PRECONDITIONING MARKOV CHAIN MONTE CARLO SIMULATIONS USING COARSE-SCALE MODELS*

Y. EFENDIEV†, T. HOU‡, AND W. LUO‡

Abstract. We study the preconditioning of Markov chain Monte Carlo (MCMC) methods using coarse-scale models with applications to subsurface characterization. The purpose of preconditioning is to reduce the fine-scale computational cost and increase the acceptance rate in the MCMC sampling. This goal is achieved by generating Markov chains based on two-stage computations. In the first stage, a new proposal is first tested by the coarse-scale model based on multiscale finite volume methods. The full fine-scale computation will be conducted only if the proposal passes the coarse-scale screening. For more efficient simulations, an approximation of the full fine-scale computation using precomputed multiscale basis functions can also be used. Comparing with the regular MCMC method, the preconditioned MCMC method generates a modified Markov chain by incorporating the coarse-scale information of the problem. The conditions under which the modified Markov chain will converge to the correct posterior distribution are stated in the paper. The validity of these assumptions for our application and the conditions which would guarantee a high acceptance rate are also discussed. We would like to note that coarse-scale models used in the simulations need to be inexpensive but not necessarily very accurate, as our analysis and numerical simulations demonstrate. We present numerical examples for sampling permeability fields using two-point geostatistics. The Karhunen–Loève expansion is used to represent the realizations of the permeability field conditioned to the dynamic data, such as production data, as well as some static data. Our numerical examples show that the acceptance rate can be increased by more than 10 times if MCMC simulations are preconditioned using coarse-scale models.

1. Introduction. Uncertainties on the detailed description of reservoir lithofacies, porosity, and permeability are major contributors to uncertainty in reservoir performance forecasting. Reducing this uncertainty can be achieved by integrating additional data in subsurface modeling. With the increasing interest in accurate prediction of subsurface properties, subsurface characterization based on dynamic data, such as production data, becomes more important.

To predict future reservoir performance, the reservoir properties, such as porosity and permeability, need to be conditioned to dynamic data, such as production data. In general it is difficult to calculate this probability distribution, because the process of predicting flow and transport in petroleum reservoirs is nonlinear. Instead, this probability distribution is estimated from the outcomes of flow predictions for a large number of realizations of the reservoir. It is essential that the permeability (and porosity) realizations adequately reflect the uncertainty in the reservoir properties; i.e., the probability distribution is sampled correctly. This problem is challenging because the permeability field is a function defined on a large number of grid blocks.

The Markov chain Monte Carlo (MCMC) method and its modifications have been used previously to sample the posterior distribution. In this paper, we design a two-stage MCMC method which employs coarse-scale models based on multiscale finite volume methods.

The direct MCMC simulations are generally very CPU demanding because each proposal requires solving a forward coupled nonlinear PDE over a large time interval. The forward fine-scale problem is usually formulated on a large number of grid blocks, which makes it prohibitively expensive to perform a sufficient number of MCMC simulations. There have been a few attempts to propose MCMC methods with high acceptance rates, for example, the randomized maximum likelihood method [19, 20]. This approach uses unconditional realizations of the production and permeability data and solves a deterministic gradient-based inverse problem. The solution of this minimization problem is taken as a proposal, and is accepted with probability 1, because the rigorous acceptance probability is very difficult to estimate. In addition to the need of solving a gradient-based inverse problem, this method does not properly sample the posterior distribution. Thus, developing efficient rigorous MCMC calculations with high acceptance rates remains a challenging problem.

In this paper, we show that using inexpensive coarse-scale computations one can increase the acceptance rate of MCMC calculations. Here the acceptance rate refers to the ratio between the number of accepted permeability samples and the number of times of solving the fine-scale nonlinear PDE system. The method consists of two stages. At the first stage, using coarse-scale runs we determine whether or not to run the fine-scale simulations. If the proposal is accepted at the first-stage, then a fine-scale simulation is performed at the second stage to determine the acceptance probability of the proposal. The first stage of the MCMC method modifies the proposal distribution. We show that the modified Markov chain satisfies the detailed balance condition for the correct distribution. Moreover, we point out that the chain is ergodic and converges to the correct posterior distribution under some technical assumptions. The validity of the assumptions for our application is discussed in the paper. We would like to note that two-stage MCMC algorithms have been used previously (e.g., [2, 15, 21, 10]) in different situations.

In this paper, we use a coarse-scale model based on multiscale finite volume methods. Note that it is essential that these coarse-scale models are inexpensive but not necessarily very accurate. The main idea of multiscale finite volume methods is to construct multiscale basis functions that contain the small scale information. Constructing these basis functions based on the single-phase flow is equivalent to single-phase flow upscaling, provided the transport equation is solved on a coarse grid. This method is inexpensive, since the basis functions are constructed only once, and the transport equation is solved on the coarse grid. The use of multiscale finite volume methods has another advantage: it can be further used as an accurate approximation for the production data if the transport equation is solved on the fine grid. For this purpose, one needs to compute the fine-scale velocity fields from the precomputed multiscale basis functions and solve the saturation on the fine grid. This provides an accurate approximation for the production data [12, 13, 1]. Since one can reuse the basis functions from the first stage, the resulting method is very efficient. We would like to note that upscaled models are used in MCMC simulations in previous findings. In an interesting work [9], the authors employ error models between coarse- and fine-scale simulations to quantify the uncertainty.

Numerical results for permeability fields generated using two-point geostatistics are presented in the paper. Using the Karhunen–Loève expansion, we can represent

the high dimensional permeability field by a small number of parameters. Further-more, static data (the values of permeability field at some sparse locations) can be easily incorporated into the Karhunen–Loève expansion to further reduce the dimen-sion of the parameter space. Numerical results are presented for both single-phase and two-phase flows for side-to-side and corner-to-corner flows. In all the simulations, we observe more than 10 times an increase in the acceptance rate. In other words, the preconditioned MCMC method can accept the same number of samples with much fewer fine-scale runs.

The paper is organized in the following way. In the next section, we briefly describe the model equations and their upscaling. Section 3 is devoted to the analysis of the preconditioned MCMC method and its relevance to our particular application. Numerical results are presented in section 4.

2. Fine and coarse models. We consider two-phase flows in a domain Ω under the assumption that the displacement is dominated by viscous effects. We neglect the effects of gravity, compressibility, and capillary pressure. The two phases are referred to as water (aqueous phase) and oil (nonaqueous phase liquid), designated by sub-scripts w and o, respectively. We write Darcy's law, with all quantities dimensionless, for each phase as follows:

$$(2.1) \qquad \boldsymbol{v}_j = -\frac{k_{rj}(S)}{\mu_j}\boldsymbol{k} \cdot \nabla p,$$

where $\boldsymbol{v}_j$, $j = w, o$, is the phase velocity, $\boldsymbol{k}$ is the permeability tensor, k_{rj} is the relative permeability of the phase j, S is the water saturation (volume fraction), and p is the pressure. In this work, a single set of relative permeability curves is used and $\boldsymbol{k}$ is taken to be a diagonal tensor. Combining Darcy's law with a statement of conservation of mass allows us to express the governing equations in terms of the so-called pressure and saturation equations:

$$(2.2) \qquad \nabla \cdot (\lambda(S)\boldsymbol{k}\nabla p) = q,$$

$$(2.3) \qquad \frac{\partial S}{\partial t} + \boldsymbol{v} \cdot \nabla f(S) = -q_w,$$

where $\lambda(S)$ is the total mobility, q and q_w are the source terms, $\boldsymbol{v}$ is the total velocity, and $f(S)$ is the flux function, which are, respectively, given by

$$(2.4) \qquad \lambda(S) = \frac{k_{rw}(S)}{\mu_w} + \frac{k_{ro}(S)}{\mu_o},$$

$$(2.5) \qquad \boldsymbol{v} = \boldsymbol{v}_w + \boldsymbol{v}_o = -\lambda(S)\boldsymbol{k}\,\nabla p,$$

$$(2.6) \qquad f(S) = \frac{k_{rw}(S)/\mu_w}{k_{rw}(S)/\mu_w + k_{ro}(S)/\mu_o}.$$

The above description is referred to as the fine model of the two-phase flow problem. For the single-phase flow, we have $\lambda(S) = 1$ and $f(S) = S$. Throughout, the porosity is assumed to be constant and the permeability is assumed to have the form $\boldsymbol{k}(\boldsymbol{x}) = k(\boldsymbol{x})\boldsymbol{I}$, where $\boldsymbol{I}$ is a unit matrix and $k(\boldsymbol{x})$ is a scalar function.

The proposed coarse-scale model consists of upscaling the pressure equation (2.2) to obtain the velocity field on the coarse grid and then using it in (2.3) to resolve the
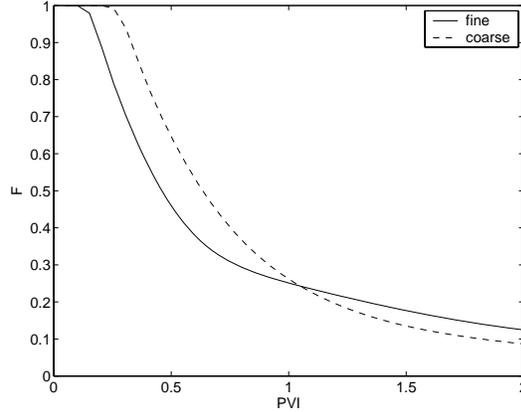
FIG. 2.1. *Typical fine- and coarse-scale fractional flows.*

saturation on the coarse grid. The pressure equation is upscaled using the multiscale finite volume method. The details of the method are presented in Appendix A. Using the multiscale finite volume method, we obtain the coarse-scale velocity field, which is used in solving the saturation equation on the coarse grid. Since no subgrid modeling is performed for the saturation equation, this upscaling procedure introduces errors. In Figure 2.1, we present a comparison of the typical fractional flows computed by fine- and coarse-scale models. The fractional flows are plotted against the dimensionless time "pore volume injected." The pore volume injected (PVI) at time T is defined as $\frac{1}{V_p} \int_0^T q_t(\tau) d\tau$, where q_t is the combined flow rates of water and oil at the production edge, and V_p is the total pore volume of the system. PVI provides the dimensionless time for the flow displacement. The fractional flow $F(t)$ (denoted simply by F thereafter) is the fraction of oil in the produced fluid and is defined as $F = q_o/q_t$, where $q_t = q_o + q_w$, with q_o and q_w denoting the flow rates of oil and water at the production edge of the model. More specifically,

$$F(t) = 1 - \frac{\int_{\partial \Omega^{out}} f(S) \, v_n \, dl}{\int_{\partial \Omega^{out}} v_n \, dl},$$

where $\partial \Omega^{out}$ is the outflow boundary and $v_n = \boldsymbol{v} \cdot \boldsymbol{n}$ is the normal velocity on the boundary. In future analysis, the notations q_o, q_w, or q_t will not be used, and q will be reserved for the proposal distributions. The proposed coarse-scale model is somewhat similar to the single-phase flow upscaling [4]. One can improve the accuracy of the above coarse model by solving the transport equation on the fine grid using the fine-scale velocity field which can be computed employing precomputed multiscale basis functions. This makes solving the coarse model more expensive because the transport update is performed on the fine grid with smaller time steps. However, it can provide an efficient numerical solver for the second stage of the preconditioned MCMC method, as we will discuss later.

3. Preconditioning MCMC simulations using coarse-scale models.

3.1. Problem setting. The problem under consideration consists of sampling the permeability field given fractional flow measurements. Typically, the permeability field is known at some sparse locations. This information should be incorporated into

the prior models (distributions) of the permeability. Since the fractional flow is an integrated response, the map from the permeability field to the fractional flow is not one-to-one. Hence this problem is ill-posed in the sense that there exist many different permeability realizations for the given production data.

From the probabilistic point of view, this problem can be regarded as sampling the permeability field conditioning on the fractional flow data with measurement errors. Consequently, our goal is to sample from the conditional distribution $P(k|F)$, where k is the fine-scale permeability field and F is the fractional flow curve measured from the production data. Using the Bayes theorem we can write

$$(3.1) \qquad P(k|F) \propto P(F|k)P(k).$$

In the above formula, $P(k)$ is the prior distribution of the permeability field, which is assumed to be log-normal. The prior distribution $P(k)$ will also incorporate the additional information of the permeability field at the sparse locations. The likelihood function $P(F|k)$ denotes the conditional probability that the outcome of the measurement is F when the true permeability is k.

In practice, the measured fractional flow F contains measurement errors. Denote the fractional flow for a given k as F_k. F_k can be computed by solving the model equation (2.1)–(2.3) on the fine grid. The computed F_k will contain a modeling error as well as a numerical error. In this paper, we assume that the combined errors from the measurement, modeling, and numerics satisfy a Gaussian distribution. That is, the likelihood function $P(F|k)$ takes the form

$$(3.2) \qquad P(F|k) \propto \exp\left(-\frac{\|F - F_k\|^2}{\sigma_f^2}\right),$$

where F is the observed fractional flow, F_k is the fractional flow computed by solving the model equations (2.1)–(2.3) on the fine grid for a given k, and σ_f is the precision associated with the measurement F and the numerical solution F_k. Since both F and F_k are functions of t, $\|F - F_k\|^2$ denotes the L_2 norm

$$\|F - F_k\|^2 = \int_0^T [F(t) - F_k(t)]^2 \, dt.$$

It is worth noting that the method discussed in this paper does not depend on the specific form of the error functions. A more general error model can be used in the simulations. We would like to emphasize that different permeability fields may produce the same fractional flow curve. Thus, the likelihood distribution $P(F|k)$ is a multimodal function of k with multiple local maxima.

Denote the posterior distribution as

$$(3.3) \qquad \pi(k) = P(k|F) \propto \exp\left(-\frac{\|F - F_k\|^2}{\sigma_f^2}\right) P(k).$$

Sampling from the distribution $\pi(k)$ can be accomplished by using the MCMC method. The main idea of the MCMC method is to generate a Markov chain with $\pi(k)$ as its stationary distribution. A key step to this approach is to construct the desired transition kernel for the Markov chain. In this paper, we use the Metropolis-Hasting algorithm. Suppose $q(y|x)$ is a general transitional probability distribution, which is

easy to sample and has an explicit form. The Metropolis-Hasting MCMC algorithm (see, e.g., [22]) consists of the following steps.

ALGORITHM (METROPOLIS-HASTING MCMC [22]).
- Step 1. At state k_n generate k from $q(k|k_n)$.
- Step 2. Accept k as a sample with probability

$$(3.4) \qquad p(k_n, k) = \min\left(1, \frac{q(k_n|k)\pi(k)}{q(k|k_n)\pi(k_n)}\right);$$

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

Starting with an arbitrary initial permeability sample k_0, the MCMC algorithm generates a Markov chain $\{k_n\}$. At each iteration, the probability of moving from state k_n to a next state k is $q(k|k_n)p(k_n, k)$, so the transition kernel for the Markov chain $\{k_n\}$ is

$$K(k_n, k) = p(k_n, k)q(k|k_n) + \left(1 - \int p(k_n, k)q(k|k_n)dk\right)\delta_{k_n}(k).$$

Using the explicit formula of the transition kernel, it is not difficult to prove that the target distribution $\pi(k)$ is indeed the stationary distribution of the Markov chain $\{k_n\}$. As a result, we can take k_n as samples of the distribution $\pi(k)$ after the chain reaches the steady state.

3.2. The preconditioned MCMC method. In the above Metropolis-Hasting MCMC algorithm, the major computational cost is to compute F_k in the target distribution $\pi(k)$, which involves solving the coupled nonlinear PDE system (2.1)–(2.3) on the fine grid. Generally, the MCMC method requires thousands of iterations before it converges to the steady state. To quantify the uncertainty of the permeability field accurately, one also needs to generate a large number of different samples. However, the acceptance rate of the direct MCMC method is very low, due to the large dimensionality of the permeability field. The algorithm needs to test many proposals to accept only a few permeability samples. Most of the CPU time is spent on simulating the rejected samples. That makes the direct (full) MCMC simulations prohibitively expensive in practice.

One way to improve the direct MCMC method is to increase its acceptance rate by modifying the proposal distribution $q(k|k_n)$. In this paper, we propose an algorithm in which the proposal distribution $q(k|k_n)$ is adapted to the target distribution using the coarse-scale model. Instead of testing each proposal by fine-scale computations directly, the algorithm first tests the proposal by the coarse-scale model. This is achieved by comparing the fractional flow curves on the coarse grid first. If the proposal is accepted by the coarse-scale test, then a full fine-scale computation will be conducted and the proposal will be further tested as in the direct MCMC method. Otherwise, the proposal will be rejected by the coarse-scale test and a new proposal will be generated from $q(k|k_n)$. The coarse-scale test filters the unacceptable proposals and avoids the expensive fine-scale tests for those proposals. The filtering process essentially modifies the proposal distribution $q(k|k_n)$ by incorporating the coarse-scale information of the problem. That is why the modified method is called a *preconditioned MCMC method*.

Recall that the fine-scale target distribution is given by (3.3). We approximate

the distribution $\pi(k)$ on the coarse scale by

$$(3.5) \qquad \pi^*(k) \propto \exp\left(-\frac{\|F - F_k^*\|^2}{\sigma_c^2}\right) P(k),$$

where F_k^* is the fractional flow computed by solving the coarse-scale model of (2.1)–(2.3) for the given k, and 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 numerical simulations. (cf. Figure 3.1 and later discussion). Using the coarse-scale distribution $\pi^*(k)$ as a filter, the preconditioned MCMC method can be described as follows.

ALGORITHM (PRECONDITIONED MCMC).
- Step 1. At k_n, generate a trial proposal k' from distribution $q(k'|k_n)$.
- Step 2. Take the real proposal as

$$k = \begin{cases} k' & \text{with probability } g(k_n, k'), \\ k_n & \text{with probability } 1 - g(k_n, k'), \end{cases}$$

where

$$(3.6) \qquad g(k_n, k') = \min\left(1, \frac{q(k_n|k')\pi^*(k')}{q(k'|k_n)\pi^*(k_n)}\right).$$

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

$$(3.7) \qquad Q(k|k_n) = g(k_n, k)q(k|k_n) + \left(1 - \int g(k_n, k)q(k|k_n)dk\right)\delta_{k_n}(k).$$

- Step 3. Accept k as a sample with probability

$$(3.8) \qquad \rho(k_n, k) = \min\left(1, \frac{Q(k_n|k)\pi(k)}{Q(k|k_n)\pi(k_n)}\right);$$

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

In the above algorithm, if the trial proposal k' is rejected by the coarse-scale test (Step 2), k_n will be passed to the fine-scale test as the proposal. Since $\rho(k_n, k_n) \equiv 1$, no further (fine-scale) computation is needed. Thus, the expensive fine-scale computations can be avoided for those proposals which are unlikely to be accepted. In comparison, the regular MCMC method requires a fine-scale simulation for every proposal k, even though most of the proposals will be rejected at the end.

It is worth noting that there is no need to compute $Q(k|k_n)$ and $Q(k_n|k)$ in (3.8) by the integral formula (3.7). The acceptance probability (3.8) can be simplified as

$$(3.9) \qquad \rho(k_n, k) = \min\left(1, \frac{\pi(k)\pi^*(k_n)}{\pi(k_n)\pi^*(k)}\right).$$

In fact, (3.9) is obviously true for $k = k_n$ since $\rho(k_n, k_n) \equiv 1$. For $k \neq k_n$,

$$Q(k_n|k) = g(k, k_n)q(k_n|k) = \frac{1}{\pi^*(k)}\min\left(q(k_n|k)\pi^*(k), \; q(k|k_n)\pi^*(k_n)\right)$$

$$= \frac{q(k|k_n)\pi^*(k_n)}{\pi^*(k)}g(k_n, k) = \frac{\pi^*(k_n)}{\pi^*(k)}Q(k|k_n).$$

Substituting the above formula into (3.8), we immediately get (3.9).

Since the computation of the coarse-scale solution is very cheap, Step 2 of the preconditioned MCMC method can be implemented very quickly to decide whether or not to run the fine-scale simulation. The second step of the algorithm serves as a filter that avoids unnecessary fine-scale runs for the rejected samples. It is possible that the coarse-scale test may reject an individual sample which will otherwise have a (small) probability to be accepted in the fine-scale test. However, that does not play a crucial role, since we are interested only in the statistical property of the samples. We will show later that the preconditioned MCMC algorithm converges under some mild assumptions.

We would like to note that the Gaussian error model for the coarse-scale distribution $\pi^*(k)$ is not very accurate. We use it only in the filtering stage to decide whether or not to run the fine-scale simulations. The choice of the coarse-scale precision parameter σ_c is important for increasing the acceptance rate. If σ_c is too large, then too many proposals can pass the coarse-scale tests and the filtering stage will become less effective. If σ_c is too small, then eligible proposals may be incorrectly filtered out, which will result in biased sampling. Our numerical results show that the acceptance rate is optimal when σ_c is of the same order as σ_f. The optimal value of σ_c can be estimated based on the correlation between $\|F - F_k\|$ and $\|F - F_k^*\|$ (cf. Figure 3.1).

Based on the Gaussian precision models (3.3) and (3.5), the acceptance probability (3.9) has the form

$$(3.10) \qquad \rho(k_n, k) = \min\left(1, \frac{\pi(k)\pi^*(k_n)}{\pi(k_n)\pi^*(k)}\right) = \min\left(1, \frac{\exp\left(-\frac{E_k - E_{k_n}}{\sigma_f^2}\right)}{\exp\left(-\frac{E_k^* - E_{k_n}^*}{\sigma_c^2}\right)}\right),$$

where

$$E_k = \|F - F_k\|^2, \qquad E_k^* = \|F - F_k^*\|^2.$$

If E_k^* is strongly correlated with E_k, then the acceptance probability (3.10) could be close to 1 for certain choice of σ_c. Hence a high acceptance rate can be achieved at Step 3 of the preconditioned MCMC method. To demonstrate that E_k^* is indeed strongly correlated with E_k, we compute E_k and E_k^* for many different permeability samples k (see the second example of section 4, Figure 4.7, for details of the permeability field) and plot E_k against E_k^* in Figure 3.1. We find that the correlation coefficient between E_k^* and E_k is approximately 0.9. If the correlation between E_k and E_k^* is strong, we can write

$$E_k \simeq \alpha E_k^* + \beta.$$

Substituting this into (3.10) and choosing $\sigma_c^2 = \sigma_f^2/\alpha$, we can obtain the acceptance rate close to 1 in Step 3. In practice, however, one does not know a priori the correlation constant α. The approximate value of α can be estimated by a priori numerical simulations where E_k and E_k^* are computed for a number of permeability samples.

The preconditioned MCMC method uses the coarse-scale distribution (3.5) with the reference fractional flow being the observed fine-scale fractional flow. One can also use a different reference fractional flow curve in Step 2 of the preconditioned MCMC method to improve the acceptance rate. In our numerical simulations (not presented here), we have used the coarse-scale fractional flow corresponding to the
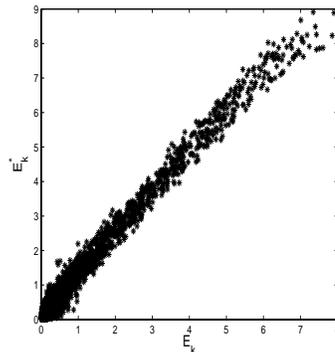
FIG. 3.1. *Cross-plot between* $E_k = \|F - F_k\|^2$ *and* $E_k^* = \|F - F_k^*\|^2$.

observed fractional flow as the reference fractional flow in the preconditioned MCMC simulations. We have observed similar numerical results. Since the coarse-scale fractional flow corresponding to the observed fractional flow is generally not known, we do not present these numerical results here. However, we note that one can possibly improve the preconditioning by a careful choice of the reference fractional flow.

The preconditioned MCMC method employs multiscale finite volume methods in the preconditioning step. If a proposal is accepted by the coarse-scale test (Step 2), one can use the precomputed multiscale basis functions to reconstruct the velocity field on the fine scale. Then the transport equation can be solved on the fine grid coupled with the coarse-grid pressure equation [6, 12, 13, 1]. This approach provides an accurate approximation to the production data on the fine grid and can be used to replace the fine-scale computation in the second stage (Step 3). In this procedure, the basis functions are not updated in time, or updated only in a few coarse blocks. Thus the fine-scale computation in the second stage of the preconditioned MCMC method (Step 3) can also be implemented quickly. Since the basis functions from the first-stage are reused for the fine-scale computation, this combined multiscale approach can be very efficient for our sampling problem.

3.3. Analysis of the preconditioned MCMC method. Next we will analyze the preconditioned MCMC method in more details. Denote

(3.11)
$$\begin{aligned}
\mathcal{E} &= \big\{k; \ \pi(k) > 0\big\}, \\
\mathcal{E}^* &= \big\{k; \ \pi^*(k) > 0\big\}, \\
\mathcal{D} &= \big\{k; \ q(k|k_n) > 0 \text{ for some } k_n \in \mathcal{E}\big\}.
\end{aligned}$$

The set $\mathcal{E}$ is the support of the posterior (target) distribution $\pi(k)$. $\mathcal{E}$ contains all the permeability field k which has a positive probability of being accepted as a sample. Similarly, $\mathcal{E}^*$ is the support of the coarse-scale distribution $\pi^*(k)$, which contains all the k acceptable by the coarse-scale test. $\mathcal{D}$ is the set of all the proposals which can be generated by the instrumental distribution $q(k|k_n)$. For the preconditioned MCMC method to work properly, the conditions $\mathcal{E} \subseteq \mathcal{D}$ and $\mathcal{E} \subseteq \mathcal{E}^*$ must hold (up to a zero measure set) simultaneously. If one of these conditions is not true, say, $\mathcal{E} \nsubseteq \mathcal{E}^*$, then

there will exist a subset $A \subset (\mathcal{E} \setminus \mathcal{E}^*)$ such that

$$\pi(A) = \int_A \pi(k) dk > 0 \qquad \text{and} \qquad \pi^*(A) = \int_A \pi^*(k) dk = 0,$$

which means no element of A can pass the coarse-scale test and A will never be visited by the Markov chain $\{k_n\}$. Thus, $\pi(k)$ cannot be sampled properly.

For most practical proposals $q(k|k_n)$, such as the random walk samplers and independent samplers, the conditions $\mathcal{E}, \mathcal{E}^* \subset \mathcal{D}$ can be naturally satisfied. By choosing the parameter σ_c in $\pi^*(k)$ properly, the condition $\mathcal{E} \subset \mathcal{E}^*$ can also be satisfied (see the discussion below). As a result, we have $\mathcal{E} \subset \mathcal{E}^* \subset \mathcal{D}$. In this case, $\mathcal{E}^*$ is identical to the support of the effective proposal $Q(k|k_n)$:

$$\mathcal{E}^* = \{k; \ Q(k|k_n) > 0 \text{ for some } k_n \in \mathcal{E}\}.$$

Due to the high dimension of the permeability field k, the support $\mathcal{E}$ of the target distribution $\pi(k)$ is much smaller than the support $\mathcal{D}$ of the proposal $q(k|k_n)$ distribution. For all the proposals $k \in (\mathcal{D} \setminus \mathcal{E})$, they will never be accepted as samples in the MCMC method since $\pi(k) = 0$. In the preconditioned MCMC algorithm, the effective proposal distribution $Q(k|k_n)$ samples from a much smaller support $\mathcal{E}^*$, and hence avoids solving the fine-scale problems for all $k \in (\mathcal{D} \setminus \mathcal{E}^*)$. Suppose that we sample the posterior distribution $\pi(k)$ by both the regular MCMC method and preconditioned MCMC method. For each proposal k generated from $q(k|k_n)$, the regular MCMC method accepts it as a sample with probability $p(k_n, k)$ as defined by (3.4), while the preconditioned MCMC method accepts it with probability $g(k_n, k)\rho(k_n, k)$, where $g(k_n, k)$ is the acceptance probability (3.6) of the coarse-scale test and $\rho(k_n, k)$ is the acceptance probability (3.8) of the fine-scale test. When $g(k_n, k) < 1$ and $\rho(k_n, k) < 1$, which is true for most proposals k, it is easy to show that $g(k_n, k)\rho(k_n, k) = p(k_n, k)$. That is, the two methods accept k as an example with the same probability. In numerical experiments, both methods indeed accept approximately the same amount of proposals for a fixed number of iterations. However, the regular MCMC method needs to solve a fine-scale problem for each MCMC iteration, while the preconditioned MCMC method solves only the fine-scale problem when the proposal passes the coarse-scale test. For all the proposals $k \in (\mathcal{D} \setminus \mathcal{E}^*)$, they will be rejected directly by the coarse-scale criteria and do not require fine-scale computations. For each iteration, the preconditioned MCMC solves only the fine-scale problem r time in average, where

$$r = \int_{\mathcal{E}^*} g(k_n, k) q(k|k_n) dk < 1.$$

Note that $\int_{\mathcal{D}} q(k|k_n) dk = 1$ and $g(k_n, k) \leq 1$. If $\mathcal{E}^*$ is close to $\mathcal{E}$, and hence much smaller than $\mathcal{D}$, then $r \ll 1$. Therefore, the preconditioned MCMC method requires much less fine-scale simulation while still accepting approximately the same amount of proposals. In other words, the preconditioned MCMC method can achieve a much higher acceptance rate for each fine-scale computation.

Next we will discuss the stability property of the preconditioned MCMC method. We shall show that the preconditioned MCMC method shares the same convergence property as the regular MCMC method. Denote by K the transition kernel of the Markov chain $\{k_n\}$ generated by the preconditioned MCMC method. Since its effec-

tive proposal is $Q(k|k_n)$ as defined by (3.7), we get

$$
(3.12) \qquad\qquad K(k_n, k) = \rho(k_n, k)Q(k|k_n) \qquad \text{for } k \neq k_n,
$$

$$
(3.13) \qquad\qquad K(k_n, \{k_n\}) = 1 - \int_{k \neq k_n} \rho(k_n, k)Q(k|k_n)dk.
$$

That is, the transition kernel $K(k_n, \cdot)$ is continuous when $k \neq k_n$ and has a positive probability for the event $\{k = k_n\}$.

As in the regular MCMC method, it is easy to show that $K(k_n, k)$ satisfies the detailed balance condition

$$
(3.14) \qquad\qquad \pi(k_n)K(k_n, k) = \pi(k)K(k, k_n)
$$

for any $k, k_n \in \mathcal{E}$. In fact, the equality (3.14) is obviously true when $k = k_n$. If $k \neq k_n$, then from (3.12) we have

$$
\pi(k_n)K(k_n, k) = \pi(k_n)\rho(k_n, k)Q(k|k_n) = \min\Big(Q(k|k_n)\pi(k_n),\ Q(k_n|k)\pi(k)\Big)
$$

$$
= \min\left(\frac{Q(k|k_n)\pi(k_n)}{Q(k_n|k)\pi(k)},\ 1\right)Q(k_n|k)\pi(k) = \rho(k, k_n)Q(k_n|k)\pi(k) = \pi(k)K(k, k_n).
$$

So the detailed balance condition (3.14) is always satisfied. Using (3.14), we can easily show that $\pi(A) = \int K(k, A)dk$ for any $A \in \mathcal{B}(\mathcal{E})$, where $\mathcal{B}(\mathcal{E})$ denotes all the measurable subsets of $\mathcal{E}$. Thus, $\pi(k)$ is indeed the stationary distribution of the transition kernel $K(k_n, k)$.

In the regular MCMC method, the proposal $q(k|k_n)$ is usually chosen to satisfy

$$
(3.15) \qquad\qquad q(k|k_n) > 0 \text{ for any } (k_n, k) \in \mathcal{E} \times \mathcal{E},
$$

which guarantees that the resulting regular MCMC method is irreducible. A similar statement is true for the preconditioned MCMC method.

LEMMA 3.1. *If the proposal distribution $q(k|k_n)$ satisfies (3.15) and $\mathcal{E} \subset \mathcal{E}^*$ holds, then the chain $\{k_n\}$ generated by the preconditioned MCMC method is strongly π-irreducible.*

Proof. According to the definition of strong irreducibility, we need only to show that $K(k_n, A) > 0$ for all $k_n \in \mathcal{E}$ and any measurable set $A \subset \mathcal{E}$ with $\pi(A) > 0$. Note that

$$
K(k_n, A) \geq \int_{A \setminus k_n} K(k_n, k)dk = \int_{A \setminus k_n} \rho(k_n, k)Q(k_n, k)dk
$$

$$
= \int_{A \setminus k_n} \rho(k_n, k)g(k_n, k)q(k|k_n)dk.
$$

In the above inequality, the equal sign holds when $k_n \notin A$. Since $\pi(A) = \int_A \pi(k)dk > 0$, it follows that $m(A) = m(A \setminus k_n) > 0$, where $m(A)$ is the Lebesgue measure. Since $A \subset \mathcal{E}$ and $\mathcal{E} \subset \mathcal{E}^*$, both $\rho(k_n, k)$ and $g(k_n, k)$ are positive for $k \in A$. Combining the positivity assumption (3.15), we can easily conclude that $K(k_n, A) > 0$. $\qquad\square$

Most practical proposal distributions, such as random walk samplers or independent samplers, satisfy the positivity condition (3.15). Thus condition (3.15) poses only a mild restriction in practice. As we will see later, the proposals used in our numerical experiment naturally satisfy the condition (3.15).

Based on the stability property of Markov chains [22, 18], the following convergence result is readily available.

THEOREM 3.2 (see [22]). *Suppose* (3.15) *is true and* $\mathcal{E} \subset \mathcal{E}^*$ *holds; then the preconditioned Markov chain* $\{k_n\}$ *is ergodic: for any function* $h(k)$,

$$(3.16) \qquad \lim_{N \to \infty} \frac{1}{N} \sum_{n=1}^{N} h(k_n) = \int h(k)\pi(k)dk.$$

If the chain $\{k_n\}$ *is also aperiodic, then the distribution of* k_n *converges to* $\pi(k)$ *in the total variation norm*

$$(3.17) \qquad \lim_{n \to \infty} \sup_{A \in \mathcal{B}(\mathcal{E})} \left| K^n(k_0, A) - \pi(A) \right| = 0$$

for any initial state k_0.

To get the convergence property (3.17), we need to show that the Markov chain $\{k_n\}$ generated by the preconditioned MCMC method is aperiodic. Recall that a simple sufficient condition for aperiodicity is that $K(k_n, \{k_n\}) > 0$ for some $k_n \in \mathcal{E}$. In other words, the event $\{k_{n+1} = k_n\}$ happens with a positive probability in the preconditioned MCMC method. From the definition (3.13), we have

$$K(k_n, \{k_n\}) = 1 - \int_{k \neq k_n} \rho(k_n, k)Q(k|k_n)dk = 1 - \int_{k \neq k_n} \rho(k_n, k)g(k_n, k)q(k|k_n)dk.$$

Consequently, $K(k_n, \{k_n\}) \equiv 0$ requires $g(k_n, k) = 1$ and $\rho(k_n, k) = 1$ for almost all $k \in \mathcal{D}$, which means that all the proposals generated by $q(k|k_n)$ are correct samples of distributions $\pi(k)$ and $\pi^*(k)$. This is obviously not true in practice. Thus, the practical preconditioned MCMC method is always aperiodic and converges to the target distribution $\pi(k)$ in the sense of (3.17).

Next we discuss the necessary condition $\mathcal{E} \subseteq \mathcal{E}^*$, which is essential to guarantee the convergence of the preconditioned MCMC method. Due to the Gaussian form of the posterior distribution, $\pi(k)$ and $\pi^*(k)$ do not have a compact support and the domain $\mathcal{E}$ (or $\mathcal{E}^*$) is the whole space spanned by all k. However, if the precision parameters σ_f and σ_c are relatively small, then $\pi(k)$ and $\pi^*(k)$ are very close to zero for most proposals. From the numerical point of view, the proposal k is very unlikely to be accepted if $\pi(k)$ or $\pi^*(k)$ is close to zero. Consequently, the support of the distributions should be interpreted as $\mathcal{E} = \{k; \pi(k) > \delta\}$ and $\mathcal{E}^* = \{k; \pi^*(k) > \delta\}$, where δ is a small positive number.

If $k \in \mathcal{E}$, then $\pi(k) > \delta$ and $\|F_k - F\|^2/\sigma_f^2$ is not very large. To make $k \in \mathcal{E}^*$, $\|F_k^* - F\|^2/\sigma_c^2$ should not be very large either. If $\|F_k^* - F\|^2$ is bounded by $\|F_k - F\|^2$ up to a multiplicative constant, then the condition $\mathcal{E} \subseteq \mathcal{E}^*$ can be satisfied by choosing the parameter σ_c properly. For most upscaled models, the coarse-scale quantity is indeed bounded by the corresponding fine-scale quantity. For example, the upscaled velocity $\boldsymbol{v}^*$ in the saturation equation is obtained by averaging the fine-scale velocity $\boldsymbol{v}$ over the coarse-grid blocks

$$\boldsymbol{v}^*(x) = \sum_i \left(\frac{1}{|\Omega_i|} \int_{\Omega_i} \boldsymbol{v}(y)dy \right) 1_{\Omega_i}(x),$$

where $\Omega_i \subset \Omega$ are the coarse blocks. It is easy to show that

$$
\|\boldsymbol{v}^*\|^2_{L^2(\Omega)} = \sum_i \frac{1}{|\Omega_i|} \left(\int_{\Omega_i} \boldsymbol{v}(y) dy \right)^2
$$

(3.18)

$$
\leq \sum_i \frac{1}{|\Omega_i|} \left(\int_{\Omega_i} 1(y)^2 dy \right) \left(\int_{\Omega_i} \boldsymbol{v}^2(y) dy \right) = \|\boldsymbol{v}\|^2_{L^2(\Omega)}.
$$

Thus, the coarse-scale velocity is bounded by the corresponding fine-scale velocity. We would like to remark that for some nonlinear averaging operators, one can also show that the coarse-scale quantities are bounded by the corresponding fine-scale quantities. One of the examples is the homogenization operator for linear elliptic equations.

In general, it is difficult to carry out such estimates for fractional flows. However, coarse-scale fractional flows can be interpreted as some type of average of the corresponding fine-scale fractional flows. Indeed, the fine-scale fractional flow curve can be regarded as the travel times along the characteristics of the particles that start at the inlet. The coarse-scale fractional flow, on the other hand, represents an average of these travel times over characteristics within the coarse domain. In general, the estimation similar to (3.18) does not hold for fractional flow curves, as our next counterexample shows. For simplicity, we present the counterexample for the single-phase flow in porous media with four layers. This example can be easily generalized. Denote by t_i, $i = 1, 2, 3, 4$, the breakthrough times for the layers. Consider two fine-scale (with four layers) permeability fields with breakthrough times $t_1 = T_1$, $t_2 = T_2$, $t_3 = T_1$, $t_4 = T_2$ and $t_1 = T_1$, $t_2 = T_1$, $t_3 = T_2$, $t_4 = T_2$, respectively. These two fine-scale permeability fields will give the same fractional flows, since the times of the flights are the same up to a permutation. Now we consider the upscaling of these two fine scale permeability fields to two-layered media. Upscaling is equivalent to averaging the breakthrough times. Consequently, the breakthrough times for the corresponding upscaled models are $t_1^* = 0.5(T_1 + T_2)$, $t_2^* = 0.5(T_1 + T_2)$ and $t_1^* = 0.5(T_1 + T_1) = T_1$, $t_2^* = 0.5(T_2 + T_2) = T_2$, respectively. Thus, the coarse-scale models give different fractional flows, even though the fractional flows are identical for the fine-scale models. However, this type of counterexample can be avoided in practice, because the near-well values of the permeability are known, and consequently, permutation of the layers can be avoided.

4. Numerical results. In this section we discuss the implementation details of the preconditioned MCMC method and present some representative numerical results. Suppose the permeability field $k(\boldsymbol{x})$, where $\boldsymbol{x} = (x, z)$, is defined on the unit square $\Omega = [0, 1]^2$. We assume that the permeability field $k(\boldsymbol{x})$ is a log-normal process and its covariance function is known. The observed data include the fractional flow curve F and the values of the permeability at sparse locations. We discretize the domain Ω by a rectangular mesh, and the permeability field k is represented by a matrix (thus k is a high dimensional vector). As for the boundary conditions, we have tested various boundary conditions and observed similar results for the preconditioned MCMC method. In the following numerical experiments we assume $p = 1$ and $S = 1$ on $x = 0$ and $p = 0$ on $x = 1$ and no flow conditions on the lateral boundaries $z = 0$ and $z = 1$. We call this type of boundary condition side-to-side. We have chosen this type of boundary condition because they provide large deviations between coarse- and fine-scale simulations for the permeability fields considered in the paper. The other

type of boundary condition is set by specifying $p = 1$, $S = 1$ along the $x = 0$ edge for $0.5 \leq z \leq 1$ and $p = 0$ along the $x = 1$ edge for $0 \leq z \leq 0.5$. On the rest of the boundaries, no-flow boundary conditions are assumed. We call this type of boundary condition corner-to-corner. We will consider both single-phase and two-phase flow displacements.

Using the Karhunen–Loève expansion [17, 23], the permeability field can be expanded 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. To impose the hard constraints (the values of the permeability at prescribed locations), we will find a linear subspace of the random parameter space (a hyperplane) which yields the desired permeability fields satisfying the hard constraints.

We first briefly recall the basic idea of the Karhunen–Loève expansion. Denote $Y(\boldsymbol{x}, \omega) = \log[k(\boldsymbol{x}, \omega)]$, where the sample variable ω is included to remind us that k is a random field. Suppose $Y(\boldsymbol{x}, \omega)$ is a second order stochastic process; that is, $Y(\boldsymbol{x}, \omega) \in L^2(\Omega)$ with probability 1. Without loss of generality, we assume that $E[Y(\boldsymbol{x}, \omega)] = 0$. Given an arbitrary orthonormal basis $\{\phi_k\}$ in $L^2(\Omega)$, we can expand $Y(\boldsymbol{x}, \omega)$ in Fourier series

$$Y(\boldsymbol{x}, \omega) = \sum_{k=1}^{\infty} Y_k(\omega) \phi_k(\boldsymbol{x}),$$

where

$$Y_k(\omega) = \int_{\Omega} Y(\boldsymbol{x}, \omega) \phi_k(\boldsymbol{x}) d\boldsymbol{x}, \qquad k = 1, 2, \ldots,$$

are random variables with zero means. We are interested in the special L^2 basis $\{\phi_k\}$ which makes Y_k uncorrelated: $E(Y_i Y_j) = 0$ for all $i \neq j$. Denote the covariance function of Y as $R(\boldsymbol{x}, \boldsymbol{y}) = E[Y(\boldsymbol{x})Y(\boldsymbol{y})]$. Then such basis functions $\{\phi_k\}$ satisfy

$$E[Y_i Y_j] = \int_{\Omega} \phi_i(\boldsymbol{x}) d\boldsymbol{x} \int_{\Omega} R(\boldsymbol{x}, \boldsymbol{y}) \phi_j(\boldsymbol{y}) d\boldsymbol{y} = 0, \qquad i \neq j.$$

Since $\{\phi_k\}$ is complete and orthonormal in $L^2(\Omega)$, it follows that $\phi_k(\boldsymbol{x})$ are eigenfunctions of $R(\boldsymbol{x}, \boldsymbol{y})$:

$$(4.1) \qquad \int_{\Omega} R(\boldsymbol{x}, \boldsymbol{y}) \phi_k(\boldsymbol{y}) d\boldsymbol{y} = \lambda_k \phi_k(\boldsymbol{x}), \qquad k = 1, 2, \ldots,$$

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

$$(4.2) \qquad R(\boldsymbol{x}, \boldsymbol{y}) = E[Y(\boldsymbol{x})Y(\boldsymbol{y})] = \sum_{k=1}^{\infty} \lambda_k \phi_k(\boldsymbol{x}) \phi_k(\boldsymbol{y}).$$

Denote $\theta_k = Y_k / \sqrt{\lambda_k}$; then θ_k satisfy $E(\theta_k) = 0$ and $E(\theta_i \theta_j) = \delta_{ij}$. It follows that

$$(4.3) \qquad Y(\boldsymbol{x}, \omega) = \sum_{k=1}^{\infty} \sqrt{\lambda_k} \theta_k(\omega) \phi_k(\boldsymbol{x}),$$

where ϕ_k and λ_k satisfy (4.1). We assume that the eigenvalues λ_k are ordered $\lambda_1 \geq \lambda_2 \geq \cdots$. The expansion (4.3) is called the Karhunen–Loève expansion (KLE) of

the stochastic process $Y(\boldsymbol{x}, \omega)$. For finite discrete processes, the KLE reduces to the principal component decomposition.

In (4.3), the L^2 basis functions $\phi_k(\boldsymbol{x})$ are deterministic and resolve the spatial dependence of the permeability field. The randomness is represented by the scalar random variables θ_k. Generally, we need only to keep the leading order terms (quantified by the magnitude of λ_k) and still capture most of the energy of the stochastic process $Y(\boldsymbol{x}, \omega)$. For a N-term KLE approximation $Y_N = \sum_{k=1}^{N} \sqrt{\lambda_k} \theta_k \phi_k$, we define the energy ratio of the approximation as

$$e(N) := \frac{E\|Y_N\|^2}{E\|Y\|^2} = \frac{\sum_{k=1}^{N} \lambda_k}{\sum_{k=1}^{\infty} \lambda_k}.$$

If $\lambda_k, k = 1, 2, \ldots$, decay very fast, then the truncated KLE would be good approximations of the stochastic process in the L^2 sense.

Suppose the permeability field $k(\boldsymbol{x}, \omega)$ is a log-normal homogeneous stochastic process. Then $Y(\boldsymbol{x}, \omega)$ is a Gaussian process and θ_k are independent standard Gaussian random variables. We assume that the covariance function of $Y(\boldsymbol{x}, \omega)$ has the form

$$(4.4) \qquad R(\boldsymbol{x}, \boldsymbol{y}) = \sigma^2 \exp\left(-\frac{|x_1 - y_1|^2}{2L_1^2} - \frac{|x_2 - y_2|^2}{2L_2^2}\right).$$

In the above formula, L_1 and L_2 are the correlation lengths in each dimension, and $\sigma^2 = E(Y^2)$ is a constant. In our first example, we set $L_1 = 0.2$, $L_2 = 0.2$, and $\sigma^2 = 2$. We first solve the eigenvalue problem (4.1) numerically and obtain the eigenpairs $\{\lambda_k, \phi_k\}$. In Figure 4.1 the first 50 eigenvalues are plotted. As we can see, the eigenvalues of the KLE decay very fast. It has been shown in [8] that the eigenvalues decay exponentially fast for the covariance function (4.4). Therefore, only a small number of terms need to be retained in the truncated expansion (4.3). We can sample $Y(\boldsymbol{x}, \omega)$ easily from the truncated KLE (4.3) by generating independent Gaussian random variables θ_k.

It is worth noting that for a different covariance function such as $R(\boldsymbol{x}, \boldsymbol{y}) = \sigma^2 \exp\left(-\frac{|x_1 - y_1|}{L_1} - \frac{|x_2 - y_2|}{L_2}\right)$, the eigenvalues of the integral equation (4.1) may decay slowly (only algebraically [8]). To achieve the same accuracy, more terms should be retained in the truncated expansion (4.3), which will increase the dimension of the parameter space to represent the permeability. As a result, sampling the permeability from the distribution will be more expensive for both the direct MCMC method and the preconditioned MCMC method. However, small parameter space does not favor the preconditioned MCMC method, and the preconditioning technique is applicable independent of the problem dimension. For permeabilities with higher dimensional parameters, the acceptance rates of the direct MCMC method will be even lower. Consequently, the preconditioned MCMC method will be more preferable since its filtering procedure can increase the acceptance rates dramatically. Note that if the permeability field is not a log-normal process, then θ_k in the expansion (4.3) are not necessarily Gaussian random variables. However, we can still sample the permeability field from the truncated expansion (4.3) by sampling the random variables θ_k.

In the numerical experiments, we first generate a reference permeability field using all eigenvectors and compute the corresponding fractional flows. To propose permeability fields from the prior (unconditioned) distribution, we keep 20 terms in the KLE. Suppose the permeability field is known at eight distinct points. This
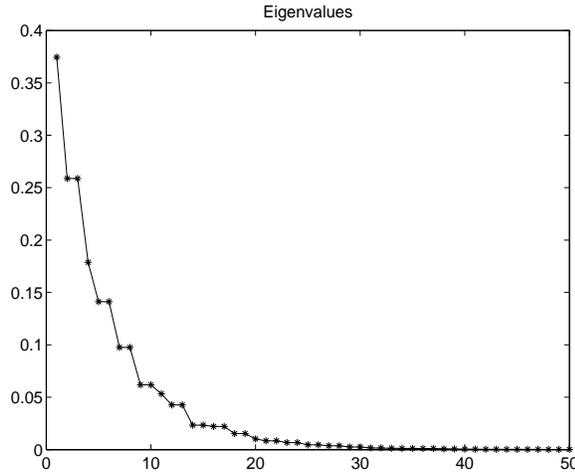
FIG. 4.1. *Eigenvalues of the KLE for the Gaussian covariance with $L_1 = L_2 = 0.2$. The eigenvalues decay very fast.*

condition is imposed by setting

$$(4.5) \qquad \sum_{k=1}^{20} \sqrt{\lambda_k} \theta_k \phi_k(\boldsymbol{x}_j) = \alpha_j,$$

where α_j $(j = 1, \ldots, 8)$ are prescribed constants. For simplicity, we set $\alpha_j = 0$ for all $j = 1, \ldots, 8$. In the simulations we propose 12 θ_i and calculate the rest of θ_i by solving the linear system (4.5). In all the simulations, we test 50000 proposals and iterate the Markov chain 50000 times. Because the direct MCMC computations are expensive, we do not select the large model problems and consider only 40×40 and 60×60 fine-scale models. However, the preconditioned MCMC method is applicable independent of the size of the permeability field.

We have considered two types of instrumental proposal distributions $q(k|k_n)$: the independent sampler and the random walk sampler. In the case of the independent sampler, the proposal distribution $q(k|k_n)$ is chosen to be independent of k_n and equal to the prior (unconditioned) distribution. In the random walk sampler, the proposal distribution depends on the previous value of the permeability field and is given by

$$(4.6) \qquad k = k_n + \epsilon_n,$$

where ϵ_n is a random perturbation with prescribed distribution. If the variance of ϵ_n is chosen to be very large, then the random walk sampler becomes similar to the independent sampler. Although the random walk sampler allows us to accept more realizations, it often gets stuck in the neighborhood of a local maximum of the target distribution. For both proposal distributions, we have observed consistently more than 10 times an increase in the acceptance rate when the preconditioned MCMC method is used.

For the first set of numerical tests, we use a 40×40 fine-scale permeability field and 10×10 coarse-scale models. The permeability field is assumed to be log-normal, with $L_1 = L_2 = 0.2$ and $\sigma^2 = 2$ for the covariance function (4.4). In Figure 4.2, the acceptance rates are plotted against different coarse-scale precisions, σ_c. Here
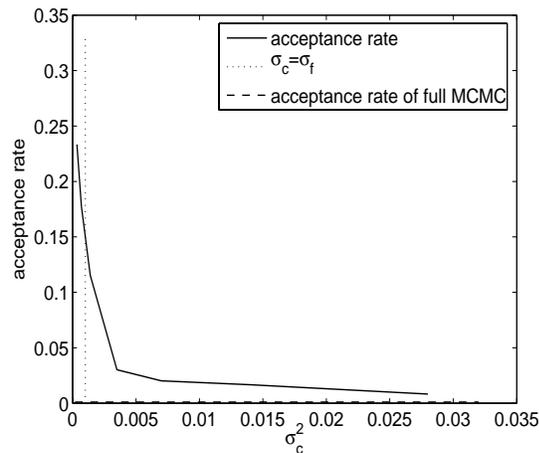
FIG. 4.2. *Acceptance rate versus different coarse-scale precisions for the preconditioned MCMC method. Single-phase flow and $\sigma_f^2 = 0.001$.*

the acceptance rate refers to the ratio between the number of accepted permeability samples and the number of fine-scale simulations that are performed. The acceptance rate for the direct (full) MCMC method is plotted using the dashed line, and it is equal to 0.001. The vertical dotted line marks the coarse-scale precision $\sigma_c = \sigma_f$. If σ_c is very small, then the total number of accepted realizations is also small, even though the acceptance rate is higher. We have found that if σ_c is of the same order as σ_f, then the preconditioned MCMC method accepts almost the same number of proposals as the direct MCMC method but requires only 10% of the fine-scale runs. Note that as σ_c increases the acceptance rate decreases and reaches the acceptance rate of the full MCMC method. Indeed, if σ_c is very large, then all the proposals will be accepted by the coarse-scale test, and there is no gain in preconditioning. In general, one can estimate the optimal σ_c based on a limited number of simulations, prior to the full simulations as described above.

In Figure 4.3 we plot the fractional flows of the accepted permeability realizations. On the left plot, the cross-plot between the reference fractional flow and the sampled fractional flows (of accepted realizations) is plotted. Since the reference fractional flow is the same for every accepted sample, the curve has jumps in the vertical direction. On the right plot, fractional flows of accepted samples are plotted using dotted lines. The bold solid line is the reference fractional flow curve. As we can see from these figures, the fractional flows of accepted realizations are very close to the observed fractional flow, because the precision is taken to be $\sigma_f^2 = 0.001$. In Figure 4.4, we plot the fractional flow error $E_k = \|F - F_k\|^2$ of the accepted samples for both the direct and preconditioned MCMC methods. We observe that the errors of both of the Markov chains converge to a steady state within 20 accepted iterations (corresponds to 20000 proposals). Note that we can assess the convergence of the MCMC methods based on the fractional flow errors. This is a reasonable indicator for the convergence and is frequently used in practice. Given the convergence result of the MCMC method, a longer chain can be easily generated when it is needed. We present a few accepted permeability realizations generated by the preconditioned MCMC method in Figure 4.5. The first plot is the reference (true) permeability field and the others are the last
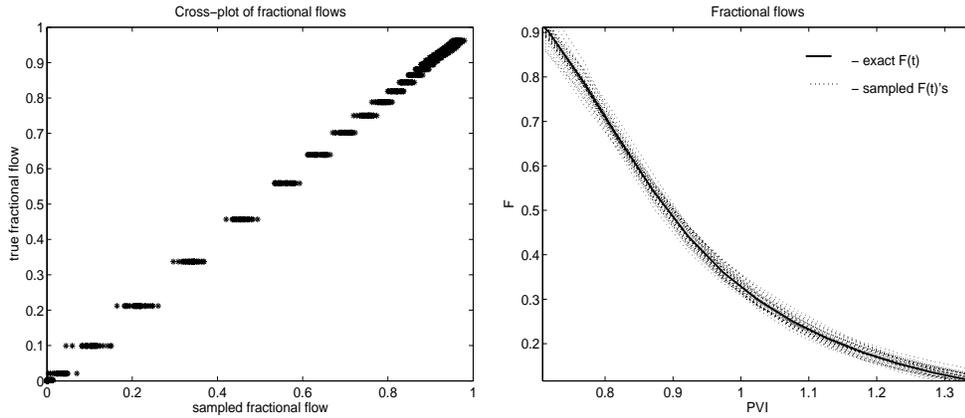
FIG. 4.3. *Fractional flow comparisons. Left: Cross-plot between the reference fractional flow and sampled fractional flows. Right: the solid line designates the fine-scale reference fractional flow, and dotted lines designate fractional flows corresponding to sampled permeability fields.*
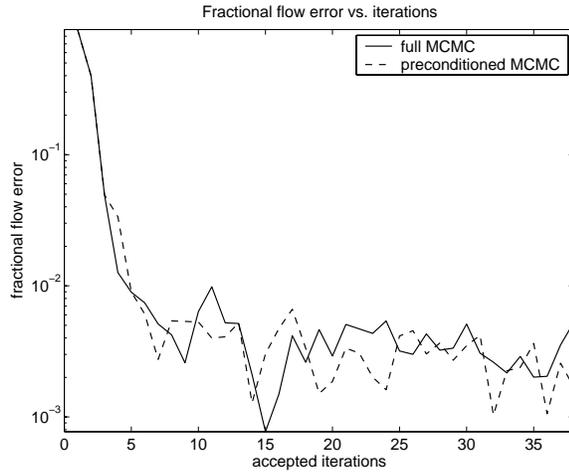

FIG. 4.4. *Fractional flow errors versus accepted iterations.*

five accepted permeability realizations. Some of these realizations closely resemble the true permeability field. Note that the fractional flows of these accepted realizations are in good agreements with the reference (true) fractional flow. One can use these samples for the uncertainty estimation.

For the next set of numerical examples, we consider an anisotropic permeability field with $L_1 = 0.4$, $L_2 = 0.05$, and $\sigma^2 = 2$ defined on a 60×60 fine grid. As in the previous example, we use eight conditioning points and truncate the KLE expansion of the permeability field with 20 terms to maintain a sufficient accuracy. In Figure 4.6, we plot the acceptance rates for 6×6 and 10×10 coarse-scale models against different choices of σ_c^2. The acceptance rate for the direct (full) MCMC method is 0.0008, and it is designated by the dashed line. The acceptance rate is increased by more than 10 times in the preconditioned MCMC method when σ_c is slightly larger than σ_f (the vertical dotted line marks the choice $\sigma_c = \sigma_f$). We also observe a higher acceptance
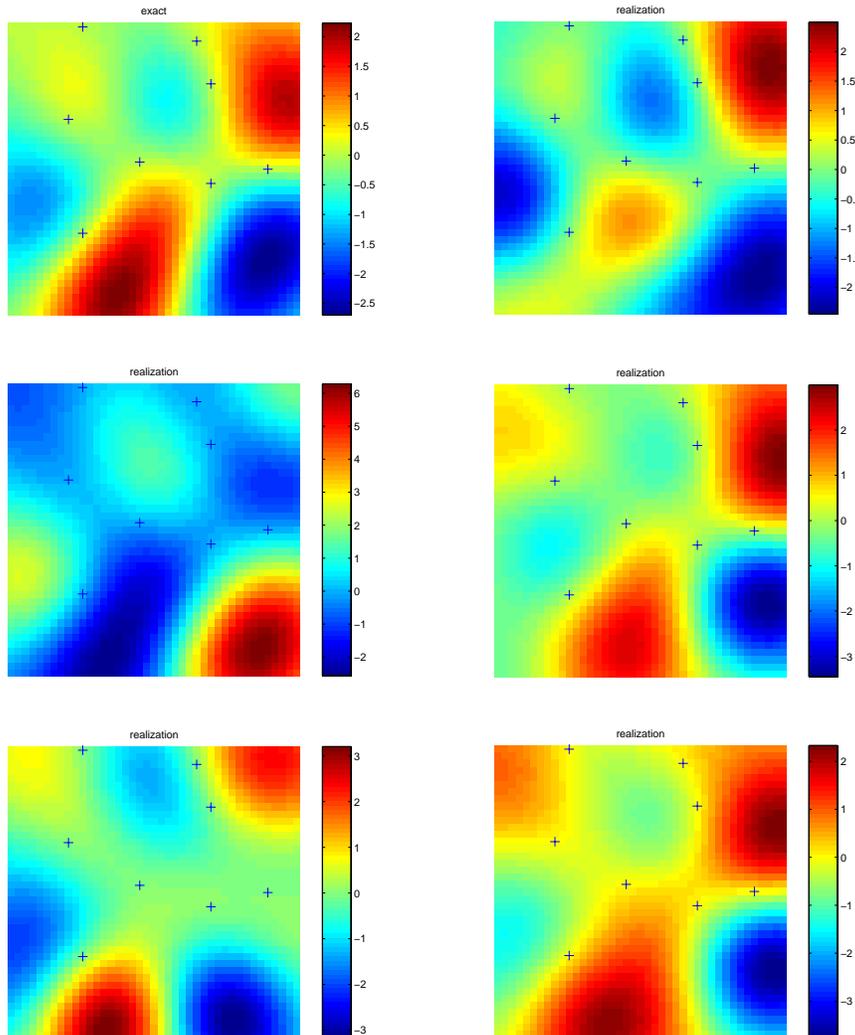
FIG. 4.5. *The last five accepted realizations of the log permeability field. The "+" sign marks the locations of the hard data.*

rate for the 10×10 coarse-scale model than for the 6×6 coarse-scale model. This is because the 10×10 coarse-scale model provides more accurate predictions of the fine-scale results compared to the 6×6 coarse-scale model. As in the previous cases, when the σ_c is slightly larger than σ_f, the preconditioned MCMC method can accept the same number of samples as the underlying full MCMC method but performs only 10% of the fine-scale simulations. Moreover, we have observed that both the direct (full) MCMC method and the preconditioned MCMC method converge to the steady state within 20 accepted iterations, which indicates that both chains have similar convergence properties. In Figure 4.7, we plot the last five accepted samples of the permeability field generated by the preconditioned MCMC method using the 6×6 coarse-scale model. Some of these samples closely resemble the reference (true)
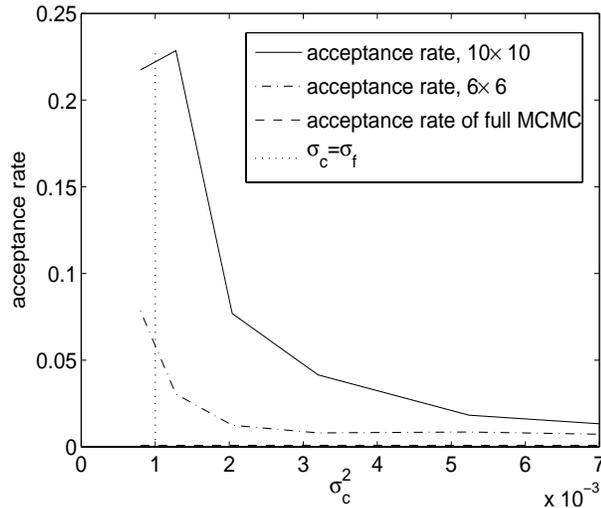
Fig. 4.6. *Acceptance rate versus different coarse-scale precisions of the MCMC method using* 6×6 *and* 10×10 *coarse-scale models. Anisotropic single-phase flow and* $\sigma_f^2 = 0.001$.

permeability field.

Our next set of numerical experiments are for the two-phase flow simulations. We have observed very similar results for two-phase flow simulations, and thus restrict our numerical results to only a few examples. We consider $\mu_w/\mu_o = 5$ and $k_{rw}(S) = S^2$, $k_{ro}(S) = (1 - S)^2$. Typically, one observes similar behavior between the upscaling errors for single- and two-phase flows. We consider a 40×40 fine-scale log-normal permeability field with $L_1 = L_2 = 0.2$ and 10×10 coarse-scale models. In Figure 4.8, the acceptance rate for $\sigma_f^2 = 0.001$ is plotted. As in the case of the single-phase flow simulations, we observe more than 10 times an increase in the acceptance rate. The preconditioned MCMC method accepts the same amount of samples as in the full MCMC method with less than 10% of the fine-scale runs. To study the relative convergence of the preconditioned MCMC method, we plot the fractional flow error for both full and preconditioned MCMC simulations in Figure 4.9. It can be seen from this figure that both the full and preconditioned MCMC methods reach the steady state within 20 accepted iterations. This indicates that the direct and preconditioned MCMC methods have similar convergence properties. The typical samples for the two-phase flow are very similar to those for the single-phase flow, and we do not present them here.

Next we present some numerical results using the random walk sampler (4.6) as the instrumental proposal distribution. The random walk sampler proposes new permeability fields in a neighborhood of the previously accepted permeability field. This improves the acceptance rate in general, though the random walk sampler can get stuck in the neighborhood of the local maxima of the distribution. As a result, the MCMC method will accept a large number of realizations, but it takes a long time for the Markov chain to go from one local maxima to another local maxima. We consider 60×60 fine-scale permeability fields, with $L_1 = 0.4$, $L_2 = 0.05$, and $\sigma^2 = 2$ for the covariance function (4.4). In the preconditioning step, 10×10 and 6×6 coarse-scale models are used. In Figure 4.10, we present the acceptance rates for
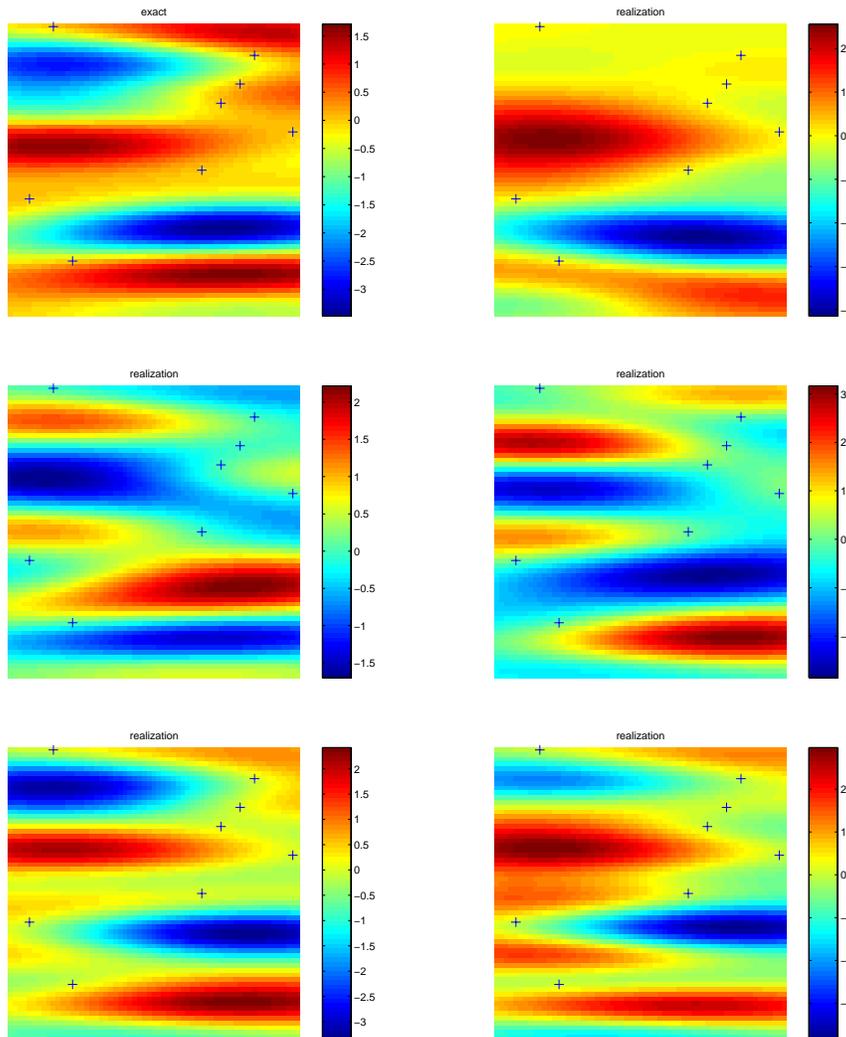
FIG. 4.7. *The last five accepted realizations of the log permeability field for the anisotropic case. The "+" sign marks the locations of the hard data.*

both coarse-scale models when the side-to-side boundary condition is used. In both cases, the acceptance rates are increased several times. In particular, the acceptance rate reaches its peak for σ_c close to σ_f and decreases as σ_c increases. We find that the generated chain k_n has a long correlation length and the nearby accepted permeability realizations are similar to each other. This indicates that the permeability realizations are sampled from a neighborhood of a local maxima, and consequently many proposals are required for a proper sampling. Next we study the convergence of the direct (full) and preconditioned MCMC methods using the random walk sampler (4.6). Figure 4.11 is the plot of the fractional flow errors against accepted iterations. As we can see from this figure, both the full MCMC method and the preconditioned MCMC method converge within 20 accepted iterations.
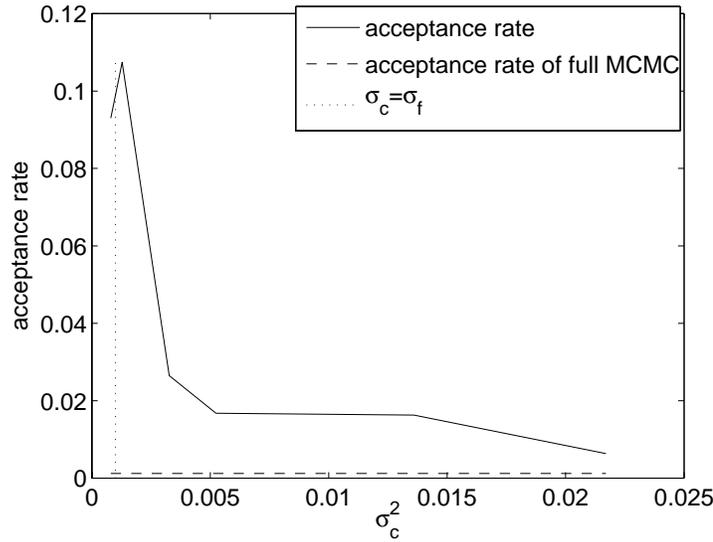
FIG. 4.8. *Acceptance rate versus coarse-scale precision of the MCMC method. Two-phase flow and $\sigma_f^2 = 0.001$.*
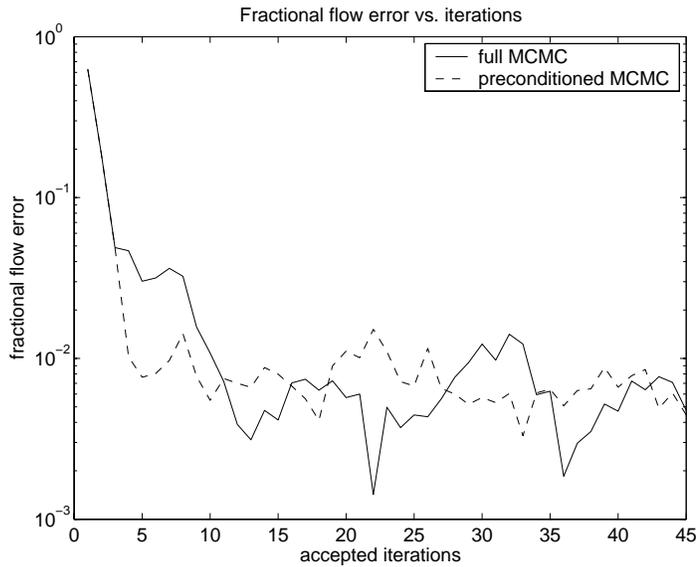

FIG. 4.9. *Fractional flow errors vresus accepted iterations in two phase-flow.*

Finally, we test the preconditioned MCMC method when different boundary conditions are used. In Figure 4.12, we compare the acceptance rates using 10×10 and 6×6 coarse-scale models for the side-to-side and the corner-to-corner boundary conditions. We obtain similar increases of the acceptance rates in the preconditioned MCMC method for the different boundary conditions. We have tested the preconditioned MCMC algorithm with more complicated boundary conditions involving mul-
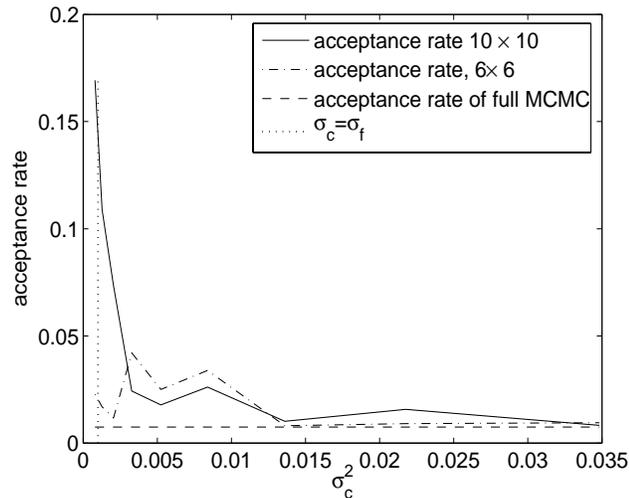
FIG. 4.10. *Acceptance rate versus coarse-scale precision of the MCMC method using 6×6 and 10×10 coarse-scale models. Single-phase flow and $\sigma_f^2 = 0.001$. The random walk sampler is used as the proposal distribution.*

tiple wells (source terms) that arise in petroleum applications (see [5]). In these numerical tests, the single-phase flow upscaling is used (as in [4]) since the multiscale finite volume methods require additional modifications to take into account the well information. The resulting preconditioned MCMC method can increase the acceptance rates by several times. In general, we have found the multiscale finite volume methods to be more accurate for coarse-scale simulations, and they can be further used for efficient and robust fine-scale simulations.

As we mentioned earlier, the full MCMC method and the preconditioned MCMC method accept approximately the same amount of samples for a fixed number of tested proposals. Denote N as the total number of tested proposals; then the direct MCMC method requires exactly N number of fine-scale simulations. Suppose $M < N$ is the number of fine-scale simulations conducted in the preconditioned MCMC method. Denote t_f and t_c as the CPU times for a single fine-scale and coarse-scale forward simulation. Then the computational costs for the direct MCMC method and the preconditioned MCMC method would be Nt_f and $Nt_c + Mt_f$, respectively. Therefore, the CPU cost for the preconditioned MCMC method is only $\frac{t_c}{t_f} + \frac{M}{N}$ that of the direct MCMC method. The coarse-scale computational cost t_c is usually much smaller than the fine-scale computational cost t_f. Suppose the fine-scale model is upscaled 5 times in each direction. Then solving the pressure equation at each time step is about 25 times faster on the coarse grid than on the fine grid. Moreover, the saturation equation is also solved on the *coarse grid* and with *larger time steps*. This makes the overall coarse-scale computations of the two-phase flow equation at least 25 times faster than the fine-scale computations, i.e., $t_c \approx 0.04 t_f$. If the acceptance rate is increased by more than 10 times in the preconditioned MCMC method, as in our numerical experiments, then $\frac{M}{N} < 0.1$, and the overall CPU cost of the preconditioned MCMC method would be only 10% of the CPU costs of the direct MCMC method. Note that using very coarse-scale models (fewer coarse blocks) reduces t_c but increases the
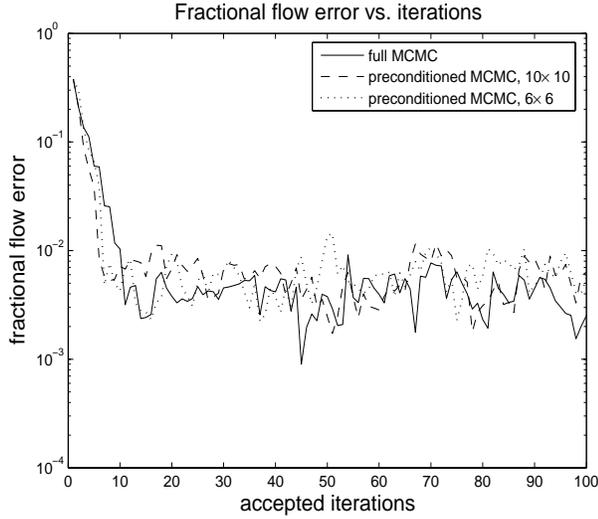
FIG. 4.11. *Fractional flow errors versus accepted iterations for 6×6 and 10×10 coarse-scale models. Single-phase flow and $\sigma_f^2 = 0.001$. The random walk sampler is used as the proposal distribution.*
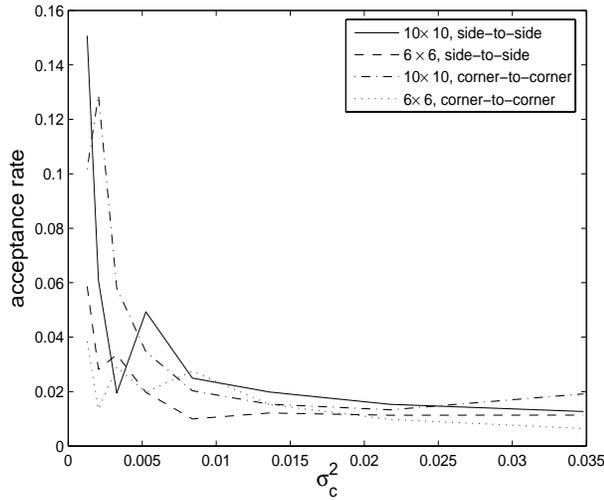

FIG. 4.12. *Acceptance rates of the preconditioned MCMC method using 6×6 and 10×10 coarse-scale models for side-to-side and corner-to-corner boundary conditions. Single-phase flow and $\sigma_f^2 = 0.001$. The random walk sampler is used as the proposal distribution.*

fine-scale run ratio $\frac{M}{N}$. On the other hand, using finer coarse-scale models reduces the ratio $\frac{M}{N}$ but increases t_c. Consequently, a somewhat moderate coarsening (5–10 times coarsening in each direction for large-scale fine models) can provide an optimal choice in the preconditioning of the MCMC simulations.

One can use cruder approximation methods instead of physics-based upscaling methods in preconditioning the MCMC simulations. Next we discuss applying simple

averaging methods in the preconditioned MCMC method. Suppose that the proposal $k(\boldsymbol{x})$ can be represented by the KLE $\log(k(\boldsymbol{x})) = \sum_{k=1}^{n} c_k \phi_k(\boldsymbol{x})$. Denote $\phi_k^*(\boldsymbol{x})$ as the spatial average of $\phi_k(\boldsymbol{x})$ on the coarse grid

$$\phi_k^*(\boldsymbol{x}) = \sum_i \left(\frac{1}{|\Omega_i|} \int_{\Omega_i} \phi_k(\boldsymbol{x}) \, d\boldsymbol{x} \right) 1_{\Omega_i}(\boldsymbol{x}),$$

where Ω_i are the coarse blocks. Then $k^*(\boldsymbol{x}) = \exp(\sum_{k=1}^{n} c_k \phi_k^*(\boldsymbol{x}))$ is a coarse-scale approximation of $k(\boldsymbol{x})$. We can use $k^*(\boldsymbol{x})$ in the coarse-scale simulations to determine whether or not to run the fine-scale simulations. We would like to note that this type of averaging is less expensive compared to the upscaling method used in the paper because it involves only volume average and it is performed only once. However, in general this type of averaging does not represent the correct average flow properties, and consequently the strong correlation between the fine-scale and coarse-scale quantities is not guaranteed. Our numerical results show that using simple averaging methods, such as the one presented here, can give an incorrect sampling. We have observed that averaging the KLE eigenfunctions leads to more uniform permeability fields. Consequently, the first stage of the preconditioned MCMC method restricts the proposal permeability to the more uniform fields and leads to incorrect sampling of the multimodal target distribution.

Finally, we would like to point out that the coarse-scale approximation techniques can also be efficiently used for other instrumental distributions. In our recent work [3], we have used coarse-scale approximations based on the multiscale finite volume methods in Langevin MCMC algorithms. In the Langevin MCMC algorithms, the gradient of the posterior distribution is used in the instrumental proposal distribution. The computation of the gradient of the posterior distribution is very expensive. We have employed the coarse-scale model in approximating the gradient and used a two-stage MCMC method in filtering these proposals. We have shown that one can achieve the acceptance rate comparable to the fine-scale Langevin MCMC with much less CPU time.

5. Conclusion. In this paper, we study the preconditioning of MCMC simulations using inexpensive coarse-scale runs in inverse problems related to subsurface characterization. For each MCMC proposal, a coarse-scale simulation is performed to decide whether or not to run the fine-scale simulations. The coarse scale simulation, which is based on the multiscale finite volume methods, filters unlikely acceptable proposals and avoids expensive fine-scale simulations for them. The filtering process takes into account the coarse-scale information of the problem and modifies the Markov chain generated by the MCMC method. We formulate the conditions which guarantee that the modified chain will converge to the correct posterior distribution. We also discuss the applicability of these conditions to the commonly used upscaling methods. Numerical examples show that we can achieve more than 10 times an increase in the acceptance rate if the MCMC simulations are preconditioned using coarse-scale models. The sampled realizations of the permeability field can be used in uncertainty quantification.

Appendix A. Coarse-scale models using multiscale finite volume methods. In this appendix, we discuss the details of the upscaled model used in the paper. The key idea of the method is the construction of the finite element basis functions on the coarse grids, such that these basis functions capture the small scale information on each coarse block. The method we will use follows its finite element counterpart

presented in [11]. The basis functions are constructed from the solution of the leading order homogeneous elliptic equation on each coarse element with carefully chosen boundary conditions. For a coarse element K with d vertices, the local basis functions $\phi^i, i = 1, \ldots, d$, satisfy the following elliptic problem:

$$-\nabla \cdot (k \cdot \nabla \phi^i) = 0 \quad \text{in } K,$$

(A.1)

$$\phi^i = g^i \quad \text{on } \partial K$$

for some functions g^i defined on the boundary of the coarse element K. Hou and Wu [11] have demonstrated that a careful choice of the boundary condition would guarantee that the basis functions capture the local information of the solution, and hence improve the accuracy of the method. The function g^i for each i varies linearly along ∂K. Thus, ϕ^i will reduce to a standard linear/bilinear basis function for a constant diagonal tensor. Note that as usual we require $\phi^i(\xi_j) = \delta_{ij}$. Finally, a nodal basis function associated with each vertex ξ is constructed from the combination of the local basis functions that share this ξ. These nodal basis functions are denoted by $\{\psi_\xi\}_{\xi \in Z_h^0}$.

Denote by V^h the space for the approximate pressure solution which is spanned by the basis functions $\{\psi_\xi\}_{\xi \in Z_h^0}$. Based on (2.2), a statement of mass conservation is formed on each control volume V_ξ, where the approximate solution is expressed as a linear combination of the basis functions. Assembly of this mass conservation statement on all control volumes would give rise to a linear system of equations that can be solved accordingly. The resulting linear system has incorporated the fine-scale information through the involvement of the nodal basis functions on the approximate solution. To be more specific, the problem now is to seek $p^h \in V^h$ with $p^h = \sum_{\xi \in Z_h^0} p_\xi \psi_\xi$ such that

$$\text{(A.2)} \qquad \int_{\partial V_\xi} \lambda(S) k \cdot \nabla p^h \cdot \vec{n} \, dl = \int_{V_\xi} f \, dA$$

for every control volume $V_\xi \subset \Omega$. Here $\vec{n}$ denotes the unit normal vector on the boundary ∂V_ξ of the control volume, and S is the fine-scale saturation field at this point. We note that concerning the basis functions, a vertex-centered finite volume difference is used to solve (A.1).

Once the pressure solution is available, it can be used to compute the total velocity field at the coarse-scale level, denoted by $\overline{\boldsymbol{v}} = (\overline{v}_x, \overline{v}_z)$ via (2.5). In general, the following formulas are used to compute the velocities in the horizontal and vertical directions, respectively:

$$\text{(A.3)} \qquad \overline{v}_x = -\frac{1}{h_z} \sum_{\xi \in Z_h^0} p_\xi \left(\int_E \lambda(S) k_x \frac{\partial \psi_\xi}{\partial x} dz \right),$$

$$\text{(A.4)} \qquad \overline{v}_z = -\frac{1}{h_x} \sum_{\xi \in Z_h^0} p_\xi \left(\int_E \lambda(S) k_z \frac{\partial \psi_\xi}{\partial z} dx \right),$$

where E is the edge of V_ξ. Furthermore, for the control volumes V_ξ adjacent to the Dirichlet boundary (which are half control volumes), we can derive the velocity approximation using the conservation statement derived from (2.2) on V_ξ. One of the terms involved is the integration along part of the Dirichlet boundary, while the other

three terms are known from the adjacent internal control volumes calculations. The analysis of the two-scale finite volume method can be found in [7].

As for the upscaling of the saturation equation, we use the coarse-scale velocity to update the saturation field on the coarse grid, i.e.,

$$
(A.5) \qquad \frac{\partial \overline{S}}{\partial t} + \overline{\boldsymbol{v}} \cdot \nabla f(\overline{S}) = 0,
$$

where $\overline{S}$ denotes the saturation on the coarse grid. In this case the upscaling of the saturation equation does not take into account the subgrid effects. As we mentioned above, one can reconstruct the velocity field and solve the saturation equation on the fine grid. The latter, though more expensive, provides an accurate approximation of the production data.

Acknowledgments. The authors would like to thank the referees for valuable comments and suggestions and Victor Ginting for his help in preparing this manuscript.

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 Model. Simul., 2 (2004), pp. 421–439.

[2] J. A. CHRISTEN AND C. FOX, *Markov Chain Monte Carlo using an approximation*, J. Comput. Graph. Statist., 14 (2005), pp. 795–810.

[3] P. DOSTERT, Y. EFENDIEV, T. HOU, AND W. LUO, *Coarse-gradient Langevin algorithms for dynamic data integration and uncertainty quantification*, J. Comput. Phys., to appear.

[4] L. J. DURLOFSKY, *Numerical calculation of equivalent grid block permeability tensors for heterogeneous porous media*, Water Resour. Res., 27 (1991), pp. 699–708.

[5] Y. EFENDIEV, A. DATTA-GUPTA, V. GINTING, X. MA, AND B. MALLICK, *An efficient two-stage Markov chain Monte Carlo method for dynamic data integration*, Water Resour. Res., W12423.

[6] Y. EFENDIEV, V. GINTING, T. HOU, AND R. EWING, *Accurate multiscale finite element methods for two-phase flow simulations*, J. Comput. Phys., to appear.

[7] V. GINTING, *Analysis of two-scale finite volume element method for elliptic problem*, J. Numer. Math., 12 (2004), pp. 119–142.

[8] P. FRAUENFELDER, C. SCHWAB, AND R. A. TODOR, *Finite elements for elliptic problems with stochastic coefficients*, Comput. Methods Appl. Mech. Engrg., 194 (2005), pp. 205–228.

[9] J. GLIMM AND D. H. SHARP, *Prediction and the quantification of uncertainty*, Phys. D, 133 (1999), pp. 152–170.

[10] D. HIGDON, H. LEE, AND Z. BI, *A Bayesian approach to characterizing uncertainty in inverse problems using coarse and fine-scale information*, IEEE Trans. Signal Process., 50 (2002), pp. 388-399.

[11] T. Y. HOU AND X. H. WU, *A multiscale finite element method for elliptic problems in composite materials and porous media*, J. Comput. Phys., 134 (1997), pp. 169–189.

[12] P. JENNY, S. H. LEE, AND H. TCHELEPI, *Multi-scale finite volume method for elliptic problems in subsurface flow simulation*, J. Comput. Phys., 187 (2003), pp. 47–67.

[13] P. JENNY, S. H. LEE, AND H. A. TCHELEPI, *Adaptive multiscale finite-volume method for multiphase flow and transport in porous media*, Multiscale Model. Simul., 3 (2005), pp. 50–64.

[14] P. KITANIDIS, *Quasi-linear geostatistical theory for inversing*, Water Resour. Res., 31 (1995), pp. 2411–2419.

[15] J. S. LIU, *Monte Carlo Strategies in Scientific Computing*, Springer, New York, 2001.

[16] O. LODOEN, H. OMRE, L. DURLOFSKY, AND Y. CHEN, *Assessment of uncertainty in reservoir production forecasts using upscaled models*, in Proceedings of the Seventh International Geostatistics Congress, Banff, AB, Canada, 2004.

[17] M. LOEVE, *Probability Theory*, 4th ed., Springer, Berlin, 1977.

[18] S. P. MEYN AND R. L. TWEEDIE, *Markov Chains and Stochastic Stability*, Springer, London, 1996.

[19] D. OLIVER, L. CUNHA, AND A. REYNOLDS, *Markov chain Monte Carlo methods for conditioning a permeability field to pressure data*, Math. Geol., 29 (1997), pp. 61–91.

[20] D. OLIVER, N. HE, AND A. REYNOLDS, *Conditioning permeability fields to pressure data*, in Proceedings of the 5th European Conference on the Mathematics of Oil Recovery, Leoben, Austria, 1996.

[21] H. OMRE AND O. P. LODOEN, *Improved production forecasts and history matching using approximate fluid-flow simulators*, SPE Journal, (2004), pp. 339–351.

[22] C. ROBERT AND G. CASELLA, *Monte Carlo Statistical Methods*, Springer, New York, 1999.

[23] E. WONG, *Stochastic Processes in Information and Dynamical Systems*, McGraw–Hill, New York, 1971.