]. At the same time, this kind of approaches gives better approximation for mean than variance estimations. It would be interesting to develop more suitable algorithm for variance computation. Besides the advanced sampling methods we used in this dissertation, other powerful techniques can be further adopted, for example Langevin, Riemann manifold Langevin and Hamiltonian methods.

REFERENCES

[1] J. E. Aarnes, On the use of a mixed multiscale finite element method for greater flexibility and increased speed or improved accuracy in reservoir simulation, Multiscale Modeling and Simulation 2 (2004) 421-439.

[2] J. E. Aarnes, Y. Efendiev, Mixed multiscale finite element for stochastic porous media flows, SIAM Journal on Scientific Computing 30 (2008) 2319-2339.

[3] J. E. Aarnes, Y. Efendiev, L. Jiang, Mixed multiscale finite element methods using limited global information, Multiscale Modeling and Simulation 7(2) (2008) 655-676.

[4] J. Aarnes, T. Hou, Multiscale domain decomposition methods for elliptic problems with high aspect ratios, Acta Mathematicae Applicatae Sinica, English Series 18(1) (2002) 63-76.

[5] J. E. Aarnes, S. Krogstad, K.-A. Lie, A hierarchical multiscale method for two-phase flow based upon mixed finite elements and nonuniform grids, Multiscale Modeling and Simulation 5(2) (2006) 337-363.

[6] A. Abdulle, B. Engquist, Finite element heterogeneous multiscale methods with near optimal computational complexity, Multiscale Modeling and Simulation 6(4) (2007) 1059-1084.

[7] G. Allaire, R. Brizzi, A multiscale finite element method for numerical homogenization, Multiscale Modeling and Simulation 4(3) (2005) 790-812.

[8] T. Arbogast, Implementation of a locally conservative numerical subgrid upscaling scheme for two-phase Darcy flow, Computational Geosciences 6 (2002) 453-481.

[9] T. Arbogast, G. Pencheva, M. F. Wheeler, I. Yotov, A multiscale mortar mixed finite element method, Multiscale Modeling and Simulation 6(3-4) (2007) 319-346.

[10] M. Balhoff, S. Thomas, M. Wheeler, Mortar coupling and upscaling of pore-scale models, Computational Geosciences 12(1) (2007) 15-27.

[11] V. Barthelmann, E. Novak, K. Ritter, High dimensional polynomial interpolation on sparse grids, Advanced in Computational Mathematics 12 (2000) 273-288.

[12] L. Berlyand, H. Owhadi, A new approach to homogenization with arbitrary rough high contrast coefficients for scalar and vectorial problems, (2009) arXiv: 0901.1463v3d.

[13] F. Brezzi, Interacting with the subgrid world, in: Numerical Analysis 1999 (Dundee), Chapman & Hall/CRC, Boca Raton, FL, (2000) 69-82.

[14] J. Caers, Petroleum Geostatistics, SPE Interdisciplinary Primer Series, Society of Petroleum Engineers, 2005.

[15] R. E. Caflisch, W. Morokoff, A. Owen, Valuation of mortgage-backed securities using brownian bridges to reduce the effective dimension, Journal of Computational Finance 1 (1997) 27-46.

[16] G. Caumon, S. Strebelle, J. Caers, A. Journel, Assessment of global uncertainty for early appraisal of hydrocarbon Fields, SPE 89943, Annual Technical Conference and Exhibition, 26-29 September, Houston, Texas, 2004.

[17] Z. Chen, T. Y. Hou, A mixed multiscale finite element method for elliptic problems with oscillating coefficients, Mathematics of Computation 72 (2003)

541-576.

[18] Y. Cheng, G. Lin, An improved method of VFSA for global optimization, in progress (2012).

[19] E. Chung, Y. Efendiev, Reduced-contrast approximations for high-contrast multiscale flow problems, Multiscale Modeling and Simulation 8(4) (2010) 1128-1153.

[20] C. V. Deutsch, A. G. Journel, GSLIB: Geostatistical Software Library and Users Guide, Oxford University Press, New York, NY, 1998.

[21] A. Datta-Gupta, M. J. King, Streamline Simulation: Theory and Practice, Society of Petroleum Engineers, Richardson, TX, 2007.

[22] O. Dubrule, Geostatistics in Petroleum Geology, American Association of Petroleum Geology, Course Note Series 38, 1998.

[23] L. J. Durlofsky, Numerical calculations of equivalent gridblock permeability tensors for heterogeneous porous media, Water Resources Research 27(5) (1991) 699-708.

[24] L. J. Durlofsky, Coarse scale models of two-phase flow in heterogeneous reservoirs: volume averaged equations and their relation to existing upscaling techniques, Computational Geosciences 2 (1998) 73-92.

[25] L. J. Durlofsky, Y. Efendiev, V. Ginting, An adaptive local-global multiscale finite volume element method for two-phase flow simulations, Advances in Water Resources 30 (2007) 576-588.

[26] Y. Efendiev, J. Galvis, A domain decomposition preconditioner for multiscale high-contrast problems, in Domain Decomposition Methods in Science and Engineering XIX, Huang, Y.; Kornhuber, R.; Widlund, O.; Xu, J. (Eds.), Volume 78 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, Part 2 (2011) 189-196.

[27] Y. Efendiev, J. Galvis, X. H. Wu, Multiscale finite element methods for high-contrast problems using local spectral basis functions, Journal of Computational Physics 230 (2011) 937-955.

[28] W. E, B. Engquist, The heterogeneous multiscale methods, Communications in Mathematical Sciences 1(1) (2003) 87-132.

[29] Y. Efendiev, A. Datta-Gupta, V. Ginting, X. Ma, B. Mallick, An efficient two-stage Markov chain Monte Carlo method for dynamic data integration, Water Resources Research 41 (2005) W12423.

[30] Y. Efendiev, J. Galvis, F. Thomines, A systematic coarse-scale model reduction technique for parameter-dependent flows in highly heterogeneous media and its applications, preprint (2012).

[31] Y. Efendiev, V. Ginting, T. Hou, R. Ewing, Accurate multiscale finite element methods for two-phase flow simulations, Journal of Computational Physics 220(1) (2006) 155-174.

[32] Y. Efendiev, T. Y. Hou, Multiscale Finite Element Methods: Theory and Applications, Springer, New York, NY, 2009.

[33] Y. Efendiev, T. Hou, V. Ginting, Multiscale finite element methods for nonlinear problems and their applications, Communications in Mathematical Sciences 2

(2004) 553-589.

[34] Y. Efendiev, T. Hou, W. Luo, Preconditioning Markov chain Monte Carlo simulations using coarse-scale models, SIAM Journal on Scientific Computing 28(2) (2006) 776-803.

[35] T. Egeland, F. Georgsen, R. Knarud, H. More, Multifacies modelling of fluvial reservoirs, SPE 26502 in proceedings of the SPE Annual Technical Conference and Exhibition, 3-6 October, Houston, Texas, 1993.

[36] R. Fisher, Statistical Methods for Research Workers, Oliver and Boyd, Edinburgh, UK, 1925.

[37] H. P. Flath, L. C. Wilcox, V. Akcelik, J. Hill, B. Van Bloemen Waanders, O. Ghattas, Fast algorithms for Bayesian uncertainty quantification in large-scale linear inverse problems based on low-rank partial hessian approximations, SIAM Journal on Scientific Computing 33(1) (2011) 407-432.

[38] J. Y. Foo, G. E. Karniadakis, Multi-element probabilistic collocation in high dimensions, Journal of Computational Physics 229 (2009) 1536-1557.

[39] J. Galvis, Y. Efendiev, Domain decomposition preconditioners for multiscale flows in high contrast media, Multiscale Modeling and Simulation 8(4) (2012) 1461-1483.

[40] J. Galvis, Y. Efendiev, Domain decomposition preconditioners for multiscale flows in high-contrast media: Reduced dimension coarse spaces, Multiscale Modeling and Simulation 8(5) (2010) 1621-1644.

[41] I. G. Graham, P. O. Lechner, R. Scheichl, Domain decomposition for multiscale PDEs, Numerische Mathematik 106(4) (2007) 589-626.

[42] M. Grigoriu, Stochastic Calculus: Applications in Science and Engineering, Birkhäuser, Boston, MA, 2002.

[43] H. Haldorsen, E. Damsleth, Stochastic modeling, JPT, SPE 20321 (1990) 404-412.

[44] U. Hetmaniuk, R. Lehoucq, Special finite element methods based on component mode synthesis techniques, ESAIM: Mathematical Modeling and Numerical Analysis 44(3) (2010) 401-420.

[45] T. Y. Hou, X. H. Wu, A multiscale finite element method for elliptic problems in composite materials and porous media, Journal of Computational Physics 134 (1997) 169-189.

[46] X. Hu, G. Lin, T.Y. Hou, P. Yan, An ANOVA-based data-driven stochastic multiscale method for elliptic PDE with random coefficients, preprint (2012).

[47] S. S. Hubbard, Y. Rubin, E. Majer, Spatial correlation structure estimation using geophysical and hydrogeological data, Water Resources Research 35(6) (1999) 1809-1826.

[48] T. Hughes, G. Feijoo, L. Mazzei, J. Quincy, The variational multiscale method - a paradigm for computational mechanics, Computer Methods in Applied Mechanics and Engineering 166 (1998) 3-24.

[49] P. Jenny, S. H. Lee, H. Tchelepi, Multi-scale finite volume method for elliptic problems in subsurface flow simulation, Journal of Computational Physics 187 (2003) 47-67.

[50] L. Jiang, I. Mishev, Y. Li, Stochastic mixed multiscale finite element methods and their applications in reandom porous media, Computer Methods in Applied

Mechanics and Engineering 119 (2011) 2721-2740.

[51] C. E. Koltermann, S. M. Gorelick, Heterogeneity in sedimentary deposits: a review of structure-imitating, process-imitating, and descriptive approaches, Water Resources Research 32(9) (1996) 2617-2658.

[52] H. Lee, D. Higdon, Z. Bi, M. Ferreira, M. West, Markov random field models for high-dimensional parameters in simulations of fluid flow in porous media, Technometrics 44(3) (2002) 230-241.

[53] O. P. LeMaître, O. M. Knio, Spectral Methods for Uncertainty Quantification: with Applications to Computational Fluid Dynamics, Springer, New York, 2010.

[54] F. Liang, Annealing stochastic approximation Monte Carlo for neural network training, Machine Learning 68(3) (2007) 201-233.

[55] F. Liang, Annealing evolutionary stochastic approximation Monte Carlo for global potimization, Statistics and Computing 21 (2010) 375-393.

[56] F. Liang, C. Liu, R. J. Carroll, Advanced Monte Carlo Methods: Learning from Past Samples, Wiley, Hoboken, NJ, 2010.

[57] F. Liang, W.H. Wong, R. J. Carroll, Stochastic approximation in Monte Carlo computation, Journal of the American Statistica Sinica 10(2) (2008) 317-342.

[58] G. Lin, G. E. Karniadakis, Sensitivity analysis and stochastic simulations of non-equilibrium plasma flow, International Journal for Numerical Methods in Engineering 80 (2009) 738-766.

[59] G. Lin, C.-H. Su, G. E. Karniadakis, Predicting shock dynamics in the presence of uncertainties, Journal of Computational Physics 217 (2006) 260-276.

[60] G. Lin, C.-H. Su, G. E. Karniadakis, Stochastic modeling of random roughness in shock scattering problems: theory and simulations, Computer Methods in Applied Mechanics and Engineering 197 (2008) 3420-3434.

[61] J. S. Liu, Monte Carlo Strategies in Scientific Computing, Springer, New York, NY, 2001.

[62] I. Lunati, P. Jenny, Multiscale finite-volume method for compressible multiphase flow in porous media, Journal of Computational Physics 216 (2006) 616-636.

[63] X. Ma, N. Zabaras, An adaptive hierarchical sparse grid collocation algorithm for solution of stochastic differential equations, Journal of Computational Physics 228 (2009) 3084-3113.

[64] X. Ma, N. Zabaras, An efficient Bayesian inference approach to inverse problems based on an adaptive sparse grid collocation method, Inverse Problems 25 (2009) 035013.

[65] X. Ma, N. Zabaras, An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations, Journal of Computational Physics 229 (2010) 3884-3915.

[66] X. Ma, N. Zabaras, A stochastic mixed finite element heterogenours multiscale method for flow in porous media, Journal of Computational Physics 230 (2011) 4696-4722.

[67] A. Mondal, A. Datta-Gupta, Y. Efendiev, B. Mallick, Bayesian uncertainty quantification for flows in heterogeneous porous media using reversible jump

Markov Chain Monte Carlo methods, Advances in Water Resources 33 (2010) 241-256.

[68] Y. M. Marzouk, H. N. Najm, Dimensionality reduction and polynomial chaos acceleration of Bayesian inference in inverse problems, Journal of Computational Physics 228 (2009) 1862-1902.

[69] Y. M. Marzouk, H. N. Najm, L. A. Rahn, Stochastic spectral methods for efficient Bayesian solution of inverse problems, Journal of Computational Physics 224 (2007) 560-586.

[70] Y. M. Marzouk, D. Xiu, A stochastic collocation approach to Bayesian inference in inverse problems, Communications in Computational Physics 6(4) (2009) 826-847.

[71] T. P. A. Mathew, Domain decomposition methods for the numerical solution of partial differential equations, volume 61 of Lecture Notes in Computational Science and Engineering, Springer-Verlag, Berlin, 2008.

[72] R. G. Regis, C. A. Schoemaker, A stochastic radial basis function method for the global optimization of expensive functions, INFORMS Journal on Computing 19(4) (2007) 497-509.

[73] R. G. Regis, C. A. Schoemaker, Parallel stochastic global optimization using radial basis functions, INFORMS Journal on Computing 21(3) (2009) 411-426.

[74] F. Nobile, R. Tempone, C. G. Webster, A sparse grid stochastic collocation method for partial differential equations with random input data, SIAM Journal on Numerical Analysis 46 (2008) 2309-2345.

[75] F. Nobile, R. Tempone, C. G. Webster, An anisotropic sparse grid stochastic collocation method for elliptic partial differential equations with random input data, SIAM Journal on Numerical Analysis 46 (2008) 2411-2442.

[76] J. Nolen, G. Papanicolaou, O. Pironneau, A framework for adaptive multiscale method for elliptic problems, Multiscale Modeling and Simulation 7 (2008) 171-196.

[77] J. M. Nordbotten, Adaptive variational multiscale methods for multiphase flow in porous media, Multiscale Modeling and Simulation 7(3) (2009) 1455-1473.

[78] M. Ohlberger, A posterior error estimates for the heterogeneous mulitscale finite element method for elliptic homogenization problems Multiscale Modeling and Simulation 4(1) (2005) 88-114.

[79] D. S. Oliver, Y. Chen, Recent progress on reservoir history matching: a review, Computational Geosciences 15(1) (2011) 185-221.

[80] D. S. Oliver, A. C. Reynolds, Z. Bi, Y. Abacioglu, Integration of production data into reservoir models, Petroleum Geoscience 7 (2001) S65-S73.

[81] S. Osher, R. Fedkiw, Level Set Methods and Dynamic Implicit Surfaces, Springer, New York, NY, 2003.

[82] S. Osher, J.A. Sethian, Front propogation with curvature dependent speed: algorithms based on Hamilton-Jacobi formulations, Journal of Computational Physics 56 (1988) 12-49.

[83] H. Owhadi, L. Zhang, Metric based up-scaling, Communications on Pure and Applied Mathematics 60(5) (2007) 675-723.

[84] C. Pechstein, R. Scheichl, Analysis of FETI methods for multiscale PDEs, Numerische Mathematik 111(2) (2008) 293-333.

[85] M. Peszynska, M. F. Wheeler, I. Yotov, Mortar upscaling for multiphase flow in porous media, Computational Geosciences 6 (2002) 73-100.

[86] P. Popov, Y. Efendiev, G. Qin, Multiscale modeling and simulations of flows in naturally fractured karst reservoirs, Communications in Computational Physics 6(1) (2009) 162-184.

[87] C. Robert, G. Casella, Monte Carlo Statistical Methods. Springer-Verlag, New-York, 1999.

[88] G. Sangalli, Capturing small scales in elliptic problems using a residual-free bubbles finite element method, Multiscale Modeling and Simulation 1 (2003) 485-503.

[89] C. Schwab, R. Todor, Karhunen-Loève approximation of random fields by generalized fast multipole methods, Journal of Computational Physics 217 (2006) 100-122.

[90] J. Sethian, Level Set Methods and Fast Marching Methods, Cambridge University Press, Cambridge, UK, 1999.

[91] T. Strinopoulos, Upscaling immiscible two-phase flows in an adaptive frame, Ph.D. dissertation, California Institute of Technology, 2005.

[92] A. M. Stuart, Inverse problems: a Bayesian perspective, Acta Numerica 19 (2010) 451-559.

[93] D. Svyatskiy, D. Moulton, K. Lipnikov, Multilevel multiscale mimetic (M3) method for two-phase flows in porous media, Journal of Computational Physics 227(14) (2008) 6727-6753.

[94] X. C. Tai, T. Chan, A survey on multiple level set methods with applications for identifying piecewise constant functions, International Journal of Numerical Analysis & Modeling 1 (2004) 25-47.

[95] M. Tatang, G. McRae, Direct treatment of uncertainty in models of reaction and transport, Technical report, MIT Tech. Rep., 1994.

[96] A. Toselli, O. Widlund, Domain decomposition methods – algorithms and theory, volume 34 of Springer Series in Computational Mathematics, Springer-Verlag, Berlin, 2005.

[97] D. W. Vasco, S. Yoon, A. Datta-Gupta, Integrating a dynamic data into high resolution reservoir models using streamline-based analytical sensitivity coefficients, SPE Journal 4(4) (1999).

[98] D. W. Vasco, A. Datta-Gupta, R. Behrens, P. Condon, J. Rickett, Seismic imaging of reservoir flow properties: time-lapse amplitude changes, Geophysics 69 (6) (2004) 1425-1442.

[99] R. Villegas, O. Dorn, M. Moscoso, M. Kindelan, Simulations characterization of geological shapes and permeability distributions in reservoirs using the level set method, SPE 100291, SPE Europec/EAGE Annual Conference and Exhibition held in Vienna, Austrua, 12-15 June 2006.

[100] K. J. Weber, Influence of common sedimentary structures on fluid flow in reservoir models, Journal of Petroleum Technology (1990) 665-672.

[101] A. Wilson, Y. Rubin, Characterization of aquifer heterogeneity using indicator variables for solute concentrations, Water Resources Research 38 (2002) 1283.

[102] E. Wong, Stochastic Processes in Information and Dynamical Systems, MCGraw-Hill, New York, NY, 1971.

[103] X. H. Wu, Y. Efendiev, T. Y. Hou, Analysis of upscaling absolute permeability, Discrete and Continuous Dynamical Systems Series B 2 (2002) 185-204.

[104] D. Xiu, Numerical Methods for Stocahstic Computations: A Spectral Method Approach, Princeton University Press, Princeton, NJ, 2010.

[105] D. Xiu, J. S. Hesthaven, High order collocation methods for differential equations with random inputs, SIAM Journal on Scientific Computing 27 (2005) 1118-1139.

[106] J. Xu, L. Zikatanov, On an energy minimizing basis for algebraic multigrid methods, Computing and Visualization in Science 7 (2004) 121-127.

[107] X. Yang, M. Choi, G. Lin, G. E. Karniadakis, Adaptive ANOVA decomposition of stochastic incompresible and compressible flows, Journal of Computational Physics 231(4) (2012) 1587-1614.

[108] Z. Zhang, M. Choi, G.E. Karniadakis, Anchor points matter in anova decomposition, in: Spectral and High Order Methods for Partial Diferential Equations Lecture Notes in Computational Science and Engineering 76, Springer (2011) 347-355.